

头戴式光场显示技术研究进展

华宏*

3D Visualization and Imaging Systems Laboratory, Jame C. Wyant College of Optical Sciences, The University of Arizona,
Tucson 85721, Arizona, USA

摘要 光场显示器旨在通过重建三维场景在不同方向发出的几何光线来渲染三维场景的视觉感知,从而为人的视觉系统提供自然舒适的视觉体验,解决传统平面立体三维显示器中的聚散调节冲突问题。近年来,多种光场显示方法被尝试应用到头戴式显示技术中。本文对头戴式光场显示器的最新发展进行全面概述。

关键词 头戴显示器; 光场显示; 虚拟现实; 增强现实

中图分类号 TH74 **文献标志码** A

DOI: 10.3788/AOS230959

1 引言

人类视觉系统利用各种深度信息来感知三维空间和形状。这些深度信息可以分为心理和生理两类。心理类深度信息包括透视、遮挡、空中透视、阴影、表观大小、纹理梯度,而生理类深度信息包括双眼视差、运动视差、视网膜模糊效应、视距调节和聚散度。其中,视距调节信息是指眼睛的聚焦动作,睫状肌改变晶状体的屈光力并最小化景物注视深度的模糊量。与眼睛视距调节相关的是视网膜图像模糊效应,由于眼睛的景深有限,视距调节深度大的物体在视网膜的成像较模糊,而图像模糊的程度随着物体聚焦深度变化。视距调节和视网膜图像模糊效应一起被称为聚焦信号。聚散度是指双眼视轴协同运动,使视轴向内或向外在近距离或远距离处与感兴趣的三维物体相交的能力。

多年来,人们已经提出许多不同的三维显示技术,其中大部分是基于立体显示原理。传统的立体三维显示器向每只眼睛分别呈现一幅二维透视图像,它们包含从两个略微不同的位置观看三维场景所产生的双眼视差和其他图形深度信息。虽然立体显示技术可以渲染令人信服的深度感知,但这类显示器无法为三维场景呈现正确的聚焦视觉信息,包括视距调节和视网膜模糊效果,因此会受到聚散调节冲突(VAC)问题的影响。VAC可能表现为眼睛的视距调节深度和聚散度不匹配,或者渲染的三维场景和真实世界三维场景的视网膜模糊效果不匹配。这些矛盾通常是与观看立体三维显示器相关的各种视觉伪影的关键促成因素,例如扭曲的深度感知和视觉不适。

近年来,人们已经提出了几种可能解决VAC问题

的显示方法,包括全息显示、体积显示、变焦或多焦平面显示以及光场显示^[1]。在这些不同的显示方法中,三维光场显示技术旨在通过再现三维场景中每个点发出的几何光线来实现对三维场景的感知。几何光线的重建方式可以基于放置在不同深度的一叠离散焦平面,也可以基于对不同眼位观察的光线进行角度采样而得到的一组数据阵列。光场显示技术被认为是最有前途的三维显示技术之一,它能够解决VAC问题,同时提供了极强的灵活性和可扩展性。

在过去的几十年中,学术界已经探索出许多不同的方法来渲染光场效果并实现了不同的保真度。特别是近年来,学术界针对头戴式显示器(HMD)探索了多种光场显示方式。本文着重概述头戴式光场显示器的架构方法和最新研究进展。

2 头戴式光场显示方法综述

如图1(a)所示,自然的三维场景在宽光谱范围内的所有可能方向上发射或反射光线,并且可能随时间动态变化,这些光线信息被称为光场。人类观察者通过对从不同方向和不同位置进入眼睛瞳孔的光线进行采样来观看自然三维场景。接收的光场随观察者相对于三维场景的空间位置而变化,感知到的视网膜图像取决于观察者眼睛的视距调节状态。例如,如果眼睛聚焦在较远的深度:使魔方位于清晰的焦点上,则沙滩球会出现模糊,如图1(b)所示;使沙滩球位于清晰的焦点上,则魔方会出现模糊,如图1(c)所示。一般来说,每条光线的辐射亮度可以表示为光线角度 (θ, ϕ) 、波长 (λ) 、时间 (t) 和观察位置 (V_x, V_y, V_z) 的函数,这些参数共同描述为七维全光场函数^[2]。

收稿日期: 2023-05-10; 修回日期: 2023-06-22; 录用日期: 2023-07-11; 网络首发日期: 2023-07-20

通信作者: *hhua@optics.arizona.edu

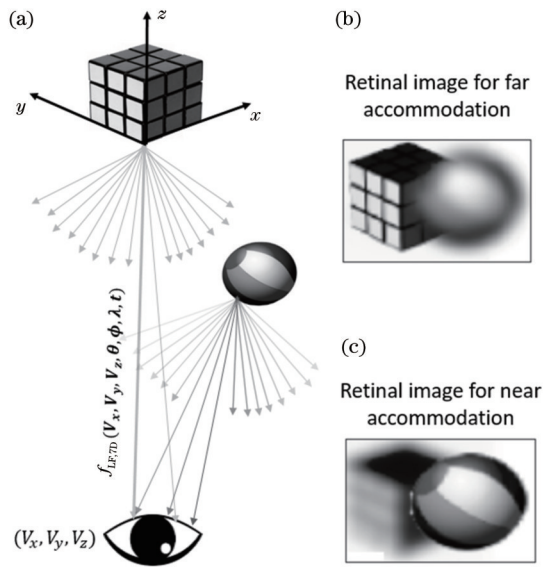


图 1 自然三维场景的光场示意以及光场函数表达。(a)光场示意图;(b)远聚焦时的视网膜成像;(c)近聚焦时的视网膜成像
Fig. 1 Light field of a natural 3D scene and 7D plenoptic function. (a) Schematic of light field; (b) retinal image for far accommodation; (c) retinal image for near accommodation

现有的光场显示方法通常采用不同的采样机制对全光场函数的参数进行简化。实际系统的设计通常伴随着各个指标的权衡。在依据特定应用场景进行系统指标权衡时,一个重要的属性为显示器是用于多用户

“直接观看”,还是仅限于单个用户头戴式(近眼)观看。

在头戴式光场显示中,显示源和光学元件被佩戴在观看者的头部。三维场景的光场可以通过沿视轴放置在不同深度的一组离散焦平面进行采样,相应技术被称为多焦平面头戴式显示技术^[3-5];或者通过光学阵列对由三维场景发出的不同角度方向的光线进行采样,相应技术被称为基于集成成像的光场显示技术^[6];或通过多层空间光调制器计算重建光场函数,相应技术被称为计算多层式光场显示技术^[7]。

