

VR、AR、MR

VR (Virtual Reality), 即虚拟现实, 是利用计算机图形学、传感器、人机工程等多学科协作, 模拟一个三维空间的虚拟世界, 提供用户逼真的、沉浸式的、多通道的感官的模拟, 使用户可以即时地、没有限制地观察虚拟世界的三维空间内的事物的技术。

AR (Augmented Reality), 即增强现实, 指利用透过摄像机影像的位置及角度配合图像分析技术, 让屏幕上的虚拟世界能够与现实世界场景进行结合和交互的技术(构建虚拟影像和现实世界的映射关系)。

MR(Mixed Reality), 即混合显示, 指结合真实和虚拟世界创造了新的环境和可视化, 物理实体和数字对象共存并能实时相互作用, 以用来模拟真实物体的技术。



VR

核心技术包括：计算机图形学、三维建模、虚拟图像实时生成（即 CG，Computer Graphics）、立体显示、立体声合成、触觉反馈等。

VR 技术在实际应用上目前可以分为两支，一种是需要佩戴 VR 眼镜实现的体验，最常见的就是游戏上的应用，如节奏光剑（Beat Saber）；一种是不需要佩戴 VR 眼镜就能实现的，如全景照片、全景视频等（商业化的有 VR 看房这样的应用）。

目前的 VR 应用场景包括，游戏、教育、商业广告、科研等。

但目前 VR 遭遇的主要问题是“杀手级”的应用，即并没有某个特定的应用场景时只能由 VR 完成或 VR 能带来极大体验提升的。此外，另一个出现在 VR 眼镜上的问题就是眩晕问题。

3| 特性

Immersion，沉浸性。

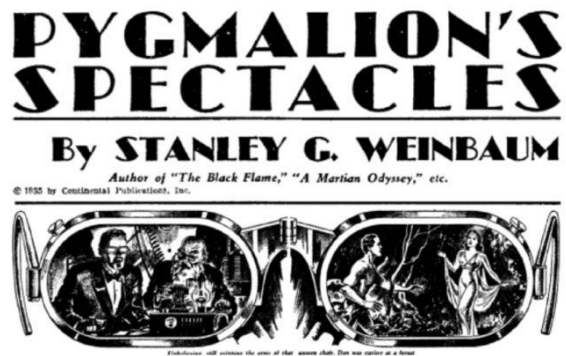
Imagination，想象性。

Interaction，交互性。

发展

1935 年，美国小说家斯坦利·温鲍姆（Stanley Weinbaum）的小说《皮格马利翁的眼睛》中，首次出现了 VR 的概念。

批注 [刘1]: 参考：
<https://zhuanlan.zhihu.com/p/26592125>、



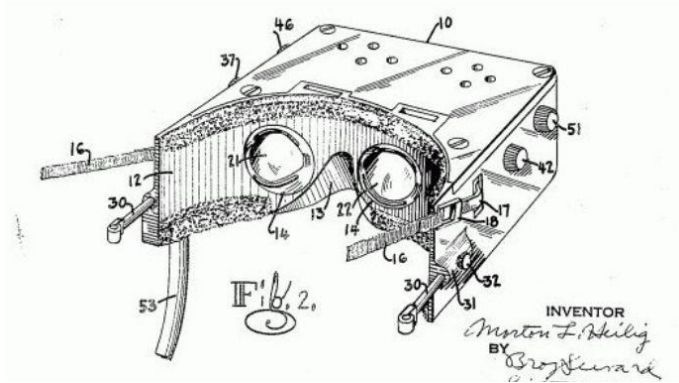
Stanley Weinbaum 的小说《皮格马利翁的眼睛》

20 世纪 50 年代 ,VR 原型机 Sensorama 被美国摄影师莫顿·海里格(Morton Heilig) 发明 , Sensorama 能够实现 3d 立体声、3d 显示、振动座椅、风扇、气味生成。该设备被认为是 VR 设备的鼻祖。但这一产品由于体型过大 , 并没有实现商品化。



Morton Heilig 发明的 VR 原型机 sensorama

进一步的 , Morton Heilig 在 1960 年设计了一个简易的头戴式的 VR 眼睛。但受当时图形显示技术的限制 , 这更像是一个头戴式的 3d 电影眼睛。



Morton Heilig 在 1960 年设计的 VR 眼睛

1968 年，美国计算机科学家伊万·萨瑟兰（Ivan Sutherland，没错就是之前《人机交互接口》中提到的那个发明 CAD 的人）发明了一款 VR 设备，但由于其质量过大，需要有机械臂悬吊在人的头顶，因此被戏称为“达摩克利斯之剑”。该设备通过超声波加机械轴，实现了初步的姿态检测功能，当用户头部发生姿态变化时，计算机会计算新的图像并显示给用户。



Ivan Sutherland 发明的 VR 眼镜

1995 年，任天堂推出 VR 产品 **Virtual Boy**。从它的设计中已经能看出很多今天 VR 眼睛的影子了。但这款产品由于软硬件双方面的问题，最终黯淡收场。同时期，世嘉（sega）、美国的 Forte Technologies（不是福特）等等厂商都推出了各种 VR 眼睛，但最终都没有取得显著的商业成绩。

批注 [刘2]: 参考：

<https://www.bilibili.com/video/BV1iJ411F7Am>.



任天堂 Virtual Boy

2012 年，Oculus Rift 作为一款众筹产品问世，重新使 VR 出现在人们的视野之中。



Oculus Rift 的早期模型

2014 年，谷歌提出了纸板 VR——CardBoard，使得用户能以极低的价格体验到 VR。

眼镜本身就是一个纸板制成的装置，通过把带有陀螺仪的智能手机放入其中作为双眼显示工具，即可实现 VR 的体验。我一开始是挺不认可这类产品的，但后来发现其实这种产品的主要目的并非实现优秀的 VR 体验，它更像是 21 世纪的万花筒，用来作为一个简单的 VR 图片查看装置也许才是这个产品的主要目的。

批注 [刘3]: 参考：
<https://www.bilibili.com/video/BV19s411C7pQ>。

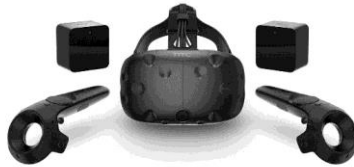


谷歌的 CardBoard

2016 年被称为 VR 元年，Oculus Rift、HTC Vive（2015 年 3 月）、索尼 PS VR、三星 Gear VR（2015 年 11 月）等产品的接连上市，使 VR 市场重新被推到风口浪尖。



Oculus Rift



HTC vive



PS VR



Gear VR

然而在短暂的流行之后，VR 市场很快又陷入了寒冬。过大过重的显示装置、可能存在的眩晕、使用场景受限、过高的价格、缺少核心应用场景等问题使得人们再次意识到，目前的技术所达到的 VR 仍然与理想的、沉浸式的虚拟现实体验有很大差距。

之后，在 2019 年，随着性能更优的无线 VR 眼镜的出现，VR 市场又慢慢走出了之前的寒冬。Oculus Quest 的出现使得 VR 设备的购买门槛大大降低。而 Valve Index 在设备

性能的优化也是的之前 VR 设备中出现的低帧率、镜头畸变等问题得到了解决。设备从有线变为无线更是给用户提供了极大的操作便捷。VR 市场在经历了大浪淘沙之后，开始慢慢地向更优秀的方向发展。



Oculus Quest



Valve Index

头戴式 VR 设备定位技术的实现方式

批注 [刘4]: 参考：

<https://www.zhihu.com/question/46422259>。

不同版本、不同品牌的 VR 产品在定位方式上有着很大的差异，目前来看一共可以分为 4 类：灯塔定位（Light House）、OC（Oculus）星座、从内向外的图像识别、从外向内的图像识别。

