



# VIRTUAL REALITY

Steven M. LaValle

(中译版)



## 前言

《Virtual Reality》的中文译版

英文原版: <http://vr.cs.uiuc.edu/>

历经前前后后三个多月, 《Virtual Reality》的中文翻译和校验的工作终于基本完成。此书英文原版作者为 Steven M. LaValle, 他是伊利诺伊大学的教授, 同是华为科技有限公司VR/AR/MR 首席科学家。此外, 还要感谢大朋VR的陈玉琨先生为本书修改提出了诸多宝贵的建议。目前市面上 VR/AR 方向的中文好书较为稀缺, 而近年来 VR 产业发展迅猛, 我们特翻译此书来让读者更早些更方便的学到相关的知识。

《Virtual Reality》这本书涵盖了虚拟现实系统的基础知识, 包括几何建模、转换、图形渲染、光学、人类视觉、听觉和前庭系统, 跟踪系统, 界面设计以及一些技术问题等。读者可以自取所需。

尽管翻译版本还有很多问题, 但至少大部分内容是可读的, 并且是准确的。我们在翻译的过程中尽可能地保留了原书中的意思。由于能力有限, 仍然还会有一些影响阅读的小错误, 需要读者的建议和帮助, 一起减小翻译的偏差。读者可以以中文版为主、英文版为辅来阅读学习, 但我们仍建议研究者阅读原版。

翻译与校验人员: 上海交通大学媒体技术实验室(<http://medialab.sjtu.edu.cn>)

宋利: 审核与指正所有章节翻译内容, 负责定稿校验工作。

罗莹: 负责所有章节的一部分翻译, 与第 1-5 章的校验工作。

刘煦: 负责单数章节的一部分翻译, 与第 6-9 章的校验工作。

朱辰: 负责双数章节的一部分翻译, 与第 10-13 章的校验工作。

周圣宇: 负责双数章节的一部分翻译。

王思懿: 负责单数章节的一部分翻译。

董瑜: 负责网站管理与资料上载。

注意:

各种问题或者建议可以提 issue, 建议使用中文。

或联系邮箱: [sjtu\\_medialab@163.com](mailto:sjtu_medialab@163.com)

可用于学习研究目的, 不得用于任何商业行为。谢谢!

# 目 录

第一章 虚拟现实概览 .....	1
1.1 虚拟现实是什么? .....	1
1.2 现代 VR 体验 .....	6
1.3 VR 的历史发展 .....	14
第二章 总体架构 .....	22
2.1 硬件基础 .....	22
2.2 软件 .....	29
2.3 人体生理学与认知 .....	33
第三章 虚拟世界的几何模型 .....	40
3.1 几何模型 .....	40
3.2 改变位置和方向 .....	43
3.3 旋转的轴-角表示 .....	50
3.4 视角变换 .....	53
3.5 转换级联 .....	57
第四章 光与光学 .....	61
4.1 光的基本行为 .....	61
4.2 镜头 .....	66
4.3 光学畸变 .....	70
4.4 人眼 .....	74
第五章 人类视觉生理学 .....	79
5.1 从角膜到感光体 .....	79
5.2 从感光体到视觉皮层 .....	82
5.3 眼部运动 .....	86
5.4 对 VR 的影响 .....	90
第六章 视觉感知 .....	97
6.1 感知深度 .....	97
6.2 运动感知 .....	105
6.3 颜色感知 .....	110
6.4 结合信息来源 .....	114

第七章 视觉渲染.....	119
7.1 光线跟踪和着色模型 .....	119
7.2 光栅化.....	123
7.3 光学畸变的修正 .....	130
7.4 减小时延&提升帧率 .....	131
7.5 沉浸式照片和视频.....	138
第八章 真实世界和虚拟世界中的运动 .....	142
8.1 速度和加速度 .....	142
8.2 前庭系统 .....	145
8.3 虚拟世界中的物理.....	149
8.4 不匹配的运动和观察 .....	156
第九章 追踪 .....	162
9.1 2D 方向追踪 .....	162
9.2 3D 方向追踪 .....	166
9.3 追踪位置和方向 .....	172
9.4 追踪附属体.....	180
9.5 环境三维扫描 .....	185
第十章 交互 .....	189
10.1 运动程式和重映射 .....	189
10.2 运动 .....	192
10.3 操作 .....	198
10.4 社交互动 .....	200
10.5 其他交互机制 .....	204
第十一章 音频 .....	207
11.1 声音的物理学 .....	207
11.2 人类听觉的生理学 .....	210
11.3 听觉感知 .....	212
11.4 听觉渲染 .....	216
第十二章 VR 系统和体验评估 .....	222
12.1 感知训练 .....	222
12.2 针对开发人员的建议 .....	226
12.3 舒适度和虚拟现实病 .....	228

12.4 人类受试者的实验 .....	233
第十三章 未来前景 .....	240
13.1 触觉和本体感受 .....	240
13.2 嗅觉和味觉 .....	244
13.3 机器人接口 .....	246
13.4 人脑-机器接口 .....	250
参考文献 .....	254

# 第一章 虚拟现实概览

## 1.1 虚拟现实是什么？

虚拟现实（VR）作为前沿科技，近年来发展十分迅速，这也导致设备更新换代的速率越来越快。如果我们着眼于具体的设备来研究虚拟现实，则可能很快就会被时代淘汰。因此，在这本书中，我们重点关注虚拟现实的基本理论。首先，我们必须要搞明白虚拟现实的概念，对虚拟现实进行定义要抓住它的根本，要保证定义的准确性不会随着技术的发展而轻易的改变。此外，对 VR 做出的定义还应该具有普适性，不仅要适用于如今我们对虚拟现实的理解，也应包含我们对其未来形态的预测。

在开始介绍虚拟现实技术之前，我们先举两个具有代表性的例子：

- 1) 一个人通过挥动“翅膀”体验飞跃旧金山的感觉(图 1.1)
- 2) 老鼠通过在跑轮上跑动来探索投影在四周的虚拟迷宫(图 1.2)

在 1.2 节中我们还会介绍更多的样例，我们希望这些样例都能够包含在我们对虚拟现实的定义中。



图 1.1 苏黎世艺术学院的“鸟类体验”装置，用户穿戴 VR 头盔，挥动“翅膀”体验飞跃旧金山的感觉，平台和风扇用来提供额外的感官刺激。右图展示了每只眼睛看到的画面。

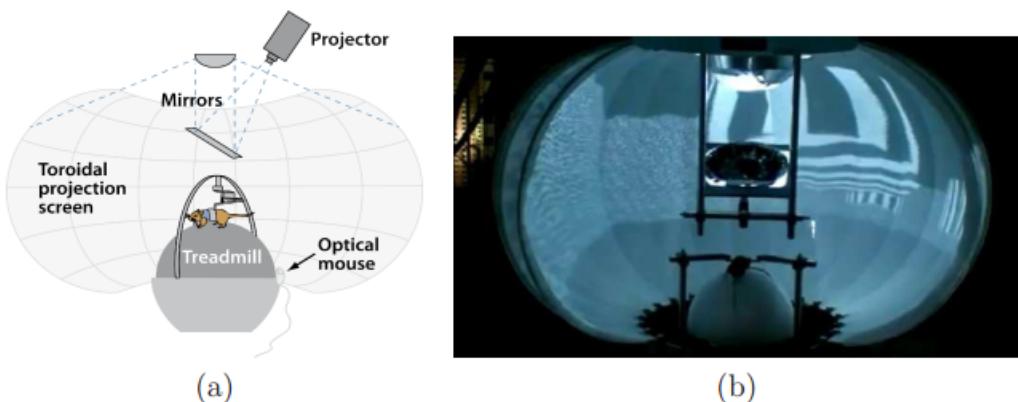


图 1.2 (a) 慕尼黑大学的神经生物学家制作的实验装置，将小鼠放在中心的跑轮上，随着小鼠的运动，四周墙壁上的画面不断变化，为其提供视觉刺激（图源于 Kay Thurley） (b)类似的实验装置图，由普林斯顿大学的研究人员制作

## 虚拟现实的定义

**虚拟现实定义:** 利用人工设计的感官刺激, 引诱生物体产生一种正在进行某种特定行为的感觉, 而整个过程中该生物体几乎无法意识到人工干扰的存在。

定义中出现的四个关键词:

**特定行为:** 生物体正在进行由人工设计的“动作”, 比如说飞行、散步、探索、看电影、社交等等。

**生物体:** 可能是你, 他人, 或是其他物种, 像是果蝇, 蟑螂, 鱼, 老鼠, 猴子等等 (科学家发现虚拟现实技术对上述生物体有影响)

**人工感官刺激:** 通过工程学的技术, 将生物体内的一种或多种感官劫持, 然后使用人工的刺激来替代原有的感官输入

**意识:** 整个过程中, 生物几乎没有感受到人工干扰的存在, 从而成功地被“骗入”虚拟世界。这会使得生物体认为自身正处于全新的世界当中。

## 作用范围

显然, 图 1.1 和图 1.2 的例子, 以及穿戴 VR 头盔体验虚拟场景等等均包含在我们的定义当中。那么, 定义的作用范围到底有多大? 用耳机听音乐算不算? 在剧场看电影算不算? 显然, 电影和音频系统的技术已经可以给予人工的感官刺激。再进一步说, 欣赏画展中的画作算是虚拟现实吗? 它有着包含涂料、画布等技术。甚至看小说是不是也可以被认为是虚拟现实, 它也包含了书写和印刷等技术。这些刺激虽然都是视觉上的, 但是似乎不像电影场景和音频系统那样直接。其实, 我们并不需要给虚拟现实定下一个明确的界限, 上述的例子都值得我们讨论研究。而且, 从历史角度来看的话, 它们也是很好的切入点, 我们会在 1.3 节详细介绍。

## 谁被骗了?

让我们回到虚拟现实的定义, “欺骗”生物体听上去似乎不切实际, 但是, 神经生命学的相关研究告诉我们, 这是可行的。从原理上来看, 当生物来到新的环境时, 首先会构建一系列神经结构, 这些结构由位置细胞组成, 用来记忆周围的空间信息, 如图 1.3(a)。当生物回到之前的环境, 关于这个地方的所有位置细胞会被立即激活。虽然我们还没有完全弄懂原理, 但格点细胞记忆位置信息的方式类似于笛卡尔坐标系, 如图 1.3(b)。研究表明, 进行 VR 体验时, 生物体的神经系统也有可能构建一系列用来记忆环境的神经结构。换句话说, 我们的大脑对虚拟的环境也会做出反应, 生成相应的位置细胞。这表明虚拟现实或多或少地欺骗了我们的大脑。你可能会想, 那么小说中描绘的环境又是否会让我们的大脑做出反应从而生成位置细胞呢。

我们甚至还会想, 我们是否一直在被欺骗, 现实的环境是否只是一场梦境, 是否还存在着更大的现实尚未被我们揭示。古今中外很多哲学家都讨论过这个问题, 最早可以追溯到古希腊时期柏拉图《理想国》中的洞穴寓言。在寓言中, 苏格拉底是这样描述的, 一群整日束缚在洞穴之中的囚徒, 他们面临着一堵空白的墙壁, 他们身后有一堆火在燃烧, 火和囚徒之间有一些人拿着器物走动, 只有当人们经过时才会看到投影到墙壁上的阴影。他解释说, 哲学家就像这些洞穴人, 只有从洞穴中解脱出来才能看到现实的本质, 而不能仅通过预测来认知现实。这个观念一直广为流传, 也与宗教和灵性信仰有着千丝万缕的联系。1641 年, 勒内·笛卡尔提出, 我们的所见所想都是由一个全知全能的恶魔创作的幻象。1973 年, 吉尔伯特介绍了“缸中的大脑”(图 1.4)的思想实验, 来具象化“笛卡尔恶魔”的行为。这听起来很像是《黑客帝国》, 没错, 这部电影正是在这个思想实验的影响下创作出来的。电影中, 一个名为“矩阵”的计算机人工智能系统控制着人们的身体, 并将人们的大脑连接到一个模

拟世界当中。主角 Neo 必须决定是否要回到真正的现实，或是喝下记忆消除药丸幸福地活在梦境之中。

### 有关“现实”的术语

“虚拟现实”的概念最早由德国哲学家伊曼努尔·康德提出，尽管当时还未出现相关的技术。20世纪80年代由Jaron Lanier进行进一步的推广。现如今，虽然虚拟现实已经广泛应用到各行各业，但是一些术语仍然有所争议。比如说虚拟环境，它早于虚拟现实被广泛使用，也更受到研究人员偏爱。它通常被认为与虚拟现实同义，然而，在本书中我们强调，这两个词还是有差异的，在虚拟现实中，我们对环境的感知不都是源于完全合成的世界，也有可能来自镜头捕捉到的“真实”世界，因此，这个环境并不一定是“虚拟”的。增强现实(AR)指的是视觉刺激通过镜头或照相机直接进入眼睛，并且叠加一些额外的内容。混合现实(MR)的概念有时被认定为一种技术组合，包含虚拟现实、增强现实、和现实世界，但近年来VR/AR/MR的含义也可以互相指代。远程呈现指的是让用户感觉自己在现实世界中的其他地方的系统。在此基础上，如果用户可以控制一些设备，比如无人机等，那么用远程操控描述更加合适。按照我们的说法，虚拟环境、增强现实、混合现实、远程呈现、远程操控都可以被视为虚拟现实完美的应用实例。必须要指出，虚拟现实的关键在于通过工程学方法改变用户对现实的感知，而不是用户认为环境是真实的还是虚拟的，因此，我们还可以用感知工程来描述。

从字面上来看，“虚拟”和“现实”似乎是矛盾的，这看起来像是一个哲学问题，【3】中也将其修正为虚拟。尽管如此，我们仍要继续使用虚拟现实这个词语，同时，我们要注意区分：真实世界指的是包含用户的物理世界，虚拟世界指的是在虚拟现实体验中所感知到的世界。

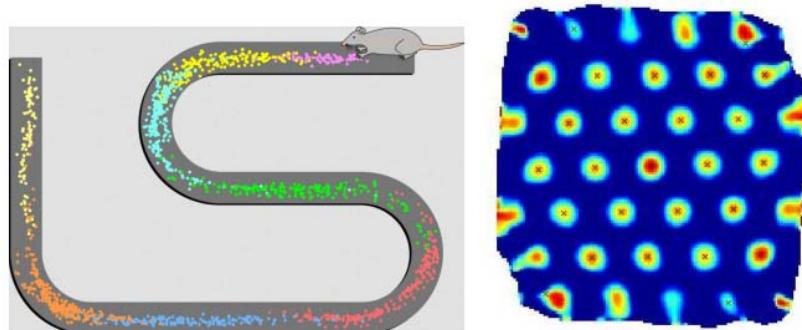


图 1.3 (a) 动物体将神经元分化为位置细胞，当其回到有记忆的地方时会被激活。图中描绘了大鼠沿着特定路线进行往返运动时，脑中 8 个位置细胞的空间分布。(图源于 Stuart Layton) .(b) 在空间中均匀分布的格点细胞，用于对位置坐标进行编码记忆 (图源于 Torkel Hafting)。



图 1.4：关于虚拟现实的思想实验：缸中的大脑，Gilbert Harman 于 1973 年提出（图源于 Alexander Wivel）

## 交互性

虚拟现实系统还有另一个特性：交互性。感官刺激是否依赖于机体采取的行动？如果不是，那么 VR 系统是一个开环系统；否则，它是闭环的。在闭环 VR 系统中，机体对刺激有一定程度的反馈，表现在刺激可能会随着身体的运动而不断变化，包括眼睛、头部、手部或腿部的运动；也可能由于语音命令、心率、体温和皮肤电导性（是否出汗？）等原因而有所变化。

## 第一人称 VS 第三人称

如果你正在读这本书，我猜测你很可能想要开发 VR 系统。那么，你一定要注意关于第一人称/第三人称的事情。研究人员设计生物实验时，分工是清晰的：实验体是体验者，是第一人称；研究人员是观察者，是第三人称。他们通过设计 VR 系统，进行实验来验证自己的科学假设。比如说，抑制大鼠脑中的部分神经元，观察它的导航能力是否受到影响？如何被影响？与研究人员不同，当工程师或开发人员设计 VR 系统时，他们通常将自己代入到系统中，以便于对系统进行评估和改进，也就是说，在“实验体”和“研究人员”两种角色间不断进行切换。读完整本书你会发现，这种做法非常的糟糕。这些设计者想要取得效果的同时避免重复的实验，但是，他们本身知道系统会给予人一种怎样的体验，会有什么样的效果，这种意识会导致他们的感受与首次接触的人的感受相差很大。更糟糕的是，设计者会逐渐地在生理或心理上适应系统的某些缺陷，导致这些缺陷不能被及时发现并解决。日常生活中也有类似的经历，比如说，我们很难预测其他人会对自己的文章做出怎么的反应；还有，相比他人的文章，我们更难以正确地审视自己的文章。在虚拟现实的研究中，这些影响会变得更加严重，你必须时刻注意这一点。当你那些始终信任的感官被劫持时要小心，你将会进入到一个完全未知的领域。

## 比现实更“真实”？

VR 的体验到底应该有多么真实？毫无疑问，让它无限接近我们的现实世界是很诱人的想法。在 10.1 节中，我们会详细介绍通用模拟原理，是指现实世界中的任何交互机制都可以在虚拟现实中进行模拟。我们的大脑对这些机制最为熟悉，因此很容易进行模拟。这种理念在游戏行业中非常火热，例如，在第一人称射击游戏（FPS）的开发中，随着显卡性能的不断提升，游戏画面也被渲染的非常漂亮。但这并不意味着卡通式的环境一无是处，有的时候，简单的环境也许是更可取的选择。纵观历史，有很多这样的例子，我们会在 1.3 节中详

详细介绍。

作为设计整个 VR 系统的人，你需要考虑你想要用户获取什么样的信息，会做出怎样的反应，以及期望的效果是怎样的。经过精心的设计，你甚至可以使得整个体验比现实世界还好。用户可以体验什么？上数学课？观看文艺表演？开发软件？设计房屋？异地情侣之间的交流？玩游戏？冥想放松？到地球甚至宇宙中去旅行？对于每一种情况，不同的需求也会有产生不同的体验，要结合实际需求考虑。例如，在 VR 环境中开发软件。我们现如今开发软件的方法是在电脑里面敲代码，但要知道，在 20 世纪 50 年代这是难以想象的。通过 VR，我们可以完全模拟如今的开发环境，让程序员认为自己坐在办公桌前，面对电脑屏幕敲代码。然而，这有些“大材小用”，、我们可以在 VR 中创建几乎任何东西，无需拘泥于现有的思维定式。我们可以设计一个全新的环境，不是在办公桌前，而是让屏幕漂浮在海滩上、倒映在森林中。此外，我们还可以设想，是否可以让调试器把整个程序的运行轨迹显示出来？怎样显示效果会更好？

## 合成 VS 捕捉

设计 VR 系统，我们首先要构建一个虚拟世界。构建的方法主要分为两种，第一种，我们可以构建一个合成世界，完全由几何原型和物理仿真制作，这在视频游戏中很常见，我们可以认为，这是几十年前 VR 体验的主要方式；另一种，我们可以用镜头捕捉真实场景，再用成像技术将其还原出来。在一个多世纪前，人们已经做到了将摄像机拍摄到的画面投射到屏幕上观看。如今，我们捕捉全景图像和视频，通过 VR 系统可以从任何角度观看，这也是摄影技术的延伸。然而很多情况下，将真实场景投影到相机传感器上会丢失很多图像信息，当用户的头部位置和观察区域发生改变时场景会怎样变化？我们应该捕捉更多的信息来尽可能的还原真实场景。通过深度传感器和 SLAM（即时定位与地图构建）技术，我们可以获取到空间定位的信息，完成地图的构建或场景的生成。然而，想要还原出完全精确的真实场景非常困难，除非整个空间是专门为此设计的。（例如，动作捕捉工作室）

与用户进行互动的时候，跟踪他们的动作并及时作出反应是 VR 系统的关键，这也是我们需要捕捉的信息。穿戴 VR 头盔时候他们的面部表情是怎样的？我们是否应该弄懂他们的手势？他们的情绪是怎样的？他们是否将注意力集中系统中？必要的话，我们可以在虚拟世界中构造一个化身，以合成的方式模拟注意力或情绪状态，此外通过运动跟踪和其他属性来增强这个化身，使其更接近用户，以此来进行人机交互并且在一定程度上保护了用户的隐私。然而，这由引出了另一个著名的哲学问题——恐怖谷理论，指的是当机器人与人类相像程度超过一定程度，由于机器人与人类在外表、动作上都相当相似，所以人类亦会对机器人产生正面的情感。直至到了一个特定程度，人类的反应便会突然变得极之反感。哪怕机器人与人类有一点点的差别，都会显得非常显眼刺目，让整个机器人显得非常僵硬恐怖。目前，用户还没办法很好地把自己在 VR 系统中的所见所感分享给其他人，当然，他们可能也不想这么做。

## 健康和安全

随着目标的不同，我们对 VR 系统可能会有各种各样的要求，但唯一不变的就是用户的健康和安全。与广播或电视等传统媒体不同，VR 有可能会影响到感官和大脑，导致疾病疲劳。几十年前，已经有学者在研究模拟晕动症，本书中，我们将 VR 应用中产生的不良症状统称为 VR 疾病。尽管有时候 VR 疾病是由于低端的软硬件导致的，但大多数情况是由于开发人员没有重视到 VR 的副作用，这也正是为什么这本书的大部分内容都是有关人类生理学和感知心理学的。为了开发出令人舒适的 VR 体验，这些都是必须要了解的。从原理上讲，疲劳是由于给予感官的异常刺激难以被大脑整合理解导致的。有时候，接受到与预期不一致

的刺激，或是其他感官给予的刺激，甚至会导致头晕恶心。

导致疲劳的另一个因素是 VR 互动需要大量的肌肉运动，比方说，你想要移动沙盒游戏中的物体，那么你必须不停地摆动你的胳膊，这会导致疲劳和大猩猩臂效应，即人很难在手臂往外展开的姿势下长时间操作。鼠标的发明正是源于此，通过小而轻微的操作来控制屏幕上光标的移动。经过一段时间的适应期，我们的大脑就会将这两种动作结合起来，使整个过程变得自然，并且大大的减少了疲劳。

## 1.2 现代 VR 体验

当下的 VR 时代是由智能手机行业的显示，传感和计算技术进步而带来的。从 Palmer Luckey 2012 年设计的 Oculus Rift 到简单地在智能手机上构建此功能[10,16,19]，随着 VR 头戴设备的大量生产并发布世界各地，这个世界已经迅速发生了变化。这种趋势在许多方面与家用电脑和网络浏览器革命相似；随着越来越多的人接触到这项技术，他们也正利用 VR 来创造更多可能。

这节内容能让你快速了解当今人们正利用 VR 做哪些事，并且提供一个找寻在互联网方面相似的切入点。在此，我们只能用文字和图片来描述体验，它还是会与你自己亲身体验所带来的惊喜相差甚远。书类的纸质媒体对于充分传达虚拟现实技术来说还是极其不足的。也许这就像 1890 年代在报纸上描述一个电影院是什么样的！如果可能的话，我们强烈建议你亲身去尝试体验一些 VR 技术，从而形成一种对其直观的印象并且激发你的想象力去实现更好的功能。

### 视频游戏

人们几十年来一直梦想着进入视频游戏的世界。在 1982 年以前，这种观念已经因为被迪士尼电影 Tron 而流行起来。图 1.5 展示了一些 VR 游戏的体验。大多数玩家现在都追求通过一个虚拟实景探索更大，更现实的世界。图 1.5(a)为 Valve's Portal 2，它是在 HTC Vive 头戴式设备上开发的一个冒险类的解谜游戏。图 1.5(b)为一个全方位的跑步机外设装置，当人们的脚在地面上滑动时，它可以提供给使用者一种走路的感觉。这两个例子都提供了用户一种当事人的视角。相反，图 1.5(c)展示了 Lucky's Tale，它相反地提供了一个舒适第三方视角当用户看起来像是要漂浮在她控制的角色之上。图 1.5(d)与其他图片都不同，它被明确地设计用来利用 VR 的力量。

### 沉浸式影院

好莱坞电影持续提供升级版的现实体验。为什么不让观看者感觉到他们就是场景中的一部分呢？图 1.6 展示了一个沉浸式的小故事。电影导演已经进入到一个关于电影的迷人的新时代。20 世纪学到的技巧需要被重新研究，因为它们建立在电影摄影师控制摄影机视角的假设上。在 VR 中，观看者们可以选择任何的视角，甚至可以在场景中走动。为什么他们能够这样做呢？你怎么确保他们不会错过故事中的部分场景？这个故事应该是固定的吗，还是应该对观看者的行为有所调整适应？观看者在影片中应该是当事人吗，还是一个对别的人物不可见的旁观者？怎么能让一群好友同时体验 VR 影片？什么时候模拟相比于真实场景的捕捉会更真实合适？

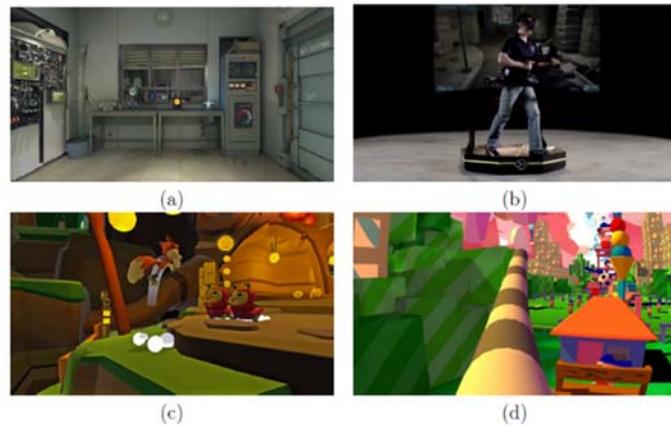


图 1.5: (a) Valve's Portal 2 为 HTC Vive 头戴式设备设计的演示是一款在虚拟世界中解谜的体验游戏。(b) 支持第一人称穿越射击游戏的 Virtuix Omni 跑步机。(c) 当玩家在角色上方浮动时, Oculus Rift 上支持的的 Lucky's Tale 保持第三人称视角。(d) 在 DePaul 大学开发的 Dumpy 游戏中, 玩家似乎有一个大象鼻。游戏的目的是通过摆动的象鼻将东西击倒, 从而享受这种不寻常的体验。

解决这些现存的以及之后会出现的无数问题需要花费很多年的时间。在这期间, VR 也可以用来作为一种现存电影的“包装”。图 1.7 展示了 VR 影院的应用, 它允许用户在虚拟的影院中任意选择座位。无论在用户硬盘上的是何种标准的电影或是视频文件, 都可以被传输到影院的屏幕。他们可以是 2D 也可以是 3D 的。背后的投影仪发出闪烁的灯光并且音频会被调整以模仿真实剧场的音效。它提供了一个快速的途径来转化所有原本创作用来在屏幕上观看的内容, 并且将它们变为 VR。许多简单的扩展可以在不修改影片内容的前提下实现。举个例子, 在一个关于僵尸的电影中, 一些虚拟的僵尸能进入到剧场并开始追逐你。在一个关于龙卷风的电影中, 也许剧场会被撕裂。你还可以体验社交。想象跟来自世界各地的朋友在虚拟的剧场中坐在一起, 共同度过一个“电影之夜”。你们甚至可以在剧院里体会到捣乱的快感, 而不会被赶出去。

### 远程呈现

营造一种我们仿佛在其他地方的第一步是捕捉远程环境的全景(图 1.8)。谷歌的街道场景和地球应用就是依赖于从全世界各地几百万个地方捕捉的的全景图像。查询街景服务器的简单 VR 应用程序会使用户直接感觉他仿佛站在这些位置, 同时能轻松地在附近的位置之间转换(图 1.9)。全景视频的拍摄更加引人注目。图 1.10 展示了一个沉浸式摇滚音乐会的框架。更好的方式是提供现场全景视频界面, 人们可以通过该界面参加体育赛事和音乐会。实时界面使交互变为可能。人们可以把视频会议发展至一个让遥远的参会者仿佛在场出席会议的新阶段。通过在机器人上装备全景相机, 用户甚至可以在遥远的地方自由移动。现在的 VR 技术让我们可以虚拟地去到许多遥远的地方并且能进行大多数以往只能通过亲身到场才能进行的交互活动。这会提供给大家远程办公的机会。这最终可能有助于扭转 19 世纪工业革命引发的城市化趋势, 当人们更均匀地分布在世界各地市实现去城市化。



图 1.6: Oculus Story Studio 制作了艾美奖获奖作品 Henry, 这是一个简短的沉浸式故事, 讲述的是一只不受喜爱的刺猬, 希望能结交一位新朋友, 就是屏幕前的观众。



图 1.7: 由 Joo-Hyung Ahn 为 Oculus Rift 制作的 VR 电影。 你可以选择座位并观看喜欢的任何电影。



图 1.8: 实现远程呈现的重要组件是捕获全景图: (a) 顶部有摄像头和深度传感器的汽车, Google 使用它来制作街景视图。(b) Bublcam 是一种便宜便携的方式来捕捉和流式传输全景视频。



图 1.9: 简单的 VR 体验, 通过 VR 头戴式设备呈现 Google 街景图像: (a) 巴黎的熟悉场景。 (b) 在设备内部创建左眼和右眼视图, 同时还考虑到用户的观看方向。



图 1.10: Jaunt 拍摄了 Paul McCartney 音乐现场的全景视频。Let Die 提供了让用户感到他们仿佛与摇滚明星一同在舞台上的 VR 体验。



图 1.11: 机器人头戴设备的例子: (a) 宾夕法尼亚大学的 DORA 机器人模仿用户的头部运动, 允许他在保持立体视图的同时在遥远的世界中环视 (全景是单色的)。(b) Plexidrone, 一款低成本飞行机器人, 专为流式全景视频而设计。

## 虚拟社会

当远程呈现技术让我们感觉到仿佛处于物理世界的另一部分时, 虚拟现实还允许我们构建出整个社会, 提醒我们这个物理世界的存在, 但实际上只是包含与真人相关的虚拟形象的合成世界。图 1.12 显示了人们通过头戴设备在幻想世界中互动的第二人生场景; 此类体验最初设计为在屏幕上观看, 但现在可以通过 VR 体验。由于各种原因, 一群人可以在这些空间中共度时光, 包括共同的特殊爱好, 教育目标或者只是想逃离普通的生活。

## 同理心

VR 提供的第一人称视角是一个让人们对别人的状况有同理心的有力工具。世界持续在为使不同种族, 宗教, 年龄, 性别, 性取向, 社会地位和教育被接受和获得平等而斗争, 而进步的最大障碍是大多数人无法理解拥有不同身份的感觉。图 1.13 展示了一个由联合国赞助的 VR 项目旨在设身处地体会那些陷入 2015 年叙利亚危机的人们的感受。我们之中的一些人可能会对别人的不幸感到同情, 但是真正设身处地体会到他们的挣扎将会是一种更强烈的感受。图 1.14 是一个允许男人和女人互换身体的 VR 系统。通过虚拟的社会, 我们可以探索更多的可能。如果你比其他任何人都矮十公分将会怎样? 如果你以不同的性别教授你的课程将会怎样? 如果你是警察种族歧视的受害者呢? 使用虚拟现实技术, 我们可以想象许

多如果你不在“合适的”群体中，是无法体会到的“生命游戏”。



图 1.12：虚拟社会通过在虚拟世界中相互作用的交互设备发展，这些虚拟世界由一台通用服务器维护。第二世界的快照如图所示。



图 1.13：在西德拉云上，电影制片人克里斯牛奶为叙利亚难民的痛苦提供了第一人称视角。

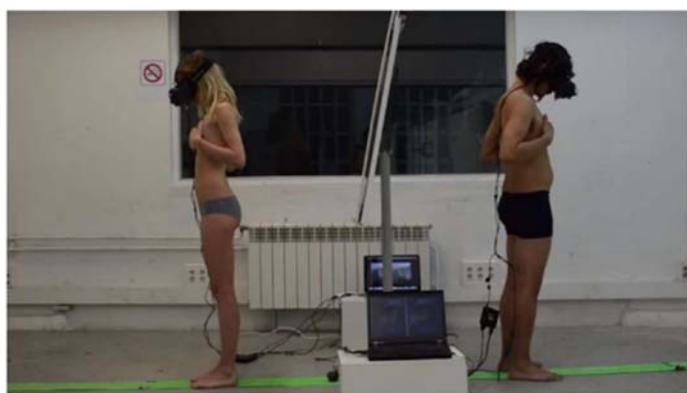


图 1.14：巴塞罗那的学生提供了一个可以与其他性别交换身体的体验。每个人都戴着一台 VR 头戴式设备，前端安装有摄像头。因此，每个人都从另一个人的近似观点看世界。他们被要求以协调一致的动作移动双手，以便他们看到他们新身体的适当移动。

## 教育

除了教会同理心，第一视角也可以对教育的许多方面进行革新。在工程学、数学、和科学方面，虚拟现实提供了把难以解释的概念或数据中几何关系可视化的机会。此外，虚拟现实技术自然适合于实践培训，因为在现实的虚拟环境中开发的技能可能自然地转移到真实的环境中。由于有时真实环境的成本高昂或会产生健康风险，所以开发此技术的动机尤其强烈。VR中最早也是最常见的训练例子之一是飞行模拟（图 1.15）。其他例子包括消防，核电站安全，搜救，军事行动和医疗程序。



图 1.15：美国空军使用的飞行模拟器（Javier Garcia 拍摄）。用户坐在物理驾驶舱中，同时被显示环境的显示器包围。



图 1.16：参观亚瑟国王 Ashurnasirpal II 的 Nimrud 宫殿，这是一个由 Learning Sites Inc. 和伊利诺斯大学开发的 VR 体验。

除了这些 VR 的常见用途之外，也许对于 VR 教育来说最大的机会在于人文学科，包括历史学、人类学和外语的学习。考虑一下读一本英国维多利亚时代的书籍和有机会漫游于由历史学家精心拟建的 19 世纪伦敦街头之间的区别。我们甚至可以去到一个由废墟重建出的古老城市（图 1.16）。迷人的可能性体现在我们既可以通过虚拟现实界面观看实体博物馆，也可以直接在虚拟博物馆中扫描和展示文物。这些例子都属于数字遗产。

## 虚拟原型

在现实世界，我们建立原型来理解一个提出的设计模型具体是怎样工作的。由于 3D 打印跟相关技术的出现，这项工作比之前容易了许多。同时，虚拟原型设计使设计师能够在包含原型的虚拟世界中居住（图 1.17）。他们可以快速地进行交互并修改。它们也能够带客户进入他们构造的虚拟世界以便交流想法。想象你想要重新改造你的厨房。你可以在 VR 中构

建一个模型并向承包人解释它看起来应该是怎么样的。VR 的虚拟原型设计在商业中也有重要的用途，包括房地产、建筑业、以及飞机，航天器，汽车，家具，服装和医疗器械的设计。

## 健康医疗

虽然健康和安全是颇具挑战性的 VR 话题，这些技术还是能够帮助我们改善健康状况。分布式医疗的发展趋势越来越明显，医生们培训人们在遥远的世界各地执行常规医疗程序。医生们能通过远程呈现提供指导建议，并且运用 VR 技术来培训。在 VR 的另一用途中，医生们可以让自己沉浸在利用医疗扫描数据生成的 3D 器官模型中（图 1.18）。通过在手术前简短地了解病人的身体状况，能让他们更好地规划和准备医疗步骤。这也能更好地向病人或其家人解释医疗方面的一些选择以便他们做出更加知情可靠的决定。当下一个另外的应用是 VR 可以直接提供理疗以帮助病人，包括通过重复地暴露克服恐惧症和压力失调，改善或维持认知技能以对抗老化，并且改善运动技能以克服平衡，肌肉或者神经系统紊乱失调等问题。虚拟现实系统有一天还可以通过使老年人虚拟旅行，进行有趣的物理治疗，并通过与亲朋好友通过界面进行互动，使他们在远程活动中感受到自己的存在和参与，来克服孤独感从而延长寿命。

## 增强和混合现实

在一些应用中，用户可以通过叠加一些附加图形来观看真实世界，以增强其外观；如图 1.19。这被称为增强现实或者混合现实（我们在此书中都将它们归为 VR 的一部分）。通过在现实世界中放置文本、图标以及其他图形，用户可以利用互联网的力量来辅助许多操作，例如导航社交互动和机械维护。迄今为止的许多应用都旨在帮助企业更有效地开展业务。想象一下，在一个工厂的环境，工人可以在需要组装的零件上看到标识，或者他们可以直接看到机器的内部来确定潜在的需要替换的零件。

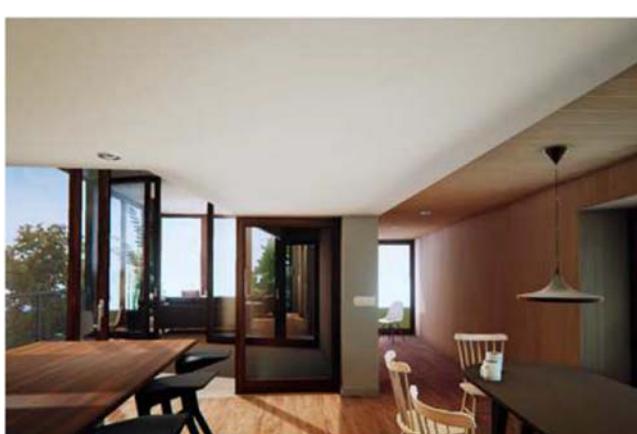


图 1.17：建筑是虚拟原型具有无价价值的最佳例子。这个名为 Ty Hedfan 的演示由设计师 Olivier Demangel 创建。真正的厨房在上面，虚拟厨房在下面。



图 1.18：基于真人心脏图像的心脏可视化系统。这是由 Jump Trading 模拟和教育中心以及伊利诺伊大学开发的。

这些技术都十分依赖于先进的计算机视觉技术。在确定如何绘制仿佛像自然嵌入的虚拟对象之前，必须先识别对象，重构形状，并识别真实世界中的光源。由于视觉算法在不可预见的环境中经常出现错误，所以实现高度的可靠性就成为了一项挑战。必须先估计真实世界的光照条件来确定如何绘制虚拟物体，以及它们可能投射到环境的实际部分以及其他虚拟物体上的任何阴影。此外，在某些应用实例中，真实和虚拟的物体都需要被完美地进行排列，这给追踪和计算机视觉系统都会带来很大的负担。

在虚拟现实中还存在这几种可能。一个固定屏幕可以显示通过 3D 眼镜增强的图像。数字投影机可以通过将光照射到物体上来增强环境，为它们提供新的颜色和纹理，或者在真实世界中放置文本。作为智能手机或平板电脑一部分的手持式屏幕可以用作增强或混合世界的窗口。这是流行的任天堂 Pokemon Go 游戏的基础;图 1.20。与本书更相关的案例涉及将显示器安装在头上。在这种情况下，存在两种主要的方法。在透视显示器中，用户只需要通过透明材料就能看到大部分真实世界，而虚拟对象出现在显示器上以更改部分视图。近期采用先进透视显示技术的头戴式设备包括 Google Glass，Microsoft Hololens 和 Magic Leap。实现高分辨率，宽视野和阻挡入射光的能力仍然是经济实惠的消费级设备面临的重大挑战;然而，它可能会在几年内得到很好的解决。另一种方法是直通显示器，它将图像从外置摄像头发送到头戴式设备内部的标准屏幕。直通显示器克服了当前的透视显示问题，但是却存在延迟，光学失真，颜色失真和有限的动态范围等问题。



图 1.19：微软 Hololens 使用先进的透视显示技术将图形图像叠加到普通的物理世界中，如透过眼镜看到的那样。



图 1.20：任天堂的 Pokéモン Go 是一款网络游戏，它允许用户想象一个叠加在现实世界中的虚拟世界。他们只能通过他们的智能手机屏幕才能看到口袋妖怪角色。



图 1.21：(a) Epic Games 通过虚拟客厅创造了一个狂野的过山车。(b) 断头台模拟器由 Andre Berlemont, Morten Brunbjerg 和 Erkki Trummal 制作。参与者在刀片掉落时被朋友击中了脖子，他们可以在头部滚动时获得适当的视角。

### 新的人类体验

最后，VR 的重点可能只是提供一种新的人类体验。通过远程呈现，人们可以通过机器人或其他人的眼睛体验不一样的世界。然而，我们还可以进一步给予人们在现实世界中不可能（或者致命）的体验。大多数情况下，艺术家们是作出这一努力的人。如同鸟一般的人类飞行体验（图 1.1）就是一个很好的例子。图 1.21 显示了另外两个。如果我们改变规模会怎么样？想象一下，身高只有 2 毫米，看着正对着脸上的蚂蚁。相比身高 50 米，高耸于城市而人们尖叫着跑过你。如果我们模拟系统中药物的作用会怎样？如果你能成为你最喜欢的动物呢？如果你成了一块食物呢？艺术家的创作可能性似乎是无止境的。我们只是受限于身体机能，令人兴奋的冒险其实就在眼前！

## 1.3 VR 的历史发展

### VR 究竟从何而来？

VR 的历史可以从一个很早的时间开始说起，这个时间虽然早于普遍意义上理解的 VR 开端的时间，但其中包含了许多对 VR 来说至关重要的方面。很久以前，我们的祖先看着墙

壁的画，想象着一个三维的故事世界，图 1.22 显示了一些这样的例子。如图 1.22 (a) 所示为距离 3 万年前洞穴绘画，图 1.22 (b) 展示了欧洲中世纪的一幅画，与洞穴绘画类似，但它之中有更多的绘画场景细节，因此留给观赏者的想象力较少。图 1.22 (c) 显示了一幅后期绘画，其中透视图被画家精心诠释，塑造出一座美丽的宫殿景观，它甚至不需要人们的想象就可以将其视为 3D 的场景。到了 19 世纪，许多艺术家已经厌倦了这种现实主义，并开始了有争议的印象派运动，图 1.22 (d) 展示了其中的一个例子，这些绘画留给观看者的想象更多，就像早期的洞穴绘画一样。

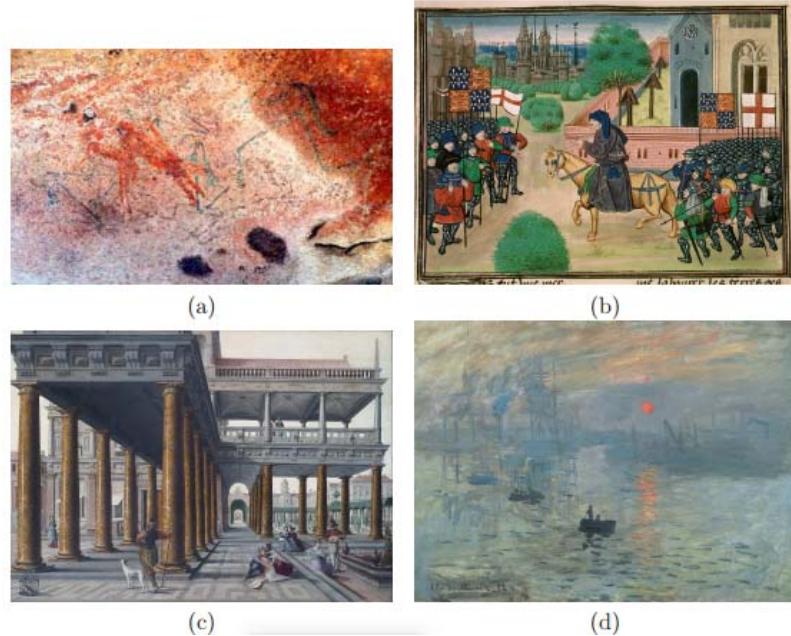


图 1.22：(a) 一幅来自印度 Bhimbetka 岩石庇护所的 30,000 年的绘画作品（印度考古调查照片）。(b) 约 1470 年左右的一幅英国画，描绘了约翰·鲍尔鼓励泰勒泰勒反叛者（未知艺术家）。(c) 1596 年 Hans Vredeman de Vries 的一幅画。(d) 1874 年 Claude Monet 的印象派绘画作品。

### 运动的画面

频闪观运动现象是我们今天称之为电影或动态图像的基础。通过一系列图片快速翻转会产生运动错觉，即使速度低至每秒两张。若每秒超过 10 张照片，动作甚至看起来是连续的，而不是被视为单独的照片。最早的例子之一就是 1878 年爱德华·慕布里奇（Eadward Muybridge）创作的赛马电影，见图 1.23。

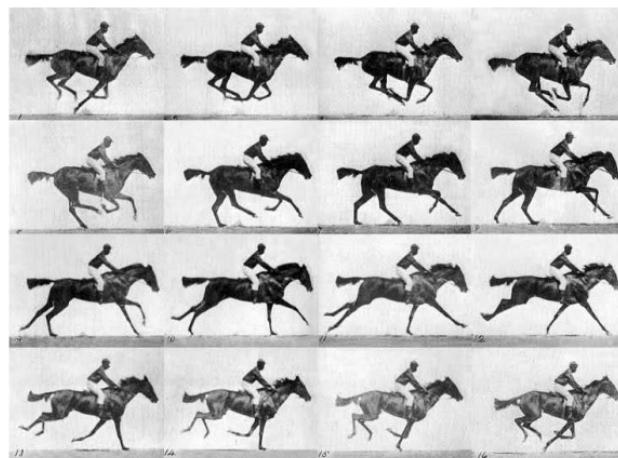


图 1.23: Eadward Muybridge 拍摄的这部 1878 年的动画电影是通过沿轨道均匀放置 24 个摄像机并在马匹经过时通过绊索触发它们而创建的。这部动画是在电影放映机的前身 zoopraxiscope 上播放的，但机械上类似于电唱机。

电影技术迅速得到了改善。到 1896 年，电影院里已经有一些 3D 电影的雏形，比如在一部影片中，一列火车进入火车站的场景会让观众们相信火车即将撞上它们（图 1.24(a)），但是那时候还没有声音。随着观众期望的增加，特殊效果所产生的现实主义程度也随之增加。1902 年，观众可以在电影上观看到“月球之旅”，虽然画质较为拙劣（如图 1.24(b) 所示），但到了 2013 年，要保持观众相信的必要条件是高度的现实还原主义（如图 1.24(c) 和 1.24(d) 所示）。

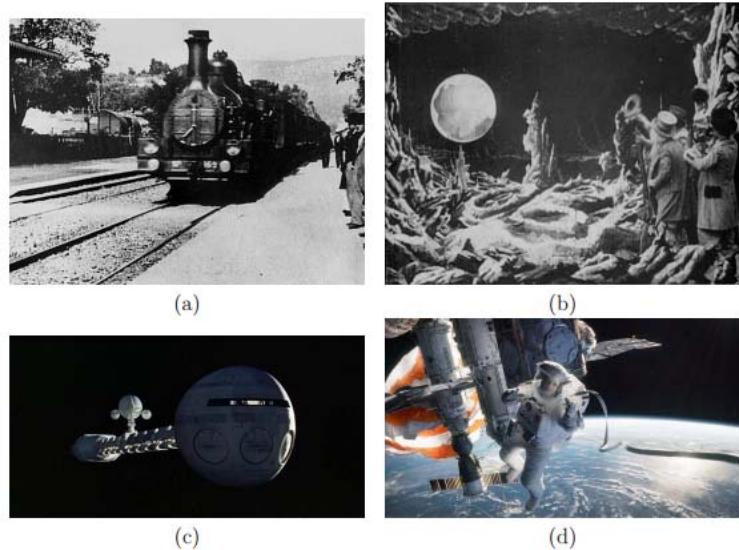


图 1.24: 一系列特殊效果: (a) 1896 年在拉西奥塔火车站列车。(b) 1902 年的月球之旅。(c) 1968 年的 movie2001。(d) 重力, 2013 年

与此同时，电影观众也愿意接受现实主义程度较低的作品。就如同绘画一样，它们之间相同的原理就是留给观众的想象力更多。动画（也称为动漫或漫画）的流行就印证了这个道理（图 1.25）。从 1908 年 Fantasmagorie 的简单线条图起，动画开始萌芽（图 1.25(a)）；随着 1928 年米老鼠的推出，动画出现更多细节部分的元素（图 1.25(b)）；到 2003 年，动画电影取得了更高程度的现实主义（图 1.25(c)）。不过，简单的动画也受到广泛的欢迎（如图 1.25(d)）。

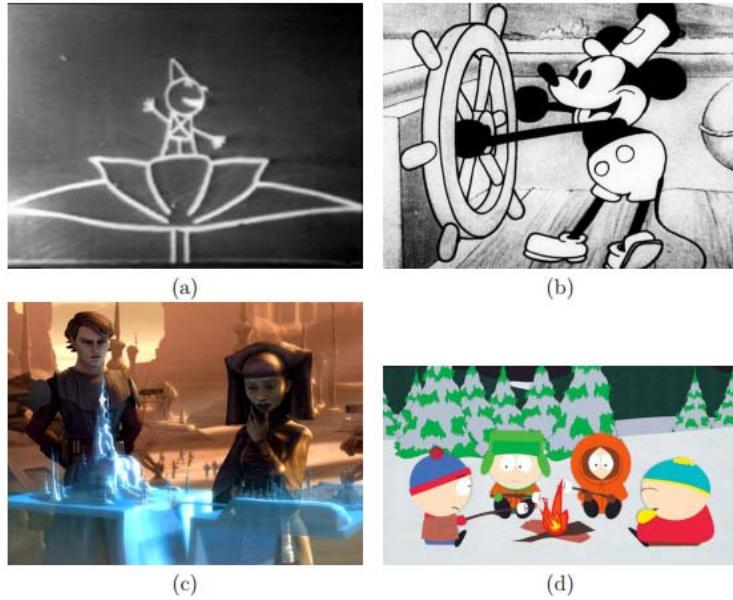


图 1.25：漫画的进展：(a) Emile Cohl, Fantasmagorie, 1908 年。(b) 1928 年 Steamboat Willie 的米老鼠。  
(c) 克隆战争系列, 2003 年。(d) 南方公园, 1997 年

### 成本与便携性

成本和便携性是让人进一步接受较低程度的现实主义的主要原因。如图 1.26 所示，家人愿意聚集在电视机前观看家中的免费广播节目，尽管他们知道可以在剧院观看高分辨率以及全景的 3D 电影。但对于目前的预期，这种显示在小小电视机屏幕上的画面依旧是观看体验上的缺陷。因此迫切亟待的要求是：1) 必须去某个地方观看它，2) 能够在家中观看它，3) 能够随身携带它。无论是图片，电影，电话，电脑还是视频游戏，以上提到的三种需求都是发展的趋势。因此我们期望 VR 系统也是如此的发展趋势。与此同时，我们期望从便携式设备获得的质量也需要达到相同的高质量水平。



图 1.26：尽管有大屏幕的电影院可用，但家庭也很愿意聚集在电视机周围，这些电视机产生的观看质量在目前的标准下是无法忍受的，如 1958 年的这张照片所示。

## 视频游戏

运动图像从某种程度上会产生一种被动的旁观者体验，相比之下，拥有交互性质的视频游戏更接近主观体验。视频游戏是闭环 VR 的重要一步，而动态图像则是开环的。如图 1.27 所示，我们看到了从简单到增进现实，然后回归简单的趋势。最早的视频游戏如 Pong 和 Donkey Kong，给人留下了很大的想象空间。像 Doom 这样的带入性质的射击游戏为玩家提供了主人公视角，并对接下来的十年中的游戏发展造成了深远影响，这种游戏的重点旨在提高图像质量和逼真度。刺客信条展示了一个现代的、现实的视频游戏的典型场景。与此同时，关注简单这个性质，也出现了很多广受欢迎的游戏。愤怒的小鸟看起来像 20 世纪 80 年代的游戏，Minecraft 允许用户创建和居住由方块组成的世界。减少现实主义往往导致更简单的任务要求。在 2015 年，先进的 FPS 游戏可能需要强大的 PC 和图形卡，而简单的游戏则可以在基本的智能手机上运行。因此，游戏做的越逼真不一定越好。

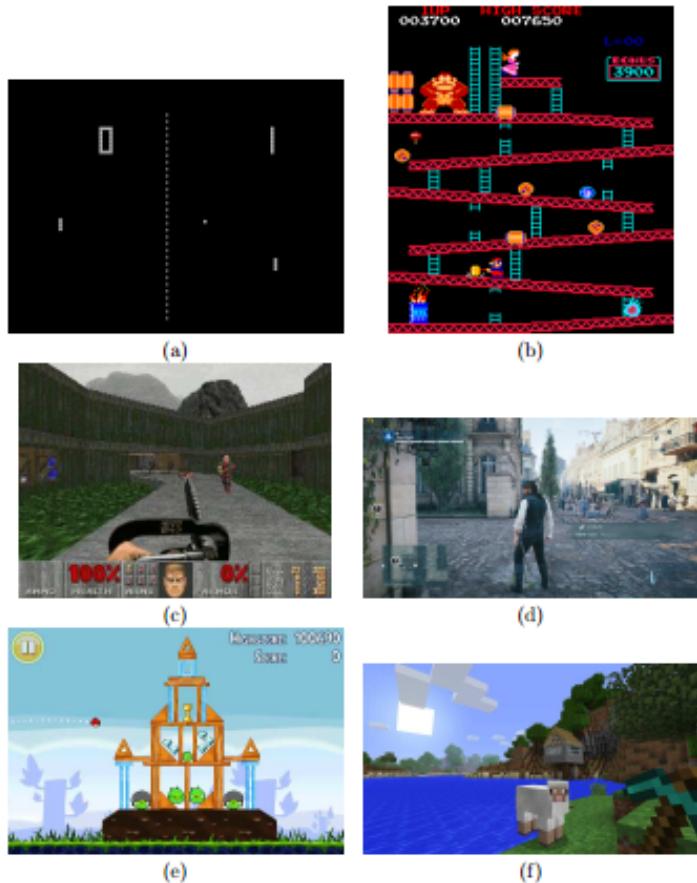


图 1.27：视频游戏的进展：(a) Atari 的 Pong，1972 年。(b) 任天堂的 Donkey Kong，1981 年。(c) id Software 的 Doom，1993 年。(d) Ubisoft 的刺客信条团结，2014 年。(e) Rovio Entertainment 愤怒的小鸟，2009 年。(f) Markus “Notch” Persson’s Minecraft，2011

## 观看视野的突破

到目前为止，我们所提到的观看概念仍然是停留在一个矩形的平面上。若是想要突破这个概念，有两个重要的方面需要考虑：1) 向每只眼睛呈现单独的图片以引起“3D”效应。2) 增加视野，使用户不会被边界影响而分散注意力。我们的大脑从眼睛推断物体的距离的一种方式是立体视觉，通过观察和匹配世界上左右眼可见的特征来获取信息。它们在视网膜上的图像之间的差异产生关于距离的信息。1838 年，Charles Wheatstone 在一个称为立体镜的系统中进行了立体视觉的 3D 效果（图 1.28 (a)）。到 20 世纪 30 年代，便携式立体镜成为当

今被称为 View-Master 的成功商业产品(图 1.28(b))。沿着这个想法进一步产生了 Sensorama，它增加了视觉，声音，振动，甚至是气味的不同感官体验 (图 1.28 (c))。这些设计的一个不幸的局限是要求视点相对于图片是固定的。如果设备太大，那么用户的头部也会变为固定。自 20 世纪 50 年代以来，电影院已经有 3D 电影的趋势。当观看者佩戴特殊眼镜，使用偏振滤光片为每只眼睛选择不同的图像时，立体视觉就会实现。这种流行的 3D 电影，在今天的剧院中都以同样的方式观看。

另一种增加沉浸感和深度感的方式是增加视野范围。20 世纪 50 年代的 Cinerama 系统提供了一个弯曲的宽视场，类似于今天提供的弯曲的大型 LED (发光二极管) 显示器 (图 1.28 (d))。沿着这些路线，我们可以在我周围放置屏幕。沿着这个思路一个重要的虚拟现实系统系列诞生了，称为 CAVE。它于 1992 年在伊利诺伊大学推出[49] (图 1.29 (a))。用户进入一个房间，视频投影到几面墙壁上。CAVE 系统还通过使用偏振光和特殊眼镜向每只眼睛呈现不同图像来提供立体观看，通常还会执行头部追踪以允许视点相关的视频出现在墙上。

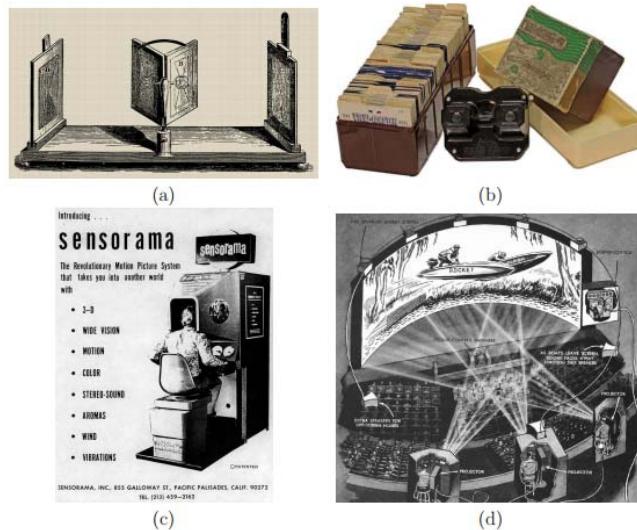


图 1.28: (a) 1838 年由 Charles Wheatstone 开发的第一台立体镜，使用镜子为每只眼睛呈现不同的图像；不久之后镜片被镜片替代。(b) View-Master 是一种大规模生产的立体镜，自 20 世纪 30 年代以来一直可用。(c) 在 1957 年，莫顿海利格的 Sensorama 增加了电影，声音，震动，甚至是气味。(d) 在与立体观看竞争中，全景电影提供了更大的视野。更大的电影屏幕导致 3D 电影的流行在 20 世纪 50 年代衰落

## VR 头戴式设备

VR 头戴式设备再一次显示了便携性的趋势。1968 年，VR 的一个重要里程碑是引入了伊凡萨瑟兰的达摩克利斯之剑，它利用了现代显示器和计算机的原理(图 1.29(b))[315,316]，构建了被广泛认为是第一个 VR 头戴式设备。当用户转动头部时，屏幕上呈现的图像会进行调整以进行补偿，以便虚拟物体看起来像在空间中固定。这让我们首先看到了一个重要概念：平稳感。为了使物体在移动时看起来是静止的，产生刺激的装置必须改变其输出来补偿运动。这就要求传感器和跟踪系统成为 VR 系统的一部分。20 世纪 80 年代，Jaron Lanier 的 VPL 公司开始出现商用 VR 头戴式设备，从而推广了护目镜和手套的交互设备，如图 1.29 (c)。在 20 世纪 90 年代，基于 VR 的视频游戏出现在市场 (图 1.29 (d)) 和家庭 (图 1.29 (e)) 中，但在那时候，使用体验并不令人十分满意，没有吸引大众的兴趣。目前的 VR 技术由于智能手机行业的发展，以及高分辨率屏幕和传感器的广泛应用，已经产生了可以提供轻量级，低成本，高视野的头戴式设备，如 Oculus Rift (图 1.29 (f))。这极大地提高了 VR 体验的质量，同时显着降低了开发者和业余爱好者的入门门槛。这也引起了近来对 VR 技术和应用的

兴趣。

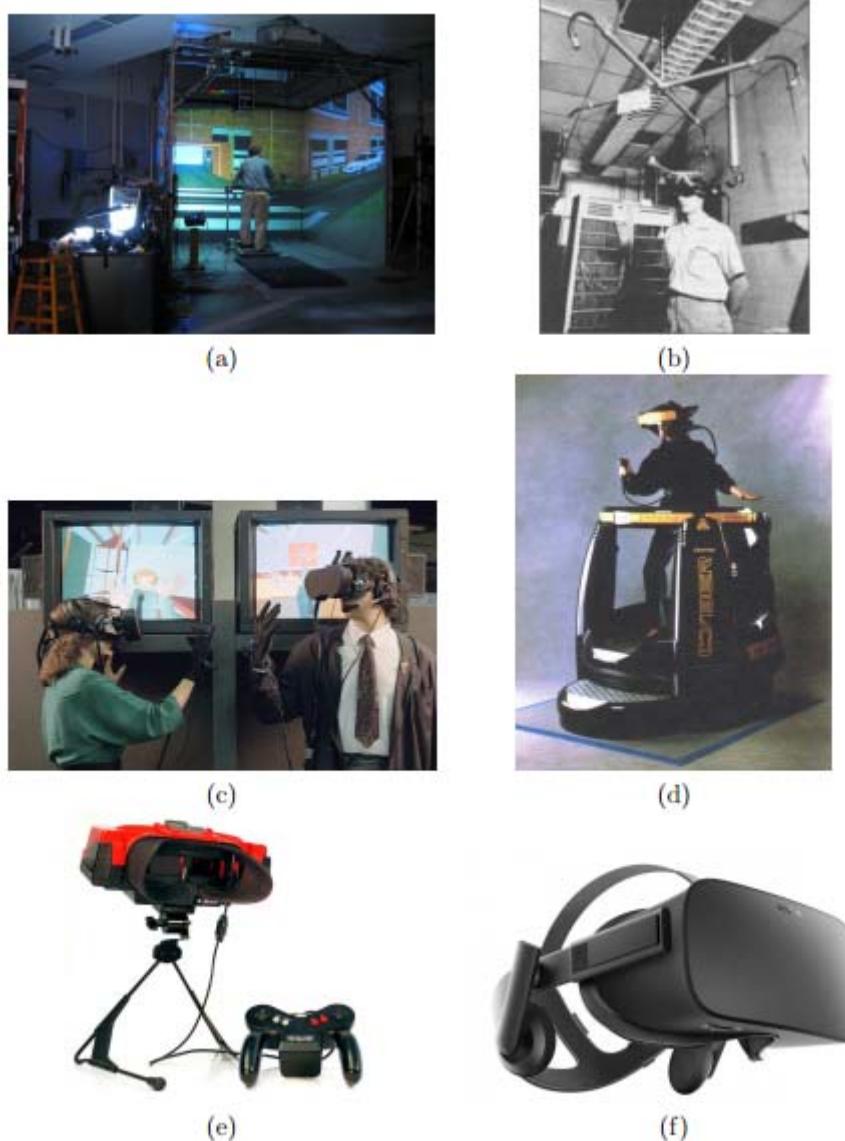


图 1.29: (a) CAVE VR, 1992 年。 (b) 达摩克利斯之剑, 1968 年。 (c) VPL 眼镜, 1980 年代。 (d) 90 年代的虚拟游戏。 (e) 任天堂虚拟男孩, 1995 年。 (f) Oculus Rift, 2016 年。

### 群体聚集体验

群体聚集体验是一个容易忽视的重要方面，即人与人之间或社会之间的相互作用。我们使用诸如现场演出，教室或演讲厅等方式为人们提供交流或娱乐。我们写作和阅读小说，互相讲故事。在写作之前，熟练的讲故事者会将经验传播给其他人。几个世纪以来我们可以通过写信给对方进行沟通。而眼下更新的技术使我们能够立即进行直接交互。音频部分已通过电话传输了一个多世纪，现在视频部分也通过互联网上的视频会议传输。与此同时，简单的短信已成为我们互动的宝贵部分。自 20 世纪 70 年代以来，通过因特网进行交互的在线用户群体一直在增长。在游戏的背景下，早期的多用户地下城（MUD）已经发展成为今天的大规模多人在线游戏（MMORPG）。在教育方面，伊利诺伊大学的 PLATO 系统是第一个计算机辅助教学系统，包括留言板，即时消息，屏幕共享，聊天室和表情符号。这是许多基于社区的在线学习系统的前身，如 Khan 学院和 Coursera。今天最大的在线社交互动是通过 Facebook 进行的，这些应用程序可以通过文字直接沟通，并共享图片，视频和链接。



图 1.30: Second Life 于 2003 年推出, 通过化身进行社交并基本上建立一个虚拟世界的生活方式。作者在 2014 年开放社区大会上发表主题演讲。开发人员构建开源软件工具, 用于构建和托管这些虚拟人物社区, 以及身后的真人。

回到虚拟现实部分, 我们可以在虚拟环境中创建自己的虚拟形象并“共存”, 就像 Second Life 和 Opensimulator (如图 1.30) 一样。一部大众科幻小说描绘了一个令人激动的反乌托邦式的未来世界, 在这个世界里面每个人都喜欢通过虚拟现实进行互动。但未来到底会是怎样的趋势, 这还仍有待观察。

随着目前科学技术的发展, 在制作 VR 体验的时候发现其实不是描绘的越真实就越好, 给用户留足想象力的空间也是很重要的环节。尽管随着技术演变的变化, 人类的设计美感仍然是一样的, 需要了解我们感官、大脑和身体如何工作, 这对于 VR 今后的需求发展至关重要。

### 进一步阅读

本书的每一章都会对前面正文中未提及的其他相关文献进行总结。目前, VR领域已经有具有指导意义的书籍。消费级VR革新之前的几本关键教科书是W. R. Sherman 和 A. B. Craig 编写, 2002 出版的《理解虚拟现实》[288] 和 D. A. Bowman 等人编写, 2005 出版的《3D 用户界面》[31]。基于当前技术的书籍包括[135, 183]。有关增强现实概念的调查, 请参阅[348]。有关增强现实的最新报道超出了本书的范围, 请参阅[280]。

目前也有大量关于VR的研究文献。不幸的是, 目前的消费级VR行业似乎已经忘却了VR研究的长期历史, 导致学术与商业之间存在相当大的认识差距。当今提出的许多问题和工业中引入的方法在几十年前就得到了很好的解决。早期的大部分工作放在今天仍然有用, 因此值得仔细研究。一个很好的参考读物是2015年出版《虚拟环境手册》[108], 其中包含了数十篇最近的调查文章和数千篇相关的参考文献。最近的资料可以在发表VR相关论文的网站、论坛、出版物中找到。阅读这些最新的会议、出版物可能很有用: IEEE 虚拟现实 (IEEE q VR), IEEE 混合和增强现实国际会议 (ISMAR), ACM q SIGGRAPH 会议, ACM 应用感知研讨会, ACM qSIGCHI 会议, IEEEq3D 用户界面研讨会, 《Journal of Vision》, 《Presence: Teleoperators and Virtual Environments》。

## 第二章 总体架构

本章阐述了 VR 系统从硬件（第 2.1 节）到软件（第 2.2 节）再到人类感知（第 2.3 节）的概述。其目的是迅速地提供一个全面的视角，以便在更大范围内理解其余章节中的详细主题。12.2 节为 VR 开发人员提供了一些建议，有助于更深刻的理解 VR。这些章节中的基本概念将主要面向工程和科学背景的读者。此外，本书的读者读完本书后应该能够具备开发新技术并向他人提出他们自己的建议的能力，以便高效学习 VR 系统，并有好的学习体验。

### 2.1 硬件基础

理解 VR 如何工作，第一步是考虑整个 VR 系统的构成。人们很容易认为它只是硬件组件，例如电脑、头戴式设备和控制器，这是不全面的。如图 2.1 所示是 VR 系统的组织构成。硬件会产生覆盖用户感官的刺激。在第 1.3 节的达摩克利斯之剑（图 1.29 (b)）中，回想一下，我们需要用跟踪并根据人体的运动来调整刺激。VR 硬件通过使用自己的传感器来完成这一点，从而跟踪用户的运动。头部跟踪是最重要的，但跟踪也可能包括按钮按压，控制器移动，眼球运动或任何其他身体部位的移动。最后，将周围的物理世界视为 VR 系统的一部分也很重要。尽管 VR 硬件提供了刺激，但用户总是会有其他感官响应来自现实世界的刺激，他还有能力通过身体运动改变他的环境。VR 硬件还可能跟踪用户以外的对象，特别是如果与他们进行交互是 VR 体验的一部分。通过自动化界面，VR 硬件也可能改变现实世界，比如通过 VR 界面遥控机器人。

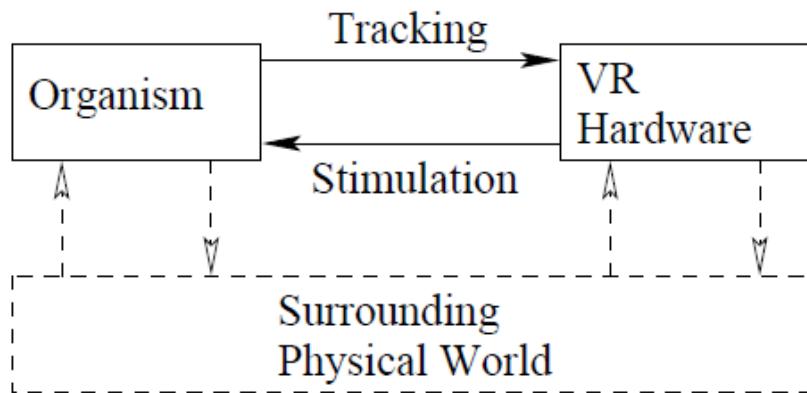


图 2.1 VR 系统的第三人称视角。设计的硬件和软件就是完整的 VR 系统，这种假设是错误的：人及其与硬件的交互同样重要。此外，在 VR 体验中，与周围物理世界的交互仍然存在。

