

第三章 显示与交互设备

The screen is a window through which one sees a virtual world. The challenge is to make that world look real, act real, sound real, feel real. ——Ivan Sutherland

如果说眼睛是人观察现实世界的窗户，那显示设备就是观察数字世界的窗口。显示设备将数字信号转变为人能观察到的光照、声音、触觉等信号，在日常生活中有广泛的应用。如十字路口的交通信号灯，计算器、电脑、手机的显示屏幕，新兴的虚拟现实和增强现实眼镜等等。显示设备让世界变得更加丰富多彩，丰富了人们获取信息的渠道，也大大增加了每天获取的信息量。在显示的同时，交互设备也在改变着人们日常体验，手机从按键到触摸屏，显示从平面到立体且可以与人进行交互，智能音箱智能家居，让生活变得越来越便捷，显示与交互设备一起，改变着新时代人们的生活方式。

3.1 显示设备

数字设备里的信息通过显示设备转化成图像、声音乃至体感信息传达给用户。本章将介绍常见的显示设备、技术及其应用场景和背后的原理。

显示设备的制造依托于显示基元的选取，常见的显示基元有像素 (pixel) 和矢量 (vector) 两种。基于像素的显示方法以其成本低廉、适用广泛的优势，逐渐成为显示设备的主流。目前，以矢量作为基元的显示设备以绘图仪为主，包括笔式绘图仪 (pen plotter)、Vinyl 切割式绘图仪等。此外亦有对于矢量激光投影仪的研究，仍在每年 SIGGRAPH 会议中崭露头角¹。有关像素和矢量表示的具体不同，我们将在图像表示与处理一章中详细讨论。此外，对于基于像素的显示方式，图形处理器，也即 GPU 的发明大大提升了显示前的处理效率。与 CPU 类似，GPU 拥有自己的多级高速缓存和多块并行计算核心；然而，GPU 的处理单元泛化性弱于 CPU，却具有远多于 CPU 的单元数目。上述特性使得 GPU 更适于面对图形显示等高并行计算任务。我们将在图形管线一章中详细介绍。

3.1.1 2D 显示

数码管 数码管，也称作辉光管，是一种可以显示数字和其他信息的电子设备，被广泛应用于仪表，时钟，车站，家电等场合，如篮球比赛的计分板、广告牌等都是采用的数码管发光显示。如图3.1所示的 LED 数码管包括 7 段和小数点，一般有两种驱动方式，静态显示和动态显示，静态显示中每一个段都由一个单片机的 I/O 端口驱动，优点是编程简单，显示亮度高，但需要占用较多 I/O 端口；而动态驱动则是通过电路分时轮流控制各个数码管，每个数码管 1 2ms，由于视觉暂留的效果和余辉效应，视觉上不会有闪烁感，这种方式能够节省大量 I/O 端口，且功耗更低。

¹<https://dl.acm.org/doi/10.1145/3407982.3407990>

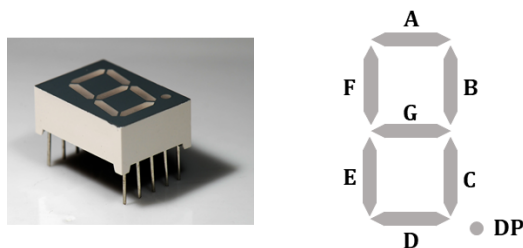


图 3.1: LED 显示元件.

阴极射线显像设备 阴极射线管显示器，又叫 CRT(cathode-ray tube) 显示器，是实现最早、应用最为广泛的一种显示技术，具有技术成熟、图像色彩丰富、还原性好、全彩色、高清晰度、较低成本和丰富的几何失真调整能力等优点，主要应用于电视、计算机显示器、工业监视器、投影仪等终端显示设备。

图 3.2给出了 CRT 的基本工作原理，由电子枪发射出的电子束，通过聚焦系统和偏转系统打到涂有荧光层的屏幕上的指定位置。电子束的强度由控制栅极控制，通过施加负电压，控制发射的电子数量，电子束轰击荧光层会发光，光的强度与轰击的电子数量有关，因此可以通过控制栅极控制显示屏的亮度；聚焦系统用来控制电子束轰击荧光层时会聚到一个小点；电子束的偏转受电场或磁场控制，CRT 颈部包括两对线圈，一对实现水平偏转，一对实现垂直偏转，通过调节经过线圈的电流强度，可以得到适当的偏转量；电子束轰击在荧光屏上，激活荧光层发光，轰击产生的发光随时间衰减，需要反复发射电子束。这类显示器称为刷新式 CRT，在屏幕上画图的频率称为刷新频率。CRT 无重叠显示的最多点数称为分辨率。它常常被简述为每个方向的总点数，但更精确的定义是在水平和垂直方向上每英寸可绘制的点数。