定位技术的主要指标有三个：定位精度、定位频率、跟踪范围。

定位精度就是指对 VR 设备的具体位置的判断精确程度，精度过低会导致使用时手柄或主视角的显示出现位置上的偏差。

定位频率就是指定位设备每次对 VR 眼镜及手柄的定位的时间间隔。定位频率过低，VR 设备的动作延迟就会过高，导致眩晕感强烈。

跟踪范围就是指定位设备能对 VR 设备进行持续定位的范围。当跟踪范围过小时，用户很有可能在使用时离开 VR 设备的跟踪范围，导致设备无法继续正常使用。

灯塔定位

这种定位方式被 HTC vive、Valve Index、小派 5K+ 使用，是最为精确但也最为昂贵

的。

这种定位方式与激光雷达类似，使用了机械式的、两个同步高速旋转的激光发射器，分别横纵向放置，VR 眼镜和手柄上都安装有激光接收器。当横向的激光发射器开始转动后，当转动到特定角度时，发射出的激光会被设备上的接收器所捕获。由于激光发射器的转动是匀速的，因此通过确定激光信号从发出到被捕获的时间就可以计算出发射器转动角度，进而可以确定接收器在横向上的位置。同样的道理可以确定发射器在纵向上的位置。灯塔定位通常需要 2 个激光发射基站，当 2 个基站都在同一个周期内确定了接收器的相对位置时，就可以确定出接收器相对于基站的空间位置，从而完成定位。

灯塔定位的优点在于，其定位精度和定位频率都很高；机械式构造寿命长；只要特定空间内放置两个基站，就能实现对 VR 设备的全方位跟踪；如果在身上增加更多的定位设备，还能在 VR 中实现更多部位的控制。

灯塔定位的缺点在于，其机械式的构造一般造价较高，大量的反光平面可能会产生反射，干扰激光定位。

星座系统定位

这一定位方式被应用在第一代 Oculus Rift 中。（这一方式其实和后面的由外向内的图像识别几乎是一样的。）

这种定位的主要原理是使用红外发射器和红外摄像头，配合图像识别完成定位。VR 的眼镜和手柄上都安装了大量的红外发射器，不同设备的红外发射器频率可能有所不同。使用时在使用者前后放置共 3 个红外摄像头，摄像头会不断拍摄图像，捕捉设备发出的红外灯在图中的位置，然后根据 cv 算法实现建模计算，确定每个设备所在的位置。

星座系统定位的定位精度受图像分辨率与图像识别的误差决定，虽然没有灯塔定位的精度高，但并不影响正常的使用体验。

批注 [刘5]: 参考：
https://www.sohu.com/a/85699098_114877。

由于算法进行位置识别时需要花费的时间更多,因此星座系统的定位频率不如灯塔定位高。

通常,只要有 3 个红外摄像头,就可以实现一定空间内的持续跟踪。

不过这一定位方式在一代之后就被舍弃了,个人猜测可能是 cv 算法的发展和设备成本等原因导致 Oculus 决定更换定位方式的。

从内向外的图像识别

这一定位方式被应用在微软的 Hololens (虽然这个其实更像 AR 设备) 和 Oculus 最新的 Oculus Quest 及 Oculus rift 中。

这一定位方式是性价比最高的定位方式。它通过在 VR 眼睛上安装摄像头来拍摄周围的画面,然后根据周围画面的特征进行 cv 计算,构建空间三维模型。在由陀螺仪、加速度计等传感器的信息获得运动信息后,经过模拟的三维空间校准,实现定位。

这类定位方式对 VR 手柄的定位也是通过内置惯性传感器加图像识别实现的。手柄上通常有一些特征点 (或者有些厂商会在手柄上放置红外 LED), 方便 cv 算法的定位。

由内向外的图像识别在定位精度和定位频率上都只能说做到刚刚够用。而且由于它对手柄的定位同样依靠图像识别,因此一旦手柄移动到摄像头拍摄范围的外部就会出现跟丢的情况。依靠图像识别的定位使得这类设备可以在一般正常的起居室环境中良好工作,但在过大的开阔空间(不过好在一般 VR 的使用环境也就在 10m*10m 的范围以下) 及光线过量或过暗的房间中会出现较大问题。该方法可能是目前定位的主流发展方向。

其最显著的优点就是价格低廉。由于不需要额外的基站,图像识别定位技术也在今年有了很大发展,因此设备的成本得到了很好的控制。由于没有基站的限制,设备的使用情景也得到了很大的扩展,用户甚至可以躺在床上使用这一设备。

批注 [刘6]: 参考:

<https://www.bilibili.com/video/BV1SJ411K7D6?from=search&seid=16511678261391686946>。

由外向内的图像识别

只有 PlayStation VR 采用这种由外向内的图像识别定位技术。

这类定位技术其实与星座系统定位很类似（原理几乎是一样的），只不过 PS VR 只有一个双目 RGB 摄像头用于进行定位（星座定位则使用多个红外摄像头），且它是依靠一般的 RGB 图像通过寻找 VR 眼镜和手柄上的 LED 光点实现定位的。

这一定位方式使得 PS VR 的定位精度极低，尤其是在室内光线明亮的情况下。当处于黑暗环境中时，PS VR 的定位精度能略微有所改善。PS VR 的定位频率由图像识别速度决定。而由于它只有一个位置的摄像机，因此用户并不能在佩戴 VR 眼镜后随意移动。PS VR 更像是为 Play Station 用户提供了一个额外的视觉输出补充的设备，而非一款主流的 VR 设备。

PS VR 的价格相比其他设备都要更低，且可以搭配已有的 Play Station 摄像头使用。

VR 设备中的输入元件

手柄

激光接收器（或红外接收器）

以灯塔定位为定位方式的 VR 设备中都有激光接收器用于接收基站发出的信号。我们通常会看到 VR 手柄都是一个头顶带环的设计，其原因就是为了在环中放置激光（或红外）接收器（或者对于主动进行图像识别实现定位的设备而言是红外发射器，这个后面再讲）。

电容压力传感器

VR 手柄上的按钮基本都是电容式的按压按钮。

一些 VR 手柄（如 HTC vive 和 Valve Index）会采用触控板来实现按压输入，那些触控板也是电容压力传感器。此外，Valve Index 的侧边触控板还可以对单独的手指按压进行

批注 [刘7]: 输入和下面的输出都参考：

<https://cloud.tencent.com/developer/article/1123309> ,
https://www.sohu.com/a/125549843_223764。（上面两个都是对 Oculus 手柄的拆解，不过两边似乎都不完整）

批注 [刘8]: 参考：

<https://www.bilibili.com/video/BV1Zt411N7Tk?from=search&seid=7579867663494981893>。

追踪，因此它可以实现在游戏中对每根手指进行单独控制。这使得 Valve Index 允许大量的手势输入。

VR 眼镜

惯性传感器

一般指 3 轴陀螺仪加 3 轴加速度传感器。VR 眼镜中的惯性传感器用于实时判断用户头部的运动姿态，以保证画面与立体声的正确输出和设备定位。

摄像头

采用由内向外的图像识别来实现定位的 VR 眼镜需要利用摄像头来捕捉周围环境，以辅助惯性传感器实现设备定位；同时通过分析 VR 手柄发射出的红外光，完成对手柄的定位。

激光接收器（或红外接收器）

采用灯塔定位的 VR 设备需要激光接收器辅助基站实现定位。为保证 VR 眼镜能够实现定位，眼镜中也会植入激光接收器（或红外接收器）。

数据接收模块

以有线或无线的方式，从计算机中获取用于输出的视频流信息。

眼动检测模块

目前我所知的装有眼动检测的 VR 眼镜是小派的产品。通过在眼镜中内置眼动检测来实现优化的注视点渲染，保证在渲染计算量下降的同时，用户的使用体验能仍然保持良好。此外，用眼动装置也能帮助解决视觉辐辏失调（在下面的眩晕问题-认知原因中提到）的问题。

VR 设备中的输出元件

手柄

红外发射器

利用由内向外的图像识别原理的 VR 设备需要在手柄上放置红外发射器，借此 VR 眼镜才能通过红外摄像头实现对手柄的定位。



用红外摄像头观察到的初代 Oculus 手柄上红外发射器

震动马达

通过在手柄中放置震动马达，实现触觉反馈。这一功能实际上就是一般主机手柄的震动反馈在 VR 设备上的复制。

蓝牙模块

用于将手柄的输入信息传递给头显设备，再由头显设备直接实现信息处理或传入连接的电脑设备进行信息处理。

VR 眼镜

处理器

有一些 VR 眼镜（如 Oculus Quest）是无需外接计算机的，因此 VR 眼镜会将处理器内置，从而实现程序运行及图形显示所需要运算。但由于 VR 眼镜的设备大小和散热能力有