在多焦平面头戴式光场显示中,三维场景的光场由一组离散焦平面采样,这些焦平面沿视轴放置在不同深度,如图 2 所示。当焦平面足够密集时,三维场景的每个空间位置都由相应平面上的像素唯一采样,将每个像素称为体素,这种配置称为体积显示,可以从大范围的观察点进行观察;当焦平面稀疏且相邻焦平面间隔较大时,在空间内具有一定深度的三维物体被这些焦平面分成多个焦区,每个焦平面从固定视点渲染以焦平面的焦点为中心、相应焦深为半径的三维子体积的二维投影。将这些二维投影叠加起来可重建出从给定观察点看到的场景的光场。在这种配置中,用于重建三维场景的光线是从同一视点采样的,光线位置在深度上有所不同,因此这种配置被称为固定视点体积显示。这种简化的采样机制的好处是使用少量的焦平面就能以相对较低的深度分辨率为代价渲染具有高空间分辨率的大深度体积。

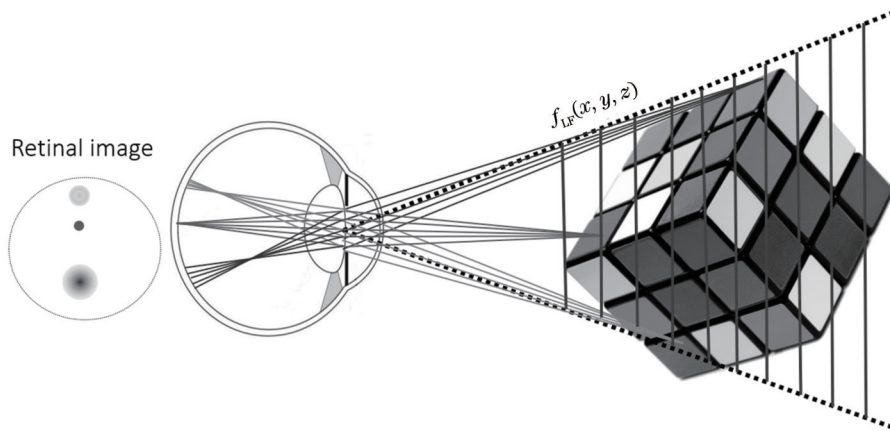


图 2 离散多焦面光场采样示意
Fig. 2 Illustration of light field sampling through a stack of discrete planes

在基于集成成像原理的头戴式光场显示中,利用二维光学阵列元件(例如微透镜阵列或针孔阵列)对三维场景发出的几何光线进行角度采样,从而重建三维场景的光场。如图 3 所示,一个简单的基于集成成像的光场显示系统由一个微型显示面板和一个直接放置在观看者眼前的微透镜阵列(MLA)组成。显示器呈现一组二维图像阵列,图像单元代表三维场景的不同视角。重建的三维场景的光场被直接耦合到观看者的眼睛中,因此观看者的眼睛接收光场的视窗被限制在

一个小区域。

计算多层头戴式光场显示的原理与压缩显示器相同,它通过均匀或定向背光照明的多层像素阵列对定向光线进行采样^[7]。如图 4 所示,三维场景的光场通过算法被分解为多个衰减掩模,每个掩模表示每层光衰减器的透射率或反射率。从背光源进入眼睛的每条光线的强度值是光线在各衰减层的相交处像素值的乘积。由于重建光线具有准直特性,计算多层显示器通过定向光线的集成来渲染三维场景的光场。

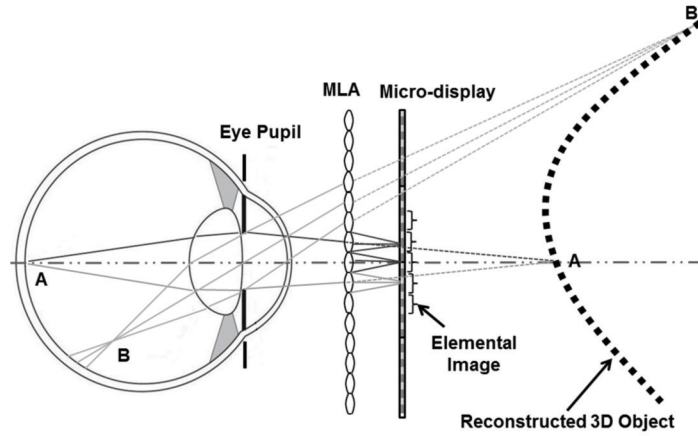


图 3 通过光学元件阵列采样光场示意图

Fig. 3 Schematic illustration of light field sampling through an array of optical elements

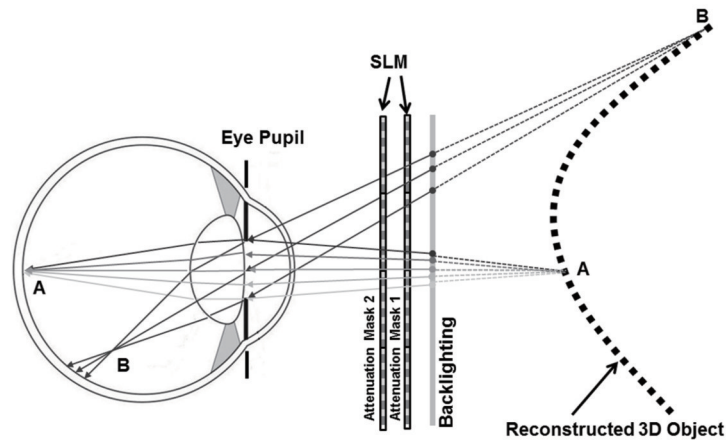


图 4 计算多层头戴式光场显示系统(LF-HMD)采样示意图

Fig. 4 Schematic illustration of computational multi-layer LF-HMD

3 基于多焦面的光场显示技术

基于多焦面的光场显示技术始于空间复用方法。Marran 和 Schor^[8]提出使用双焦镜头为头戴式显示系统中视野的不同部分创建不同的焦深。Akeley 等^[5]展示了第一个基于空间复用三焦平面显示器样机,相邻焦平面的间距为 $0.67 D$ ($1 D = 1 \text{ m}^{-1}$)。该样机通过放置在与观察者不同距离处的 3 个分束器将平板显示器分成 3 个焦平面,从而覆盖 $0.311 \sim 0.536 \text{ m}$ 的固定深度范围。2010 年, Schowengerdt 等^[9]利用光纤阵列制备出多焦点光束的空间复用视网膜扫描显示器。2013 年, Cheng 等^[10]通过堆叠两个自由棱镜设计了一个空间复用的双焦平面系统。

通常情况下,空间复用多焦面方法允许并行渲染多个焦平面并降低对显示技术的速度要求。此外,由于缺乏具有高透射率的光调制器,以及同时渲染多层二维图像的计算能力需求,空间复用多焦面方法的实施仍具有挑战性。由于光路复杂性较高,且通过多层光调制器的通光效率较低,实现大量焦平面的难度较大。该方法的另一个缺点是改变焦平面位置和间距的