### 传感器和感官

传感器和感觉器官如何从现实世界提取信息？显然，这对于 VR 系统至关重要。在工程中，传感器是指将能量从一种形式转换为另一种形式的装置。传感器是一种特殊的转换器件，可将接收到的能量转换为电信号，这取决于电路类型，可能是模拟或数字信号。传感器通常具有收集用于转换的能量接收器，有机体以相似的方式工作。“传感器”被称为感官，比如眼睛和耳朵。因为我们的“电路”是由相互连接的神经元形成的，所以感官将能量转化为神经冲动。在阅读本书时，请留心传感器与感官之间的相似之处，他们测量的东西相同，有时甚至以类似的方式运作。这不应感到意外，因为我们和我们的工程设备共享相同的物理世界：

物理和化学定律保持不变。

### 感觉器官的空间构造

当用户在物理世界中移动时，他的感觉器官随之移动。此外，一些感觉器官相对于身体骨骼移动，例如我们的眼睛在其眼窝内旋转。每个感官器官都有一个空间构造，它对应于所有可能的方式来进行转换或配置。这个最重要的方面是感官自由度的数量。第3章将全面介绍这一点，但现在要注意的是，在普通空间中移动的刚体有六个自由度。三个自由度与其在空间中的变化位置相对应：1) 侧向运动，2) 垂直运动，以及3) 远近运动。另外三个自由度对应于物体可能的旋转方式；换句话说，需要恰好三个独立的参数来指定对象如何定向。这些被称为偏航，俯仰和滚转，并在3.2节中介绍。

举个例子，考虑你的左耳。当你旋转你的头部或者在空间中移动你的身体时，耳朵的位置以及它的方向都会改变，这产生六个自由度。右眼也是如此，但它也能够独立于头部旋转。请记住，我们的身体有更多的自由度，这会影响我们感官的构造。跟踪系统可能是必要的，以确定接收人工刺激的每个感觉器官的位置和方向，接下来会尽快对其进行解释。

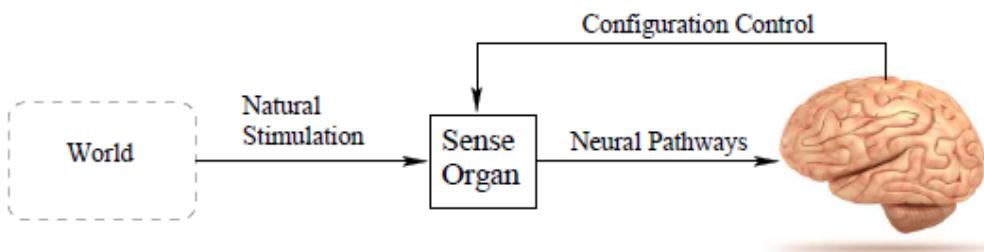


图 2.2：在正常情况下，大脑（和身体部位）在接受来自周围物理世界的自然刺激时控制感觉器官（眼睛，耳朵，指尖）的配置。

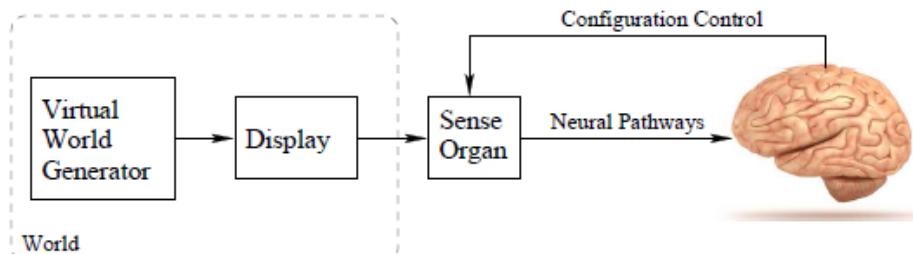


图 2.3：与图 2.2 相比，VR 系统通过用称为显示器的硬件提供的人工刺激代替自然刺激来“劫持”每种感觉。通过使用计算机，虚拟世界生成器维持接通的虚拟世界。这个虚拟世界以适当的“视图”来呈现给显示器。

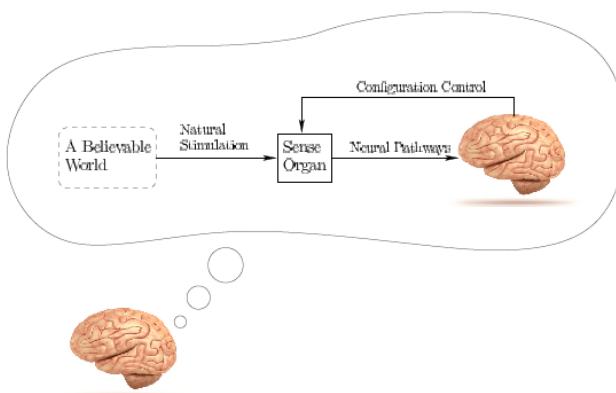


图 2.4：如果做得好，大脑会被“愚弄”并相信虚拟世界实际上是周围的物理世界，自然的刺激源于其中。

## 抽象视图

图 2.2 说明了在没有 VR 硬件的干扰下，我们的感觉器官的正常操作。大脑控制着它的构形，而感觉器官将来自环境的自然刺激转化为发送到大脑的神经冲动。图 2.3 显示了它如何体现在 VR 系统中。VR 硬件包含了几部分，后续我们将对其进行讨论。虚拟世界生成器 (VWG) 运行在计算机上并产生“另一个世界”，这可能有许多可能性，例如合成世界的完全模拟，对真实世界的记录或与另一部分真实世界的实时连接。人类通过每个使用显示器的目标感觉器官来感知虚拟世界，该感官感知到专门设计用于模拟在无 VR 的情况下将出现的刺激类型的能量。将来自 VWG 的信息转换为显示输出的过程称为渲染。对于人眼，显示器可能是智能手机屏幕或视频投影仪的屏幕。对于耳朵，这被称为扬声器。（即使这是日常生活中常见的用法，显示器也不一定是可见的。）如果 VR 系统是有效的，那么大脑有希望被“愚弄”，如图 2.4 所示。用户应该相信感官的刺激是自然的，并且来自合理的世界，与一些过去的经验一致。

## 听觉：世界固定与用户固定

从第 1.3 节中回想一下这样一种趋势：从不得不去某个地方体验一下，到可以在家中体验，再到最终让它变得完全便携。为了理解 VR 系统的这些选择及其对技术的影响，比较一个更简单的情况是有帮助的：音频或听觉系统。

图 2.5 显示了杜比 7.1 环绕声系统的扬声器设置和听众位置，可安装在剧院或家庭娱乐室。分布在房间周围的七个扬声器产生大部分声音，而低音音响（“7.1”的“1”）提供最低频率的部件。因此，听觉显示是全局固定的。将此与佩戴耳机的听众相比较，如图 2.6 所示。在这种情况下，听觉显示是对象固定的。希望你已经多次经历过类似这样的设置。

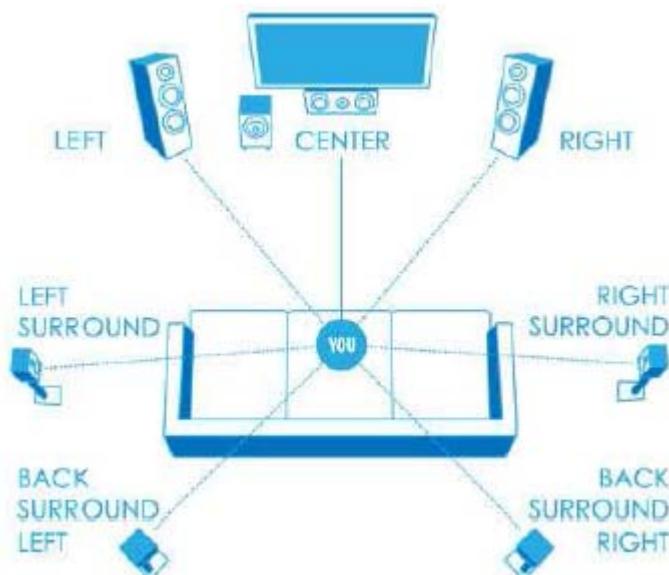


图 2.5：在环绕声系统中，当用户从中心收听时，听觉显示器（扬声器）是全局坐标中固定的。



图 2.6：使用耳机，显示器是用户固定的，与环绕声系统不同。

它们之间关键的区别是什么？除了头戴式耳机明显的便携性之外，下面很快就会想到：在环绕声系统中，生成的声音（或刺激）远离耳朵，但耳机实则非常接近。

距离差异的含义是与远端扬声器相比，若产生相等的感知响度水平，耳机需要的功率要少得多。距离的另一个意义是耳机佩戴者可允许的隐私程度。高音量的环绕声系统可能会引起街坊邻居的愤怒。长时间戴头戴电子设备可能会感到不舒服，从而更偏爱环绕声。可同时几个人在环绕声系统中享受相同的体验（尽管他们不能全部坐在最佳位置）。如果使用耳机，那他们将需要将音频源同时分配到各自的耳机里。根据制造难度和可用的组件技术，它们可能会有不同的成本。目前，耳机比一套环绕声扬声器成本低得多（尽管可以花费大量的金钱）。所有这些差异都会传递到 VR 系统。这应该不会让人感到意外，因为根据第 1.1 节的定义，我们可以轻松地将纯音频体验视为一种特殊的 VR 体验。

在听音乐的同时，闭上你的眼睛，想象你周围环绕着艺术家，你正在现场欣赏演奏。你认为艺术家和他们的乐器在哪里？他们在你周围，还是他们似乎在你的头顶？使用耳机，他们很可能在你的脑袋里。在环绕声系统中，如果正确记录并将其显示，声音应该看起来好像来自它们原来位于头部以外的位置。它们可能感觉起来不太自然，然而却进入你的水平面。

耳机的这个缺点目前尚未得到广泛认可，但对于包含比如视频显示器的 VR 系统而言，这个问题变得尤为严重。如果你想保留你对声音来自哪里的感知，那么耳机就需要考虑你的耳朵在空间中的配置来相应地调整输出。例如，如果您以“不”手势左右摇头，则需要调整呈现给每只耳朵的声音，以便模拟声源以相反的方向旋转。在环绕声系统中，扬声器不会跟随您的头部，因此不需要旋转。如果扬声器随着您的头部旋转，则需要反向旋转以“抵消”你的头部旋转，以使声源位置被认为是静止的。

### 视觉：世界固定与用户固定

现在考虑添加一个可视化显示。在听音乐时，您可能不太担心艺术家和乐器的位置，但您很快就会注意到他们的位置是否显得不正确。我们的视觉感比我们的听觉感更强大和复杂。图 2.7 (a) 显示了一个 CAVE 系统，它在很多方面与环绕声系统相似。用户再次坐在中央，周围的显示器给他的眼睛带来视觉刺激。扬声器被视频屏幕取代。图 2.7 (b) 显示了佩戴与头戴式耳机相似的 VR 头盔的用户。

假设用户眼前的屏幕显示了在耳机中固定的图像，如果用户旋转他的头部，则图像将被视为与头部相连。例如，如果在使用 Viewmaster 时旋转头部（回忆图 1.28 (b)），就会发

生这种情况。如果您想将图像视为周围固定环境的一部分，那么在您旋转头部时，头戴式耳机中的图像必须更改以进行补偿。周围的虚拟世界应该是反向旋转的，其意义将在第 3.4 节中更加精确。



图 2.7: (a) 在英国 Teesside 大学开发的 CAVE VR 系统。 (b) 2013 年, 一位 90 岁女性 (Rachel Mahassel) 戴着 Oculus Rift DK1 耳机。

一旦我们认可这种转变是必要的，那么估计已发生的头部和眼球运动的量并且以及时和准确的方式应用适当的转换便成为了重大的工程挑战。如果处理得不好，那么用户可能会有不好的用户体验。更糟糕的是，他们可能会成为 VR 疾病的牺牲品。这是 VR 头盔在九十年代衰退的主要原因之一，而且组件技术还不够好。幸运的是，目前情况有了很大改善。对于音频而言，似乎很少有人对这种转变感到困扰，但对于视觉对象来说，这绝对是关键。最后一点需要注意的是，如果我们希望屏幕上的图像根据房间内眼睛位置的变化而改变，那么 CAVE 系统中的跟踪和应用转换也是必要的。

现在您已经对通用硬件架构有了高层次的了解，我们将深入研究构建 VR 系统时可用的硬件组件。预计这些变化很快，成本下降，性能提高。我们也预计未来几年将有许多新设备出现在市场中。尽管如此，本书的基本面仍然没有改变。依据当前技术的了解提供具体的例子，使 VR 基本概念更加清晰。

VR 系统的硬件组件可以方便地分类为：

- 显示 (输出):** 刺激感官的装置。
- 传感器 (输入):** 从现实世界提取信息的设备。
- 计算机:** 依次处理输入和输出的设备。

## 显示器

显示器的目的是为目标感觉器官生成刺激。视觉是我们的主要感官，任何为眼睛构造的显示都必须在视网膜上形成所需的图像。由于这个重要性，第 4 章和第 5 章将分别解释光学系统和人类视觉系统。对于 CAVE 系统，则对数字投影仪和镜像进行了组合。由于成本暴跌，一系列大屏幕显示器可能会被采用。头戴设备可以将智能手机显示屏靠近眼睛放置，并为每只眼睛使用一个放大镜将其聚焦。屏幕制造商目前正在通过利用智能手机行业的最新 LED 显示技术制造 VR 头戴式定制显示器。有些厂商则瞄准每个眼睛配置一个显示器，帧速率高于 90Hz，单眼超过两百万像素。第 5 章解释了这个原因。



图 2.8: 触觉反馈设备的两个例子。 (a) Geomagic Punk 允许用户用真实的触针刺入虚拟物体时感受到强大的阻力。机器人手臂提供适当的力量。 (b) 有些游戏控制器偶尔产生振动。

为 VR 定制的新显示技术正在迅速兴起。通过使用微型投影技术（包括 DLP（数字光处理），LCD（液晶显示器）和 LCoS（硅上的液晶））可以直接实现视网膜刺激。使用此技术的产品包括 Google Glass，Microsoft Hololens 和 Avegant Glyph。为了解决舒适性问题，如辐辏调节不匹配（参见第 5.4 节），更多的显示技术已经进行了原型设计，但仍然可以在像素密度，视场，帧率，可制造性，成本等方面与其他显示技术相媲美。两个主要的类群是光场显示器[74,161,198]和多焦平面显示器[4,127,187]。

现在设想展现其他感官。使用经典扬声器技术将声音传送到耳朵，也可以使用骨传导方法，其使头骨振动并将波传播至内耳；这种方法被运用在 Google Glass 上。第 11 章详细介绍了 VR 的听觉部分。对于触觉而言也会有触觉的显示，图 2.8 给出了两个例子，触觉反馈可以以振动，压力或温度的形式给出。有关触觉，味觉和嗅觉的更多细节，请参阅第 13 章。

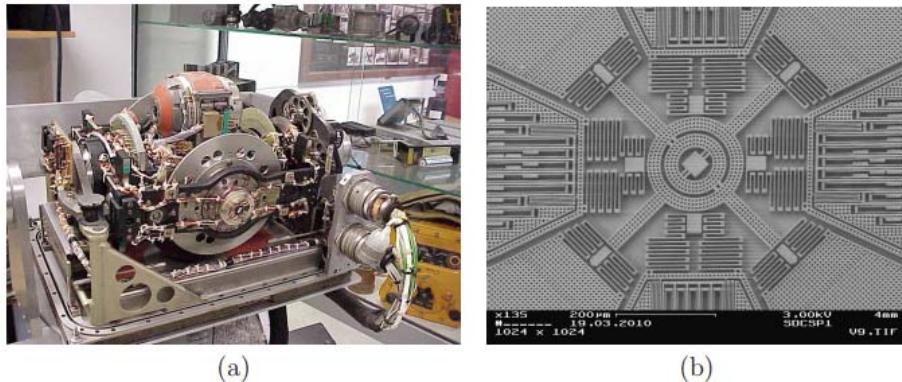


图 2.9 惯性测量单元 (IMU) 已经从大型重型机械系统转向廉价微型 MEMS 电路。(a) Litton Industries 在 20 世纪 60 年代开发的 LN-3 惯性导航系统。(b) MEMS 陀螺仪的内部结构，其总宽度小于 1mm。

## 传感器

让我们考虑一下 VR 硬件的输入端。这里只是给出一个简要概述，我们会在第 9 章详细的介绍传感器和跟踪系统。对于视觉和听觉的显示器，人体感官的位置和方向必须由传感器进行跟踪以适当调整刺激。定向部分通常由惯性测量单元或 IMU 完成。主要部件是一个陀螺仪，它测量自己的旋转速度；该速率被称为角速度并且具有三个分量。来自陀螺仪的测量值随时间增长以获得方位累积变化的估计值。所导致的误差称为漂移误差，如果不使用其他传感器，其值将逐渐增加。为了减少漂移误差，IMU 还包含一个加速计和一个磁力计。多年来，IMU 已经从曾经只存在于飞机和导弹等大型机械系统中到变成了手机中的小型设备，

见图 2.9。由于其体积小，重量轻，成本低，IMU 可以轻松嵌入可穿戴设备中。它们是当前 VR 头盔最重要的支持技术之一，主要用于跟踪用户的头部方向。

数码相机为跟踪系统提供了另一个重要的信息来源。与 IMU 类似，由于智能手机行业的需求，它们已变得越来越便宜和便携，同时也提高了图像质量。相机运用了“视线可见性”的跟踪方法，这个方法可识别图像中用作移动物体或静止背景参考点的特征或标记。这种可见性约束严重限制了可能的对象位置和方向。标准相机通过将光线聚焦于光学系统而形成图像，非常像人眼的成像原理。一旦摄像机校准参数已知，则可以利用观察到的标记去定位空间中的光线。照相机通常用于跟踪眼睛，头部，手部，整个人体以及物理世界中的其他物体。目前面临的主要挑战之一是：当不在用户或场景周围放置特殊标记时，也可获得可靠和准确的性能。

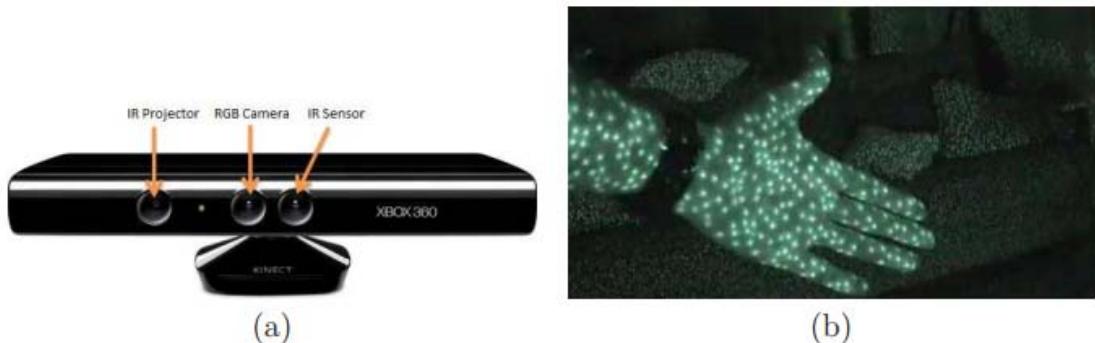


图 2.10：(a) Microsoft Kinect 传感器收集普通的 RGB 图像和深度图（每个像素距离传感器的距离）。(b) 通过观察从 IR 相机获得的图像中投影的 IR 点的位置来确定深度。

与标准相机相反，深度相机通过将光投影到场景中并观察其在影像中的反射来工作。这通常是在红外（IR）光谱中完成的，因此人眼不会注意到，见图 2.10。

除了这些传感器之外，我们还非常依赖古老的机械开关和电位计来创建键盘和游戏控制器，光学鼠标也是常用的。这些熟悉设备的一个优点是用户可以利用他们现有的知识快速输入数据或控制他们的角色。一个缺点是，如果他们的面部被头戴式耳机覆盖，他们可能很难互动。

## 计算机

虚拟世界来自于计算机。而这台电脑应该在哪里？尽管对于“world-fixed”的显示器并不重要，但对于“body-fixed”的显示器而言，该位置至关重要。如果在系统中需要一台独立的 PC，则必须在头戴式设备和 PC 之间提供快速，可靠的通信。这种连接目前通过电线进行，而有线总是带来体验的不适感，但目前的无线速度又是不够的。正如你已经注意到的，大多数传感器都存在于智能手机上，以及功能适中的计算机上。因此，可以将智能手机放入带有镜头的盒子中，以提供几乎不增加成本的 VR 体验（图 2.11）。但是，这样的做法的一个限制是模拟出的虚拟世界必须比在 PC 上运行要更简单，以便它可以在不太强大的计算硬件上运行。在不久的将来，我们希望看到一体式的无线多功能头戴式设备，以提供 VR 体验。在这种一体机中将消除智能手机中不必要的组件，而是专门为 VR 夺身定制光学器件，微芯片和传感器。



图 2.11：通过将智能手机放入盒子创建 VR 体验的两款头戴式设备。（a）Google Cardboard，适用于各种智能手机。（b）三星 Gear VR，针对特定的智能手机进行了优化（本例中为三星 S6）。

在 VR 系统中，除了主要的计算系统之外，还需要使用专门的计算硬件。图形处理单元（GPU）已经经过优化可以快速将图形渲染到屏幕上，并且目前正在适应处理 VR 的特定性能需求。而且，显示接口芯片将输入视频转换为一条条的显示命令，最后，微控制器从传感设备收集信息并使用标准协议（如 USB）将它们发送到主计算机。



图 2.12：Oculus Rift DK2 耳机的拆卸（摄图为 ifixit）

图 2.12 显示了 Oculus Rift DK2 的硬件组件，该组件在 2014 年底推出。在图中的左下角可以看到一个智能手机屏幕作为显示屏，在图中上方还包含 IMU，显示接口芯片，USB 驱动芯片，用于驱动耳机上的 LED 以追踪 LED 的芯片集以及可编程微控制器的电路板，右下角是镜头，它可以使智能手机屏幕看起来“无限远”，但仍填满了用户视野的大部分区域。右上方显示的是灵活的电路，为嵌入在耳机中的 IR LED 提供电源（它们隐藏在红外透明塑料后面）。相机用于跟踪，其部件显示在上图的中心部分。

## 2.2 软件

从开发人员的角度来看，通过提供高级描述和让软件自动确定所有低级细节来编程实现 VR 系统将是一个理想的选择。在一个近乎真实的虚拟世界里，会有一个 VR 引擎，它的用途类似于当今用于制作视频游戏的游戏引擎。如果开发者在实现之前遵循许多模式，那么只需调用精心设计的软件库中的函数即可避免许多复杂的细节。但是，如果开发人员想要尝试

一些原创的东西，那么将不得不从头开始设计这些功能。这需要对 VR 基础知识有更深入的了解，同时也熟悉底层级别的系统操作。

不幸的是，我们距离拥有功能齐全的通用 VR 引擎还有很长的路要走。随着虚拟现实技术的应用越来越广泛，专门的 VR 引擎也可能出现。例如，一个可能是针对沉浸式电影摄影，而另一个则是面向工程设计。哪些组件将变得更像 VR“操作系统”的一部分，并将成为更高级别的“引擎”组件？鉴于目前的情况，开发人员可能会从零开始实施其 VR 系统的大部分功能。这可能涉及利用软件开发工具包（SDK）处理特定头戴式设备，以处理最底层级别的操作，例如设备驱动程序，头部跟踪和显示输出。另外，他们可能会发现自己使用最近已编译 VR 的游戏引擎，尽管它基本上是为屏幕上的视频游戏而设计的。这可以首先避免大量的工作，但是当有人想要实现不属于标准视频游戏时，这可能会比较困难。

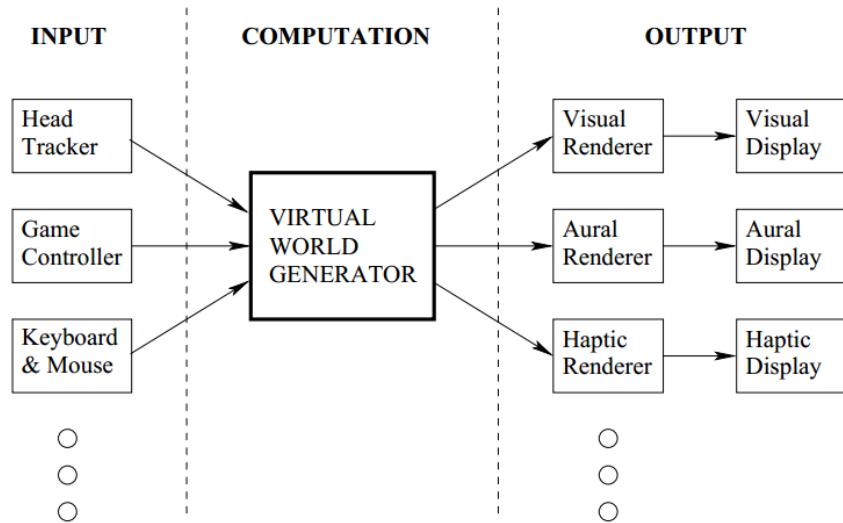


图 2.13：虚拟世界生成器（VWG）维护着另一个世界，它可能是合成的，真实的或者某种组合。从计算角度来看，输入是从用户和他的周围环境中接收到的，并且适当的场景被呈现给显示器。

需要哪些软件组件来产生 VR 体验？图 2.13 展示了一个高级视图，强调了虚拟世界生成器（VWG）的核心作用。VWG 接收来自低级系统的输入，表明用户在现实世界中正在做什么。头部跟踪器提供用户头部位置和方向的及时估计。键盘，鼠标和游戏控制器事件到达可准备处理的队列。VWG 的关键作用是保持足够的内部“现实”，以便渲染者可以提取他们需要的信息来计算其显示的输出。

### 虚拟世界：真实与合成

虚拟世界可以被完全合成。在这种情况下，许多三角形态在 3D 空间中定义，以及指示它们如何与光，声音，力等相互作用的材质属性。在计算机图形学领域较为认可的一个观点是，在合成模型中计算机生成的图像对 VR 来说非常重要，可参见第 7 章。另一方面，虚拟世界刻画了一个还原纪实的物理世界，它运用了现代相机，计算机视觉和同步定位和映射技术（SLAM），如图 2.14。极端之间存在许多可能性。例如，可以拍摄真实物体的相机图像，然后将其映射到虚拟世界中的合成物体上，这被称为纹理映射，是计算机图形学中的一种常见操作，见第 7.2 节。

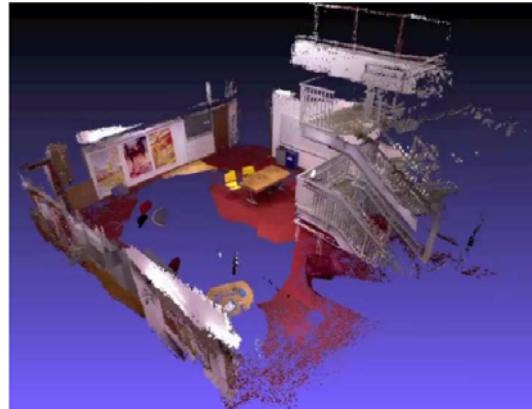


图 2.14：使用摄像机的颜色和深度信息，可以使用同步定位和映射（SLAM）技术自动提取 3D 模型。

### 匹配运动

VWG 的最基本操作是保持现实世界中的用户运动与虚拟世界之间的对应关系，见图 2.15。在现实世界中，用户的运动被限制在一个安全区域，我们称之为匹配区域。想象一下，匹配区域是真实和虚拟世界完美对齐的地方。最大的挑战之一就是障碍的不匹配：如果用户在虚拟世界中碰到了障碍而在现实世界中没有的话，该怎么办？在一个有座椅设备的体验中，用户戴着头戴式设备坐在椅子上。在这种情况下，匹配区域是一个小区域，例如一立方米，用户可以在其中移动头部。头部动作在两个世界中应该是相匹配。如果用户不受限于座位，则匹配的区域可以是整个房间或室外场地。请注意这里安全性成为一个问题，因为用户可能会饮酒，撞墙，或陷入只存在于现实世界中但在虚拟世界中不可见的凹坑。较大的匹配区域往往导致更大的安全问题。用户必须确保匹配的区域在现实世界中没有危险，否则开发人员应该在虚拟世界中使其可见。

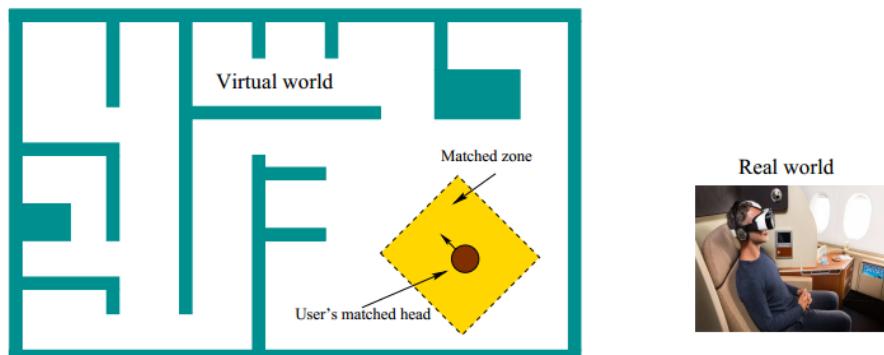


图 2.15：匹配区域是指用户在现实世界和虚拟世界中的一个显示关系。用户可以通过使用诸如游戏控制器等在虚拟世界中移动，而用户在现实世界中不会进行相应地移动。

虚拟世界中哪些动作应该反映在虚拟世界中？这在不同的 VR 体验中各不相同。在向头部显示图像的 VR 头戴式设备中，头部运动必须匹配，以便视觉渲染器在虚拟世界中使用正确的视点。身体的其他部位的匹配不那么重要，但如果用户需要进行手眼协调或查看身体的其他部位并期望实现自然移动时，则会变得十分重要。

### 用户运动

在许多 VR 体验中，用户想要在匹配区域外移动。这意味着用户在虚拟世界中自我移动，而这个移动在现实世界中是不匹配的。想象一下，你想要在真实世界中坐下来探索虚拟城市。

这应该如何实现？你可以拉一张地图，并指向你想要去的地方，用一个快速的传送操作把你送到目的地。一种流行的选择是通过操作游戏控制器，鼠标或键盘来在虚拟世界中移动自己。通过按下按钮或移动旋钮，你在虚拟世界中实现步行，跑步，跳跃，游泳，飞行等等操作。你也可以坐上虚拟世界中的一辆车，并操纵其控制来移动自己。这些操作当然很方便，但由于平衡和视觉感官不匹配，往往会导致晕眩，具体请参阅第 2.3, 10.2 和 12.3 节。

## 物理学

VWG 通过应用适当的数学变换来处理运动的几何方面。另外，VWG 通常实施一些物理学，以便随着时间的推移，虚拟世界的行为就像现实世界一样。在大多数情况下，力学的基本定律控制物体在虚拟世界中的移动方式。例如，如果你放下一个物体，那么它应该由于作用在其上的重力而加速到地面。一个重要的组成部分是碰撞检测算法，它确定两个或多个物体是否在虚拟世界中相交，如果发生碰撞，则需要适当的响应。例如，假设用户在虚拟世界中穿过墙壁探出头部。虚拟世界中的头部是否应该停下来，即使它在现实世界中继续前进？为了使它更复杂，如果你将装满篮球的大货车卸载到虚拟世界的一条繁忙街道上会发生什么？模拟物理学可能变得非常具有挑战性，并且这本身就是一门学科。有关虚拟世界物理的更多信息，请参见第 8.3 节。

除了处理运动物体的运动特征之外，物理学还必须考虑如何创建不可见的介质。比如，光线如何在环境中传播？光线如何与虚拟世界中的表面相互作用？什么是光源？声音和气味如何传播？这些与渲染问题相对应，分别在视觉和音频案例的第 7 章和第 11 章中介绍。

## 网络体验

在互联网 VR 体验的情况下，共享虚拟世界由服务器维护。每个用户都有一个独特的匹配区域。它们的匹配区域可能在现实世界中重叠，但必须小心，以避免不必要的碰撞。大多数情况下，这些区域是分离的，分布在地球周围。在虚拟世界中，用户交互（包括冲突）必须由 VWG 管理。如果多个用户在社交场合中互动，则匹配动作的负担可能会增加。当用户相互见面时，他们可以期望看到眼部、面部表情和肢体语言，可参见 10.4 节。

## VWG 的开发人员选择

总而言之，开发人员可以从 VR 头戴式设备供应商的基本软件开发套件（SDK）开始，然后从头开始构建自己的 VWG。SDK 应提供基本的驱动程序和接口，以访问跟踪数据并调用图形渲染库。在这种情况下，开发人员必须从头开始构建虚拟世界的物理模型，处理诸如头像移动，碰撞检测，照明模型和音频等问题。这为开发人员提供了最大限度的控制和优化性能的权限，然而在有些情况下它可能会换来实施上的困难，但在一些特定的情形中它可能又不会变的太过复杂。例如，对于 Google Street 浏览器（如上一章的图 1.9），其物理模型很简单：观看位置需要以方便的方式在全景图像之间来回切换，同时又要保持观看者在地面的位置感。在使用机器人进行远程呈现的情况下，VWG 将不得不考虑在物理世界中的移动。如果未能处理碰撞检测，可能会导致机器人破损，甚至人受伤。

另一方面，开发人员可以使用现成的 VWG，通过选择菜单选项和编写高级脚本来定制 VWG，以制作特定的 VR 体验。目前可用的例子有 OpenSimulator，World-Viz 的 Wizard，Unity 3D 和 Epic Games 的虚幻引擎。后两者是适用于 VR 的游戏引擎，并且是目前 VR 开发者中最受欢迎的游戏引擎。OpenSimulator 是 Second Life 的替代产品，用于构建虚拟社会。如前所述，使用这种更高级别的引擎可以让开发人员在短时间内轻松制作 VR 体验；然而，其缺点是难以做出制造商无法想象的高度原创体验。

## 2.3 人体生理学与认知

人类的身体并非是为 VR 而设计的。通过对感官进行人工刺激，我们正在破坏生物机制的运作，这些机制经历了数亿年时间在自然环境中演变而来。我们也向大脑提供与现实体验不完全一致的信息。在某些情况下，我们的身体可能会适应新的刺激，这可能会使得我们无法察觉 VR 系统中的缺陷。在另一些情况下，我们可能会提高解读含义不清的 3D 场景的能力。但不幸的是，在有很多情况下，我们的身体会产生疲劳或头痛等症状，部分原因是大脑比平常更高速地运转，以理解这类刺激。而最糟糕的情况是 VR 疾病的发作，通常会出现眩晕和恶心的症状。人体感知心理学是理解大脑如何将感官刺激转化为感知现象的科学。以下是 VR 中出现的一些典型问题，属于这个范畴：

- 该对象看起来有多远？
- 需要多少视频分辨率才能避免看到像素？
- 每秒多少帧的刷新率可以将运动视为连续性的？
- 用户的头部是否出现在虚拟世界的适当高度？
- 虚拟声音来自哪里？
- 为什么会感到眩晕？
- 为什么一个人的体验比另一个人更累？
- 什么是存在？

要回答以上甚至更多的问题，我们必须了解以下几点：1) 人体的基本生理结构，主要包括感官和神经通路；2) 基于实验的感知心理学的关键理论和见解；3) VR 系统对人类通常的感知过程产生的影响或副作用。

虚拟现实的感知往往无法引起开发人员的重视。在现实世界中，感知过程大多是我们看不见的。想想认出一个家庭成员需要花费多少精力，当你看到一个你熟悉的人时，这个过程会自动开始，立即结束，似乎不需要付出任何代价。科学家们进行了实验，揭示了与此类似的感知过程中究竟发生了多少工作。通过脑部神经研究，他们能够看到一小部分脑功能不正常时的影响。有些人患有面部失认症，这使得他们即使在几乎所有其他功能都正常时，也无法认出熟悉的人的面部，包括他们自己在镜子中的形象。科学家们也能够对大脑中某个神经元的状态进行记录，主要在动物身上进行实验，这可以揭示单个神经元对感官刺激的反应。

### 视觉错觉

认知人类感知复杂性的最普遍的一种方式是研究视觉错觉。视觉错觉会产生人们平常毫不在意却令人惊讶的结果。该项研究通过提供与日常生活中普通刺激不太相同的刺激来揭示人类视觉系统的一些缺陷。下图展示了其中的两种错觉。这些结果或许会激励你去了解人体感官和神经结构每一次感知所要进行的工作，并根据生活经验和生物结构对此进行解释。当然，在不理解感知过程的情况下对其进行干扰是十分不明智的。

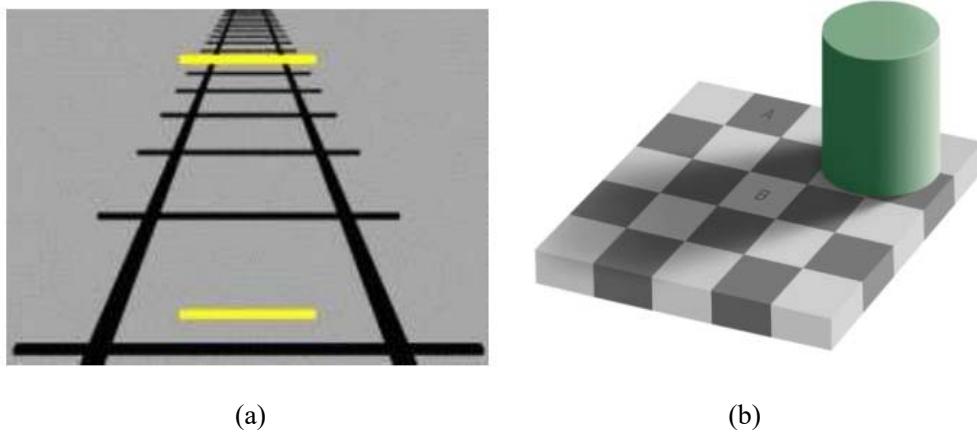


图 2.16 视觉错觉是一种非同寻常的视觉刺激，凸显出了人类视觉系统的局限性。 (a) Ponzo 错觉导致上部线段比下部线段显得更大，即使它们的长度相同。 (b) 阴影错觉使得 B 瓦片比 A 瓦片看起来更亮，尽管它们完全相同的灰度

### 感知分类

感知与错觉不仅限于人眼。图 2.17 显示了基本感觉的分类。其中传感器将能量源转换为电路中的信号。就我们的身体而言，这意味着刺激会转化为神经冲动。对于每种感觉，图 2.17 说明了刺激和将刺激转化为神经冲动的受体类型。将每个受体看作是针对特定类型刺激的传感器。这被称为感知系统选择性。在单个眼睛内，有超过 1 亿个光感受器精确捕捉可见光频率范围内的电磁能量。不同种类的光感受器甚至可以针对不同的颜色和光照水平，具体见 5.1 节。听觉，触觉和平衡感涉及运动，振动和重力，这些刺激被机械感受器感知。有关听觉的生理知识还会在 11.2 节和 11.3 节中介绍。触觉还包括温度的感知，有关触觉的内容还将在 13.1 节中介绍。平衡感有助于我们知道头部朝向哪个方向，包括“向上”的方向，将在 8.2 节继续介绍。最后，我们将味觉和嗅觉归为一类，称为化学感觉，依赖于化学感受器，这类信号的产生基于在舌头或鼻腔中物质的化学成分，详见 13.2 节。

感觉种类	刺激形式	接收单元	感官
视觉	电磁能量	感光器	眼睛
听觉	空气压力波	机械感受器	耳朵
触觉	组织扭曲	机械感受器	皮肤，肌肉
		a 热觉感受器	皮肤
平衡感	引力， 加速	机械感受器	前庭器官
味道/气味	化学成分	化学感受器	嘴，鼻子

图 2.17 人体感觉分类

要注意的是，这些感官具有等价物体，其中大部分会出现在 VR 系统中。想象一下，假如你设计了一款远程呈现的人形机器人，并期望通过 VR 设备进行交互。你可以通过代理机器人体验生活。数码相机将成为它的眼睛，而手机将成为耳朵。机器人身上可以安装压力传感器和温度计以提供触感。至于平衡感，可以安装一个 IMU。事实上，人类前庭器官和现代 IMUs 在产生信号方面有惊人的相似之处，参见 8.2 节。我们甚至可以安装化学传感器，如 pH 计，测量化学成分的各个方面，以提供味道和气味。

### 大脑

感觉发生在感官将刺激转换为神经冲动之后。根据最新研究[15]，人体含有约 860 亿个

神经元。大约有 200 亿个神经元为大脑皮层工作，处理知觉和许多其他更高级的功能，如注意力，记忆力，语言和意识。它是一个约三毫米厚的神经中枢，大量折叠以适配头骨结构。再如其他生物，蛔虫，果蝇和大鼠分别有 302,10 万和 2 亿个神经元。一头大象拥有超过 2500 亿个神经元，这比人类的还多！

只有哺乳动物有一个大脑皮层。老鼠的大脑皮层有约 2000 万个神经元。猫和狗分别为 300 万和 1.6 亿。大猩猩有大约 40 亿。一种名为长肢领航鲸的海豚，其大脑皮层中估计有 370 亿个神经元，大约是人类大脑皮质的两倍。然而，科学家声称更多的神经元并不意味着更良好的认知能力[220,273]。

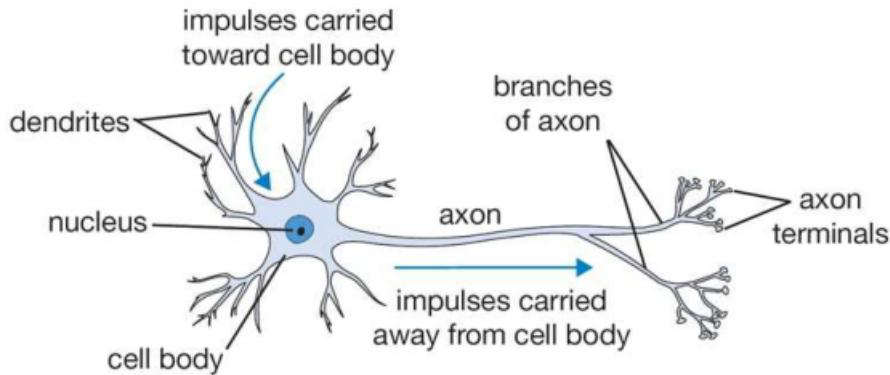


图 2.18 典型神经元，通过树突接收信号

感知和整体认知能力的另一个重要因素是神经元之间的相互联系。想象一个巨大的有向图，具有通常的节点和有向边。每个神经元的核或细胞体便是某种“处理”的节点。图 2.18 为我们展示了一种典型的神经元。树突是神经元的输入边缘，而轴突是输出边缘。通过树突网络，神经元可以汇集来自众多其他神经元的信息，这些神经元本身可能已经汇总了来自其他神经元的信息。结果通过轴突发送给一个或多个神经元。对于连接的轴突-树突对，信息交互发生在称为突触的中间结构中，其中传递电信号或化学信号。人脑中的每个神经元与其他神经元平均具有约 7000 个突触连接，这导致我们巨大的脑图中有约  $10^{15}$  个边缘！

### 分层处理

离开感官受体后，信号在神经元之间传播，最终到达大脑皮层。在传播途径上，实现分层处理，具体过程见图 2.19。通过选择性，每种受体对刺激在时间、空间、频率等各种方面的响应范围都变得相当窄。通过几个神经元后，将同时处理来自多个受体的信号，因而这种机制可以在刺激中检测到复杂的形式。在视觉感知情况下，特征检测器出现在早期的分层阶段，使人们能够检测到如边缘，角落和运动等特征。一旦进入大脑皮层，来自传感器的信号就会与我们生活中经历过的任何其他信息相结合，这些信息可能与刺激的解释有关，由此出现各种感知现象，例如识别脸部或识别歌曲。最终出现在大脑皮层的信息便映射出了我们周围甚至全景环境的情况。令人惊讶的是，有方法揭示，在神经元分布的某些情况下，受体间的空间关系可以维持不变。此外，在此前的 1.1 节中也已提到过空间识别细胞和定位细胞可以识别出熟悉环境的空间关系。

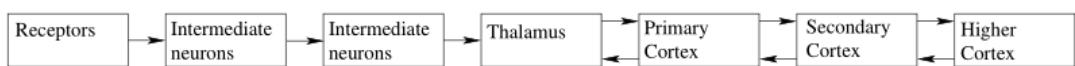


图 2.19 刺激信号的传播受体捕获的刺激通过神经元的分级网络传递。早期阶段，信号由多个受体组合而成，并向上传播。后期阶段，信息双向流动。

## 本体感觉

除了感官和记忆的信息之外，人体还具备本体感觉，这是一种能够感受到身体各部位相对位置并移动相应肌肉的行为。闭上你的眼睛，在空旷的地方移动手臂，尽管你可能无法精确地用手指进行触摸，但你应该大致知道你的手臂位于哪里。这种可以控制人体运动的信息对大脑来说非常重要，它向大脑其他部分发送一种名为“*efference copies*”的信号，传达身体已执行的运动。将这种感觉继续与机器人进行比较，它的等价物是在关节上安装的编码器，用来得到机器人移动的距离。因为本体感觉，从而有了以下这种有趣的现象：人们并不能把自己挠痒，因为我们知道自己的手指在哪里移动，但其他人却极有可能把你挠痒。所以这种本体运动信息对搔痒感的产生是极为重要的。

## 感觉融合

来自多种感觉和本体感觉的信号可以被同时处理，并与我们的生活经历相结合。在日常生活中，没有虚拟现实或药物，我们的大脑对这些输入组合以一种连贯，一致和熟悉的方式进行解释。任何干扰这个处理过程的行为都可能导致感官数据的不匹配，此时大脑会以各种方式作出紧急反应，最后可能使得我们没有意识到这场冲突，但我们会感到疲惫不堪。更糟的是我们可能会出现头晕或恶心的症状。在其他一些情况下，大脑可能会让我们有意识地感受到冲突，以明白这是人为体验，这与目前 VR 体验还未能令人沉浸于虚拟世界的情况相对应。为了获得有效且舒适的 VR 体验，VR 试验最重要的是要了解大脑对于虚拟场景是如何反应的。

在虚拟现实的背景下，最糟糕的一种感官冲突就是动感，这是自我运动的幻觉。当你的视觉向大脑告知你正在加速时，冲突便会产生，此时你的平衡感会认为你并没有在动。而当人们真正走在大街上时，他们的平衡感和视觉是匹配的。即使没有 VR，我们也可能在之前就经历过这种感受。当你遇到交通拥堵看到正在向前移动的车辆的同时，你可能会觉得自己正在向后移动。在 19 世纪 90 年代，亚玛利雅湖建造了一套游乐设施，该设施包括一个始终保持静止的秋千，而围绕秋千的整个房间进行前后摇摆（图 2.20）。在 VR 中，相对运动错觉是由 2.2 节中描述的运动行为所引起的。例如我们使用控制器加速前进，那么只有我们的眼睛会感受到加速度，前庭器官并不能。有关缓解此问题的方法会在后续 10.2 节中介绍。

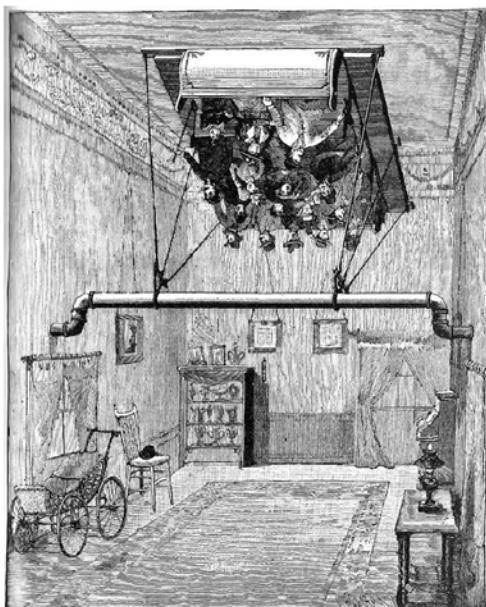


图 2.20 十九世纪九十年代，通过旋转周围的房间来创造虚拟的摇摆体验。这被称为摇摆幻觉。尝试过它的人既觉得有趣又会感到恶心。

## 适应性

人体感官系统的一个普遍特征是适应性，这意味着刺激的感知效果会随着时间而改变。例如，飞机或汽车的电机噪声感知到的响度会在几分钟内逐渐减小。对于视觉而言，人眼光学系统和感光器会调节以适应环境亮度。经过长时间的感知训练，适应能力也可以提升，参见 12.1 节。在军事训练中，士兵所经历的疾病似乎低于预期，可能是长期的风吹日晒所致 [171]。有趣的是，具有丰富经验的视频游戏玩家似乎也是如此。那些投入了大量时间和精力在第一人称射击游戏上的玩家，也可以更快地适应 VR 场景。

因此适应性也是 VR 的关键因素。通过反复的体验，开发者可能会对一个原本感到恶心的体验逐渐适应，但这也会对他们面向大众的开发产生不利。但从另一方面来讲，通过重复的有针对性的培训，开发人员在调试时会有更敏锐的观察能力，常见的例子包括：

- 大量跟踪延迟，这会对视野的平稳性产生影响。
- 左右眼视图已交换。
- 物体只存在于某只眼睛的视野中。
- 双眼显示图像的延迟不同。
- 直线扭曲。

## 心理物理学

心理物理学是对由物理刺激产生的感知现象的科学研究。例如，在什么情况下人们会说某个物体是“红色”的？刺激与进入眼睛的光线相对应，感知现象是大脑中形成“红色”概念的过程。感知现象的其他例子有“直线”，“更大”，“更响”，“发痒”，“酸”等等。图 2.21 展示了一个心理物理实验的典型场景。由于某个参数是可以变化的，例如光的频率，通常存在一定范围内的值，受试者不能可靠地对现象进行分类。例如，可能在某个区域内人们并不能确定光线是否为红色。在某种极端情况下，他们可能将其归类为“红色”，而在另一种极端情况下，他们则认为光线“不是红色”，对于介于两者之间的情况，记录检测的概率。在 12.4 节中还将继续讨论如何设计和实施这些实验。

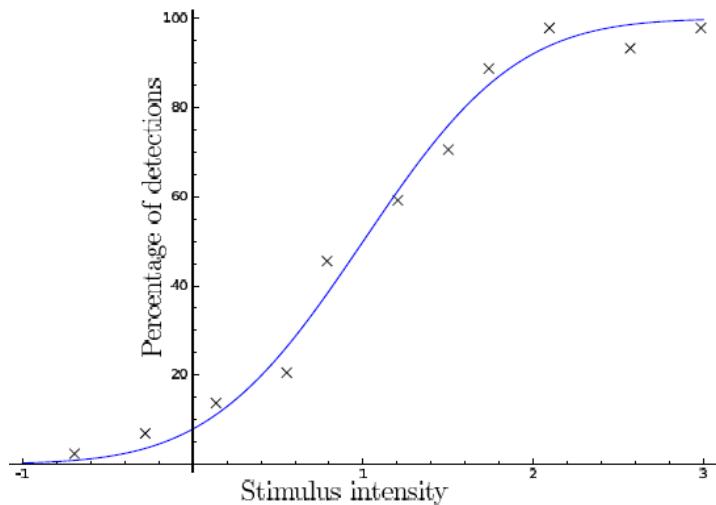


图 2.21 最基本的心理测量函数。就该函数而言，随着刺激强度的增加，被检测到实验现象的人的比例增加。沿着曲线对应于 50% 的点表示刺激强度的临界阈值。

## 史蒂文斯幂律

心理物理学中最著名的规律之一是史蒂文斯幂律，它描述了物理刺激强度与其感知强度之间的关系[311]，认为两者存在指数关系：

$$p=cmx \quad (2.1)$$

其中：

- $m$  是刺激强度，
- $p$  是感知强度，
- $x$  将刺激强度与感知强度相关联，是等式中最重要的部分，
- $c$  是一个无关紧要的常数，取决于单位。

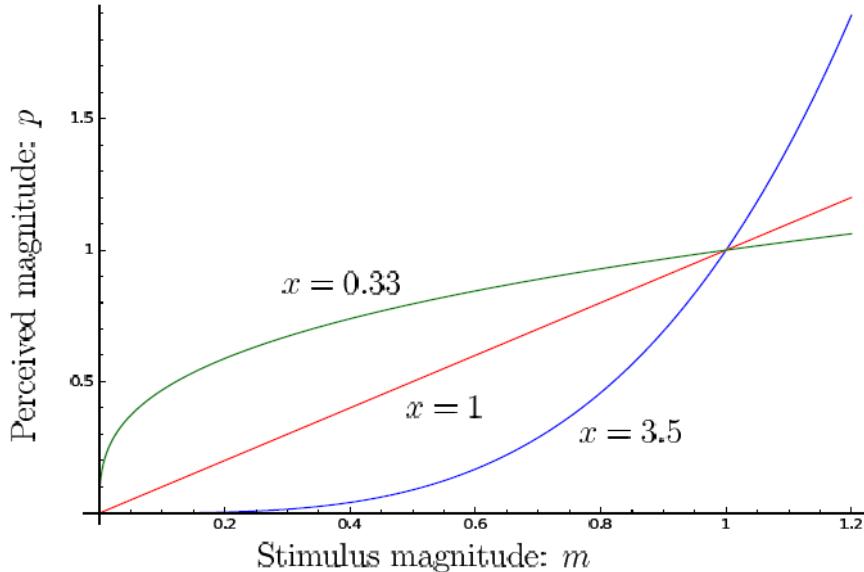


图 2.22 史蒂文斯幂律捕捉到了刺激强度与其感知强度之间的关系。该模型是指数曲线，具体参数取决于刺激类型。

结合图 2.22 来看，当  $x=1$  时，公式 (2.1) 为线性关系， $p=cm$ 。这方面的一个例子是我们对眼前孤立线段长度的感知，我们感知的长度与其实际长度是成正比的。但更有趣的情况是当  $x \neq 1$  时，对于黑暗中感知目标亮度的情况， $x=0.33$ ，这意味着感受到的光强度变化比实际的亮度变化更小。另一方面，人们感受通过手指的电流时的  $x=3.5$ ，因而微弱的电击就可以带来巨大的疼痛感。

### 最小可觉差

另一个重要的心理物理学概念是最小可觉差 (JND)，表示人类或者动物对于某一特定的感官刺激所能察觉的最小改变，在具体实验中，需要至少 50% 的测试者感受到这一最小变化。对于大的刺激变化，几乎所有测试者都会察觉到变化。反之如果变化太小，几乎没有测试者会注意到。因而实验的难点就在于找到这个 50% 变化的测试点。

考虑不同强度刺激下的 JND，例如亮度。那么 JND 本身是如何随着强度的变化而变化的呢？根据韦伯定律得到了如下关系：

$$\frac{\Delta m}{m} = c \quad (2.2)$$

式中  $\Delta m$  代表 JND， $m$  为刺激强度， $c$  为常数。

## 实验设计

VR 破坏了用户通常的感知过程。从本章前述中应该明白一点，人们所提出的 VR 系统和 VR 体验需要从用户的角度进行评估，以了解 VR 技术是否产生了期望的效果，同时根据评估结果去避免不良影响。这需要应用科学方法进行观察，制定假设，并设计验证其有效性的实验。当以为受试者时，这将变得极具挑战性。有多少测试样本才足够？当他们适应了实验环境之后会发生什么？他们的先前经验会对实验产生怎样的影响？如果他们在测试的那一天身体有恙，该怎么办？他们早餐吃了什么，会有什么影响？这些问题的答案都可能会影响最终的结果。

再深入思考下去情况将变得更为糟糕。假设测试者已经知道了实验的假设，这很可能会影响他们的反馈。另外，实验数据应是什么形式？是要求他们填写一份问卷，还是根据头部运动，心率和皮肤电导等测量数据推断他们在测试时的状态？这些选择也很关键。有关这个主题的更多信息，参见 12.4 节。

## 拓展阅读

本章介绍的软硬件技术正在迅速发展。我们可以随时在互联网上查找到 VR 开发者所设计的最新型的头戴式设备和相关工具。但是，VR 所包含的核心概念到目前为止也几乎没有改变，这也会在接下来的章节中进行介绍。为了更全面地认识人类的感知过程，请参阅[204]和其他相关书籍中“感觉与感知”的部分。

# 第三章 虚拟世界的几何模型

2.2节介绍了虚拟世界生成器(VWG)，它维护着虚拟世界的几何和物理规律。本章涵盖几何部分，关于制作模型与按需求改换。这些模型可能包括建筑物的墙壁，家具，天空中的云彩，用户的虚拟形象等等。第3.1节介绍了关于如何定义一致且有用的模型的基础知识。第3.2节解释了如何应用能够在虚拟世界移动它们的数学变换。这涉及两个组成部分：转换（更改位置）和旋转（改变方向）。第3.3节介绍了最佳的表达并操纵3D旋转的方式，这是移动模型中最复杂的部分。第3.4节将介绍如果我们试图从特定的角度“看”虚拟世界，它将如何呈现。这是视觉渲染中的几何部分，在第7章中将会介绍。最后，第3.5节将所有的变换结合在一起，以便您可以看到如何从定义模型转到让它出现在显示屏的正确位置。

如果你使用高级引擎来构建VR体验，那么大多情况下本章中的概念似乎不是必需的。你可能只需要从菜单中选择选项并编写简单的脚本。然而，理解基本的转换，例如如何表达3D旋转或移动相机视角，对于使用软件做你想做的事情还是至关重要的。此外，如果你想从零开始构建虚拟世界，或者至少想了解在软件引擎的背后发生了什么，那么本章至关重要。

## 3.1 几何模型

我们首先需要一个虚拟世界来包含几何模型。为此，一个具有笛卡尔坐标的3D欧几里德空间就足够了。因此，让 $\mathbb{R}^3$ 表示虚拟世界，其中每个点都表示为一个三元组实值坐标： $(x, y, z)$ 。我们的虚拟世界的坐标轴如图3.1所示。我们在此书中将始终使用右手坐标系，因为它们代表了整个物理学和工程学的主要选择；然而，左手系统也在一些地方出现，最值得注意的是微软的DirectX图形渲染库。在这些情况下，与右手系统中的方向相比，三个轴中的一个指向相反的方向。这种不一致可能导致在编写软件时几个小时的疯狂；因此，要了解它们之间的差异和要求如果您混合使用两种软件或模型。如果可能，避免混合使用右手和左手系统。

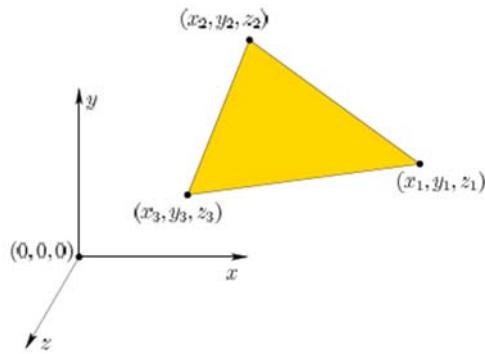


图3.1：虚拟世界中的点在y轴向上的右手坐标系中给出坐标。原点 $(0, 0, 0)$ 位于轴相交的点上。还显示了一个由其三个顶点而决定的3D三角形，每个顶点都是 $\mathbb{R}^3$ 中的一个点。

几何模型由 $\mathbb{R}^3$ 中的曲面或实心区域组成，并包含无限多点。由于计算机中的表示是有限的，模型是根据基元来定义的，其中每个基元表示一组无限点。最简单和最有用的基元是一个3D三角形，如图3.1。对应于“内部”所有点和三角形边界上的平面由三角形顶点的坐标完全指定：

$$((x_1, y_1, z_1), (x_2, y_2, z_2), (x_3, y_3, z_3)). \quad (3.1)$$

为了模拟虚拟世界中的复杂物体，我们可以将大量的三角形排列成一个网格，如图 3.2 所示。这引发了很多重要的问题：

1. 我们如何确定，VR 用户查看每个三角形时，它们是什么样子的？
  2. 我们如何让对象“移动”？
  3. 如果物体表面是尖锐弯曲的，那么我们应该使用弯曲的基元而不是试图用微小的三角形拟合弯曲的物体吗？
  4. 物体的内部是模型的一部分吗，还是只有它的表面？
  5. 是否有有效的算法来确定哪些三角形与沿面指定的三角形相邻？
  6. 我们应该避免重复许多相邻三角形共有的顶点坐标吗？
- 我们以相反的顺序来解决这些问题。

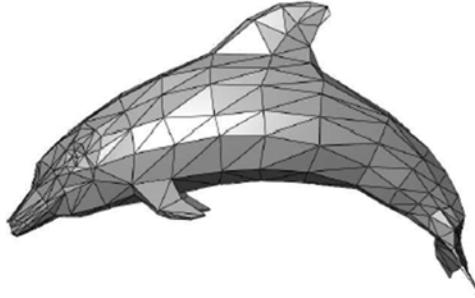


图 3.2：一只海豚的几何模型，由 3D 三角形的网格组成。（来自维基百科用户 Chrschn）

## 数据结构

考虑列出文件或存储数组中的所有三角形。如果三角形形成一个网格，那么在多个三角形中，大部分或全部顶点将会被共享。这显然是浪费空间。另一个问题是我们会经常想要在模型上执行操作。例如，当移动一个物体后，我们能确定它是否会与另一个对象发生碰撞吗（在 8.3 节会提及）？一个典型的低级任务可能是确定哪些三角形与给定的三角形共享一个共同的顶点或边。这可能需要在三角形列表中线性搜索以确定它们是共享一个顶点还是两个。如果有数以百万计的三角形，这并不罕见，那么它会因为重复执行此操作而花费许多。

由于以上及更多原因，几何模型通常利用精炼的数据结构编码。数据结构的选择应该取决于将在模型上执行的操作。其中最有用和最常见的是双连接边列表，也称为半边数据结构 [2, 10]。如图 3.3。在这个和与其类似的数据结构中，有三种数据元素，面、边和定点。它们分别在模型中代表着二维、一维和零维部分。在我们的例子中，每个面元素代表一个三角形。每条边代表无重复的一个或两个三角形的边界。每个顶点都是一个或多个三角形之间共享的，同样没有重复。数据结构包含相邻面，边，和顶点之间的指针，以便算法可以按照与它们如何连接在一起的方式快速遍历模型组件。

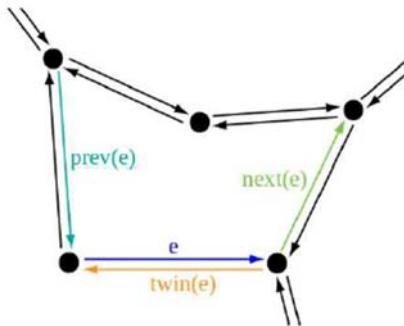


图 3.3：这里显示了一个五条边界表面的双连接边列表的一部分。每个半边结构  $e$  存储着沿着表面边界指向前一条和下一条边的指针。它还存储一个指向它自己的孪生半边结构的指针，这是相邻面边界的一部分。  
(图来自维基百科用户 Accountalive)。

### 内部 vs 外部

现在考虑物体的内部是否属于模型的一部分（回忆图 3.2）。假设网格中三角形完美匹配，因此每条边都恰好与两个三角形相交，并且除非三角形沿表面相邻，否则它们不会相交。在这种情况下，模型会形成一个完整的在物体内部与外部之间的屏障。如果我们假设用气体填充内部，并且它们不会泄漏到外部。这是一个连贯模型的例子。如果内部或外部的概念对 VWG 至关重要，则需要使用此类模型。例如，一分钱可能在海豚的内部，但是它不与任何边界三角形相交。它始终需要被检测吗？如果我们移除一个三角形，那么假想的气体就会泄漏。物体的内外将不再有明显的区别，它使得这个关于硬币和海豚的问题难以回答。在极端情况下，在空间中可以有一个三角形。它显然没有天然的内部或外部。在极端情况下，模型可能会像多边形汤一样糟糕，它们是一堆不能很好匹配的，并且甚至有内部相交的三角形。总之，在建模时要多加小心，以便你稍后要执行的操作在逻辑上保持清晰。如果你要使用高级设计工具，比如 Blender 或者 Maya 来建模，那么连贯模型将会被自动建立。

### 为什么是三角形

通过上述问题，继续向前思考。使用三角形是因为它们对于算法来说是最方便处理的，特别在硬件实施方面。GPU 实现倾向于偏向于较小的表示，以便可以并行地将多个指令的紧凑列表应用于多个模型部分。当然可以使用更复杂的原型，如四边形，样条线和半代数曲面 [3, 4, 9]。这可能导致更小的模型尺寸，但通常会因处理这样的原型而带来更大的计算开销。例如，两个样条曲面很难确定是否碰撞，相比于两个 3D 三角形来说。

### 固定模型 vs 可移动模型

以后再虚拟世界中将会有两种模式，它们被嵌入在  $\mathbb{R}^3$  中：

- 固定模型，它们永远保持相同的坐标。典型的例子包括街道，地板和建筑物。
- 可移动模型，它们可以转换为多种位置和方向。例子包括车辆，虚拟人像和小家具。运动可以通过多种方式引起。使用跟踪系统（第 9 章），模型可能会移动以匹配用户的动作。或者，用户可能操作控制器来移动虚拟世界中的对象，包括自己的形象。最后，物体可以根据虚拟世界中的物理规律自行移动。第 3.2 节将涵盖将模型移动到他们想要的位置的数学运算，第 8 章将描述速度，加速度以及其他关于运动的物理概念。

### 选择坐标轴

一个经常被忽略的问题就是根据位置和规模，为模型选择坐标轴。如果这些在一开始就

被很好地定义了，那么可以避免许多繁琐。如果虚拟世界与真实环境中熟悉的环境相对应，那么轴的缩放应该与常见单位相匹配。例如， $(1, 0, 0)$  应该对应  $(0, 0, 0)$  右边一米。放置原点  $(0, 0, 0)$  在一个方便的位置也是明智的。通常， $y = 0$  对应建筑物的底层或者海平面。 $x = 0$  和  $z = 0$  的位置可能是虚拟世界的中心，以便根据符号能很好地划分象限。另一个常见的选择是当从上面观看世界时，将其放在左上角，使得所有  $x$  和  $z$  坐标都是非负的。对于可移动模型，原点的位置和轴的方向非常重要，因为它们会影响模型的旋转。这部分在 3.2 和 3.3 节介绍旋转后应该会变得更清楚。

### 查看模型

当然，VR 最重要的方面之一是当在显示器上观看时，模型看上去如何。这个问题分为两部分。第一部分涉及确定虚拟世界中的点应该出现在显示器上的什么位置。这是通过查看第 3.4 以及第 3.5 节的其他转换产生的最终结果完成的。第二部分涉及考虑虚拟世界中定义的光源和表面介质之后模型中每个部分应该如何呈现。这是渲染问题，这第 7 章中会有介绍。

## 3.2 改变位置和方向

假设可移动模型已被定义为三角形网格。为了移动它，我们对每个三角形的每个顶点应用单一变换。本节首先考虑变换的简单情况，其次是相当复杂的旋转情况。通过结合变换和旋转，模型可以放置在任何地方，以及虚拟世界  $\mathbb{R}^3$  中的任何方向。

### 变换

考虑以下 3D 三角形，

$$((x_1, y_1, z_1), (x_2, y_2, z_2), (x_3, y_3, z_3)), \quad (3.2)$$

顶点坐标表示为通用常量。

设  $x_t$ ,  $y_t$  和  $z_t$  是分别沿  $x$ ,  $y$  和  $z$  轴的我们想要改变的三角形位置量。改变位置的操作称为变换，由下式给出

$$\begin{aligned} (x_1, y_1, z_1) &\rightarrow (x_1 + x_t, y_1 + y_t, z_1 + z_t) \\ (x_2, y_2, z_2) &\rightarrow (x_2 + x_t, y_2 + y_t, z_2 + z_t) \\ (x_3, y_3, z_3) &\rightarrow (x_3 + x_t, y_3 + y_t, z_3 + z_t) \end{aligned} \quad (3.3)$$

其中  $a \rightarrow b$  表示在变换之后  $a$  被  $b$  替换。将 (3.3) 应用于模型中的每个三角形将会将其全部转换至所需的位置。如果三角形排列成网格，那将变换单独应用于顶点就足够了。所有的三角形将保留它们原有的大小和形状。

### 相关性

在变换变得太复杂之前，我们想要提醒你要能正确解释它们。图 3.4 (a) 和 3.4 (b) 展示了一个例子：一个三角形以  $x_t = -8$  和  $y_t = -7$  变换。该顶点坐标与在图 3.4 (b) 和 3.4 (c) 中是相同的。图 3.4 (b) 显示了我们到目前为止打算覆盖的情况：三角形被解释为在虚拟世界中发生移动。但是，图 3.4 (c) 显示了另一种可能：虚拟世界的坐标已被重新分配，三角形更加接近原点。这相当于让整个世界移动了，而三角形是唯一不移动的部分。在这种情况下，变换应用于坐标轴，但它们是负的。当我们应用更多一般变换时，它扩展了，因此变换坐标轴会导致相应地移动模型的变换的逆。否定就是在变换情况下的逆。

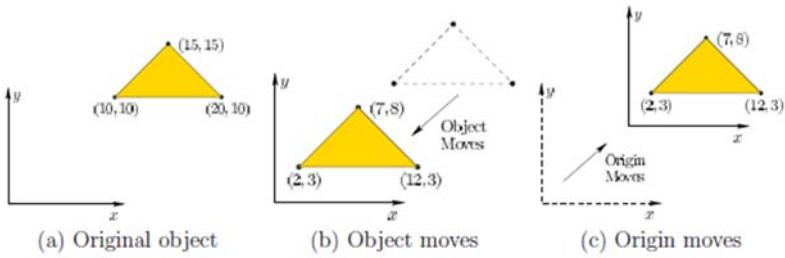


图 3.4: 尽管数学上是一样的, 每个转换还是有两种可能的解释。这是一个 2D 的例子, (a) 中定义了一个三角形。我们可以通过  $x_t = -8$  和  $y_t = -7$  来变换这个三角形以获得 (b) 中的结果。如果我们想要固定三角形, 但将原点在  $x$  方向上往前移动 8, 在  $y$  方向上往前移动 7, 那么三角形顶点的坐标改变方式完全相同, 如 (c) 所示。