批注 [刘9]: 参考：
<https://www.bilibili.com/video/BV1L7411r7tL>。

限，因此此类设备为了实现设备一体化会牺牲部分显示效果。

透镜组

严格意义上说，透镜组并不是一个输出设备，它只是辅助显示器完成视觉输出的工具。

由于 VR 眼镜中的屏幕离人眼极近，正常情况下人眼是无法看清对这么近距离的东西的（晶状体变焦能力有限）。因此，要实现清晰的显示，必须通过透镜组，提前修正光线的路径，从而保证画面的清晰。

显示器

显示器可以说是 VR 中最主要的输出设备，同样，它也是 VR 设备目前遇到的诸多瓶颈的主要问题来源。

VR 眼睛是通过对人的左右眼显示不同画面，来实现 3d 显示的。

目前 VR 眼镜的显示器遇到的一大问题是显示分辨率问题。这里要先引入 PPD（Pixel per Degree）的概念，即每 1° 视场角中所包含的像素点个数。一般情况下，如果要实现视网膜屏幕，需要满足 60PPD（实际中其实 30PPD 也差不多可以满足），这等效于将 300PPI 的屏幕放于离眼睛 40 厘米的距离。但由于 VR 设备离人眼距离极近，屏幕的 PPD 会迅速下降，这导致人眼能够重新分辨出屏幕的像素颗粒。分辨率不足的问题就带来了纱窗效应的问题（之后再展开讲）。

从 PPD 的角度重新审视需要的分辨率，假设单眼的 FOV（Field of View，视角场）大小为 100°，配合 PPD 计算所需的分辨率就是 6000，双眼所需分辨率就是约 12k（注意，虽然 VR 眼镜中左右眼看到的内容是不同的，但其实一块屏幕）。而目前主流的 VR 眼镜屏幕大多只有 2k，最高的也不过 8k。

对显示器分辨率的高要求同时意味着对画面渲染的高要求、对视频解码的高要求以及对无线传输速率的高要求（后来我算了一下，12k 的 12hz 的视频数据传输量为

批注 [刘10]: 参考：
<https://zhuanlan.zhihu.com/p/36959531>。

$\frac{12000 \times 12000 \times 3 \times 240}{1024 \times 1024 \times 1024} = 48\text{GB/s}$ ，以无线速率 800Mbps 来看，还是游刃有余的。从量上看，无线传输不存在问题，但由于 VR 需要小于 20ms 的延迟，如何实现低延迟的传输才是无线传输的最大问题。如果要逐一实现这些要求，最终将导致 VR 设备本身成本极高，同时配套的硬件设备价格也过高。

音频输出模块

有些 VR 眼镜自带了输出的麦克风，有些则配备了耳机接口，用于实现音频信息的输出。立体声音效合成需要依赖 HRTF (Head Related Transfer Function)，即头部相关变化函数。

数据发送模块

用无线或有线的方式，将 VR 眼镜的运动状态和 VR 手柄的操作信息发送回电脑，使电脑内的程序做出对应的反应。

辅助设备

下面这块我想简单拓展一下，介绍一些现有的一些拓展设备。

全向跑步机

现有的 VR 体验仅依靠手柄和 VR 眼镜是无法提供优秀的移动体验的（大量的游戏都是固定玩家位置的），因此就出现了类似《头号玩家》电影中那样的能支持全向跑动的设备，使得用户在实际体验时能够用自然地走路方式实现在虚拟世界中的位移。

这类跑步机有两种设计思路，一种是让用户站在一个碗型的空间中，利用特质的鞋底带有滑轮的鞋子控制并记录用户的运动；一种利用双向跑步机，通过将一般跑步机的普通履带换为许多另一个方向的小履带拼合而成的履带实现双向运动。

批注 [刘11]: 之后有空可以展开一下。

关于 HRTF，可以参考苹果 airpods 的立体声功能的相关内容：

<https://www.bilibili.com/video/BV1qr4y1w7Lo>。

批注 [刘12]: 参考：

<https://www.zhihu.com/market/pub/119600010/manuscript/1084158757281468416>，
<https://www.zhihu.com/question/41906960>。

比较有名的全向跑步机有美国的 Virtuix Omni，价格在 26000 元左右。

VR 座椅

通过让用户坐在座椅设备中,使得用户在体验驾驶类的游戏时能够实时地感受到由座椅的旋转和震动带来的反馈。

这类设备在实际体验上效果还不错,可以被视为模拟驾驶类游戏在“方向盘+踏板+屏幕+座椅”的组合之外走出的另一条可行的道路。但设备本身的成熟度还不高,在安装、调试等方面的表现还不尽如人意。

批注 [刘13]: 参考:

<https://www.bilibili.com/video/BV1JK41157Ha>。

全身追踪装置

通常情况下,VR 设备只能实现对头、手以及部分手指的运动跟踪。但如果添加额外的追踪装置,就能实现对身体更多部位的追踪,如腰、腿、脚等。这样的额外追踪提供了额外的交互入口,使得更多的交互组合成为可能。

但这一设备的主要问题也是安装、匹配、校准的过程太过复杂。此外,设备本身大概是手掌大小,佩戴方式是条带状的魔术贴。如果能进一步缩减大小,应该能带来更好的用户体验。

批注 [刘14]: 参考:

<https://www.bilibili.com/video/BV1zZ4y1s7Rt>。

嗅觉输出装置

有一些 VR 设备还会配备面罩,实现嗅觉输出,但应用很有限,因此过多展开。

批注 [刘15]: 参考:

https://www.sohu.com/a/285951453_100023985,
<https://www.bilibili.com/video/BV1NA411479q>。



带有嗅觉面罩的 VR 眼睛。

环式屏幕

环形屏幕其实是一个脱离 VR 眼镜的虚拟现实展示方式。简单地说就是将用户放在一个的环形屏幕空间内，通过全视角的屏幕空间，实现虚拟现实化的体验。

这一做法目前已经被迪士尼应用在《曼达洛人》电视剧的拍摄中，用 270°的环形屏幕配合游戏引擎的场景渲染来替代以前的绿幕加后期合成。这一做法可以大大地减少场景搭建成本和因绿幕溢光而造成的后期返工问题。

批注 [刘16]: 参考：

https://www.bilibili.com/video/BV1ei4y1x7Em/?spm_id_from=333.788.videocard.0。

VR 所遇到的问题与解决

这一部分我会把问题和解决合到一起，先描述具体问题，然后直接给出解决方案。

眩晕

VR 的眩晕问题是一个很常见，但问题归因又很不明确的问题。

从硬件角度看，可以包括分辨率、屏幕刷新率、延迟等。

从认知角度看，可以分为动静失调、远近失调、视野大小失调等。

分辨率

眩晕的造成可能是由于分辨率不高，导致人眼能够分辨画面中的像素点。这一问题进一步会导致人眼看到的原本应该是平滑的线条会呈现出锯齿状。而在用户的头部发生移动时，对锯齿边缘的重复渲染会导致其锯齿形状由于采样规则的不同而时刻发生变化。这就是纱窗效应（人眼就好像在透过纱窗看东西一样）。

纱窗效应的存在就导致画面中有大量细节在时刻产生变化，而人类对于同一画面中的变化内容又有极高的敏感度，这就使得用户会不自觉的投入大量精力处理这些细节变化的信

批注 [刘17]: 纱窗效应：

<https://www.zhihu.com/question/39696826>。

息，进而导致疲惫。

对分辨率问题的解决，最好的做法就是之前提到的提高屏幕的分辨率。另一种可行的替代方案是使用注意点渲染技术，即只对视觉焦点处的图像进行清晰渲染，而对焦点区域以外的图像进行逐步的模糊，通过模糊来掩盖细节的变化。