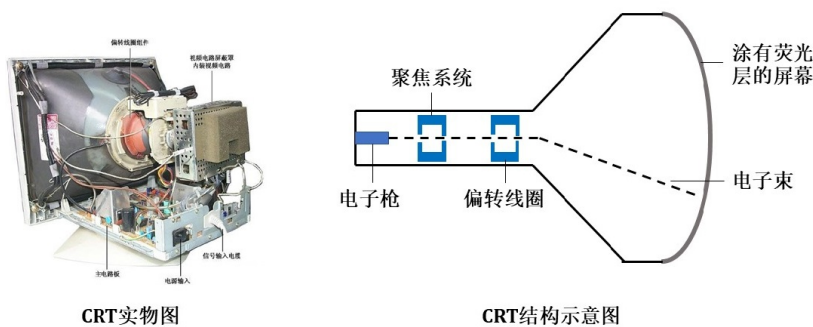


图 3.2: CRT 显示器和 CRT 结构示意图.

使用 CRT 的普通图形监视器通常是**光栅扫描显示器**。电子束横向扫描屏幕，一次一行，从顶到底依次进行，每一行称为一个扫描行。图形定义保存在**帧缓存**中，保存了一组对应屏幕所有点的强度值。每个可由电子束点亮的屏幕点称为一个像素。在部分光栅扫描系统中，采用了**隔行扫描**的方式分两次显示每一帧，一次从顶到底，一行隔一行扫描，另一次扫描另一半扫描线。通过隔行扫描技术，可以把帧率提高一倍，是避免闪烁且提供相邻扫描线包含类似信息的有效技术。

CRT 监视器能利用发射不同颜色光的荧光层的组合来显示彩色图形，不同荧光层的发射光组合起来，就可以形成特定的颜色。通过在屏幕上涂上多层不同的荧光粉，发射颜色由电子束在荧光层中的穿透深度决定，这种方法称为**电子束穿透法**。另一种**荫罩法**能产生的

彩色范围要大得多，每一个像素位置有三个荧光彩色点，分别能发出红、绿、蓝三种颜色的光，电子枪也有三支，与这些点一一对应，荫罩栅格位于屏幕之后，上面包含了一系列孔，对应屏幕上的像素。屏幕能显示的颜色种类与每只电子枪的强度等级有关，有的系统只能控制电子枪的开关，能显示的颜色也就只有八种，而高质量的光栅图形系统每只电子枪能设置 256 级电压，每个像素有近 1700 万种颜色，这种称为**全彩色系统**。

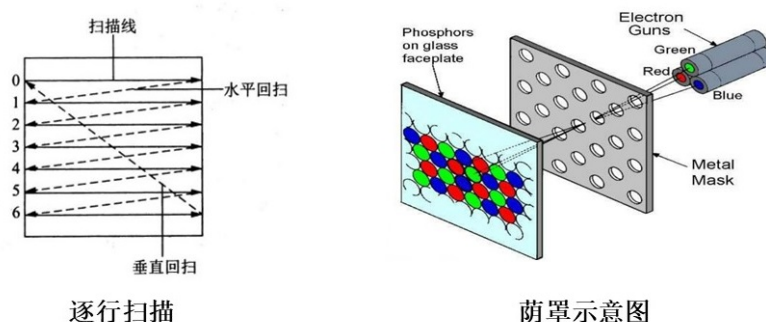


图 3.3: 逐行扫描显示和荫罩栅格。

除了某些特殊的场合还在使用 CRT 显示器外，在绝大多数场合，它已经被轻薄、低功耗、低成本的液晶显示器所取代。随着科技的发展，液晶显示器的应用日益广阔，已经广泛应用于各种仪表、计算器、液晶电视、计算机、掌上电子玩具、手机等许多方面。尽管液晶显示器已经全面取代 CRT 成为电脑装机的首选，但是在一些对色彩还原要求较高的行业，如医疗、冶金等，仍需要使用 CRT 显示器进行作业。

液晶显示器 各种视频信号和电脑数据信息的终端显示，实现屏幕平板化、平板大型化与显示高清晰化，一直是显示技术开发追求的目标。液晶显示器 (LCD) 作为一种“非发射型”平板显示器，较之 CRT 具有屏幕平板化、工作电压低、无电磁辐射、能量消耗低、轻巧便携、使用寿命长、易于集成组装，以及可以实现数字式接口等优点，目前已在笔记本电脑、智能手机、触摸屏显示、液晶电视，以及导航装置、全球定位系统，以及公共信息服务显示等方面取得了广泛的应用，另外，人们日常生产中的许多家用电器今天也离不开 LCD，如微波炉、空调、冰箱、洗衣机、空气清洁装置等，以及应用于许多人所共知的其他方面，如液晶广告宣传牌、液晶计时钟表、液晶游戏机、液晶仪表计量、液晶传感器等。