因此, 我们有一种“相对性”: 物体是移动的还是它周围的世界在移动? 这个想法在 3.4 节会变得很重要当我们想改变视角。如果我们站在原点看三角形, 那么结果在任何一种情况下看起来都是相同的; 但是, 如果原点移动, 那么我们也会随之移动。一个深刻的知识问题在此出现。如果我们认为自己已经移动了, 那么 VR 的缺陷可能会增加即使它是移动的对象。换句话说, 我们的大脑会最佳地猜测哪种类型的动作发生, 但有时还是会出错。

### 为旋转作准备

我们如何让汽车上的车轮滚动? 或者把桌子翻一个面? 为了完成这些, 我们需要改变模型在虚拟世界中的定位。这个改变定位的操作改称为旋转。不幸的是, 三维旋转比变换复杂许多, 这给工程师和开发人员带来无数挫折。为了让 3D 旋转概念更为清晰, 我们首先从一个更简单的问题开始: 二维线性变换。

考虑一个 2D 虚拟世界, 其中的点坐标  $(x, y)$ 。你可以把它想象成我们最初的 3D 虚拟世界中的一个垂直平面。现在考虑一个通用的  $2 \times 2$  矩阵

$$M = \begin{bmatrix} m_{11} & m_{12} \\ m_{21} & m_{22} \end{bmatrix} \quad (3.4)$$

其中四项的每一个可以是任何实数。我们看看当这个矩阵乘以点  $(x, y)$  后会发生什么, 当它被写为列向量时。

执行乘法, 我们有

$$\begin{bmatrix} m_{11} & m_{12} \\ m_{21} & m_{22} \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} x' \\ y' \end{bmatrix}, \quad (3.5)$$

其中  $(x', y')$  是变换点。使用简单的代数, 矩阵乘法产生

$$\begin{aligned} x' &= m_{11}x + m_{12}y \\ y' &= m_{21}x + m_{22}y. \end{aligned} \quad (3.6)$$

使用 (3.3) 中的符号,  $M$  是  $(x, y) \rightarrow (x', y')$  的变换。

### 将二维矩阵应用于点

假设我们放置两个点  $(1, 0)$  和  $(0, 1)$  在平面上。它们分别位于  $x$  和  $y$  轴上, 距离原点的距离  $(0, 0)$  一个单位。使用向量空间, 这两点就是标准单位矢量 (有时写为  $i$  和  $j$ )。如

果我们将它们替换成 (3.5)，看看会发生什么：

$$\begin{bmatrix} m_{11} & m_{12} \\ m_{21} & m_{22} \end{bmatrix} \begin{bmatrix} 1 \\ 0 \end{bmatrix} = \begin{bmatrix} m_{11} \\ m_{21} \end{bmatrix} \quad (3.7)$$

和

$$\begin{bmatrix} m_{11} & m_{12} \\ m_{21} & m_{22} \end{bmatrix} \begin{bmatrix} 0 \\ 1 \end{bmatrix} = \begin{bmatrix} m_{12} \\ m_{22} \end{bmatrix}. \quad (3.8)$$

这些特殊点只选择 M 上的列向量。这意味着什么？如果 M 应用于模型转换，则 M 的每一列确切地表示每个坐标轴是如何改变的。

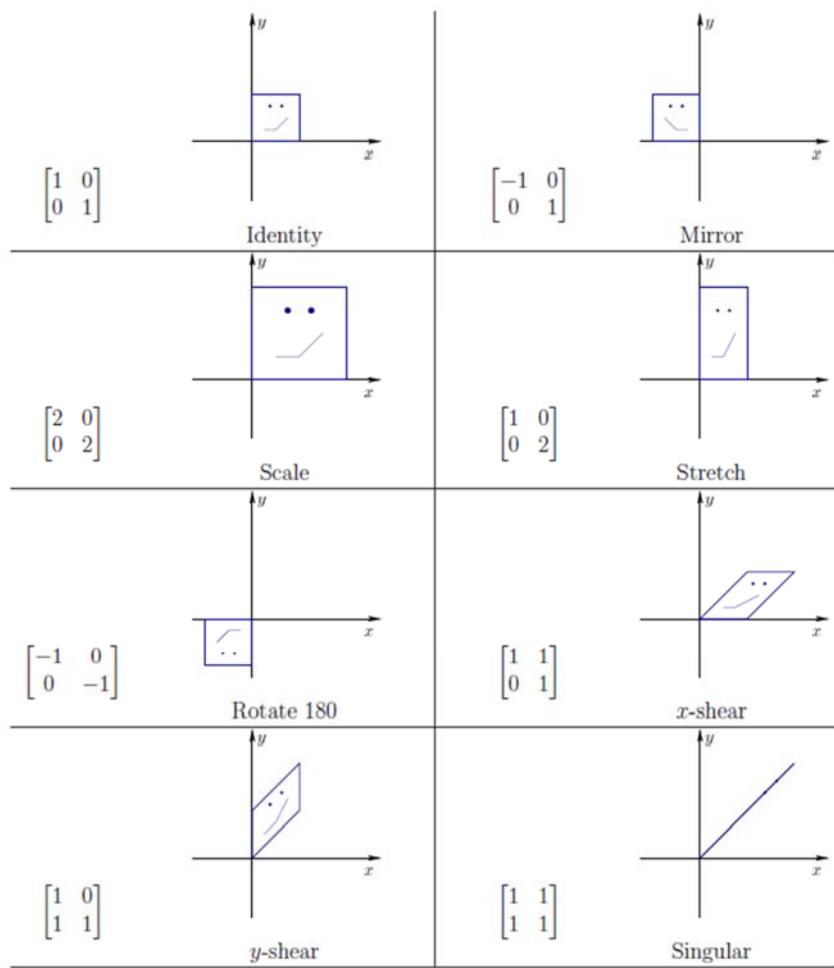


图 3.5：八个应用于变换方形面的不同矩阵。这些例子很好地从定性的角度涵盖了所有可能的情况。

图 3.5 说明了将各种矩阵 M 应用于模型的效果。从右上角开始，单位矩阵不会导致坐标改变： $(x, y) \rightarrow (x, y)$ 。第二个例子会导致翻转，就像镜子被放置在 y 轴上一样。在这种情况下， $(x, y) \rightarrow (-x, y)$ 。第二行是缩放的例子。左边的矩阵产生  $(x, y) \rightarrow (2x, 2y)$ ，其中尺寸加倍了。右边的矩阵只在 y 方向延伸模型，导致纵横比失真。在第三行中，似乎左边的矩阵产生了关于 x 和 y 的镜像轴。这是事实，除了镜像的镜像恢复了原始图像。因此，这对应于 180 度 ( $\pi$  弧度) 旋转的情况，而不是镜像。右边的矩阵产生沿 x 方向的剪切： $(x, y) \rightarrow (x + y, y)$ 。位移量与 y 成比例。

在最下面一行中，左边的矩阵显示在  $y$  方向的偏斜。最终的矩阵起初可能会引起更多的偏差，但它是退化的。当应用  $M$  时，二维形状会折叠成单一的维度： $(x, y) \rightarrow (x + y, x + y)$ 。这对应于奇异矩阵的情况，意味着它的列不是线性独立的（它们实际上是相同的）。一个矩阵当且仅当它的行列式为零时才是奇异的。

### 只有一些矩阵产生旋转

图 3.5 中的例子涵盖了各种  $2 \times 2$  矩阵  $M$  之间的主要定性差异。其中两个是旋转矩阵：0 度旋转，和 180 度旋转的单位矩阵。在所有可能的  $M$  集合中，哪些可能是有效的旋转？我们必须确保模型不会被扭曲。这是通过确保  $M$  满足以下规则来实现的：

1. 没有伸展的轴。
2. 没有剪切。
3. 没有镜像。

如果没有违反这些规则，那么结果就是旋转。

为了满足第一条规则， $M$  的列必须有单位长度：

$$m_{11}^2 + m_{21}^2 = 1 \text{ and } m_{12}^2 + m_{22}^2 = 1. \quad (3.9)$$

图 3.5 中的缩放和剪切变换违反了这一点。

为了满足第二条规则，坐标轴必须保持垂直。否则，会发生剪切。由于  $M$  的列表示轴是如何变换的，那么此规则意味着它们的内部（点）积为零：

$$m_{11}m_{12} + m_{21}m_{22} = 0. \quad (3.10)$$

图 3.5 中的剪切变换违反了这个规则，这明显导致模型中的直角被破坏。

满足第三条规则要求  $M$  的行列式是正数。当满足前两条规则后，唯一可能的行列式是 1（正常情况下）和 -1（镜像情况）。因此，该规则意味着：

$$\det \begin{bmatrix} m_{11} & m_{12} \\ m_{21} & m_{22} \end{bmatrix} = m_{11}m_{22} - m_{12}m_{21} = 1. \quad (3.11)$$

图 3.5 中的镜像示例导致  $\det M = -1$ 。

第一个约束 (3.9) 表示每列必须都被选择以使它的组件位于一个以原点为中心的单位圆上。在标准的平面坐标，我们通常将这个圆的方程写作  $x^2 + y^2 = 1$

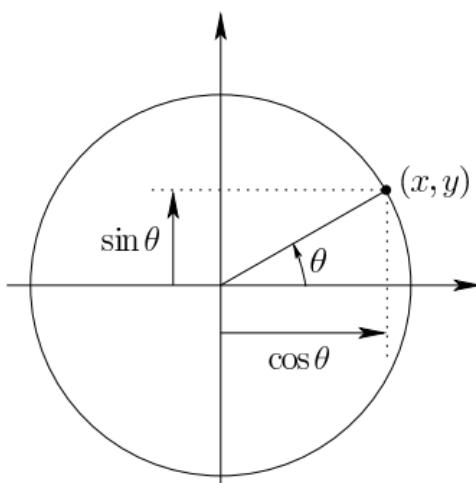


图 3.6 在单位圆中，可以用参数  $\theta$  ( $0 \sim 2\pi$ ) 来表示圆上点的坐标

第一个约束条件 (3.9) 表示每一列的元素都要落在单位圆上，标准形式写为  $x^2+y^2=1$ 。我们也可以用角度 $\theta$ （弧度  $0 \sim 2\pi$ ）来表示坐标  $(x, y)$ ，如图 3.6 所示：

$$x = \cos\theta \quad y = \sin\theta \quad (3.12)$$

我们用角度 $\theta$ 代替点的坐标  $(x, y)$ ，令  $m_{11}=\cos\theta \quad m_{21}=\sin\theta$ ，式 (3.4) 可以写为

$$\begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} \quad (3.13)$$

$m_{12}, m_{22}$  由式 (3.10) 和 (3.11) 唯一确定。通过 $\theta$ 的变化 ( $0 \sim 2\pi$  范围内)，可以表示所有 2D 旋转。

接下来我们讨论旋转的自由度 (DOF)。一开始的时候，矩阵  $M$  的四个元素是可以任意取值的，也就是说有 4 个自由度。(3.9) 中的约束条件减少两个自由度，(3.10) 的约束条件也减少一个自由度，(3.11) 并没有减少自由度，它只去掉了半可能的变换，因为这些变换是另一半的镜像翻转。最后，我们得出结论，2D 旋转只有一个自由度，由参数 $\theta$ 决定；而且，这些旋转都可以用单位圆的点来表示。

### 3D 旋转

现在我们尝试将 2D 旋转的模式扩展到三维空间中，式 (3.4) 矩阵的三维形式如下所示，包含 9 个元素：

$$M = \begin{bmatrix} m_{11} & m_{12} & m_{13} \\ m_{21} & m_{22} & m_{23} \\ m_{31} & m_{32} & m_{33} \end{bmatrix} \quad (3.14)$$

因此，一开始 3D 旋转有着 9 个自由度，同样的，我们也要知道一个合法的 3D 旋转需要哪些限制条件。首先，每一列必须保证是单位长度。例如， $m_{11}^2 + m_{21}^2 + m_{31}^2 = 1$ 。这就是说每一列中的元素必须落在单位球面上。因此，单位长度的限制将 DOF 减少到 6 个。然后，根据正交轴定理，取矩阵的任意两列，它们的内积必须为 0。例如，我们选择前两列

$$m_{11}m_{12} + m_{21}m_{22} + m_{31}m_{32} = 0 \quad (3.15)$$

同理，对于后两列、一三列同样有如上的约束。每一种情况减少一个自由度，我们还剩下 3 个自由度。为了避免镜像出现，矩阵  $M$  的行列式要为 1，但这不减少自由度。

最后，我们得到了一系列满足代数约束的矩阵，然而，和 2D 旋转不同，它并不能用单位球面上的点来表示。我们只知道它有 3 个旋转自由度，意味着它应该可以由 3 个独立的参数来表示，(3.14) 中的 9 个元素均可由这 3 个参数计算出。

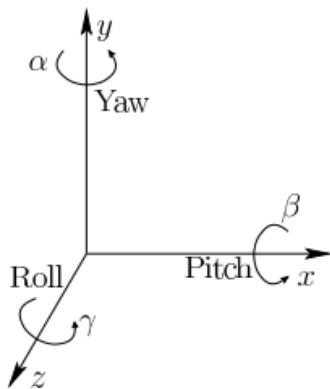


图 3.7 用偏航角 (yaw)、俯仰角 (pitch)、翻滚角 (roll) 可以描述任意三维旋转

## 偏航角、俯仰角、翻滚角

如何确定这 3 个参数呢？我们很容易想到，是不是可以构建几个 2D 旋转变换来描述 3D 旋转，如图 3.7 所示。先考虑绕 z 轴的旋转。我们用翻滚角 (*roll*) 来表示绕 z 轴逆时针方向的旋转  $\gamma$ 。旋转矩阵如下

$$R_z(\gamma) = \begin{bmatrix} \cos \gamma & -\sin \gamma & 0 \\ \sin \gamma & \cos \gamma & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad (3.16)$$

可以看出，矩阵的左上部分就是 2D 旋转矩阵，仅仅是  $\theta$  变成了  $\gamma$ 。这是因为它实际上就是 xy 平面上的 2D 旋转。矩阵的剩余项看上去是一个单位阵，这使得经过翻滚角旋转后 z 方向保持不变。

相似地，我们用俯仰角 (*pitch*) 来表示绕 x 轴的逆时针旋转  $\beta$ 。在这种情况下，x 轴保持不变而在 yz 平面进行 2D 旋转。

$$R_x(\beta) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos \beta & -\sin \beta \\ 0 & \sin \beta & \cos \beta \end{bmatrix} \quad (3.17)$$

最后，我们用偏航角 (*yaw*) 来表示绕 y 轴的逆时针旋转  $\alpha$ ，类似地，在 xz 平面进行 2D 旋转而 y 方向不变。

$$R_y(\alpha) = \begin{bmatrix} \cos \alpha & 0 & \sin \alpha \\ 0 & 1 & 0 \\ -\sin \alpha & 0 & \cos \alpha \end{bmatrix} \quad (3.18)$$

## 旋转组合

(3.16) (3.17) (3.18) 分别代表旋转中的一个自由度。偏航角、俯仰角、翻滚角经过组合可以表示所有可能的 3D 旋转：

$$R(\alpha, \beta, \gamma) = R_y(\alpha)R_x(\beta)R_z(\gamma) \quad (3.19)$$

在式 (3.19) 中， $\alpha$  和  $\gamma$  的取值范围都是  $0 \sim 2\pi$ ，然而  $\beta$  的取值范围是  $-\pi/2 \sim \pi/2$ ，尽管如此，它已经可以表示所有可能的 3D 旋转。

务必注意式 (3.19) 中的顺序不可颠倒，因为它不满足交换律。例如，先旋转  $\pi/2$  的偏航角再旋转  $\pi/2$  的俯仰角与先旋转俯仰角再旋转偏航角的结果是不一样的。你可以将  $\pi/2$  代入到 (3.17) (3.18) 中，看看经过顺序不同的矩阵乘法的结果是怎样的。2D 旋转满足交换律是因为旋转轴总是相同的，所以旋转角可以加性结合。编写 VR 软件时把旋转矩阵的次序弄错了是令人崩溃的。

## 矩阵乘法是“后向的”

当我们对矩阵进行乘法运算时我们要注意什么？考虑对点  $p = (x, y, z)$  进行旋转，我们有两个旋转矩阵  $R$  和  $Q$ 。如果我们用  $R$  进行旋转，可以得到  $p' = Rp$ ，然后继续用  $Q$  进行旋转，可以得到  $p'' = Qp'$ 。现在，我们希望将两个旋转组合起来从而直接从  $p$  得到  $p''$ 。程序员通常将它们组合成  $RQ$  因为我们是从左到右读的，也是这样理解的。然而，在线性代数中，它是后向的。因为  $Rp$  是先进行的运算，因此，它应该是从右向左读的。因此，我们必须将其组合成  $QR$ ，从而得到  $p'' = QRp$ 。在本章的后面内容中，我们会将几个旋转矩阵链式结合在一起，要记得从右向左读来理解它们做了什么。

## 用单一矩阵完成平移和旋转

如果可以通过一次操作同时完成平移和旋转，那我们使用起来将会方便很多。我们用一

个旋转矩阵  $R$  进行旋转，紧接着平移  $(x_t, y_t, z_t)$ ，代数形式如下所示：

$$\begin{bmatrix} x' \\ y' \\ z' \end{bmatrix} = R \begin{bmatrix} x \\ y \\ z \end{bmatrix} + \begin{bmatrix} x_t \\ y_t \\ z_t \end{bmatrix} \quad (3.20)$$

尽管我们没有办法通过一个  $3 \times 3$  的矩阵完成如上操作，但是我们可以增加矩阵的维度，用一个  $4 \times 4$  的齐次变换矩阵来表示。

$$T_{rb} = \begin{bmatrix} R & \begin{matrix} x_t \\ y_t \\ z_t \\ 1 \end{matrix} \\ \begin{matrix} 0 & 0 & 0 \end{matrix} & 1 \end{bmatrix}. \quad (3.21)$$

符号  $T_{rb}$  指代一个做刚体变换的矩阵，也就是说，不包含扭曲变形。一个齐次变换矩阵可能包含其他种类的变换，我们会在 3.5 节介绍。

式 (3.20) 表示的变换等价于原始点乘上 (3.21) 的变换，如下所示：

$$\begin{bmatrix} R & \begin{matrix} x_t \\ y_t \\ z_t \\ 1 \end{matrix} \\ \begin{matrix} 0 & 0 & 0 \end{matrix} & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix} = \begin{bmatrix} x' \\ y' \\ z' \\ 1 \end{bmatrix}. \quad (3.22)$$

因为存在额外的维度，我们对点  $(x, y, z)$  扩展一个维度得到  $(x, y, z, 1)$ 。要十分注意，因为平移和旋转是不能交换的，所以变换矩阵一定要满足 (3.21) 的格式。

## 反变换

我们经常会想要进行反变换，对于平移变换  $(x_t, y_t, z_t)$ ，只需简单地对其取相反数  $(-x_t, -y_t, -z_t)$  即可；对于一般的矩阵变换  $M$ ，我们取它的逆  $M^{-1}$ （如果存在）。这计算起来通常很复杂，但幸运的是，对于我们所研究的情况，它们的逆计算起来要简单很多。对于旋转矩阵  $R$ ，取逆的过程相当于取它的转置  $R^{-1}=R^T$ 。对于齐次变换矩阵 (3.21)，它的反变换写为：

$$\begin{bmatrix} R^T & \begin{matrix} -x_t \\ -y_t \\ -z_t \\ 1 \end{matrix} \\ \begin{matrix} 0 & 0 & 0 \end{matrix} & 1 \end{bmatrix}. \quad (3.23)$$

这会将平移和旋转复原，然而，这个顺序是错的。要记住，我们必须正确处理操作的顺序，因为它们是不可交换的，如图 3.8。对应一般的矩阵（非交换群理论中的一部分），对矩阵的乘积取逆时要将顺序颠倒过来：

$$(ABC)^{-1} = C^{-1}B^{-1}A^{-1} \quad (3.24)$$

我们可以将逆接在乘积的后面，即

$$ABCC^{-1}B^{-1}A^{-1} \quad (3.25)$$

首先， $C$  被自己的逆抵消，接着是  $B$ ，最后是  $A$ 。如果顺序有错的话，就不能产生抵消的效果。

矩阵  $T_{rb}$  (式 3.21) 先旋转再平移，而式 (3.23) 同样是先旋转再平移，未将顺序反过来。因此， $T_{rb}$  的逆应该是

$$\begin{bmatrix} R^T \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & -x_t \\ 0 & 1 & 0 & -y_t \\ 0 & 0 & 1 & -z_t \\ 0 & 0 & 0 & 1 \end{bmatrix}. \quad (3.26)$$

右面的矩阵先抵消平移操作，然后左面的矩阵抵消旋转。

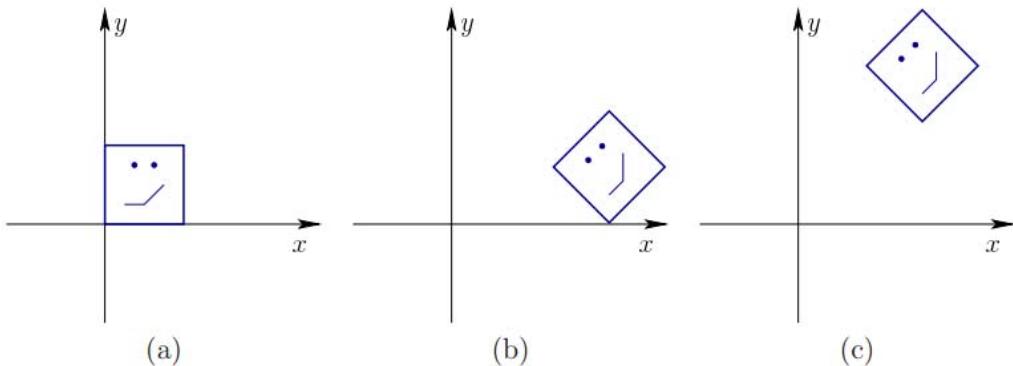


图 3.8 (a)一个  $1 \times 1$  的刚体正方形 (b)将图形平移  $x_t=2$  再旋转  $\pi/4$  ( $45^\circ$ ) (c)颠倒顺序：先旋转  $\pi/4$  再平移  $x_t=2$

### 3.3 旋转的轴-角表示

如 3.2 节所示，3D 旋转是很复杂的，原因有三：(1) 参数化的过程不简单，仅由 3 个独立的参数来代替旋转矩阵中 9 个元素 (2) 每次的旋转轴是不同的。(3) 运算是不可交换的，所以顺序非常重要。这些问题在 2D 旋转中都不存在。

#### 运动奇异点

当我们使用偏航角、俯仰角、翻滚角（以及相关的欧拉角变量）时，有一个严重的问题。即使它们直觉上容易理解，但是一些具象化特征会有所退化，使得运动奇异点很难被发现。举一个小例子，考虑我们是如何表示地球上点的位置，我们用经纬度坐标来表示这些三维的点，就像是对于偏航角和俯仰角的限制，经度的范围是  $0 \sim 2\pi$ ，而纬度的范围只有  $-\pi/2 \sim \pi/2$ （通常我们用东西经  $0 \sim 180^\circ$  来描述经度，这是等价的）。当我们从一个地方到另一个地方时，经纬度坐标看上去就像是  $xy$  坐标；然而，这是位置点离极点较远的情况。当我们靠近北极点，虽然纬度还能保持正常，但是经度会发生变化，我们移动一点点距离，经度就会发生很大的变化。想想世界地图中极点周围是怎样的，巨大的格陵兰岛以及围绕整个底部的南极洲（假定在投影中经度仍然保持直线）。这些极点就是运动奇异点：在这些特殊的点中，你可以改变经度，但是在地球中的位置并不会改变。在这些点上两个自由度中似乎有一个不起作用了。

同样的问题也出现在 3D 旋转中，但由于额外的自由度，它变得更难理解。如果俯仰角保持  $\beta = \pi/2$ ，就像是“北极点”一样，虽然  $\alpha$  和  $\gamma$  的变换之间没有关系，但是只会产生一个自由度（就像是经纬度，虽然有一个参数在变化，但是它并没有实际效果）。我们看下偏航角、俯仰角、翻滚角是怎么组合的：

$$\begin{bmatrix} \cos \alpha & 0 & \sin \alpha \\ 0 & 1 & 0 \\ -\sin \alpha & 0 & \cos \alpha \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} \cos \gamma & -\sin \gamma & 0 \\ \sin \gamma & \cos \gamma & 0 \\ 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} \cos(\alpha - \gamma) & \sin(\alpha - \gamma) & 0 \\ 0 & 0 & -1 \\ -\sin(\alpha - \gamma) & \cos(\alpha - \gamma) & 0 \end{bmatrix} \quad (3.27)$$

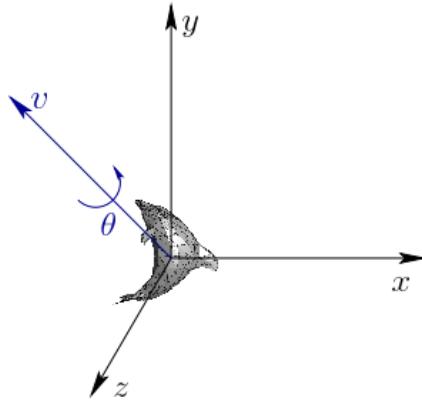


图 3.9：欧拉旋转理论指出，给定单位方向矢量  $v = (v_1, v_2, v_3)$ ，每一个 3D 旋转都可以视为一个经过原点的角度  $\theta$  的旋转

上式第二个矩阵对应着俯仰角  $\beta = \pi/2$  (3.17)。经过矩阵乘法和减法三角恒等式，我们得到右面的结果。可以发现，矩阵中的元素是一个关于  $\alpha$  和  $\gamma$  的函数，但是仅有  $\alpha$  和  $\gamma$  的差对结果有影响，所以它只有一个自由度。在视频游戏行业，大家对于这个问题的重视程度也不尽相同。在第一人称射击游戏 (FPS) 中，玩家是不会将头部的俯仰角一直维持  $\pm\pi/2$ ，正好避免了这个问题。但在虚拟现实中，用户是可以让他的头部垂直向上/下的，所以这个问题一直存在，运动奇异点经常会使得可视区域的旋转不可控。这种现象也出现在用机械系统传感和控制航天器的方向时，称为万向锁问题。

这个问题用轴-角表示方法可以轻易的解决。它比起欧拉角来说难于学习，但是因为它可以避免这些问题，所以也是值得研究的。此外，很多现有的软件库和游戏引擎都直接应用了这种表示方法。因此，要有效率的使用这些开源库，你需要理解它的内容。

解决运动奇异点问题最重要的就是欧拉旋转定理，如图 3.9 所示。即使多次旋转组合之后旋转轴可能会变，但欧拉发现任何 3D 旋转都可以被表示为一个过原点的旋转  $\theta$ 。这与旋转的 3 个自由度是匹配的：它用两个参数来指示方向，一个参数来指示  $\theta$ 。唯一的缺点就是在旋转矩阵和轴-角表示之间来回变换有些不方便。这推动了一个新的数学概念的产生，它和轴-角表示类似，同样可以模拟 3D 旋转的代数形式，甚至可以直接应用到旋转模型中。叫做四元数。

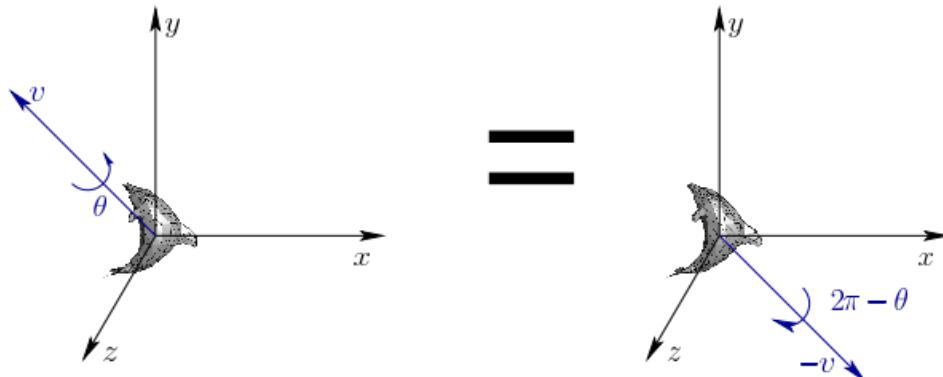


图 3.10：按照轴-角表示，同样的旋转可以用两种方式来表示，对应着方向  $v$  和  $-v$

## 2 对 1 问题

在介绍四元数之前，我们要指出欧拉旋转定理中的一个问题。如 3.10 所示，轴-角表示不是唯一的。事实上，除了恒等变换，所有 3D 旋转都存在着两种表示方法。这是因为轴的方向是分正负的。我们可以规定轴总是指向一个方向，如 $+y$ ，但是这不能完全解决问题，因为存在着边界条件（水平轴）。除了这个无法避免的问题之外，四元数可以很好的解决的 3D 旋转的所有问题。

四元数由 William Rowan Hamilton 于 1843 年提出。第一次看到的时候，很多人无法理解它独特的代数形式。因此，我们先搞清楚四元数与其准确对应的 3D 旋转之间的关系。然后，我们介绍有限四元数代数；这在 VR 系统开发中用处不大，除非你想要自己实现 3D 旋转的工具库。然而，四元数和 3D 旋转的对应关系是非常重要的。

四元数  $q$  是一个 4D 的向量：

$$q = (a, b, c, d) \quad (3.28)$$

$a, b, c, d$  均取实数。因此， $q$  可以认为是 4 维空间的一个点。经过证明，我们仅能用单位四元数来表示 3D 旋转，也就是说要保证

$$a^2 + b^2 + c^2 + d^2 = 1 \quad (3.29)$$

可以看出，这是一个在单位球面上的方程，只是维度要更高一些。通常的球面是一个 2D 平面，而单位四元数是在一个 3D 的“超平面”上，通常也叫做流形【1, 5】。我们将用单位四元数空间来表示所有的 3D 旋转。它们都具有 3 个自由度，看上去似乎是合理的。

Quaternion	Axis-Angle	Description
$(1, 0, 0, 0)$	$(\text{undefined}, 0)$	Identity rotation
$(0, 1, 0, 0)$	$((1, 0, 0), \pi)$	Pitch by $\pi$
$(0, 0, 1, 0)$	$((0, 1, 0), \pi)$	Yaw by $\pi$
$(0, 0, 0, 1)$	$((0, 0, 1), \pi)$	Roll by $\pi$
$(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}, 0, 0)$	$((1, 0, 0), \pi/2)$	Pitch by $\pi/2$
$(\frac{1}{\sqrt{2}}, 0, \frac{1}{\sqrt{2}}, 0)$	$((0, 1, 0), \pi/2)$	Yaw by $\pi/2$
$(\frac{1}{\sqrt{2}}, 0, 0, \frac{1}{\sqrt{2}})$	$((0, 0, 1), \pi/2)$	Roll by $\pi/2$

图 3.11：这些情况下，可以很清楚地看出 3D 旋转四元数表示和对应的轴-角表示

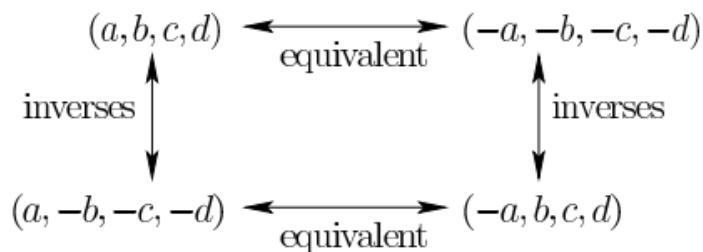


图 3.12：相等的四元数及它们的相反数示意图

令  $(v, \theta)$  为 3D 旋转的轴-角并表示，如图 3.9 所示。它可以转化为下面的四元数形式：

$$q = (\cos \frac{\theta}{2}, v_1 \sin \frac{\theta}{2}, v_2 \sin \frac{\theta}{2}, v_3 \sin \frac{\theta}{2}) \quad (3.30)$$

将  $q$  考虑成一个表示 3D 旋转的数据结构，同样的，也可以将  $q$  转换为  $(v, \theta)$

$$\theta = 2\cos^{-1}a \text{ and } v = \frac{1}{\sqrt{1-a^2}}(b, c, d) \quad (3.31)$$

如果  $a=1$ , 那么式 (3.31) 不成立, 这也正对应了恒等旋转。

你现在可以在  $(v, \theta)$  和  $q$  之间相互转换。图 3.11 展示了几个简单的, 经常会出现的例子。此外, 图 3.12 展示了四元数及其对应旋转之间的关系。水平箭头的  $q$  和  $-q$  表示相同的旋转, 对应着图 3.10 中的两种表示方式; 从式(3.30)中也可以看出它们之间的相等关系。竖直箭头对应着反向旋转, 是因为轴的方向反向了 (旋转  $\theta$  变成了旋转  $2\pi-\theta$ )。

我们怎么将四元数  $h=(a,b,c,d)$  应用到旋转模式? 一种方式是如下面所示, 将四元数变换为旋转矩阵:

$$R(h) = \begin{bmatrix} 2(a^2 + b^2) - 1 & 2(bc - ad) & 2(bd + ac) \\ 2(bc + ad) & 2(a^2 + c^2) - 1 & 2(cd - ab) \\ 2(bd - ac) & 2(cd + ab) & 2(a^2 + d^2) - 1 \end{bmatrix} \quad (3.32)$$

也可以不用这么麻烦, 我们定义四元数乘法, 对于任意两个四元数  $q_1$  和  $q_2$ , 乘积用  $q_1 * q_2$  表示, 定义如下:

$$\begin{aligned} a_3 &= a_1 a_2 - b_1 b_2 - c_1 c_2 - d_1 d_2 \\ b_3 &= a_1 b_2 + a_2 b_1 + c_1 d_2 - c_2 d_1 \\ c_3 &= a_1 c_2 + a_2 c_1 + b_1 d_2 - b_2 d_1 \\ d_3 &= a_1 d_2 + a_2 d_1 + b_1 c_2 - b_2 c_1 \end{aligned} \quad (3.33)$$

换句话说,  $q_3 = q_1 * q_2$  由式 (3.33) 定义。

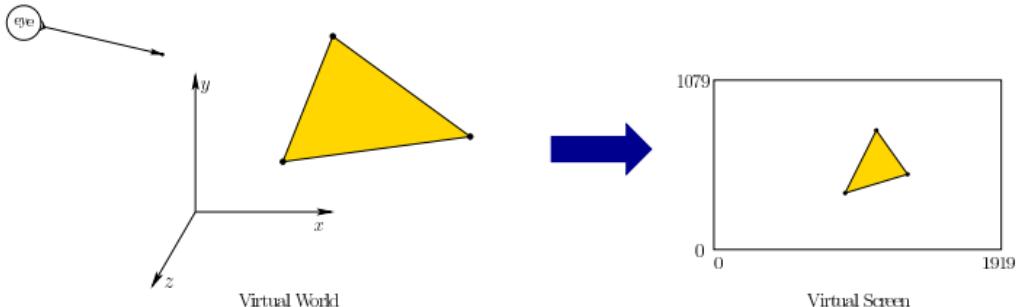


图 3.13: 如果我们将虚拟眼睛或相机放入虚拟世界, 它会看到什么? 3.4 节提供了基于虚拟眼睛的特定视点, 能将物体从虚拟世界放置到虚拟屏幕上。由于工程和历史原因, 虚拟屏幕选择矩形的形状, 尽管它与视网膜的形状不匹配。

我们使用关于  $h$  的旋转表达式来对  $(x, y, z)$  点进行旋转。令  $p=(x, y, z, 1)$ , 这是为了给点  $p$  与四元数相同的维数。通过应用四元数乘法来旋转这个点, 如下式所示:

$$p' = q * p * q^{-1}, \quad (3.34)$$

其中  $q^{-1} = (a, -b, -c, -d)$  (从图 3.12 中可得到)。旋转后的点是  $(x', y', z')$ , 它取自结果  $p' = (x', y', z', 1)$ 。

我们举一个简单的例子: 点  $(1, 0, 0)$ 。假设  $p = (1, 0, 0, 1)$ , 并考虑执行偏航旋转  $\pi/2$ 。根据图 3.11, 相应的四元数是  $q = (0, 0, 1, 0)$ 。 $q^{-1} = (0, 0, -1, 0)$ 。应用式 (3.33) 计算式 (3.34) 后, 结果为  $p' = (0, 1, 0, 1)$ 。因此, 在偏航角为  $\pi/2$  时, 旋转得到的点是  $(0, 1, 0)$ 。

### 3.4 视角变换

本节介绍如何转换虚拟世界中的模型, 以便它们出现在虚拟屏幕上。主要目的是为图形

渲染奠定基础，从而能增加光照、材料特性和量化效果，最终可以实现在物理显示屏上显示。这些转换的另一方面也解释了相机如何形成图像，至少是一个理想的成像过程。将这些转换部分看作描述在虚拟世界中的虚拟相机，那么用该相机拍摄的虚拟照片应该是什么样子？为了使 VR 正常工作，“相机”实际上应该是置于虚拟世界中的两个虚拟眼中一个。因此，根据虚拟世界中的位置和方向，虚拟眼睛应该看到什么？这里并不是指在视网膜上会出现什么，这部分将会在 4.4 节之后讲解，这里我们只是计算模型顶点在虚拟世界中的平面矩形屏幕上的出现位置。见图 3.13。

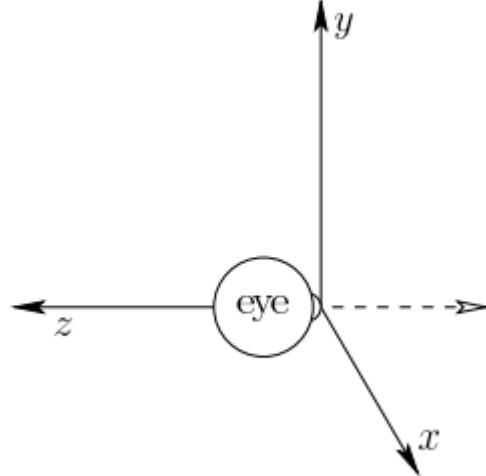


图 3.14：考虑一个在负方向上向下看 z 轴的眼睛。模型的起源是光线进入眼睛的点。

### 人眼观看

图 3.14 显示了一个俯视负 Z 轴的虚拟眼睛。它的放置位置从眼睛的角度来看，x 向右增加，y 向上，这对应于一般的笛卡尔坐标。替代方案可以是：1) 在正 z 方向上朝向眼睛，这使得 xy 坐标平面向后推移，或者 2) 颠倒 z 轴，这将导致左手坐标系。因此，我们做出选择可以来避免了更糟的情形。

假设眼睛是我们想要放置在三维虚拟世界中的物体模型，我们可以在某个位置  $e = (e_1, e_2, e_3)$  和矩阵给出的方向，如下式所示：

$$R_{eye} = \begin{bmatrix} \hat{x}_1 & \hat{y}_1 & \hat{z}_1 \\ \hat{x}_2 & \hat{y}_2 & \hat{z}_2 \\ \hat{x}_3 & \hat{y}_3 & \hat{z}_3 \end{bmatrix} \quad (3.35)$$

如果图 3.14 中的眼球是由三角形组成的，那么  $R_{eye}$  的旋转和  $e$  的平移将应用于所有的三维空间的顶点中。

然而，这并不能解决虚拟世界如何出现在眼前的问题。我们需要将虚拟世界中的所有模型移动到眼睛的参照系，而不是在虚拟世界中移动眼睛。这意味着我们需要应用逆变换。旋转的逆是  $R_{eye}^T$ ， $R_{eye}$  的转置。 $e$  的倒数是  $-e$ 。应用 (3.26) 做适当的变换：

$$T_{eye} = \begin{bmatrix} \hat{x}_1 & \hat{x}_2 & \hat{x}_3 & 0 \\ \hat{y}_1 & \hat{y}_2 & \hat{y}_3 & 0 \\ \hat{z}_1 & \hat{z}_2 & \hat{z}_3 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & -e_1 \\ 0 & 1 & 0 & -e_2 \\ 0 & 0 & 1 & -e_3 \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad (3.36)$$

请注意，如 (3.35) 中所示的  $R_{eye}$  已被转置并放置在上面的左侧矩阵中。另外，转换顺序和旋转顺序已经交换过，如 3.2 节所述，逆顺序是必需的。

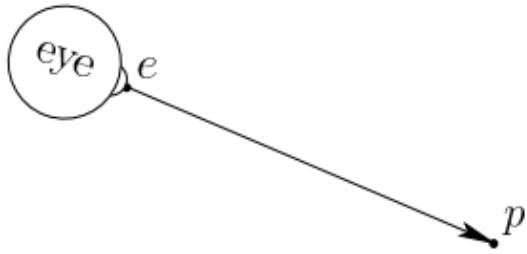


图 3.15：从眼睛位置  $e$  到它正在查看的点  $p$  的矢量被归一化为 (3.37) 中的  $c$ 。

在图 3.4 后，式 (3.36) 有两种可能的解释。如前所述，这可能对应于移动所有虚拟世界模型（对应于图 3.4 (b)）。但在当前的设置中，更合适的解释是虚拟世界的坐标系正在移动，以便它与图 3.14 中的眼睛帧相匹配。这对应于图 3.4 (c) 的情况，不是 3.2 节中的适当解释。

从 VR 的外观来看，虚拟世界中眼睛的位置和方向是由跟踪系统和可能的控制器作为输入。相比之下，在计算机图形学中，通常首先描述眼睛的位置以及眼睛所看的方向。这被称为“look-at”，具有以下部分：

1. 眼睛的位置：  $e$
2. 眼睛的中心方向：  $\hat{c}$
3. 向上的方向：  $\hat{u}$ 。

$\hat{c}$  和  $\hat{u}$  都是单位向量。第一个方向  $\hat{c}$  对应于视角的中心。无论  $\hat{c}$  指向什么都应该最终在显示屏的中心。如果我们希望这是三维空间中的特定点  $p$ （见图 3.15），则  $c$  可以计算为

$$\hat{c} = \frac{p - e}{\|p - e\|}, \quad (3.37)$$

其中  $\| \cdot \|$  表示矢量的长度。结果是从  $e$  到  $p$  的矢量并做归一化。

第二个方向  $\hat{u}$  指示哪个方向为向上的方向。想象一下拿着相机，就好像你要拍一张照片然后进行旋转一样。你可以使图像中的水平地面看起来倾斜或甚至颠倒。因此， $\hat{u}$  表示虚拟相机或眼睛的向上方向。

我们现在从式 (3.36) 构造  $T_{\text{eye}}$  变换。转换部分已经由  $e$  来确定，这是在“look-at”中给出的。我们只需要确定旋转  $R_{\text{eye}}$ ，如 (3.35) 所示。回想一下在 3.2 节中矩阵列指出了矩阵如何转换坐标轴（参考 (3.7) 和 (3.8)）。这简化了确定  $R_{\text{eye}}$  的问题。每个列向量计算为

$$\begin{aligned}\hat{z} &= -\hat{c} \\ \hat{x} &= \hat{u} \times \hat{z} \\ \hat{y} &= \hat{z} \times \hat{x}.\end{aligned}\quad (3.38)$$

计算  $\hat{z}$  会出现负号，因为眼睛正在向下看  $z$  轴。 $\hat{x}$  方向使用标准叉积  $\hat{z}$  来计算。对于第三个方程，我们可以使用  $\hat{y} = \hat{u}$ ；然而， $\hat{z} \times \hat{x}$  会巧妙地纠正  $\hat{u}$  通常指向上方但不垂直于  $\hat{c}$  的情况。式 (3.38) 中的单位矢量代入 (3.35) 可得到  $R_{\text{eye}}$ 。因此，我们拥有构建  $T_{\text{eye}}$  所需的全部信息。

### 正交投影

当  $T_{\text{eye}}$  已知，设  $(x, y, z)$  表示任意点的坐标。如果我们强制把每个  $z$  坐标取值为 0，

然后将所有点直接投影到垂直  $xy$  平面中，将会发生什么？换句话说，也就是  $(x, y, z) \rightarrow (x, y, 0)$ ，这就是所谓的正交投影。如果我们将  $xy$  平面想象为模型的虚拟展示，那么会出现以下几个问题：

1. 混杂的物体将被叠加，而不是物体之间有相互遮挡的关系。
2. 显示画面将在所有方向无限延伸（ $z$  除外）。如果显示器是  $xy$  平面上的一个小矩形，则超出范围的模型部分可以被消除掉。
3. 距离较近的物体应大于距离较远的物体，这存在于现实世界中。我们可以回想一下第 1.3 节（图 1.22 (c)）正确处理视角的绘画。

前两个问题是重要的图形操作，这部分的介绍将在第 7 章讲述。我们要解决第三个问题。

### 透视投影

我们定义一个透视投影，而不用上述提到的正交投影。对于每个点  $(x, y, z)$  它位于通过原点的一条线上。下式表示所有点的集合：

$$(\lambda x, \lambda y, \lambda z), \quad (3.39)$$

其中  $\lambda$  可以是任何实数。换句话说， $\lambda$  是一个能够到达线上所有点（包含  $(x, y, z)$  和  $(0, 0, 0)$ ）的参数，见图 3.16。

现在我们可以在虚拟世界的任何地方放置一个平面的“电影屏幕”，并查看所有线条穿透它的位置。为了简化这个数学表达式，我们选择  $z = -1$  平面将我们的虚拟屏幕直接放在眼睛的前方，见图 3.17。使用 (3.39) 的第三个分量，我们有  $\lambda z = -1$ ，这意味着  $\lambda = -1/z$ 。使用 (3.39) 的前两个分量，屏幕上点的坐标计算为  $x' = -x/z$ ,  $y' = -y/z$ 。请注意，由于  $x$  和  $y$  对于每个轴的缩放量  $z$  相同，因此它们的宽高比将保留在屏幕上。

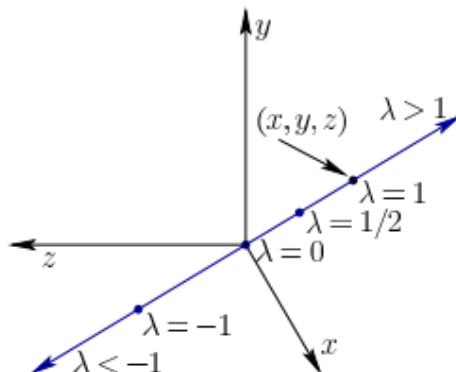


图 3.16：从任意点  $(x, y, z)$  开始，通过原点的线可以使用参数  $\lambda$  形成。它是任何实数  $\lambda$  的形式  $(\lambda x, \lambda y, \lambda z)$  的所有点的集合。例如， $\lambda = 1/2$  对应于  $(x, y, z)$  和  $(0, 0, 0)$  沿线的中点。

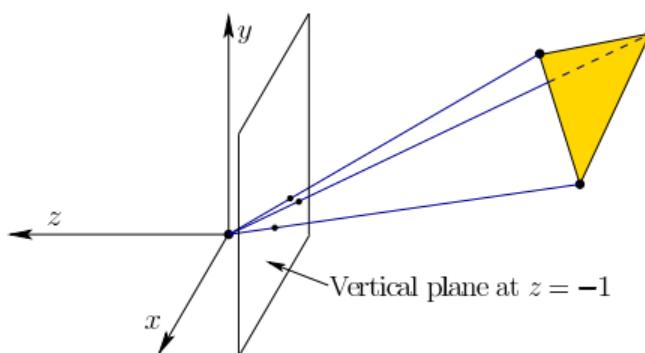


图 3.17：透视投影的例证。模型顶点通过绘制直线和原点  $(0,0,0)$  投影到虚拟屏幕上。虚拟屏幕上的点的“图像”对应于线条与屏幕的交点。

更一般地说，假设垂直屏幕放置在沿着  $z$  轴的某个位置  $d$  处。在这种情况下，我们可以获得更多的通用表达式来显示屏幕上某个点的位置：

$$\begin{aligned}x' &= dx/z \\y' &= dy/z.\end{aligned}\tag{3.40}$$

这是通过对  $\lambda$  求解  $d = \lambda z$  并将其代入 (3.39) 得到的。

这就是我们将点投影到虚拟屏幕上，同时考虑到各种距离下物体的缩放属性时所需要做的所有事情。在虚拟现实中正确使用这一点有助于理解深度和规模，这在第 6.1 节中介绍。在 3.5 节中，我们将使用变换矩阵来调整 (3.40)。此外，只有位于眼前区域的点才会投射到虚拟屏幕上。太靠近，太远或超出正常视野范围的点将不会在虚拟屏幕上呈现；这在第 3.5 节和第 7 章中讨论。

### 3.5 转换级联

本节将本章的所有转换级联在一起，同时稍微调整其形式以适应 VR 和计算机图形工业目前使用的形式。本节中出现的一些矩阵可能看起来有些复杂。原因是因为表达式是由算法和硬件问题二者同时驱动的，而不是仅是数学表达式那么简单。特别是，即使在不是线性的透视投影（式 (3.40)）的情况下，也存在偏向于将每个变换放入  $4 \times 4$  齐次变换矩阵。以这种方式，可以在矩阵链上迭代有效的矩阵乘法算法以产生结果。

该级联的一般表达式如下：

$$T = T_{vp}T_{can}T_{eye}T_{rb}.\tag{3.41}$$

当  $T$  应用于点  $(x, y, z, 1)$  时，会生成屏幕上点的位置。请记住，这些矩阵乘法不可交换，并且操作顺序是从右向左。第一个矩阵  $T_{rb}$  是应用于可移动模型上的点的刚体变换 (3.21)。对于模型中的每个刚体， $T_{rb}$  保持不变，然而，不同的物体通常会放置在不同的位置和方向。例如，“虚拟汽车的轮子移动”将会不同于“阿凡达的头部”。根据 (3.36)，在应用  $T_{rb}$  之后， $T_{eye}$  将虚拟世界转换为眼睛的坐标系。在一个固定的瞬间，虚拟世界中所有点的  $T_{eye}$  和所有剩余的变换矩阵是相同的。在这里，我们假定眼睛位于两个虚拟眼之间的中点，从而形成一个环视点。在本节的后面，我们将把它扩展到左右眼的情况，以便构建立体视点。

#### 规范化视图变换

下一个变换， $T_{can}$  执行 3.4 节中所述的透视投影；然而，我们必须解释它是如何“强迫”进入  $4 \times 4$  矩阵的。再次受到工业需求的驱动，我们希望变换的结果是一种看起来没有单位的规范形式，因此， $T_{can}$  被称为规范化视图变换。图 3.18 显示了一个基于矩形虚拟屏幕的四个角的视锥。 $z = n$  和  $z = f$  分别为近平面和远平面。请注意，这些情况下  $z < 0$ ，因为  $z$  轴指向相反的方向。虚拟屏幕包含在近平面中。透视投影应该将截锥内的所有点放置在以近平面为中心的虚拟屏幕上。运用到式 (3.40) 我们发现  $d = n$ 。

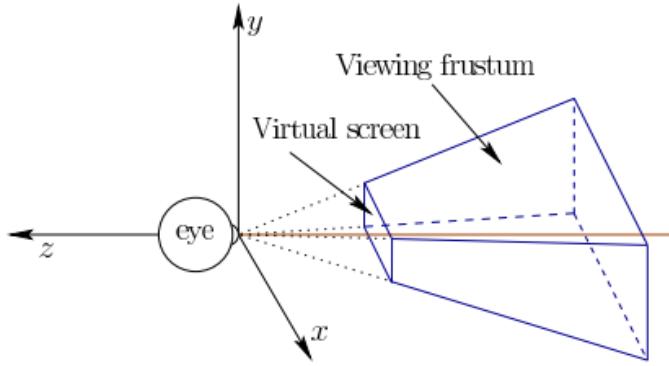


图 3.18 截掉的椎体视角

我们现在想用矩阵来重新推导出式 (3.40)。应用以下矩阵乘法的结果为：

$$\begin{bmatrix} n & 0 & 0 & 0 \\ 0 & n & 0 & 0 \\ 0 & 0 & n & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix} = \begin{bmatrix} nx \\ ny \\ nz \\ z \end{bmatrix} \quad (3.42)$$

在前两个坐标中，我们得到 (3.40) 的分子。(3.40) 的非线性部分是  $1/z$  因子。为了处理这个问题，第四个坐标用来表示  $z$ ，而不是 1，以此来处理  $T_{rb}$ 。从这一点开始，所得到的 4D 矢量被解释为一个 3D 矢量，该矢量通过划分其第四个分量来缩放。例如，( $v_1, v_2, v_3, v_4$ ) 被解释为

$$(v_1/v_4, v_2/v_4, v_3/v_4) \quad (3.43)$$

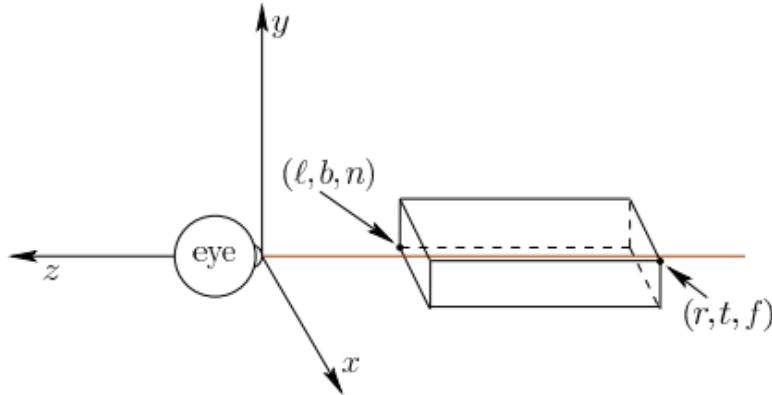


图 3.19：由  $T_p$  转换后的视锥角形成的矩形区域。所选对角的坐标在  $T_{st}$  中使用的六个参数， $l, r, b, t, n$  和  $f$ 。

因此，(3.42) 的结果被解释为：

$$(nx/z, ny/z, n), \quad (3.44)$$

其中前两个坐标与  $d=n$  匹配 (3.42)，第三个坐标是沿着  $z$  轴的虚拟屏幕的位置。

## 跟踪深度

以下矩阵常用于计算机图形学，并将在我们的级联中用于此处：

$$T_p = \begin{bmatrix} n & 0 & 0 & 0 \\ 0 & n & 0 & 0 \\ 0 & 0 & n+f & -fn \\ 0 & 0 & 1 & 0 \end{bmatrix} \quad (3.45)$$

除了转换 z 坐标的方式外，它与 (3.42) 中的矩阵相同。在虚拟屏幕上放置点，这种做法是不必要的，因为我们已经知道它们全部放置在  $z=n$  处。因此，z 坐标被用于另一个目的：跟踪每个点与眼睛的距离，以便图形算法可以确定哪些对象位于其他对象的前面。矩阵  $T_p$  计算第三个坐标为：

$$(n+f)z - fn \quad (3.46)$$

当将式 (3.46) 除以 z，我们不保留精确的距离，但图形方法（其中一些在第 7 章中介绍）仅需要保留距离排序。换句话说，如果点 p 比点 q 离眼睛更远，那么即使距离失真，点 p 仍然会比 q 远。但是，它确实保留了两种特殊情况下的距离： $z=n$  和  $z=f$ 。这可以通过将这些代入 (3.46) 并除以 z 来看出。

## 额外的转换和缩放

应用  $T_p$  后，平截头体的 8 个角被转换成矩形框的角，如图 3.19 所示。以下内容对 z 轴进行了一个简单的转换，并进行了一些重新调整，使其在原点处居中，其角点坐标为  $(\pm 1, \pm 1, \pm 1)$ ：

$$T_{st} = \begin{bmatrix} \frac{2}{r-\ell} & 0 & 0 & -\frac{r+\ell}{r-\ell} \\ 0 & \frac{2}{t-b} & 0 & -\frac{t+b}{t-b} \\ 0 & 0 & \frac{2}{n-f} & -\frac{n+f}{n-f} \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad (3.47)$$

如果平截头体在 xy 平面上完全居中，则最后一列的前两个分量变为 0。最后，我们将 (3.41) 中的规范视图变换  $T_{can}$  定义为：

$$T_{can} = T_{st}T_p \quad (3.48)$$

## 视口变换

要在式 (3.41) 中应用的最后一个转换是视口转换  $T_{vp}$ 。在应用了  $T_{can}$  之后，x 和 y 坐标的每个范围都从 -1 到 1。需要最后一个步骤才能将投影点带到用于对物理显示中的像素进行索引的坐标。令 m 为水平像素的数量，n 为垂直像素的数量。例如，对于 1080p 显示器， $n=1080$ ， $m=1920$ 。假设显示索引的行数从 0 到  $n-1$ ，列数从 0 到  $m-1$ 。此外， $(0, 0)$  位于左下角。在这种情况下，视口变换是：

$$T_{vp} = \begin{bmatrix} \frac{m}{2} & 0 & 0 & \frac{m-1}{2} \\ 0 & \frac{n}{2} & 0 & \frac{n-1}{2} \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad (3.49)$$

## 左右眼

我们现在讨论如何改变转换式 (3.41) 进行立体观看。设  $t$  表示左右眼之间的距离。其在现实世界中的值因人而异，其平均值在  $t=0.064$  米左右。为了处理左眼视图，我们需要简单地将环眼（中心）眼睛水平向左移动。回想一下第 3.4 节，实际上反过来也是适用的。模型需要向右移动。所以，

$$T_{left} = \begin{bmatrix} 1 & 0 & 0 & \frac{t}{2} \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad (3.50)$$

当从眼睛的角度看时，这对应于模型的右移。这个转换放在  $T_{eye}$  之后再来调整它的输出。对 (3.41) 做适当的修改：

$$T = T_{vp} T_{can} T_{left} T_{eye} T_{rb}. \quad (3.51)$$

通过对称性，通过用 (3.51) 替换  $T_{left}$  类似地处理右眼：

$$T_{right} = \begin{bmatrix} 1 & 0 & 0 & -\frac{t}{2} \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad (3.52)$$

这就解释了在虚拟世界中放置和移动模型的整个转换公式，然后让它们出现在显示器的正确位置。在阅读第 4 章之后，很明显，在整个公式应用之后可能需要进行最后的转换。这是为了补偿由于VR头盔中广角镜头引入的非线性光学畸变。

## 拓展阅读

大多数矩阵变换出现在标准的计算机图形文本中。本章中的介绍严格遵循[202]。有关四元数及其相关代数性质的更多详细信息，请参见[157]。机器人文本通常涵盖了刚体和链体的三维变换，并且还考虑了单元的奇异性；见[163, 303]。

## 第四章 光与光学

了解光在物理世界中的传播方式对理解虚拟现实至关重要。一个原因是这关系着视觉的显示和光线如何在我们的人眼界面成像。光线从显示器发出，以重现光线的方式到达视网膜。在目前这一代 VR 头戴式设备中，工程镜头和人眼系统是成像的研究对象。研究光传播的另一个原因是虚拟世界的构建。第 3 章涵盖了理论的几何建模方面。接下来就是在虚拟世界里模拟物理世界的光传播，这将在第 7 章继续讲解，第七章介绍了什么内容应该在视觉显示器上呈现。最后，光线传播也有助于理解相机如何工作，这为提供虚拟世界提供了另一种方式：通过全景视频。相机对于追踪也很重要，这将在第 9.3 节讨论。

4.1 节介绍了光的基本物理特性，包括它与材料及其光谱特性的相互作用。第 4.2 节提供了镜头如何工作的理想模型。然后，4.3 节将展示镜头行为偏离理想模型的许多方式，这从而降低 VR 体验。在第 5 章中详细介绍了眼睛和人类视觉之前，4.4 节介绍了人眼作为镜头的光学系统，第 4.5 节介绍了可视为工程眼睛的相机。

### 4.1 光的基本行为

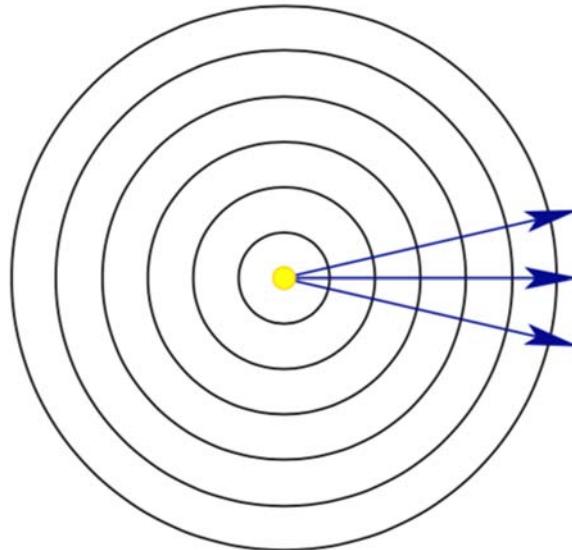


图 4.1：从点光源发出的波和可见光线。

光可以用三种看起来相互矛盾的方式来描述：

1. 光子：在空间中高速移动的微小能量粒子（本书不需要量子力学！）。当考虑传感器或接收器接收的光子数量时，这种解释是有帮助的。
2. 光波：通过空间的波纹，类似于在水面上传播的波浪，但是是三维的。波长是峰值之间的距离。在研究颜色的光谱时，这种解释是有帮助的。
3. 光线：光线追踪单个假想光子的运动。方向垂直于波前（见图 4.1）。这个解释在解释镜头和定义可见性概念时很有用。

幸运的是，现代物理学已经证明了这些解释实际上是如何相容的；在本书中都很有用。

## 扩散波

图 4.1 显示了波如何从一个假想点光源传播。所有方向的密度都是相同的（径向对称），但随着距离光源越来越远，密度会越来越小。重新命名半径为  $r$  的球体的表面积为  $4\pi r^2$ 。考虑围绕光源对准球形屏幕，每秒击中半径为 1 的屏幕的光子总数应该与半径为 2 的屏幕相同；然而，密度（每秒每个区域的光子数）应该减少  $1/4$ ，因为它们分布在面积的 4 倍以上。因此，光子密度随距离点光源的距离而二次减小。

随着点光源变得越来越远，波前的曲率也慢慢减小。如果光波无限地传播，那么它们将完全变平，如图 4.2 所示。这导致平行波前的重要情况。没有镜头或镜子的帮助，实际上不可能从物理世界中的微小光源获得这种情况，因为它不可能如此遥远；然而，它既是远距离光源的有用近似，也是数学描述透镜的理想方法。请记住，在与点光源有限的距离内，光线始终发散；没有镜头或镜子的帮助，就不可能使它们汇合。

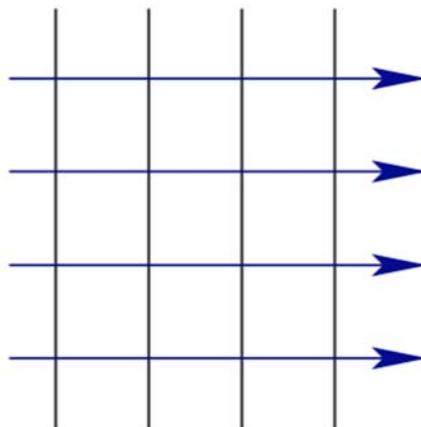


图 4.2：如果点光源离“无限远”，那么将获得平行波前。此设置的其他名称包括：平行光，平行光线，无限远光线，无穷远光线和零聚散度。

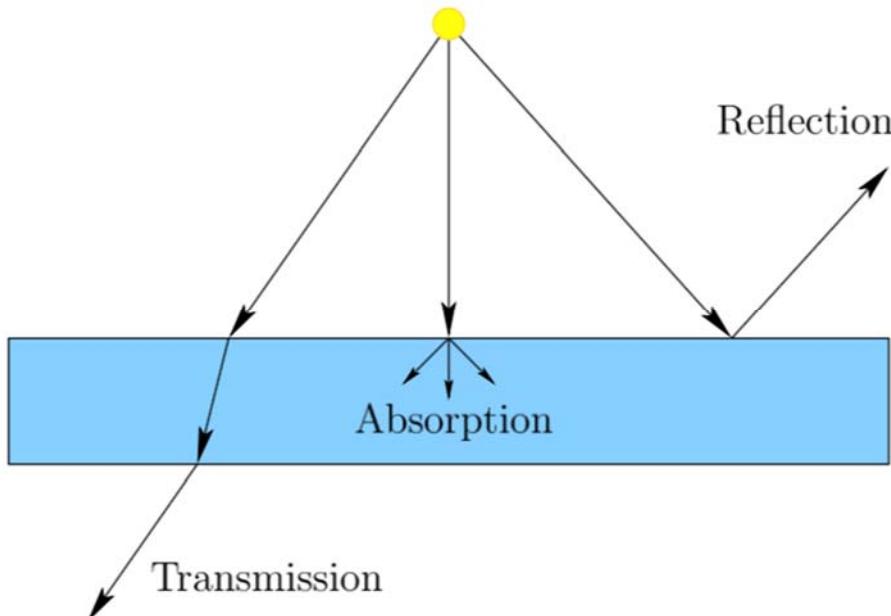


图 4.3：当光能碰到不同介质的边界时，有三种可能性：透射，吸收和反射。

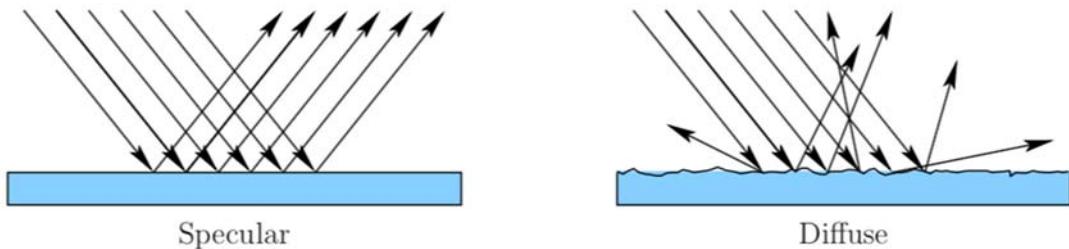


图 4.4: 显示了两种极端的反射模式。镜面反射意味着所有光线都以相同的角度反射。漫反射意味着光线散射的方式与其接近角度无关。镜面反射对于抛光表面(例如镜子)是常见的,而漫反射对应于粗糙表面。

### 与材料的相互作用

当光线撞击材料表面时,可能会发生三种行为之一,如图 4.3 所示。在透射的情况下,能量穿过材料并从另一侧离开。对于像玻璃这样的透明材料,透射光线会按照斯涅尔定律减速并弯曲,这将在 4.2 节中介绍。对于不透明的半透明材料,射线在离开之前会散射到各个方向。在吸收的情况下,当光被捕获时,能量被材料吸收。第三种情况是反射,其中光线从表面偏转。沿着完美光滑或抛光的表面,光线以相同的方式反射:出射角等于入射角。如图 4.4。这种情况称为镜面反射,与漫反射相反。反射光线在任意方向上散射。通常,所有三种传播,吸收和反射的情况同时发生。在这些情况之间划分的能量数量取决于许多因素,例如接近角度,波长以及两种相邻材料或介质之间的差异。

### 一系列波长

图 4.1 给出了一个过于简单化的视图,可以很容易地理解 4.2 节中的理想镜头。不幸的是,它忽略了许多在其他设置中变得重要的细节,例如了解镜头像差(第 4.3 节)或光与物理世界中的材料如何相互作用。因此本节的其余部分考虑了出现的各种现实出现的问题。

### 相干与混杂的光线

第一个复杂因素是光源通常不会发出相干光,这个术语意味着波前完全在时间和空间上对齐。激光是一种确实会产生相干光的特殊情况。它发出一个恒定波长的平行波,这些波也在时间上同步,以便它们的峰在它们传播时对齐。常见的光源,如灯泡和太阳,反而会发出一系列具有不同波长的波,并且没有对齐它们的峰。

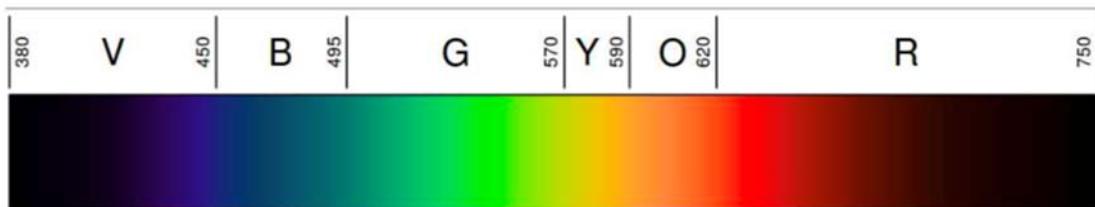


图 4.5: 可见光谱对应于波长在 400nm 和 700nm 之间的电磁波的范围。(图自维基百科的大卫埃克尔斯)