屏幕刷新率与延迟

当屏幕刷新率过低（低于 60Hz）时，画面无法形成连贯的影像，进而导致用户在进行旋转或运动时视角出现大幅度跳跃（而非细小的、连贯的移动），产生眩晕。

延迟主要是指用户的操作与实际视频输出的延迟（其中包括数据处理延迟，信号传输延迟），即动态显示延迟（MTP，Motion To Photons latency）。VR 体验要求将延迟控制在 20ms 之内。过长的延迟会导致用户难以将自己的动作与视觉反馈同步，产生认知上的失调。

认知原因

VR 本身沉浸式的体验既是优点也是缺点。由于 VR 提供的只有视觉的回馈，这就导致用户在感受到视觉上的位置变化的同时，无法在身体的其他感觉器官那里同步地感受变化（这一点似乎正好与晕车的原理相反），这就造成了用户在认知上产生了矛盾，进而导致了眩晕。

此外，视觉辐辏失调（VAC，Vergence Accommodation Conflict）也是一个重要的原因。简单的说就是，在我们日常观察远近事物时，我们的晶状体会进行调节，以使得不同距离的物体都能够正常地在成像在视网膜上。但在 VR 眼镜中，实际上所有物体都是在屏幕上的，它们到人眼的距离是统一的。因此当人眼希望自然地通过调节晶状体去实现对远近物体的观察时，这种调节并不能真正地起到效果（VR 中所有的画面都是清晰的，不会因为观察近物导致远景模糊，也不会因为观察远景导致近物模糊），这就是视觉辐辏失调。这样的失调给用户带来了不适和疲劳。（不过说句实话，这个问题其实自有屏幕开始就一直存在。

批注 [刘18]: 视觉辐辏失调（VAC）：

<https://www.youtube.com/watch?v=YWA4gVibKJE>

。视觉辐辏失调的一些其他解决方法（看看就好）：

<https://zhuanlan.zhihu.com/p/83361657>。

我们在使用屏幕的过程中也会慢慢习惯这种全清晰的显示。)

另一点原因是视角上的，人眼通常的视角可以达到 200° ，但实际 VR 眼睛的视角只能做到 100° 。视角的缩小导致人眼所习惯的视野内的清晰、模糊的分布产生了变化，视野外的暗区也使用户产生了不习惯的感受。

从认知出发解决眩晕问题

对于动静反馈所带来的认知问题，可以通过添加额外的辅助设备(如之前提到过的座椅)来进行弥补。

视觉辐辏失调问题可以通过在 VR 加装眼动装置，通过检测眼动判断目前眼睛的聚焦点，进而对画面进行部分的模糊处理。此外，还有利用光场显示等技术来尝试解决的，但技术都不成熟，因此不多展开。

视角的问题可以通过增大显示屏幕来实现。国内的 VR 厂商小派所推出的产品就是以极广的视角和极高的分辨率为主要卖点的。从实际用户的体验来看，加大 FOV 确实能够给用户带去更加优化的体验。

批注 [刘19]: 参考：
<https://www.bilibili.com/video/BV1Tx41177UP>。

从交互角度解决眩晕问题

眩晕问题的两个重要原因是用户频繁地进行转动及场景本身会产生以用户为中心的高速运动(简单的说就是发生了快速的视角变化或视野内容变化)，因此通过减少需要用户发生头部转动的交互操作，减少可能会导致眩晕的运动场景或降低运动速度，也是解决眩晕问题的一种思路。

设备体积与价格

随着近些年 VR 眼镜都逐渐无线化，VR 眼镜以前由于数据线带来的障碍已经不复存在。但作为一个可穿戴设备，VR 无论是体积还是重量都还是过大了一些，设备本身的佩戴会对

用户带来一些负担。此外,依靠灯塔技术的 VR 设备还需要对基站进行布置。使用一般的 VR 设备还必须由一台足以让 VR 流畅运行的电脑。这一系列硬件的价格成本和使用成本都很大程度上限制了 VR 的进一步普及。

而这一问题的解决也是难以一蹴而就的。设备的微型化依赖屏幕技术的革新,设备价格的下降更是整个硬件行业都在努力的方向。

软件数量

目前 VR 的主流应用还是在游戏层面,且并没有什么仅在 VR 平台发售的 3A 大作。VR 的普及率问题本身成为游戏开发者不敢踏足这一领域的一大担忧,而缺少优秀的游戏作品又使得 VR 设备本身难以进一步获得对用户的吸引力,形成恶性循环。

VR 的软件问题本质上其实还是 VR 设备的成本问题。简单地说,现在的 VR 有点类似当年的第一代 iPhone,以革命性产品的身份登场,但却缺少足够的应用支持。但除非 VR 设备能证明自己将是下一个智能手机,或 VR 设备的生产价格大幅下降,否则 VR 产业的问题将一直持续下去。

VR 的拓展

这一部分会简单介绍一些以全景视野为核心的产品或技术。

所谓全景,就是通过将多个摄像机拍摄同一个位置的周围信息,将所得到的信息经过裁切、扭曲、匹配后组合到同一个大平面内,然后将平面映射到一个球体的内侧。用户可以通过移动视角,观察当前位置的各个方向的视野。

全景相机

全景相机可以分为单摄像头和多摄像头的。由于全景的核心是裁切和匹配,因此在位置

不变的条件下，只要通过转换摄像头的角度，拍摄多张（通常为 26 张）不同角度的照片，在进行集中合成即可。目前的无人机全景拍摄采用的大多都是这种方式。但这一方法的一个巨大弊端就是它只能拍摄全景照片，无法拍摄全景视频。

如果需要全景视频的拍摄，则需要多摄像头的全景相机。通过多摄像头同步记录全景信息，再进行合成，即可实现全景视频的拍摄。

VR 应用

批注 [刘20]: 参考：
<https://www.zhihu.com/question/40479545>。

游戏

游戏可以说是目前 VR 集中发展的领域。但当前的 VR 游戏与电影《头号玩家》中所描述的那种游戏体验还有很大差异。目前的 VR 游戏还停留在轻量化的、休闲的游戏领域，但 VR 所提供的真实场景确实为游戏的发展提供了极大的可能性。

职业培训

飞行员、医生等职业可以通过在 VR 中模拟现实工作场景来进行职业技能练习。目前飞行员已经有利用一般的屏幕和其他输入设备进行飞机操作练习的模拟，相信在于 VR 结合后这一模拟的真实度能得到更好的提升。

另外，听说美军已经在新兵训练时引入了 VR 设备以实现对战场的加强模拟了。

艺术创作

艺术创作中，最能从 VR 中获益的莫过于 3d 建模中的细节雕刻环节。以往的建模都是在屏幕上通过旋转摄像机视角实现建模的，在 VR 的加持下，艺术家能更直观地在空间中感受到艺术作品的存在形式。

教育

3d 化的教育资源能使原先只能从书本上获取 2d 信息的我们对事物产生更直观的认识。

旅游与商业

目前在许多租房售房网站上已经出现了 VR 看房的功能。其本质就是房间不同区域中的全景照片的组合。与之类似的还有更早的百度全景地图。

在未来，也许 VR 设备会成为人类出游或实地考察的一种新方式。在 5G 实时信息传递的条件下，实现足不出户、游遍全球的梦想。

零散内容的补充

定时锚点

这里对老师提到的 VR 的交互中的内容进行一下补充。

在进行 VR 体验时，有时会遇到需要点击按钮的问题。如何实现按钮的点击是交互方式上的一个重要问题。如果只通过手柄的按键操作，会导致效率过低的问题；如果试图用手柄去触碰空间中的按钮，会导致交互不便利的问题（这其实就是目前 hololens 遇到的一大交互问题）。