灵活性较低。

空间多路复用方案的替代方法是使用变焦元件(VFE)的时间多路复用方法,其中不同深度的三维物体按其在二维焦平面的顺序进行渲染,VFE 元件的焦距随着物体的深度同步调整。如果能够把所有物体都以无闪烁的高频进行显示,这些物体就可以同时为人眼提供正确的聚焦深度信号。在过去的十多年中,曾出现过多个时分复用多焦面样机。McQuaide 等^[4]提出了双焦平面视网膜扫描显示器,其中一种方法是通过光栅扫描的方式把调制激光束逐个像素地生成二维图像,每个像素的聚焦深度信号则通过一个可变形膜反射镜装置使激光束散焦。这种方法可以通过调制 3 个 RGB 激光器的强度来描绘全彩色像素。Love 等^[11]展示了另一种样机设计,该设计以双折射透镜为 VFE,并以高刷新率阴极射线管(CRT)为图像源来生成 4 个固定焦平面。Liu 和 Hua^[12]使用液体镜头作为 VFE,OLED 微显示器作为图像源,展示了第一个双焦平面光学透视增强现实(AR)显示器样机,它可以保持对现实世界的无障碍透视。在该设计中,前后焦平面分别位于距眼睛 $5 D$ 和 $1 D$ 的深度处。图 5 展示了

相机从显示器的出瞳处拍摄的两张照片,比较了 5 D 和 1 D 下的相机对焦。虚拟物体显示出与物理参照物相当的模糊信号,该模糊程度随物体深度的变化而变化。近年来,基于时间复用原理的多焦面显示技术取得了不少研究成果。Chang 等^[13]根据 VFE 虽然需要较长的稳定时间却具有 1000 Hz 频率响应的特性,通过三角波驱动 VFE 的方式使其快速扫描整个焦距

范围,从而实现每秒渲染多达 1600 个焦平面的效果。Rathinavel 等^[14]采用类似的三角波驱动方式,并结合将真彩色图像根据深度分解成 280 个不同深度的二进制图像的渲染方法,实现多达 280 个焦平面的真彩色三维显示。Zhang 等^[15]基于偏振光复用原理和双极性立体超表面技术实现了一个多达 20 个焦平面的多焦面三维显示系统。

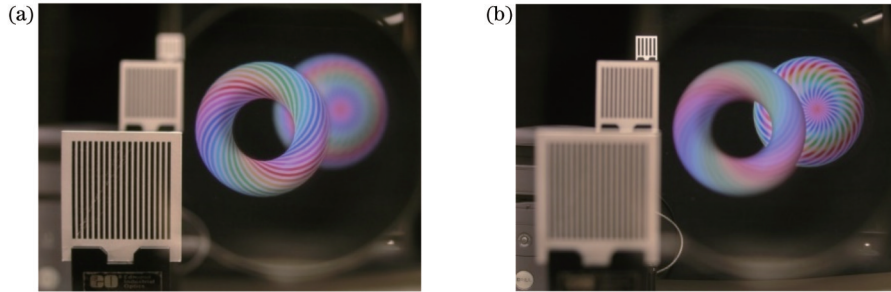


图 5 通过一个双焦面光学透视式头盔样机拍摄的视图^[12]。(a)拍摄相机聚焦在位于 5 D 的近端物体;(b)拍摄相机聚焦在位于 1 D 的远处物体

Fig. 5 View examples captured through a dual-focal plane optical see-through HMD prototype^[12]. (a) Camera focused at the front resolution target located at 5 D distance; (b) camera focused at the back resolution target located at 1 D distance

对于空间复用和时间复用方法来说,大量的焦平面和小的屈光间距对于创建具有准确聚焦信号和高图像质量的大场景体积非常重要。Rolland 等^[3]对 MFP-HMD 的工程要求进行了深入分析,所得出的结论是:需要 14 个等距(屈光间距)焦平面来覆盖从无限远到 0.5 m 的聚焦范围;如果要实现 1' 的角度分辨率,则需要多达 28 个焦平面才能覆盖从无限远到 4 D 的深度范围。然而,在当前的技术条件下,实现这么多的焦平面实际上非常具有挑战性。在不显著影响显示图像质量和渲染焦点信号准确性的情况下减少焦平面的数量则具有非常重要的实际意义。

为了在获得良好的图像质量的前提下显著地减少焦平面的数量,可通过结合深度融合三维(DFD)感知^[16]的原理进一步改进多焦平面方法。这是因为人眼会把显示在不同深度的两幅重叠图像融合感知为一幅单一深度的图像。可以通过深度加权融合函数来调节两幅图像之间的亮度比,进而优化融合图像的对比度和梯度变化,调节融合图像的感知深度,并为人眼提供有效的聚焦调节信号。这种结合深度融合效应的多焦平面显示器被称为深度融合多焦平面显示器(DFD-MFP 显示器)^[17-19]。这种方法将必要的焦平面数量显著减少到可承受的水平,以便在大深度范围内为观者呈现正确或接近正确的聚焦信号^[17-19]。

相邻焦平面之间的最佳距离以及最佳融合函数取决于多个因素。Akeley 等^[5]建议在相邻的焦平面上使用一对线性深度混合函数来减轻由焦平面的离散采样引起的图像伪影。为了获得正确或几乎正确的聚焦信号并最大限度地降低 VAC,在选择融合函数时需要考虑几个关键因素,例如对焦点信号错误的容忍度、焦平

面间距、显示特性和眼模型参数。Liu 和 Hua^[17]以及 Hu 和 Hua^[18-19]提出了 DFD-MFP 系统设计的优化框架,该方法通过优化融合的视网膜图像的对比度变化幅度和梯度来选择最优的焦平面间距和融合函数的目的,有助于刺激或稳定眼睛聚焦调节反应。为了证明 DFD-MFP 方法及其融合功能的有效性,他们进一步设计并构建了一个具有 6 个焦平面的 OST-HMD 样机^[18-19]。

Hu 和 Hua^[18-19]提出的深度加权多焦面融合方法,专注于优化视网膜图像质量,并在通过几个具有均匀间距的离散焦平面渲染连续三维场景体积时最大限度地降低聚焦信号冲突。Narain 等^[20]提出一种替代的深度加权混合方案,通过计算优化亮度权重来准确地再现遮挡、反射和其他非局部效应的散焦行为作为调节的函数。Wu 等^[21]分析了根据要渲染的三维内容的特性优化和调整有限数量焦平面的动态配置的优势,他们优化了一个目标函数,该函数表征了渲染三维场景的整体感知质量——多焦点场景散焦质量。他们建议,通过显式减少渲染图像体积内聚合适应状态的整体对比度损失,可以显著提高内容的感知视觉质量。

总的来说,以上提及的样机设计充分展示了 DFD-MFP 显示方法具有呈现几乎正确的聚焦信号的能力以及避免调节-收敛冲突的潜力。然而,该技术在成为真正可穿戴光场 AR 显示器的可行解决方案方面存在几个关键技术障碍:第一个主要障碍是技术小型化。由于高速显示器和变焦元件等支持技术的限制,目前的样机是以台式机的形式实现的,比如 Liu 和 Hua^[17]、Hu 和 Hua^[19]展示的样机占据了近 500 mm×

300 mm×100 mm 工作台空间的体积。第二个主要障碍是实时渲染和显示。受高速显示技术和显示计算机接口可用性的限制,大多数样机无法以比标准单帧显示速度快数倍的速度渲染和刷新多帧高分辨率全彩图像。相反,待显示的预计算图像需要被预加载到专用显卡可用的内存缓冲区中,且颜色深度是受损的。将这种显示方式转变为紧凑型可穿戴解决方案需要多项技术创新。