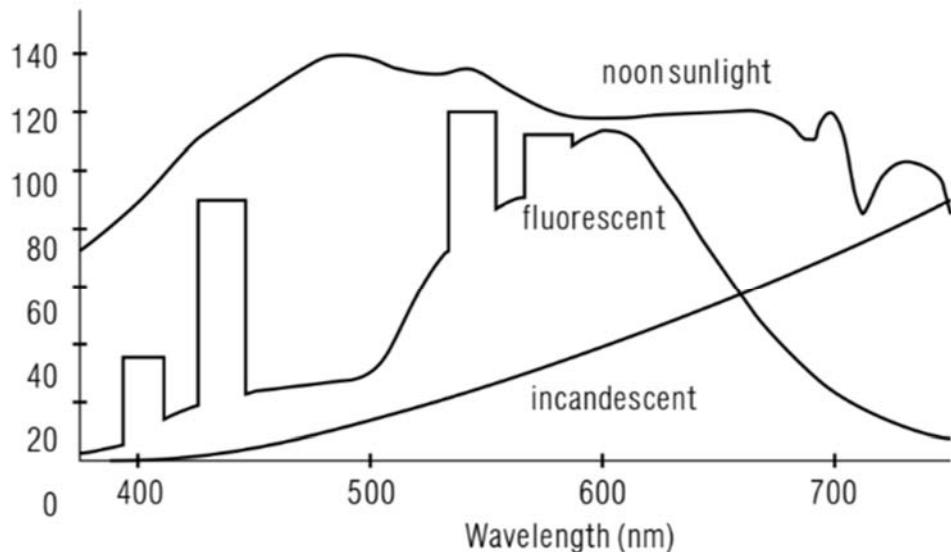


图 4.6: 一些常见光源的光谱功率分布。 (图[292])。

### 波长和颜色

为了从混合的波中理解，我们将描述它们在波长方面如何分布。图 4.5 显示了人类可见的波长范围。每个波长对应一个光谱颜色，这就是我们用一个固定在那个波长上的相干光源会感知到的。700 到 1000nm 之间的波长称为红外线，这对我们来说是不可见的，但有些相机可以感知它们（参见第 9.3 节）。100 到 400nm 之间的波长称为紫外线；它们也不是我们可见光谱的一部分，但是一些鸟类，昆虫和鱼类可以感知超过 300nm 的紫外线波长。因此，我们关于可见光的概念已经与人类的看法联系在一起。

### 光谱功率

图 4.6 显示了如何为常见光源分配波长。一个理想的光源将具有所有可见波长表示的相同的能力，从而产生理想化的白光。相反的是完全黑暗，这是黑色的。我们通常不允许光源将光线直接传播到视网膜上（不要盯着太阳！）。相反，我们观察周围物体反射的光线，使我们感觉到它们的颜色。每个表面都有自己的反射波长分布。反射回来的光能部分取决于波长，导致如图 4.7 所示的曲线。对于我们把物体表面视为红色，红色波长必须包含在光源中，表面必须强烈反射红色波长。其他波长也必须被抑制。例如，光源可能是白色（包含所有波长），并且物体可能强烈反射所有波长，导致表面呈现白色而不是红色。6.3 节将提供色彩感知的更多细节。

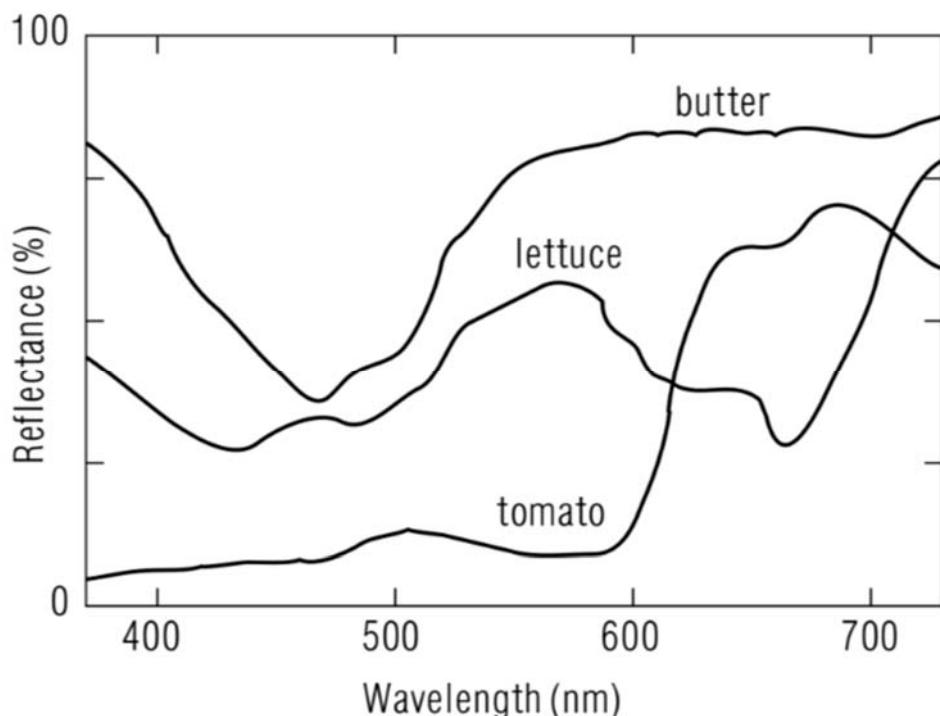


图 4.7：一些常用材料的光谱反射功能。 (图[292])。

### 频率

通常情况，谈论频率而不是波长是有用的。频率是波峰每秒通过固定位置的次数。使用波长  $\lambda$  和速度  $s$ ，频率  $f$  计算如下：

$$f = \frac{s}{\lambda}. \quad (4.1)$$

光在真空中的速度通用常数  $c$  表示，其值约等于  $3 \times 10^8 \text{ m/s}$ 。在这种情况下，(4.1) 中的  $s = c$ 。光在真空中的传播速度比在空气中大约快 0.03%，这使得在大多数工程计算中差异被忽略。空气中可见光的频率范围大约为 400 至 800 太赫兹，这通过应用 (4.1) 获得。随着光线通过密度较大的介质（如水或透镜）传播， $s$  明显变小；这种差异是光学系统的基础，下面将会介绍。

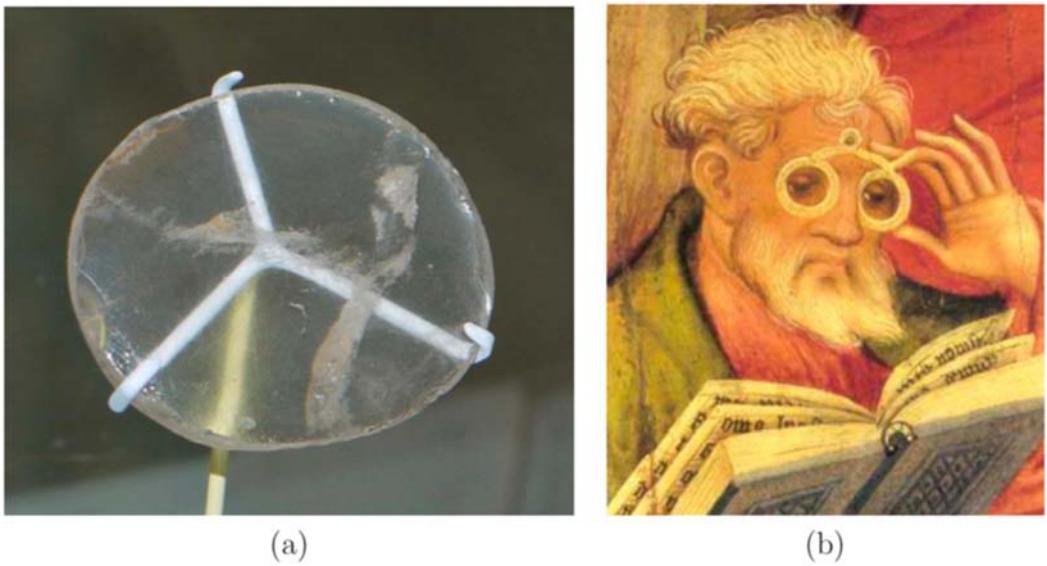


图 4.8: (a) 最古老的已知人造透镜，在古代亚述尼姆鲁德的公元前 750 年至前 710 年制作而成。目前还不知道这个神器是纯粹的装饰物还是用来制作聚焦的图像的。图片来自大英博物馆。(b)1403 年由 Conrad von Soest 绘制的一幅画，显示老年男性使用老花镜。

## 4.2 镜头

镜头已经制造了数千年，其中最古老的已知工件如图 4.8 (a) 所示。它在公元前 700 年建于亚述尼姆鲁德。无论是透明材料还是用作反射镜的抛光表面，透镜都会弯曲光线，从而形成聚焦图像。几个世纪以来，它们的使用造就了几种著名的设备，如眼镜（图 4.8 (b)），望远镜，放大镜，望远镜，照相机和显微镜。因此，光学工程充满了设计模式，以说明如何优化这些易于理解的设备的设计。VR 头盔与传统光学设备不同，导致许多超出了几个世纪以来存在的标准模式的新的挑战。因此，VR 的镜头设计模式仍在编写中。解决当前挑战的第一步是了解简单镜头的工作原理。

### 斯内尔定律

镜片的工作原因是斯涅耳定律，它表达了当进入或退出透明材料时光线有多少弯曲。回想一下，介质中的光速小于真空中的速度  $c$ 。对于给定的材料，将其折射率定义为：

$$n = c / s \quad (4.2)$$

其中  $s$  是介质中的光速。例如， $n = 2$  意味着光线穿过介质的时间是真空的两倍。对于一些常见的例子，空气  $n = 1.000293$ ，水  $n = 1.33$ ，冠玻璃  $n = 1.523$ 。

图 4.9 显示了入射光波和射线的情况。假设在这个例子中，光从空气传播到玻璃，所以  $n_1 < n_2$ 。令  $\theta_1$  表示相对于表面法线的入射角，并让  $\theta_2$  表示通过材料时得到的角度。斯内尔定律将四个量相关联：

$$n_1 \sin \theta_1 = n_2 \sin \theta_2.$$

典型地，给出  $n_1 / n_2$  和  $\theta_1$ ，使得 (4.3) 求解  $\theta_2$  来获得：

$$\theta_2 = \sin^{-1} \left( \frac{n_1 \sin \theta_1}{n_2} \right).$$

如果  $n_1 < n_2$ , 则  $\theta_2$  比  $\theta_1$  更接近于垂直。如果  $n_1 > n_2$ , 则  $\theta_2$  离垂直更远。 $n_1 > n_2$  的情况也是有趣的, 因为如果入射角  $\theta_1$  太大, 光可能不能穿透表面。 $\sin^{-1}$  的范围是 0 到 1, 这意味着 (4.4) 仅在以下公式才提供  $\theta_2$  的解:

$$(n_1/n_2) \sin \theta_1 \leq 1.$$

如果上述条件不成立, 那么光线就会从表面反射出来。这种情况发生在水下并仰望表面时。取决于观察角度, 游泳者可能会看到反射, 而不是能够看到上面的世界。

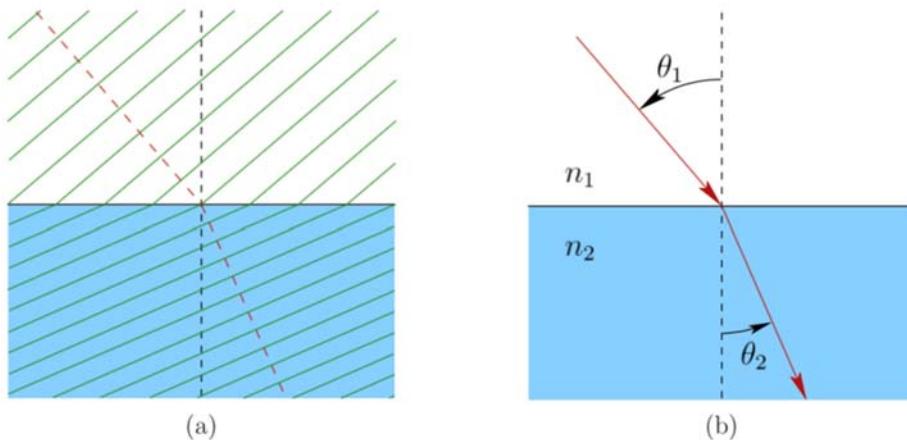


图4.9: 从折射率较低的介质 (如空气) 传播到具有较高折射率的介质 (如玻璃)。(a) 当波前进入较低介质时, 显示较慢传播对波前的影响。(b) 这显示了总是垂直于波前的光线弯曲的结果。斯涅耳定律将折射率和角度与  $n_1 \sin \theta_1 = n_2 \sin \theta_2$  相关联。

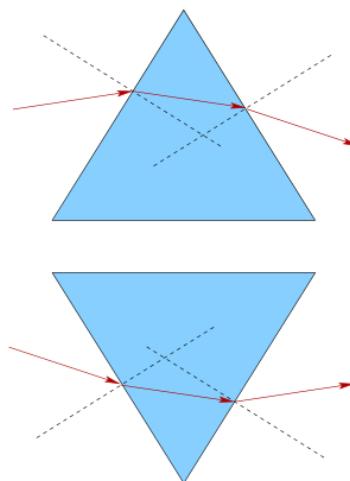


图 4.10: 上半部分说明只要入射光线的斜率不太高, 一个简单的棱镜可将向上的光线弯曲成向下的光线。这是通过在进出边界应用斯内尔定律实现的。将棱镜倒置会导致向下光线上升。把它们放在一起, 我们就会看到, 一个透镜就是棱镜的组合, 它们迫使发散的光线通过折射实现会聚。

## 棱镜

设想利用如图 4.10 的方法通过棱镜照射激光束。斯内尔定律可以用来计算光线进入和退出棱镜后的弯曲方式。对于垂直棱镜，略微向上的光线经过棱镜后向下弯曲，棱镜的折射率越大，会导致更大的弯曲。通过将棱镜倒置，向下的光线会向上弯曲。一旦折射率固定，弯曲仅取决于射线进入和离开表面的角度，而不取决于棱镜的厚度。为了生成一个透镜，我们将利用这个原理并构建一种图 4.10 的曲面版本。

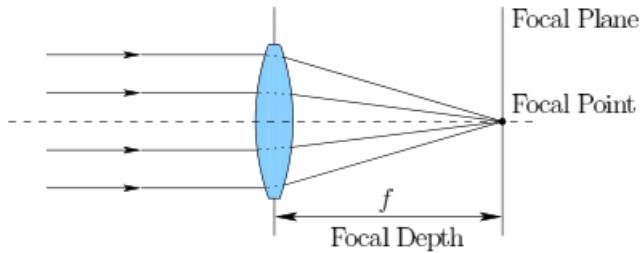


图 4.11：一个简单的凸透镜会使平行光线聚焦在焦点处。虚线为光轴，它垂直于镜头并穿过其中心。

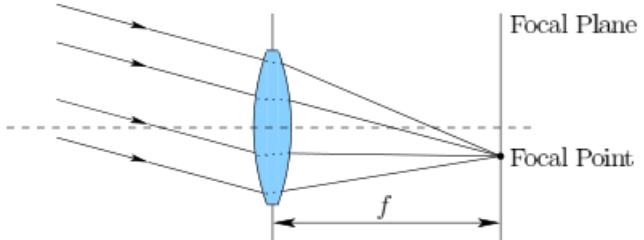


图 4.12：如果光线不垂直于透镜，则焦点偏离光轴。

## 简单的凸透镜

图 4.11 展示了一个简单的凸透镜。镜头表面不再是菱形，而是球状弯曲的，这使得平行的水平入射光线会聚到镜头另一侧的一个点上。这个特殊的聚合地点被称为焦点。它与镜头中心的距离称为焦距。

图 4.11 的入射光线有两个特点：1) 它们是平行的，因此对应于无限远的光源；2) 它们垂直于透镜的中心平面。如果光线平行但不垂直于透镜平面，则焦点相应地移动，如图 4.12 所示。在这种情况下，焦点不在光轴上。平行入射光线方向有两个自由度，对应了包含所有可能焦点的聚焦平面。不幸的是，这种平面度只是一个近似值；4.3 节则解释了这其中实际发生的事情。基于这种理想化的设定下，真实图像生成在图像平面上，就好像是一个投影屏幕，显示了世界在镜头前的样子（假设世界上的所有物体都很遥远）。如果光线不平行，那么仍然可以将它们聚焦成真实图像，如图 4.13 所示。假设透镜焦距为  $f$ 。如果将光源放置在与透镜距离为  $s_1$  的位置，当且仅当满足以下等式（斯内尔定律）时，经过该透镜的光线才会聚焦：

$$\frac{1}{s_1} + \frac{1}{s_2} = \frac{1}{f} \quad (4.6)$$

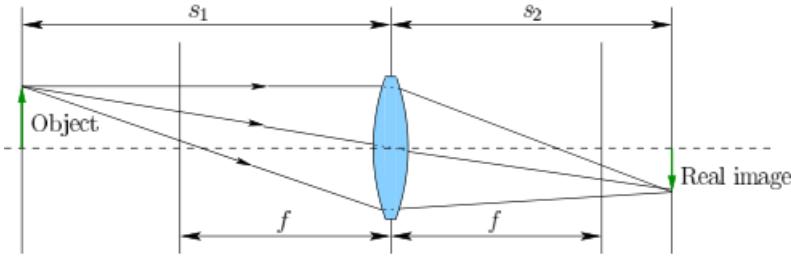


图 4.13：在现实世界中，一个物体并不会是无穷远的。如果使用 (4.6) 计算，当物体距离透镜  $s_1$  时，实际图像形成平面位于透镜后方  $s_1 > f$  处。

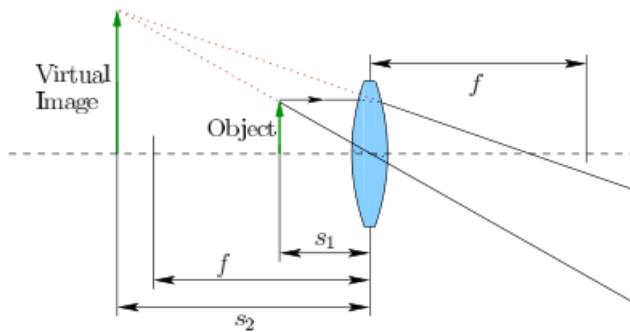


图 4.14：如果物体非常接近透镜，那么透镜不能强制出射光线会聚至某一点。但在这种情况下会出现一个虚拟图像，透镜的作用为放大镜。这是通常用于 VR 设备的方式。

图 4.11 对应于  $s_1$  无穷大的理想情况，对此求解 (4.6) 得到  $s_2 = f$ 。如果被观察物体位于与镜头垂直的平面中但不完全平坦，该怎么办？实际上在这种情况下，不存在可以将整个物体聚焦单个平面，我们必须接受这样一个事实，即只有大部分内容可以被汇聚起来。更糟糕的是，现实世界中绝大部分都是这样的情况，包括我们眼睛的会聚功能（见 4.4 节）。

如果光源放置得离镜头太近，那么出射光线可能会发散，以至于镜头不能强制会聚。如果  $s_1 = f$ ，则出射光线将平行 ( $s_2 = \infty$ )。如果  $s_1 < f$ ，则根据 (4.6) 得到  $s_2 < 0$ 。在这种情况下，不形成真实图像；但此时会发生一种有趣的现象：放大现象，也就是当我们看着镜头时会出现虚像，如图 4.14 所示。这正是图 2.11 所示的 View-Master 和 VR 头戴式设备发生的情况。

### Lensmaker 方程

对于一个给定的简单透镜，焦距  $f$  可以使用 Lensmaker 等式计算得到：

$$(n_2 - n_1) \left( \frac{1}{r_1} + \frac{1}{r_2} \right) = \frac{1}{f} \quad (4.6)$$

参数  $r_1$  和  $r_2$  分别表示透镜前后表面的曲率半径。该公式假设透镜很薄，这意味着镜头厚度相对于  $r_1$  和  $r_2$  来说非常小。此外，通常假定  $n_1 = 1$ ，为空气的折射率。

## 凹面镜片

为了完整起见，我们还给出了简单凹型透镜的情况，如图 4.15 所示。平行光线经过凹型透镜被迫发散，而不是会聚；然而，该情况仍可以通过反向延伸发散射线找到焦点，因而也有负焦距的概念。该情况下，Lensmaker 公式（4.7）需要稍调才可计算  $f$  [104]。

## 屈光度

VR 光学系统中，单个镜头通过连续组合几个透镜的方式构成。这样组合的效果是什么呢？眼科医生则提出了一种用简单算术来解决这个问题。该想法是定义一个屈光度  $D$ ， $D = 1/f$ ，即焦距的倒数。如果透镜将平行光线聚焦在镜头后方 0.2m 处，则  $D = 5$ 。较大的屈光度表示较强的聚焦能力。对于凹透镜而言， $D < 0$ ，则较低的数值意味着较强的发散能力。为了连续组合多个透镜，我们只需利用到它们的屈光度，便可确定其作为单个简单镜片时的等效功率。图 4.16 展示了一个简单的例子。

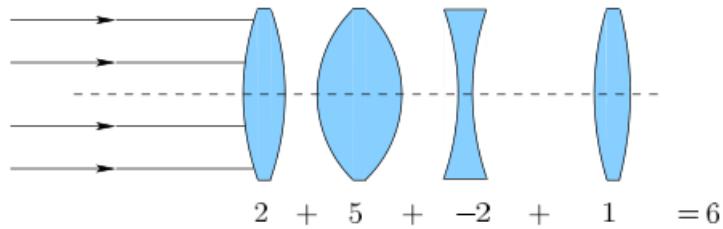


图 4.16：计算透镜链光强度的方式十分简单：将它们各自的屈光度代入加法计算。这四个透镜的配置相当于一个 6 屈光度的镜头，其焦距为 0.1667 米

## 4.3 光学畸变

如果现实世界中的镜头完全按照 4.2 节中的描述进行光线会聚或发散，那么 VR 系统将简单得多，且令人印象深刻。事实上，许多统称为畸变的缺陷会降低由镜头形成的图像质量。由于这些问题在日常使用中都很明显，例如通过 VR 设备观看内容，因而这些问题非常重要，需要采取相关的补偿措施并应用到 VR 系统中。

## 色差

回顾一下 4.1 节的内容，光通常是一束具有波长光谱的波。你可能看到过，当白光通过棱镜折射时，整个可见光谱按颜色被很好地分开。这是一个美丽的光学现象，但对于镜头来说，这是非常糟糕的，因为它分散了图像的各种颜色成分。这个问题被称为色差。

问题的实质在于通过介质的光速取决于波长。因此，我们设定介质的折射率  $n(\lambda)$  以表示它是  $\lambda$  的函数。图 4.17 展示了简单凸透镜的色差现象。此时，焦距为波长的函数。如果我们沿着相同的光线将红色、绿色和蓝色激光直接照射到镜头中，则每种颜色的光线会在不同的位置穿过光轴，产生红色、绿色和蓝色焦点。

回顾 4.1 节的光谱功率分布和反射函数。对于常见的光源和介质，透过镜头的光线会形成连续的焦点集合。图 4.18 显示了一个具有色差伪像的图像。因而可以通过组合不同介质的凸透镜和凹透镜减少色差，使得发散的射线被强制收敛 [298]。

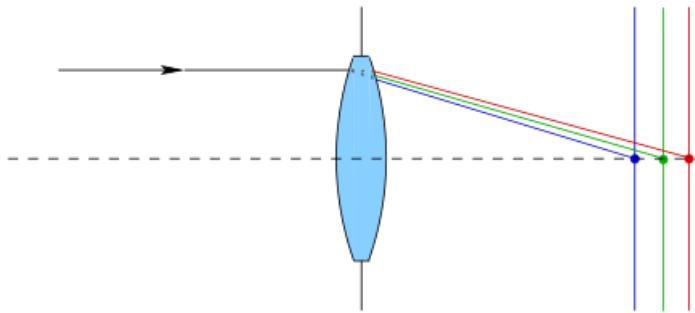


图 4.17：色差是因不同波长的光线在同一介质内的不同光速所引起的，导致每种颜色的光线都有不同的聚焦平面。



图 4.18：上子图被适当地聚焦，而下子图遭受了色差问题。（图由 Stan Zurek 提供）

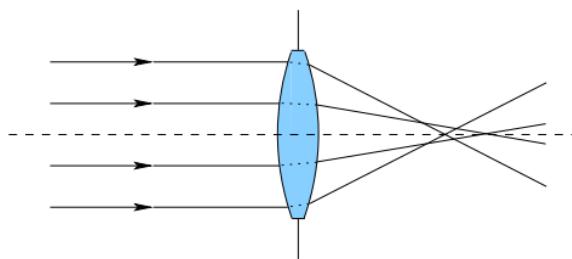


图 4.19：球面像差引起焦点无法聚集，因为远离光轴的光线比周边的光线折射更厉害。

### 球面像差

图 4.19 显示了球面像差，这是由远离镜头中心的光线折射得比中心附近的光线多。其结果类似于色差，但这种现象是单色像差，因为它与光波长无关。输入的平行光线集中在不同的深度，而不是集中在一个点上。得到的结果自然是成像模糊，而且无法通过移动物体、镜头或图像平面来补偿。此外，图像还可能会聚焦到一个曲面上，称为 Petzval 曲面，而不是图像平面。这种像差是由于球形透镜造成的。非球面透镜更复杂，它具有非球面表面，专门用于消除球面像差并减少其他像差。

## 光学失真

即使图像本身投射到图像平面上，它也可能在外围发生扭曲。假定透镜是径向对称的，则这种失真是随着距离光轴越远而图像拉伸或压缩越严重的失真现象。图 4.20 举了两个例子来解释这种失真现象：桶形失真和枕形失真。对于具有宽视场的镜头，失真更强，特别是鱼眼镜头这种极端情况。图 4.21 显示了一个桶形失真较强的图像。纠正这种失真对于目前具有宽视场的 VR 头戴式设备至关重要，否则，呈现的虚拟世界就会变形。

## 像散性

图 4.22 描述了像散性，这是对不垂直于晶状体的入射光线发生的晶状体像差。直到现在，我们的镜头图纸都是 2D 的，然而，需要引入第三个维度来了解这种新的畸变。射线可以在一维上发生轴偏移，但在另一维中对齐。通过沿着光轴移动图像平面，不可能使图像成为焦点。相反，这样会出现水平和垂直震源深度，如图 4.23 所示。

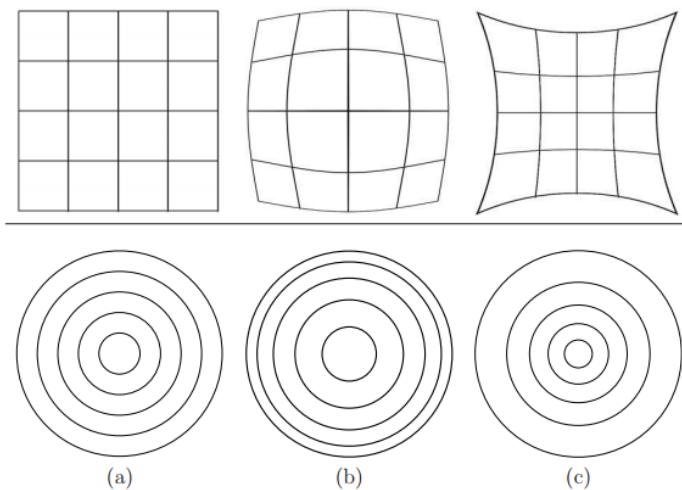


图 4.20：常见的光学畸变。（a）原始图像。（b）桶形失真。（c）枕形失真。对于第一行的畸变，网格变得非线性扭曲。第二行说明它仍然保持圆对称。



图 4.21：一个带有桶形失真的图像，由鱼眼镜头拍摄。（图片由维基百科用户 Ilveon 提供。）

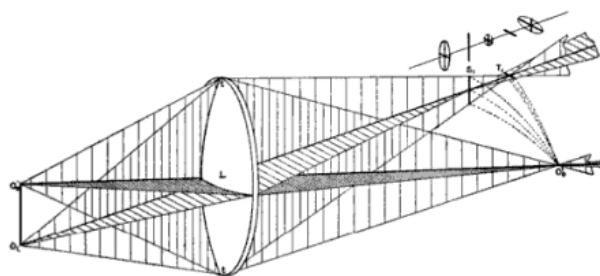


图 4.22：像散性主要是由于入射光线在一个平面内偏离轴线，而在另一个平面内接近垂直。

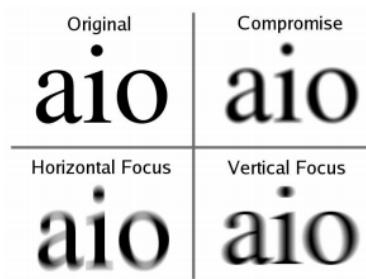


图 4.23：由于像散性，无法将图像完美地聚焦。在一个深度上，它可能会水平聚焦，而另一个深度上则是垂直聚焦。因此我们被迫需要选择妥协。

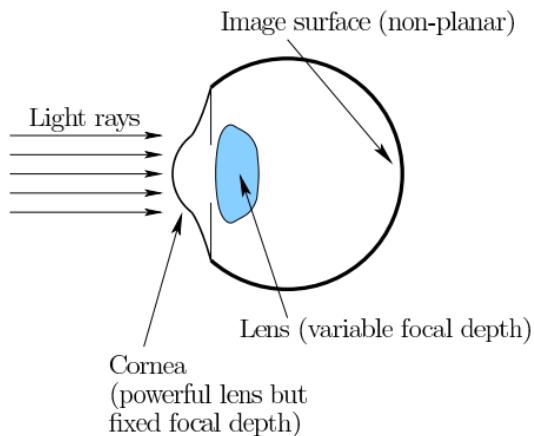


图 4.24：人眼光学系统的简化视图

### 昏迷和眩晕

最后，晕眩也是一个失真。在这种情况下，由于光线远离镜头垂直，图像放大倍数变化很大。结果是图像平面中的“彗星”图案。另一种现象是镜头眩光，非常明亮的光线通过镜片散射并且显示圆形图案。这经常出现在电影中，当视点通过太阳或星星时产生，有时候是人为添加的。

此节的所有像差都会使系统复杂化或降低 VR 设备的体验；因此，大量的努力都花在了缓解这些问题上。

## 4.4 人眼

我们已经在本章中以足够的概念来描述人眼的基本操作，这显然是任何 VR 系统中的重要组成部分，被视为镜头和图像光学系统的一部分。人类视觉的生理和感知部分我们将在第 5 章介绍。

图 4.24 显示了人眼的横截面。平行光线从左边进入，与图 4.11 相比，它显示了工程凸透镜的相似情况。虽然眼睛的操作与设计的环境相似，但在这个阶段会出现几个重要的区别。焦平面被视网膜的球形曲面所取代。视网膜包含将光转换为神经脉冲的光感受器；这在 5.1 节和 5.2 节中有介绍。眼球的内部实际上是液体，而不是空气。沿着从外部空气到视网膜路径结构的折射率如图 4.25 所示。

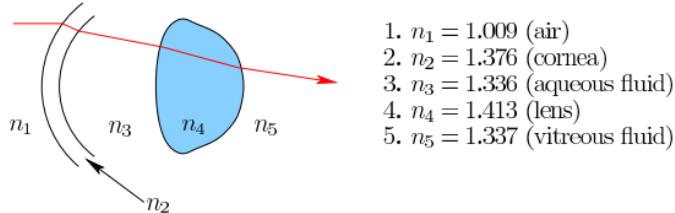


图 4.25：一束光在击中视网膜之前通过五个媒质，它们的折射率已在图中标出。考虑到斯涅尔定律，由于从空气到角膜的过渡，将发生最大弯曲。请注意，一旦光线进入眼睛，它只能通过液体或固体介质。

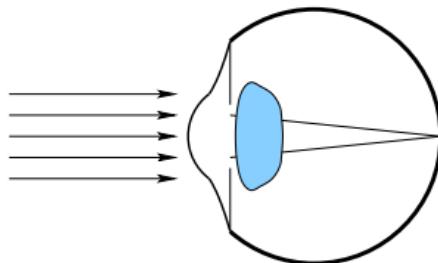


图 4.26：带有晶状体的正常眼睛。

眼睛的光焦度眼球的外径大约为 24mm，这意味着至少需要 40D 的镜片才能使平行光线聚焦到眼睛内部的视网膜中心（回忆 4.2 节中的屈光度）。实际上眼睛有两个凸透镜：角膜和晶状体。角膜是最先进入的眼睛的最外面部分，具有最大的光焦度，约 40D。眼睛晶状体不那么强大，提供额外的 20D。通过添加屈光度，角膜和晶状体的组合是 60D，这意味着平行光线聚焦到距离外角膜大约 17mm 距离的视网膜上。图 4.26 显示了这个系统如何作用于正常视力的人的平行光线。远处物体的图像由此聚焦到视网膜上。

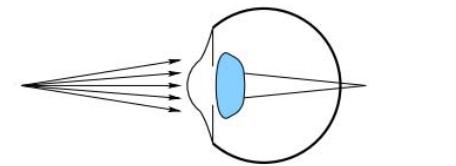


图 4.27：一个近物会产生发散的光线，但如果晶状体松弛，图像在视网膜上模糊不清。

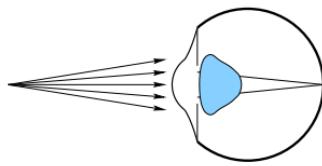


图 4.28：调节过程：眼睛肌肉拉动晶状体，使其增加聚光力量并将图像聚焦在视网膜上。

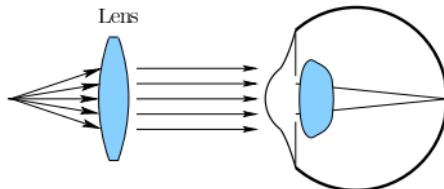


图 4.29：在眼前放置凸透镜是提升聚光能力的另一种方式，以便附近的物体可以在眼睛中汇聚焦点。这就是老花镜的原则。

### 焦距适应

当我们想把注意力集中在一个附近的物体上，而不是一个“无限远”的地方时，会发生什么？如果不改变光学系统，图像在视网膜上会模糊，如图 4.27 所示。幸运的是，晶状体改变了屈光度以适应更近的距离。这个过程被称为适应，如图 4.28 所示。屈光度的变化是通过拉动晶状体改变其形状的肌肉来实现的。在年幼的儿童中，晶状体可以增加 15 到 20D 的力量，这就解释了为什么一个孩子可能会把一些东西放在你面前，并希望你专注于它，因为他们能！在 20D 时，这对应于聚焦在离角膜仅 5cm 的物体上。年轻人已经失去了这种能力，可以容纳 10D 左右。因此，在正常的视力下，他们可以读一本书到 10 厘米左右的距离（眼睛会有些疲劳）。一旦成年人达到 50 岁，几乎没有适应调节能力。这种情况称为老花眼。图 4.29 显示了最常见的治疗方法，即戴上老花镜。

### 视力异常

视力问题将陪伴一个人的一生。出现异常的一个可能是光学系统没有足够的力量来将平行光线聚到视网膜上。这种状况被称为远视。可以用眼镜来调节。简单的解决方法是在眼睛前面放置一个凸透镜（正屈光度），就像老花镜一样。相反，一些眼睛有太多的力量。这种情况称为近视，可将一个凹透镜（负屈光度）置于眼前，以适当降低聚光力。回想一下，我们有两只眼睛，而不是一只眼睛。这使得每只眼睛都有可能出现不同的问题，导致每只眼睛的镜片屈光度不同。其他视力问题可能不只是聚光力的问题。最常见的是散光，这在第 4.3 节中有介绍。在人眼中，这是由于角膜过于椭圆形而不是径向对称造成的。需要特殊的镜片来纠正这种情况。您可能还想知道第 4.3 节中的像差是否出现在人眼中，如色差。他们会发生，但是他们会被我们的大脑自动纠正，因为我们已经在生活中学会了如何去解释这些有缺陷的图像。

### 简单的 VR 头戴式设备

现在假设我们正在通过放置一个非常接近眼睛的屏幕来构建 VR 头戴式设备。如果距离接近 10 厘米，年轻人已经无法把它放在焦点上。我们希望将其放近点以便填满用户的视角区域。因此，通过使用凸透镜来增加光焦度，其功能与老花镜相同，见图 4.30。这也是第 4.2 节放大的过程。镜头通常放置在焦点处。使用 (4.6)，这意味着  $s_2 = -f$ ，而导致  $s_1 = \infty$ 。

屏幕显示为无限远的巨大虚拟图像。但请注意，真实图像仍会投射到视网膜上。我们不会察觉身边的世界，除非视网膜上形成了身边的真实图像。

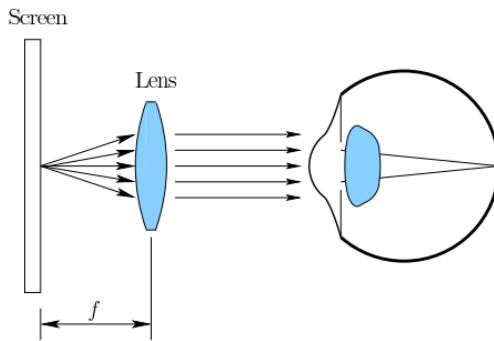


图 4.30：在 VR 头戴式设备中，镜头放置得使屏幕看起来无限远。

为了解决某些人的视力问题，头戴式设备上可能会出现对焦旋钮，这会改变镜头和屏幕之间的距离。这调整了光焦度，使得镜片和角膜之间的光线不再平行。它们可以被制造为会聚，从而帮助远视的人。或者，它们可以分散，这有助于近视的人。这样，人们可以取下眼镜看屏幕，而不用将镜片放在眼睛之前。但是，如果每只眼睛需要不同的屈光度，则每只眼睛都需要一个对焦旋钮。此外，如果他们有散光，那么这个情况不能被纠正。将眼镜放在头戴式耳机内部可能是唯一剩下的解决方案，但它可能不舒服，并且可能会缩小视野。

本节中许多细节已被略过或简化。VR 头戴式设备的一个重要细节是每个镜头应该完全居中在角膜前方。如果两个镜头之间的距离永久固定，那么这样就不能满足使用头盔的每个人的需求。瞳距是人眼中心之间的距离。人类的平均瞳距约为 64 毫米，但其种族、性别和年龄（以儿童为例）差别很大。为了能够将镜头对准每个人，镜头中心之间的距离应该可以从 55 到 75 毫米调整。这是双筒望远镜的常见范围。不幸的是，情况并非如此简单，因为我们的眼睛也在眼窝内旋转，这改变了角膜相对于晶状体的位置和方向。这放大了第 4.3 节介绍的光学像差问题。第 5.3 节将介绍眼球运动。另一个重要细节是保真度：放置在我们眼前的屏幕需要什么分辨率，以免我们注意到像素？一个类似的问题是，在一张印刷好的纸上需要多少个每英寸点数 (DPI)，以便即使在放大镜下观察时也不会看到这些点？我们在 5.1 节回到这个问题。

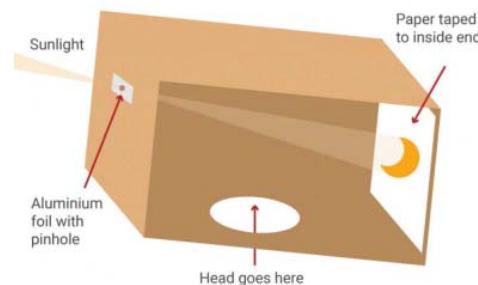


图 4.31：用于查看日食的针孔摄像机。（图来自于 [TimeAndDate.com](http://TimeAndDate.com)。）

## 4.5 相机

现在我们已经解决了人眼问题，描述工程化的眼睛模型似乎很自然，这被称为相机。人们已经发明并使用了数百年的相机，从暗箱开始，允许光线通过针孔，呈现有真实图像的表面。图 4.31 显示了一个可以用于观看日食的例子。（回想一下第 3.4 节中的透视变换数学）。十八世纪的艺术家使用了一面镜子和一张描图纸来转换图像并使其完美复制到平面上。在整

个 19 世纪，各种基于化学的技术被开发出来，直到发现了光子撞击成像表面会自动蚀刻出图像。在整个 20 世纪，电影被广泛使用，直到数码相机通过使用传感器以电子方式捕获图像，从而完全避免了蚀刻过程。两种流行的技术是电荷耦合器件（CCD）阵列和 CMOS 有源像素图像传感器，如图 4.32（a）所示。这种数字技术记录沿着图像击中每个像素位置的光量，从而直接产生捕获的图像。这些设备的成本近年来一落千丈，业余爱好者可以购买像图 4.32（b）所示的相机模块，价格低于 30 美元。

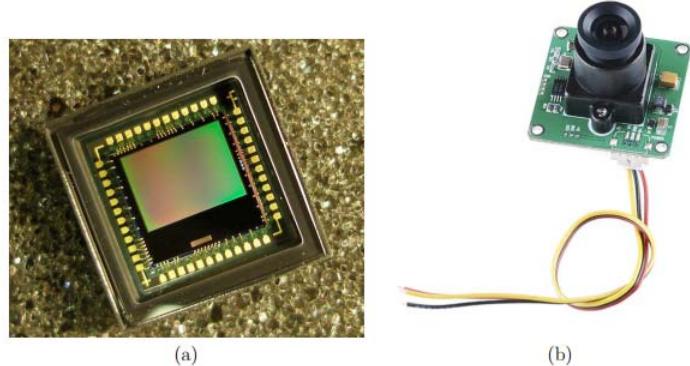


图 4.32：(a) 一个 CMOS 有源像素图像传感器。 (b) 低成本 CMOS 摄像头模块 (SEN-11745)，适用于爱好者项目。

## 快门

拍摄数字图像时出现几个实际问题。图像是像素的二维阵列，每个像素都具有红色 (R)，绿色 (G) 和蓝色 (B) 值，其值通常在 0 到 255 之间。请考虑命中图像平面的总光量。对于更高分辨率的相机，每个像素通常会有更少的光子，因为像素更小。每个感测元件（每个像素一个颜色一个）可以被想象成一个收集光子的桶，就像雨滴一样。要控制光子的数量，快门会阻挡所有的光线，打开一段固定的时间，然后再次关闭。长时间间隔（低快门速度），收集更多光线，然而，缺点是场景中的运动物体将变得模糊，并且传感元件可能会被太多的光线饱和。摄影师必须在确定快门速度时考虑到场景中的光线量，感光元件的灵敏度以及相机和物体在场景中的移动情况以达到平衡。

与快门相似，CMOS 传感器的工作是逐行发送图像信息。因此，传感器与滚动式快门连接，在信息发送之前，每个线路都允许光线进入。这意味着捕捉不会在整个图像上同步，从而导致奇怪的伪像，如图 4.33 所示。与滚动快门和运动一起工作的图像处理算法通常会转换图像以纠正这个问题。CCD 传感器一次抓取并发送整个图像，产生全局快门。CCD 一直比 CMOS 传感器更昂贵，这导致智能手机中滚动快门相机的普遍出现；然而，全局快门相机的成本迅速下降。



图 4.33：由于滚动式快门效应，飞行直升机的机翼明显向后弯曲。

光圈光学系统也影响到达传感器的光量。如图 4.31 所示，使用针孔，光线会落到图像传感器上，但对于大多数目的而言（除了观看日食以外）它不够明亮。因此，使用凸透镜代替，使得多条光线会聚到图像平面中的相同点，如图 4.11。这为每个感测元件产生更多的光子。主要缺点是镜头在单一深度地聚焦物体，同时模糊其他物体，如图 4.6。在针孔的情况下，所有的深度基本上都是“对焦”的，但是可能没有足够的光线。因此，摄影师希望调整光学系统的行为更像一个针孔或更像一个完整的镜头，取决于预期的结果。结果是一个可控的光圈（图 4.34），它出现在镜头后面并设置光线进入的孔的大小。小半径通过阻挡除镜头中心之外的所有物体来模仿针孔。较大的半径允许光线穿过整个镜头。我们的眼睛以类似的方式通过收缩或扩大我们的瞳孔来控制光照水平。最后，请注意，光圈越大，第 4.3 节中所述的像差越会干扰成像过程。

### 拓展阅读

大部分基本镜头和光学系统概念都包含在大学物理入门教材中。对于更高级的概念，尤其是镜头像差，请参阅经典的光学工程文本：[298]。快速覆盖光学几何的便捷指南是[104]。[156]中详细介绍了利用电子，激光和 MEMS 的光学系统。这为理解下一代视觉显示技术提供了基础。一本考虑人眼与工程光学元件结合的出色书籍[296]。相机涵盖了许多不同的角度，包括计算机视觉[111,318]，相机工程[126]和摄影[297]。成像的数学基础在[19]中有全面的介绍。

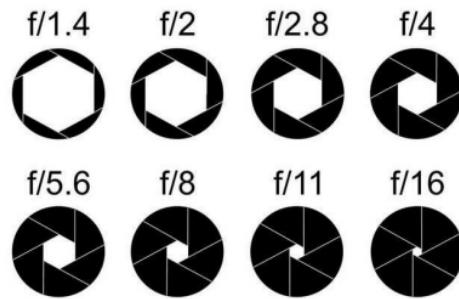


图 4.34：光圈设置的光谱，控制进入镜头的光量。显示的数值称为焦比或光圈。

# 第五章 人类视觉生理学

你所感知到的世界其实都是“你脑海中的世界”。从第四章，尤其是 4.4 节中，你应该了解到，光线进入到我们的视网膜，从而形成图像，来呈现物理世界中的颜色、运动以及空间关系。对于视力正常的人，图像的视野广阔，清晰度、速度、准确性和分辨率看上去也是理想的。然而事实上并不是这样的，我们被视觉系统所“欺骗”了。在本章中，你将会学到，视觉上近乎完美的图像呈现其实是一种错觉，我们的神经结构在图像中填充了貌似真实的细节，从而生成一幅符合我们经验的正常图像。搞清楚视觉系统的工作原理对我们构建 VR 系统是很重要的，比方说我们要知道，现如今的显示技术并不能完美的复制周围场景，用虚拟现实系统“欺骗”视觉神经的话也会产生一些副作用。

上接 4.4 节，我们在 5.1 节中会介绍人眼的解剖结构，主要关于感光体，用来匹配“输入像素”和 VR 数字显示的“输出像素”；5.2 节从神经科学角度解释了从感光体到视觉皮层传递视觉信息的过程；5.3 节解释了我们的眼睛如何移动，这样做的话会不断干扰我们视网膜中的图像；5.4 节通过应用视觉生理学的知识来确定 VR 的需求，例如屏幕分辨率等。

## 5.1 从角膜到感光体

### 眼睛的构成

图 5.1 显示了人眼的生理结构。形状像是一个直径约 24mm 的球体。角膜是一个坚硬、透明的表面，光通过它进入人眼并提供最大的光焦度（详见 4.4 节）。外表面的其余部分由称为巩膜的坚硬白色层保护。眼睛内部主要由玻璃体液组成，一种透明的胶状物质，允许光线穿透而几乎不产生变形衰减。

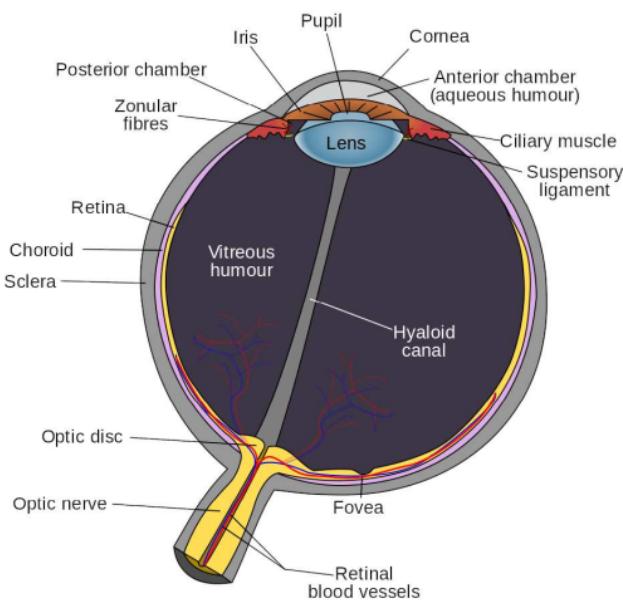


图 5.1：人眼的生理结构，图中所示的是右眼的水平横截面（鼻子在左边）（源自维基用户 Rhcastilhos）

当光线穿过角膜时，它们会经过含有房水的小室，这是另一种透明胶状物质。之后，光线经过瞳孔进入晶状体。瞳孔的大小由称为虹膜的盘状结构控制，用来调节允许通过的光量。

睫状肌可以改变晶状体的光焦度。穿过晶状体后，光线通过玻璃体液射向内层的视网膜。图 5.1 中显示的是 2D 横截面，因此视网膜呈弧形；然而要记住它实际上是一个 2D 表面。我们可以把它想象成一个与视觉显示对应的曲面。为了捕捉来自“输出像素”的光线，它与感光体排成一行，像“输入像素”一样。视网膜中最重要的部分是中央凹部分；视敏度是衡量视觉锐利度或清晰度的标准，由视象与中心凹之间的距离决定。视神经盘是视网膜上的一个小孔，神经脉冲通过其传送到眼外。它恰好位于鼻子中央的位置。

## 感光体

视网膜包含两种感光体：1) 视杆细胞，对弱光敏感；2) 视锥细胞，强光刺激才能引起兴奋，具有分辨颜色的能力。如图 5.2 所示，我们介绍一下感光体的大小，最小的视锥细胞的宽度约为 1000 纳米，很接近可见光的波长，意味着感光体不需要再小了。每个视网膜都包含大约 1.2 亿个视杆细胞和 600 万个密集分布在视网膜中的视锥细胞。图 5.3 展示了每种感光体的检测能力。视杆细胞的光谱灵敏度峰值在 498 纳米处，对应光谱中蓝色和绿色之间。视锥细胞可分类三种类型，基于其是否对蓝光、绿光或红光有所感应。

感光体在大的动态范围内感应光照。图 5.1 展示了几个常见的例子。亮度 (luminance) 由国际单位坎德拉/平方米 ( $\text{cd}/\text{m}^2$ ) 表示，对应每个区域内的发光强度。范围跨越 7 个数量级，从每 100 秒被 1 个光子击中到每秒被 100,000 个光子击中。在光照不强的时候，只有视杆细胞被触发，我们之所以无法在夜间区分颜色正是由于视杆细胞不能区分颜色所导致的。我们的眼睛可能需要长达 35 分钟才能完全适应弱光，导致一种称为暗视觉的单色模式。相反地，视锥细胞需要更强的光线才会变得活跃。适应这种称为明视觉的三色模式可能需要长达 10 分钟的时间（比如说，晚上你躺在床上，突然有人打开了灯，你无疑需要一段调整时间）

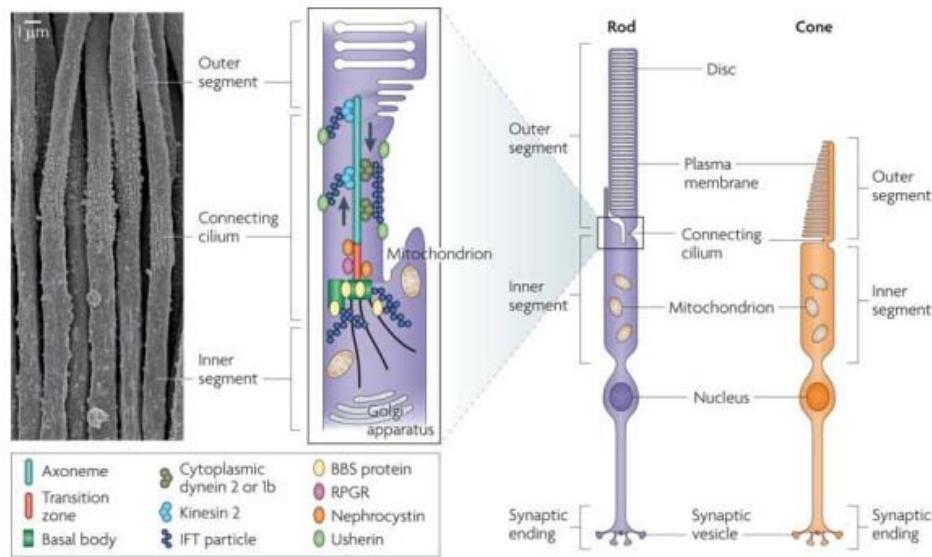


图 5.2 左图是感光体的电镜图；右图是视杆细胞和视锥细胞的组成和结构，外部片段含有光色素，在受到光子击中时可以产生电化学响应。

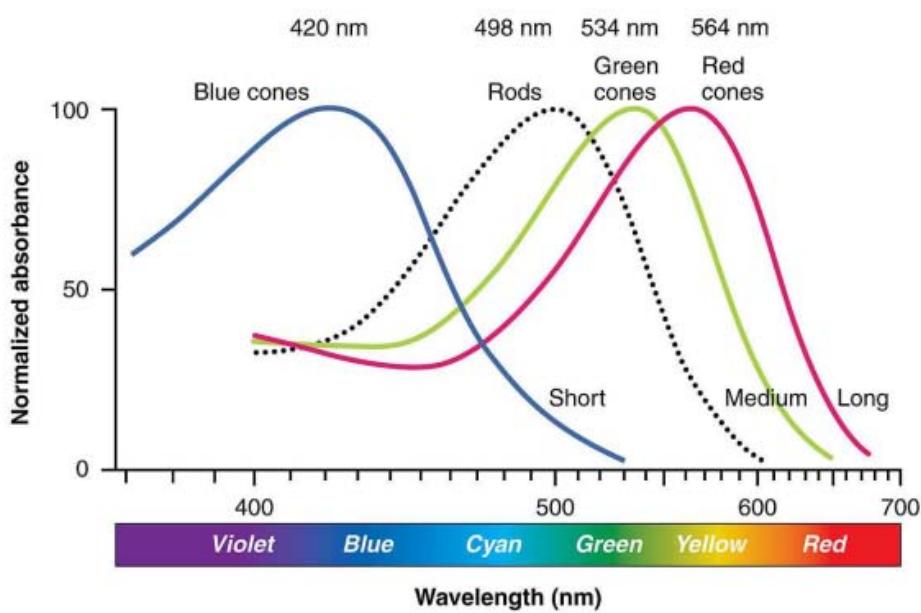


图 5.3 视杆细胞和视锥细胞对光波长的敏感度曲线（图由 OpenStax College 提供）

Light source	Luminance ( $\text{cd}/\text{m}^2$ )	Photons per receptor
Paper in starlight	0.0003	0.01
Paper in moonlight	0.2	1
Computer monitor	63	100
Room light	316	1000
Blue sky	2500	10,000
Paper in sunlight	40,000	100,000

图 5.4 一些熟悉的场景以及其对应的每秒击中感光体的光子数（图由【17, 22】提供）

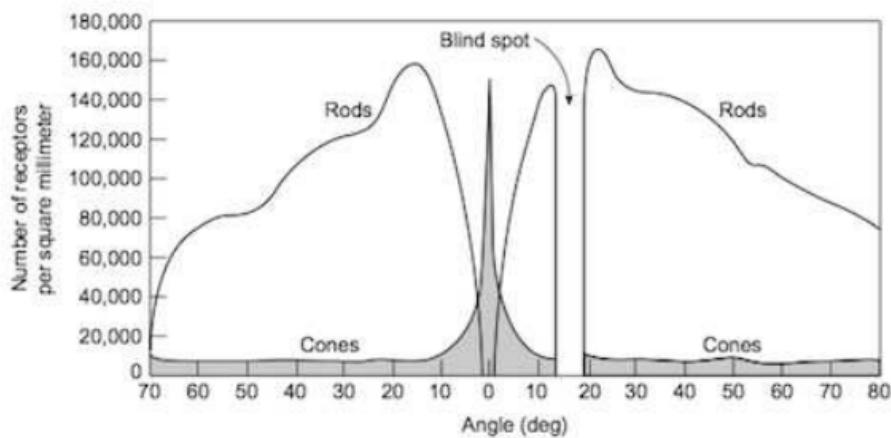


图 5.5 感光体密度和角度的关系函数，中间的空白是鼻侧（对应着从相反方向进入的光线）

### 感光体密度

如图 5.5 所示，视网膜上感光体分布密度不均，中心凹是感光体最集中的地方。中心凹最内部的直径仅为 0.5mm，角度范围为  $\pm 0.85^\circ$ ，包含了几乎全部视锥细胞。这意味着眼睛必须直接对准目标才能感知到清晰的彩色图像。整个中心凹的直径为 1.5mm（角度范围

$\pm 2.6^\circ$ ），外圈主要是视杆细胞集中的区域。从侧面进入角膜的光线落在视网膜上，光密度较低，视锥细胞密度也较低，这对应了周边视觉。对于这种情形，相较于有效地区分颜色，我们对周边环境的运动检测要好很多。经推测，对周边环境的运动检测可能会帮助我们的祖先避免被食肉动物抓到，因此这个特性被遗传下来，保留至今。最有意思的部分是视网膜的盲点区域，它不包含任何感光体。由于人眼的视神经是在视网膜前面，它们汇集到一个点上穿过视网膜连接大脑，如果一个物体的像刚好落在这个点上就不会被看到，我们称这个点为盲点，在 5.2 节中会详细介绍。

图 5.5 所示的感光体密度给我们带来了难题。如果裸眼视力正常，我们感知这个世界应该像是我们的眼睛在广角范围内捕捉鲜明多彩的图像。然而，这几乎是不可能的，因为我们智能在小角度范围内感知到清晰的彩色图像。此外，盲点还在我们感知到的图像中放置了一个“黑洞”。然而，不可思议的是，我们的感知过程会出现错觉，从而填补出一幅完整的图像。这种错觉的产生有几点原因导致：一是利用上下文信息来填充缺失的细节，在 5.2 节中会详细介绍；二是通过频繁眼球运动来完成，会在 5.3 节中介绍。如果你仍然不相信你所看到的完整图像是由大脑产生的错觉，你可以尝试一下“盲点实验”，如图 5.6 所示。



图 5.6 盲点实验。闭上右眼，一直盯着右面的“X”，不断变化眼睛到纸张（或屏幕）的距离。在一定范围内，你会发现左边的点消失了。你可以在有纹理的表面（如方格纸）上重复该实验，当左边点消失时你可能会发现在该处的纹理与其他地方无异。这表明你的大脑会在盲点区域填充上预期的纹理。

## 5.2 从感光体到视觉皮层

感光体是将光能刺激转化为叫做神经冲动的电信号的传感器，从而将关于外界的信息传入到神经结构中。回忆 2.3 节中介绍的，信号从感光体到视觉神经皮层是以分层的方式向上传播的（图 2.19）。考虑每个感光体对神经元网络的影响，图 5.7 展示了一个简化的模型。随着层数的增加，受影响的神经元也迅速增加。图 5.8 展示了相同的过程，但以不同的方式突出展示了影响单个神经元的感光体数量随着层级的增加而增加。最低层的神经元可以对来自相邻感光体的信号作出简单的比较；随着层数的增加，神经元可能会对较大的视网膜图像做出反应。在本节中我们可以看到很多神经结构，可以使我们更好的理解这个原理。最后，当信号到达最高层时（远远超出图中所示的层数），将来自之前记忆的信息与从感光体传入的信息结合在一起。当大脑进行处理时，会产生感知的结果，例如识别人脸或判断树的大小，整个过程大约需要 100 多毫秒。

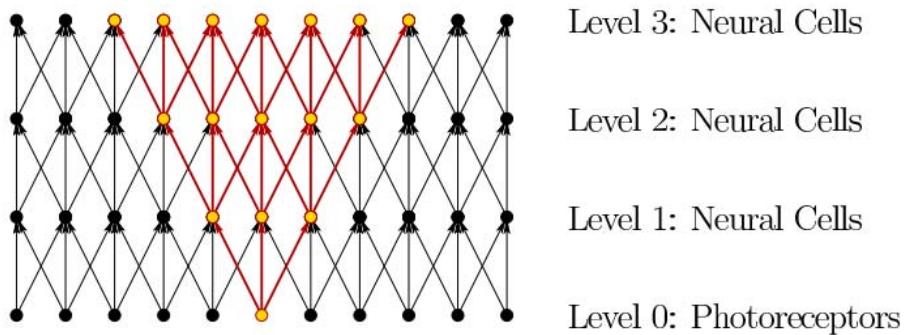


图 5.7 简单层次结构的四个层级。每个圆点对应一个神经细胞或感光体，箭头表示信息流。感光体在 0 级产生信息。在整个非常简化和理想化的视图中，每个感光体和神经元恰好连接到下一层的三个感光体或神经元。红色和金色部分突出显示了单个感光体随着层数增加而可能产生的增长区域。

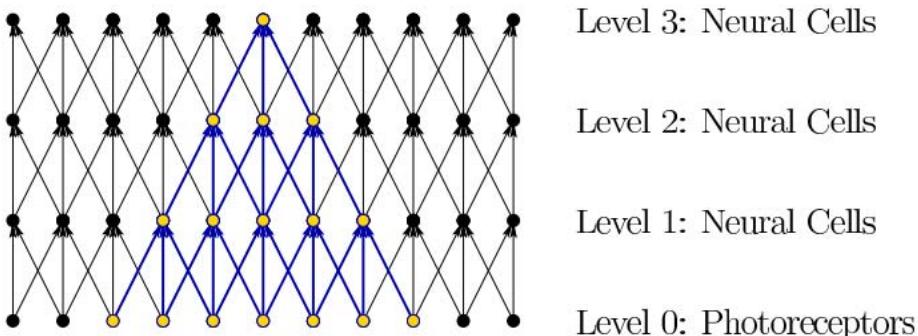


图 5.8 与图 5.7 相同，只是突出展示了进入单个神经元的信息。考虑参与单个神经细胞反应的一组感光体，称为感受野。随着层级的提高，感受野的大小迅速增加。由于感光体的空间排列，意味着每个神经元对视网膜上图像中不断增长的区域有所反应。区域随着层级的增加不断增长。

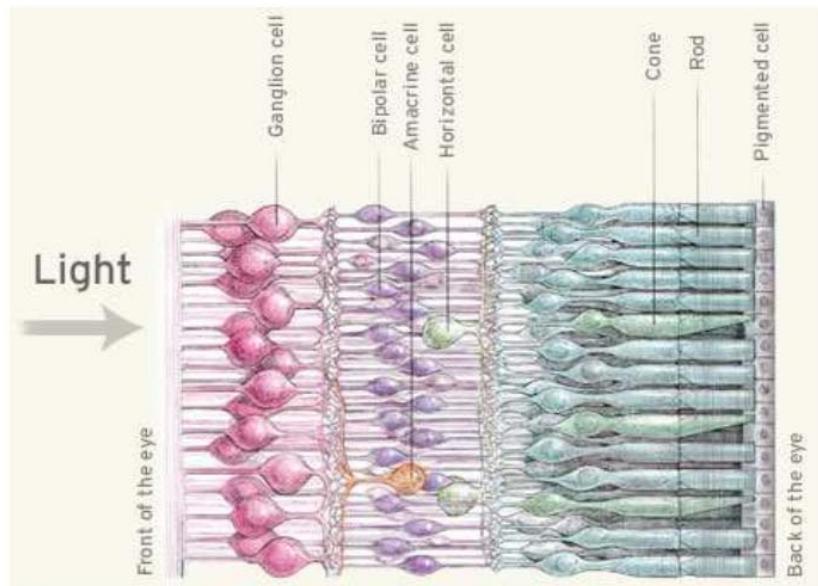


图 5.9：光线经过视杆细胞和视锥细胞之前要先穿过一些神经层（Institute for Dynamic Educational Advancement 提供）

现在更仔细地考虑第一层神经元，如图 5.9 所示。信息从右向左发送，经过视杆细胞和视锥细胞传递到双极细胞、无长突细胞和平滑细胞。这三种细胞位于内核层。信号通过那里到达神经节细胞，用来形成神经节细胞层。要注意，光传递的方向似乎是反的，在到达感光体之前要先经过这些神经细胞。这是由于人眼视网膜是从内到外的，如图 5.10 所示。像是章鱼和其他头足类，光是直接进入感光体的。视网膜从内到外的结构会导致神经节细胞的轴突不能直接连到视神经（图 5.10 中的第三项），从而无法直接将信号发送到眼外。因此，我们的视网膜上有一个小孔被“刺破”，这样来自神经节细胞的“信号流”可以被传到眼外（图 5.10 中的第四项），这也正是图 5.6 所示盲点的成因。

研究过图 5.9，我们应该了解到神经细胞的排列方式并不是如图 5.8 所示的理想方式。双极细胞将来自感光体的信号传递给神经节细胞。一些双极细胞仅仅与视锥细胞相连，每个双极细胞连接 1 到 10 个视锥细胞。其余的仅与视杆细胞相连，每个双极细胞连接 30 到 50 个视杆细胞。

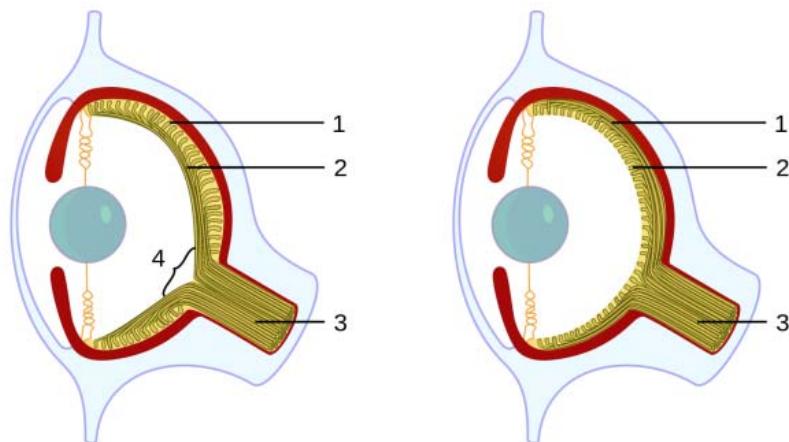


图 5.10：脊椎动物（包括人类）具有内外视网膜，这导致出现盲区并且也导致光感受器远离入射光。左边为脊椎动物的眼睛，右边为头足类眼睛，可以解释：光感受器面向光线，没有盲点。（图由 Jerry Crimson Mann 提供。）

根据功能的不同存在两种类型的双极细胞。ON 双极性在其连接的感光器中光子吸收率增加时激活。一个 OFF 双极在减少光子吸收时激活。连接到视锥细胞的两极有两种类型，然而，视杆细胞的双极只有 ON 双极。细胞两级的连接被认为是垂直的，因为它们直接从感光器连接到神经节细胞。这与内核层中剩余的两种细胞类型形成对比。水平细胞通过输入（树突）连接到半径高达 1mm 的感光器和双极细胞。它们的输出（轴突）被输入到感光器中，引起横向抑制，这意味着一个感光器的激活倾向于减少其周围细胞的激活。最后，无长突细胞和平滑细胞和其他无长突细胞水平连接，与神经节细胞之间垂直连接。连接方式有几十种类型，他们的功能不能轻易理解。因此，即使在最底层，科学家也不能完全理解人类的视觉。尽管如此，对于我们设计有效的 VR 系统和预测其他人类对视觉刺激的反应的能力，我们了解得很清楚。

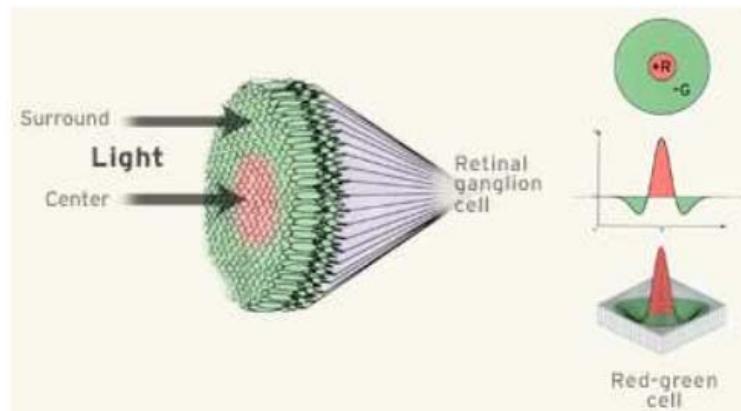


图 5.11：中心神经节细胞的感受野（图由动态教育发展研究提供）

在神经节细胞层，有几种细胞处理部分视网膜图像。每个神经节细胞都有一个大的感受野，它对应于如图 5.8 所示的有助于其活化的感光器。神经节细胞中最常见且理解最透彻的三种细胞称为小侏儒，阳伞和双分化细胞。他们基于光感受器上刺激物的空间，时间和光谱（颜色）变化对其接受区域执行简单的滤波操作。图 5.11 举了一个例子。当中心检测到红色但周围区域不是绿色时将触发神经节细胞。这种情况是空间对称性的一个例子，神经结构被设计用于检测局部图像变化。因此，将神经节细胞视为微小图像处理单元，可以根据时间，空间和/或颜色的局部变化来处理。他们可以检测和强调简单的图像特征，如边缘。一旦神经节轴突通过视神经离开眼睛，已经执行了大量的图像处理以帮助视觉感知。完全基于光子撞击感光器的原始图像从未离开眼睛。

视神经连接到称为外侧膝状核（LGN）的一部分丘脑，见图 5.12 所示。LGN 主要充当路由器，将感官信号发送到大脑，但也执行一些处理。LGN 将图像信息发送到位于大脑后部的初级视觉皮层（V1）。图 5.13 中突出显示的视觉皮层包含几个互相连接的区域，每个区域都执行特定的功能。图 5.14 显示了一个由视觉皮层进行的深入研究的操作。第 6 章将描述视觉感知，视觉感知是视觉皮层处理的有意识的结果，基于神经回路，刺激视网膜，来自其他感官的信息以及基于先前经验的预期。描述所有这些过程如何运作并整合在一起仍然是一个活跃的研究领域。