LCD 是一种采用**液晶**为材料的显示器。液晶 (Liquid Crystal, LC) 是一类介于固态和液态间的有机化合物，加热会变成透明液态，冷却后会变成结晶的混浊固态。在电场作用下，液晶分子会发生排列上的变化，从而影响入射光束透过液晶产生方向上的变化。在液晶前后分别加上互相垂直的偏振片，当不施加电压时，光线经过液晶发生偏转通过两个偏振片，亮度大，而施加电压后，光线经过液晶不发生偏转，无法通过第二个偏振片，亮度低。值得指出的是，液晶材料因为本身并不发光，所以 LCD 通常都需要为显示面板配置额外的光源，主要光源系统称之为**背光模组**。目前 LCD 采用的背光模组可分为侧光式背光模组和直射式背光模组两大类。手机、笔记本电脑主要采用侧光式背光模组，而液晶电视大多采用直射式背光模组光源。

彩色 LCD 的面板中，每一个像素由 3 个液晶单元格构成，每一个单元格前面分别有红色、绿色和蓝色的三色滤光片，通过不同单元格的光线就可以显示出不同的颜色，在高分辨率环境下创造色彩斑斓的画面。

人类视觉器官(眼睛)对动态影像的感知存在所谓“视觉残留”的现象,即高速运动的画面在人脑中会形成短暂的印象。早期的动画片、电影,一直到现在最新的游戏节目正是应用了“视觉残留”的原理,让一系列渐变的图像在人眼前快速连续显示,便形成动态的影像。当多幅影像产生的速度超过 24 帧/s,人的眼睛会感觉到连续的画面。这也是电影每秒 24 帧播放速度的由来。响应时间是 LCD 的一个特殊指标。LCD 的响应时间指的是显示器各像素点对输入信号反应的速度,就是液晶由“暗转亮”或由“亮转暗”的反应时间。此值是越小越好,足够快的响应时间才能保证画面的连贯。如果响应时间太长了,就有可能使 LCD 在显示动态图像时,有尾影拖曳的感觉。LCD 一般的响应时间在 2-5ms。

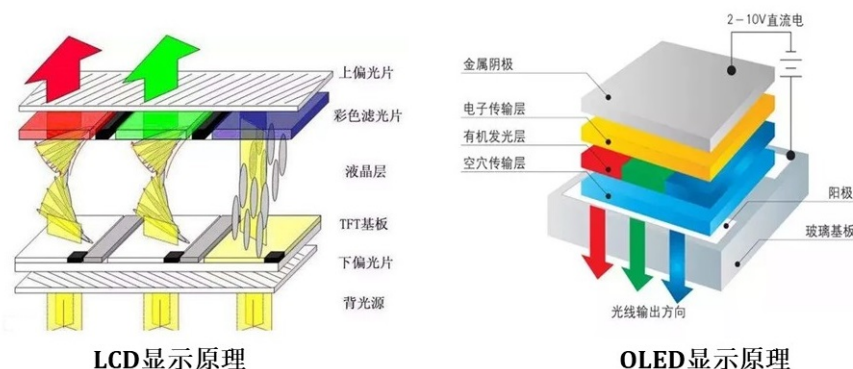


图 3.4: 液晶显示和 OLED 显示发光元件示意图。

OLED 显示器 LCD 本身不能发光,需要背光模组作为外部光源,存在可视角度小,响应时间长,漏光等缺点。有机电致发光器件(OLED)是一种低电压式发光器件,具有发光二极管的性质,相比较 LCD, OLED 的发光器件核心厚度约为十万分之一毫米,比 LCD 要小得多,亮度更高,对比度更大,色彩效果更加丰富,由于不需要背光源, OLED 更加节能,且没有视角问题。以上种种特性,对于手机、数码相机、笔记本电脑等要求轻薄、长续航的显示设备来说,据有很大的优势。同时, OLED 的组成为全固态结构,无真空腔,无液态成分,抗震动性强,可实现柔软显示, OLED 显示屏理论上可以折叠或卷起来,但由于加工工艺还未成熟,目前还没有商业化。

2019 年,在中国面板市场上, LCD 占据 78%, 而 OLED 占 20%, 尽管 OLED 有以上的许多优点,但还有许多关键性缺点没有解决。OLED 最大缺点是寿命低,因为本身是有机材料,电导率低,电阻大,在工作中产生较多热量,加速了器件的化学反应,从而工作寿命低。同时,由于工艺还不够成熟, OLED 的散热,大面积制作等方面还有很多问题有待克服。

显示器颜色指标 在第二节讨论了颜色空间,而对于显示设备来说,也有一套自己的颜色指标。色深代表画面颜色数量的多少,数值越大色彩过渡越平滑自然。目前主流色深是 6bit、8bit 和 10bit, 8bit 指的是红、绿、蓝三原色显示时各有 $2^8 = 256$ 种,总的颜色数量就是 256^3 种。