4 基于集成成像原理的光场显示技术

自 1908 年 Lippmann^[6] 提出集成成像(InI)方法以来,集成成像技术已被广泛用于捕捉真实场景的光场以及用于裸眼自动立体显示器。众所周知,传统的集成成像方法存在低横向和纵向分辨率、窄景深和窄视角的局限性。然而,与其他非立体三维显示技术相比,集成成像技术的简单光学架构使其与头戴式显示光学系统的集成并创建可穿戴光场显示器具有吸引力。

一个简单的基于集成成像原理的头戴式光场显示由一个微型显示面板和一个直接放置在观众眼前的 MLA 组成。显示器用来渲染一组二维元素图像,图像代表三维场景的不同视角。元素图像中相应像素发射

的锥形光束相交并叠加,从而形成似乎在发光并占据三维空间的三维场景。重建的三维场景的光场直接耦合到观看者的眼睛中。例如,Lanman 等^[22] 通过直接在眼前放置一个 MLA 和一个 OLED 微型显示器,搭建了一种非透视头戴式光场显示器,该显示器可为虚拟现实(VR)渲染三维场景的光场所用,其视场约为 29°×16°,空间分辨率为 146 pixel×78 pixel。这种方法的主要优点是器件结构紧凑轻薄;除了不透明的明显局限性外,这种方法还存在横向和纵向分辨率低以及景深浅的主要局限性。受限于独立 MLA 的成像能力,重建的场景似乎离眼睛只有几 inch(1 inch=2.54 cm)。

为了克服上述简单设计方法的局限性,可以在观察者的眼睛和由微型显示器和 MLA 组成的微型集成成像单元之间插入观察光学元件。观察光学系统对由微型集成成像单元重建的微型三维场景进行放大,并使放大的虚拟三维场景呈现出让观看者的眼睛感到舒适的扩展深度体积。如图 6(a)~(d) 的示例所示,观察光学器件可以是简单的放大镜型目镜、透视目镜^[23]、中继光学器件和透视目镜的组合^[24]、准直透镜和基板引导组合器的组合^[25]。最后 3 种配置为 AR 应用提供了光学透视功能。图 6(b) 所示的自由曲面目镜等透视目镜可以实现虚拟三维场景和现实世界场景光线的叠加。例如,Hua

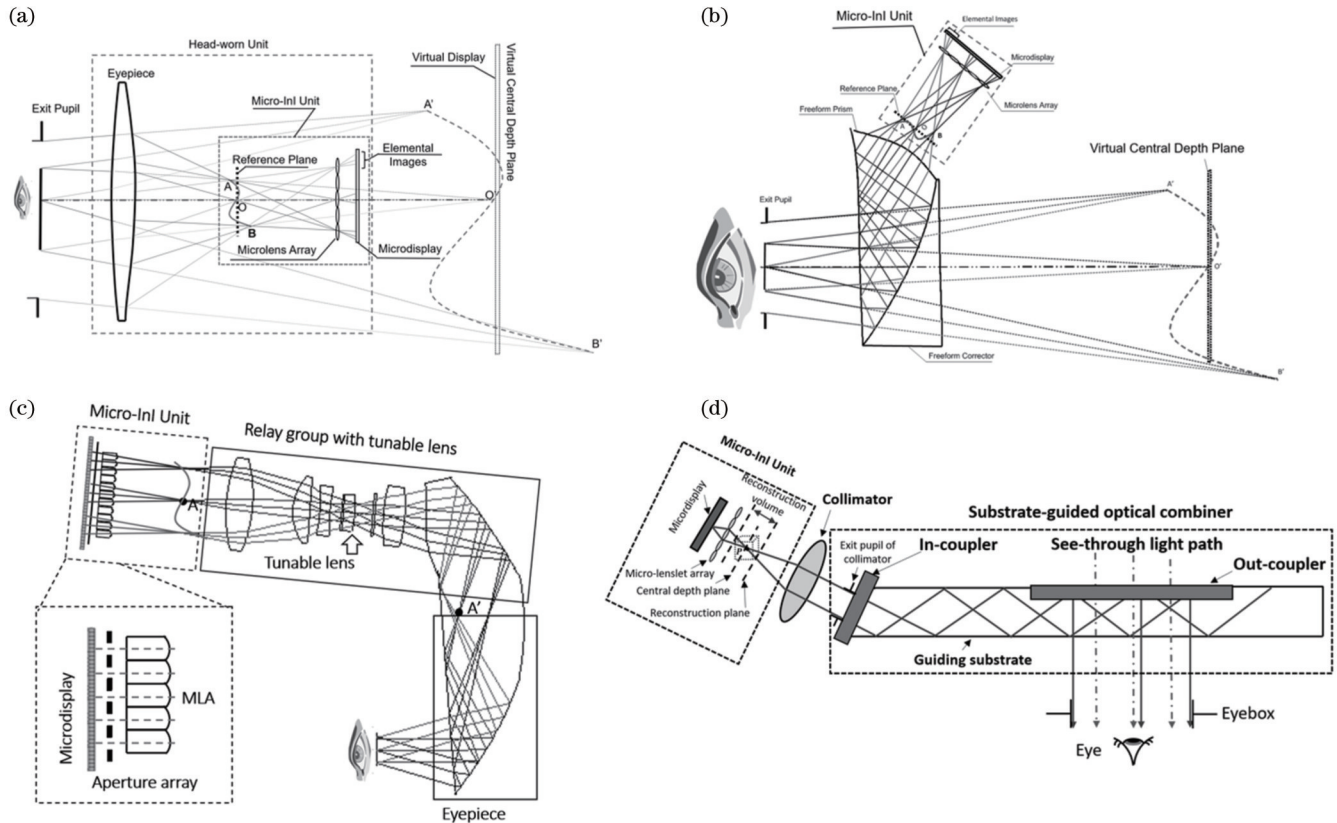


图 6 基于集成成像方法的头戴式光场显示系统(InI-HMD)的不同光学架构。(a)简单放大镜结构; (b)透视式自由曲面目镜^[23]; (c)可变焦中继透镜和目镜结合的透视式光学系统^[24]; (d)准直透镜和波导元件相结合的透视式光学系统^[25]

Fig. 6 Different optical configurations of InI-HMD. (a) Magnifier-type eyepiece; (b) a see-through eyepiece^[23]; (c) combination of a relay optics and eyepiece^[24]; (d) combination of a collimation lens and substrate-guided combiner^[25]

和 Javidi^[23]展示了第一个基于集成成像的光学透视头戴式光场显示的样机设计,它集成了用于全视差三维场景可视化的微型集成成像方法和自由曲面透视目镜光学器件。该样机能重建的三维场景的对角线视场角为 33.4° ,中心深度的最大角分辨率为 $2.7 (')/\text{pixel}$,中心深度位于约 1 m 处,深度范围为 33 cm~4 m。