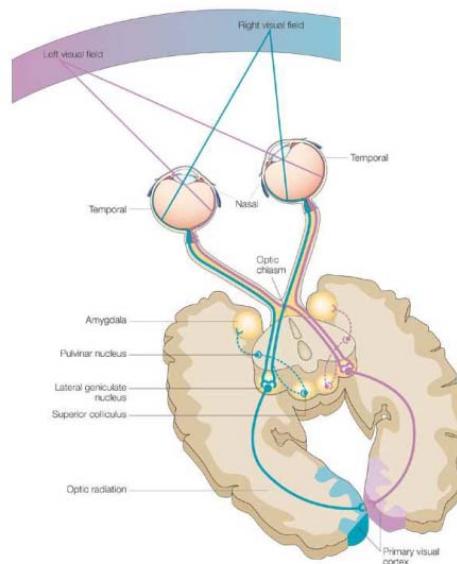


图 5.12：从眼睛到 LGN 到视觉皮层的视觉通路。请注意，来自视野的左侧和右侧的信息在皮质中交换。  
(图片来自 Nature Reviews: Neuroscience)

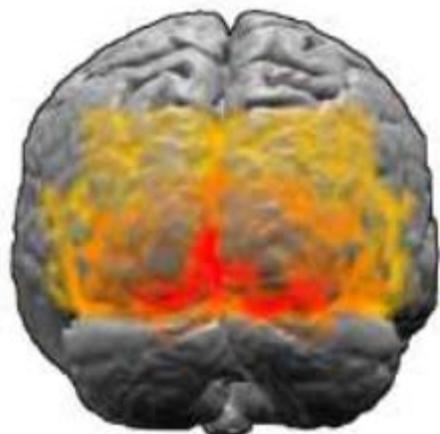


图 5.13：视觉皮层位于头部后方（图由华盛顿·欧文提供）

### 5.3 眼部运动

眼球旋转是人类视觉的一个复杂且不可分割的部分。它们既是自发地也可以非自发地发生，并且允许人们注视世界上的特征，即使他的头部或目标特征正在移动。眼球运动的主要原因之一是将感兴趣的特征定位在中央凹上。回顾第 5.2 节，只有中央凹可以感觉到密集的彩色图像，不幸的是它跨越了一个非常狭窄的视野。为了获得一个大型物体的连贯而详细的视图，眼睛在注视兴趣点的同时快速扫描它。图 5.15 举了一个例子。眼动的另一个原因是我们的光感受器由于其化学性质而对刺激响应缓慢。它们需要长达 10ms 才能完全响应刺激并产生长达 100ms 的响应。眼球运动有助于将图像固定在同一组光感受器上，以便它们可以载入电荷。这类似于低照度和慢快门速度下相机中出现的图像模糊问题。眼睛运动的其他原因是保持立体视图并防止适应持续刺激。为了解释最后的原因，实验证明，当眼球运动完全被抑制时，视觉感知完全消失[116]。当随着眼球运动的结合形成一个连贯的视角时，科学家很难预测和解释人们如何响应某些刺激。例如，当我们的眼睛在图 5.16 上面扫视时，会产生错觉，认为图似乎在移动。

#### 眼睛肌肉

每只眼睛的旋转由六个肌肉控制，每个肌肉通过肌腱附着于巩膜（外眼球表面）。图 5.17 和 5.18 显示了他们的名字和排布。肌腱以相反的方式牵拉眼睛。例如，为了执行偏转（左右旋转），内侧直肌和外侧直肌上的张力是变化的，而其他肌肉基本不受影响。为了引起俯仰运动，每只眼睛涉及四个肌肉。所有六个肌肉都参与执行俯仰和偏航，例如，向上和向右看。然而，我们的眼睛通常不是设计成有很多滚动的模式。如果存在许多滚动的话，想象一下，你把你的眼球上下颠倒！因此，在大多数情况下，将眼旋转估计为仅包含偏航和俯仰的 2D 集合是合理的，而不是 3.2 节中关于刚体旋转获得的全部 3D 自由度。

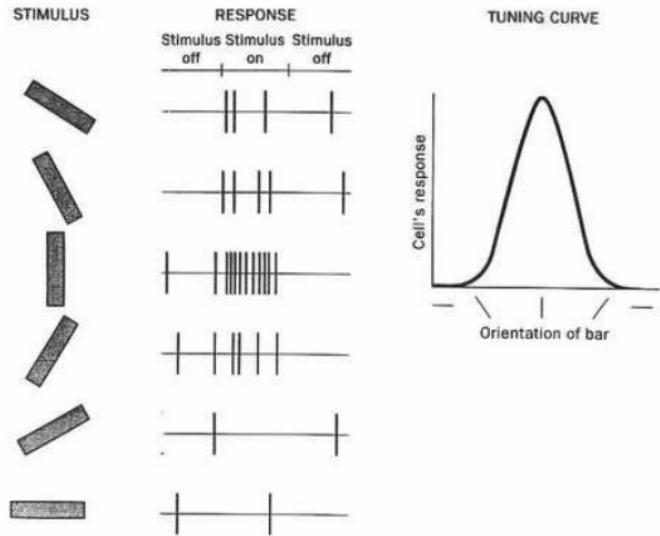


图 5.14：视觉皮层功能的一个示例是定向调整，其中单位记录由皮质中的单个神经元组成。当一个长条形的物体在眼前旋转时，神经元的响应变化如图所示。它强烈支持一个特定的方向。



图 5.15：使用扫视法扫描人脸的痕迹。

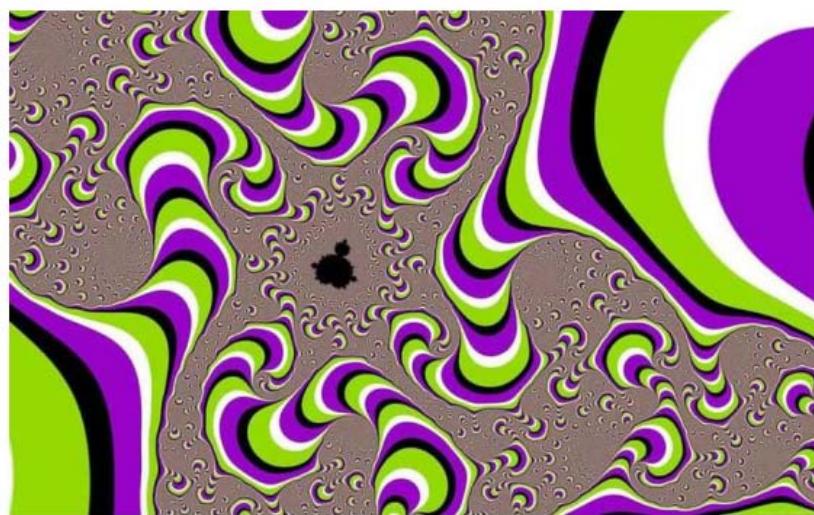


图 5.16：分形图似乎在移动，直到您仔细注意到单个零件位置。

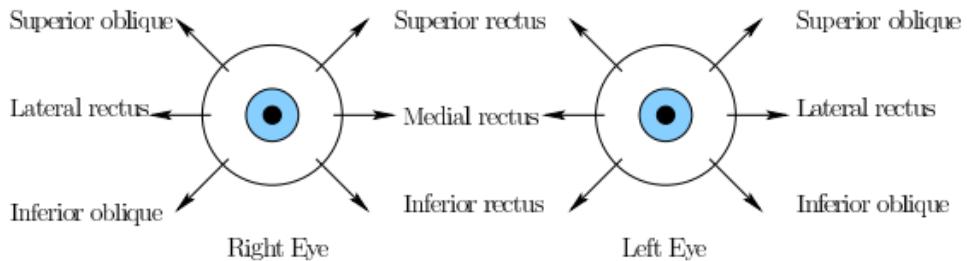


图 5.17：每只眼睛有六块肌肉，每块肌肉都能将瞳孔拉向其位置。

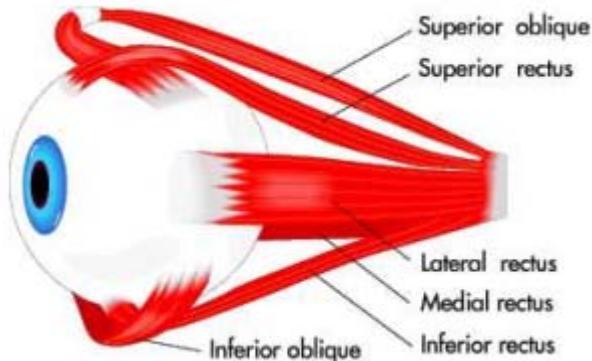


图 5.18：六条肌肉肌腱附着在眼睛上，这样可以实现偏航，俯仰和少量滚动。

### 眼球运动类型

我们现在考虑基于其目的的眼球运动，一共有六种类型：1) 快速扫视，2) 平滑移动，3) 前庭眼反射，4) 视动反射，5) 聚散，以及 6) 微跳动。所有这些动作都会使双眼大致以同样的方式旋转，除了聚散度外，它会导致眼睛朝相反的方向旋转。我们将跳过第七类叫做快速眼动 (REMs) 的动作，因为它们只发生在我们睡觉的时候，因此对 VR 体验没有贡献。现在将详细讨论六个类别。

### 快速扫视

眼睛可以以一种称为扫视的快速动作移动，该动作持续时间小于 45ms，每秒旋转约 90°。其目的是快速重新定位中心凹，使场景中的重要特征以最高的视敏度感知。图 5.15 举了一个例子，其中通过快速连续地固定各种特征来扫描脸部。功能之间的每次转换均通过扫视完成。有趣的是，我们的大脑使用扫视掩蔽来隐藏从我们的记忆中出现扫视的时间间隔。这会导致时间感知扭曲，就像第二只手在模拟时钟上点击位置时一样。扫视的结果是我们在大角度范围内获得了高敏锐度的幻觉。尽管扫视常常发生，但我们很少或根本没有意识到，但我们有能力自觉地控制它们。

### 平稳移动

在追求平稳的情况下，眼睛缓慢旋转以追踪移动的目标特征。例如一辆汽车，一个网球或者一个走过的人。旋转速率通常小于每秒 30°，这比快速扫视慢得多。平稳移动的主要功能是减少视网膜上的运动模糊，这也被称为图像稳定。如第 5.1 节所述，模糊是由于光感受器响应时间缓慢造成的。如果目标移动速度太快，那么可能会间歇性地将扫视器插入追击动作中以追上目标。

## 前庭眼反射

VR 最重要的动作之一是前庭眼反射或 VOR。将手指放在脸前舒适的距离并注视。接下来，来回摇头，每次左右转动 20 或 30 度左右。您可能会注意到，您的眼睛毫不费力地旋转以抵消头部的旋转，以便您的手指保持在视野内。眼动是不由自主的。如果你不相信，那么尽量避免旋转你的眼睛，同时注意你的手指和旋转头部。它被称为反射，因为运动控制绕过了更高的大脑功能。图 5.19 显示了这个回路是如何工作的。根据前庭器官感知的角加速度，信号被发送到眼肌以提供适当的反向运动。VOR 的主要目的是提供图像稳定功能，就像平稳移动的情况一样。有关前庭器官的更多细节，请参阅第 8.2 节。

## 视动反射

下一个类别称为视动反射 (optokinetic reflex)，在快速物体快速移动时发生。这种情况发生在观看快速行驶的列车时站在附近的固定地面上。眼睛快速和不自主地选择跟踪物体的特征，而在平滑移动和扫视运动之间交替。

## 聚散

立体视觉是指双眼被固定在相同物体上的情况，导致单个感知图像。出现两种聚散运动来将眼睛与物体对齐。见图 5.20。如果物体比先前的固定更近，则会发生收敛运动。这意味着眼睛在旋转，这样瞳孔就越来越近了。如果物体更远，则发生发散运动，这导致瞳孔进一步分开。聚散运动产生的方向提供了有关物体距离的重要信息。

## 微跳动

第六类运动被称为微跳动，这是微小的，不到一度的不自主的抽动，追寻不规则的路径。它们被认为可以增强许多其他过程，包括控制注视，减少因适应引起的感知褪色，改善视力以及解决感知模糊[269]。尽管这些运动自 18 世纪以来就已知晓[53]，但它们的行为非常复杂，并且没有完全理解。微跳动是知觉心理学，生物学和神经科学研究的一个活跃话题。

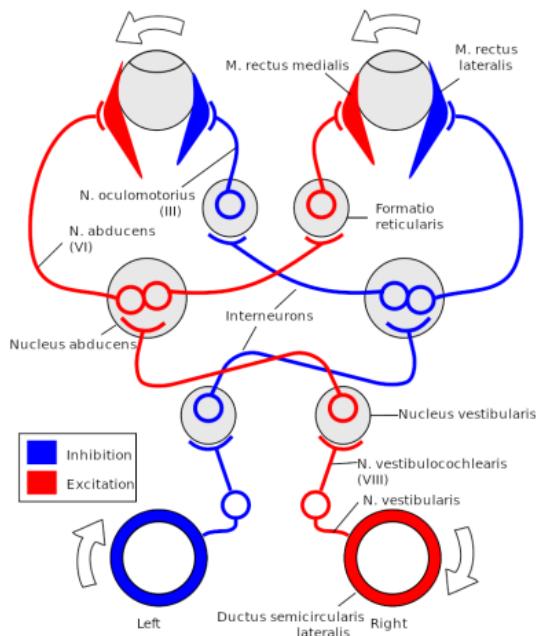


图 5.19：前庭眼反射 (VOR)。眼球肌肉与前庭器官内的角加速度计相连，以反向的眼球运动来抵抗头部运动，延迟小于 10 毫秒。眼睛和前庭器官之间的连接由专门的前庭和眼外运动核提供，从而绕过更高的脑功能。

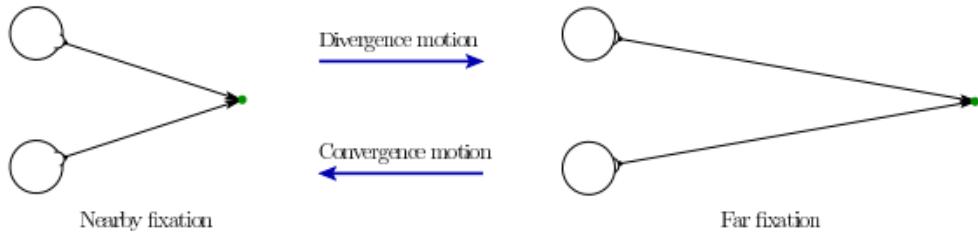


图 5.20：在立体视觉的过程中，双眼都凝视着世界上的同一个特征。要从近距离过渡到远距离特征，会发生背离运动。收敛运动则反之。

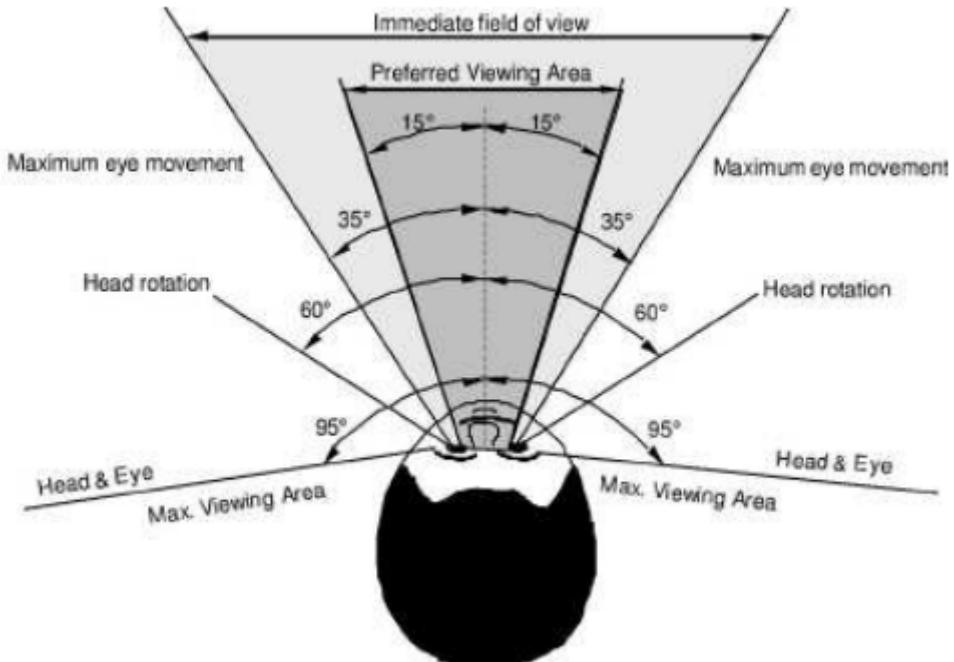


图 5.21：头部和眼睛一起旋转以固定新的或移动的目标。

### 眼睛和头部运动结合

虽然本节主要关注眼球运动，但重要的是要了解大部分时间眼睛和头部在一起移动。图 5.21 显示了头部和眼睛偏航旋转的角度范围。尽管左眼或右眼  $35^\circ$  左右对称，但双眼倾斜时却不是。人眼可以俯仰  $20^\circ$  和俯仰  $25^\circ$ ，这表明当眼睛直视前方时，将 VR 显示器置于瞳孔下方的中心位置可能是最佳选择。在 VOR 的情况下，控制眼球旋转以抵消头部运动。在平稳移动的情况下，头部和眼睛可以一起移动以将移动目标保持在首选观看区域中。

## 5.4 对 VR 的影响

到目前为止，本章介绍了人类视觉的“硬件”。诸如光感受器密度或 VOR 回路的基本生理特性直接影响视觉显示的要求。视觉系统必须足够好才能足以欺骗我们的感官，但它们不需要超出我们接收器极限的质量水平。因此，VR 显示器理想情况下应该设计成完美匹配它试图欺骗的表现。

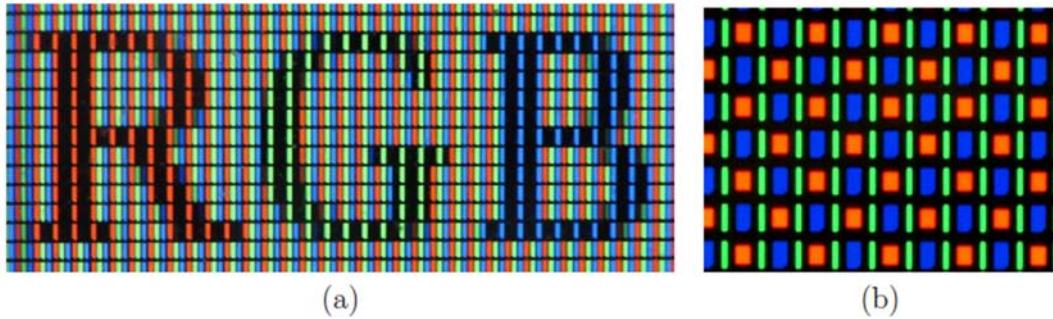


图 5.22：显示过程中，像素以与光感受器分解成红、蓝，绿色成分相同的方式分解成子像素。 (a) 一个 LCD 显示。（图片来自 Luis Flavio Loureiro dos Santos）(b) 来自 Nexus One 智能手机显示的一个 AMOLED PenTile。（图片来自 Matthew Rollings）

## VR 视觉显示器性能需要有多好？

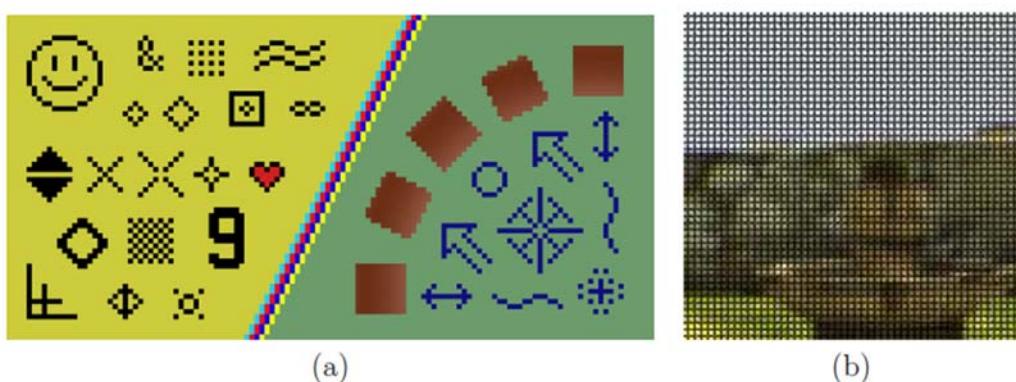
三个关键因素是：

1. **空间分辨率**: 每平方面积需要多少像素？
2. **强度分辨率和范围**: 可以生成多少个强度值，最小和最大强度值是什么？
3. **时间分辨率**: 显示器需要多快才能改变其像素？

空间分辨率因素将在下一段中讨论。第二个因素也可以称为颜色分辨率和范围，因为每个红色，绿色或蓝色子像素的强度值都在颜色空间中产生点；参见章节 6.3。回想图 5.4 中触发光感受器的强度范围。感光器可以跨越 7 个数量级的光强度。然而，显示器中每种颜色只有 256 个强度等级来覆盖这个范围。使用目前的显示技术，进入暗视视觉模式似乎不可能，因为在极低的光照水平下，需要高强度的分辨率。时间分辨率是非常重要的，但会推迟到第 6.2 节，在运动知觉的背景下介绍。

## 多少像素密度是足够的？

我们现在来解决空间分辨率的问题。对所需空间分辨率的洞察是通过感光器密度获取的。如图 5.22 所示，当显示器高度放大时，我们会看到各个灯光。当它被缩小时，我们仍然可能将锐利的对角线视为锯齿状，如图 5.23 (a) 所示；这种现象被称为走样。另一个神器是纱窗效应，如图 5.23 (b) 所示；这通常在由数字 LCD 投影仪产生的图像中被注意到。显示像素密度需要是多少才能使我们不会感觉到单个像素？2010 年，乔布斯的苹果公司宣称，每英寸（PPI）326 像素就足够了，实现了他们所称的视网膜显示<sup>1</sup>。这是否合理？它与 VR 又有什么关系？



<sup>1</sup> 密度相当于每平方毫米 165 个像素，但是我们会使用线性英尺因为它是显示比较方面的国际标准。

图5.23：(a) 由于像素的原因，我们获得了锯齿状而不是尖锐的直线的坏结果（称为走样更为正式）。  
 (图来自维基百科用户Jmf145) (b) 在纱窗效应中，在像素周围可见黑色网格。

假设中心凹陷处直接被指向显示器以提供最佳效果感知。第一个问题是红色，绿色和蓝色圆锥体被排列一个镶嵌图上，如图 5.24 所示。这个模式比在图 5.22 中的工程版本更不稳定。视觉科学家和神经生物学家已经研究过通过视力测量的有效或感知输入分辨率[139]。通常会要求研究中的受试者表明他们是否可以检测或识别一个特定的目标。在检测的情况下，例如，科学家可能会想了解打印到表面时可以感知的最小点。在显示方面，类似的问题是：像素需要多小才能使黑色背景上的单个白色像素无法被检测？在识别的情况下，一个熟悉的例子是试图阅读一个显示各种大小，任意字母的视力表。就显示而言，这可能相对应尝试阅读不同大小，分辨率和字体的文本。许多因素有助于敏锐度任务，比如亮度，对比度，眼球运动，时间曝光，以及被刺激的视网膜部分。

最广泛使用的概念之一是每度的周期数，大致可以看作是沿观看弧线分离的条纹（或正弦峰）的数量；见图 5.25。被验光师广泛使用斯内伦视力表，被设计以便患者从 20 英尺（或 6 米）之外识别印刷的字母。一个拥有“正常”20/20（或 6/6 英寸）视力的人几乎不会在字母“E”中形成横条纹，如图 5.25 所示。这是假设他直接看着这些字母，用的是在中央凹陷处的光感受器。图表上的 20/20 线就是设计成，当眼睛距离视力表 20 英尺时，字母高度对应于每度 30 个周期。

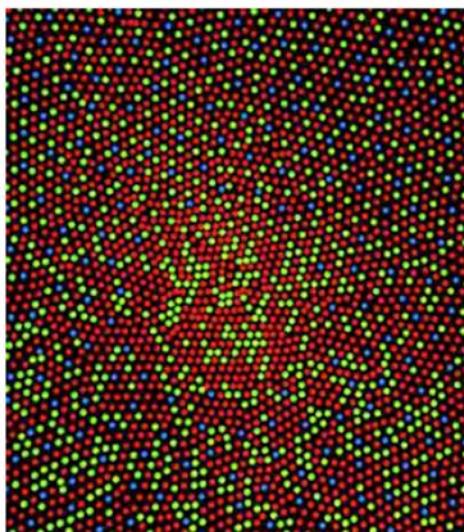


图 5.24：红色，绿色和蓝色光锥感光器分布在中心凹处复杂的马赛克中。 (图来自 Mark Fairchild)

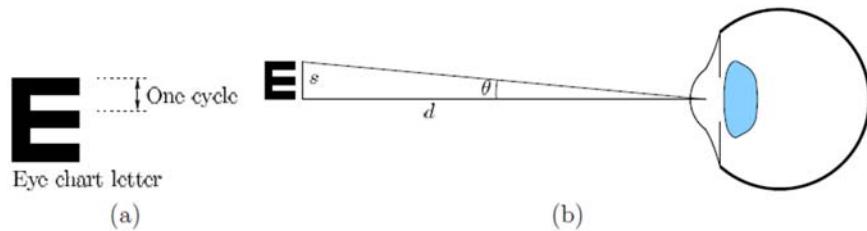


图 5.25：(a) 视力表上的单个字母。(b) 字母（或其他感兴趣特征）的大小  $s$ ，观看者的距离  $d$  和视角  $\theta$  以  $s = d \tan \theta$  联系。

“E”的总高度是 1/6 度。请注意，每个条纹是一半周期。如果主体离视力表仅 10 英尺远，会发生什么？这些字母大概会变为两倍大。

使用简单的三角函数，

$$s = d \tan \theta, \quad (5.1)$$

我们可以确定对于一个离眼睛距离为  $d$  的观看视角  $\theta$ ，某个特征的大小  $s$  应该是多少。对于非常小的  $\theta$ ， $\tan \theta \approx \theta$ （以弧度表示）。对于视力表的例子， $s$  可以对应一个字母的高度。加倍距离  $d$  和尺寸  $s$  应该保持  $\theta$  大致固定，这对应于视网膜上图像的大小。

我们现在回到视网膜显示的概念。假设一个拥有 20/20 视力的人正在观看距离它 20 英尺（6.096 米）的大屏幕。为了产生每度 30 个周期，每度必须至少有 60 个像素。使用 (5.1)， $s = 20 * \tan 1^\circ = 0.349\text{ft}$ ，相当于 4.189in。因此， $60 / 4.189 = 14.32\text{PPI}$  就足够了。现在假设智能手机屏幕距使用者眼睛 12 英寸。在这种情况下， $s = 12 * \tan 1^\circ = 0.209\text{in}$ 。这个要求屏幕至少有  $60 / 0.209 = 286.4\text{ PPI}$ ，它由 Apple 最初声称的 326 PPI 所实现。

在 VR 的情况下，用户不像在智能手机情况下直接看屏幕。通过插入放大镜头，显示器可以更接近眼睛。这通常由 VR 头显设备完成，如图 4.30 所示。假设镜头位于远离屏幕的焦距处，为了举例，它只有 1.5 英寸（这与当前 VR 头显设备是相当的）。在这种情况下， $s = 1 * \tan 1^\circ = 0.0261\text{in}$ ，并且显示器必须至少有 2291.6 PPI 以实现每度 60 个周期！如今密度最高的智能手机显示屏是三星 S6 的 Super AMOLED 1440x2560 5.1 英寸屏幕，它用于 Gear VR 系统。它只有 577 PPI，其中意味着 PPI 需要增加大约四倍才能获得 VR 头显设备的视网膜显示分辨率。

这不是完整的故事，因为有些人，特别是年轻人，有好于 20/20 的视力。基于光感受器密度和神经元过程[38,51]，视敏度的界限已经确定为每度大约 60 到 77 个周期；然而，这是基于将激光直接照射到视网膜上，当光线通过眼睛时会绕过许多光学像差问题。少数人（可能有百分之一）的敏锐度高达每度 60 个周期。在这种极端情况下，显示密度需要达到 4583 PPI。因此，确定足够的 VR 分辨率涉及许多因素。可以说如今在消费类 VR 头显设备中的分辨率是不足的，直到 PPI 增长至几倍高时视网膜显示分辨率才会实现。

## 多少视野是足够的？

如果屏幕更接近眼睛，填补更多的视野会怎么样？基于图 5.5 的感光体密度图和图 5.21 所示的眼旋转限制中，最大视场似乎在  $270^\circ$  左右，这比平面屏幕（小于  $180^\circ$ ）可以提供的更大。通过拉近屏幕来增大视野需要更高的像素密度，但周边的镜头像差（第 4.3 节）可能会限制有效的视野。此外，如果镜头太厚且太靠近眼睛，那么睫毛可能会刮到它；菲涅耳透镜可能提供一个薄的选择，但是会引入缺陷。因此，寻求 VR 视网膜显示器可能会以光学系统质量和人眼限制之间的平衡结束。弯曲的屏幕可能有助于缓解一些问题。

## 动画渲染

这个分析的一个挫折是，我们无法利用光感受器密度远离中心凹处减少的事实。我们必须保持像素密度高因为我们无法控制用户将看到的显示器的哪一部分。如果我们可以跟踪眼睛正在看的位置，并有一个始终定位在瞳孔前方的小型的可移动零延迟显示屏，那么我们需要的像素就会变少。这将大大减少图形渲染系统的计算负担（第 7 章介绍）。这个过程可以通过保持固定的显示但只对眼睛正在聚焦地方进行图形渲染而不是移动一个小屏幕来进行模拟，这被称为 foveated 渲染，它在工作[105]有所讨论，但目前成本太高，而且延迟太大以及眼球运动和显示更新之间的其他差异。在不久的将来，它可能成为大众市场的有效途径。

## VOR 增益适应

VOR 增益是比较眼睛旋转的比率（分子）来抵消头部的旋转和平移率（分母）。由于头部运动有六个自由度，因此将增益分成六个部分是合适的。在头部俯仰和左右摇晃的情况下，VOR 增益接近 1.0。例如，如果你以每秒 10° 的速度将头向左摇，那么眼睛以每秒 10° 的速度向相反方向偏转。VOR 滚动增益是非常小的，因为眼睛的滚动范围很小。VOR 平移增益取决于到特征的距离。

回顾第 2.3 节，适应是我们感官系统的一个普遍特征。VOR 增益也不例外。对于那些戴眼镜的人来说，VOR 增益必须根据第 4.2 节所述的光学变换有所适应。镜头影响视野并且感知物体的大小和距离。VOR 通过改变增益来轻松地适应这个问题。现在假设你戴上可能遭受缺陷的 VR 头显设备，诸如光学系统不完善，跟踪延迟以及屏幕上对象的错误渲染。在这种情况下，随着大脑试图调整其对平稳性的看法来弥补缺陷，适应可能会发生。在这种情况下，你的视觉系统可以说服你的大脑，设备运行正常，然后你的平稳感觉在现实世界中会变得扭曲，直到你重新适应。例如，在虚幻的虚拟现实体验之后，你可能会在现实世界中摇头并有一种静止物体在前后滑动的感觉<sup>2</sup>。

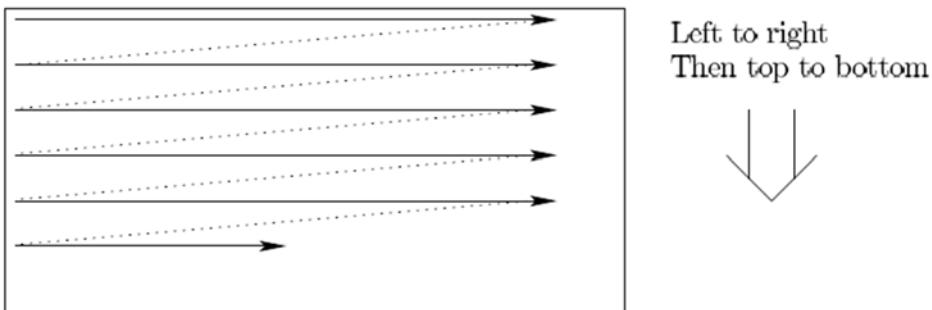


图 5.26：大多数显示器仍像旧电视机和 CRT 显示器一样工作：逐行更新像素。对于具有 60 FPS（每秒帧数）的显示器，这可能需要高达 16.67 毫秒。

## 显示扫描

回顾第 4.5 节，根据传感元件是逐行扫描还是并行扫描，摄像机具有滚动或全局快门。显示器以相同的方式工作，但相机是输入设备，而显示器是输出模拟。今天的大多数显示器都是滚动扫描（称为光栅扫描），而不是全局扫描。这意味着像素逐行更新，如图 5.26 所示。这个过程是旧电视机和监视器的一个缺陷，每个电视机都有一个阴极射线管（CRT），屏幕上布满了荧光体元件。电子束被电磁体弯曲，因此它会反复击打和刷新发光的荧光粉。

由于感光器的电荷和响应时间较慢，因此我们在正常使用过程中不会察觉到扫描图案。然而，当我们的眼睛，场景中的特征或两者都在移动时，滚动扫描的副作用可能会变得明显。考虑逐行打印机的操作，例如收银机上的收据打印机。如果我们在打印时拉扯录像带，那么这些线条就会变得分开。如果无法一次打印一条线，那么这些线本身会变得倾斜。如果我们可以打印时将磁带拉到一边，整个页面就会变得倾斜。你也可以通过用铅笔反复画一条水平线，同时用另一只手在特定方向轻轻拉纸来达到此效果。这个比喻在这篇文章中是视网膜和铅笔对应于试图充电感光器的光线。图 5.27 显示了矩形在平滑追踪和 VOR 情况下是如何变形的。一种可能性是通过渲染一个失真的图像来解决这个问题，这个图像将由逐行扫描产生的失真进行校正[212]（稍后在[1]中提出了这一点）。构建这些图像需要精确计算扫描时间。显示器的另一个问题是，像素可能花费很长时间才能切换（最长 20ms），以致锐利的边缘看起来模糊。我们将在第 6.2 节在运动观念的背景下，以及 7.4 节

渲染的背景下继续讨论这些问题。

<sup>2</sup>这在作者开发和测试 Oculus Rift 时经常发生。

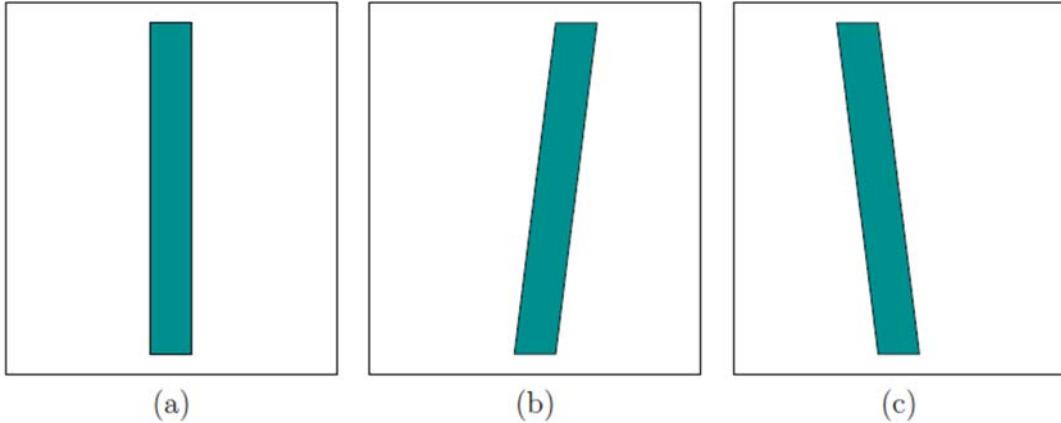


图 5.27：由于显示扫描造成的伪影：(a) 场景中的垂直矩形。(b) 矩形在虚拟世界中向右移动时，如何在平滑追踪期间扭曲。(c) 在使用 VOR 进行补偿时，如果将头旋转到右侧时固定矩形会如何变形。如果在每种情况下运动的方向相反，则 (b) 与 (c) 的情况交换。

### 视网膜图像滑动

回想一下，眼动既有助于维持视网膜上一个固定位置的目标（平滑追踪，VOR），也有助于略微改变其位置以减少感知褪色（微网络）。在日常活动（不是 VR）中，由于运动和光学失真，眼睛移动和特征的图像可能在视网膜上稍微移动。这被称为视网膜图像滑动。一旦使用 VR 头戴式设备，视网膜上的图像特征运动可能与现实世界中发生的不匹配。这是由于已经提到的许多因素，例如光学失真，跟踪延迟和显示扫描。因此，由 VR 伪影引起的视网膜图像滑动与在真实世界中遇到的视网膜图像滑动不匹配。这种情况的后果几乎没有被证实，科学性的定性更少。他们可能会导致疲劳，并可能导致 VR 不适。作为这个问题的一个例子，有证据表明，微网络是由缺乏视网膜图像滑动触发的[71]。这意味着由于 VR 使用导致的视网膜图像滑动的差异可能会干扰微网络运动，这些运动尚未被完全了解。

### 辐辏调节不匹配

回顾第 4.4 节，调节是改变眼睛镜片的光学能力的过程，以便可以将近距离物体聚焦。这通常发生在双眼固定在相同的物体上，导致立体视图被引入焦点。在现实世界中，眼睛的聚散运动和镜头的调节紧密耦合。例如，如果你将手指放在脸前 10 厘米处，则在眼睛强烈聚焦时，你的眼睛会尝试增加镜头的力量。如果将镜头放在离屏幕一定距离的焦距处，然后用正常的眼睛在眼睛放松时总是聚焦（回忆图 4.30）。如果一个物体被渲染到屏幕上让它看起来只有 10 厘米的距离，该怎么办？在这种情况下，眼睛强烈地聚焦，但他们不需要改变眼睛镜头的光焦度。尽管如此，眼睛可能会尝试适应，这会产生模糊感知图像的效果。结果被称为聚散度-调节不匹配，因为 VR 提供的刺激与现实世界不一致。即使眼睛习惯了不匹配，使用者在长时间使用后也可能会感到额外的紧张或疲劳[246,289]。眼睛基本上正在接受训练以获得新的自由度：将聚散度与调节分开，而不是将它们耦合起来。新的显示技术可能会缓解这个问题，但目前它们太昂贵且不精确。例如，通过使用眼动追踪来估计聚散度并随后改变光学系统的功率，可以大大减少失配[4,187]。

## 扩展阅读

5.1 节至 5.4 节的大部分概念都出现在有关感觉和感知的标准教科书中[97,204,350]。[204]的第 7 章包含的内容比本章中的更多。关于感光器结构的更多细节出现在[51,225,337]。眼睛和工程光学系统之间的接口在[296]中有介绍，其中数字光学系统也是相关的[156]。

[184]中提供了广泛的眼球运动。从神经科学的角度来看眼动，参见[177]。在[58,91,284]中研究了 VOR 增益适应。微网络函数的理论在[269]中讨论。[73]解释了平滑追踪和扫视之间的协调。头部和眼睛运动的协调在[162,247]中进行了研究。请参阅[17,246,289]关于性能不匹配的舒适性问题。

# 第六章 视觉感知

本章从第 5 章的结尾继续，介绍从人类视觉的生理学转向感知的问题。如果将我们比喻成电脑的话，那么这种转变看起来像是从低级硬件到更高级别的软件和算法。尽管我们的生物硬件有局限性，我们的大脑如何有效地解释我们周围的世界？要理解我们如何可能地被显示器呈现的视觉刺激所迷惑，你必须首先了解我们在正常情况下如何看待或解释真实世界。我们看到的内容并不总是那么清楚。我们已经看到了几种视错觉 VR 本身可以被认为是一个宏大的视错觉。它会在什么条件下成功或失败？

第 6.1 节涵盖了对我们眼睛的物体距离的感知这也与物体尺度的感知有关。第 6.2 节解释了我们如何感知运动。其中一个重要部分是我们从视频中看到的运动幻觉，视频仅仅是一系列的图片。第 6.3 节涵盖了颜色的感知，这可能有助于解释为什么显示器只使用三种颜色（红色，绿色和蓝色）来模拟光线的整个光谱功率分布（请参阅第 4.1 节）。最后，第 6.4 节介绍了一个基于统计的模型，说明如何将信息从多个来源组合到一起以产生感知体验。

## 6.1 感知深度

本节介绍人类如何使用视觉判断从他们的眼睛到现实世界中物体的距离。感知距离是可测量的，这意味着获得绝对距离的估计。例如，一座房子看起来可能距离大约 100 米。或者，距离信息可以是顺序，这意味着可以推断出可见物体的相对排列。例如，如果一间房屋部分遮挡了另一间房屋的视野，则该房屋似乎比另一间房屋更近。

### 单眼与立体线索

从感官刺激获得的与感知相关的信息称为感官线索或线索。在本节中，我们只考虑深度线索，这对深度感知有贡献。如果深度线索来自感光器或单眼的运动，那么它被称为单眼深度线索。如果需要双眼，那么它是一个立体深度线索。单眼深度信息比立体更多，这就解释了为什么我们能够从单张照片中推断出如此多的深度信息。图 6.1 显示了一个例子。此外，从图 6.2 中我们可以看出，即使是简单的线条图也足以提供强有力的线索。有趣的是，人类使用的线索也适用于计算机视觉算法，以从图像中提取深度信息[318]。



图 6.1：这幅画使用了一种称为纹理渐变的单眼深度线索来增强深度感知：随着深度的增加，砖块变得越来越小。其他线索来自透视投影，包括视野中的高度和视网膜图像大小。（“Paris Street, Rainy Day”，1877 年的古斯塔夫凯勒博特，芝加哥艺术学院。）

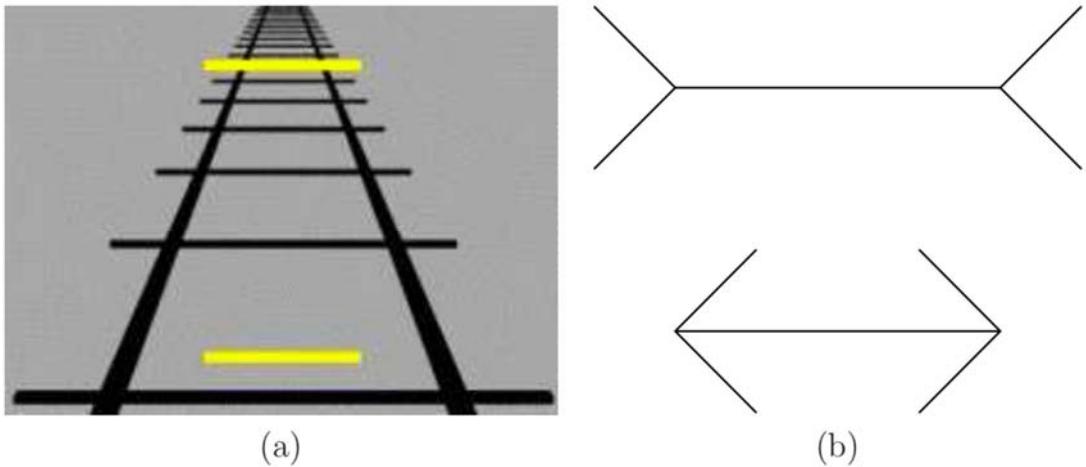


图 6.2: 即使是简单的线条图也能提供重要的线索。 (a) 蓬佐错觉: 上面的黄色条看起来更长, 但两者长度相同。 (b) 米勒莱尔错觉: 下面的水平段似乎比上面的短, 但它们的长度相同。

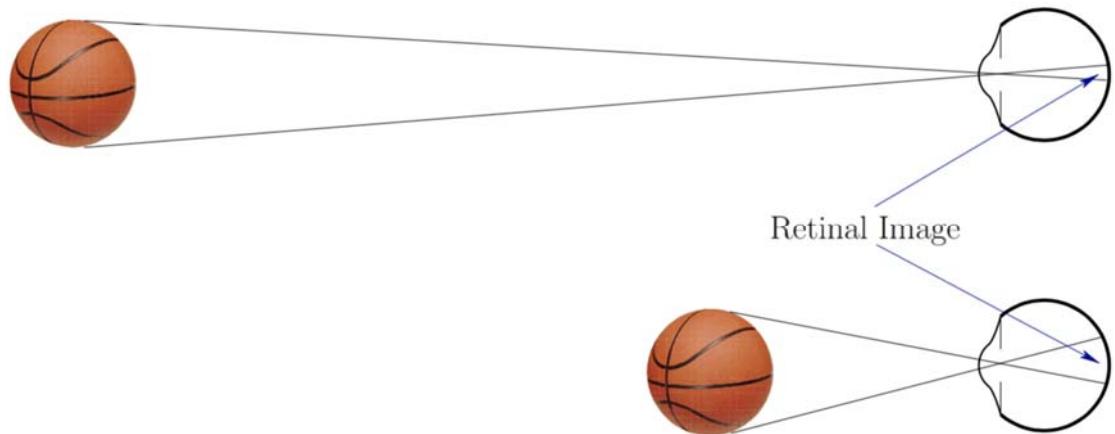


图 6.3: 一个熟悉物体的视网膜图像大小是一个强大的单眼深度线索。较近的物体投射到覆盖视网膜较大部分的大量感光器上。

### 6.1.1 单目深度线索

视网膜图像尺寸许多线索源于透视投影引起的几何失真; 回顾一下图 1.22(c)的“3D”外观。对于一个熟悉的物体, 比如人类, 硬币或篮球, 我们经常通过“有多大”来判断它的距离。回顾第 3.4 节的透视投影数学, 视网膜上图像的大小与  $1/z$  成正比, 其中  $z$  是与眼睛的距离 (或所有投影线的公共汇聚点)。见图 6.3。使用相机拍摄照片时也会发生同样的情况: 篮球照片会占据图像的较大部分, 覆盖更多像素, 因为它更接近相机。这种线索被称为视网膜图像大小, 并在 [96] 中进行了研究。

该情况存在两个重要因素。首先, 观众必须熟悉该物体, 以便知道其实际尺寸。对于熟悉的物体, 例如人或汽车, 当人走近, 我们的大脑通过假定物体的距离, 而不是大小, 来稳定地对大小进行比例缩放。尺寸常数落在主观恒定的总标题上, 这通过感知的许多方面出现, 包括形状, 大小和颜色。第二个因素是, 物体必须自然出现, 以免与其他深度线索发生冲突。

如果对象的大小存在很大的不确定性, 那么对它距离的了解应该有助于估计它的大小。

这受到尺寸感知的影响，这与深度感知密切相关。在第 6.4 节讨论的方式中，每个线索影响另一个。

一个有争议的理论是我们感知的视角与实际的视角不同。视角与视网膜图像大小成正比。这个理论被用来解释月球在接近地平线时似乎变得更大的错觉。如图 6.4。

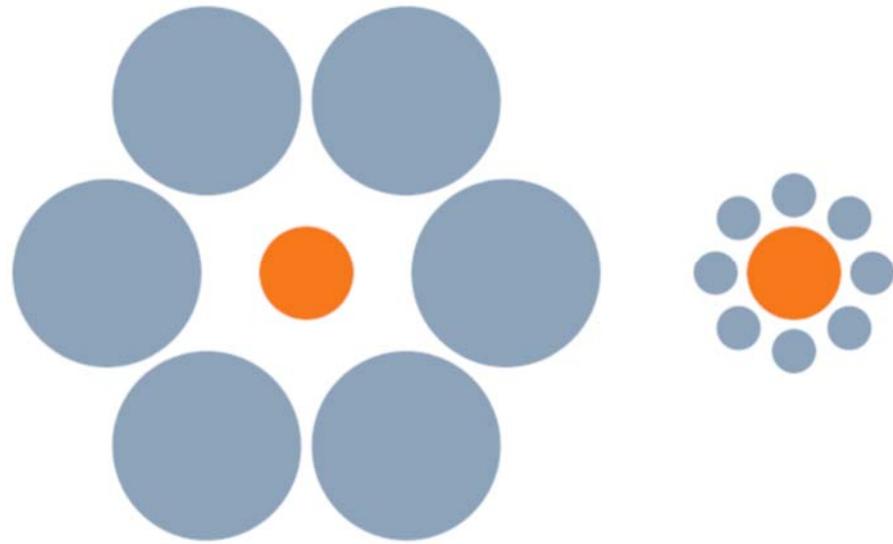
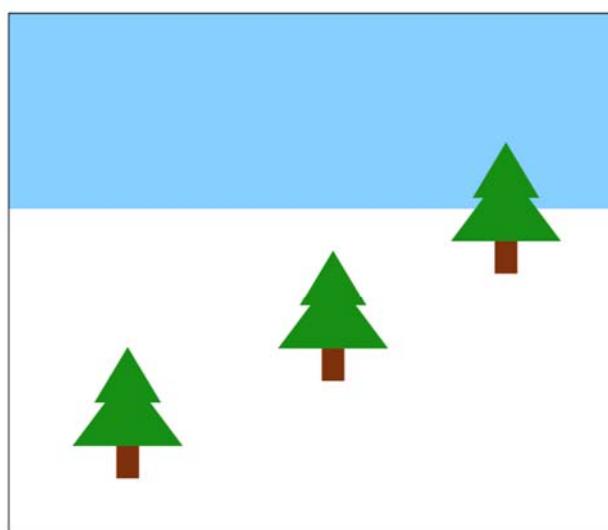


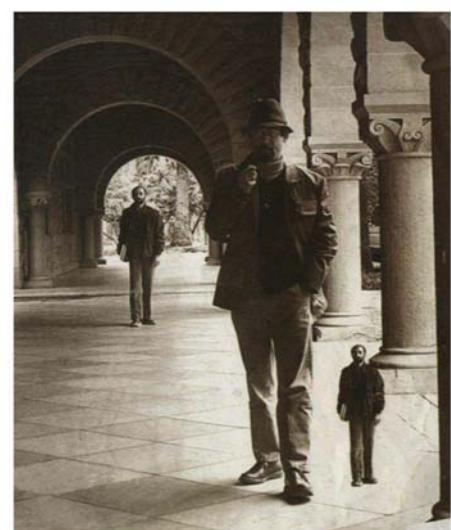
图 6.4：对于艾宾豪斯错觉，当小圆盘包围时，内圆盘看起来更大。无论哪种情况，内盘的尺寸都是相同的。这可能是真实视角（或视网膜图像大小）与感知视角之间差异的证据。

### 视野中的高度

图 6.5 (a) 展示了另一个重要的线索，它是视野中物体的高度。图 6.2 (a) 中的蓬佐错觉利用了这种线索。假设我们可以远距离看到没有障碍的地方。由于透视投影，地平线是将视野分成两半的线。上半部分被认为是天空，下半部分是地面。由于透视投影，物体与地平线之间的距离直接对应于它们的距离：越接近地平线，感知距离越远。如果可用，尺寸恒定标度与视野中的高度相结合，如图 6.5 (b) 所示。



(a)



(b)

图 6.5: 视野中的高度。 (a) 接近地平线的树木似乎更远，尽管所有树木都产生相同的视网膜图像尺寸。

(b) 人们在视野中不正确的位置说明了尺寸的恒定缩放，其与深度线索紧密结合。

## 适应

回顾第 4.4 节，人眼晶状体可通过适应过程改变其光焦度。对于年轻人来说，变化量大约为 10D (屈光度)，但对于 50 岁以上的成年人而言，变化量会减少到小于 1D。睫状肌控制晶状体，并通过电动机控制信号的复制副本将其张力水平报告给大脑。这是不依赖于光感受器产生的信号的第一个深度线索。

## 运动视差

直到现在，深度信号还没有利用到运动。如果你曾经从快速移动的车辆的侧窗看过去，你可能已经注意到附近的物体比其他物体的速度快得多。速度的相对差异称为视差，是一个重要的深度线索；见图 6.6。即使是在短时间内从不同角度观看的两幅图像，也可以提供强大的深度信息。想象一下，尝试模拟一台立体摄像机，我会拍摄一张照片，然后快速移动相机以拍摄另一张照片。如果世界其他地方是静止的，那么结果大致相当于有两个并排的摄像机。鸽子经常来回摆动头部，以获得比他们的眼睛提供的更深的深度信息。最后，与运动视差密切相关的是光流，这是表征特征在视网膜上移动的速率的表征。这些内容将在 6.2 节和 8.4 节中重新讨论。

## 其他单眼线索

图 6.7 显示了其他单眼线索。如图 6.7 (a) 所示，源投射的阴影遇到一个对象提供了一个重要的线索。图 6.7 (b) 显示了一个简单的图，它通过指示哪些对象位于其他对象之前提供了一个称为插入的序数深度线索。图 6.7 (c) 说明了图像模糊线索，其中根据不同的焦点清晰度推断出深度。图 6.7 (d) 显示了一个大气线索，其中空气湿度导致远处的景物具有较低的对比度，因此看起来更远。

### 6.1.2 立体深度线索

正如你所预料的那样，将双眼聚焦在同一个物体上可以提高深度感受。人类在太空中感知一个单一的聚焦图像，称为双眼视界；见图 6.8。回想一下 5.3 节的聚合运动。与容纳线索情况类似，对于聚散运动，眼部肌肉的运动控制向大脑提供关于会聚量的信息，从而提供对距离的直接估计。每只眼睛提供了不同的视点，这导致了视网膜上的不同图像。这种现象称为双目视差。回想 3.5 节 (3.50) 中的观点，将视点向右或向左移动，为每只眼睛提供横向偏移。这种转变本质上将虚拟世界转移到任何一方。现实世界中的并列相机立体装置也会发生同样的转变。然而，人类的双眼视差是不同的，因为除了具有横向偏移之外，眼睛可以旋转来进行聚集。因此，当注视一个物体时，左右眼之间的视网膜图像可能会略有不同，但这仍然提供了大脑使用的强大线索。

此外，当在一个深度上聚焦在一个物体上时，我们会感知其他深度的物体的双重图像(尽管我们通常不关注它)。

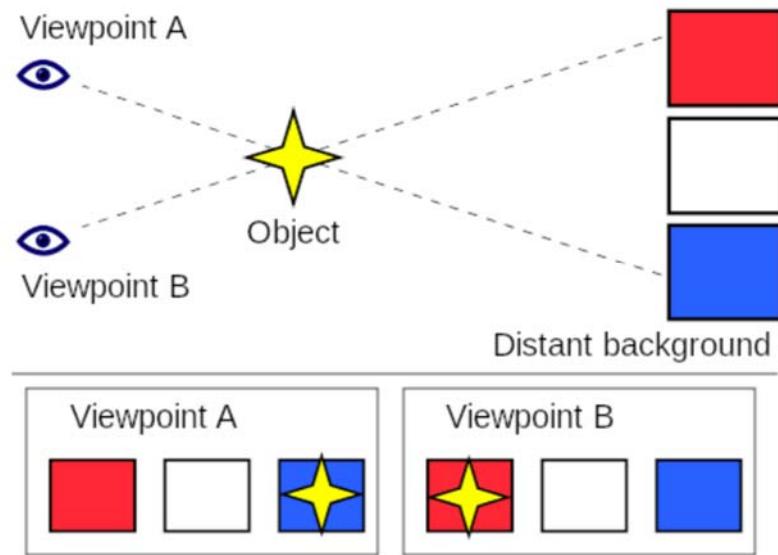


图 6.6: 运动视差: 当视角横向变化时, 较近的物体比其他物体具有更大的图像位移。(图来自维基百科。)

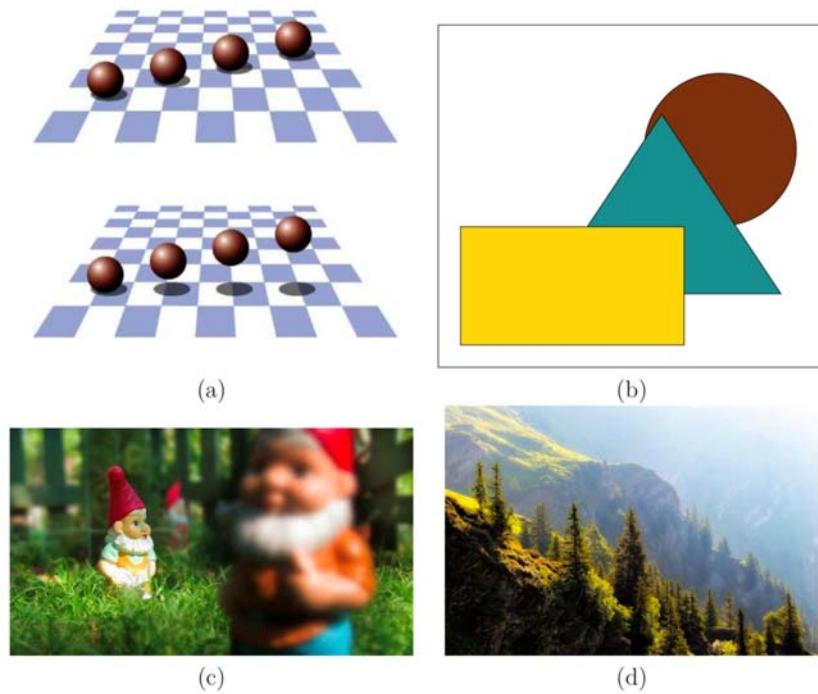


图 6.7: 更多的单眼深度线索: (a) 阴影解决球和幻影中不明确的深度。 (b) 物体的插入提供了一个有序的深度线索。 (c) 由于图像模糊, 一个侏儒似乎比其他人更近。 (d) 这个场景提供了一个大气线索: 由于对比度较低, 所以有些景观被认为距离较远。

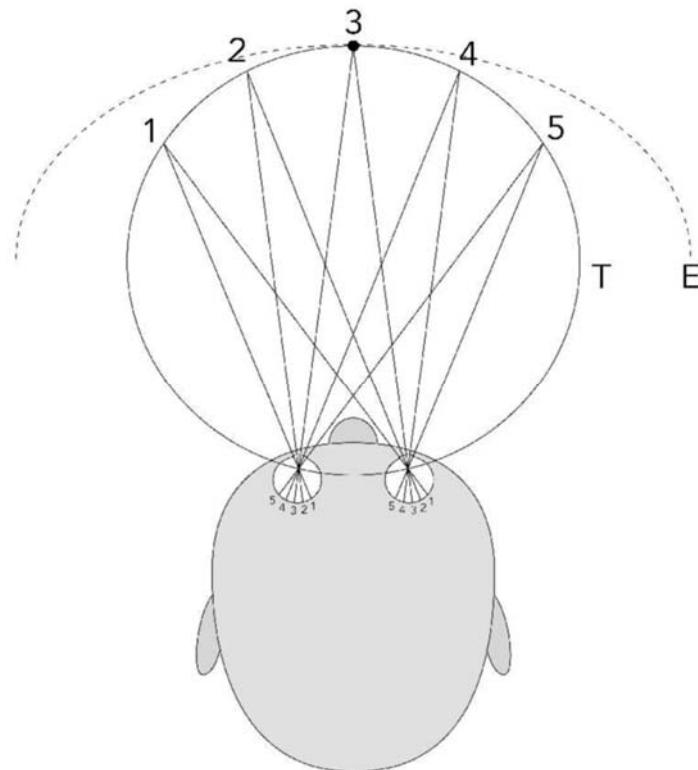


图 6.8：双眼视界是眼睛可以聚焦并聚焦在单个深度上的点的轨迹。T 曲线显示了基于简单几何理论的双眼直角坐标。E 曲线显示了经验性的双眼视界，其范围大得多并且对应于感知单个聚焦图像的区域。（图由 Rainer Zenz。）

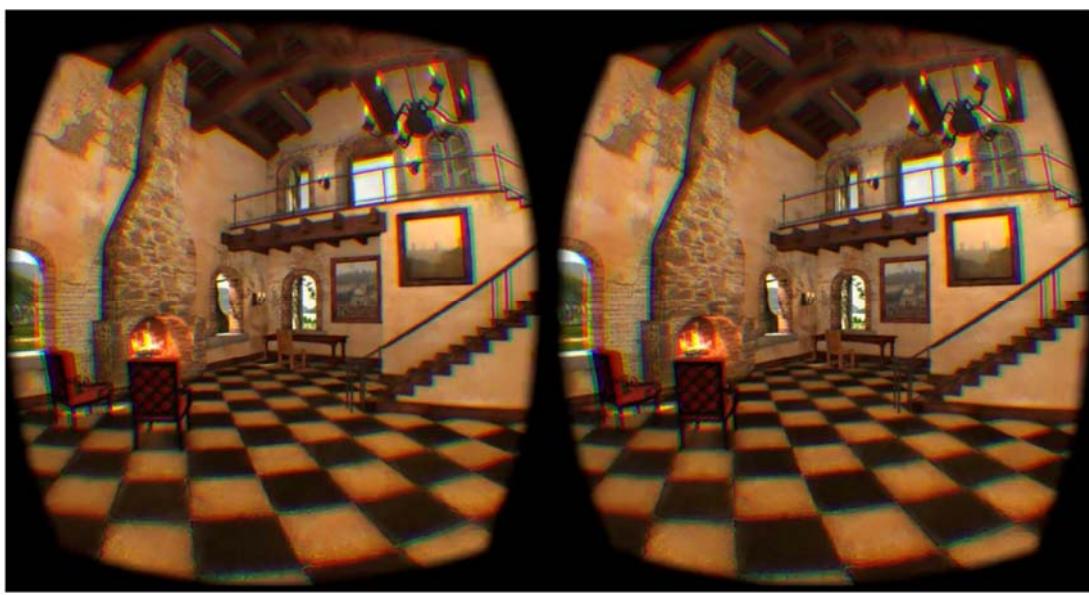


图 6.9：在 Oculus VR 的托斯卡纳演示中，没有足够熟悉的对象来精确解析深度和大小。你曾经去过这样的别墅吗？地砖是否是熟悉的尺寸？桌子是否太低了？

这种双重图像效果被称为复视。你可以通过将你的手指放在你的脸前约 20 厘米并将它聚拢来感知它。在注视你的手指时，你应该感觉到周围其他物体的双重图像。你也可以在保

持手指放在同一个地方的同时注视距离。你应该看到你的手指的双重形象。如果你额外地来回晃动头部，则应该看起来好像你的左右手指相对于彼此上下移动。这些对应于视网膜图像中的显着差异，但是我们通常感知不到，因为我们将两个视网膜图像视为单个图像。

### 6.1.3 对 VR 的影响

#### 不正确的比例感知

虚拟世界可能充满了我们在现实世界中不熟悉的对象。在许多情况下，它们可能与熟悉的物体相似，但其确切比例可能难以确定。考虑 Oculus VR 的托斯卡纳演示世界，如图 6.9 所示。虚拟别墅被设计为与人类居住，但由于没有足够的熟悉物体，很难判断物体的相对尺寸和距离。使问题进一步复杂化的原因是用户在 VR 中的身高可能不符合他在虚拟世界的身高。用户太小，还是世界太大？一个常见和令人困惑的事情是，用户可能坐在现实世界中，但站在虚拟世界中。如果瞳孔间距离（从第 4.4 节回忆）与现实世界不匹配，则会发生额外的并发症。例如，如果用户的瞳孔在现实世界中相距 64 毫米，但在虚拟世界中相距仅 50 毫米，则虚拟世界看起来会更大，这会显著影响深度感知。同样，如果瞳孔相距甚远，用户可能感觉到巨大或虚拟世界看起来很小。想象一下模拟哥斯拉体验，用户身高 200 米，整个城市似乎是一个模型。在 VR 中进行这种尺度和深度扭曲是很好的尝试，但了解它们对用户感知的影响很重要。

#### 失谐

在现实世界中，所有深度线索协调一致地工作。我们有时被视觉幻觉愚弄，这些幻觉旨在故意引起线索之间的不一致。有时候简单的绘图就足够了。图 6.10 显示了一个精巧的幻想，需要在现实世界中构建一个扭曲的房间。它的设计非常完美，因此当从一个位置的透视投影下观看时，它看起来是一个矩形框。一旦我们的大脑接受了这一点，我们意外地感觉到人们在房间里走动时人的大小会发生变化！这是因为所有基于透视的线索似乎都运作正常。6.4 节可以帮助你了解如何解决多个线索，即使在不一致的情况下也是如此。

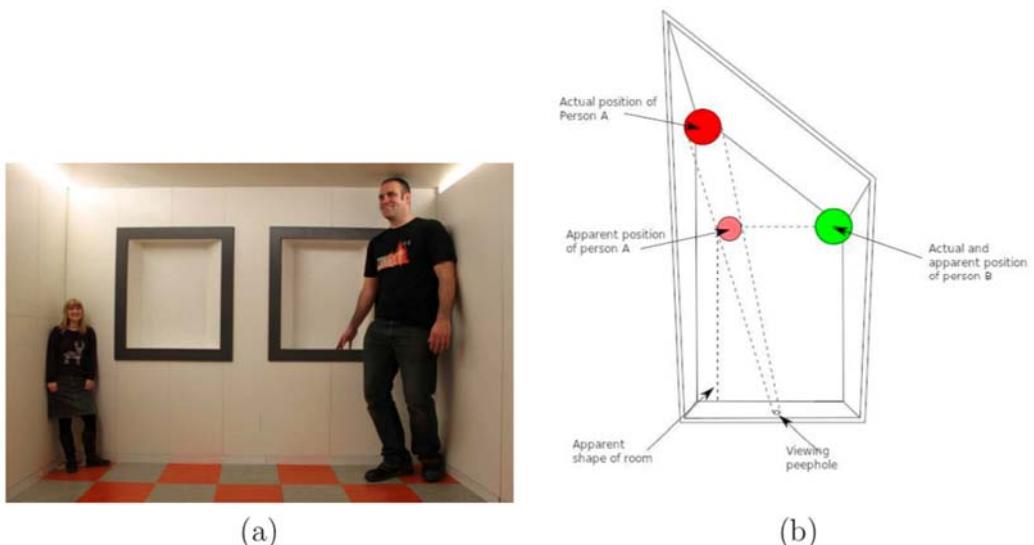


图 6.10：艾姆斯房间：(a) 由于深度线索不正确，导致尺寸检测结果不正确。(b) 房间被设计成在透视投影后只呈现矩形。一个人实际上远比另一个人远。（图由亚历克斯 Valavanis。）

在 VR 系统中，很容易造成不匹配，并且在许多情况下它们是不可避免的。回顾第 5.4 节，VR 头盔中会出现辐辏调节不匹配的情况。不完美的头部追踪可能会导致另一个失配的来源。如果存在显着的延迟，则视觉刺激将不会在预期的时间出现在正确的位置。此外，许多追踪系统只追踪头部方向。这使得如果用户没有任何旋转地从一侧移动到另一侧，则不可能使用运动视差作为深度线索。为了保留基于运动的大部分深度线索，除了定位之外，重要的是追踪头部位置；见 9.3 节。光学失真可能会导致更多失配。

### 单目线索非常强大！

普通公众常见的误解是：深度感知是由立体线索单独引起的。目前市场上的“3D”电影和立体显示器深受大家喜爱。最常见的例子是在电影院中使用圆偏振三维眼镜，这样当看着屏幕时，每只眼睛都会看到不同的图像。虚拟现实也正是利用了这种常见的错觉。CAVE 系统提供了内置主动式快门的 3D 眼镜，因此左右眼镜框之间可以交替显示。请注意，这会使得帧速率降低一半。既然我们拥有舒适的头戴显示器，为每只眼睛提供不同的视觉刺激就更简单了。一个缺点是渲染工作（第 7 章的主题）翻了一番，尽管这可以通过一些特定于上下文的技巧来改善。

正如你在本节中看到的，还有更多单眼深度线索比立体线索更多。因此，假设仅当存在立体图像时才将世界视为“3D”是错误的。这种见解对于利用来自现实世界的捕获数据特别有用。回想第 1.1 节，虚拟世界可能是合成的或被捕获的。创建合成世界通常成本更高，但生成立体视点非常简单（以更高的渲染成本）。另一方面，捕捉全景，单色图像和电影的速度快而且价格低廉（例子见图 1.8）。目前已有智能手机应用将图片拼接在一起制作全景照片，并且在几年内直接捕获全景视频很可能成为智能手机的标准功能。由于广泛的视野和单目深度线索，认识到这个内容足够“3D”，因此它成为创建 VR 体验的有力途径。Google Street View 中已经有数亿张图片，如图 6.11 所示，可以使用 Google Cardboard 或其他头戴设备轻松查看。即使没有立体显示，它们也提供了具有深度感的高度身临其境的体验。甚至有强有力的证据表明立体显示会造成显着的疲劳和不适，特别是对于深度较深的物体 [245, 246]。因此，应该慎重考虑使用立体显示。在许多情况下，当 VR 任务或体验可能已经有足够的单眼线索时，可能会花费更多的时间，成本和麻烦，而不值得去获得立体线索。