色域指的是屏幕能呈现的色彩范围,色域越广,能呈现的色彩就越丰富,目前常见的色彩空间有 Adobe RGB、sRGB、CMYK 等。一般来说,显示器中的色彩的覆盖范围 Adobe RGB > sRGB > CMYK。目前用的最广的颜色标准是 sRGB,所以很多显示器的色域标示

会以 sRGB 为基准，如 100%sRGB，125%sRGB 等等，这里的范围一般是指色域容积（不然覆盖率最高只能有 100%），因此 125%sRGB 也不能保证能完全覆盖 sRGB。

高动态范围成像（High Dynamic Range，简称 HDR）是用来实现比普通数位图像技术更大曝光动态范围的技术。和普通的图像相比，它可以提供更多的动态范围和图像细节，与之对应的是**标准动态范围成像**（Standard Dynamic Range，简称 SDR）。根据融合不同曝光时间的低动态范围图像（LDR），利用每个曝光时间相对应最佳细节的 LDR 图像，合成 HDR 图像，能更好地反映出真实环境中的视觉效果。

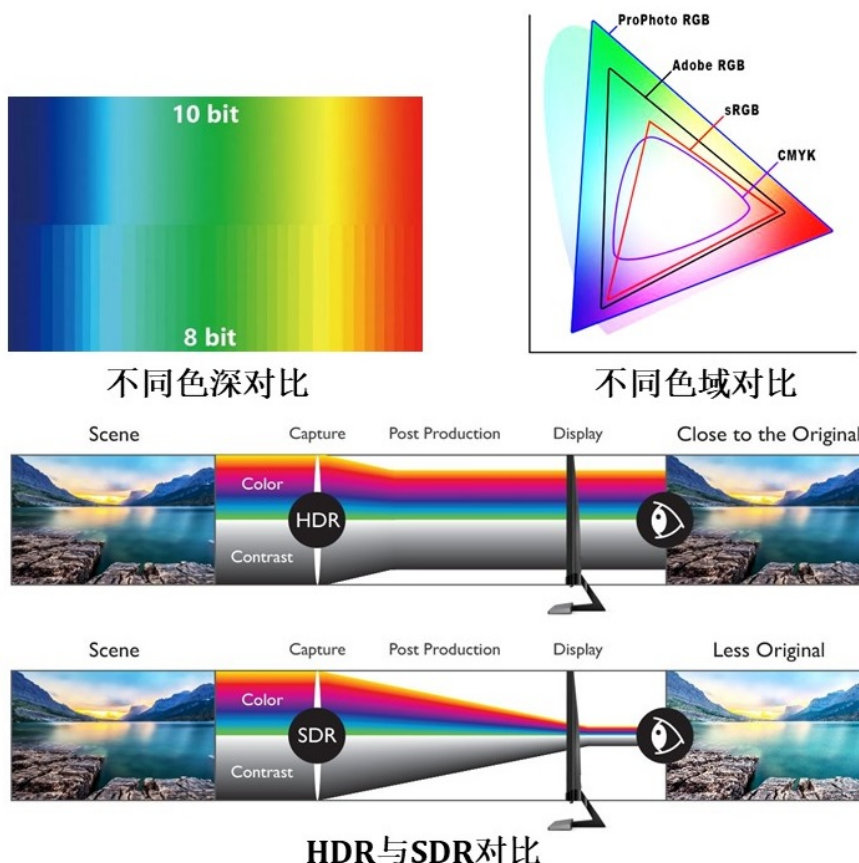


图 3.5: 色深、色域和 HDR.

3.1.2 3D 显示

3D 显示最基本的原理是利用人眼的**双目视差**，由于两只眼睛的瞳距和注视角度不同造成左右眼看物体时存在一定程度的水平差异，经过大脑对不同的图像信息进行分析，就能推测物体到眼睛的距离和自身的形状，也就是**立体感**。3D 显示正是利用这样的原理，通过给两只眼睛展示不同的图像，从而让人觉得观察的目标是立体的而非平面的。第 12 章介绍的三维重建，也是利用多视角图片的视差，来估计重建对象各个部分在空间中的三维位置。

3D 眼镜 电影院里 3D 眼镜用得比较多，常见的有**红蓝眼镜**、**偏振镜**和**液晶快门眼镜**。

红蓝眼镜（色分法）是一般左眼红色镜片右眼蓝色，画面也是红蓝错位的，红色镜片只有红光能通过，蓝色镜片只有蓝光能通过；这样左眼的镜片会过滤掉画面的红内容，只看到蓝色部分，右眼过滤掉蓝色，只看到红色部分，左右眼看到视差画面而产生立体感。这种眼