基于图 6(a)、(b)所示配置的 InI-HMD 通常在对应微显示器目视光学器件的光学共轭平面处具有最高空间分辨率,该平面被称为虚拟中心深度平面(CDP)。由于图像单元引起散焦效应,重建物体的空间分辨率会随着其偏离 CDP 而降低。为了突破这个局限,如图 6(c)的示例所示,可以在目镜和微型集成成像单元之间插入一个中继光学器件,它可以包含一个可调焦光学元件,使得虚拟 CDP 的深度位置可以动态控制,以匹配重建的三维物体的深度范围。这样一来,头戴式光场显示系统的有效景深可以显著扩展,使得系统的空间分辨率可以在非常大的范围内保持与 CDP 相同。2018 年, Huang 和 Hua^[24]展示了一种基于集成成像的高性能头戴式光场显示器,其中使用可调透镜连续调整 CDP 位置并将系统的景深扩展到 3.5 D 以上。当可调焦光学元件和微显示器以相对较低的速度(例如 60 frame/s)运行时,该系统按照可变深度的单一 CDP 模式运行。此外,该系统也可以采用时分复用方法在多 CDP 模式下运行,这就要求可调焦光学元件和微显示器以高出数倍的速度运行,多个不同深度的虚拟 CDP 被时分复用于创建瞬时扩展的深度体积。例如,在 Huang 和 Hua^[24]展示的样机中,两个虚拟 CDP 被时间复用于创建超过 3 D 的扩展深度体积,而重建对象的空间分辨率几乎保持不变。2020 年, Xu 和 Hua^[25]探究了将微型集成成像单元与波导光学组合器集成以实现紧凑系统的可行性, Wang 和 Hua^[26]用数字可切换双焦 MLA 来代替结构复杂、体积庞大的中继光学系统,同时将景深扩展到类似的范围。

为了解决前面提到的 VAC 问题,基于集成成像的头戴式光场显示器需要创建多个不同的图像单元,以便当人眼瞳孔位于不同位置时都能够具有观看自然三维场景的视觉效果并刺激眼睛以适应三维重建对象的深度,而不是光线实际源自的单个图像单元的深度。因此,所重建的虚拟物点的信息在视网膜上形成的图像为单个图像单元的视网膜图像的集成,其视觉质量会随着人眼晶状体的深度调节而变化。

基于集成成像的头戴式光场显示器能否正确刺激眼睛的聚焦调节反应取决于重建光场的质量是否满足要求,这在很大程度上取决于三维光场显示系统的采样特性。与采样属性仅取决于渲染图像的等效像素分辨率的传统立体显示器不同,光场显示器的采样属性可以通过两个采样分辨率来表征:光线在眼瞳箱平面上的方向采样分辨率和光线在图像采样屏幕上的位置

采样分辨率。

Takaki^[27]建议定向光束的角分辨率在 $0.2^\circ\sim 0.4^\circ$,以允许每只眼睛有两个以上的视图样本才能够形成眼睛的聚焦信号。Kim 等^[28]实验测量了人眼观看真实物体和通过集成成像渲染的数字三维物体时的聚焦调节反应,结果表明 71 名参与者中超过 73% 的人能够适应渲染物体的深度而不是显示屏的深度。Stern 等^[29]试图将主要影响人类视觉系统感知的要求与分析工具相结合,以构建用于建立可感知光场和确定显示设备规格的分析框架。为了填补缺乏系统方法来量化聚焦信号渲染的准确性与每个瞳孔区域的样本数量之间的关系以及视图数量与视网膜图像质量之间的权衡关系的关键空白, Huang 和 Hua^[30-31]开发了一个通用框架来对光场显示技术的图像形成过程进行建模。基于此框架,该团队通过考虑视觉和显示因素来模拟和表征视网膜图像质量以及模拟人眼对光场显示器感知的聚焦调节响应。该团队进一步使用该框架来研究视图采样属性(例如视图密度和填充因子)以及位置采样分辨率对三维重建对象的质量以及光场显示器呈现的聚焦调节信号精度的影响。重建质量包括重建对象的空间分辨率、图像对比度、图像伪影和景深,而聚焦信号的准确性受到聚焦对比度变化幅度、离焦对比度梯度和调节误差的影响。

Huang 和 Hua^[30-31]开发了一个通用的光场显示原理模型。在该模型中,用出瞳面上每单位面积所采样的不同视角的密度来标定光线方向采样分辨率,该密度等同于瞳孔同时接收的不同元素图像的数量^[30]。位置采样的特征可以定义为每单位距离的像素密度或每个像素所对应的视角^[31]。总的来说,光线的方向采样和位置采样在影响三维光场显示器的视网膜图像质量和眼睛调节反应方面都起着重要作用。Huang 和 Hua^[30-31]从模型中得出的结论是,视图密度维持在 $0.5\sim 1.3 \text{ mm}^{-2}$ 范围内能够很好地平衡视网膜图像质量和聚焦信号的准确性,这个视图密度所对应的等效视图分辨率相当于在 3 mm 的眼瞳直径范围渲染 2×2 和 3×3 个不同的视图。在无限高采样分辨率的假设下,视图密度在 $0.5\sim 1.3 \text{ mm}^{-2}$ 范围的三维光场显示器所产生的截止角分辨率分别约为 $0.75'$ 和 $1'$ 。一般来说,较低的视图密度会导致较高的图像对比度以及较窄的景深和较差的聚焦信号准确性,而较高的视图密度会导致较低的图像对比度以及较宽的景深和更好的聚焦信号准确性。更高的位置采样分辨率会提高感知的视网膜图像质量和聚焦信号准确性。实际上, 2×2 和 3×3 视图密度以及 $3 (')/\text{pixel}$ 或更高的角分辨率分别产生 0.14 D 和 0.08 D 的平均调节误差^[30-31]。在不影响空间分辨率的前提下, Wang 和 Hua^[32]利用时分复用法来改善投射到人眼入瞳上的视图密度,提高聚焦信息的逼真度。

5 计算多层式光场显示技术

与基于集成成像原理不同,计算多层式光场显示器不是通过针孔阵列或微透镜阵列对三维光场发出的定向光线进行角度采样,而是通过多层像素阵列对定向光线进行采样。计算多层式光场显示技术的原理与压缩显示器相同,它通过均匀或定向背光照明的多层像素阵列对定向光线进行采样^[7]。它通常主要由多层光衰减层和一个均匀或定向背光照明层构成,通过计算优化的方式将三维场景的光场分解为多个衰减掩模,表征每层光衰减器的透射率或反射率。从背光源进入眼睛的每条光线的强度值是光线相交处的衰减层像素值的乘积。由于重建光线具有准直特性,计算多层式光场显示器通过定向光线的集成来渲染三维场景的光场。2020年,Xu和Hua^[33]对计算多层式光场显示系统的特性和性能进行了全面的分析研究。

一般可以通过3种方法来构建计算多层头戴式光

场显示系统。如图7(a)~(c)所示,第一种方法是直接将多个空间光调制器叠起来直接放在眼睛前面,第二种方法是在多层空间光调制器和眼睛之间插入观察光学器件,最后一种方法是把一个空间光调制器和一个点光源阵列组合起来。