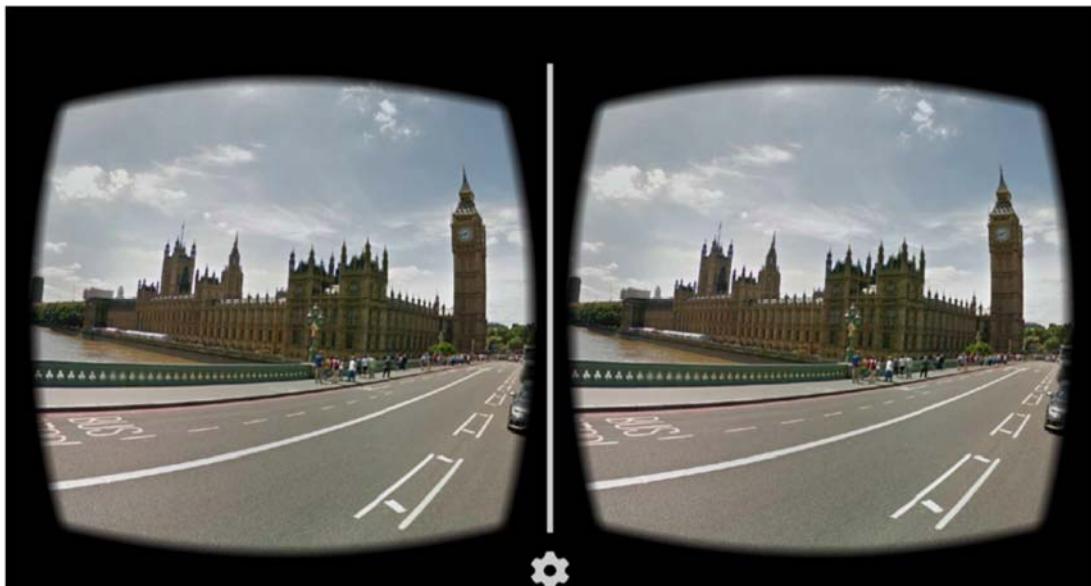


图 6.11：在 Google Cardboard 和其他 VR 头戴设备中，可以查看数亿个全景街景图像。由于单视觉深度线索，即使将相同的图像呈现给双眼，也具有显着的深度感知。

## 6.2 运动感知

我们往往依靠视力来感知运动，其中的一种情况是将移动的图形与静止的背景分开。例如，森林中的伪装动物只有在移动时才会变得明显。无论人类是猎杀者或被猎杀者，这种感知都是非常有用的。运动还可以帮助人们评估物体的三维结构。试想一下通过旋转并观察一种水果来评估其市场价值。运动感知的另一个用途是指导人类行动。尽管目前的技术有限，实际上 VR 系统在虚拟世界中对于复制这些用途的要求是很高的。与运动感知同样重要的是非运动的感知，我们在 2.3 节称之为平稳感知。例如，如果我们通过转动头部来应用 VOR，那么虚拟世界对象是否能在显示屏上正确移动以使得它们看起来是静止的？时间或图像位置中的轻微错误可能会无意中触发运动感知。

### 6.2.1 检测机制

#### Reichardt 探测器

图 6.12 显示了一个神经电路模型，称为 Reichardt 探测器，它响应人类视觉系统中的定向运动。神经节层和 LGN 中的神经元检测视网膜图像中不同点的简单特征。在较高的水平上，当特征从视网膜上的一个点移动到另一个附近点时，存在运动检测神经元。运动检测神经元激活一个特征速度，这取决于从其输入神经元的路径长度差异。它也基于输入神经元的感受区域的相对位置，对运动的特定方向敏感。由于运动检测器的简单性，它很容易被欺骗。图 6.12 显示了一个从左到右移动的功能。假设一列特征从右向左移动。根据这些特征的速度以及它们之间的间距，检测器可能会无意中产生作用，导致运动被感知到相反的方向。这是货车轮效应的基础，即根据速度，人眼看到的车轮可能会朝相反的方向旋转。这个过程可以通过引起眼睛共鸣而中断 [276]。这模拟了在第 6.2.2 节讨论过的频闪状况。此外，运动探测器具有适应性。因此存在一些幻觉，例如瀑布幻觉 [18] 和螺旋效应，其由于持续固定的后果而感知到不正确的运动 [18, 205]。

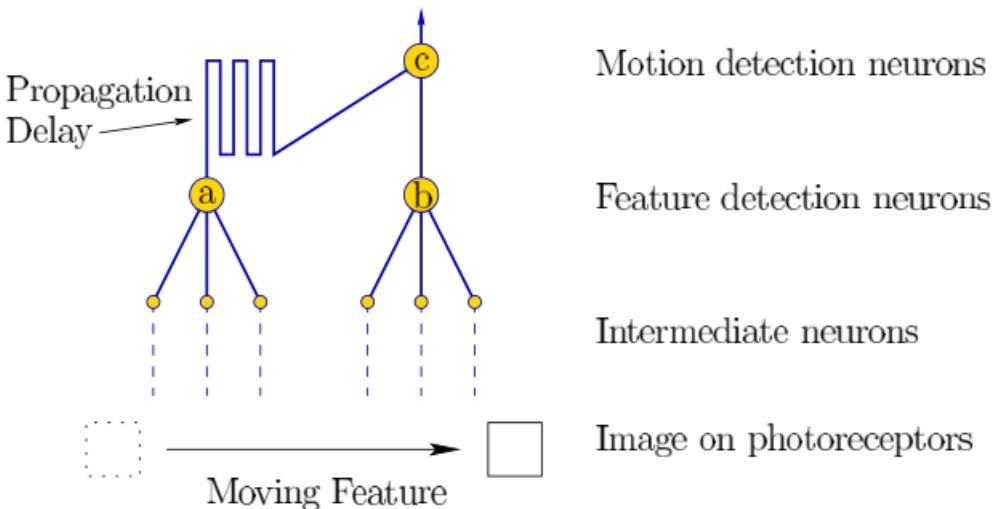


图 6.12：神经通路支持运动检测。随着图像特征穿过视网膜，附近的特征检测神经元（标记为 a 和 b）连续激活。它们的输出连接到运动检测神经元（标记为 c）。由于从 a 和 b 到 c 的路径长度不同，激活信号在不同的时间到达。因此，c 在 b 检测到特征时已经激活。

#### 局部数据到整体结论

运动探测器是局部的，视野的一小部分会激活各个探测器。在大多数情况下，检测整个视野的各探测器数据被整合，以指示刚体运动。根据第 3.2 节的方程式，一个刚体的所

有部分都在空间中移动。我们的视觉系统则预计这种协调一致的动作。在一些情况下，人类的视觉会产生孔径问题，如图 6.13 所示。描述整个视网膜全局运动的一种简洁数学方法是通过矢量场，即在每个位置处指定一个速度矢量。全局结果被称为光流，它为物体运动和自身运动提供了强大的线索。后一种情况会导致相对运动错觉，这是 VR 疾病的主要原因之一。详情请参阅第 8.4 节和第 10.2 节。

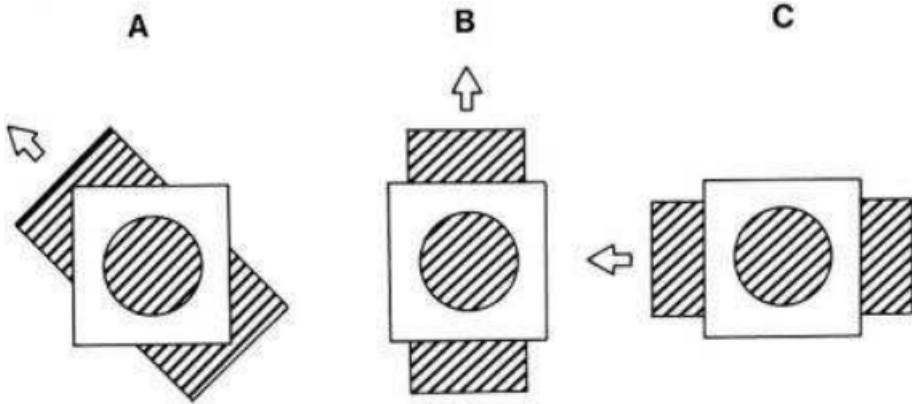


图 6.13：由于运动探测器的局部性质，会导致孔径问题。较大身体的运动在通过小孔感知时是模糊的，因为大量可能的运动在洞内会产生相同的效果，因而导致不正确的运动推断。

### 区分物体运动和观察者运动

图 6.14 显示了随时间推移在视网膜上产生相同图像的两种情况。在图 6.14 (a) 中，物体移动时眼睛被固定。在图 6.14 (b) 中，情况相反：对象是固定的，但眼睛在移动。大脑会使用几个线索来区分这些情况。第 5.3 节提到的 Saccadic 抑制在运动过程中隐藏了视觉信号，这可能抑制第二种情况下的运动检测器。另一个线索是由本体感觉提供的，这是身体由于运动命令而估计其自身运动的能力。这包括在第二种情况下使用眼部肌肉。最后，信息通过大范围运动得到。如果整个场景看起来都在移动，那么大脑就会假设最有可能的解释，那就是用户正在移动。这就是为什么如图 2.20 所示的秋千错觉非常有效。

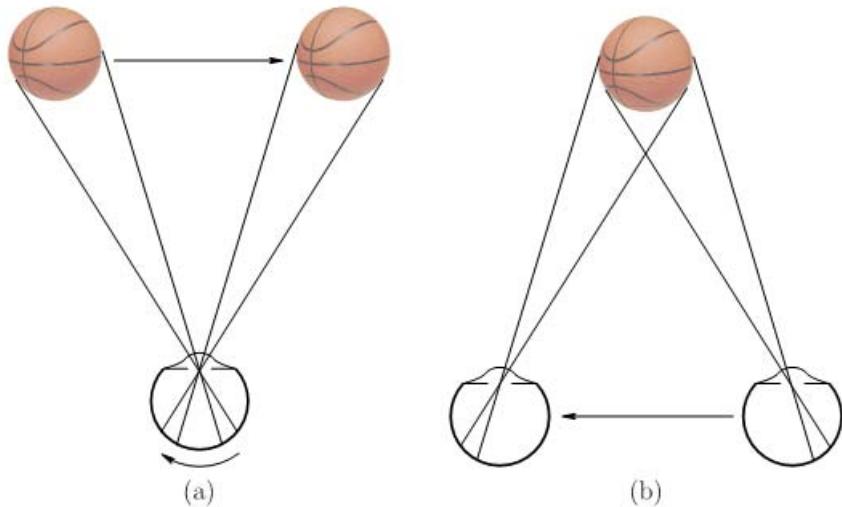


图 6.14：导致图像在视网膜上等效移动的两种运动：(a) 眼睛被固定并且物体移动；(b) 当物体固定时眼睛移动。

### 6.2.2 频闪视运动

无论是通过电视，智能手机还是电影屏幕，几乎地球上的每个人应该都看过电影。我们在电影中看到的运动是一种幻觉，因为其只是一系列静止图像放映于屏幕上。这种现象被称为频闪视运动，它在 19 世纪被发现和完善。如图 6.15 所示，西洋镜大约在 1834 年开发出来。它由一个带有狭缝的旋转鼓组成，当鼓旋转时，每个框架都可以看见一瞬间。在 1.3 节中，图 1.23 展示了 1878 年的马动画电影。

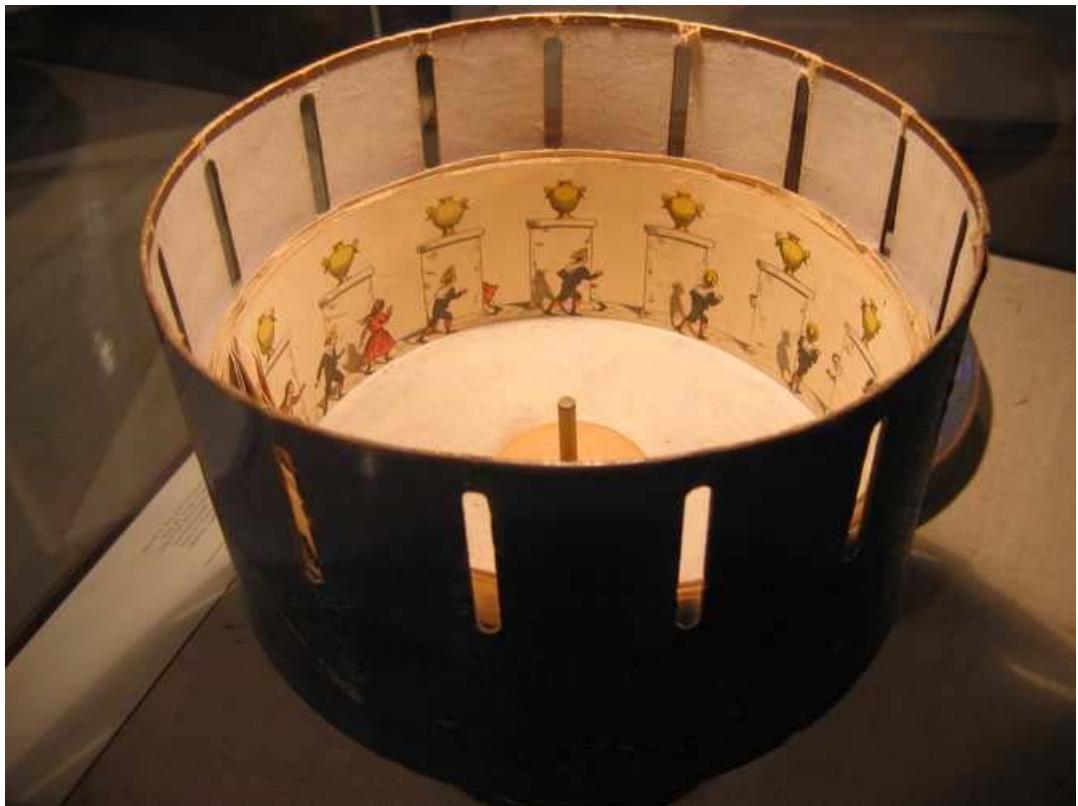


图 6.15：西洋镜在 1830 年代发展起来，并提供频闪视运动，通过旋转盘中的狭缝可以看到图像。

为什么这种运动幻觉可以起作用？近年来被驳斥的一个早期理论被称为视觉持久性理论。该理论指出，图像在帧之间的间隔期间持续存在于视觉系统中，从而导致它们看起来是连续的。反对这一理论的一个证据是，图像在视觉皮层中持续约 100ms，这意味着 10 FPS（每秒帧数）是频闪视运动的最慢速度。另一个反对视觉持续性的证据是存在无法用它来解释的频闪视运动。phi 现象和 beta 运动是闪烁灯运动感知的例子，而不是闪烁的帧（见图 6.16）。频闪视运动起作用的最可能原因是它触发了图 6.12 所示的神经运动检测电路 [204, 211]。

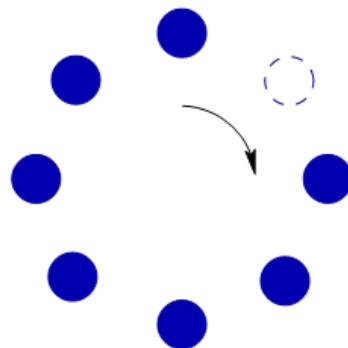


图 6.16: phi 现象和 beta 运动在生理学上是不同的, 可以感知运动[347, 307]。在点序列中, 任何时候都会关闭一个点。按照顺时针方式, 每帧中关闭一个不同的点。在非常低的速度下 (2 FPS), beta 运动会触发后一关闭点的运动感知。以更高的速度, 例如 15 FPS, 整体看上去就像是一个移动的洞, 这对应于 phi 现象。

## 帧率

每秒多少帧适用于电影? 答案取决于预期用途。图 6.17 显示了从 2 到 5000 的重要帧速率表。频闪视运动从 2 FPS 开始。想象一下以这样的速度观看视频, 很容易区分出各帧, 同时人的动作也可以感知。一旦达到 10 FPS, 运动会显得更加平滑, 我们开始失去区分单个帧的能力。早期无声电影的范围为 16 到 24 FPS。帧率经常波动, 并以比拍摄速度更快的速度播放。一旦将声音添加到胶片中, 就不再允许不正确的速度和速度波动, 因为声音和视频都需要同步。回放的固定速率为 24 FPS, 电影行业目前仍在使用它。到 20 世纪 70 年代, 个人摄像机保持在 16 或 18 FPS。著名的 1930 年肯尼迪遇刺的扎普鲁德电影以 18.3 FPS 拍摄。尽管 24 FPS 可能足以平滑地感知运动, 但由于大部分摄影致力于确保运动速度不会太快, 因而帧速率较低, 有时会看到跳跃现象。

FPS	Occurrence
2	Stroboscopic apparent motion starts
10	Ability to distinguish individual frames is lost
16	Old home movies; early silent films
24	Hollywood classic standard
25	PAL television before interlacing
30	NTSC television before interlacing
48	Two-blade shutter; proposed new Hollywood standard
50	Interlaced PAL television
60	Interlaced NTSC television; perceived flicker in some displays
72	Three-blade shutter; minimum CRT refresh rate for comfort
90	Modern VR headsets; no more discomfort from flicker
1000	Ability to see zipper effect for fast, blinking LED
5000	Cannot perceive zipper effect

图 6.17: 各种帧速率和相应的适用情况。单位是每秒帧数 (FPS)。

由于低帧率导致的闪烁现象, 几种解决方法应运而生。在电影放映机的情况下, 有人发明了两叶片和三叶片百叶窗, 以便它们分别显示两帧或三帧。这使得电影能够以 48 FPS 和 72 FPS 显示, 从而减少闪烁造成的不适。根据不同国家的标准, 20 世纪的模拟电视广播采用 25 (PAL 标准) 或 30 FPS (NTSC 标准)。为了加倍帧频并减少感知闪烁, 他们使用隔行扫描在一帧时间内画出一半图像, 然后在另一帧时间内画出一半图像。前一部分绘制一半的水平线, 其余线绘制在第二部分中。这将电视屏幕上的帧率提高到 50 和 60 FPS。游戏行业已经使用 60 FPS 的标准来实现流畅的游戏。

当人们在 20 世纪 90 年代初开始坐在大型 CRT 显示器前观看时, 闪烁问题再次出现。因为在周围对闪烁的敏感性更强。此外, 即使闪烁不能被直接察觉, 也可能导致疲劳或头痛。因此, 帧率增加到更高的水平。大型 CRT 显示器的最低可接受人体工程学标准为 72 FPS, 广泛认为 85 至 90 FPS 足够高以消除大部分闪烁问题。心理学家在闪烁融合阈值的主题下仔细研究了这个问题。闪烁可察觉或导致疲劳的程度取决于除 FPS 之外的许多因素, 例如视网膜上的位置, 年龄, 颜色和光强度。因此, 实际可以改善的方面在于显示器的种类, 尺寸, 规格, 使用方式以及使用方式。用作电视机, 电脑屏幕和智能手机屏幕的现代 LCD 和 LED 显示屏具有 60, 120 甚至 240 FPS。

故事并没有结束。如果将 LED 连接到脉冲发生器 (串联一个电阻), 那么可以以更高的

速率感知到闪烁。设置脉冲发生器产生几百赫兹的方波。然后去一个黑暗的房间，握住你的手中的 LED。如果你四处挥舞，以至于你的眼睛无法追踪它，那么闪烁就会变成拉链图案。这被称为拉链效应。发生这种情况是因为每次发光时，它都会在视网膜上的不同位置成像。如果没有图像稳定功能，它会显示为一组灯光。运动速度越快，图像形成位置就会越远。脉率（或 FPS）越高，图像越接近。因此，要以非常高的速度查看拉链效应，你需要快速移动 LED。有可能看到上千 FPS 的效果。

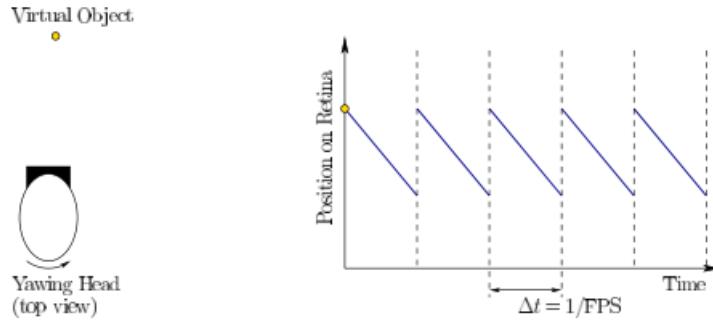


图 6.18：在频闪视运动下观察平稳性的问题：随着 VOR 的执行，特征的图像以重复模式滑过视网膜。

### 6.2.3 对 VR 的影响

不幸的是，VR 系统需要比平常更高的显示性能。我们在第 5.4 节已经看到，其需要更高的分辨率，以使得像素和混叠伪像不可见。因而与 24 FPS 甚至 60 FPS 的普通电视或电影标准相比，VR 系统需要更高的帧速率。要理解其原因，请参见图 6.18。这个问题最容易理解，就 2.3 节中提到的平稳感而言，如果人们注释一个附近物体，然后向左摇晃头部。由于 VOR，你的眼睛应该向右旋转以将物体保持在视网膜上的固定位置（5.3 节）。如果你在戴着虚拟现实头戴式显示器并注视虚拟世界中的物体时也这样做，则在转动头部时，物体的图像需要在屏幕上移动。假设像素在每个新帧时间点时瞬间变化，虚拟物体的图像将滑过视网膜，如图 6.18 所示。其结果是一种抖动，其中的物体会高频率但小幅度地左右晃动。

其问题在于每个特征在屏幕上固定时间太长，理想情况下应该在屏幕上连续移动。在 60 FPS 时，在每帧期间固定为 16.67 毫秒（理想化设置，忽略 5.4 节中的扫描问题）。如果屏幕仅在每帧的一或两毫秒内打开，然后在剩余时间内变为黑色，则视网膜图像滑动量将大大减少。这种显示模式是低持久性的，如图 6.19 (a) 所示。显示器被照亮的短时间足以使感光器收集足够多的光子以使图像被感知。问题是，在低余辉模式下 60 FPS 时，会感觉到闪烁，从而导致疲劳或头痛。在 Samsung Gear VR 耳机的明亮场景中，可以很容易地看到这一点。如果帧速率提高到 90 FPS 或更高，那么闪烁的副作用几乎可以消除。

如果帧频增加到 500 FPS 或更高，那么它甚至不需要闪烁，如图 6.19 (b) 所示。

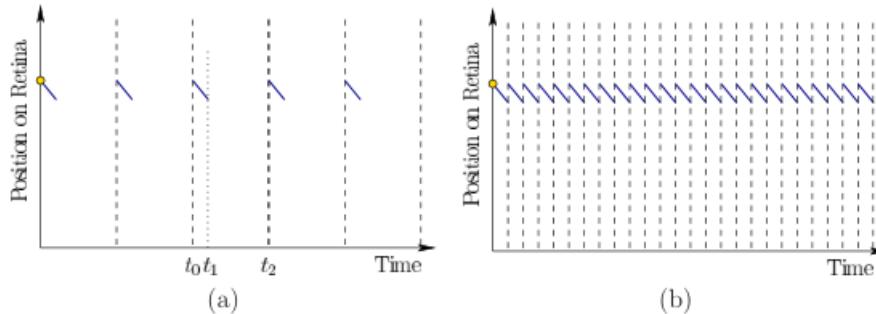


图6.19：减少视网膜图像滑动的工程解决方案：(a) 使用低余辉，显示器点亮足够短的时间以激活感光器( $t_1 - t_0$ )，然后在剩余时间( $t_2 - t_1$ )内产生空白。 $t_1 - t_0$ 通常为1到2毫秒。(b) 如果帧率非常大(至少500 FPS)，则不需要空白间隔。

最后一点是图 6.19 中隐含了像素切换速度。在现代 OLED 显示面板中，像素可以在小于 0.1ms 内达到其目标强度值。但是，许多 LCD 显示器以更慢的速度改变像素值。根据强度变化的数量和方向，达到目标强度的延迟可能长达 20ms。在这种情况下，一个固定的虚拟物体会在运动方向上变得模糊。这在使用 LCD 显示面板的 Oculus Rift DK1 中很容易观察到。

### 6.3 颜色感知

物体是如何呈现出“紫色”，“粉红色”或“灰色”？色觉纯粹是我们视觉生理学和神经结构的结果，而不是物理世界中可以测量的东西。换句话说，“一切都在你的脑海中”。如果两个人拥有相似的色彩感知系统，那么他们可以使用普遍认同的名称来讨论色彩，他们则认为物体具有相同的颜色。这与其他感知主题（如运动，深度和尺度）形成了对比，因为除了色觉的主题都与周围世界的可测量指标相对应。仪器可以确定物体的大小或相对于某个帧的运动速度。无论人们的个体感知系统如何运作，人们的感知结果将被迫与测量的数值结果达成一致。



图 6.20：2014 年，由于人们无法认同它是“蓝黑”还是“白金”，这件连衣裙照片成为互联网的一个轰动事件。

#### 连衣裙

图 6.20 说明了衣服的颜色感知错觉。由于数百万人争论关于这件衣服的颜色，其成为了一个互联网的重大事件之一。基于颜色和光照条件的组合，其外观落在人类色彩感知系统处理的边界上。大约 57% 的人认为它是蓝黑色（正确），30% 认为它是白金色，10% 认为蓝色和棕色，10% 可以在几种颜色组合之间切换[159]。

#### 降维

回想一下 4.1 节，光能是具有光谱能量分布的不同波长或波强度的集合。图 4.6 提供了一个例子。当我们看到物体时，根据光谱分布函数（图 4.7），环境中的光线根据波长以不同方式从表面反射回来。当光线通过人眼并聚焦到视网膜上时，每个光感受器都会接收

到包含许多波长的混合光能。由于功率分布是波长的函数，所有可能分布的集合就是一个函数空间，这个空间一般是无限维的。我们有限的硬件无法感知整个功能。棒状和锥状光感受器则以偏向某个目标波长的方式进行取样，如 5.1 节的图 5.3 所示。其结果在有关降维的研究中被总结为一个很好的理论。功率分布的无限维空间折叠成 3D 色彩空间。人眼恰好具有三种类型的锥体并非巧合，RGB 显示目标与感光体的颜色是相同的。

### 黄色=绿色+红色

为了帮助理解这种降维效果，让我们考虑“黄色”的感觉。根据可见光谱（图 4.5），黄色的波长约为 580nm。假设我们有一个纯粹的光源，它将 580nm 波长的光照射到我们的视网膜上，而没有其他波长。光谱分布函数在 580nm 处有一个尖峰，在其他地方为零。如果我们有一个在 580nm 处具有峰值检测并且对其他波长不敏感的圆锥体，那么它将完美地检测到黄色。相反，我们通过激活绿色和红色锥体感知黄色，因为它们的灵敏度区域（图 5.3）包括 580nm。其情况可以通过发射包含两种波长：1) 533nm 处的“绿色”光和 2) 564nm 处的“红色”光的混合光来产生相同的光感受器响应。如果调整绿色和红色光的大小，使得绿色和红色圆锥以与纯黄色相同的方式激活，那么我们的视觉系统就不可能将纯绿色/红色的混合光从纯黄色中分辨出来。两者都会被视为“黄色”。这种来自红色、绿色和蓝色组合的颜色匹配被称为同色异谱。这种混合正是在 RGB 显示器上完成的，以产生黄色效果。假设每种颜色的强度范围从 0（暗）到 255（亮）。红色由  $\text{RGB} = (255, 0, 0)$  产生，绿色为  $\text{RGB} = (0, 255, 0)$ 。这种显示方式通过激活每一个 LED（或 LCD）的颜色，从而产生纯红色或绿色。如果两者都打开，则会感知到黄色。因此，黄色为  $\text{RGB} = (255, 255, 0)$ 。

### 色彩空间

为了方便起见，常常定义参数化的色彩空间。计算机图形学中最常见的一种色彩空间称为 HSV，它具有以下三个元素（图 6.21）：

- **色调：**直接对应于感知的颜色，如“红色”或“绿色”。
- **饱和度：**即颜色的纯度。换句话说，除了色调的波长之外，多少能量来自波长？
- **值：**与亮度相对应。

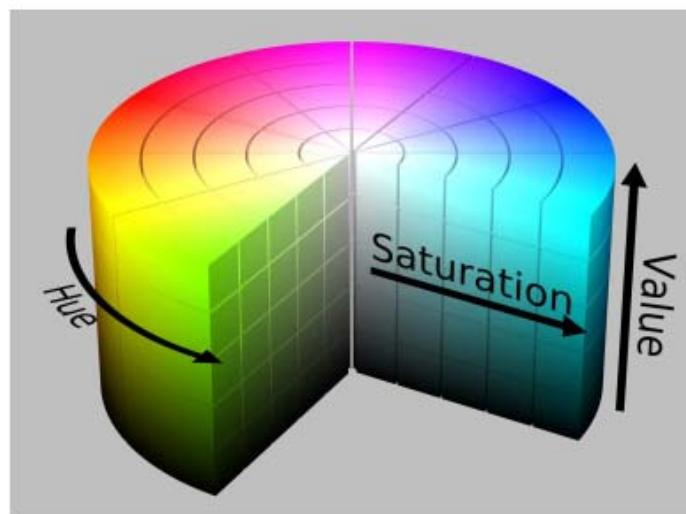


图 6.21：HSV 色彩空间的一种表示，它涉及三个参数：色调，饱和度和值（亮度）。（由维基百科用户 SharkD 提供）

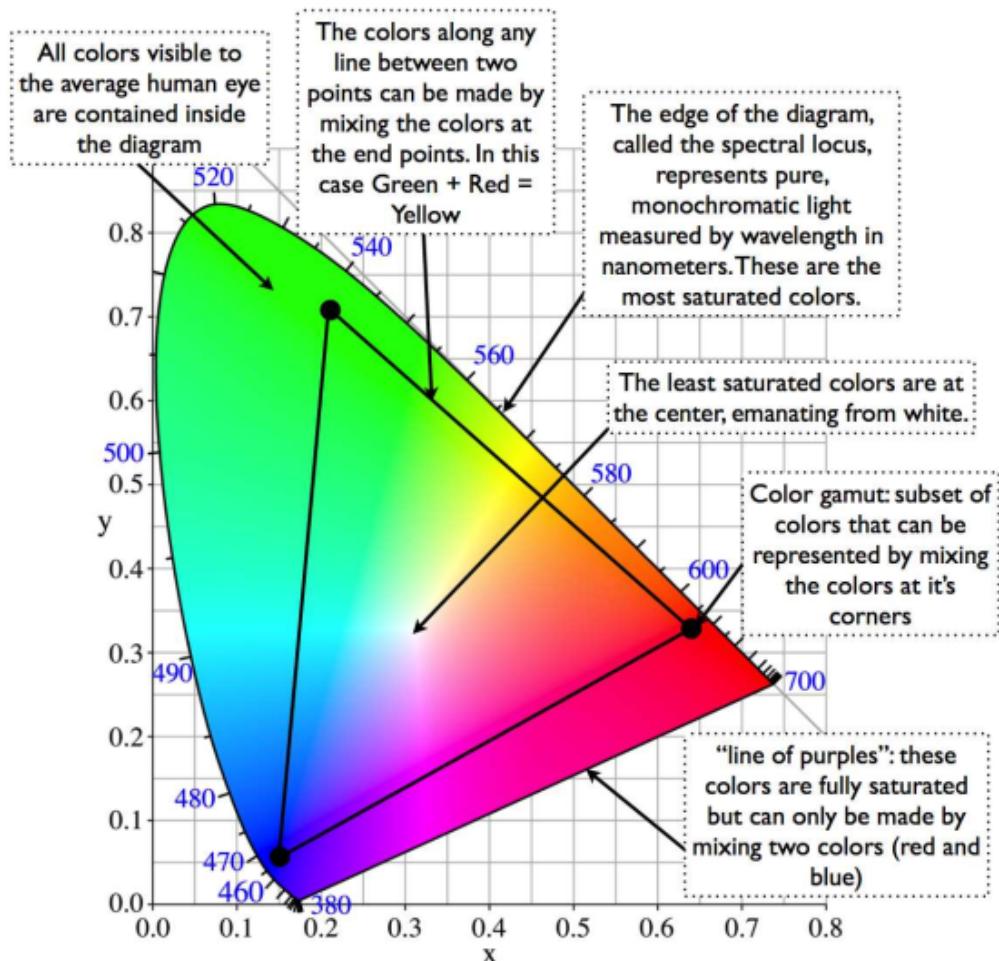
有很多方法可以缩放 HSV 坐标，这将以各种方式扭曲色彩空间。可以选择使用 RGB 值，

不过这导致有时候人们更难以理解。

一种理想的表示差异量感知的特性是两点之间的距离。换句话说，随着两点距离越来越远，我们区分它们的能力就会增加。距离应该直接对应于可区分的数量。视觉科学家们设计了一种表现形式来实现这一点，从而产生了如图 6.22 所示的 1931 年 CIE 色彩标准。因此，从感性的角度来看，CIE 被认为是不失真的。它只是二维的，因为它忽略了亮度分量，根据颜色匹配实验[204]，这与亮度分量无关。

### 混合颜色

假设我们有三种纯光源，就像由 LED 产生的光源一样，采用红色，蓝色和绿色。我们已经讨论过如何通过混合红色和绿色来产生黄色。一般来说，大多数可感知的颜色可以通过三种混合物来匹配。这被称为三色理论（或 Young-Helmholtz 理论）。实现这一点的一组颜色被称为原色。混合所有三个均匀产生感知的白光，其在显示器上被实现为  $RGB = (255, 255, 255)$ 。黑色是相反的： $RGB = (0, 0, 0)$ 。这种轻混合物遵循线性特性。假设原色用于感知两个不同光源的功率分布，如果将光源合并，则只需添加它们的原色强度即可获得组合的感知匹配。此外，总体强度可以通过乘以红色，绿色和蓝色分量来调整，而不会影响感知的颜色。只有感知的亮度可能会改变。



Anatomy of a CIE Chromaticity Diagram

图 6.22：1931 年 CIE 颜色标准与 RGB 三角形。这种表示在感知颜色之间的距离方面是正确的。（由杰夫 Yurek 提供）

迄今为止的讨论都集中在可添加的混合物。当混合颜料或印刷书籍时，由于光谱反射功能会改变，颜色也会进行减法混合。当用白色纸张时，几乎所有的波长都被反射。在页面上绘制绿线可防止除绿色以外的所有波长在该点上反射，当去除所有波长时会产生黑色。印刷机不是使用 RGB 组件，而是基于 CMYK，它对应于青色，品红色，黄色和黑色。前三种是原色的两两混合。包括黑色成分以减少通过使用其他三种颜色减色产生黑色而浪费的墨水量。请注意，只有入射光包含目标波长时才能观察到目标颜色。绿色线条在纯绿色光线下显示为绿色，但在纯蓝色光线下可能显示为黑色。

## 恒定性

图 6.20 中的连衣裙显示了一种极端的情况，由于奇怪的光照条件导致人们的颜色混淆。通常，人的色彩感觉对于颜色来源有惊人的鲁棒性。无论晚上在室内灯光下还是在阳光直射下，红色衬衫都呈红色。这些对应于在到达视网膜的光谱功率分布方面极其不同的情况。我们将物体感知为在各种光照条件下具有相同颜色的能力称为颜色恒定性。几种感知机制允许这种情况发生。其中之一是色彩适应，由于长时间暴露于特定颜色，导致感知颜色发生变化。感知颜色的另一个因素是对周围物体颜色的期望。此外，关于环境中物体通常如何着色的常识也会影响我们的理解。

在不考虑特定的颜色情况下，恒定性原则也会出现。我们的感知系统也保持亮度恒定性，使整体亮度水平看起来不变，即使在照明条件发生显着变化之后，见图 6.23 (a)。根据比例原理理论，只有场景中物体之间的反射率比例是可感知的，而反射强度的总体数量是不可感知的。更复杂的是，由于场景包含不均匀的照明，因此我们对物体的亮度和颜色的看法才得以维持。从一个对象投射到另一个对象上的阴影提供了一个清晰的情形。我们的感知系统解释了阴影并调整了我们对物体阴影或颜色的感知。图 6.23 所示的阴影错觉是由于阴影造成的补偿。

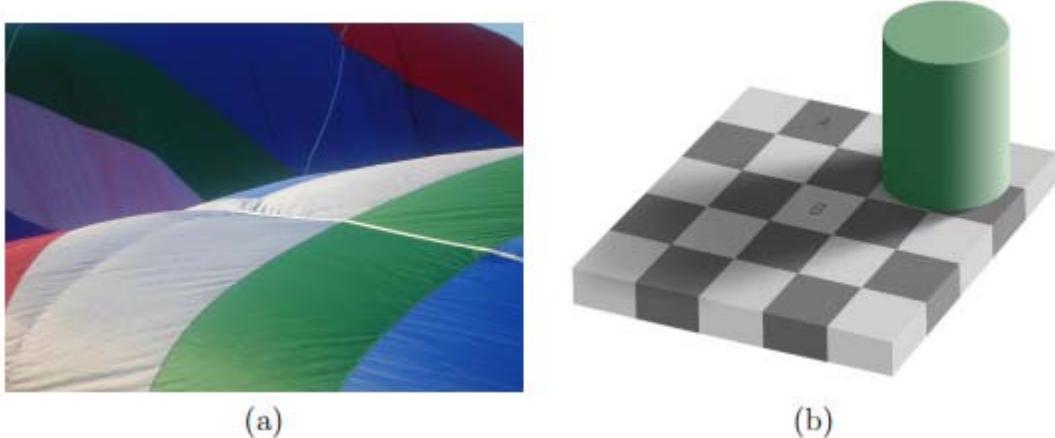


图 6.23：(a) 无论在阳光直射或阴影中，感知到的热气球颜色都是相同的。（图由维基百科用户 Shanta 提供）(b) 第 2.3 节中的阴影错觉由亮度常数原理解释，因为阴影会促使对感知亮度进行补偿。（图由 Adrian Pingstone 提供。）

## 显示问题

显示器通常使用 RGB 灯来生成颜色和亮度的调色板。回想一下图 5.22，它显示了某些常见显示器的各个组件颜色的亚像素镶嵌。通常，每个 R, G 和 B 值的强度都是通过选择 0 到 255 之间的整数来设置的。这是对亮度级数的严格限制，如 5.4 节所述。人们不可能希望密集地涵盖可感知的光强度的所有七个数量级。增强整个范围内对比度的一种方法是执行伽

马校正。在大多数显示器中，图像用约 0.45 的伽玛编码，并用 2.2 的伽玛解码。

另一个问题是，所有可用颜色的集合位于由 R, G 和 B 顶点形成的三角形内部。图 6.22 中解释了 sRGB 标准的这种限制。大多数 CIE 都被覆盖，但许多人类能够感知的颜色不能在显示器上生成。

## 6.4 结合信息来源

在本章中，我们已经看到了结合多种来源信息的感知过程。这些可以从同样的意义上得到线索，如用于判断深度的众多单眼线索。感知还可以结合来自两种或更多种感官的信息。例如，人们在面对面讲话时通常结合视觉和听觉信息。来自这两个来源的信息使得更容易理解某人，特别是如果背景噪音很大。我们也看到，随着时间的推移，信息被整合在一起，例如在人眼快速扫视时注视几个对象特征的情况下。最后，我们的常识和对周围世界行为的总体期望影响了我们的结论。因此，信息是从先前的期望以及接受许多暗示中综合而来的，这些暗示可能来自不同时期的不同感觉。

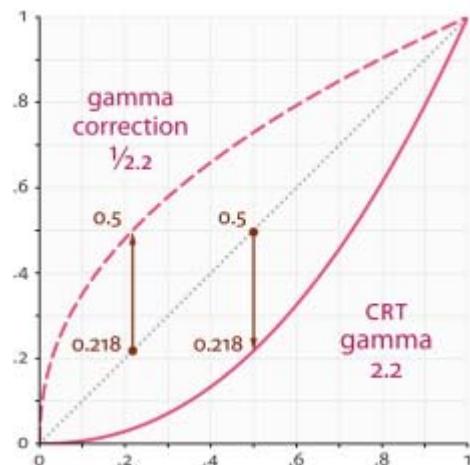


图 6.24：虽然位数有限，但伽马校正用于跨越更多数量级。变换是  $v' = cv^\gamma$ ，其中  $c$  是常数（通常  $c = 1$ ），并且  $\gamma$  控制校正或失真的非线性。

统计决策理论提供了一个有用且直接的数学模型，用于进行包含相关的观测数据的先前偏差和来源的选择。它已经应用于许多领域，包括经济学，心理学，信号处理和计算机科学。一个关键组成部分是贝叶斯原则，该原则指定如何根据新的观察结果更新先前的信念，以获得后验置信。更正式地说，“置信”被称为概率。如果概率考虑到来自先前信息的信息，则它被称为条件概率。在这里没有适当引入概率论的空间，只有基本的想法才能提供一些没有严谨的直觉。如需进一步学习，请查找在线课程或经典教科书（例如，[272]）。

让

$$H = \{h_1, h_2, \dots, h_n\} \quad (6.1)$$

为一组假设（或解释）。同样，让

$$C = \{c_1, c_2, \dots, c_m\} \quad (6.2)$$

$C$  是检测器的一组可能的输出。例如，检测器可以输出当前可见的脸部的眼睛颜色。在这种情况下， $C$  是一组可能的颜色：

$$C = \{\text{棕色, 蓝色, 绿色, 淡褐色}\}。 \quad (6.3)$$

对人脸识别器建模,  $H$  将对应于与设定的人相似的一组假设。

我们要计算  $H$  中每个假设的概率值。每个概率值必须介于 0 到 1 之间, 并且  $H$  中每个假设的概率值之和必须总和为 1。在任何信息之前, 我们先从称为先验分布的值的分配开始, 将其写为  $P(h)$ 。“P”表示它是概率函数或分配;  $P(h)$  表示赋值已经应用于  $H$  中的每个  $h$ 。必须使赋值成为

$$P(h_1) + P(h_2) + \dots + P(h_n) = 1, \quad (6.4)$$

并且对于从 1 到  $n$  的每个  $i$ ,  $0 \leq P(h_i) \leq 1$ 。

先验概率通常以分散的方式分布在假设上, 图 6.25 (a) 举了一个例子。在任何信息之前, 任何假设的可能性与其自然发生的频率成正比, 这取决于进化和人的经历的寿命。例如, 如果你在生活中的随机时间睁开眼睛, 看到人类与野猪的可能性有多大?

在正常情况下(不是 VR!), 我们预计随着信息的到来, 正确解释的可能性会增加。正确假设的概率应该向上直到 1, 有效地从其他假设中偷取概率, 这将推动它们的值朝向 0, 见图 6.25 (b)。“强”线索应该比“弱”线索更快地向上提出正确的假设。如果一个假设的概率值接近 1, 那么分布被认为是高峰的, 这意味着高置信度; 见图 6.25 (c)。另一方面, 不一致或不正确的线索会影响两个或更多假设之间的概率。因此, 正确假设的可能性可能会降低, 因为其他假设可能被认为是合理的并且接受更高的值。由于不能从给定线索解决模糊性, 两种替代假设也可能保持较强鲁棒性, 见图 6.25 (d)。

为了考虑来自线索的信息, 定义了条件分布, 其被写为  $P(h|c)$ 。这被称为“给定  $c$  的概率”。这对应于假设和线索的所有可能组合的概率分配。例如, 如果存在至少两个假设和五个线索, 则它将写为  $P(h_2|c_5)$ 。继续回到我们的脸部识别器, 这将看起来像  $P(\text{奥巴马}|\text{布朗})$ , 应该是大于  $P(\text{巴拉克·奥巴马}|\text{蓝色})$ (他有棕色的眼睛)。

我们现在得出基本问题, 即在条件到达后计算  $P(h|c)$ 。这是由贝叶斯公式完成的:

$$P(h|c) = \frac{P(c|h)P(h)}{P(c)}. \quad (6.5)$$

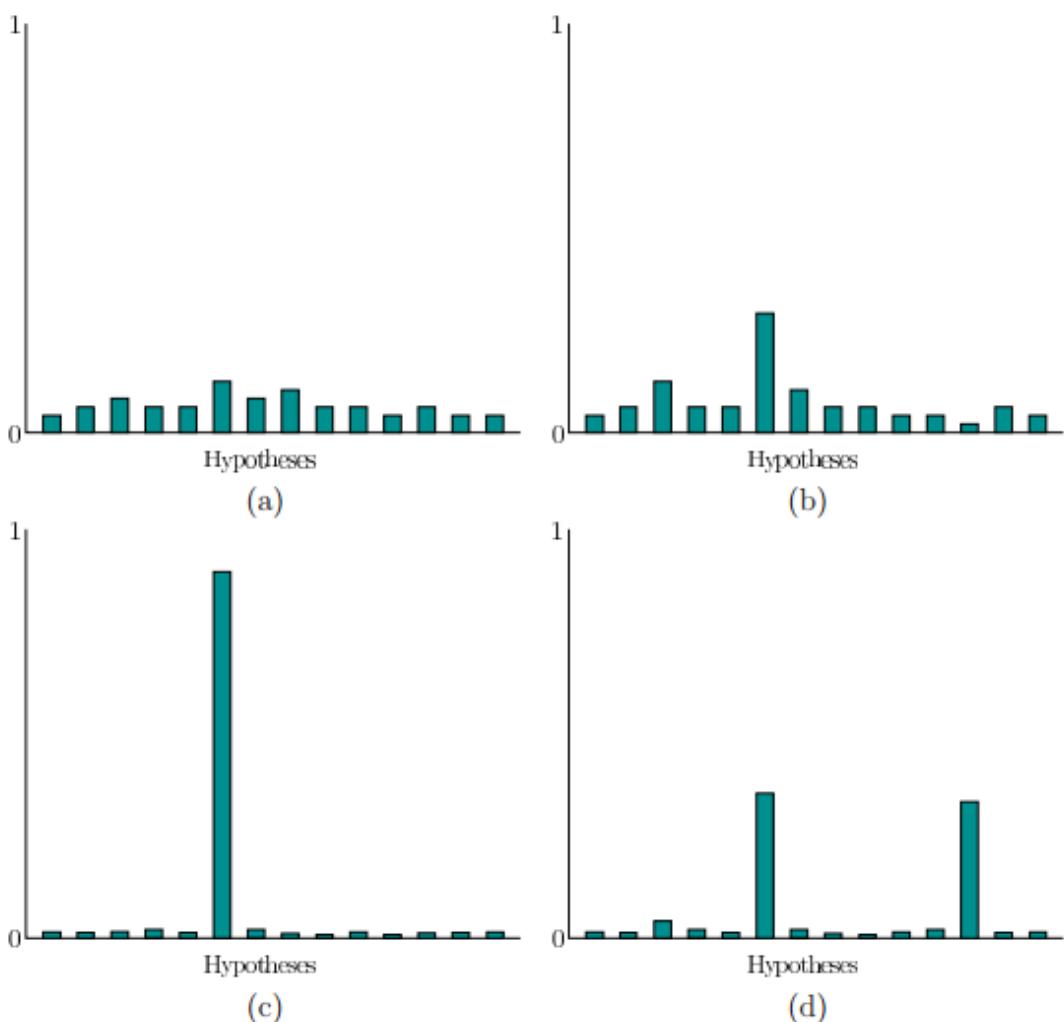


图 6.25：示例概率分布：(a) 可能的先验分布。 (b) 在条件后开始出现一种假设。 (c) 强大一致的线索导致峰值分布。 (d) 模糊性可能会导致两个（或更多）假设比其他假设更受欢迎；这是多重感知的基础。

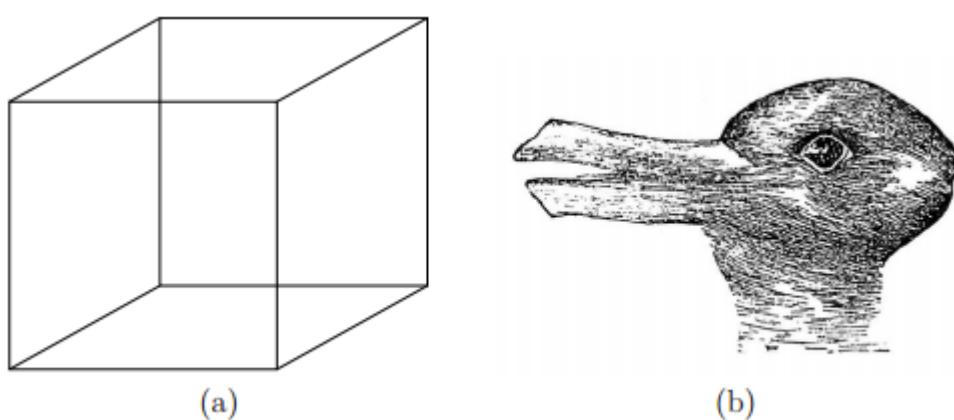


图 6.26：(a) 1832 年由瑞士晶体学家路易斯·阿尔伯特·内克研究的 Necker 立方体。 (b) 1892 年 10 月 23 日 Fliegende Blatter 的兔子鸭幻觉。

分母可以表示为

$$P(c) = P(c | h_1)P(h_1) + P(c | h_2)P(h_2) + \cdots + P(c | h_n)P(h_n), \quad (6.6)$$

或者可以忽略它作为一个归一化常数，此时只计算相对可能性而不是适当的概率。

贝叶斯原则完成的唯一事情是用先验分布  $P(h)$  和新的条件分布  $P(c|h)$  表示  $P(h|c)$ 。在建模方面，新的条件分布很容易处理。它表征了每个特定线索出现的可能性，假设该假设是真实的。

如果信息从第二个检测器到达，会怎么样？在这种情况下，再次应用(6.5)，但是  $P(h|c)$  现在被认为是关于新信息的先验分布。令  $D = \{d_1, d_2, \dots, d_k\}$  表示新检测器的可能输出。贝叶斯公式变成了

$$P(h | c, d) = \frac{P(d | h)P(h | c)}{P(d | c)}. \quad (6.7)$$

以上， $P(d|h)$  产生了所谓的条件独立假设： $P(d|h) = P(d|h, c)$ 。从建模的角度来看这更简单。更一般地说，(6.7) 的所有四个条件部分都应该包含  $c$ ，因为它是在  $d$  已知之前给出的。随着来自更多线索的信息变得可用，贝叶斯原则会根据需要重复使用。实践中出现的一个困难是模型认知偏差，这与人们做出不合理判断的众多方式相对应，尽管数据的概率影响如此。

### 多重感知

在某些情况下，我们的感知系统可能会在两个或更多个结论之间交替。这就是所谓的多重感知，对此，两种结论的特例称为双稳态感知。图 6.26 (a) 显示了两个众所周知的例子。对于 Necker 立方体而言，与观看平面平行的立方体表面位于前景中是不明确的，可以在两种解释之间切换，导致双稳态感知。图 6.26 (b) 显示了另一个例子，人们可以在不同时间看到一只兔子或一只鸭子。另一个著名的例子是来自 Nobuyuki Kayahara 的称为纺纱舞者幻觉。在这种情况下，会显示旋转舞者的轮廓，并且可以将动作解释为顺时针或逆时针。

### 麦克尔克效应

麦克尔克效应是一个实验，通过混合视觉和听觉信息来清楚地表明整合的力量[207]。一个人说话的视频会显示在配音的音轨上，以便整合出说出的声音与视频不匹配的情形。然后观察到两种类型的幻觉。如果听到“ba”并且显示“ga”，则大多数会感知到“da”被说出。这对应于解释错配的合理声音融合，但不符合任何原始线索。或者，声音可以结合起来，在声道上“ga”和视觉轨道上“ba”的情况下产生感知的“bga”。

### 对 VR 的影响

并非所有的感官都被 VR 所完美替代。因此，由于真实和虚拟世界之间的不匹配，冲突将会出现。如前所述，最常见的问题是视力障碍，这是视觉和前庭信息之间引起晕眩的矛盾，这些视觉和前庭线索是由 VR 中明显的自身运动引起的，同时在现实世界中保持静止，见 8.4 节。作为失配的另一个例子，用户的身体可能会感觉到它坐在椅子上，但 VR 体验可能涉及步行。然后，真实世界和虚拟世界之间会出现高度不匹配，以及基于本体感觉和触觉的不匹配。除了感官之间的不匹配之外，VR 硬件，软件，内容和界面中的缺陷与实际体验相比会导致不一致。其结果是可能会出现不正确的解释。更糟糕的是，这种不一致可能会增加疲劳，因为人类神经结构使用更多能量来解释混淆组合。鉴于麦克尔克效应，很容易认为许多意想不到的解释或看法可能来自 VR 系统，它提供完全不一致的线索。

VR 也非常有能力产生新的多重观念。其中一个实际发生在 VR 行业的例子涉及设计一个弹出式菜单。假设用户被置于黑暗的环境中，并且有大量菜单涌向他们。用户可能会感觉到以下两种情况之一：1) 菜单接近用户，或 2) 用户正冲向菜单。前庭感应足以解决用户是否在移动，但视觉感受到了压倒性的影响。关于正在发生的事先知识有助于产生正确的看法。不幸的是，如果做出错误的解释，那么由于感官冲突而导致 VR 晕眩增加。这，我们的感知系统可能被欺骗成一种对我们的健康更糟的解释！常识是第 12.3 节讨论的许多 VR 晕眩因素之一。

### 拓展阅读

与第 5 章一样，本章的大部分材料都出现在关于感觉和知觉的教科书中[97, 204, 350]。对于一组光学幻象及其解释，参见[233]。有关运动检测的更多信息，请参见[204]的第 7 章。与此相关的是电影的历史[32, 28]。

为了更好地理解将多个来源的线索进行组合的数学基础，请查找有关贝叶斯分析和统计决策理论的书籍。例如，参见[267]和[163]的第 9 章。一个重要的问题是通过重复使用来适应虚拟现实系统的缺陷[282, 345]。这会显着影响感知结果和错配导致的疲劳，是感知学习的一种形式，将在 12.1 节中讨论。

# 第七章 视觉渲染

本章介绍了视觉渲染，它涵盖了视觉显示器通过虚拟世界生成器（VWG）界面显示的内容。第 3 章已经介绍了数学部分，我们可以表示出虚拟世界中的物体应该出现在屏幕上的位置。这部分基于几何模型，刚体变换和视点变换。我们接下来需要根据光传播，视觉生理学和视觉上的知识来确定这些物体应该如何出现。这些分别是第 4,5,6 章的主题。因此，视觉渲染是目前涵盖上述所有内容的顶层内容。

7.1 节和 7.2 节介绍了基本概念；这是计算机图形的核心，但也会出现特定的 VR 问题，主要针对综合形成的虚拟世界进行渲染的情况。7.1 节解释了如何根据光源以及纯粹存在于虚拟世界中的材料的反射特性来确定应该出现在像素上的光。第 7.2 节解释了光栅化方法，它有效地解决了渲染问题，并广泛用于称为 GPU 的专用图形硬件。第 7.3 节解决了由光学系统缺陷引起的 VR 特定问题。第 7.4 节着重于延迟减少问题，这对 VR 来说至关重要，要保证虚拟物体在正确的时间出现在正确的位置。否则，许多副作用可能会出现，如 VR 不适，疲劳，对缺陷的适应性，或获得不具说服力的经历。最后，第 7.5 节解释了被捕捉的，而不是合成的虚拟世界的渲染。这包括从全景照片和视频形成的 VR 体验。

## 7.1 光线跟踪和着色模型

假设一个虚拟世界已经以三角形基元的形式被定义。此外，放置一双虚拟的眼睛以从一些特定的位置和方向对虚拟世界进行查看。使用第 3 章中的全变换链，每个三角形都被正确地放置在虚拟屏幕上（如图 3.13 所示）。接下来的步骤是确定哪些屏幕像素被变换后的三角形覆盖，然后根据虚拟世界的物理原理照亮它们。

还必须检查一个重要的条件：对于每个像素，三角形是否对眼睛可见，还是会被另一个三角形的一部分阻挡？这种经典的可视性计算问题极大地使渲染过程复杂化。一般问题是确定虚拟世界中的任何一对点，连接它们的线段是否与任何物体（三角形）相交。如果发生交叉，则两点之间的视线可视性被阻止。两种主要渲染方法的主要区别是如何处理可视性。

### 物体顺序 vs 图像顺序

对于渲染，我们需要考虑物体和像素的所有组合。这表明了一个嵌套循环。解决可视性的一种方法是迭代所有三角形的列表并尝试将每个三角形渲染到屏幕上。这被称为物体顺序渲染，并且是第 7.2 节的主题。对于落入屏幕视场中的每个三角形，仅当三角形的相应部分比目前为止渲染的任何三角形更接近眼睛时才更新像素。在这种情况下，外部循环遍历三角形，而内部循环遍历像素。另一种方法称为图像顺序渲染，它颠倒循环的顺序：遍历图像像素，并且对于每一个像素，确定哪个三角形会影响其 RGB 值。为了实现这一点，进入每个像素的光波路径通过虚拟环境被追踪。这种方法将首先被介绍，其中的许多组件也适用于物体顺序渲染。

### 光线追踪

为了计算像素处的 RGB 值，从放置在虚拟世界中的屏幕上的焦点通过像素的中心绘制观察光线；如图 7.1。该过程分为两个阶段：

1. 光线投射，定义了观察射线并计算虚拟世界中所有三角形之间的最近交点。

2. 阴影，根据照明条件和交点处的材料属性计算像素 RGB 值。

第一步完全基于虚拟世界几何。第二步使用虚拟世界的模拟物理。物体的材料属性和光照条件都是人工的，并且可以被选择来产生所需的效果，无论是现实还是幻想。请记住，最终的判定者是通过感知过程解读图像的用户。

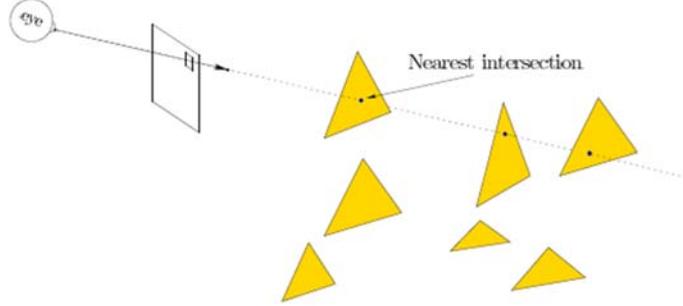


图 7.1：光线跟踪方法的第一步称为光线投射，它扩展了与图像上特定像素相对应的观察光线。射线从焦点处开始，焦点是眼睛变换  $T_{eye}$  被应用后的原点。我们的任务是确定虚拟世界模型的哪一部分是可见的。它是观察射线和所有三角形集合之间最接近的交点。

## 光线投射

如果我们忽略计算性能，计算观察光线在离开图像像素后碰到的第一个三角形（图 7.1）很简单直接。从三角形坐标，焦点和射线方向（矢量）开始，封闭形式的解决方案包含来自解析几何的基本操作，包括点积，交叉积和平面方程[320]。对于每个三角形，必须确定射线是否与其相交。如果不是，则考虑下一个三角形。如果相交，那么仅当交叉点比迄今为止遇到的最近交叉点更近时，交叉点才被记录为候选解决方案。在考虑所有三角形之后，将最近的交点将被找到。虽然这很简单，但将三角形排列到 3D 数据结构中效率更高。这样的结构通常是分层的，因此许多三角形可以通过快速坐标测试来消除。流行的例子包括 BSP 树和 Bounding Volume Hierarchies [42,85]。对几何信息进行排序以获得更高效的算法通常被归类为计算几何[54]。除了从快速测试中消除许多三角形外，许多计算光线三角交点的方法也被开发以减少操作次数。其中最受欢迎的是 Möller-Trumbore 相交算法[217]。

## 朗伯阴影

现在考虑点亮每个像素的情况并回顾 4.1 节中光的基本行为。虚拟世界模拟现实世界的物理学，其中包括光谱功率分布和光谱反射功能。假设一个点光源被放置在虚拟世界中。使用 6.3 节的三原色理论，其光谱功率分布可用 R, G 和 B 值充分表示。如果观察射线如图 7.2 所示撞击表面，那么物体应该如何出现？阴影模型考虑了关于光谱反射函数的假设。最简单的情况是朗伯阴影，其中观察射线撞击表面的角度与所得到的像素 R, G, B 值无关。这对应于漫反射的情况，适用于“粗糙”表面（回忆图 4.4）。表面相对于光源的角度是至关重要的。

设  $\mathbf{n}$  是外表面法线，并且  $\ell$  是从表面交点到光源的矢量。假设  $\mathbf{n}$  和  $\ell$  都是单位矢量，令  $\theta$  表示它们之间的角度。点积  $\mathbf{n} \cdot \ell = \cos\theta$  表示由于表面相对于光源的倾斜而产生的衰减量（0 和 1 之间）。思考三角形的有效面积是如何由于倾斜而减小的。朗伯阴影模型下的像素按如下公式点亮

$$\begin{aligned} R &= d_R I_R \max(0, \mathbf{n} \cdot \ell) \\ G &= d_G I_G \max(0, \mathbf{n} \cdot \ell) \\ B &= d_B I_B \max(0, \mathbf{n} \cdot \ell), \end{aligned} \tag{7.1}$$

其中  $(d_R, d_G, d_B)$  表示材料的光谱反射特性（三角形）， $(I_R, I_G, I_B)$  表示光源的光谱功率分布。在典型的白光情况下， $I_R = I_G = I_B$ 。对于白色或灰色材质，我们也有  $d_R = d_G = d_B$ 。

使用矢量符号，可以将 (7.1) 压缩成

$$L = dI \max(0, n \cdot \ell) \quad (7.2)$$

其中  $L = (R, G, B)$ ,  $d = (d_R, d_G, d_B)$  和  $I = (I_R, I_G, I_B)$ 。假定每个三角形都位于物体的表面上，而不是物体本身。因此，如果光源在三角形后面，那么三角形不应该被照亮，因为它背对着光线（它不能从后面照亮）。为了处理这种情况，最大函数出现在 (7.2) 中以避免  $n \cdot \ell < 0$ 。

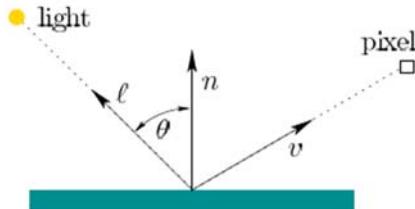


图 7.2：在朗伯阴影模型中，到达像素的光线取决于入射光与表面法线之间的角度  $\theta$ ，但与视角无关。

### Blinn-Phong 阴影

现在假设物体是“闪亮的”。如果它是完美的镜像，那么只有当它们完全对齐时，来自光源的所有光才会反射到像素；否则，没有光线会反射。如果  $v$  和  $l$  相对于  $n$  形成相同的角度，则会发生这种全反射。如果两个角度接近但不完全匹配呢？Blinn-Phong 阴影模型提出一定量的光线会被，这取决于表面闪光的量以及  $v$  与  $l$  之间的差异[24]。见图 7.3。等分线  $b$  是平均  $l$  和  $v$  后获得的矢量：

$$b = \frac{\ell + v}{\|\ell + v\|}. \quad (7.3)$$

使用压缩矢量标记，Blinn-Phong 着色模型将 RGB 像素值设置为

$$L = dI \max(0, n \cdot \ell) + sI \max(0, n \cdot b)^x. \quad (7.4)$$

由于漫反射和镜面反射两种成分，这种叠加考虑了阴影。第一项只是朗伯阴影模型 (7.2)。随着  $b$  变得更接近于  $n$ ，第二部分导致越来越多的光被反射。指数  $x$  是表示表面光泽度的材料属性。较低的值（例如  $x = 100$ ）表示轻微的光泽度，而  $x = 10000$  会使表面几乎像镜子一样。该着色模型不直接对应于光与表面之间相互作用的物理学。这只是一个方便而有效的启发式方法，但在计算机图形学中被广泛使用。

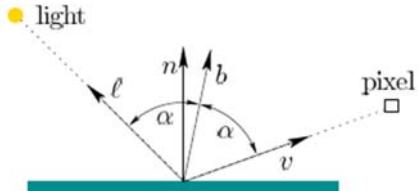


图 7.3：在 Blinn-Phong 阴影模型中，到达像素的光线取决于法线  $n$  与  $l$  和  $v$  的等分线  $b$  之间的角度，如果  $n = b$ ，则获得理想的反射，就像一面镜子。

## 环境阴影

另一个启发式想法是环境阴影，它会导致在不被光源照亮的情况下物体发光。这照亮了落入所有灯光阴影中的表面；否则，它们将完全变黑。在现实世界中，不会发生光线在物体之间互相反射以照亮整个环境。到目前为止，这种传播在阴影模型中还没有被考虑到，因此需要黑客来修复它。添加环境阴影

$$L = dI \max(0, n \cdot \ell) + sI \max(0, n \cdot b)^x + L_a, \quad (7.5)$$

其中  $L_a$  是环境光分量。

## 多个光源

通常，虚拟世界包含多个光源。在这种情况下，来自每个像素的光在像素处相加合并。对于  $N$  个光源的结果是

$$L = L_a + \sum_{i=1}^N dI_i \max(0, n \cdot \ell_i) + sI_i \max(0, n \cdot b_i)^x, \quad (7.6)$$

其中  $I_i$ ,  $\ell_i$  和  $b_i$  对应于每个源。

## BRDFs

迄今为止提出的着色模型由于其简单性和高效率而被广泛使用，尽管它们忽略了大多数物理学。为了以更精确和一般的方式考虑阴影，建立了双向反射分布函数（BRDF）[231]；见图 7.4。 $\theta_i$  和  $\theta_r$  参数分别表示相对于表面的光源角度和观察光线角度。 $\phi_i$  和  $\phi_r$  参数的范围是从 0 到  $2\pi$ ，代表在向下看表面时光线和观察矢量所形成的角度（矢量  $n$  指向你的眼睛）。

BRDF 是如下形式的一个函数

$$f(\theta_i, \phi_i, \theta_r, \phi_r) = \frac{\text{radiance}}{\text{irradiance}}, \quad (7.7)$$

其中辐亮度是从方向  $\theta_r$  和  $\phi_r$  从表面反射的光能，辐照度是从方向  $\theta_i$  和  $\phi_i$  到达表面的光能。这些被表示为不同级别，大致相当于一个无限小的表面补丁。通俗地说，它是出射光量与表面上一点入射光量的比值。以前的着色模型可以用简单的 BRDF 表示。对于朗伯阴影，BRDF 是恒定的，因为表面在所有方向上均等地反射。BRDF 及其扩展可以考虑更为复杂和物理上正确的照明效果，并具有各种各样的表面纹理。请参阅[5]的第 7 章以获得广泛的报道。



图 7.4：双向反射分布函数（BRDF），从所有可能的角度指定了入射和出射光能量的比率。

## 全局照明

回想一下，引入环境阴影项（7.5）是为了防止光源阴影中的表面显示为黑色。解决这个问题的计算密集但适当的方法是计算光在虚拟世界中如何反射物体之间的物体。通过这种方式，物体就会像现实世界中那样从被其他反射的光线间接照亮。不幸的是，这有效地将所有物体表面转变为潜在的光源。这意味着光线追踪必须考虑多重反射。这需要考

虑从光源到视点的分段线性路径，其中每个弯曲对应于反射。为通常需要考虑的反弹次数设定一个上限。简单的朗伯和 Blinn-Phong 模型经常被使用，但更普遍的 BDRF 也很常见。可以增加实际值的计算水平，但计算时间会相应增加。

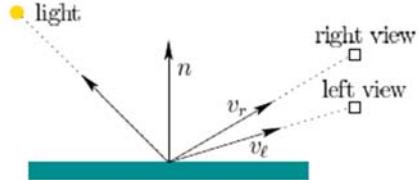


图 7.5：由于右眼和左眼的视点不同，光泽表面变得复杂。使用 Blinn-Phong 阴影模型，镜面反射对于每只眼睛应具有不同的亮度级别。为了与现实世界的行为保持一致，效果可能难以匹配。

### VR 特有的问题

VR 继承了计算机图形学中的所有常见问题，但也包含独特的挑战。第 5 章和第 6 章提到了增加的分辨率和帧率要求。这为降低渲染复杂性施加了很大的压力。此外，对于屏幕上的图形运行良好的许多启发式方法在 VR 中可能是明显错误的。高视角，分辨率，不同视角和立体图像的组合可能带来新问题。例如，图 7.5 展示了立体视图与立体视角的不同点如何影响光泽表面的外观。一般而言，一些渲染结果甚至可能导致 VR 不适。在本章的其余部分中，将会越来越多地讨论 VR 独有的复杂问题。

## 7.2 光栅化

光线投射操作很快成为瓶颈。对于 90Hz 的 1080p 图像，每秒需要执行 180 多万次，并且将针对每个三角形执行光线-三角形相交测试（尽管 BSP 等数据结构会快速消除许多考虑因素）。在大多数情况下，从这种图像顺序渲染切换到物体顺序渲染效率更高。本例中的物体是三角形，所产生的过程称为光栅化，这是现代图形处理单元（GPU）的主要功能。在这种情况下，通过遍历每个三角形并尝试对三角形落在图像上的像素着色来渲染图像。主要问题在于该方法必须解决确定哪个部分（如果有的话）最接近焦点（大致为虚拟眼睛的位置）的不可避免的问题。

解决这个问题的一种方法是按照深度顺序排列三角形，以便最近的三角形是最后一个。这使得三角形可以按照从后往前的顺序在屏幕上绘制。如果它们被正确地排序，那么任何后来渲染的三角形都会正确地将以前渲染的三角形的图像撞在相同的像素上。三角形可以一个一个绘制，完全不用考虑确定哪个最接近的问题。这被称为 Painter 的算法。但是，主要的缺陷在于深度循环的潜在存在，如图 7.6 所示，其中三个或更多三角形无法按照 Painter 算法的任何顺序正确渲染。一种可能的解决方法是检测这种情况并分割三角形。