现在采用的解决方案就是定时锚点方案。简单的说就是构造一条以手柄为起点，手柄所指的位置为终点的辅助线，帮助用户实现按钮的点击。



图中的蓝线就是用于确定按键位置的辅助线

VRS

VRS，即 Variable Rate Shading，可变速率着色。这一技术允许开发者在画面渲染时降低部分画面渲染精度，以此加快图形渲染的速度。在正常情况下，着色器必须对每个像素对应的材质进行采样，才能确定对应像素所要呈现出的颜色。如果将采样的单元从每个像素增加到每四个像素进行一次采样，计算所需的时间就缩短了 3/4。

由于 VR 眼镜中屏幕显示的画面有一部分会因为透镜组的折射而变得不那么清楚，因此，在实际渲染时就可以将这部分的画面进行简化渲染，以此减少渲染负担，降低延迟。

之前提到过多次的注视点渲染技术本身也是以可变速率着色为基础实现渲染速度的提升的。

关于头部的旋转

在肩膀和腰部都不发生转动的情况下，头部能实现的转动是，垂直上 60°、下 40°，水平左右各 55°。舒适的转动角度为，垂直上 20°、下 12°，水平左右各 30°。

关于交互距离

一般我们认为 3D 的交互分为前景、中景、远景三层。

批注 [刘21]: 参考：

<https://news.mydrivers.com/1/643/643269.htm> ,
<https://www.bilibili.com/video/BV1kE411k7Xk?from=search&seid=5385439422864517304>

,
<https://www.bilibili.com/video/BV1xk4y1z9?from=search&seid=5385439422864517304> ,
<https://zhuanlan.zhihu.com/p/73271446>。

前景距离为离人 0.5m-6m，一般的控制元件、操作要素等都应放在这个距离之内。

中景距离为离人 6m-20m，场景元素通常被放在这个距离之中，用户可能需要一定的移动才能与这些元素进行交互。

远景距离为离人 20m 以上，通常是用户无法接触的大背景。

AR

核心技术为三维追踪、三维注册、显示技术、人机交互。

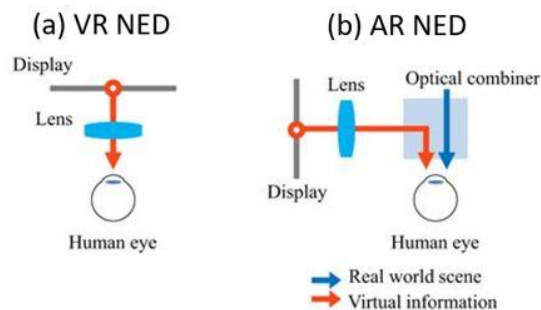
其中的三维跟踪是指通过场景中的自然特征确定空间坐标系的过程 ;三维注册是指将虚拟场景准确的定位到真实环境中的过程。

目前的 AR 体验可以根据是否通过摄像头给出现实世界的影像分为两类。依靠摄像头给出现实世界影像的 AR 包括手机、平板电脑等设备中的 AR 应用 ;不依靠摄像头 ,用光波导或光场技术将虚拟影像投射在可穿戴设备的镜片上的 AR 包括谷歌的 google glass、微软的 hololens、magic leap (不过一般 hololens 和 magic leap 也会被认作是 MR 产品) 等 , 这类 AR 体验一般都需要特定的设备来实现。

目前来看 , 直接穿戴式的 AR 设备所做的基本只有让设备在既能看到外界环境的同时 , 也能看到一些具体的平面图形界面(google glass)。在 hololens 2 和 magic leap 发售后 , 穿戴式 AR 设备也逐渐能够真正意义上实现虚拟与现实交互的功能了。

AR 中的显示技术

这里主要讨论不依靠摄像头获取现实世界的设备是如何将虚拟影像传递给人眼的。由于 AR 眼镜需要实现对真实世界的透视 , 因此需要一个或多个光学组合器 , 以层叠的方式 , 将虚拟信息与真实场景融为一体。



VR 与 AR 的显示差别，由于 AR 要能够直接透视到现实世界，因此屏幕不能放在眼睛的正前方

AR 设备的光学显示系统通常由微型显示屏和光学元件组成。

目前市面上的 AR 眼镜主要采用的显示系统就是各种微型显示屏和棱镜、自由曲面、BirdBath、光波导等光学元件的组合。

微型显示器可以选择如 OLED 或 micro-LED 这样的自发光二极管面板，也可以选择 LCD 或 LCoS 这样的需要外部照明的液晶显示屏，还可以选择基于微机电技术的 DMD 或 LBS。(不重要)

显示器（投影设备）

LCoS 即 Liquid Crystal on Silicon 硅基液晶 ;DMD 即 Digital Micromirror Device , 数字微镜器件 ; LBS , 即 Laser Beam Scanner , 激光束扫描仪。由于微型显示器的选择不是特别重要，且我暂时没有在网上找到很好的综合介绍，这里暂时先略过。

光学元件

光波导

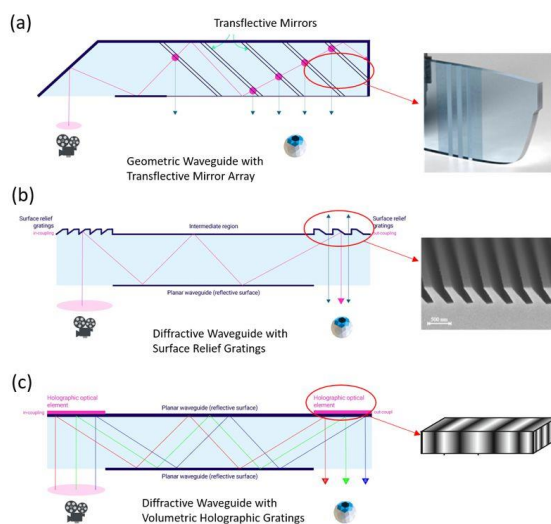
光波导的过程可以简化地描述为，视野以外的微型屏幕（或投影仪）中的像在玻璃（或其他透明材质）中发生全发射后移动到眼睛前，在反射面或衍射光栅的作用下离开玻璃，最终被人眼捕获。

批注 [刘22]: 参考：
https://www.sohu.com/a/326504900_99973031 ,
https://www.youtube.com/watch?v=q_KaZtDVGag .

在这一过程中，镜片通常都是平行于眼睛的视线的，因此核心的技术是如何让在平行玻璃中不断发生全反射的光离开平行玻璃并进入人眼。目前这一问题有两种解决方案，几何光波导和衍射光波导，其中衍射光波导是目前的主流。

在具体展开将两种光波导之前，还有一个问题要补充。由于 AR 眼睛设备大小一般有限，因此不可能在眼镜中安置巨大的屏幕。这就造成一个问题，即如何将极小的屏幕在光波导的过程中进行放大，从而使得用户能够看清屏幕上的内容。这一问题需要使用扩瞳技术(EPE , Exit Pupil Expansion)。

由于光波导一般都涉及复杂的光学元件，因此价格较高。



(a)图表示几何光波导，(b)(c)图表示衍射光波导

几何光波导

几何光波导使用嵌入到玻璃基底中、并且与传输光线形成一个特定角度的多个半透半反的镜面阵列来将全反射中的光线反射出镜片。半透半反就是指当光线通过时，一部分会被反射，一部分会经过投射继续在玻璃中前景。

之所以需要镜面是半透半反，就是因为需要对光线进行扩瞳处理。如果只有一个反射镜

面，则原先进入时宽度为 4 毫米的光束，在输出（出瞳）时也只能在 4 毫米的范围内被看到。而当镜片中存在多个平行的半透半反镜面阵列时，光线在每次接触镜面时会部分被反射，部分继续进入后一层镜面。当光线在多个镜面都被反射并出瞳时，光线的整体宽度就被延伸了。这样看貌似是让一个光点变成了多个重影，但实际上这些平行的光波最后都会在视网膜上相同位置上成像（这一块其实我也没特别搞懂）。

几何光波导的优点是，它运用传统的几何光学设计理念，并没有任何微纳米级结构，图像质量可以达到很高的对比度。

几何光波导的缺点是，虽然它没有微纳米结构，但实际生产中半透半反的镜面阵列生产流程复杂（举个例子，由于镜面要求半透半反，因此如果希望保证所有反射光的强度均匀，则不同镜面的反射透射比都要不同），总体良品率很难提高，任何一道工序的失败都可能造成背景黑色条纹、出光亮度不均匀、鬼影等问题；半透半反的镜面在关闭 VR 显示功能时仍然能够在镜片上被清楚地看到，影响了一般情况下的使用；一般只能进行一维的扩瞳，不能实现二维扩瞳，即像只在长度上被拉伸，无法在高度上被拉伸。