Maimone等^[34]开创性地将计算多层式光场显示技术应用于HMD,并展示了第一个计算多层式AR显示器。如图7(a)所示,该设计将多个透明的SLM、一个薄的透明背光源和一个高速电子快门以很小的间距叠在一起。该设计以两种模式运行:增强图像渲染模式(快门关闭)和遮挡真实世界图像形成模式(快门打开)。在增强图像模式下,真实世界的视图被遮挡,同时在SLM层上渲染一组图案,以衰减来自背光源的光线,从而再现一组在所需深度合成虚拟物体的光线。在真实世界成像模式下,背光关闭,快门打开。可以在SLM层上显示遮挡遮罩,以允许选择性地传输现实世界的光线,从而实现虚拟场景和现实世界场景之间的相互遮挡。

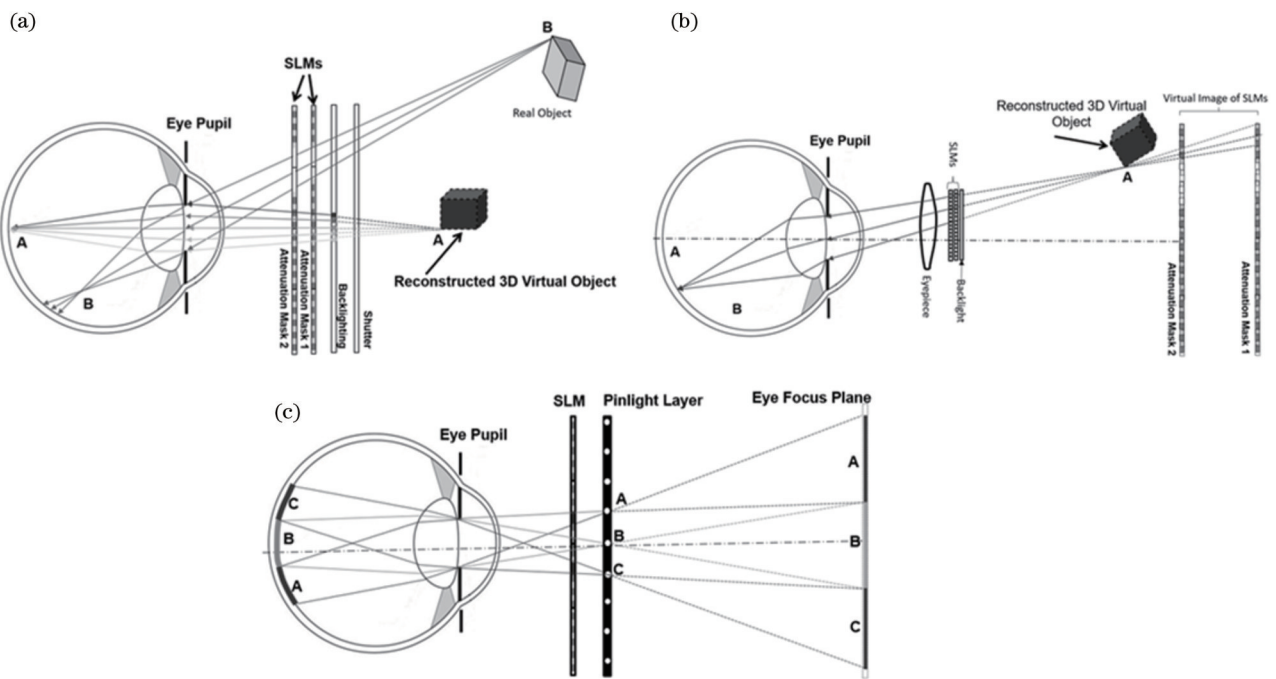


图7 计算多层式光场显示系统的光学架构。(a)无成像光学元件^[34]; (b)多层空间光调制器结合简单放大镜式目镜^[35]; (c)多层空间光调制器结合点光源阵列^[36]

Fig. 7 Multi-layer computational light field HMD configurations. (a) Without imaging optics^[34]; (b) combination of multi-layer SLM and a magnifying eyepiece^[35]; (c) combination of multi-layer SLM and a pinlight layer^[36]

如图7(b)所示,计算多层头戴式光场显示的另一配置是在多层SLM和眼睛之间插入观察光学器件。可以采用类似于图6的示例中使用的不同类型的目视光学器件。例如,Huang等^[35]展示了沉浸式VR应用的光场立体镜。该样机由两个叠放的液晶显示面板和一个简单的放大镜组成,两个液晶面板中间有1 cm的间隙,放大镜的焦距为5 cm。使用rank-1光场分解方法计算调制模式,以合成和渲染三维场景的光场。

如图7(c)所示,多层空间光调制器的一种变体是

把一个调制器与一个点光源阵列组合起来,这种配置被称为针光显示器,这个点光源阵列可以通过将紧挨着均匀背光照面板的光调制器层换成一个针孔阵列来实现,也可以直接将这层光调制器和均匀背光板用一个点光源阵列代替^[36]。在Maimone等^[36]展示的样机中,他们在一个丙烯酸树脂面板上蚀刻出一个均匀分布的针孔阵列,再把LED线阵放置在丙烯酸树脂面板的侧面,LED发出的光通过面板照亮针孔阵列,形成均匀分布的点光源阵列。从每个点光源发出的光线

被 LCD 面板的一部分调制,该 LCD 面板显示从目标图像空间分解的图像单元。调制光线经过眼睛晶状体折射,在视网膜上重聚焦为清晰的像点。所有点光源投射的图像单元平铺在一起以创建宽 FOV 图像。可以让来自多个点光源的光线在不同位置进入眼睛瞳孔并相交于视网膜上,由此就可以渲染三维场景的光场。

总的来说,基于计算多层方法的头戴光场显示技术受到许多明显的限制,需要重大创新才能实现改进。Xu 和 Hua^[33]对计算多层式光场显示技术的系统特性和性能进行了全面的分析研究。首先,由于高分辨率空间光调制器的像素孔径很小,透射式调制器的像素占空比又偏低,这些低占空比的小像素引起的衍射效应使得这类系统的空间分辨率显著降低。虽然可以通过使用具有高占空比的反射式调制器来减少衍射效应,但代价是增加了光学系统结构的复杂性和体积。其次,该技术是计算密集型的,需要显著缩短优化时间才能满足 AR 显示的低延迟应用需求。最后,由于透射式调制器的透射率相对较低,多层叠加使得系统的整体透光效率遭受严重损失。