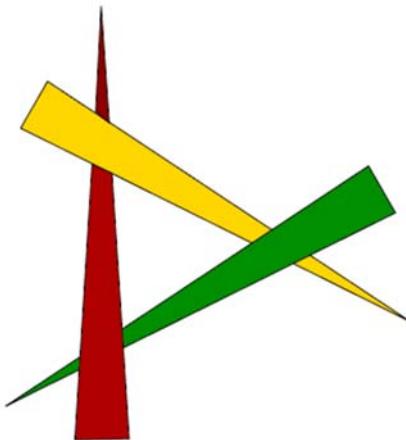


图 7.6：由于深度周期的可能性，对于距观察者的距离，物体不能按三维排序。每个物体都部分位于一个物体的前面，部分位于另一个的后面。

### 深度缓冲区

一种简单而有效的方法解决这个问题是通过维护一个深度缓冲区（也称为  $z$  缓冲区）来逐个像素地管理深度问题，对于每个像素，深度缓冲区记录三角形从焦点到在该像素处与三角形相交的射线的交点的距离。换句话说，如果这是射线投射方法，那就是沿着从焦点到交点的射线距离。使用这种方法，三角形可以以任意顺序渲染。该方法通常也适用于通过确定来自光源而不是视点的深度顺序来计算阴影的影响。靠近光线的物体会对其他物体产生阴影。

深度缓冲区在每个像素位置存储一个正实数（实际上是浮点数）。在渲染任何三角形之前，每个位置都会存储一个最大值（浮点无穷大），以反映每个像素上尚未遇到任何表面。渲染过程中的任何时候，深度缓冲区中的每个值都会记录最近渲染的三角形上的点与焦点的距离，以及图像中相应的像素。最初，所有深度都达到最大，以反映没有三角形被渲染。

通过计算完全包含它的图像的矩形部分来渲染每个三角形。这被称为边界框。通过变换三个三角形顶点来快速确定框，以确定  $i$  和  $j$ （行和列索引）的最小值和最大值。然后对边界框内的所有像素执行迭代以确定哪些位于三角形内并因此应该被渲染。这可以通过形成如图 7.7 所示的三个边缘向量快速确定

$$\begin{aligned} e_1 &= p_2 - p_1 \\ e_2 &= p_3 - p_2 \\ e_3 &= p_1 - p_3 \end{aligned} \quad (7.8)$$

点  $p$  位于三角形的内部当且仅当

$$(p - p_1) \times e_1 < 0, (p - p_2) \times e_2 < 0, (p - p_3) \times e_3 < 0, \quad (7.9)$$

其中  $\times$  表示标准向量叉积。这三个条件确保  $p$  在每个边向量的“左边”。

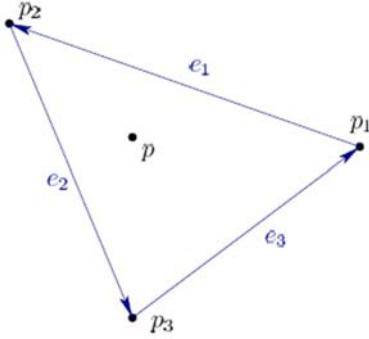


图 7.7：如果  $p$  在三角形内部，那么它必须在每个边向量  $e_1$ ,  $e_2$  和  $e_3$  的右边。重心坐标指定三角形中每个点  $p$  的位置作为其顶点  $p_1$ ,  $p_2$  和  $p_3$  的加权平均值。

### 重心坐标

随着每个三角形的渲染，来自它的信息将从虚拟世界映射到屏幕上。这通常使用重心坐标来实现（见图 7.7），该坐标将三角形内部的每个点表示为三个顶点的加权平均值：

$$p = \alpha_1 p_1 + \alpha_2 p_2 + \alpha_3 p_3 \quad (7.10)$$

其中  $0 \leq \alpha_1, \alpha_2, \alpha_3 \leq 1$  且  $\alpha_1 + \alpha_2 + \alpha_3 = 1$ 。越接近于顶点  $p_i$ , 权重  $\alpha_i$  越大。如果  $p$  位于三角形的质心，则  $\alpha_1 = \alpha_2 = \alpha_3 = 1/3$ 。如果  $p$  位于边上，则相反的顶点权重为零。例如，如果  $p$  位于  $p_1$  和  $p_2$  之间的边上，那么  $\alpha_3 = 0$ 。如果  $p$  位于顶点  $p_i$ ，那么  $\alpha_i = 1$ ，另外两个重心坐标为零。

使用 Cramer 规则计算坐标以求解所得到的线性方程组。特别地，对于  $i$  和  $j$  的所有组合，设  $d_{ij} = e_i \cdot e_j$ 。此外，让

$$s = 1/(d_{11}d_{22} - d_{12}d_{21}). \quad (7.11)$$

坐标由下式给出

$$\begin{aligned} \alpha_1 &= s(d_{22}d_{31} - d_{12}d_{32}) \\ \alpha_2 &= s(d_{11}d_{32} - d_{12}d_{31}) \\ \alpha_3 &= 1 - \alpha_1 - \alpha_2. \end{aligned} \quad (7.12)$$

可以将相同的重心坐标应用于  $R^3$  中的模型上的点或图像平面上得到的 2D 投影点（具有  $i$  和  $j$  坐标）。换句话说， $\alpha_1$ ,  $\alpha_2$  和  $\alpha_3$  是指在第 3.5 节的整个变换链之前，期间和之后模型中的相同点。

此外，考虑到重心坐标，可以通过简单地确定  $\alpha_1 \geq 0$ ,  $\alpha_2 \geq 0$  和  $\alpha_3 \geq 0$  来替代 (7.9) 中的测试条件。如果任何重心坐标小于零，则  $p$  必须位于三角形之外。

### 映射表面

重心坐标为在三角形上线性内插值提供了一种简单而有效的方法。最简单的情况是 RGB 值的传播。假设使用第 7.1 节的着色方法在三个三角形顶点计算 RGB 值。这产生对从 1 到 3 的每个  $i$  的  $(R_i, G_i, B_i)$  值。对于具有重心坐标  $(\alpha_1, \alpha_2, \alpha_3)$  的三角形中的点  $p$ ，内部点的 RGB 值被计算为

$$\begin{aligned} R &= \alpha_1 R_1 + \alpha_2 R_2 + \alpha_3 R_3 \\ G &= \alpha_1 G_1 + \alpha_2 G_2 + \alpha_3 G_3 \\ B &= \alpha_1 B_1 + \alpha_2 B_2 + \alpha_3 B_3. \end{aligned} \quad (7.13)$$

该物体不需要在整个三角形上保持相同的属性。使用纹理贴图，可以在表面上传播重复的图案，例如瓷砖或条纹[41]。

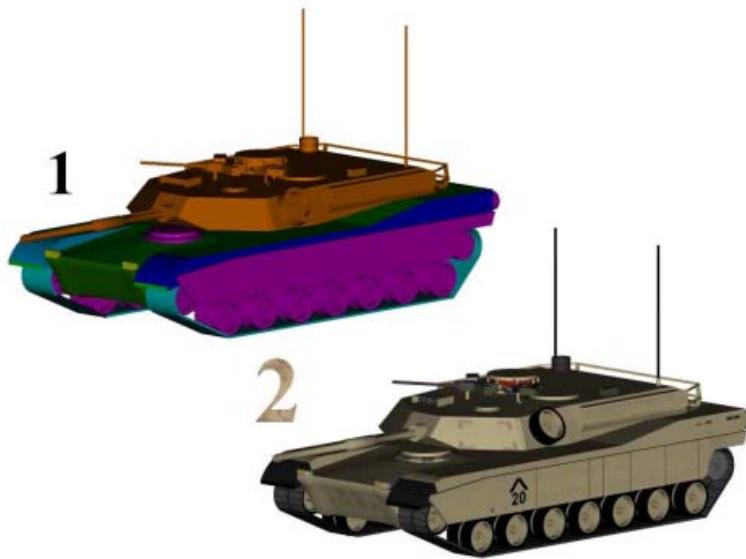


图 7.8 纹理贴图：将简单的图案或是整幅图像映射在三角形纹理中，然后在图像上进行渲染，相比于直接使用模型中的三角形纹理，这样可以提供更多细节。（图源于 Wikipedia）

通过**纹理贴图**，一些重复的图案，如瓦片或条纹等可以在物体表面进行传播[41]，如图 7.8 所示。更普遍的说，任何数字图像都可以映射到三角形上。重心坐标指的是图像中可以影响像素的点。图像，或是说“纹理”可以视为被绘制在三角形上；此外，为了更加真实，使物体有遮蔽效果，可以额外添加光照和反射属性。

另一种可行方法是**法线贴图**，通过在三角形上人工改变表面法线来改变遮蔽的过程，尽管它不能在几何上实现。在 7.1 节中的阴影模型中就使用了法线。允许其变化的话，可以在物体上添加一个仿真的曲率，法线贴图的一个重要实例叫做**凹凸贴图**，通过不规则地扰动法线使得平坦的表面变得粗糙。如果法线具有纹理的话，那么在计算阴影之后，整个表面看上去会显得很粗糙。

### 走样

由于数字图像的离散化，可能会出现一些缺陷。5.4 节中提到过走样问题，由于像素密度不足，导致一条直线看上去是阶梯状的。图 7.10 (a) 展示了在图 (7.9) 中小三角形内选中的像素。点  $p$  一般对应像素中心，如图 7.10 (b) 所示。可以看出，可能会出现该点在三角形内部而整个像素不在的情况。你可能会注意到，由于显示器中使用了亚像素镶嵌技术，图 7.10 可能不能表示真实情况。为了得到更准确的结果，走样分析也应该考虑到这一点。

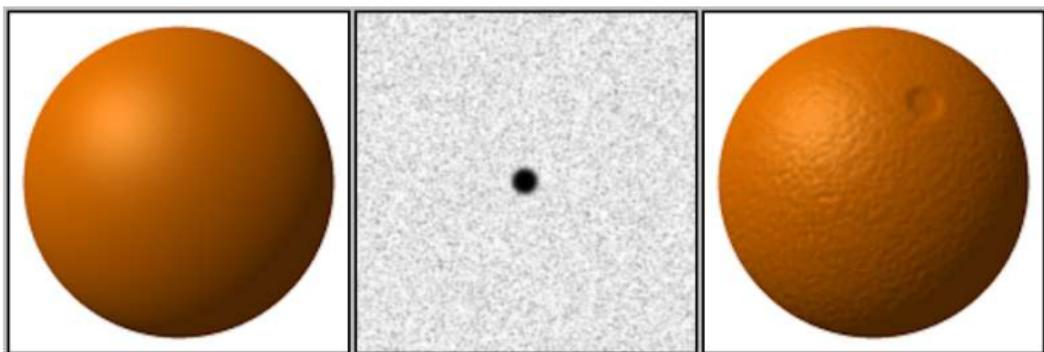


图 7.9: 凸凹贴图: 通过人工改变表面法线, 阴影生成算法会产生一个看上去粗糙的表面(图源于 Brian Vibber)

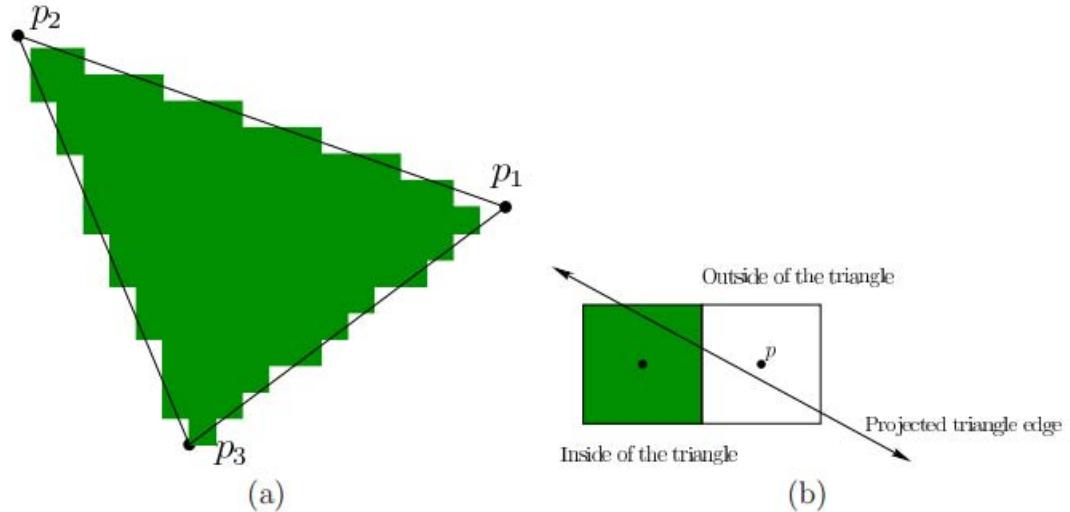


图 7.10: (a) 光栅化导致走样现象: 直线看上去像是阶梯状的 (b) 根据像素的中心点  $p$  是否在三角形内部来选取像素

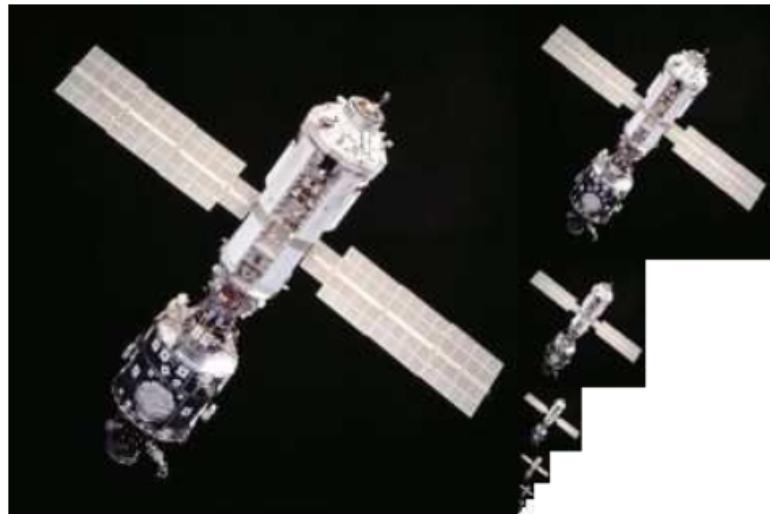


图 7.11: 纹理贴图金字塔 (mipmap), 以多个分辨率存储纹理, 使其可以适当缩放而不产生明显的走样现象。用于存储额外图像的空间通常仅为原始 (最大) 图像的 1/3 (图源于 NASA, mipmap 由 Wikipedia 用户 Mulad 创建)

仅仅通过点  $p$  的坐标来决定其是否完全包含在三角形中的话, 阶梯效应是无法避免的。更好的方法是根据三角形覆盖的像素区域来渲染像素。这样的话, 它的值可以由像素区域内多个三角形混合而成。不幸的是, 这需要超采样, 意味着要以比像素密度高得多的密度来投射射线, 从而估计三角形所覆盖的部分。这样的话代价会大大增加。通常, 可以使用一种折中的方案, 使用多重采样抗锯齿 (或 MSAA) 的方法, 在较高密度下只计算其中的一些值。一般来讲, 每个样本都要计算深度值, 但阴影不用。

空间走样的问题是由纹理贴图引起的。将虚拟世界映射到屏幕上时, 视变换会显著地减少原始纹理的大小和纵横比。如图 7.12 所示, 用来表示纹理中的重复图案的分辨率可能会有所不足。在实践中, 该问题通常通过预先计算和存储每个纹理的纹理贴图金字塔 (mipmap)

来解决，如图 7.11 所示。通过对图像进行高密度采样及存储光栅化结果，纹理可以在多分辨率下计算。基于屏幕上三角形的大小和视点，选择适当的缩放纹理图像并贴图到三角形上，可以减少走样缺陷现象。

## 剔除

在实践中，许多三角形可以在渲染之前被快速地消除。这是预处理阶段一种叫做**剔除**的方法，可以显著提高性能，也可以提高帧率。剔除操作很大程度上取决于三角形的数据结构。如果所有三角形都按照层级结构排列的话，则可以通过单次坐标比较的方式消除上千个三角形。最基本的剔除方法叫做**视域体积剔除**，可以消除完全位于视锥体之外的所有三角形（回顾图 3.18）。对于 VR 头显，由于光学系统的限制，视锥体可能会有弯曲的横截面（如图 7.13）。在这种情况下，视锥体必须被替换为一个具有合适形状的区域。对于这种情况下的截锥体，简单的几何判断可以快速消除视域外的所有物体。例如，如果

$$\frac{\sqrt{x^2+y^2}}{-z} > \tan \theta \quad (7.14)$$

$2\theta$ 指代视域的角度范围，点 $(x,y,z)$ 在锥体的外面。或者，可以在 GPU 中使用**模板缓冲**来标记镜头视域外的所有像素。在每帧画面被渲染时，通过简单的判断可以快速消除这些问题。

另一种方式叫做**背面剔除**，用于消除具有远离焦点的外表面法线的三角形。当模型持续生成时，“背面”的部分没有被渲染的必要。此外，遮蔽剔除可用于消除模型中可能会被更近的物体遮挡住的部分。这可能有些复杂，因为它要再次考虑深度顺序问题。想要了解更多细节，可以参考[5]。

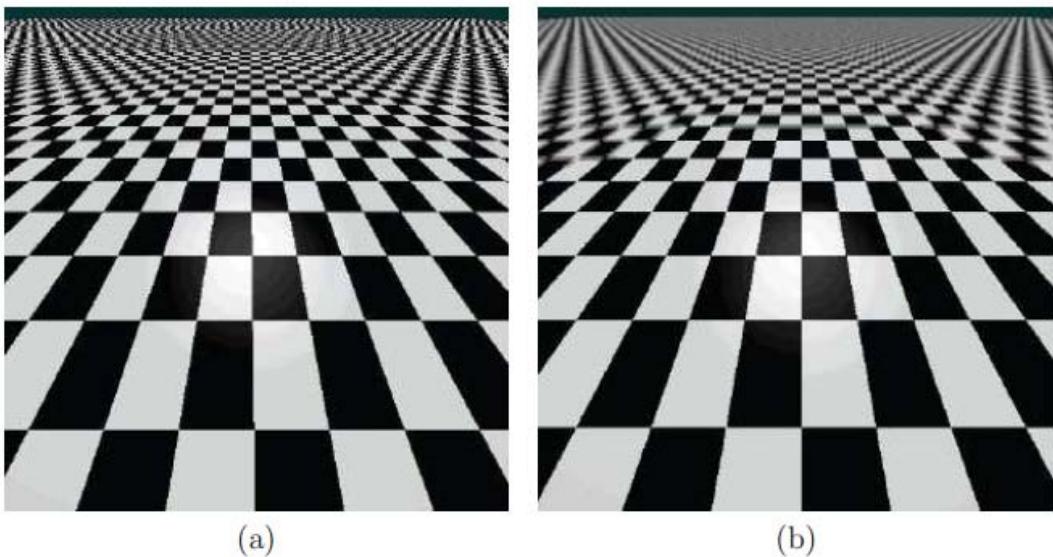


图 7.12：(a) 由于透视变换，随着深度的增加，平铺的纹理会产生**空间走样**现象 (b) 可以通过超采样来解决该问题

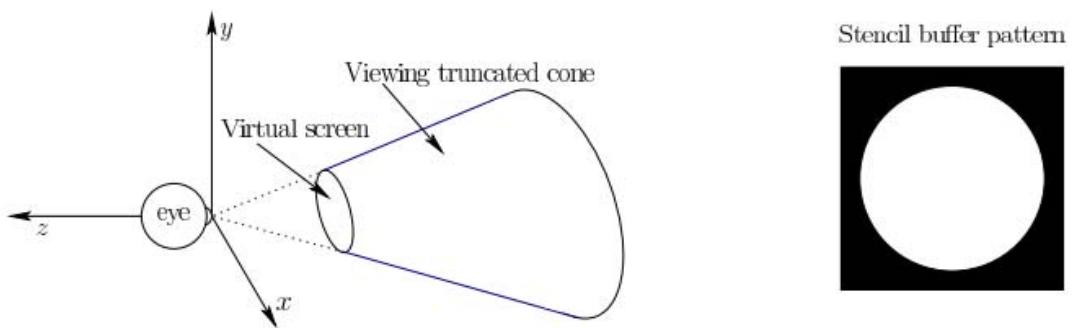


图 7.13：由于屏幕前方的光学系统，在圆形对称视图的情况下，视体将被截锥体代替。横截面形状与到每个眼睛视域的不对称性有关。（例如，鼻子会阻挡部分视域）

### 虚拟现实特有的光栅化问题

由于目前的分辨率远低于 5.4 节中所计算的视网膜显示的下限，所以由于走样造成的阶梯状问题应该会更严重。由于头部的运动导致视点不断改变，问题会变得更为严重。当用户试图注视边缘的时候，由于像素是否包含在三角形中是取决于视点的细微变化的，“阶梯”似乎更像是“自动扶梯”。作为正常感知过程的一部分，我们会被这种分散注意力的动作所吸引。对于立体的视点来说，甚至更糟：左右侧图像中的“自动扶梯”通常并不匹配。当大脑尝试将两幅图像融合成一幅连贯视图时，走样缺陷会生成强烈的运动不匹配的感觉。降低边缘的对比度并使用抗锯齿技术有助于缓解问题，但是对于 VR 来说，在显示器分辨率可以达到所需要的视网膜显示密度之前，阶梯问题一直会是一个很严重的问题。

VR 系统带来的深度感知的增强会引起一个更严重的问题。头部运动和立体视域都使用户感觉到的表面深度差异很小，这理应是一件好事，然而，在计算机图形学数十年的发展中都默认这样一种事实：将虚拟世界渲染到一个固定屏幕上时，如果从很远的距离观看，人们无法感觉到这些差异。虚拟现实会使纹理贴图看上去并不真实。例如，将地毯图片的纹理贴图到地板上时可能会让地板看上去更像是简单的涂漆。在现实世界中，我们自然可以将分辨地毯涂漆和真实的地毯。法线贴图也会出现相同的问题。由于在 VR 中两只眼睛的视线会汇聚到表面上，所以静态图像中看起来粗糙的表面经过凹凸贴图后在 VR 中有可能看上去完全平坦。因此，随着 VR 系统质量的提升，我们应该提高对渲染质量的要求，这意味着很多经典的方法应当被修改或舍弃。

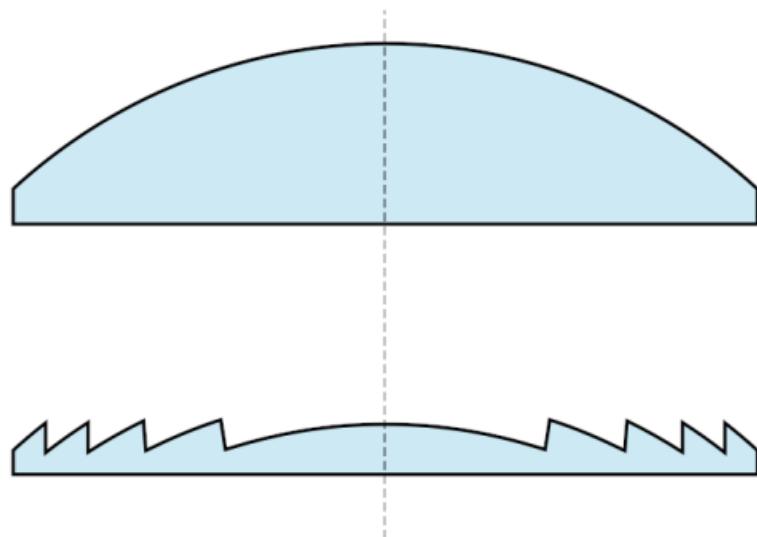


图 7.14：菲涅尔透镜，通过制造波纹表面来模拟简单的透镜，上方的凸面时由底部显示的菲涅尔透镜实现的

### 7.3 光学畸变的修正

回顾 4.3 节，对于高视场的光学系统，桶形畸变和枕形畸变是很常见的（如图 4.20）。通过 VR 头显的镜头观看时，通常会产生枕型畸变。如果图像不经过任何修正的话，那么虚拟世界看上去就会出现扭曲的现象。如果用户的头部来回转动的话，由于四周的变形比中心强烈，一些固定线条（如墙壁）的曲率会动态的改变。如果不加以修正，就没有一种静态物体的感觉，因为静态物体不应该会有动态变形。此外，这也有助于研究虚拟现实导致的一些疾病的成因，可能是由于在 VR 体验内感受到了四周异常的加速度。

那么，如何解决这个问题呢？现如今已经有很多相关的研究，可行的解决方案包括不同的光学系统及显示技术。例如，数字光处理（DLP）技术在不使用透镜的情况下直接将光投射到眼睛中。另一种实用的方法是使用菲涅尔透镜（如图 7.14），通过在较大的区域内用波纹或阶梯表面更精确地控制光线的弯曲；也可以设计成非球面的方案。例如，在 HTC Vive 的 VR 头显中就使用了菲涅尔透镜。但是它也存在副作用，当光沿着表面的突起散射的时候，会经常观察到眩光。

无论畸变的严重程度有多大，都可以通过软件进行修正。假设是默认畸变是循环对称的，这意味着畸变量仅取决于到镜头中心的距离，而不取决于中心的特定方向。即使镜头的畸变是标准的圆形对称，它也必须要放置在眼睛中央。一些头戴设备支持 IPD 调节，可以调节镜头之间的距离，使其可以匹配用户的眼睛。如果眼睛不在镜头中央，则会出现不对称畸变。这种情况并不能视为完美对称，因为随着眼睛的转动，瞳孔也会沿着球形弧面移动。随着镜头上瞳孔位置的横向变化，畸变会变得不对称。这促使厂家使用尽可能大的镜头来避免这种问题。另一个原因是，随着镜头和屏幕之间距离的变化，畸变也会变化。这种调整可以满足近视或远视用户的需求，正如三星 Gear VR 头显所做的。这种调整在双筒望远镜中也很常见，这就解释了为什么很多人在使用时不需要戴眼镜。为了正确地处理畸变问题，头戴设备应当能够准确地检测调整设置，并将这些因素考虑在内。

为了修正径向对称的畸变，我们假设将变换链  $T_{\text{can}}T_{\text{eye}}T_{\text{rb}}$  应用到几何学中，生成视域体积范式，如 3.5 节所述。所有位于视锥体内的点的  $x, y$  坐标在 -1 到 1 范围内取值。考虑用极坐标  $(r, \theta)$  描述这些点：

$$\begin{aligned} r &= \sqrt{x^2 + y^2} \\ \theta &= \text{atan2}(y, x) \end{aligned} \tag{7.15}$$

$\text{atan}$  表示  $y/x$  的反正切值。该函数通常通过编程计算，返回角度  $\theta$ ，范围在  $0$  到  $2\pi$  之间（仅使用反正切是不可以的，因为还需要知道  $(x, y)$  所在的象限）。

现在，我们通过变化半径  $r$  来表示镜头的畸变，而由于对称性，方向  $\theta$  并不会被影响。我们用  $f$  表示作用在正实数域上，变化半径为  $r$  的函数，用  $r_u$  表示未畸变的半径， $r_d$  表示畸变半径。枕形和桶形畸变通常可以使用奇数次幂的多项式来近似表示，我们定义  $f$  为：

$$r_d = f(r_u) = r_u + c_1 r_u^3 + c_2 r_u^5 \tag{7.16}$$

$c_1$  和  $c_2$  是常数。若  $c_1 < 0$ ，则为桶形畸变；若  $c_1 > 0$ ，则为枕形畸变。更高阶的项也可能会影响到，例如在后面加上  $c_3 r_u^7$  项；然而，实际中并不需要考虑。

光学畸变的修正包含两个阶段：

1. 确定特定头显的径向畸变函数  $f$ ，将特定镜头置于屏幕前固定距离的位置。这是一个回归或曲线拟合问题，步骤包含测量许多点的畸变从而确定参数  $c_1, c_2$  等，然后取最佳的

拟合结果。

2. 确定  $f$  的逆函数，使得可以在镜头产生畸变前将其应用到图像渲染上。 $f$  的逆函数按理说应该可以抵消畸变带来的影响。

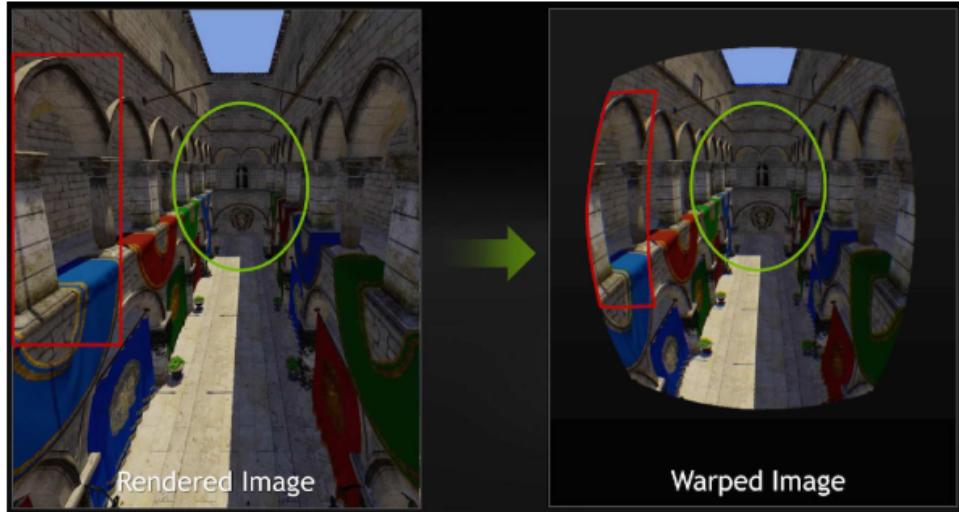


图 7.15：渲染图出现了桶形畸变。可以看到四周的分辨率有了明显的下降（图源于 Nvidia）

然而，多项式函数通常没有确定的或是闭合形式的逆函数，因此经常使用近似的结果。一种近似的方式如[118]所示：

$$f^{-1}(r_d) \approx \frac{c_1 r_d^2 + c_2 r_d^4 + c_1^2 r_d^4 + c_2^2 r_d^8 + 2c_1 c_2 r_d^6}{1 + 4c_1 r_d^2 + 6c_2 r_d^4} \quad (7.17)$$

或者，我们离线计算出精准的逆函数，然后存储在阵列中以便快速访问，每个头显只需要计算一次。使用线性插值有助于提高结果的精度。用牛顿法可以准确的计算出结果，初始值设置为交换坐标轴后的关于  $r_u$  的  $f(r_u)$  函数。

变换  $f^1$  可以直接应用于透视变换，从而用非线性运算代替  $T_p$  和  $T_{can}$ 。通过现有的图形渲染管线，它可以作为一种后处理操作。图像变换的过程有时也被称为畸变着色，因为它可以作为 GPU 中的着色操作来实现；基于逐像素操作，7.2 节中计算出的光栅化图像可以经由式 7.17 或是  $f^1$  的其他表示变换为变换后的图像。如果我们对枕形畸变进行补偿，那么所得的图像可能会出现桶形畸变，如图 7.15。为了提高 VR 的性能，Nvidia GTX 1080 型 GPU 中使用了多分辨率着色。有一个问题是，由于变换后的图像中分辨率在边缘附近会明显的下降，导致浪费了原始图像中的阴影计算。相反地，我们可以在变换之前对图像进行渲染，将最终的变换结果考虑进去。对于会被变换所压缩的部分使用较低分辨率的图像进行渲染。

本节中介绍的方法也适用于径向对称的其他光学畸变。例如，可以通过通过分别变换红、绿、蓝色子像素来校正部分色差。每种颜色径向移动不同的量以补偿基于其波长发生的径向畸变。如果正在使用色差修正，那么如果将镜头从 VR 头显中取出，则可以明显的看出呈现给显示器的图像颜色没有完美对齐。渲染系统必须根据颜色创建像素位置的变形，使得在经过镜头后可以移动到正确的位置。

## 7.4 减小时延&提升帧率

VR 头显中的系统延时指的是从响应头部方向和位置运动到更新屏幕画面所需的时间。

例如，假设用户正在注视虚拟世界中的某一固定特征点。此时如果头部偏向右侧，屏幕内对应的图像特征必须要立刻向左侧移动。否则，如果眼睛还在盯着该点，就会发现它有所移动。这破坏了平衡的感觉。

### 一个简单的例子

我们通过下面的例子来感受延迟问题。设  $d$  为屏幕的像素密度，以 pixels/rad 为单位。令  $\omega$  为头部转动的角速度，单位 rad/s； $l$  表示时延，单位 s。由于时延和角速度，头像偏移了  $d\omega l$  像素。例如， $d=40\text{pixels/rad}$ ,  $\omega=50\text{rad/s}$ ,  $l=0.02\text{s}$ , 那么图像会产生  $d\omega l=4\text{pixels}$  的错误偏移。快速的头部旋转可能会有 300rad/s 的角速度，将会产生 24 像素的错误量。

### 完美的系统

我们做一个思想实验，考虑一个完美的 VR 系统。随着头部的移动，视点可以相应的改变视觉渲染。存在一个“神谕”，可以随时指示正确的头部位置和方向。VWG 可以一直维持虚拟世界中所有物体的位置和方向。视觉渲染系统维持着所有的视角和视域变换，并且根据着色模型，光栅化过程持续对屏幕的 RGB 值进行设定。进一步设想，屏幕本身是不断更新的，并不需要花费时间来切换像素。如图 5.4 节所示，屏幕具有视网膜级别的分辨率，并且为了匹配人类的感知，具有 7 个数量级的光线输出动态范围。在这种情况下，虚拟世界给予的视觉刺激应该可以与几何类似的物理世界的情景相对应。在时间和空间上不存在误差（尽管由于对照明、阴影、材料属性、色彩空间等的假设，可能并不能够与物理世界完全相同）

### 历史问题

在实践中，完美系统是不可能实现的。所有的这些操作都需要时间来传播信息和进行计算。在早期的 VR 系统中，系统延时经常会超过 100 毫秒。在 20 世纪 90 年代，延迟减少到了 60 毫秒。时延被认为是导致 VR 疾病的重要原因之一，因此这也是过去几十年 VR 广泛运用的一大障碍。一般来说，人类可以适应固定的延迟时间，这似乎对该问题有所缓解，但他们最后不得不重新调整到现实世界中去时，又会引发一系列问题。由于难以适应，时延如果是变化的话情况会更加糟糕[68]。幸运的是，由于最新一代的跟踪、GPU 和显示技术，时延不再是大多数 VR 系统的主要问题。现在的时延可以控制在 15 到 25 毫秒左右，甚至还可以通过跟踪系统的预测方法进行补偿，达到接近 0 的有效时延。因此，现如今其他因素对 VR 疾病和 VR 疲劳的影响更大，如体感和光学像差等。

### 减少时延的方法概述

以下的方法可以一起使用，从而减少时延并最小化剩余时延带来的副作用：

1. 降低虚拟世界的复杂度
2. 提高渲染管线的性能
3. 消除从渲染图像到切换像素之间的延迟
4. 预测估计未来的视点和状态
5. 移动或扭转渲染图以补偿上一时刻的视点错误和丢失的帧

我们会相继介绍上述的方法。



图 7.16：可以使用网格简化算法来减少模型的复杂度，同时保留重要的结构。此图显示的是由开源库 CGAL 制作的手模型的简化版（图源于 Fernando Cacciola）

### 简化虚拟世界

回顾 3.1 节，虚拟世界是由几何图元组成的，通常是三维网格中排列的三角形。对于每个三角形，变换和光栅化的过程是必须的，从而导致与三角形数量成正比的计算代价。因此，一个包含数千万三角形的模型要比一个仅由数千个三角形构成的模型要花费更长时间。在很多情况下，我们得到的模型比必要的要大得多，它们通常可以做的更小（用更少的三角形），而没有明显差异，这很像图像、音视频压缩的工作原理。为什么它们一开始会很大？如果模型是由现实世界的 3D 扫描中采集的，那么它包含了高度密集的数据。使用如 FARO Focus3D X 系列之类的采集系统可以采集外部的大型世界。其他的如 Matter 和 Form MFSV1，可以通过使物体在转盘上旋转的方法来进行采集。与相机一样，自动构建 3D 模型的系统专注于生成高度精确和密集的模型，从而尺寸也会最大化。即使使用纯合成的世界，Maya 或 Blender 等建模工具也会自动在曲面上构建高度精确的三角形网格。如果不关注对后面渲染过程造成的负担，模型可能会很快变得又笨又大。幸运的是，使用网格简化算法可以减少模型的大小，如图 7.16 所示。这样做的话，需要确保简化的模型在 VR 系统所有可能的视角中都有足够好的质量。在一些系统中，如 Unity 3D，减少模型中不同材质属性的数量也可以提高性能。

除了减少渲染时间之外，简化模型还会降低虚拟世界生成器（VWG）的计算需求。对于静态世界，VWG 在初始化后不需要执行任何更新操作。用户只是简单地观察这个固定的世界。对于动态世界，VWG 要维持虚拟世界的仿真，移动世界中的几何体以满足现实世界的物理规律。它需要处理角色的动作、坠落的物体、行驶的车辆、摇曳的树木等等。当物体之间相互接触产生反应需要用到碰撞检测方法。系统可以集成用于模拟运动定律的微分方程，以便物体能够随着时间正确的移动。这些问题将在第 8 章做介绍，现在只需理解 VWG 在每次进行渲染请求时都要保持虚拟世界的一致性。因此，VWG 具有和显示器或视觉渲染系统相同的帧率。每个 VWG 框架对应着常见时刻所有几何体的位置。VWG 每秒可以更新多少次？VWG 高帧率可以保持恒定吗？当 VWG 对世界进行更新的过程中提出渲染请求会发生什么？如果渲染模块不等待 VWG 完成就执行的话，一些物体就可能会被放置在错误的

位置，因为更新过程中只有部分物体随着更新。因此，理想情况下，系统应该等待，直到 VWG 帧更新结束再开始渲染。这要求 VWG 的更新速率至少应该与渲染过程一样快，并且要做好同步，以便于为渲染提供完整的、准确的 VWG 框架。

## 改善渲染性能

任何可以提高计算机图形学领域渲染性能的技术都适用于此。然而，我们必须避免在计算机显示器上察觉不到的副作用，这在 VR 中体现的更为明显。在 7.2 节中已经提到，VR 中的纹理和法线映射方法效率不高，未来几年可能会出现更多的差异。关于 VR 独有的改进，已在第 7.2 和 7.3 节中提到，模板缓冲区和多分辨率着色可通过利用由 VR 头戴显示器中的镜头引起的形状和畸变来改善渲染性能。进一步的改进可以在 GPU 中并行执行左右眼的光栅化，每个处理器使用一个处理器。这两个过程是完全独立的。这是专门设计针对 VR 的 GPU 的重要的第一步。

## 从渲染图像到切换像素

VWG 帧不连贯的问题也出现在渲染到显示的过程中：当需要将渲染图像扫描到显示器时，VWG 帧可能还没有完成。回想一下 5.4 节，大多数显示器都有一个滚动扫描，将位于显存中的光栅化图像的行逐一绘制到屏幕上。这是由电子束在模拟电视屏幕上点燃荧光粉的运动激发的。这个动作是从左到右，从上到下，就像我们用铅笔和纸写出一页英文文本一样。由于弯曲光束的磁性线圈中存在感应惯性，因此有几毫秒的时间称为 VBLANK（垂直消隐），其中光束从屏幕的右下方移动到左上方以开始下一帧。在此期间，光束将被关闭以避免在框架上画出对角线，因此，称为“消隐”。短消隐间隔也出现在每条水平线上，以使光束从右向左返回。



图 7.17：如果在发生显示扫描时将新的帧写入显存，则会出现裂痕，其中两个或更多帧的部分同时可见。

在数字显示时代，扫描过程是不必要的，但它仍然会存在并造成一些麻烦。假设显示器以 100 FPS 运行。在这种情况下，每 10ms 发出一次绘制新渲染图像的请求。假设 VBLANK 消耗 2ms，剩下的 8ms 花费在显示器上绘制线条。如果在 2ms 的 VBLANK 期间新的光栅化图像被写入显存，那么它将在剩余的 8ms 内正确绘制。也有可能通过激光束获得额外的时间[25,212]。但是，如果正在写入新图像并通过光束扫描出来的位置，则会发生裂痕，因为它看起来好像是屏幕被撕成碎片；见图 7.17。如果 VWG 和渲染系统以 300 FPS 生成帧，则 3 或 4 张图像中的一部分可能会出现在显示屏上，因为在扫描线条时图像会发生多次变化。解决此问题的一个解决方案是使用 VSYNC（发音为“vee sink”），这是一种防止显存写入 VBLANK 间隔之外的标志。

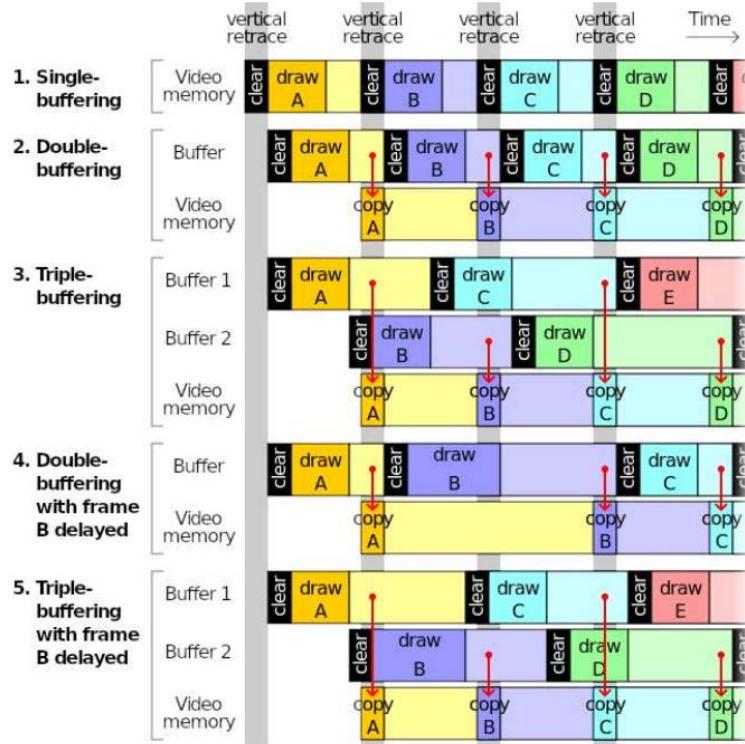


图 7.18: 缓冲通常用于可视化渲染流水线以避免裂痕和丢失帧; 然而, 它引入了更多的延迟, 这对 VR 是不利的。(图源于维基百科用户 Cmglee)

另一种避免裂痕的策略是缓冲, 如图 7.18 所示。这种方法对于程序员来说很简单, 因为它允许将帧写入未被扫描以输出到显示器的存储器中。它的一个副作用是增加了延迟。对于双缓冲, 首先将一个新帧拖入缓冲区, 然后在 VBLANK 期间传输到显存。通常很难控制产生帧的速率, 因为操作系统可能暂时中断过程或改变其优先级。在这种情况下, 三重缓冲是一种改进, 允许更多时间渲染每一帧。为了避免裂痕和提供流畅的视频游戏性能, 缓冲功能非常有用; 然而这样导致延迟增加, 这对 VR 来说是不利的。

理想情况下, 显示器应该具有全局扫描, 其中所有像素都在同一时间切换。这允许更长的时间间隔写入显存并避免裂痕。它还可以减少将第一个像素扫描到最后一个像素所需的时间。在我们的例子中, 这是一个 8ms 的间隔。最后, 显示器应尽可能减少像素切换时间。在智能手机的液晶显示屏上, 切换像素可能需要长达 20ms 的时间; 然而, OLED 像素可以在 0.1ms 以内切换。

## 预测研究

对于本节的其余部分, 我们将考虑如何处理各种情况下的延迟。作为另一个实验, 设想存在一个能够准确预测未来算命先生。有了这样的设备, 应该可以消除所有延迟问题。我们想要问算命先生如下:

1. 像素将在什么时候切换?
2. 什么是所有虚拟世界模型的位置和方向?
3. 哪个时候用户在哪里看?

让  $t_s$  来回答第一个问题。我们需要让 VWG 产生一个时间帧  $t_s$ , 然后在时间  $t_s$  对用户的视点进行视觉渲染。当像素在时间  $t_s$  切换时, 刺激将在准确的时间和地点呈现给用户。在这种情况下, 有效等待时间为零。

现在考虑实践中会发生什么。首先要注意的是, 使用来自上述所有三个问题的信息意味

着在整个 VR 系统中显着的时间同步：所有操作都必须能够访问公共时钟。对于上面的第一个问题，如果计算机足够强大并且 VR 系统具有足够的来自操作系统的控制以确保 VWG 帧将以帧速率一致地产生和渲染，那么确定  $t_s$  应该是可行的。第二个问题对于静态虚拟世界来说很简单。在动态世界的情况下，对于根据可预测的物理规律移动的所有机构而言，可能很简单。然而，很难预测人类在虚拟世界中会做什么。这使第二个和第三个问题的答案变得复杂。幸运的是，延迟非常小，动量和惯性起着重要作用，详见第 8 章。匹配区域中的物体遵循真实世界的物理运动规律。这些动作根据第 9 章中介绍的方法进行检测和跟踪。虽然可能很难预测 5 秒钟内将要看到的位置，但可以非常准确地预测头部将在 20ms 内定位和定向的位置。你没有 20ms 的自由意志！相反，动量占主导地位，头部运动可以准确预测。有些身体部位，特别是手指，惯性较差，因此变得更难以预测；然而，这些并不像预测头部运动那么重要。视角仅取决于头部运动，延迟减少在这种情况下最为关键，以避免导致疲劳和 VR 晕眩的感知问题。

Perturbation	Image effect
$\Delta\alpha$ (yaw)	Horizontal shift
$\Delta\beta$ (pitch)	Vertical shift
$\Delta\gamma$ (roll)	Rotation about image center
$\Delta x$	Horizontal shift
$\Delta y$	Vertical shift
$\Delta z$	Contraction or expansion

图 7.19：基于视点变化的六种自由度渲染图像的扭曲。前三个对应于一个方向。剩下的三个对应一个位置变化。这些操作可以通过打开数码相机并观察图像在每个扰动下如何变化来可视化。

### 后渲染图像扭曲

由于预测过程中存在等待时间和不完善之处，因此在将帧扫描到显示器之前可能需要进行最后时刻的调整。这被称为后渲染图像扭曲[201]（它在最近的 VR 行业中也被重新发现并被称为时间扭曲和异步再投影）。在这个阶段，没有时间进行复杂的阴影操作；因此，只是对图像进行简单的变换。

假设图像已经被光栅化了一个特定的视点，由位置 ( $x, y, z$ ) 和由偏航，俯仰和滚动 ( $\alpha, \beta, \gamma$ ) 给出的方位表示。如果对附近的观点进行光栅化处理，图像会有什么不同？基于视点的自由度，有六种类型的调整；见图 7.19。其中每一个都有一个未在图中指定的方向。例如，如果  $\Delta\alpha$  是正的，这对应于视点的小的逆时针偏转，则图像水平地向右移动。

图 7.20 显示了一些图像扭曲的例子。大多数情况下需要渲染的图像比目标显示更大；否则，将不会有数据转移到扭曲的图像中；见图 7.20 (d)。如果这种情况发生，那么最好从渲染的图像边缘重复像素，而不是将它们变成黑色[201]。

### 扭曲图像中的缺陷

由于方向变化导致的图像变形会产生正确的图像，因为它应该完全是从头开始为该方向渲染的图像（不考虑别名问题）。但是，位置的变化是不正确的。 $x$  和  $y$  中的扰动不考虑运动视差（从 6.1 节回忆），但是需要知道物体的深度。 $z$  中的变化会产生类似不正确的图像，因为附近的物体应该比其他物体扩展或收缩更多。更糟糕的是，视点位置的变化可能导致可视性事件，其中物体的一部分可能仅在新视点中变为可见；见图 7.21。诸如方向图[250]和可视性复合体[252]之类的数据结构被设计用于维护这些事件，但通常不包括在渲染过程中。随着等待时间变短并且预测变得更精确，扰动量减少。精密的感知研究是必要的，它可以用来评估图像扭曲错误以及可察觉到的导致不适的情况。图像变形的一种替代方法是使用并行处

理来对未来的几个视点进行采样并为所有视图渲染图像。然后可以选择最正确的图像，以大大减少图像扭曲缺陷。

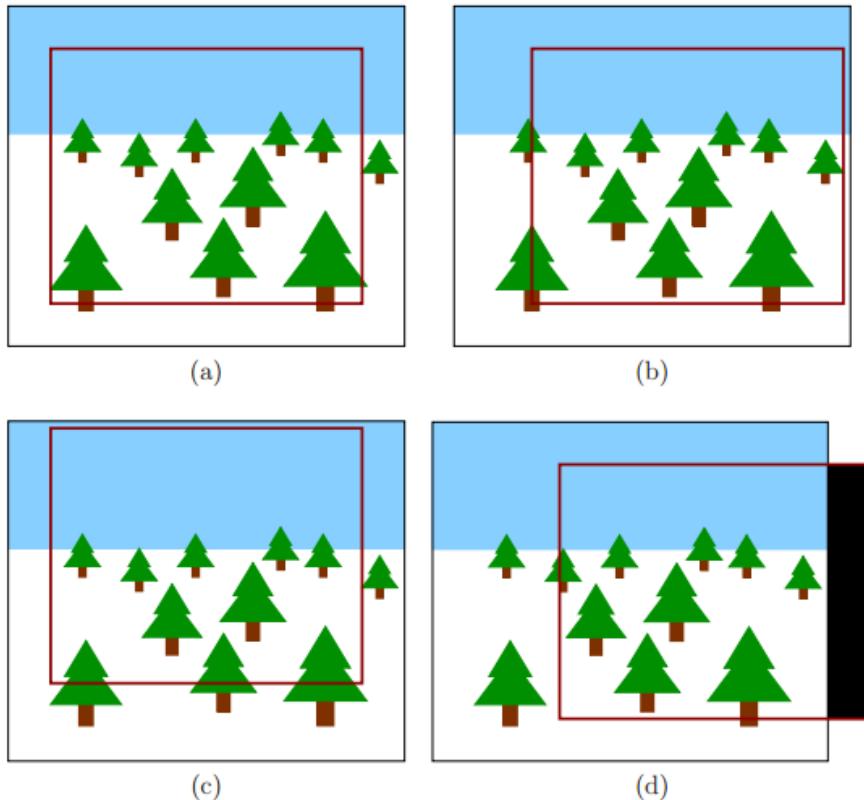


图 7.20：后渲染图像扭曲的几个示例：(a) 在变形之前，更大的图像被光栅化。红色框根据光栅化时使用的视点显示要发送到显示器的部分；(b) 负向偏航（将头部向右转）导致红色方框向右移动。图像似乎向左移动；(c) 正节距（向上看）会使盒子向上移动；(d) 在这种情况下，偏航过大，并且没有光栅化数据用于部分图像（该区域显示为黑色矩形）。

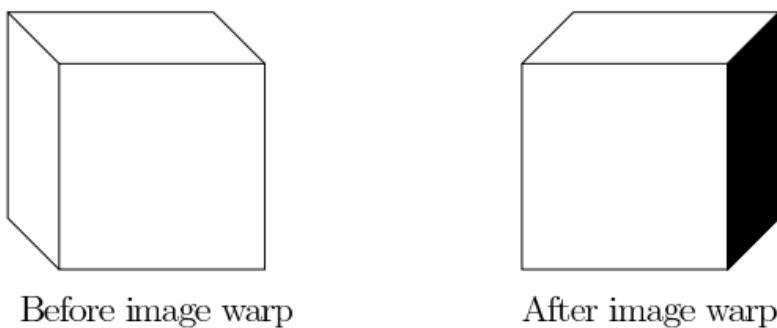


图 7.21：如果查看位置改变，则可能会遇到可视性事件。这意味着物体的一部分可能突然从新的角度变得可见。在这个示例中，视点中的水平移位揭示了最初隐藏的立方体的一侧。此外，立方体的顶部会改变其形状。

### 提高帧率

渲染后图像扭曲也可用于人为地增加帧速率。例如，假设弱电脑或 GPU 每 100 毫秒只产生一个光栅化图像。这会导致 10 FPS 时性能不佳。假设我们想将其提高到 100 FPS。在这种情况下，单个光栅化图像可以被扭曲以每 10ms 产生一帧，直到计算出下一个光栅化图像。在这种情况下，为每个正确渲染的光栅化图像插入 9 个变形帧。这个过程称为插入或补间，

并且已经使用了一个多世纪（最早的例子之一是制作幻想曲，这在图 1.25 (a) 中有描述）。

## 7.5 沉浸式照片和视频

到目前为止，本章重点介绍了从几何模型综合构建的虚拟世界。数十年计算机图形学研究所开发的方法已针对这种情况进行了研究。不过，这种趋势最近发生了变化，人们更想捕捉真实世界的图像和视频，然后轻松嵌入到 VR 体验中。这一变化主要归功于智能手机行业，这导致成千上万的人们随身携带高分辨率相机。此外，3D 相机技术不断前进，除了颜色和光线强度之外，还提供距离信息。所有这些技术正迅速地融合到全景图的情况下，全景图包含从所有可能的观察方向获取的图像数据。当前的挑战是也捕获所有可能的观看位置和方向的数据。



图 7.22：(a) 截至 2015 年，Netflix 将在线电影流式传输到大型虚拟电视屏幕上，而用户似乎坐在起居室。 (b) 电影以帧为单位纹理映射到电视屏幕上。此外，注视指针允许用户查看特定方向以选择内容。

### 纹理映射到虚拟屏幕上

将照片或视频放入虚拟世界是纹理映射的扩展。图 7.22 显示了 Netflix 通过 Samsung Gear VR 头戴显示器提供在线电影串流的商业用途。虚拟屏幕是单个矩形，可以将其视为由两个三角形组成的简单网格。照片可以映射到虚拟世界中的任何三角形网格。在电影的情况下，每个帧被视为纹理映射到网格的照片。电影的帧率通常比 VR 头盔的低很多（回忆图 6.17）。例如，假设电影以 24 FPS 录制，头盔以 96 FPS 录制。在这种情况下，每个电影帧都会在头戴显示器上呈现四帧。大多数情况下，帧速率并不是完全可分的，这导致重复帧的数量在模式中交替。一个之前的例子被称为 3:2 下拉，其中 24 FPS 电影以 30 FPS 变换为 NTSC 电视格式。有趣的是，3D 电影（立体）体验甚至可以被模拟。对于头盔显示屏上的左眼，将左眼电影帧渲染到虚拟屏幕上。类似地，右眼电影帧被呈现给头戴式显示器的右眼部分。其结果是用户将其视为 3D 电影，而不戴特殊眼镜！当然，她会戴着 VR 头盔。

### 捕捉更广阔的视野

映射到矩形上使得将用普通相机拍摄的照片或电影放入 VR 中更为简单。然而，VR 媒体本身允许扩展许多的体验。与真实世界中的生活不同，虚拟屏幕的大小可以扩展，而不会有任何显着的成本。为了填充用户的视野，将虚拟屏幕弯曲并将用户置于中心是有意义的。这种弯曲在现实世界中已经存在；例如 20 世纪 50 年代的 Cinerama 经验，如图 1.28 (d) 所示，以及现代曲面显示器。在极限情况下，我们会获得全景照片，有时称为光球。每秒显示

许多照片，导致一部全景电影，我们可以称之为电影。

回顾第 4.5 节中相机的工作方式，无法在一个瞬间从单台相机捕捉光圈，这样的话，存在两个明显的选择：

1.每次使用一台摄像机拍摄多张图像，每次指向不同方向，直至覆盖所有观看方向的整个球体。

2.使用多个摄像机，指向不同的观看方向，以便通过拍摄同步的图像覆盖所有方向。

第一种情况导致在计算机视觉和计算摄影中的图像拼接技术得到充分研究。通过将来自各种相机和时间的任意图像集拼接在一起，可以制作出硬版本。例如，从网上照片集合建立一个受欢迎的旅游网站光球。更常见的是，智能手机用户可以通过将朝向相机的方向指向足够的方向来拍摄光球。在这种情况下，软件应用程序动态构建光球，同时快速连续拍摄图像。对于硬版本，出现了一个困难的优化问题，其中需要在多个图像的重叠部分中识别和匹配特征，同时考虑未知的固有摄像机参数。必须考虑视角差异，光学像差，照明条件，曝光时间以及场景在不同时间的变化。在使用智能手机应用的情况下，正在使用相同的相机并且图像之间的相对时间很短，因此，任务要容易得多。此外，通过连续拍摄快速图像并使用内置的智能手机传感器，匹配重叠的图像部分要容易得多。这种手工制作的照相球中的大多数缺陷是由于用户无意中改变了照相机的位置，同时将其指向各个方向。

对于第二种情况，可以设计一系列相同的相机，以便覆盖所有观看方向；见图 7.23 (a)。一旦设备被校准，相机的相对位置和方向就能被精确知道，将图像拼接在一起变得简单。尽管如此，由于照明或校准的变化，修正仍然需要，否则，缝合中的接缝可能可以察觉得到。这种权衡取决于相机的数量。通过使用许多相机，可以以相对较小的光学畸变进行非常高分辨率的拍摄，因为每个相机都会为光圈提供一个窄视场图像。另一方面，只有两台摄像机就足够了，例如理光 Theta S (图 7.23 (b))。相机指向 180 度，鱼眼镜头能够捕捉大于 180 度的视角。这种设计大大降低了成本，但需要对两个捕获的图像进行重要的变形修正。



图 7.23：(a) 360Heros Pro10 HD 是一台可以在相反方向安装 10 台 GoPro 摄像机以捕获全景图像的装备。

(b) 理光 Theta S 仅使用两台摄像机拍摄全景照片和视频，每台摄像机均提供一个视场大于 180 度的镜头。

### 映射到球体上

制图中众所周知的地图投影问题是将光球映射到屏幕上，然而，在 VR 中渲染光球时不会出现这种情况，因为它直接映射到虚拟世界中的球体上。所有可能的观察方向的球体映射到虚拟世界球体而没有畸变。为了直接使用纹理映射技术，虚拟世界球可以用统一的三角形近似，如图 7.24 (a) 所示。光球本身的存储方式不会降低某些地方的分辨率。我们不能简单地使用纬度和经度坐标来索引像素，因为极点和赤道之间的分辨率差异太大。我们可以使用与四元数使用指数 ( $a, b, c$ ) 覆盖球体并且要求  $a^2 + b^2 + c^2 = 1$  的方式类似的坐标。然

而，相邻像素的结构（上，下，左，右）并不清楚。一个简单而有效的折中办法是将光球表示为六个方形图像，每个图像对应一个立方体的面。这就像一个六面 CAVE 投影系统的虚拟版本。然后，每个图像可以很容易地映射到网格上，分辨率损失很小，如图 7.24 (b) 所示。

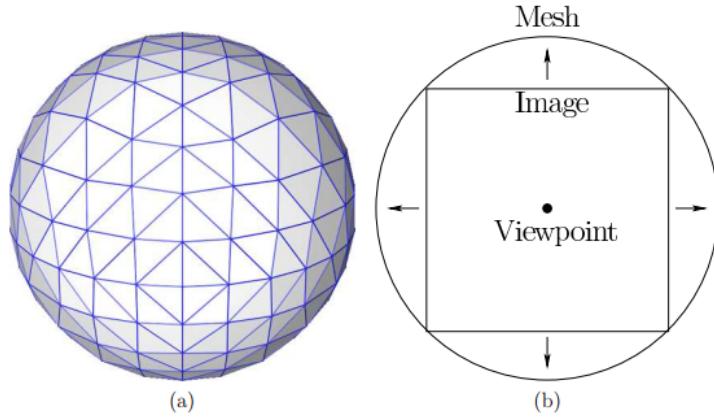


图 7.24：(a) 光斑纹理映射到球体的内部，被建模为三角形网格。 (b) 以六个图像的立方体形式存储的光球可以快速地映射到球体，分辨率损失相对较小；这里显示了横截面。

一旦光球（或电影球体）被渲染到虚拟球体上，渲染过程就非常类似于渲染后图像扭曲。呈现给用户的图像被变换为图 7.19 中描述的旋转情况。实际上，整个光栅化过程只能对整个球体执行一次，而渲染到显示器的图像则根据观察方向进行调整。进一步优化可以通过绕过网格并直接从捕获的图像形成光栅化图像来完成。

### 感知问题

观看光球或电影球时虚拟世界是否显示为“3D”？回想 6.1 节，还有更多的单眼深度线索，它们比立体线索更多。由于现代虚拟现实头戴式设备和单目深度线索的高视野，获得令人惊讶的沉浸式体验。因此，即使全景图像的相同部分呈现给双眼，它也可能比人们所期望的更“3D”。许多有趣的问题留待未来的研究关于全景的看法。如果左眼和右眼呈现不同的观点，那么虚拟球体的半径应该是多少，以获得舒适和逼真的观看效果？继续进一步，假设使用位置头部跟踪。这可能会提高观看舒适度，但由于视差不起作用，虚拟世界将显得更加平坦。例如，当头部左右移动时，更近的物体不会更快移动。可以对图像执行简单的变换，以便增强深度感知？有限的深度数据甚至可以从图像中自动提取，可以大大提高视差和深度感知？另一个问题是设计光球内的接口。假设我们想与球体内的其他用户分享经验，在这种情况下，我们如何看待插入球体的虚拟物体，如菜单或头像？虚拟激光指针如何工作以选择物体？



图 7.25: Figure Digital 的 Pantopticam 原型使用数十个相机来提高逼近更多视点的能力, 从而可以模拟立体观看和位置变化引起的视差。

### 全景的光场

全景图像构造简单, 但显然有缺陷, 因为它们没有考虑从用户移动可能获得的任何视角环绕世界将如何出现。为了准确地确定这一点, 理想的情况是捕获用户允许移动的任何观察体积内的整个光场能量。光场提供空间每个点处的光传播的光谱功率和方向。如果用户在穿着 VR 头盔时能够在物理世界中走动, 那么这似乎是不可能完成的任务。一台摄像机如何在整个房间的同一时刻在所有可能的位置捕获光能? 如果用户被限制在一个小区域内, 那么光场可以被布置在球体上的一系列摄像机大致捕获, 图 7.25 显示了这样一个原型。在这种情况下, 可能需要数十个相机, 并且使用图像变形技术来近似相机之间或来自球形内部的视点。为了进一步改善体验, 光场照相机(也称为全光照相机)能够捕捉光线的强度以及它们在太空中行进的方向。这提供了许多优点, 例如在光场已被捕获后将图像重新聚焦到不同的深度。

### 拓展阅读

VR 和计算机图形之间存在紧密的联系, 因为两者都需要将可视信息推送到显示器上, 然而, 许多细微的差异出现了, 而 VR 的发展也相当不够。对于基本的计算机图形学, 许多文本提供了本章主题的额外报告, 例如参见[202]。有关计算机图形的高性能, 高分辨率渲染的更多细节, 请参见[5]。除[5]外, BRDF 的全面报告见于[22]。

光线追踪范例可能需要重新设计 VR。从计算几何角度来看有用的算法背景可以在[336,42]中找到。有关光学畸变和校正背景, 请参阅[46,117,130,199,326,330]。色差校正出现在[230]中。[33, 299, 317]中介绍了全景图的自动拼接。

# 第八章 真实世界和虚拟世界中的运动

到目前为止，关于运动的讨论仅限于一些特定的话题。5.3 节讲述了眼球运动，6.2 节涵盖了运动的感受，第 3 章则说明了注视点是如何改变的，但未对运动进行精确的数学描述。本章中我们则想更准确地对运动进行建模，因为真实世界和虚拟世界的物理效果都会影响 VR 体验。运动物体的加速度和速度会影响 VWG 中的模拟以及用于捕捉真实世界中用户运动的跟踪方法。因此，本章提供了运动方面的一些基础知识，更深层次的内容将在第 9 章和第 10 章进行介绍。

8.1 节介绍数学和物理学的基本概念，包括速度，加速度和刚体运动。8.2 节介绍了人类前庭系统的生理学和感知问题，它可以感知速度和加速度。8.3 节描述了如何在 VWG 中描述和制作动作，其中包括数值积分和碰撞检测。8.4 节着重介绍视觉，这是因为视觉和前庭系统之间的感觉冲突是 VR 疾病产生的根源，即在前庭系统未被欺骗时，眼睛可能会感知到运动。这可以被认为是对真实和虚拟世界物理现象之间的竞争。

## 8.1 速度和加速度

### 8.1.1 一维世界

我们从最简单的情况开始，如图 8.1 所示。设想一个只能在垂直方向上运动的 1D 世界。设  $y$  为移动点的坐标。它在任何时间  $t$  的位置由  $y(t)$  表示，这意味着  $y$  是时间的函数，同时  $y$  又是一个动点，每秒都会产生许多数值。

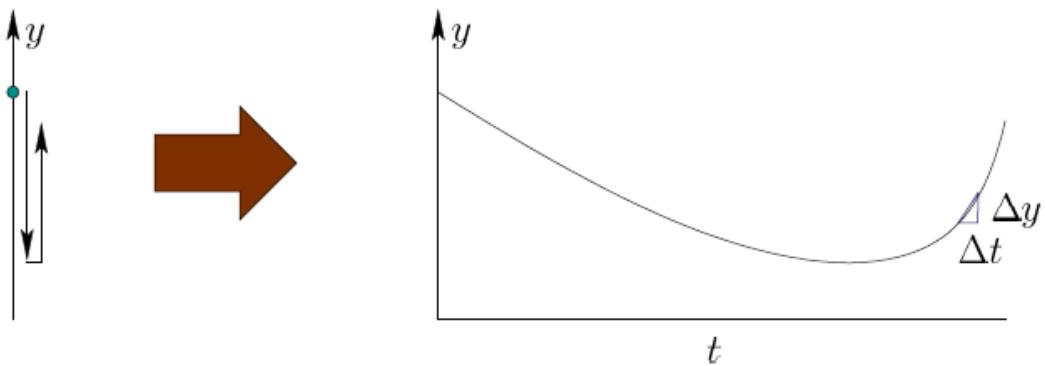


图 8.1：在一维世界中移动的一个点。

### 速度

那么这个点的移动有多快呢？通过微积分，它的速度  $v$  被定义为  $y$  关于时间的导数：

$$v = \frac{dy(t)}{dt} \quad (8.1)$$

使用数值计算， $v$  近似等于  $\Delta y / \Delta t$ ，其中  $\Delta t$  表示时间上的微小变化：

$$\Delta y = y(t + \Delta t) - y(t) \quad (8.2)$$

换句话说， $\Delta y$  是从时间开始变化到结束时  $y$  的变化。速度  $v$  可以用来估计  $y$  的变化：

$$\Delta y \approx v \Delta t \quad (8.3)$$

这种近似估计会随着  $t$  到  $t + \Delta t$  时间内  $\Delta t$  的减小和  $v$  本身变化的减小而改善。

当然，速度可能随时间的推移而改变。而任意时刻的位置可以通过积分计算得到：

$$y(t) = y(0) + \int_0^t v(s) ds \quad (8.4)$$

其中，假设  $y$  开始移动的时刻为  $t = 0$ 。如果  $v(t)$  对于任意时刻是恒定的，则用  $v$  表示，此时  $y(t) = y(0) + vt$ 。（8.4）中的积分考虑到了  $v(t)$  变化的情况。

## 加速度

下一步是通过数学方式描述速度的变化，即加速度  $a$ 。它被定义为：

$$a = \frac{dv(t)}{dt} \quad (8.5)$$

该形式与（8.1）相同，只是  $y$  被  $v$  替换。可以近似地生成近似值。近似计算的表达式也相似，即  $\Delta v \approx a \Delta t$ 。加速度本身也可以随时间变化，变为时间函数  $a(t)$ 。以下等式则为加速度和速度的积分关系：

$$v(t) = v(0) + \int_0^t a(s) ds. \quad (8.6)$$

由于加速度会变化，因此你可能会怀疑这样的命名过程是否会一直继续下去。然而在大多数情况下，高阶衍生单位是没有必要的。其中一个主要原因是经典物理学的运动通过力和加速度便能充分表示。例如，牛顿第二定律指出  $F = ma$ ，其中  $F$  是作用在一个点上的力， $m$  是物体质量， $a$  是加速度。

举一个简单例子，考虑重力加速度  $g = 9.8 \text{ m/s}^2$ ，某一点相对于地球以加速度  $g$  向下加速。使用（8.6）对其进行积分，则速度  $v(t) = v(0) - gt$ 。使用（8.4）的速度积分并假设  $v(0) = 0$ ，可以得到：

$$y(t) = y(0) - \frac{1}{2}gt^2 \quad (8.7)$$

## 8.1.2 3D 世界中的运动

### 点的移动

现在考虑 3D 世界中点的运动。想象一下 3.1 节中定义的几何模型随时间而移动。这会导致模型上每个点  $(x, y, z)$  的移动，从而每个点的坐标会是一个时间函数：

$$(x(t), y(t), z(t)). \quad (8.8)$$

因此 8.1.1 节中的速度  $v$  和加速度  $a$  须展开为三个方向的坐标。速度  $v$  由  $(v_x, v_y, v_z)$  代替以分别指示相对于  $x$ 、 $y$  和  $z$  轴的速度。该情况下， $v$  的大小称为速率：

$$\sqrt{v_x^2 + v_y^2 + v_z^2} \quad (8.9)$$

更进一步，加速度也会扩展为三个分量： $(a_x, a_y, a_z)$ 。

## 刚体运动

现在假设刚体在三维世界中移动。在这种情况下，所有点会一起移动。那么，我们怎样才能简洁地形容这一运动呢？回想一下 3.2 节的平移或旋转。首先，考虑一个简单的例子，假设旋转是禁止的，只允许刚体在空间中平移。在这种设定下，知道刚体上一点的位置便足以确定刚体上所有点随时间的位置。如果有一点已经改变其位置至 $(x_t, y_t, z_t)$ ，则所有点都进行了相同的位移。更重要的是，每个点的速度和加速度也都是相同的。