衍射光波导

衍射光波导，即在光线需要出瞳的位置对玻璃进行处理，在玻璃表面布置均匀的衍射光栅（可以是在表面雕刻出高峰和低谷，也可以是在材料内部形成明暗干涉条纹），利用光栅的透射式衍射使部分全反射光线发生特定角度的折射离开玻璃进入人眼。

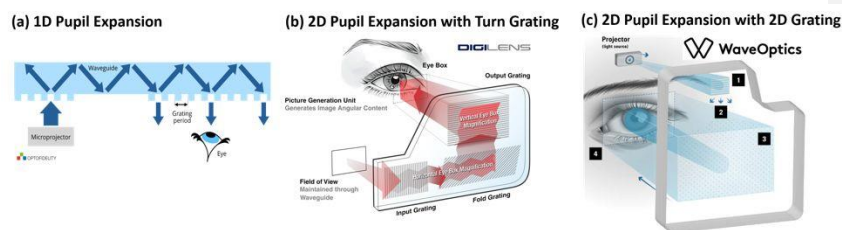
如上面的描述所说，光栅的透射式衍射只会是部分光线按特定的角度折射离开，剩余的光线将继续在玻璃中进行全反射，在下一次接触到光栅表面时，再次发生部分折射。这样的过程与之前几何光波导中的“半透半反”起到了相同的扩瞳的目的。

衍射光栅的分光能够根据光线的入射角和光栅的周期决定发生衍射的角度，在通过设计光栅中的其他参数（材料折射率、光栅形状、厚度、占空比等），可以将某一角度的衍射效

率达到最优化,使大部分光线在衍射后向这个方向传播。衍射光栅既可以用于反射,也可以用于折射。

另外,由于 AR 眼镜的进光处也会设置一个衍射光栅,而衍射光栅的折射角度与波长相关。因此如果让混合光直接进入同一个衍射光波导元件,不同波长的光会发生不同情况的全反射,最终输出也将难以统一,导致色散的问题(彩虹效应)。为了避免这种情况的发生,通常会将 RGB 三种颜色的光分别通过三层不同的衍射光波导元件。但由于即使光线颜色统一,光线内部的波长也不能完全保持统一,因此彩虹效应的问题只能得到部分的改善和优化。如果将 LED 光源换成光谱极窄的激光光源,将能够很大程度地减少彩虹效应。

之前我们提到几何光波导一般只进行一维扩瞳,在衍射光波导中,是可以实现二维扩瞳的。具体做法是,将一维扩瞳得到的光线经过转折光栅调整方向,进入另一个方向上有光栅的玻璃中(其实按同样的方法,放置两块几何光波导元件也是能实现二维扩瞳的,只是可能效果会很不好,所以参考中的笔者直接说不能实现二维扩瞳)。



衍射光波导中的扩瞳实现

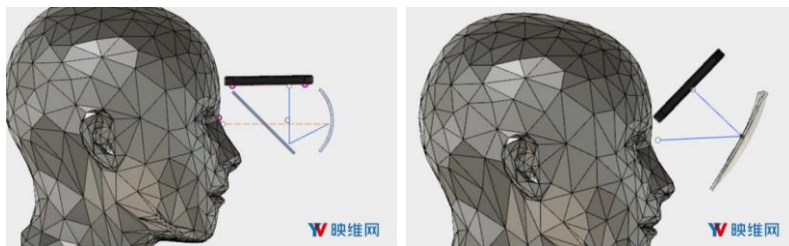
衍射光波导的优点在于,它通过二维光栅实现了二维扩瞳,使像能够在两个方向上都进行拉伸扩展;虽然制作光栅设计微纳米工艺,但其量产型和良品率都比几何光波导要高。

衍射光波导的缺点在于,它会出现之前提到的彩虹效应,导致视野内的红绿蓝三色光分布比例不同。

Birdbath、全息反射薄膜、自由曲面（自由曲面由于其设计外形独特，也被称为虫眼）

这三者的本质其实都是曾经提到过的佩珀尔幻象，即利用一块既有反射又有透射能力的镜面，利用其透射观察现实世界，利用其反射将其他位置的屏幕显示内容“投射”到镜片上。

Birdbath 和自由曲面的原理示意图如下。



左图为 birdbath 的原理，右图为自由曲面的原理



左图为 birdbath 设计，右图为自由曲面设计

这类 AR 显示技术的有点在于其原理简单，造价便宜。主要的问题在于设备体型通常较大。

这里我想稍微对一个应用全息反射薄膜的设备进行展开。由加拿大 North Focals 公司出品的智能眼镜将投影仪放置到眼镜的镜腿中，通过全息反射薄膜制成的镜片反射投影仪画面，实现 AR 显示。

之所以想特别提一下它，是因为这是目前我看到的在外形上最为小巧且最像眼镜的设备。但该设备实际的显示区域非常有限，且价格较为昂贵（首发 999 刀，目前是 599 刀）。该公司目前被谷歌收购。

批注 [刘23]: 参考：

https://www.sohu.com/a/233248382_213766。

North focals 参考：

<https://www.shenzhenware.com/articles/12980> ,
<https://www.bilibili.com/video/BV1rb411a7GL?from=search&seid=18013944728469022049>。

North focals 原理参考：

https://www.sohu.com/a/271595098_213766。



North focals 的产品图及实际佩戴效果

AR 的三维跟踪技术

AR 的三维跟踪技术有 SfM (Structure from Motion , 运动结构恢复) 和 SLAM (Simultaneous Localization And Mapping , 即使定位与地图构建) 等。其中 SfM 是相对更古老的方法。

SfM

SfM , 即 Structure from Motion , 运动结构恢复。其具体实施操作如下 : 对特定对象或空间拍摄若干张角度不同的照片 ; 用 SIFT 或其他匹配算法对这些图像中的不同特征点进行

批注 [刘24]: 参考 :

<https://www.bilibili.com/video/BV1Nk4y127qX?p=4> ,
https://blog.csdn.net/mahabharata_/article/details/70799695 ,
<https://www.bilibili.com/video/BV1Yb411L7vM?from=search&seid=3788227736947901034> ,
<https://zhuanlan.zhihu.com/p/29845703> ,
<https://blog.csdn.net/Z5122/article/details/103201716/>。

行匹配；通过匹配的特征点及其各自在像素中的位置，求解像素点对应的空间中的点的空间位置方程，继而求出空间中的点实际的空间坐标位置。

SfM 通过对大量图片中的大量特征点进行求解，确定每个特征点实际在空间中所处的位置，进而形成对特定对象或特定空间的三维重建。

SLAM

SLAM，即 Simultaneous Localization And Mapping，即时定位与地图构建。当设备被放置在未知的环境中时，会通过 SLAM 实时进行周边环境三维信息的**增量式构建**（不断扩大三维信息构建，且保留并修改之前已构建好的三维信息）。同时，设备本身会通过建立的三维信息进行空间定位。凭借对周边三维信息和自身位置信息的掌握，设备可以在空间内实现无障碍的移动。

主要应用场景

SLAM 目前的主流应用场景有应用和建模两种。

SLAM 的定位功能被广泛应用于智能驾驶、机器人、无人机等领域。之前我们曾经提到，一般通过 GPS 加惯性传感器就能解决大部分的定位问题。SLAM 之所以有其用武之地，主要有以下几个原因：1.民用的 GPS 的精度不够；2.GPS 只能在能接收信号的场所使用，地下车库、洞穴、建筑内部、水下、外太空（月球、火星）很多时候都无法良好地接收 GPS 信号，长时间处于这种环境下的设备无法依靠 GPS 实现定位校准；3.机器人或机器臂等设备可能会出现多轴向的复杂运动，这种运动需要高精度的实时校准。