6 结 论

本文总结了头戴式光场显示技术的不同光学架构以及近年来在不同架构研发方面的进展。对比不同的解决 VAC 问题的显示方法,发现头戴式光场显示技术,尤其是基于多焦面和集成成像原理的光场显示技术,不仅能够渲染接近于真实三维场景提供的视觉聚焦调节信息,从而有效地克服 VAC 问题,而且能够获得高分辨率的图像,在光学架构方面也有很高的灵活性和较强的可扩展性,可以和大多数用于传统头戴式立体显示技术的光学架构兼容。与此同时,要实现高质量、大视场角、小巧轻薄的光场显示的梦想,需要克服一些重大的技术挑战。在硬件技术方面,无论是多焦面光场显示技术还是集成成像光场显示技术都依赖于两个核心技术的发展:一个是高频、大口径、大调焦范围的 VFE;另一个是超高分辨率、高刷新频率、低功耗的微型显示技术。比如基于集成成像的头戴式光场显示器使用了一个分辨率约为 3000 PPI 的微型显示屏实现了大约 30° 的视场角和 3' 的角分辨率。在其他参数不变的情况下,需要将微型显示屏的分辨率提高到 9000 PPI 的水平才有可能将角分辨率提高到与人眼相先匹配的 1' 的水平。如果想要同时增大视场角,对显示屏的分辨率要求会显著提升。除了分辨率要求外,很多光场显示系统使用时复用技术来降低光学架构的复杂度,这就意味着要有高于普通显示屏数倍的刷新频率。除了这两个核心硬件技术之外,其他挑战包括渲染光场的巨大计算要求、传输到显示器所需的大数据带宽要求、系统的巨大功耗要求,以及在亮度、对比度、色域、动态范围和空间时间分辨率等涉及高质量图像方面的要求。尤其是消费者已经充分体验

了高质量二维显示技术带来的视觉效果,三维光场显示技术只有达到与二维显示相匹配的图像质量才能够满足消费者的期望。此外,在克服上述技术挑战的同时,还需要满足头戴式显示技术小巧轻薄的要求。随着各种硬件、软件、算法、网络以及电子技术的快速发展,相信克服这些技术挑战只是时间问题。

参 考 文 献

- [1] Hua H. Enabling focus cues in head-mounted displays[J]. Proceedings of the IEEE, 2017, 105(5): 805-824.
- [2] Adelson E H, Bergen J R. The plenoptic function and the elements of early vision[M]//Landy M, Movshon J A. Computational models of visual processing. Cambridge: The MIT Press, 1991.
- [3] Rolland J P, Krueger M W, Goon A. Multifocal planes head-mounted displays[J]. Applied Optics, 2000, 39(19): 3209-3215.
- [4] McQuaide S C, Seibel E J, Kelly J P, et al. A retinal scanning display system that produces multiple focal planes with a deformable membrane mirror[J]. Displays, 2003, 24(2): 65-72.
- [5] Akeley K, Watt S J, Girshick A R, et al. A stereo display prototype with multiple focal distances[J]. ACM Transactions on Graphics, 2004, 23(3): 804-813.
- [6] Lippmann G. Épreuves réversibles. Photographies intégrales[J]. Comptes Rendus de l'Académie des Sciences, 1908, 146(9): 446-451.
- [7] Wetzstein G, Lanman D, Hirsch M, et al. Tensor displays: compressive light field synthesis using multilayer displays with directional backlighting[J]. ACM Transactions on Graphics, 2012, 31(4): 80.
- [8] Marran L, Schor C. Multiaccommodative stimuli in VR systems: problems & solutions[J]. Human Factors, 1997, 39(3): 382-388.
- [9] Schowengerdt B T, Murari M, Seibel E J. Volumetric display using scanned fiber array[J]. SID Symposium Digest of Technical Papers, 2010, 41(1): 653-656.
- [10] Cheng D W, Wang Q F, Wang Y T, et al. Lightweight spatial-multiplexed dual focal-plane head-mounted display using two freeform prisms[J]. Chinese Optics Letters, 2013, 11(3): 31201-31204.
- [11] Love G D, Hoffman D M, Hands P J W, et al. High-speed switchable lens enables the development of a volumetric stereoscopic display[J]. Optics Express, 2009, 17(18): 15716-15725.
- [12] Liu S, Hua H. Time-multiplexed dual-focal plane head-mounted display with a liquid lens[J]. Optics Letters, 2009, 34(11): 1642-1644.
- [13] Chang J H R, Kumar B V K V, Sankaranarayanan A C. Towards multifocal displays with dense focal stacks[J]. ACM Transactions on Graphics, 2018, 37(6): 198.
- [14] Rathinavel K, Wang H P, Blate A, et al. An extended depth-of-field volumetric near-eye augmented reality display[J]. IEEE Transactions on Visualization and Computer Graphics, 2018, 24(11): 2857-2866.
- [15] Zhang X, Li X, Zhou H Q, et al. Multifocal plane display based on dual polarity stereoscopic metasurface[J]. Advanced Functional Materials, 2022, 32(52): 2209460.
- [16] Suyama S, Ohtsuka S, Takada H, et al. Apparent 3-D image perceived from luminance-modulated two 2-D images displayed at different depths[J]. Vision Research, 2004, 44(8): 785-793.
- [17] Liu S, Hua H. A systematic method for designing depth-fused multi-focal plane three-dimensional displays[J]. Optics Express, 2010, 18(11): 11562-11573.
- [18] Hu X D, Hua H. Design and assessment of a depth-fused multi-focal-plane display prototype[J]. Journal of Display Technology,

- 2014, 10(4): 308-316.
- [19] Hu X D, Hua H. High-resolution optical see-through multi-focal-plane head-mounted display using freeform optics[J]. *Optics Express*, 2014, 22(11): 13896-13903.
- [20] Narain R, Albert R A, Bulbul A, et al. Optimal presentation of imagery with focus cues on multi-plane displays[J]. *ACM Transactions on Graphics*, 2015, 34(4): 59.
- [21] Wu W M, Llull P, Tosic I, et al. Content-adaptive focus configuration for near-eye multi-focal displays[C]//2016 IEEE International Conference on Multimedia and Expo (ICME), July 11-15, 2016, Seattle, WA, USA. New York: IEEE Press, 2016.
- [22] Lanman D, Luebke D. Near-eye light field displays[J]. *ACM Transactions on Graphics*, 2013, 32(6): 220.
- [23] Hua H, Javidi B. A 3D integral imaging optical see-through head-mounted display[J]. *Optics Express*, 2014, 22(11): 13484-13491.
- [24] Huang H K, Hua H. High-performance integral-imaging-based light field augmented reality display using freeform optics[J]. *Optics Express*, 2018, 26(13): 17578-17590.
- [25] Xu M M, Hua H. Geometrical-lightguide-based head-mounted lightfield displays using polymer-dispersed liquid-crystal films[J]. *Optics Express*, 2020, 28(14): 21165-21181.
- [26] Wang X A, Hua H. Depth-enhanced head-mounted light field displays based on integral imaging[J]. *Optics Letters*, 2021, 46(5): 985-988.
- [27] Takaki Y. High-density directional display for generating natural three-dimensional images[J]. *Proceedings of the IEEE*, 2006, 94(3): 654-663.
- [28] Kim Y, Kim J, Hong K, et al. Accommodative response of integral imaging in near distance[J]. *Journal of Display Technology*, 2012, 8(2): 70-78.
- [29] Stern A, Yitzhaky Y, Javidi B. Perceivable light fields: matching the requirements between the human visual system and autostereoscopic 3-D displays[J]. *Proceedings of the IEEE*, 2014, 102(10): 1571-1587.
- [30] Huang H K, Hua H. Systematic characterization and optimization of 3D light field displays[J]. *Optics Express*, 2017, 25(16): 18508-18525.
- [31] Huang H K, Hua H. Effects of ray position sampling on the visual responses of 3D light field displays[J]. *Optics Express*, 2019, 27(7): 9343-9360.
- [32] Wang X A, Hua H. Time-multiplexed method for view density enhancement in integral imaging-based light field displays[J]. *Optics Letters*, 2022, 47(17): 4471-4474.
- [33] Xu M H, Hua H. Systematic method for modeling and characterizing multilayer light field displays[J]. *Optics Express*, 2020, 28(2): 1014-1036.
- [34] Maimone A, Wetzstein G, Hirsch M, et al. Focus 3D: compressive accommodation display[J]. *ACM Transactions on Graphics*, 2013, 32(5): 153.
- [35] Huang F C, Chen K, Wetzstein G. The light field stereoscope: immersive computer graphics via factored near-eye light field displays with focus cues[J]. *ACM Transactions on Graphics*, 2015, 34(4): 60.
- [36] Maimone A, Lanman D, Rathinavel K, et al. Pinlight displays [J]. *ACM Transactions on Graphics*, 2014, 33(4): 89.