镜历史悠久，又因其廉价、实惠、几乎不存在维护费用，适用性好的特点，被早期 3D 电影或者 3D 网络电影较多采用。但因为光通量不足，显示画面往往较暗。

偏振镜（光分法）是使用 XY 两个偏转方向，也就是通过眼镜上两个不同偏转方向的偏振镜片，让两只眼睛分别只能看到屏幕上叠加的纵向、横向图像中的一个，从而观看到立体效果。圆偏振是新一代的 3D 偏振技术，眼镜技术相比 XY 偏振要复杂许多；顾名思义，它的镜片偏振方式是圆形旋转的，一个向左旋转，一个向右旋转，这样两个不同方向的图像就会被区分开。IMAX 使用 2 个相互垂直的偏振片，放映机是 2 台，播放经过偏振的相互垂直的图像，只有偏振方向和镜片偏振片方向相同才可以看到图片，效果最好，几乎完全还原，观看时必须坐直，如果镜片有歪斜将模糊。

液晶快门眼镜（时分法）通过 3D 眼镜与显示器同步的信号来实现。3D 影片播放时，屏幕上同偏振影片一样是两幅图像，但这两幅图像是交替快速闪烁的，A 图出现则 B 图消失，B 图出现 A 图则消失。同时液晶快门眼镜会按照影片所给的信号，对应相应的 AB 二图进行同步交替的镜片开关动作，假设人眼识别为 60HZ 画面，所以时分法的屏幕画面刷新率只要达到 120Hz 即图象不会产生抖动感，可满足用户的观影体验。这种技术效果虽然不错，但设备昂贵，且有一些使用限制，长时间观看会导致眼部疲劳，信号也容易受到干扰，还需要用电。



图 3.6: 3D 眼镜。

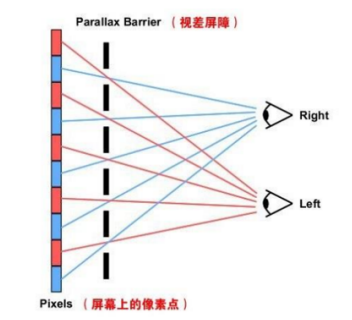
裸眼 3D 在国内的很多商圈，开始出现一种 3D 曲面屏幕，超大的屏幕中展示的画面，无需让人佩戴辅助眼镜，就能具有栩栩如生的 3D 效果。但其实这种屏幕并不算真正的裸眼 3D 技术，而是利用两块屏幕分别展示不同视角的画面，形成视差实现 3D 效果。这也就是为什么如今全国各地 3D 裸眼视频拍摄的短视频角度均一致，并且电子屏幕的视频成像都是以高倍数高清视频为主。同时，很多影像也会采用黑框等参考物，让人产生 3D 的心理“错觉”。

要想取得 3D 显示的效果，就需要形成双目视差，电影院通过过滤不同光的辅助眼镜来形成视差，而裸眼 3D 也是想办法让进入双眼的內容不一样。如图 3.8 所示，真正的裸眼 3D 技术目前有两种：

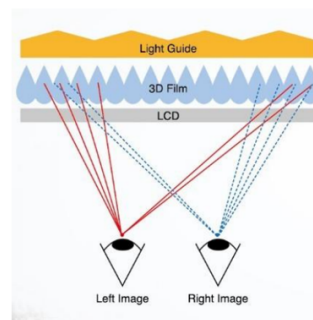
- 光屏障式技术。光屏障式 3D 技术的实现方法是使用一个开关液晶屏、偏振膜和高分子液晶层，利用液晶层和偏振膜制造出一系列方向为 90 度的垂直条纹。这些条纹宽几十微米，通过它们的光就形成了垂直的细条栅模式，称之为“视差障壁”。而该技术正是利用了安置在背光模块及 LCD 面板间的视差障壁。通过将左眼和右眼的可视画面分开，使观者看到 3D 影像。
- 柱状透镜技术。柱状透镜技术也被称为微柱透镜 3D 技术，使液晶屏的像平面位于透镜的焦平面上，这样在每个柱透镜下面的图像的像素被分成几个子像素，这样透镜就能以不同的方向投影每个子像素。于是双眼从不同的角度观看显示屏，就看到不同的子像素。柱状透镜技术并不会像光屏障式那样影响屏幕亮度，所以其比后者的显示效果要好。



图 3.7: 商圈 3D 曲面屏.



光屏障式3D技术



微柱透镜3D技术

图 3.8: 裸眼 3D 显示技术.