目前 SLAM 最广泛的商用领域是扫地机器人。扫地机可以在室内打扫的同时不断构建室内的三维模型，从而对扫地路线进行更精细的规划。这一应用的延伸是物流仓库中的无人搬运车（AGV，Automatic Guided Vehicle），通过大量的依靠 SLAM 定位的无人机器人实现货物在仓库内的取送。

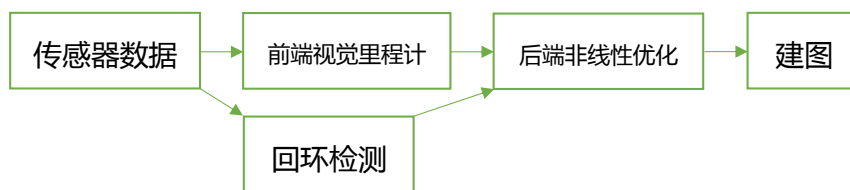
批注 [刘25]: 参考：

<https://www.bilibili.com/video/BV1C4411j7St?from=search&seid=2685064696683295047>，
<https://zhuanlan.zhihu.com/p/23247395>，
<https://www.bilibili.com/video/av22600278/>。

SLAM 的建模功能主要被用于进行实际物体或大型场景的 3D 模型重建。SLAM 在建模的应用思路与在定位的应用思路类似。这一功能的主要应用场景包括 3D 扫描、AR 程序等。

主要流程

SLAM 的主要流程可以分为前端（视觉里程计，VO，Visual Odometry）、后端（非线性优化，Optimization）、回环检测（Loop Closing）、建图（Mapping）。



SLAM 主要流程图

前端（视觉里程计），指通过分析不同时刻传感器获取的图像（或图像加深度信息）去反向求解不同时刻的相机位姿。这一过程通常使用匹配特征点实现求解（PnP 问题，Perspective n Point）；但也有不使用特征点的直接求法（ICP 问题，Iterative Closest Point）。直接法实际是对两个时刻图像间像素进行全局匹配，对计算量和图像采集率有更高要求。

后端优化问题旨在消除视觉里程计中产生的误差，使得之前建立的三维空间信息更好地拟合实际物体或环境。

回环检测是指，当设备再次回到之前走过的场景时，它能够认出这是之前到达过的场景。由于三维信息在生成的过程中会发生误差的累计，因此在重新回到之前到达过的场景时，需要对新建的三维信息与原始的三维信息进行比较、复位，同时修正之前运动过程中产生的一系列累积误差。

建图是指形成视觉化的三维场景。这一步骤在定位中不是必须的，只在某些需要进行场景还原的场合进行。

主要传感器

SLAM 使用的传感器可分为激光和视觉两大类。就视觉传感器而言，目前视觉 SLAM 主要分为三大类：单目摄像机、双目摄像机、RGBD（深度摄像机）。从实现难度区分，视觉 SLAM 的难度从高到低为：单目、双目、RGBD。

激光雷达

雷达相关的内容我在之前的“感知与传感器”中已经详细说过了，这边只会讲在 SLAM 中的应用。

激光雷达是最古老的、也就最多的 SLAM 传感器。使用激光雷达的主要目的是为了获取周围环境的障碍点与当前设备之前的距离信息。由于一般的设备（如扫地机）只在平面 2D 的范围内进行活动，因此通常会使用目数较少的 2D 激光传感器，用于扫描一个平面内的障碍物。在无人驾驶中使用的激光雷达为了获取更全面的 3D 场景信息，会采用多目激光雷达，实现对周围 3D 环境的全范围扫描。

激光雷达的有点是精度高、速度快、计算量不大，但其缺点是激光设备的价格过高。

单目摄像头

由于单目摄像头无法获得深度信息，因此单目 SLAM 想要对确定深度信息必须依靠运动中的三角测量。如果设备始终保持固定或原点旋转，单目 SLAM 将无法实现三维信息构建。不过好在一般情况下，日常使用 SLAM 时都会发生平移和旋转。

此外，由于单目摄像机在运动中获得了深度信息，这个深度信息也只是相对深度。只有确定了实际的运动距离，才能将相对深度明确为绝对深度。如果某物体到单目摄像机的距离为 A ，摄像机运动距离为 B ；如果将物体到摄像机的距离调整为 $2A$ ，摄像机运动距离改为 $2B$ ，且运动轨迹不变，则理论上这两次运动中物体在摄像机拍摄出的图像中的位置都是一致的。换言之，除非确定了移动距离，否则摄像机到物体的距离是无法具体确定的。

单目摄像机在 SLAM 的视觉里程计中可以使用稀疏图（特征点）方法，也可以使用稠密图（像素匹配）方法。稀疏方法包括：PTAM、ORB-SLAM 等方法。其特点是对画面中的几何一致性敏感（由于是对特征点进行比较，因此受光照影响小），计算量小。稠密方法包括：DTAM、LSD-SLAM。其特点是计算量大（虽然不需要提取特征点，但计算对象变多），对运动模糊抗性较好（模糊情况下，相同的特征点可能失配，但像素就不会），对画面的光照一致性敏感（由于是直接比较像素，因此光照发生变化也会导致像素发生变化，因此受光照影响大）。

双目摄像机

由于双目摄像机可以通过多个摄像机的基线估计空间点的位置，因此双目摄像机可以通过计算获得空间中某点的精确深度信息。这使得双目摄像机在静止时也能进行深度估计。

双目摄像机对深度的判断只受画面分辨率和芯片算力的限制，因此可以适当承担室外环境下的工作。

双目 SLAM 的主要方法包括：stereo LSD-SLAM、RGB-SLAM2 等。

深度（RGBD）摄像机

SLAM 中采用但一般都是红外深度摄像机，原理既有利用结构光的，也有利用 TOF 的。与双目摄像机一样，由于可以直接获取深度信息，因此 RGBD 摄像机也可以直接获得图像中各像素点到相机的距离。由于 RGBD 摄像机是直接获取深度信息的，因此不需要像双目摄像机那样通过复杂的图像算法去计算深度。

目前 RGBD 摄像机还存在测量范围窄、噪声大、视野小等问题，因此主要用于室内 SLAM。

深度摄像机 SLAM 的主要方法包括：RGB-D-SLAM、Kinect Fusion 等。

手机中的 AR

批注 [刘26]: 参考 :
<https://www.bilibili.com/read/cv6612215>。

在 2016 年前后，VR 和 AR 的概念正在风口，彼时市场上迸发出了大量的 AR 应用。应用采用的做法一般都是以摄像头拍摄的画面为背景，仅以陀螺仪检测的运动信息为线索进行 3d 空间建立。这种做法由于没有视觉信息作为校准，在应用使用过程中位置误差会迅速累计，因此体验一般都不佳。作为补偿，类似 Pokemon Go 或 Ingress 这类的游戏通过像现实地图中引入虚拟占位点和派系关系的方式，从地图层面实现另一角度的现实增强。

在 2019 年苹果发布 mac pro 时，曾使用 AR 应用对其自身的商品进行产品 3d 展示，使得用户能够通过 AR 应用现实的看到 mac pro 机箱内部的设备构成。

随着 iPhone12 的上市，通过 iPhone 自带的 LiDAR 传感器检测深度，AR 应用能够实现准确的现实测量、空间还原、实时 3d 建模等越来越多的功能。

在手机摄像头数量逐渐增加、深度传感器越来越成为主流的今天，AR 应用将逐渐由之前的简单、粗糙的体验品，慢慢转变为实用、有趣的工具。

AR 的应用场景

批注 [刘27]: 参考 :
<https://www.zhihu.com/question/40099190>。

AR 的应用与 VR 其实有很大程度的重合。

HUD

HUD，即 Head Up Display，平视显示器。这是最为传统的 AR 应用场景，通过将辅助信息投射到与视线平行的透明显示装置上，使得用户可以避免因低头检索信息而无法关注眼前的现实信息。这一类产品的具体应用包括飞行员的 HUD 显示头盔、google glass 的导航功能等。