Recent Development of Head-Mounted Light Field Display Technologies

Hua Hong*

3D Visualization and Imaging Systems Laboratory, James C. Wyant College of Optical Sciences, The University of Arizona, Tucson 85721, Arizona, USA

Abstract

Significance The human visual system (HVS) exploits a wide range of cues for the perception of three-dimensional (3D) space and shapes. Many different 3D display technologies have been developed over the years, most of which render a subset of these visual cues to facilitate 3D perception. For instance, the well-established stereoscopic 3D displays (S3Ds) render a pair of two-dimensional (2D) perspective images, one for each eye, with binocular disparities and other pictorial depth cues of a 3D scene seen from two slightly different viewing positions. Although they can create compelling depth perceptions, the S3Ds are unable to render correct focus cues for 3D scenes, including accommodation and retinal blur effects and thus are subject to a well-known vergence-accommodation conflict (VAC) problem. The VAC may be manifested as the mismatch between the depths of eye accommodation and vergence or the mismatch between the retinal blurring effects of rendered and real-world 3D scenes. These conflicts are considered the key contributing factors to various visual artifacts associated with viewing S3Ds, such as distorted depth perception and visual discomfort.

Several display methods that are potentially capable of resolving the VAC problem have been explored, among which a light field display aims to render the perception of a 3D scene by reproducing the geometric light rays apparently emitted by the 3D scene in different directions and thus deliver a visual experience that is natural and comfortable for the HVS. The geometrical light rays may be reconstructed through a stack of discrete focal planes placed at different depths or through an optics array angularly sampling the different directions of the light rays viewed from different eye positions. The light field display method is considered one of the most promising 3D display technologies capable of addressing the VAC problem while offering great flexibility and scalability.

Progress Over the last decades, many different methods have been explored and demonstrated to render light field effects

and achieved different degrees of fidelity. Light field displays may be implemented in the form of direct-view displays or head-mounted displays. With exploding interests in virtual and augmented reality applications, various light field display methods have been explored for head-mounted displays, known as light field head-mounted displays (LF-HMDs). In LF-HMDs, the display source and the optics are worn on the viewer's head. The light field of a 3D scene may be sampled by a stack of discrete focal planes placed at different depths along the visual axis, known as multi-focal-plane HMDs (MFP-HMDs), or by an array of optical elements sampling the angular directions of the light rays apparently emitted by the 3D scene, known as integral imaging (InI)-based HMDs (InI-HMDs), or a stack of multi-layer spatial light modulators which computationally reconstruct the four-dimensional light field functions, known as computational multi-layer HMDs.

In multi-focal-plane LF-HMDs, the light field of a 3D scene is sampled by a stack of discrete focal planes placed at different depths along the visual axis, as illustrated in Fig. 2. When the focal planes are adequately dense so that each spatial location of a 3D scene is uniquely sampled by a pixel on a corresponding plane, each pixel is known as a voxel, and the configuration is known as a volumetric display, which can be viewed from a large range of viewing points. In addition, when the focal planes are sparse, and the adjacent focal planes are largely separated, an extended 3D volume is divided into multiple focal zones by these focal planes, each of which renders the 2D projection of a sub-volume of the 3D scene centered at the corresponding depth of the focal plane from a fixed viewpoint. These 2D projections additively reconstruct the light field of the scene seen from the given viewing point. In this configuration, the light rays for reconstructing a 3D scene are sampled from the same viewpoint with ray positions varying in depth. Such a configuration is known as a fixed viewing point volumetric display. Such a simplified sampling mechanism gains the benefit of employing a small number of focal planes adequate for rendering a large depth volume with high spatial resolution at the cost of relatively low depth resolution.

In an InI-based LF-HMDs, the light field of a 3D scene is reconstructed by angularly sampling the geometrical light rays apparently emitted by a 3D scene through 2D optical array elements such as a lenslet array or pinhole array. As illustrated in Fig. 3, a simple InI-based LF-HMD consists of a micro-display panel and a micro-lens array (MLA) placed directly in front of a viewer's eye. The display renders a set of 2D elemental images, each of which represents a different perspective of a 3D scene. The light fields of the reconstructed 3D scene are directly coupled into a viewer's eye, and thus the view window through which the eye of a viewer receives the light fields is confined to a small region.

A computational multi-layer LF-HMD is based on the same principles as compressive displays, which samples the directional rays through multi-layers of pixel arrays illuminated by either a uniform or directional backlight. As illustrated in Fig. 4, the light field of a 3D scene is computationally decomposed into several attenuation masks representing the transmittance or reflectance of each layer of the light attenuators. The intensity value of each light ray entering the eye from the backlight is the product of the pixel values of the attenuation layers at which the ray intersects. Due to the collimated nature of the reconstructed light rays, a computational multi-layer display renders the light fields of a 3D scene through the integral sum of the directional light rays.

The author provides a comprehensive overview of different approaches and recent development of these three different optical architectures for LF-HMDs.

Conclusions and Prospects Though significant progress has been made over the last decade in developing LF-HMD technology, some significant challenges need to be overcome to fully realize light field displays. Besides the obvious challenge of achieving a glass-like form factor, some of the big challenges are huge computational requirements for rendering the light field, large data bandwidths needed for transmission to the display, significant power consumption for the system, and high image quality in terms of brightness, contrast, color gamut, dynamic range, and spatial and temporal resolution expected by consumers because of the advanced state of 2D displays. Furthermore, with the rapid development of hardware, software, computing, and electronics, we remain very optimistic about coming through all of these technical hurdles.

Key words head-mounted displays; light field displays; virtual reality; augmented reality