虚拟现实和增强现实 虚拟现实技术和增强现实技术是当下前沿的显示技术，随着元宇宙概念的大火，这两项技术也吸引了更多大众的关注。

虚拟现实（Virtual Reality，简称 VR）技术是利用计算机技术模拟产生一个三维的虚拟世界，提供使用者关于视觉等感官的模拟，让使用者感觉仿佛身历其境，可以即时、没有限制地观察三维空间内的事物。相比 3D 眼镜，虚拟现实技术创造了一个“真正”的 3D 世界供观众进行观看，在 VR 游戏中，还加入了交互功能，观众能供通过手势交互，控制手柄等方式，与虚拟世界中的人、物进行“接触”。

虚拟现实技术的核心在于“以假乱真”，首先要构建看起来足够真实的虚拟世界，这涉及到对真实世界的仿真，如保证物体的运动符合物理规律，模拟的烟雾水流等物理现象看起来足够真实；其次，需要让观察者视野里的内容符合感知规律，通过双目视差形成三维效果，当人运动时，显示的内容也应当随观察的位置、视角一起发生变化；更加复杂的虚拟现实还包括声音的三维模拟，以及交互的模拟，举例来说，当人需要与场景中物体进行交互时，需要进行精细的定位和人的运动感知；同时，所有这些显示内容都需要实时进行计算。这些都是虚拟现实的重要技术挑战。

增强现实（Augmented Reality，简称 AR）技术则是把虚拟的物体“放置”在真实世界中。通过计算机实施模拟仿真处理，将虚拟信息内容叠加在真实世界中，并让用户能够观测到，实现超越现实的感官体验。由于需要和真实世界相结合，所以 AR 技术需要对现实三维世界进行重建注册，并且能定位观察设备所处的位置，然后在沉浸式头盔或眼镜上同时显示真实世界并“放置”和“显示”虚拟物体。有些复杂的 AR 场景，还要模拟动物等复杂的对象，以及与用户进行交互，这就要求进行物理仿真的模拟和对用户行为的识别。



VR 头盔



Microsoft HoloLens

图 3.9: VR 和 AR 设备。

全息摄影 全息投影技术属于 3D 技术的一种，原指利用干涉和衍射原理记录并再现物体真实的三维图像的技术。而后随着科幻电影与商业宣传的引导，全息投影的概念逐渐延伸到舞台表演、展览展示等商用活动中。但我们平时所了解到的全息往往并非严格意义上的全息投影，而是使用佩珀尔幻像、边缘消隐等方法实现 3D 效果的一种“伪全息”投影技术。

佩珀尔幻象（Pepper's ghost）是一种在舞台上与某些魔术表演中产生幻觉的技术。这种技术借由使用一面平坦的玻璃与特定的光源技术，使物体可以出现或消失，或是变形成其他物体，所以感觉是产生了幻觉的技术。佩珀尔幻象其实是伪全息投影。因为它必须在固定的舞台上，且要在黑暗当中才能实现，而且观众必须要从特定的角度进行观看。如果角度有偏差画面就会产生变形。

真正的全息投影从任意视角都能观察画面。首先理解什么是“全息”，“全息”即“全部的信息”，光波场的复振幅表达式为 $U(P) = A(P)\exp[i\varphi(P)]$ ，其中 i 是虚数单位， P

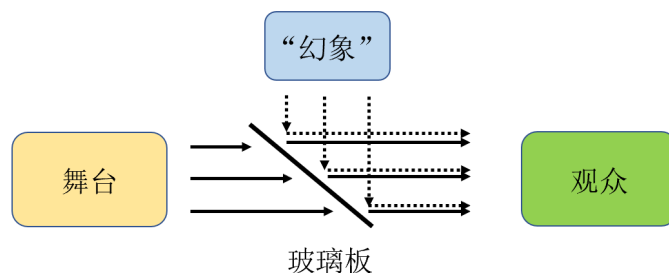


图 3.10: 佩珀尔幻象的原理: 来自明亮舞台的入射光线在穿过玻璃射向观众时发生折射, 而代表“幻象”的入射光线则被玻璃反射. 两个影像都会进入观众视线, 从而形成“幻象”在舞台上出现的效果.

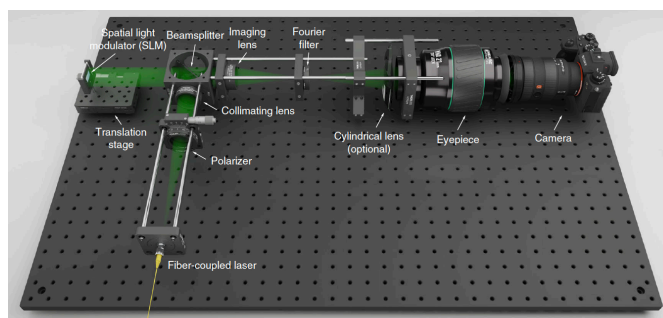


图 3.11: 全息投影设备 [5].

表示空间位置, $A(P)$ 表示振幅, $\varphi(P)$ 表示相位. 图片中存储的是光强信息 $I = U(P)^2 = |A(P)\exp[i\varphi(P)]|^2$, 由于 $\exp[i\varphi(P)]^2 = 1$, 所以相位信息在图片中丢失了. 而全息图则是利用一束已知信息的参考光和物光波产生干涉, 把相位信息也存储下来.