展示

在之前“人机交互接口”中，曾经提到过小米退出的透明显示器。通过将透明显示器与商品相结合，使用户在观察到实际产品的同时也能看到由显示器提供的产品信息，进一步扩展了商品展示的形式。

此外，在博物馆、展览会等场所，也可以通过 AR 配合实际展示模型的方法构建数字沙盘，进一步提高展品的互动性和可玩性。在 2019 年 MIT 宣传动画中也出现了配合 AR 程序，使原本静止的绘画产生动作的应用。

利用 AR 测量提高物流效率

Ebay 通过开发自动化物流系统，使用 AR 测量货品尺寸，智能选择快递包裹，提高了物流环节的效率。

线下商品体验

在线上购物极其发达的今天，人们在购买前对商品的信息却仍然只停留在图片、文字、视频的范围。当我们需要购买如旅行箱、家具等商品时，商品的现实大小、商品与周围环境的契合程度等要素都只能通过大概的估算和想象。京东目前已经提供 AR 模型展示的功能，通过将需要购买的商品用 AR 的手段放置在现实环境中，给消费者提供更感性、更直观的产品认识。

各类美颜、美图软件

包括苹果的 Memoji 在内的各种美图、美颜、滤镜软件现在大多都支持在视频中实时地添加滤镜效果。

AR 宠物

这一块是在我看到一个叫《[Hyper Reality](#)》的短片后想到的。短片中有个场景时主角进入超市后，有一个电子宠物来到主角身边，只要主角在超市期间持续“照顾”这个宠物，就能获得额外的优惠。我在想，是否可以利用这个电子宠物的要求，引导用户经过在超市的重要货架区，以此提高热门商品的购买率。

MR

目前的 MR 其实主要就是指类似 Hololens 或 Magic Leap 这样的 AR 眼镜设备。从本质上说它们其实和 AR 没有太大区别，或者说他们这种设备才是所谓的 AR（google glass 只是简单的光波导显示器）。下面说道 MR 设备，不出意外一般都是特指 Hololens。

MR 设备目前在交互方面还有很大改善余地。由于 Hololens 不带手柄，因此交互只能通过手势识别完成，然而目前的手势识别还无法做到十分精细；此外，让用户熟悉产品的手势集也是一个比较违反交互原则的做法（让设备适应人，而非人适应设备；顺带一提 hololens 从一代升级到二代的时候还做了手势集的修改）。从现实使用来看，Hololens 的手势交互很多时候会让用户做很多例如前移、前倾等不必要的多余动作（交互元素的距离离人太远），交互过程给用户的负担较重。除手势交互外，Hololens2 内置了眼动仪，允许用户通过眼动与设备进行交互，从一定程度上减轻了用户因手势交互而产生的沉重负担。由于 Hololens 本身就内置了 win10 系统，因此除了手势和眼动外，还可以通过外接蓝牙设备实现输入。

Hololens

Hololens 本身就是一个内置了 win10 系统的便携式设备，因此它本身就可以执行原先在 windows 上的大部分功能。从目前的应用来看，Hololens 并没有一款专属于该平台的应用，它更多地是让用户体验了信息与虚拟世界相结合的感受。

目前 Hololens 的应用场景主要是企业级的应用，且集中在某领域的专家与实际工人需

批注 [刘28]: 参考：

<https://www.bilibili.com/video/BV1KK4y1k7s7?from=search&seid=8844844433421413287>。

要就发生的故障或问题进行实时沟通的情况下。此外就是教育和培训的场景 ,利用 Hololens 的虚拟影像引导学生或工人学习某些较为复杂的安装、拆卸流程。就目前看来，这些场景是最适合 Hololens 所提供的信息与现实同步利用的特性的。我个人觉得 ,在一个不短未来中，Hololens 都将持续地以专业设备而非大众设备的身份继续存在。

Hololens2 的构成

Hololens2 的主要显示部分是量块光栅透镜，用于将 RGB 三种颜色的激光导入入眼中（一块用于传输红光和绿光，一块用于传输蓝光和绿光）。Hololens2 采用的光源是 MEMS Laser Display。

设备有多个摄像头（普通 RGB 摄像头和红外深度摄像头都有），用于检测周围环境进行 SLAM 定位和手势识别。此外，设备还配备了 IMU 用于检测用户的头部运动。

Hololens2 将声音处理、全息显示、信息处理等多项功能的计算和处理都集成在一个 ASIC 特殊继承电路中。

Hololens2 在机身周围布置了若干个麦克风用于语音操作，同时在耳朵上方各布置了一个扬声器用于提供音频反馈。

眼睛在鼻托的位置还放置了用于检测眼动的摄像头组。

两代 Hololens 的区别

	Hololens1	Hololens2
FOV	30°	51°
PPD	23	47
眼球追踪	不支持	支持
手势跟踪	数量不多，且仅为简单手势	支持单手 25 点跟踪 ,可以捕捉细节动作

批注 [刘29]: 参考：
<https://www.bilibili.com/video/BV14s411z7eo?from=search&seid=12646533083432531948> ,
https://zhuanlan.zhihu.com/p/99808353?from_voters_page=true。

持续工作时间	2 小时左右	3 小时左右
显示光源选择	LED+LCoS	MEMS Laser Display(激光在光波导中更不易出现彩虹效应)
重量分布	总重 579 克，几乎全都集中在前侧。	总重 566 克 前后分布均匀。
是否支持不使用时将镜片上掀	不支持	支持

对 MR 的一些看法

目前的 MR 设备其实本质上都只是 AR 设备，他们所做到的只是将虚拟的影像投射到现实之中，但并没有真正实现通过与虚拟的交互影响现实的场景。换言之，他们只将现实融合进了设备，但没有将设备再次融合进现实。

MR 作为融合虚拟与现实的下一个台阶，应该让我们真正意义上地实现现实与虚拟的双向互动。当我进入空间时，虚拟的影像能为我提供空间中的信息，当我改变虚拟的信息时，空间中的实体也应该对应地产生变化。

以智能家居为例，在 MR 眼镜中，当我的视线扫过智能家居产品时，它的信息应该能被我所访问。而当我用手势与其互动时，我可以通过上下调动的手势直接改变智能家居的某些参数（如音箱的声音和灯的亮度等）。

现有的 MR 确实是融合了光波导显示、SLAM 定位、芯片制造等等复杂技术综合产品，它的的确确走在了科技的前沿。但真正的 MR 的重点也许更应该在它与现实之间的交互之上。

VR、AR、MR 的总结

这里按照老师给出的表格，再简单的对三者进行一下区分。(这里的 MR 概念还是传统意义上的 MR，不是我在之前提到的我认为的 MR。)

	VR	AR	MR
定义	用户完全沉浸在计算机生成的虚拟环境中,是很大程度上与现实隔离的封闭式体验。	用户直接或间接观察真实场景,数字元素叠加到现实世界的对象和背景上。	讲现实场景和虚拟场景非常自然地融合在一起,他们之间可以发生实时交互。
关键要素	<div>1. 沉浸感。营造出身处虚拟场景内的感觉。</div> <div>2. 交互性。用户可以和虚拟场景中的内容发生实时交互。</div> <div>3. 想象性。可以根据设计者的想想设计出各种虚拟场景。</div>	<div>1. 现场感。直接现实真实世界现场。</div> <div>2. 增强性。对现场显示的内容增加图像、视频、声音等其他信息。</div> <div>3. 相关性。增加的内容和现场位置、内容、时间等方面有相关度。</div>	<div>1. 现场感。真实场景来自现场。</div> <div>2. 混合性。真实场景和虚拟场景自然地合在一起,发生真实的交互,如遮挡、碰撞等。</div> <div>3. 逼真性。虚拟场景的显示效果接近真实场景,不易辨别。</div>
延迟要求	小于 20ms	无特别要求	无特别要求
FOV	100°	手机设备无要求,穿戴设备要求与 MR 类似。	50°
应用举例	VR 游戏	AR 实景导航	MR 装修

典型产品	HTC Vive、Oculus、 Valve Index	Google glass	Hololens 、 Magic Leap
------	---------------------------------	--------------	--------------------------