设物光为 $U_1 = A_1\exp[i\varphi_1]$, 参考光为 $U_2 = A_2\exp[i\varphi_2]$, 两束光干涉后的光强表达式简化为:

$$I = |U_1 + U_2|^2 = A_1^2 + A_2^2 + A_1A_2(\exp[i(\varphi_1 - \varphi_2)] + \exp[i(\varphi_2 - \varphi_1)]) \quad (3.1)$$

注意到, 里面的 I, A_1, A_2, φ_2 都可以设置或者获取, 物光的相位信息从而也被存储下来. 全息图的任何局部都能再现原物的基本形状, 物体上任意点散射的球面波可抵达全息干板的每个点或每个局部, 与参考光相干涉形成基元全息图, 也就是全息图的每点或局部都记录着来自所有物点的散射光. 因此, 物体全息图每一局部都可以再现出记录时所有照射到该点局部的物点, 形成物体的像, 也就是破损后部分全息图仍能再现物体的像. 不用通过任何特殊介质, 在空气中就能显示出物体的全息影像, 人用肉眼从不同的方位和角度观察, 能看到被摄物体各个角度的立体画面, 甚至能在影像中穿梭自如.

全息技术由于它能存储更多的光场信息, 因此有很广泛的应用场景, 如显微测量、全息存储、全息计量、全息显示等领域. 对于上文提到的全息图, 只需要再用参考光进行照射, 人就可以看到原始的图片. 同样, 也可以对全息图进行数字重建, 分析物体三维信息. 对于科研中的显微成像, 普通的光学显微镜无法读取被观察物体的三维信息, 而数字全息显微镜可以在计算机中重建物体的 3D 模型, 方便研究.

计算全息则是抛弃了干涉图的记录过程, 直接将光场分布使用计算机通过数学运算计算出来. 这样做有一个巨大的好处, 那就是可以实现任意物体的全息显示, 即便这个物体在

现实中并不存在，因此许多产品的防伪标识都可以使用这种方式来实现。

图3.11中的空间光调制器（SLM）上面分布着微米量级的像素点，每个像素点都可以提供独立的相位调制，也有一些 SLM 可以实现振幅的调制，将计算出来的全息图加载到这上面，然后使用参考光束照射，就可以复现出立体图像。SLM 由于像素点尺寸比光波长大很多，显示的物体大小很受限制。但是由于 SLM 每个单元像素都可以实现独立调整，所以很适合进行动态显示。

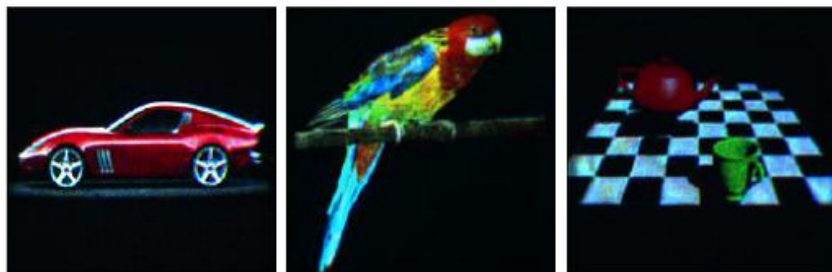


图 3.12: 使用 SLM 进行全息图显示。

超表面全息 3.13是近年来的研究热点，有各种各样的形貌。例如用于全息显示的超表面是由一系列的微纳结构实现的，这些微纳结构有着不同的尺寸或者转角，可以对光场进行振幅和相位的调制。经过计算编码，将微纳结构按照需求排布成全息图，替代全息干板的作用。它的一大优势在于单元像素点尺寸小于波长，因此可以实现放大很多倍的图像显示。

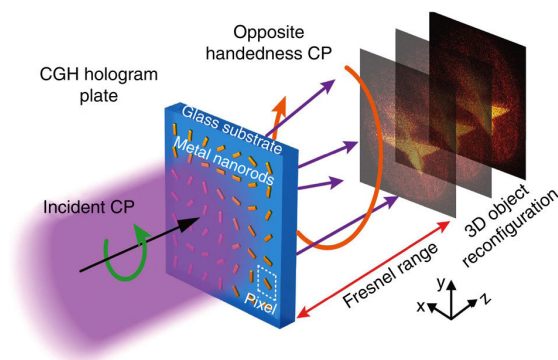


图 3.13: 该图来自于 2013 年发表在权威刊物 Nature Communications 上的一篇文章，通过超表面实现了一个飞机图像的 3D 显示。

如今全息技术正在朝向数字化、光机电一体化方向发展，作为一门新兴学科，全息技术仍处于蓬勃发展的阶段。结合光全息的原理和特点就可以看到其独特的优势，全息技术有着非常广阔的应用市场，无论在国防、医药、工业，还是公共安全等各个领域都有非常巨大的应用，其潜力仍有待挖掘。

GigaPixel Display 在电影里和展览馆里能看到通过围成一圈 2D 显示屏来实现沉浸式 3D 体验的方式，这种方式直接通过展示各个方向的视频，来实现 3D 展示的效果。这种显示方式也可以用于转播、监控等场景，利用多个屏幕观察不同的画面。



图 3.14: GigaPixel 设备.

3.2 交互设备

交互设备也指输入设备，即用户向计算机提供输入信号的设备，如常见的鼠标、键盘、手柄、触控面板、麦克风等等都是常见的输入设备。交互设备通过捕捉用户输入的机械、声音信号，转化成计算机能够识别的电信号，再通过程序进行反馈。随着技术的发展，各种传感器的出现和传感精度的提升给用户提供了更多交互的方式。一般情况下，显示和交互设备都是一起出现的，比如有电视配有遥控器调整声音，手机屏幕配有触摸屏等等。本节主要介绍几种特殊的交互方式和背后的技术细节。

3.2.1 体感交互

体感交互，顾名思义是要通过交互设备理解人体的运动。switch 舞蹈游戏《舞力全开》利用手柄内的传感器来推测人的动作，结合屏幕上的舞蹈动作进行打分，由于传感器的输入信息十分有限，所以动作的估计误差比较大。微软开发的 Kinect 设备，装有一个 RGB 摄像头，一个红外摄像头和一个结构光深度感应器，通过这些感应装置，能拍摄到用户，并通过算法估计出用户的姿态，从而实现与用户行为进行交互的功能，实现免控制器的游戏与娱乐体验。



图 3.15: Kinect 设备，上面有多个摄像头，分别可以拍摄得到彩色图、深度图、红外图等。

3.2.2 手势交互

手势交互是非接触式用户交互方式的一种，图 3.9 中的 AR 眼镜就提供了手势交互的功能。眼镜会追踪用户移动和视线，生成适当的虚拟对象，通过光线投射到用户眼中。因为设备知道用户方位，用户可以通过手势与虚拟 3D 对象交互。

手势中包含复杂的信息，如手指划动、挥动拳头、抓握等动作，由于手的运动很快，因此捕捉设备据有较高的捕捉频率和处理速度。比较有名的公司比如 Leap Motion。捕捉设备可以拍摄手势的 RGB 图片或是深度图片，然后通过算法推测出手的姿势。

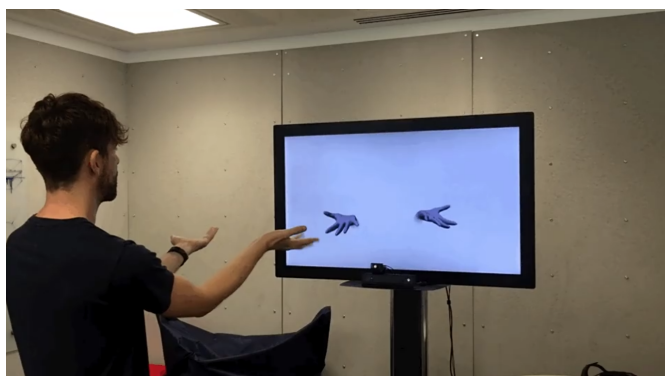


图 3.16: 通过 kinect 设备捕捉手势数据并在屏幕上进行展示。

3.2.3 动作捕捉

动作捕捉技术是影视、动画、游戏等领域的重要技术，前文提到的体感交互中就包含了动作捕捉技术。在动作捕捉技术出现之前，动画里的人物动作都是需要人工一帧一帧画上去的，无论是 2D 还是 3D，动画还是游戏，都要求动画师根据感觉和经验，一点一点手调关键帧。而有了动作捕捉技术，则可以通过真人得到运动数据，再用于控制虚拟角色生成动画。动作捕捉不止能用于肢体动作，也能用于面部表情的捕捉。具体的技术细节将会在第 ?? 章进行介绍。

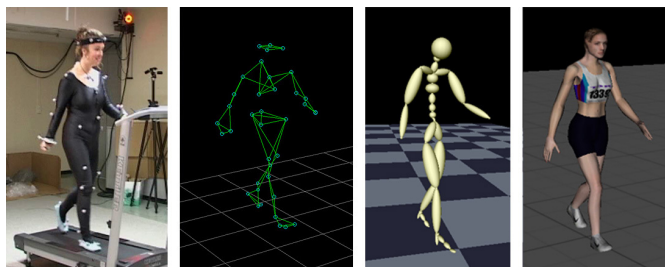


图 3.17: 通过动作捕捉设备获取运动数据，再用于生成虚拟动画角色。



图 3.18: 《猩球大战》中通过面部捕捉技术生成非常拟人化的猩猩面部表情。