一旦允许旋转，这种简单的行为不会总是发生。当刚体旋转时，各点不再保持相同的速度和加速度。这对于理解 8.4 节中的 VR 疾病以及追踪方法如何估算位置和方向来说至关重要，具体细节将在第 9 章中讨论。



图 8.2：(a) 以恒定角速度  $\omega$  旋转的“木马”。(图片来自西雅图公园和娱乐中心)(b) 从上向下的视图中，旋转木马上某一点的速度矢量  $v$  与包含它的圆相切；该圆以旋转轴为中心，加速度矢量  $a$  指向中心。

## 角速度

为了理解这些问题，考虑旋转木马的简单情况，如图 8.2 (a) 所示。它的方向在任何时候都可以用  $\theta(t)$  来描述；见图 8.2 (b)。令  $\omega$  表示其角速度：

$$\omega = \frac{d\theta(t)}{dt} \quad (8.10)$$

默认情况下， $\omega$  的单位是弧度/秒。如果  $\omega = 2\pi$ ，则刚体在一秒钟后返回至相同的位置。假设  $\theta(0) = 0$ ，并且  $\omega$  是常数，则时间  $t$  处的方向由  $\theta = \omega t$  确定。为描述刚体上一个点的运动，使用极坐标  $r$  和  $\theta$  会非常方便：

$$x = r \cos \theta \text{ and } y = r \sin \theta. \quad (8.11)$$

$\theta = \omega t$  代入得到：

$$x = r \cos \omega t \text{ and } y = r \sin \omega t \quad (8.12)$$

再对时间进行求导，得：

$$v_x = -r\omega \sin \omega t \text{ and } v_y = r\omega \cos \omega t \quad (8.13)$$

这里的速度是一个二维矢量，且该矢量与包含点  $(x, y)$  的圆相切，见图 8.2 (b)。

这是非常直观的，因为这一点正是朝着这个方向前进；然而方向也很快就会改变，并始终沿着圆移动。这种速度变化意味着产生了非零加速度。点  $(x, y)$  的加速度通过再次求导获得：

$$a_x = -r\omega^2 \cos \omega t \text{ and } a_y = -r\omega^2 \sin \omega t. \quad (8.14)$$

结果是一个指向圆心的二维加速度矢量（图 8.2 (b)），中心即是圆的旋转轴。这就是所谓的向心加速度。如果你站在那一点，你会感觉到一个相反的方向拉力，好像是“木马”试图把你从中心扔掉。这也正是旋转的空间站产生非自然重力的原因。

### 3D 角速度

现在考虑 3D 刚体的旋转。回第 3.3 节，欧拉旋转定理意味着每个 3D 旋转都可以描述为围绕一个轴  $v = (v_1, v_2, v_3)$  旋转了角度  $\theta$ 。由于刚体的方向在短时间内就发生变化，所以想象一下与旋转变化相对应的轴。在旋转木马的情况下，该轴向量将是  $v = (0, 1, 0)$ 。对于更一般的情况， $v$  可以是任何单位向量。

因此 3D 角速度表示为 3D 矢量：

$$(\omega_x, \omega_y, \omega_z). \quad (8.15)$$

这可以从 2D 情况中延伸而来，即取出 2D 情况下的  $\omega$  并将其乘以矢量  $v$ ，并根据坐标轴位置对各分量进行加权。因此，这些分量可以被认为是  $\omega_x = \omega_{v_1}$ ， $\omega_y = \omega_{v_2}$  和  $\omega_z = \omega_{v_3}$ 。

$\omega_x$ ， $\omega_y$  和  $\omega_z$  分量分别对应于俯仰角，偏航角和翻滚角的旋转速率。由于非交叉性和运动奇点，我们避免了 3.3 节中的这类表示。但事实证明，对于速度而言这些问题并不存在[303]。因此，我们可以在这个阶段暂时不使用四元数。

### 角加速度

如果允许  $\omega$  随时间变化，那么我们必须考虑角加速度。在 2D 情况下，这被定义为

$$\alpha = \frac{d\omega(t)}{dt} \quad (8.16)$$

对于三维情况，有三个分量，其表达为：

$$(\alpha_x, \alpha_y, \alpha_z). \quad (8.17)$$

三个分量可以分别解释为俯仰角，偏航角和翻滚角的加速度。

## 8.2 前庭系统

如 2.3 节所述，平衡感（或前庭感）向大脑提供有关头部如何定向或如何移动的信息。这是通过测量头部线性和角加速度的前庭器官完成的。这些器官连同其相关的神经通路，共同被称为前庭系统。该系统对涉及运动的身体功能起着至关重要的作用，从步行或跑步等普通活动到需要大量训练的活动（如体操或芭蕾舞）。回忆 5.3 节，它还可以通过 VOR 使眼球运动独立于头部运动。

前庭系统对于 VR 来说很重要，因为它常会被忽视，并导致感受的不匹配（请回忆 6.4 节中的这个问题）。当前 VR 系统中，没有任何工程设备可以将前庭信号呈现给显示器，从而精确地刺激前庭器官所应达到的值。电流式前庭刺激或许存在一些可行性，这种刺激为器官提供电刺激[80,79]。然而，这种技术可能需要更多年的研究才足够准确以及舒适，最终被批准以供大众安全使用。另一种可能性是通过低频振动来刺激前庭系统，但这或多或少会分散人们的注意力。

## 生理学

图 8.4 显示了人体头部前庭器官的位置。其像眼睛和耳朵那样，有两个对称的器官，分别位于右侧和左侧。图 8.3 显示了每个前庭器官的生理机能。耳蜗处理之后 11.2 节所述的听力部分，其余部分属于前庭系统。椭圆囊和球囊测量线性加速度；它们共同形成耳石系统。当头部未倾斜时，椭圆囊的感测表面大部分位于水平平面（或我们常见坐标系中的  $xz$  平面），而球囊的相应表面位于垂直平面中，被称为矢状面，或  $yz$  平面）。该情况下，椭圆囊感知加速度分量  $a_x$  和  $a_z$ ，球囊感知  $a_y$  和  $a_z$ 。

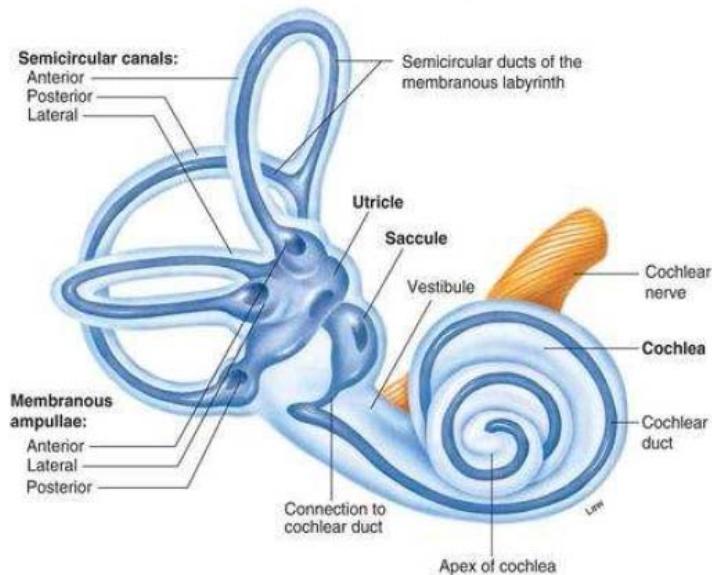


图 8.3：前庭器官

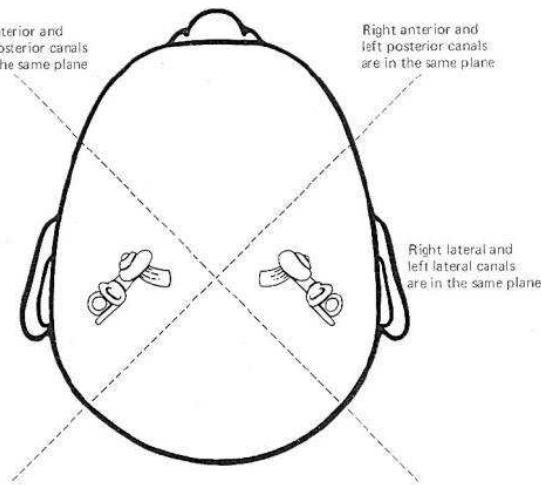


图 8.4：前庭器官位于耳后。（图片来自 CNS Clinic Jordan）

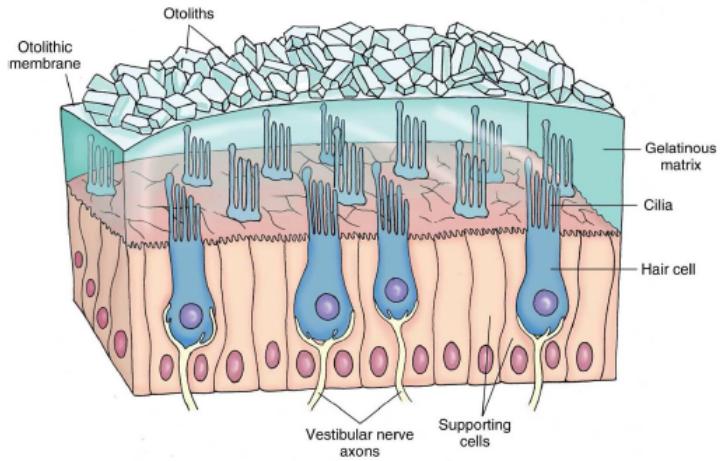


图 8.5：感知线性加速度的耳石器官（椭圆囊或球囊）

半规管测量角加速度。每个通道具有约 0.2 至 0.3mm 的直径，并且沿着直径约为 2 至 3mm 的圆弧弯曲。令人惊讶的是，这三条半规管基本上两两垂直的，因此它们可以分别测量角速度的三个分量。三个半规管的特定名称分别为前管，后管和侧管。它们与我们通常默认的 3D 坐标轴不完全一致。从图 8.4 中注意到，每组半规管相对于垂直轴旋转了 45 度。因此，左耳的前管与右耳的后管对齐。同样，左耳的后管与右耳的前管对齐。虽然在图中没有显示，但横向的半规管也偏离水平面 30 度左右。即使这样，角加速度的所有三个分量仍可以被感测到，因为这些半规管之间是互相垂直的。

### 线性加速度的感知

为了理解加速度如何被感知，我们从耳石系统的构造开始理解。图 8.5 为耳石器官的示意图，该器官可以是椭圆囊或球囊。机械性感受器以毛细胞的形式将加速度转化为神经信号。每个毛细胞都含有嵌入凝胶状基质的纤毛。重量集中于纤毛顶部，因此当产生横向加速度时，移位的重量会产生剪切力，导致纤毛弯曲。加速度越高，弯曲越大，同时神经脉冲的传输速率越高。横向偏转在两个维度均是有可能发生的。例如在椭圆囊的情况下， $xz$  平面内任何方向的线性加速度都会导致纤毛弯曲。为了区分该平面内的特定方向，纤毛被极化，使得每个细胞对一个特定方向敏感。这是通过较厚的动纤毛来实现的，而细胞的所有其他纤毛都通过一种带状物穿过尖端实现连接，从而它们可以全部弯曲在一起。

一种主要的感知限制是由于物理学的一个基本规律而来：爱因斯坦等价原理。除了前庭系统外，它还会影响 VR 追踪系统（见第 9.2 节）。该问题的关键在于重力。如果我们在太空深处，远离任何重力，那么由传感器测量的线性加速度将是相对于固定坐标系的纯加速度。在地球上，我们则受到重力，就好像我们在一艘火箭上以大约 9.8 米/秒的速度向上加速。等价原理指出，重力和真正的线性加速度对身体的影响是难以区分的。图 8.6 显示了耳石器官的感知结果，无论头部是倾斜还是线性加速，都会以相同的信号发送到大脑。如果闭上眼睛或戴上 VR 耳机，则无法区分倾斜和加速。在大多数情况下，我们并不会被这种现象所欺骗，因为前庭信号在加速时会伴随着其他刺激，如视觉和转动信号。

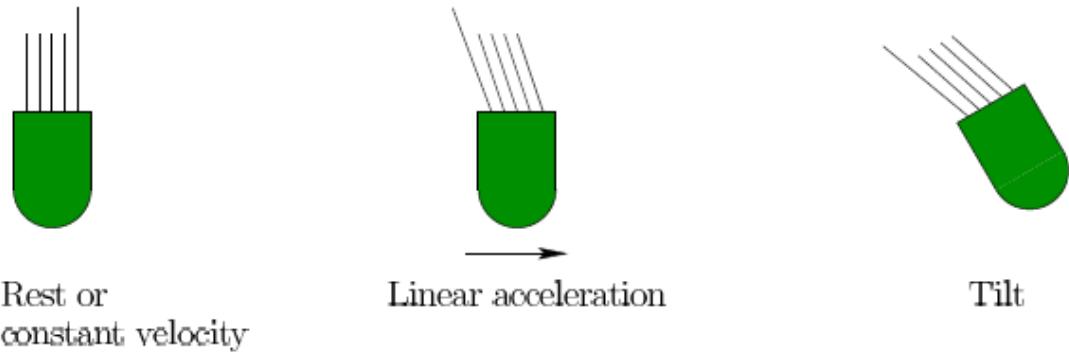


图 8.6: 由于爱因斯坦等价原理,耳石器官无法区分头部相对于重力的线性加速度与倾斜加速度。无论哪种情况,纤毛都以相同的方式偏转,向神经结构发送相同的信号。

### 感应角加速度

半规管的感知原理与耳石器官相同。它们通过在毛细胞末端弯曲纤毛来测量加速度。粘性流体在每个通道内移动。一个叫做壶腹帽的弹性结构会阻塞半规管的一小部分,并包含毛细胞;见图 8.7。比较半规管与木马的旋转。如果我们在旋转木马的周围放置充满液体的管子,那么流体将以恒定的角速度保持相当稳定的状态。然而,如果施加角加速度,由于流体与管道之间的摩擦(以及内部流体粘度),流体将开始在管内进行相对的移动。在半规管中,移动的流体向壶腹帽施加压力,使其变形并弯曲其内部毛细胞上的纤毛。请注意,恒定的角速度原则上不会对圆筒造成压力;因此,半规管测量的是角加速度而不是角速度。每个半规管极化的,它主要响应围绕垂直于包含整个半规管平面的轴的旋转。

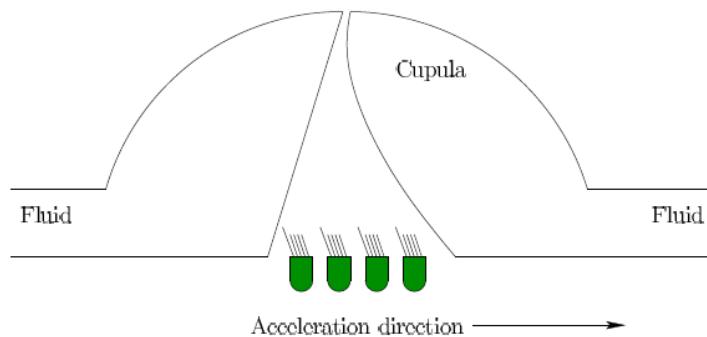


图 8.7: 壶腹帽包含一个容纳纤毛的中心膜。如果角加速度出现与导管方向一致,则压力施加到杯状物上,这导致纤毛弯曲并发送神经信号。

### 对感知的影响

来自前庭系统的感知响应与其他感觉相比通常较弱,尤其是视觉。同时,前庭系统在提供与其他系统(包括视觉和本体感觉)一致的线索时会产生良好的反应。不匹配的感知则会产生问题。例如,当前庭系统运行不正常时,有些人可能会感到眩晕。在这种情况下,他们会觉得周围的世界正在旋转或摇摆。常见的症状是恶心,呕吐,出汗和行走困难。由于 VOR,这甚至可能会影响眼球运动。后续的 8.4 节则会解释 VR 中前庭和视觉线索不匹配时所造成的不良反应。

## 8.3 虚拟世界中的物理

### 8.3.1 根据经验调整物理

如果我们想让我们的大脑相信我们生活在虚拟世界中，那么我们对现实世界的许多期望应该在虚拟世界中相匹配。我们已经看到了光的物理学（第 4 章）适用于虚拟世界的视觉渲染（第 7 章）。虚拟世界中的动作也应该以熟悉的方式表现。

这意味着 VWG 包含一个物理引擎，通过遵循物理世界的原则来管理虚拟世界中的物体的运动。作用于身体的力量，重力，流体流动和身体之间的碰撞应该以感性令人信服的方式处理。物理学引擎在整个工程和物理学的任何模拟环境中都会出现。在视频游戏，计算机图形和电影中，这些引擎执行的操作非常接近我们对 VR 的需求，这就是为什么 Unity 3D 和虚幻引擎等流行游戏引擎已经很快适应 VR 使用的原因。如第 2.2 节所述，我们还没有到达一个广泛采用 VR 引擎的时代。因此，现代游戏引擎目前还需要好好理解和运用。

为了确定什么样的物理引擎需要借鉴，改编或者从头构建，应该考虑期望的 VR 体验并确定将出现的运动类型。一些常见的一般问题是：

- 匹配区域是否保持固定，还是需要通过移动来移动用户？如果需要运动，那么用户是否会步行，跑步，游泳，驾驶汽车或飞行太空船？
- 用户是否会与对象进行交互？如果是这样，那么需要什么样的交互？可能性包括携带武器，打开门，投掷物体，倒饮料，操作机械，绘画和组装结构。
- 多个用户将共享相同的虚拟空间吗？如果是这样，那么他们的议案将如何协调或约束呢？
- 虚拟世界是否包含自行移动的实体，例如机器人，动物或人类？
- 用户是否会沉浸在熟悉或陌生的环境中？熟悉的环境可能是家庭，教室，公园或城市街道。一个充满异国情调的环境可能是水肺潜水，月球探索或穿过血管。

除了物理引擎之外，这些问题也将指导接口的设计，这将在第 10 章中介绍。

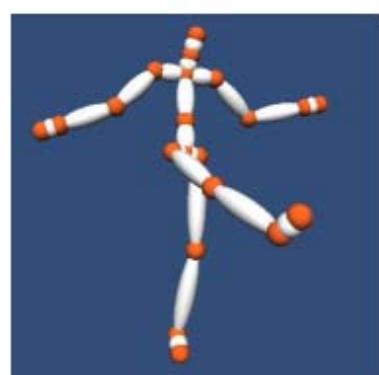
基于对上述问题的回答，物理引擎设计可能简单而高效，或者完全压倒性的。正如第 7.4 节所述，一个关键的挑战是保持经常更新的虚拟世界，以便用户和对象之间的交互很好地同步，并且为显示器提供低延迟投影。

请注意，目标可能并不总是完美匹配在现实世界中会发生的事情。在熟悉的环境中，我们可能会期望重要的匹配；然而，在陌生的环境中，做出舒适体验往往变得更加重要，而不是完美匹配现实。即使在模拟自己在世界上散步的情况下，我们也经常希望偏离现实世界的物理学，因为它会导致 VR 晕眩（见 8.4 节）。

本节的其余部分涵盖了一些常见的基本方面：1) 物理系统的数值模拟，2) 使用人类输入的系统的控制，3) 碰撞检测，它确定身体是否相互干扰。



(a)



(b)

图 8.8: (a) 在侠盗猎车手游戏中出现的虚拟车（猎豹）；它应该拥有多少自由度？(b) 通过关节连接刚体的人体骨骼通常是化身运动的基础。(图来自于 SoftKinetic)

### 8.3.2 数值模拟

#### 虚拟世界的状态

想象一个包含许多移动刚体的虚拟世界。对于每个物体，请考虑其自由度 (DOF)，这对应于唯一确定其位置和方向所需独立参数的数量。我们想知道在一次性瞬间将每个身体置于适当位置所需的所有参数列表。所有这些参数的值的规范被定义为虚拟世界的状态。然后可以将物理引擎的工作描述为计算渲染系统所需要的每个时刻或虚拟世界的“快照”的虚拟世界状态。一旦确定了状态，第 3 章的数学变换就被用来将身体正确地放置在世界中并计算它们应该如何在显示器上出现。

#### 自由度

虚拟世界模型中有多少个参数？如第 3.2 节所述，不受约束的物体具有 6 个自由度，这意味着 6 个参数可将其放置在任何位置。在很多情况下，自由度因约束而丢失。例如，在地面上滚动的球只有 5 个自由度，因为它可以实现沿地面的任何 2D 位置，并且也具有任何 3D 方向。通过指定沿着地面的位置（两个参数）和它面对的方向（一个参数）来描述具有 3 个自由度的汽车可能是足够的；见图 8.8 (a)。但是，如果赛车在空中飞行时可以在空中飞行或撞击，那么所有 6 个自由度都是需要的。

对于许多型号，刚体以允许相对运动的方式连接在一起，这被称为多体运动学[163,303]。例如，一辆汽车通常有 4 个车轮，每个车轮可以滚动以提供一个旋转的自由度。此外，前轮可以转向提供额外的自由度。转向通常会使前轮一致，这意味着一个自由度足以描述两个车轮。如果汽车有一个复杂的悬挂系统，那么它不能被视为一个单一的刚体，这将增加更多的自由度。

同样，动画角色可以通过附加刚体形成骨架来制作；见图 8.8 (b)。骨架中的每个刚体通过关节连接到一个或多个其他物体。例如，可以通过将胳膊、腿和脖子连接到刚性躯干来形成简单的人物角色。左上臂通过肩关节连接到躯干，手臂的下部通过肘关节连接，依此类推。有些关节允许比其他关节更多的自由度。例如，肩关节有 3 个自由度，因为它可以相对于躯干偏航、俯仰和滚动，但肘关节只有一个自由度。

为了充分模拟人体的灵活性，需要 244 个自由度，由 630 块肌肉控制[366]。在很多情况下，这有着太多的细节，可能导致高计算复杂性而难以实现。此外，人们应该时刻提防意外出现的未知因素（在 1.1 节中提到），其中更多的现实主义可能会导致人物感知的蠕变增加。因此，拥有更多自由度并不是明显更好，而是由 VR 内容创建者决定需要多少机动性才能使角色变为现实，并且以一种有针对性的目的引人注目。

在极端情况下，刚体不足以塑造世界。我们可能希望看到波浪在湖面上真实地荡漾起来，或者在微风中微微流淌的头发。在这些通用设置中，使用非刚性模型，在这种情况下，可以将状态视为连续函数。例如，形式为  $y=f(x, z)$  的函数可以描述水的表面。没有进行一些限制简化，结果可能会有无限多个自由度。这种设置中的运动通常使用偏微分方程 (PDE) 来描述，其在数值上被整合以获得期望时间的状态。通常，计算成本足够高，以至于禁止将它们用于交互式 VR 体验，除非通过预算算运动或快速简化模型来实现快捷方式。

#### 微分方程

我们现在介绍一些基本的微分方程来模拟运动。结果描述通常被称为动态系统。第一步

是根据状态描述刚体速度。回到涉及一个或多个刚体的模型，该状态对应于有限数量的参数。让

$$x = (x_1, x_2, \dots, x_n) \quad (8.18)$$

该式表示  $n$  维状态向量。如果每个  $x_i$  对应于刚体的位置或方向参数，则状态向量将所有物体置于其位置。让

$$\dot{x}_i = \frac{dx_i}{dt} \quad (8.19)$$

表示每个参数的时间导数或速度。

为了在任何时间  $t$  获得状态，速度需要随时间积分。在 (8.4) 之后，每个状态变量的积分决定了时间  $t$  的值：

$$x_i(t) = x_i(0) + \int_0^t \dot{x}_i(s) ds, \quad (8.20)$$

其中  $x_i(0)$  是时间  $t=0$  时  $x_i$  的值。

式 (8.20) 存在两个主要问题：

1. 积分几乎总是必须用数字来评估。
2. 速度  $x_i(t)$  必须在每个时间  $t$  被指定。

采样率对于第一个问题，时间被离散化为步骤，其中  $\Delta t$  是步长或采样率。例如， $\Delta t$  可能是 1ms，在这种情况下，可以计算时间  $t=0, 0.001, 0.002, \dots$  的状态，以秒为单位。这可以被认为是物理引擎的一种帧频。每个  $\Delta t$  对应于新帧的产生。

如第 7.4 节所述，VWG 应将虚拟世界框架的生成与渲染过程同步，以便世界不会陷入中间状态，其中一些变量更新为新时间，而其他变量则停留在前一时间。这是虚拟世界中的一种撕裂。然而，这并不意味着渲染器和物理引擎之间的帧速率是相同的。通常情况下，物理引擎的帧速率要高得多以提高数值精度。使用采样率  $\Delta t$ ，(8.20) 近似为

$$x_i((k+1)\Delta t) \approx x_i(k\Delta t) + \sum_{j=1}^k \dot{x}_i(j\Delta t) \Delta t, \quad (8.21)$$

对于每个状态变量  $x_i$ ，(8.21) 一次一个步骤更简单。令  $x_i[k]$  表示  $x_i(k\Delta t)$ ，它是时间  $t=k\Delta t$  时的状态。以下是根据旧状态  $x_i[k]$  表示新状态  $x_i[k+1]$  的更新定律：

$$x_i[k+1] \approx x_i[k] + \dot{x}_i(k\Delta t) \Delta t, \quad (8.22)$$

从  $x_i[0] = x_i(0)$  开始。

## 龙格-库塔数值积分法

(8.21) 中使用的近似称为欧拉积分。这是最简单的近似，但在许多实际环境中表现不佳。最常见的改进之一是第四代 Runge-Kutta 积分方法，它将新状态表示为

$$x_i[k+1] \approx x_i[k] + \frac{\Delta t}{6} (w_1 + 2w_2 + 2w_3 + w_4), \quad (8.23)$$

其中

$$\begin{aligned}
w_1 &= f(\dot{x}_i(k\Delta t)) \\
w_2 &= f(\dot{x}_i(k\Delta t + \frac{1}{2}\Delta t) + \frac{1}{2}\Delta t w_1) \\
w_3 &= f(\dot{x}_i(k\Delta t + \frac{1}{2}\Delta t) + \frac{1}{2}\Delta t w_2) \\
w_4 &= f(\dot{x}_i(k\Delta t + \Delta t) + \Delta t w_3).
\end{aligned} \tag{8.24}$$

虽然这比欧拉积分代价更高，但实际上提高的准确度通常是值得的。存在许多其他方法，其性能取决于  $\dot{x}$  表达式特定方式并随时间而变化[133]。

## 时不变的动态系统

(8.20) 的第二个问题是确定  $\dot{x}(t)$  的表达式。这就是物理定律，例如由于施加的力和重力引起的刚体加速。最常见的情况是时不变的动态系统，其中  $\dot{x}$  仅取决于当前状态而不取决于特定时间。这意味着每个组件  $x_i$  被表示为

$$\dot{x}_i = f_i(x_1, x_2, \dots, x_n), \tag{8.25}$$

对于某个给定的向量值函数  $f = (f_1, \dots, f_n)$ 。这可以通过使用  $x$  和  $\dot{x}$  来表示  $n$  维向量以压缩形式来表示：

$$\dot{x} = f(x). \tag{8.26}$$

上面的表达式通常称为状态转换方程，因为它表示状态的变化率。

这是一个简单的一维状态转换方程的例子：

$$\dot{x} = 2x - 1. \tag{8.27}$$

这被称为线性微分方程。速度  $x$  大致与  $x$  的值相加。幸运的是，线性问题可以在“纸上”完全解决。(8.27) 的解决方案是一般形式

$$x(t) = \frac{1}{2} + ce^{2t}, \tag{8.28}$$

其中  $c$  是一个常数，取决于  $x(0)$  的给定值。

## 相空间

不幸的是，运动通常用加速度（有时候是高阶导数）来描述，需要两次积分。这会产生高阶微分方程，难以处理。出于这个原因，相空间表示是在物理学和工程学中发展起来的。在这种情况下，状态变量的速度本身被视为状态变量。这样，加速度就成为速度变量的速度。

例如，假设位置  $x_1$  受重力作用，产生加速度  $a = -9.8 \text{ m/s}^2$ 。第二个变量  $x_2$ ，它被定义为  $x_1$  的速度。因此，根据定义， $\dot{x}_1 = x_2$ 。此外， $\ddot{x}_2 = a$ ，因为速度的导数是加速度。这两个方程都符合 (8.25) 的形式。通常情况下，状态的数量会随着加速度的增加而增加（或者甚至是更高阶的导数），但是结果的动态性以 (8.25) 的形式表示，这更容易处理。

## 处理用户输入

现在考虑用户命令对象移动的情况。例子包括驾驶汽车，飞行太空船，或者漫步一个化身。这引入了一些新的参数，称为控制，动作或动态系统的输入。包含这些新参数的微分方程称为控制系统[14]。

令  $u = (u_1, u_2, \dots, u_m)$  为控件的向量。(8.26) 中的状态转移方程简单地扩展为包含  $u$

的方程:

$$\dot{x} = f(x, u). \quad (8.29)$$

图 8.9 显示了一个有用的例子, 它举了驾驶汽车的例子, 来说明如何决定汽车的速度。例如,  $u_s=1$  正向驱动,  $u_s=-1$  反向驱动。为了得到更快的速度, 一共设置有 10 个驱动器。控制器确定前轮如何转向。状态向量是  $(x, z, \theta)$ , 它对应于汽车在水平  $xz$  平面中的位置和方向。

状态转换方程是:

$$\begin{aligned}\dot{x} &= u_s \cos \theta \\ \dot{z} &= u_s \sin \theta \\ \dot{\theta} &= \frac{u_s}{L} \tan u_\phi.\end{aligned} \quad (8.30)$$

使用 Runge-Kutta 积分或类似的数值方法, 考虑到随着时间的推移控制  $U_s$  和  $u\varphi$ , 可以计算汽车的未来状态。

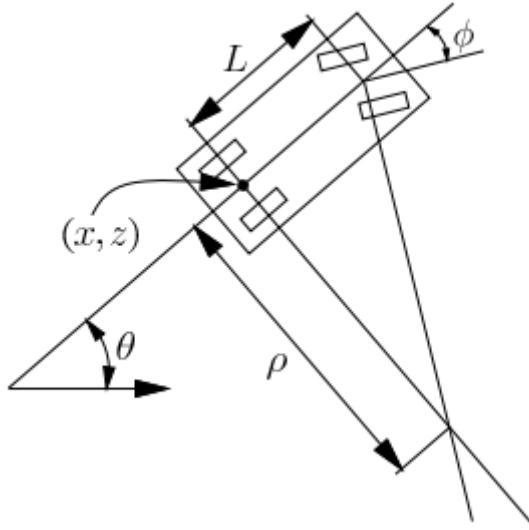


图 8.9: 简单的可操纵汽车的俯视图。它的位置和方向由  $(x, y, \theta)$  给出。参数  $\rho$  是最小转弯半径, 其取决于最大允许转向角  $\varphi$ 。通过将视角放置在后轴中心的上方, 该模型也可用于“引导”人物化身。

该模型也可用于从第一人称视角引导 VR 用户的虚拟步行。然后视点根据  $(x, z, \theta)$  改变, 而高度  $y$  保持固定。对于 (8.30) 中的模型, 汽车必须前进或后退才能改变其方向。通过将第三个分量更改为  $\theta=u\omega$ , 用户可以直接指定角速度。这会导致用户旋转到位, 就像旋转木马一样。这些更多的例子出现在[163]的第 13 章中, 包括通过加速控制的物体。

根据离散时间点来定义运动, 这在概念上有时很有用, 称为阶段。使用 (8.29) 的数值积分, 我们可以考虑在时间  $\Delta t$  上应用控制  $u$  以获得新的状态  $x[k+1]$ :

$$x[k+1] = F(x[k], u[k]). \quad (8.31)$$

函数  $F$  通过在  $\Delta t$  上积分 (8.29) 获得。因此, 如果状态是  $x[k]$ , 并且  $u[k]$  被应用, 则  $F$  计算  $x[k+1]$  作为下一阶段的状态。

### 8.3.3 碰撞检测

构建物理引擎的最大挑战之一是处理机体之间的碰撞。来自物理学或工程学的标准运动定律通常不考虑这种相互作用。因此，专门的算法用于检测何时发生这种碰撞并作出适当的响应。由于计算机图形仿真和视频游戏的广泛需求以及机器人的运动规划，碰撞检测方法和相应的软件非常丰富。

## 实体或边界模型？

图 8.10 显示了一个 2D 世界中两个三角形碰撞检测的第一个困难。前两种情况（图 8.10 (a) 和 8.10 (b)）显示明显的情况；然而，第三种情况，图 8.10 (c) 可能不明确。如果一个三角形完全位于另一个三角形内部，那么这是一次碰撞吗？如果我们将外三角解释为实体模型，那么是的。如果外三角形只是边界边缘，并且意味着内部是空的，那么答案是否定的。这就是为什么强调 3.1 节中有一个连贯的模型；否则，边界可能建立得不够好，无法区分内部和外部。

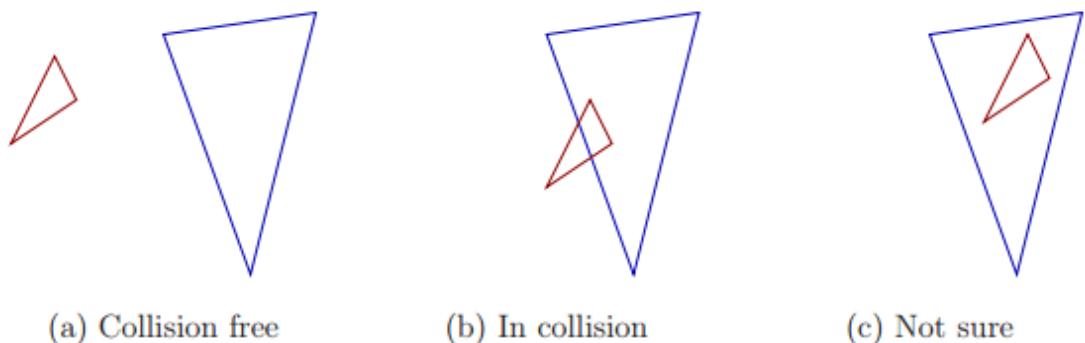


图 8.10：碰撞检测的三个有趣案例（这些是 2D 示例）。最后一种情况可能会或不会导致碰撞，具体取决于型号。

## 距离函数

许多碰撞检测方法受益于维持一个距离函数，该距离函数可以跟踪物体距离碰撞的距离。例如，让  $A$  和  $B$  表示由两个不同模型在  $\mathbb{R}^3$  中占用的所有点的集合。如果它们处于碰撞中，那么它们的交点  $A \cap B$  不是空的。如果它们没有发生碰撞，那么  $A$  和  $B$  之间的 Hausdorff 距离是最接近的一对点之间的欧几里德距离，从  $A$  和  $B$  中取一个。让  $d(A, B)$  表示这个距离。如果  $A$  和  $B$  相交，那么  $d(A, B) = 0$ ，因为  $A \cap B$  中的任何点将产生零距离。如果  $A$  和  $B$  不相交，那么  $d(A, B) > 0$ ，这意味着它们没有发生碰撞（换句话说，没有碰撞）。如果  $d(A, B)$  很大，那么即使其中一个或两个正在移动， $A$  和  $B$  在不久的将来也很可能无碰撞。这导致了一系列称为增量距离计算的碰撞检测方法，该方法假定在对算法的连续调用之间，身体只移动一小部分。在这个假设下，算法对于凸多面体[181,214]实现“几乎恒定的时间”性能。非凸体可以分解为凸面组件。

与距离有关的概念是穿透深度，这表明一个模型在另一个模型中的位置有多远[182]。这对设置允许两个物体之间有多少干扰的阈值很有用。例如，用户可能能够将他的头部戳到墙壁两厘米处，但再靠近点的话就应该采取避让措施。

## 简单的碰撞测试

在最低级别，碰撞检测通常需要测试一对模型基元来确定它们是否相交。在 3D 模型形成模型的情况下，我们需要一种方法来确定两个三角形是否相交。这与第 7.1 节中视觉渲染所需的光线三角形相交测试类似，并且涉及解析几何中的基本工具，例如交叉积和平面方程。

有效的方法在[106,216]中给出。

## 广义狭义两阶段

假设一个虚拟世界已经被定义为数百万个三角形。如果两个复杂的非凸面体被检查碰撞，那么计算成本可能很高。对于这种复杂的情况，碰撞检测往往成为一个两阶段过程：

1. 广义阶段：在广义阶段中，任务是避免为彼此远离的物体执行昂贵的计算。可以在每个物体周围放置简单的边界框，并且可以执行简单的测试以避免昂贵的碰撞检查，除非这些边框相交。在某些情况下，散列方案可以用来大大减少必须测试相交的盒子对的数量[215]。

2. 狹义阶段：在狭义阶段，单个模型部件对各自仔细检查碰撞。这涉及代价较高的测试，例如三角形 - 三角形交叉。

在广义阶段中，分层方法通常将每个身体分解成一棵树。树中的每个顶点都代表一个包含身体某个子集的边界区域。根顶点的边界区域包含整个主体。指导选择边界区域类型两个相反的标准：

1. 该区域应尽可能紧密地符合预期的模型要点。
2. 两个区域的相交测试应尽可能高效。

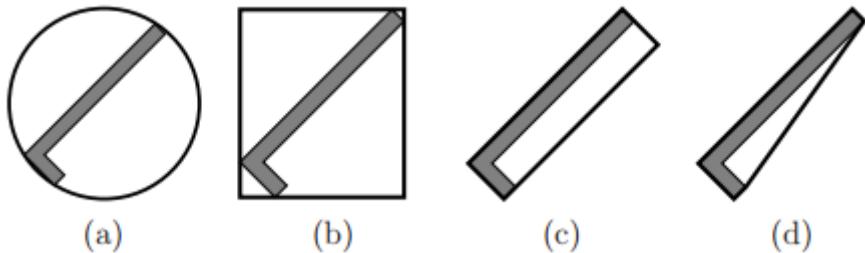


图 8.11：四种不同的边界区域：(a) 球体，(b) 轴对齐边界框（AABB），(c) 定向边界框（OBB）和 (d) 凸包。每种方法通常比前一种方法提供更严格的近似值，但测试与其他方法的交叉代价更高。

图 8.11 显示了几种流行的选择，例如 L 形机身。如 7.1 节所述，分层方法对于快速消除视觉渲染中考虑的许多三角形也很有用。

该树是为一个物体构建的，A (或者，B) 递归地如下构建。对于每个顶点，考虑包含在边界区域中的 A 中所有点的集合 X。通过定义两个较小的边界区域来构造两个子顶点，其结合覆盖 X. 分割是为了使每个孩子覆盖的部分具有相似的尺寸。如果几何模型由诸如三角形的图元组成，则可以进行分割以将三角形分成两组大致相同数量的三角形。然后为每个孩子计算边界区域。图 8.12 显示了一个 L 形主体情况下的分割示例。通过分割直到获得非常简单的组，递归地生成孩子。例如，对于空间中的三角形，除非顶点表示单个三角形，否则将进行拆分。在这种情况下，很容易测试两个三角形的交点。

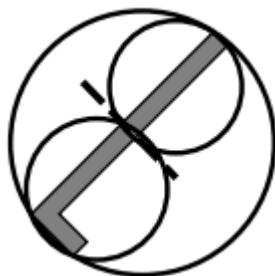


图 8.12：大圆圈显示了覆盖 L 形体的顶点的边界区域。在沿着虚线进行分割之后，使用两个较小的圆来覆盖身体的两个半部分。每个圆圈对应一个子顶点。

考虑确定身体 A 和 B 是否碰撞的问题。假设  $T_a$  和  $T_b$  树分别为 A 和 B 构建。如果  $T_a$  和  $T_b$  的根顶点的边界区域不相交，则知道  $T_a$  和  $T_b$  由于没有进行任何附加计算，则不会相互碰撞。如果边界区域确实相交，则将  $T_a$  的子区域的边界区域与  $T_b$  的边界区域进行比较。如果其中任何一个相交，则  $T_b$  的边界区域被其子区域的边界区域替换，并且该过程继续递归。只要边界区域相交，遍历树的较低层次，直到最终达到叶子。算法在叶子上测试单个三角形进行碰撞，而不是边界区域。请注意，随着树被遍历，如果  $T_a$  的顶点  $v_1$  的边界区域不与  $T_b$  的顶点  $v_2$  的边界区域相交，则  $v_1$  的子节点不必与  $v_2$  的子节点进行比较。通常，相对于测试所有交叉的三角形的方法，这大大减少了比较次数。

## 虚拟现实中不匹配的障碍

虽然碰撞检测是一个标准的，好解决的问题，但虚拟现实再一次带来了非同寻常的挑战。主要困难之一是匹配区域，其中真实世界和虚拟世界共享相同的空间。这造成了三个有趣的情况：

**真实障碍：**在这种情况下，真实世界存在障碍，但不存在于虚拟世界中。这是潜在的危险！例如，你可以移动你的手臂，碰一下真实的热咖啡，这是虚拟世界中没有的。如果你正在用 VR 耳机走路，那么想象一下如果一组真正的下楼梯没有出现，会发生什么。至少，如果用户接近匹配区域的边界，则应呈现边界。这种不匹配促使 HTC Vive 耳机中引入了 Chaperone 系统，使用朝外的相机来检测和渲染可能阻碍用户运动的真实物体。

**仅虚拟障碍：**这种情况并不危险，但可能非常影响体验。用户可以在虚拟现实环境中穿过墙壁，而不会在现实世界中感受到任何反应。这在大多数情况下不应该被允许。随着虚拟墙被联系，VWG 可以简单地停止移动虚拟世界中的观点；然而，这会在真实和虚拟运动之间产生不匹配，这对用户来说可能会感到不舒服。在试图引导他们摆脱对虚拟障碍的干扰时，保持用户的舒适性仍然是一项艰巨的挑战。

**真实和虚拟障碍：**如果障碍物在真实世界和虚拟世界中都相匹配，那么这种效果在感觉上是强大的。例如，您可能站在现实世界中略微升高的平台上，而虚拟世界显示您站在建筑物屋顶上。如果屋顶和平台边缘完美对齐，那么您可以用脚感受边缘。你会害怕走出边缘吗？更简单的情况是呈现与用户可能坐在的真正椅子相匹配的虚拟椅子。

## 8.4 不匹配的运动和观察

第 2.3 节提到的运动是一种由不同的视觉刺激引起的自我运动幻觉。换句话说，即使实际上没有发生运动，大脑也被欺骗地认为头部正在移动。图 2.20 显示了闹鬼的摇摆幻觉，它使上下颠倒的人们信服。然而，人们静止的时候房子在移动。在真实世界中没有相应运动的情况下，通过移动用户的视点，通常也会在 VR 中感知到运动。



图 8.13：由于世界上的运动而引起的图像中的特征光流。这些是使用图像处理算法自动计算的。（图来自于图宾根的马克斯普朗克智能系统研究所的 Andreas Geiger）

虚拟现实中带来的移动的认知因素是不匹配线索的一个主要例子，这在第 6.4 节中讨论过。尽管 McGurk 效应没有任何有害的副作用，但不幸的是，这个因素会让很多人都会遇到晕眩，恶心，偶尔甚至呕吐等疾病症状。因此，它应该非常谨慎地应用于 VR 体验。此外，如果使用它，应该尝试改变内容，以使副作用最小化。业界巨头经常宣称他们最新的 VR 设备已经击败了 VR 疾病问题；然而，这忽略了以下违反直觉的行为：

如果设备在空间分辨率，帧速率，跟踪精度，视场和延迟方面更好，那么由虚拟现实带来的移动认知因素和其他不匹配线索让人们感到不舒服的潜在性更高。

简单而直观地说，如果头盔能够更准确地模拟真实情况，那么感官线索就会更强大，而我们的感知系统会对错误的线索更加自信。它甚至可以模仿较差的设备，从而可以比较评估早期 VR 系统的副作用。在某些情况下，线索的不匹配可能是无害的（尽管可能导致降低存在感）。在其他情况下，由于大脑难以解决小的冲突，错配可能导致更大的疲劳。在最糟糕的情况下，VR 疾病会出现，因为虚拟现实中带来的移动的认知因素是当今 VR 体验的最大罪魁祸首。最糟糕的情况之一是第一人称射击游戏直接适用于虚拟现实，在虚拟形象探索敌对环境时，几乎一直会有虚拟现实中的移动。

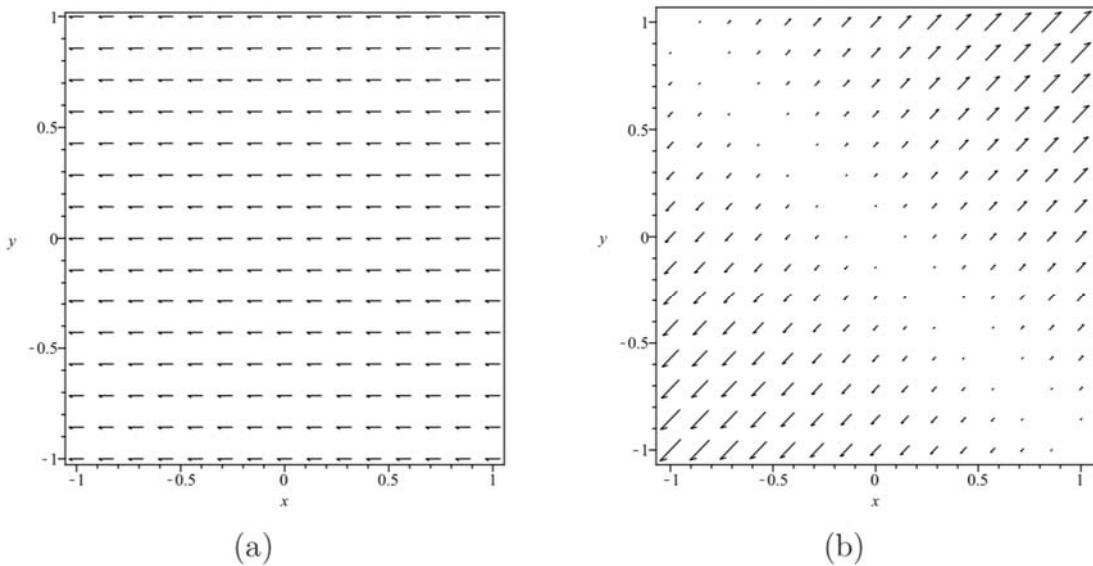


图 8.14：矢量场示例：(a) 一个常数矢量场，对于每个矢量  $(-1, 0)$ ，无论位置在哪里。 (b) 在这个矢量场中， $(x, y) \rightarrow (x + y, x + y)$ ，矢量指向远离  $(-1, 1)$  到  $(1, -1)$  的对角线，长度与距离成正比。

## 光流

从第 6.2 节回忆，人类视觉系统具有专门用于检测视场中视觉特征运动的神经结构；见图 8.13。它实际上是穿过视网膜的这些特征的图像。因此，有一个描述表面上移动点速度的数学概念是有用的。因此，我们定义了一个矢量场，它在表面上的每个点处分配一个速度矢量。如果曲面是  $xy$  平面，则是速度矢量

$$(v_x, v_y) = \left( \frac{dx}{dt}, \frac{dy}{dt} \right)$$

在每个点  $(x, y)$  被分配。例如，

$$(x, y) \mapsto (-1, 0)$$

是一个常量向量场，它每个地方都分配  $v_x = -1$  和  $v_y = 0$ ；见图 8.14 (a)。矢量场

$$(x, y) \mapsto (x + y, x + y)$$

不是常数，并且在每个点  $(x, y)$  处分配  $v_x = v_y = x + y$ ；见图 8.14 (b)。对于这个矢量场，速度方向总是对角线的，但矢量的长度（速度）取决于  $x + y$ 。

为了最准确地描述沿着视网膜的特征的运动，应该在与光感受器的位置相对应的球形表面上定义矢量场。相反，我们将在一个正方形区域描述矢量场，并且理解它应该被转换到一个球体上以获得更高的精度。

## 虚拟现实中移动的类型

虚拟现实中移动的认知因素可以由虚拟世界中视角的角速度和线速度的任意组合引起。为了有效地表征不同类型运动的影响，将视点速度分解为三个线性分量  $v_x$ ,  $v_y$  和  $v_z$  以及三个角分量  $\omega_x$ ,  $\omega_y$  和  $\omega_z$  是很方便的。因此，我们考虑这六种情况下的光流量（见图 8.15）：

**yaw 向移动：**如果视点绕  $y$  轴逆时针旋转（正  $\omega_y$ ），则所有的视觉特征以相同的速度从右向左移动，如图 8.15 (a) 所示。等同地，虚拟世界围绕用户顺时针旋转；然而，相反方向的自我运动被感知。这会让用户感觉自己仿佛置身于旋转木马（回忆图 8.2）。

**pitch 向移动：**通过绕  $x$  轴（正的  $\omega_x$ ）逆时针旋转视点，所有特征以相同的速度向下移动，如图 8.15 (b) 所示。

**roll 向移动：**逆时针旋转视角约  $z$ （光轴（正向  $\omega_z$ ）），使特征绕图像中心顺时针旋转，如图 8.15 (c) 所示。每个特征的速度与包含它的圆相切，速度与从特征到图像中心的距离成正比。

**横向移动：**在这种情况下，视点被转换到右侧，对应于正向  $v_x$ 。结果，这些特征水平移动；然而，在偏航方面有一个重要的区别：与较近物体相对应的特征比远距离物体更快。图 8.15 (d) 通过假设该特征的垂直位置对应于其深度（深度域中较低）来描绘该场。这是视差的再现，在这种情况下，会产生侧向运动的错觉，并将其与偏航运动区分开来。

**垂直移动：**视点向上平移，对应于正向  $v_z$ ，导致向下移动，如图 8.15 (e) 所示。视差再次导致特征的速度取决于相对对象的距离。这样可以将纵向切割与俯仰切割区分开来。

**向前/向后的移动：**如果视点沿着光轴从场景转移（正向  $v_z$ ），那么特征向内流向图像中心，如图 8.15 (f) 所示。它们的速度取决于它们距图像中心的距离以及它们在虚拟世界中相对对象的距离。由此产生的错觉是后退运动。在负  $z$  方向上的平移会导致向前的运动（例如在“星球大战”电影中跳转到超空间之后的千年猎鹰飞船）。

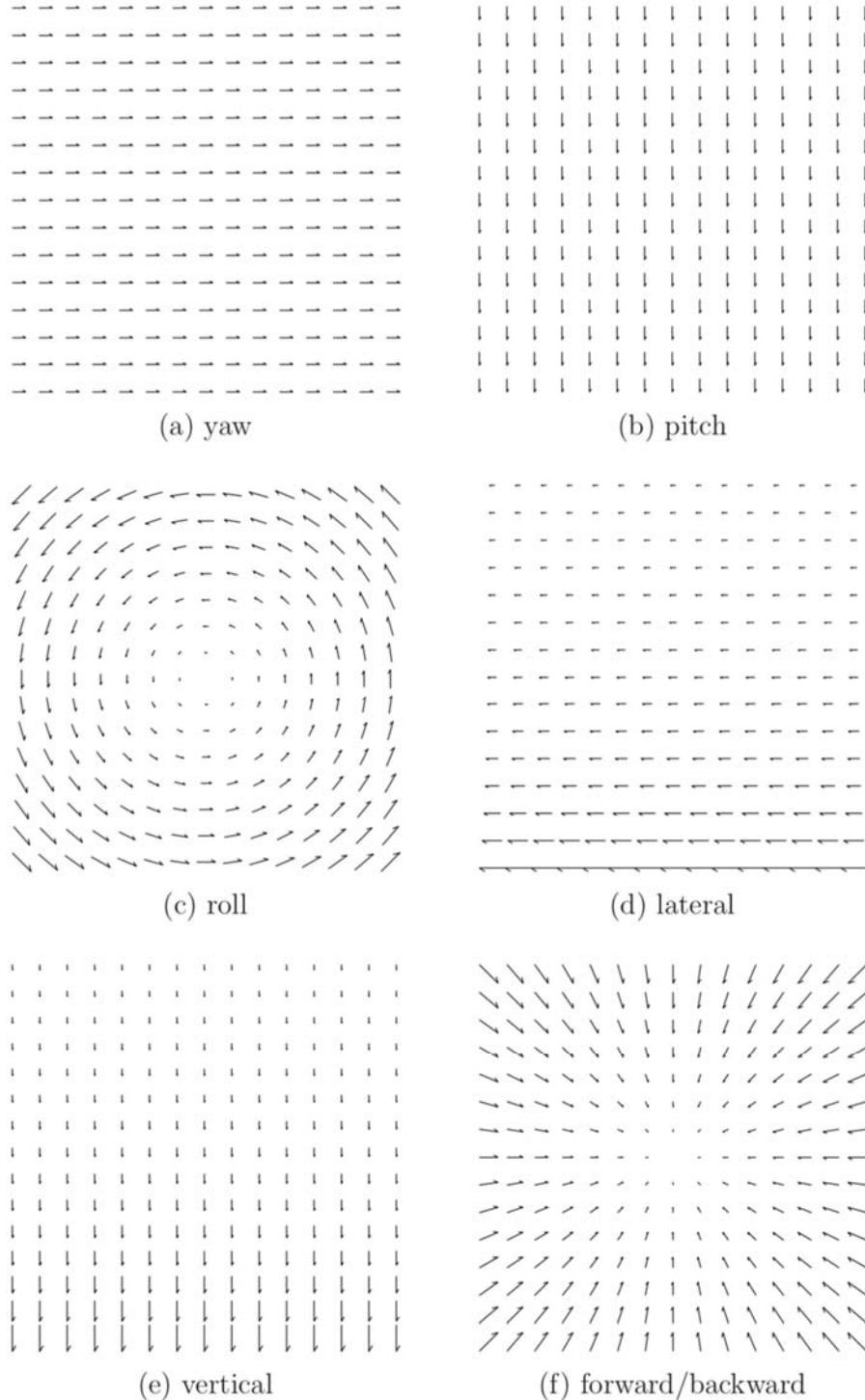


图 8.15：根据刚体运动的六个自由度，六种不同类型的光流。这些中的每一个都是 vection 的一个贡献组成部分

前两种有时被称为圆形切面，后三种被称为线性切割。由于我们的眼睛被吸引到移动的特征上，所以改变视点可能会引发顺利追踪眼球运动（请参阅第 5.3 节）。在这种情况下，图 8.15 所示的光流与视网膜上的特征不一致。因此，我们迄今为止的表征忽略了眼动，其经常被设计用于抵消光流并在视网膜上提供稳定的图像。尽管如此，由于本体感受，大脑知道这些眼球旋转，这导致了对自身运动的相同感知。

所有形式的动力都会引起感知速度，但加速度的感觉更为复杂。首先考虑视点的纯粹旋转。如果偏航角度，俯仰角度和侧倾角度的旋转速度变化，则会感知到角加速度。即使在角速度恒定的情况下，偏航，俯仰或侧倾也可以感知线性加速度。这是由于图 8.2 (b) 所示的旋转木马效应。

现在考虑纯粹的线性运动（不旋转）。视点的任何线性加速度都将被视为加速度。但是，如果视点以恒定速度移动，那么这是唯一没有感知加速度的视觉形式。在 VR 头戴式设备中，用户可能会感觉到由于光学失真或渲染和显示中的其他缺陷而导致的加速度。

## 前庭不匹配

我们尚未考虑这六种情况中每一种与前庭线索不匹配的效果。如果使用者没有相对于地球移动，那么只有重力应该被前庭器官（特别是耳石器官）感知。假设用户面向前而没有任何倾斜。在这种情况下，任何感知的加速度都会导致不匹配。例如，偏航俯仰应该引起朝向旋转中心的可感知的恒定加速度（回忆图 8.2 (b)），这与前庭重力线索不匹配。作为另一个例子，向下的垂直视觉应该使得使用者感觉他正在跌倒，但是前庭提示会另外指示。

对于在恒定角速度下的移动，俯仰和翻滚运动的情况，由于半圆形运河测量角加速度，所以与前庭器官感测到的旋转可能没有冲突。因此，视点的角速度必须改变以引起前庭系统的这部分不匹配。由于耳石器官感知到的线性加速度不匹配，可能会产生疾病。

如果头部实际在移动，那么前庭器官就会受到刺激。这种情况更难理解，因为与现实世界中线性和角度加速度相对应的前庭线索与表征不同加速度的视觉线索相结合了。在某些情况下，这些线索可能更一致，而在其他情况下，它们可能会进一步分歧。

## 影响灵敏度的因素

虚拟移动的强度受很多因素影响：

**视场角百分比：**如果只有一小部分视野在移动，那么人们倾向于认为它是由运动物体引起的。然而，如果大部分视野正在移动，那么他们就认为自己正在移动。人类视觉系统实际上包括具有感受野的神经元，其覆盖大部分视网膜以检测自身运动[35]。随着 VR 头戴式设备增加了视野，他们投射到视网膜的更大区域，从而加强了视觉线索。

**曝光时间：**由于暴露在光流中的时间而导致的对于自身运动的感知增加。如果曝光时间非常短暂，如几毫秒，则不会发生任何动作。

**空间频率：**如果虚拟世界很复杂，具有许多小的结构或纹理，那么视觉特征的数量将大大增加，并且光流变成更强的信号。随着 VR 设备显示分辨率的提高，可以生成更高的空间频率。

**对比度：**对比度越高，光流信号越强，因为更容易检测到这些特征。因此，典型的情况是强度更大。

**其他感官线索：**回顾 6.4 节，感知现象取决于许多线索的组合。可以通过提供更多一致的线索来增强观察力。例如，向前移动可能伴随着用户脸上的风扇，隆隆的引擎以及虚拟世界中的静止物体的声音。同样地，可以通过提供与没有发生相应的运动的现实世界一致的提示来减弱视觉效果。

**先验知识：**只要事先知道应该感知什么样的动作会影响移动的进行。这导致了先前的偏差，如果偏差与自身运动相反，则可能需要更长的时间才能克服，但如果与自身运动一致，则需要更少的时间来克服。先前的偏见可能来自某人告诉用户将要发生的事情，或者可能仅仅是通过用户一生中类似视觉体验的积累。此外，用户可能期望该动作是由于所采取的动作而产生的，例如转动虚拟车辆的方向盘。

**注意力：**如果用户被其他活动分散注意力，例如瞄准虚拟武器或选择菜单选项，则可能

会减轻虚拟移动及其副作用。

**事先训练或适应：**在足够的暴露下，身体可以学会将真实运动与真实运动区分开来，以使运动变得舒适。因此，许多用户可以接受培训，通过反复，长时间的接触来克服 VR 疾病。

由于所有这些因素以及现代 VR 头盔的不完善性，很难表征虚拟移动的效能及其对用户舒适性的副作用。

## 拓展阅读

有关矢量场，速度和动力学系统的基本概念，请参见[12]。机械动力系统的建模和分析见[275]。 [364,365]介绍了人体运动的具体问题。有关游戏引擎的概述，请参见[103]，包括模拟物理和碰撞检测等问题。有关特定碰撞检测算法的介绍，请参见[100, 182]。

前庭系统的一个很好的介绍，包括它作为一个动态系统的反应[150]。在文献[145]中对透视和视觉晕动病进行了深入调查，其中包括大量的参考文献供拓展阅读。一些重要的文章阐述了对动机的敏感性[6,13,69,178,179,279,340]。

# 第九章 追踪

追踪物理世界的运动对于任何 VR 系统都是关键的部分。追踪是将 VR 头显引入消费型电子产品的最大障碍之一，并且由于我们希望扩展和改进 VR 体验，它仍然会是一个重大的挑战。高精确的追踪方法主要由商用硬件组件支持，例如惯性测量单元(IMU)和由于智能手机产业发展，规格和成本急剧下降的相机。

根据目前被追踪的物体情况，VR 系统中可能会出现三类追踪：

## 1. 使用者的感官：

回想 2.1 节，人的感觉器官，比如眼睛和耳朵，都有由身体控制的自由度。如果一个显示器连接到一个感应器，它应该被视为在 VR 中被附着到周围的世界，那么器官的位置和方向就需要被追踪。我们可以将追踪变换的反方向应用于刺激来正确地“抵消”这些自由度。大部分的焦点在于头部追踪，这对 VR 的视觉和听觉成分来说已经足够了。然而，如果渲染和显示技术需要补偿 5.3 节中讨论的眼动，视觉系统可能还需要人眼追踪。

## 2. 用户的其他身体部位：

如果用户想要在虚拟世界中看到自己身体被正确地表示，那么身体的运动理应被追踪，以便它可以在匹配区域中再现。也许面部表情或手势是互动所必需的。虽然完美匹配是追踪感觉器官的理想选择，但追踪其他身体部位不是必需的。现实世界中的小动作可以转换成更大的虚拟世界动作，以便用户更不费力。在极限情况下，用户只需按下一个按钮即可更改主体配置。例如，她可以通过一次点击就能在她的虚拟手中抓住一个物体。

## 3. 其余环境：

在用户周围的现实世界中，可以追踪物理对象。对于存在于真实世界但不存在于虚拟世界中的物体，出于安全原因，系统需要提醒用户它们的存在。想象一下，用户即将撞墙，或绊倒一名幼童。在一些 VR 应用中，被追踪的物理对象可以在 VR 中匹配，以便用户在与它们交互时接收触碰反馈。在其他应用中，例如远程监控，物理世界的很大一部分可以通过实时捕捉被“引入”虚拟世界。

第 9.1 节介绍了追踪围绕单个轴的旋转的简单情况从而为将框架扩展到追踪 3D 刚体的 3-自由度方向的 9.2 节做铺垫。这主要依赖于 IMU 的角速度读数。最常见的用途是追踪穿戴 VR 头显设备的头部运动，但也可能适用于追踪手持控制器或其他设备。第 9.3 节讨论了位置和方向的追踪，这在大多数系统中要求在物理世界的固定部分和被追踪物体之间的视线可见性。9.4 节讨论追踪通过节点连接在一起的多个物体的情况。最后，第 9.5 节介绍了使用传感器构建物理世界的表示以便将其引入虚拟世界的情况。

## 9.1 2D 方向追踪

本节介绍如何使用惯性测量单元(IMU)估算刚体的方向。主要应用是确定视点方向，即第 3.4 节介绍的当用户佩戴 VR 头显设备时的  $R_{eye}$ 。另一个应用是估计手持控制器的方向。例如，假设我们想根据用户指示的方向制作一个适用于虚拟世界的激光笔。场景中亮红色点的位置将由控制器的估计方位决定。更一般地说，如果它具有附加的 IMU，则可以确定物理世界中任何人体部位或移动物体的方位。

为了估计方向，我们首先仔细考虑第 8.1.2 节的旋转木马模型来考虑 2D 情况。在这种情况下，技术问题很容易可视化，并扩展到更重要的 3D 旋转情况。因此，假设我们在旋转的旋转木马上安装陀螺仪。它的工作就是像旋转木马旋转一样测量角速度。在本章中通过区

分真实参数值和估计值将会很方便。为了实现这个目标，一个“帽子”将被放置在估计值上。因此，令  $\hat{\omega}$  对应估计或测量的角速度，其可能与真实值  $\omega$  不同。

$\hat{\omega}$  和  $\omega$  如何相关？如果陀螺仪功能完好，则  $\hat{\omega}$  等于  $\omega$ ；然而，在现实世界中，这是无法实现的。 $\hat{\omega}$  和  $\omega$  之间的差异的主要原因是标定误差。标定质量从昂贵的 IMU(数千美元)到廉价(一美元)之间有很大区别。

我们现在定义一个简单的校准误差模型。以下传感器映射表明传感器输出如何与真实角速度相关：

$$\hat{\omega} = a + b\omega. \quad (9.1)$$

上面， $a$  和  $b$  分别被称为偏移和比例。它们是干扰测量的未知常数。如果  $\omega$  被完美测量，将有  $a = 0$  和  $b = 1$ 。

考虑校准误差的影响。比较测得的角速度和真实的角速度得出：

$$\hat{\omega} - \omega = a + b\omega - \omega = a + \omega(b - 1). \quad (9.2)$$

现在想象使用传感器来估计旋转木马的方向。我们想了解真实方向  $\theta$  与使用传感器输出计算的估计值  $\hat{\theta}$  之间的差异。令  $d(t)$  表示称为漂移误差的时间函数：

$$d(t) = \theta(t) - \hat{\theta}(t). \quad (9.3)$$

请注意， $d(t)$  可能是负值，可通过应用绝对值来获得  $|d(t)|$ ，从而强制其为正值。这将被避免以简化讨论。

假设初始给定  $\theta(0) = 0$ ，为了保持简单，角速度  $\omega$  是恒定的。通过随时间积分(9.2)，漂移误差是

$$d(t) = (\hat{\omega} - \omega)t = (a + b\omega - \omega)t = (a + \omega(b - 1))t. \quad (9.4)$$

当然，随着  $a$  偏离 0 或随着  $b$  偏离 1，漂移误差会增加(正向或负向)；然而，请注意第二个分量与  $\omega$  成比例。忽略  $a$ ，这意味着漂移误差与旋转木马的速度成正比。在使用陀螺仪追踪 VR 头显设备方面，这意味着随着头部旋转更快，追踪误差以更快的速度增加[168]。

在这一点上，即使对于这种简单的情况，也必须解决四个常见问题才能建立一个有效的追踪系统：

**1. 校准：**如果有更好的传感器可用，那么两者可以紧密配对，以便更差传感器的输出进行转换以尽可能接近更好的传感器。

**2. 集成：**传感器提供离散时间点的测量结果，产生一个采样率。方向通过聚合或综合测量来估计。

**3. 配准：**初始定位必须以某种方式确定，或者通过额外的传感器，或者巧妙的默认假设或启动程序来确定。

**4. 漂移误差：**随着误差随时间的增加，需要其他传感器来直接估算并补偿。所有这些问题在本章中都会涉及到更复杂的设置。组合来自多个传感器读数的信息的过程通常称为传感器融合或过滤。

我们在 2D 情况下讨论其中的每一个问题，然后再将这些想法扩展到第 9.2 节中的 3D 情况。

## 校准

你可以购买传感器，并开始使用它，并假定它已经很好地校准。然而，对于更便宜的传

感器，校准常常不可靠。假设我们有一个昂贵的，校准良好的传感器，可以获取误差很小的角速度。设 $\hat{\omega}'$ 表示其输出，以将其与永远未知的真值 $\omega'$ 区分开来。现在假设我们要校准一堆便宜的传感器，以便它们尽可能地接近昂贵的传感器。这可以通过将它们安装在可移动表面上并比较它们的输出来完成。为了获得更高的精度和控制，最昂贵的传感器可能是完整的机械系统的一部分，例如昂贵的转盘，校准台或机器人。设 $\hat{\omega}$ 表示待校准的廉价传感器的输出；每个便宜的传感器都必须单独校准。

校准涉及多次采样，有时需要数千次，并将 $\hat{\omega}'$ 与 $\hat{\omega}$ 进行比较。一个通用的标准是平方和的误差，由下式给出

$$\sum_{i=1}^n (\hat{\omega}_i - \hat{\omega}'_i)^2 \quad (9.5)$$

对于角速度的 n 个样本。其任务是确定一个适用于廉价传感器输出 $\hat{\omega}$ 的变换，以便它尽可能接近昂贵的传感器输出 $\hat{\omega}'$ 。

使用(9.1)中的误差模型，我们可以选择优化误差的常量  $c_1$  和  $c_2$ ：

$$\sum_{i=1}^n (c_1 + c_2 \hat{\omega} - \hat{\omega}')^2. \quad (9.6)$$

这是一个经典的回归问题，称为线性最小二乘。它通常通过计算包含采样数据的非方阵矩阵的 Moore-Penrose 伪逆求解[341]。

一旦计算出  $c_1$  和  $c_2$ ，每一个未来的传感器读数都被转换为

$$\hat{\omega}_{cal} = c_1 + c_2 \hat{\omega}, \quad (9.7)$$

其中 $\hat{\omega}$ 是原始的原始传感器输出， $\hat{\omega}_{cal}$ 是校准输出。因此，校准会在便宜的传感器输出周围产生一种不可见的外壳，以便模拟昂贵的传感器。外部流程不再显示原始的廉价传感器输出。因此，本章其余部分将校准后的输出简称为 $\hat{\omega}$ 。

## 集成

传感器输出通常以常规采样率达到。例如，Oculus Rift 陀螺仪每 1ms 提供一次测量(产生 1000Hz 的采样率)。令 $\hat{\omega}[k]$ 指第 k 个样本，它在时间  $k \Delta t$  到达。

时刻  $t = k \Delta t$  时的方向  $\theta(t)$  可以通过积分来估计为：

$$\hat{\theta}[k] = \theta(0) + \sum_{i=1}^k \hat{\omega}[i] \Delta t. \quad (9.8)$$

每个输出 $\hat{\omega}[i]$ 引起 $\Delta \theta[i] = \hat{\omega}[i] \Delta t$  的旋转。有时以增量形式写入(9.8)更方便，这表示在每个新传感器输出到达后更新为 $\hat{\theta}$ ：

$$\hat{\theta}[k] = \hat{\omega}[k] \Delta t + \hat{\theta}[k-1]. \quad (9.9)$$

对于第一种情况， $\hat{\theta}[0] = \theta(0)$ 。

如果 $\omega(t)$ 在 $\theta(k \Delta t)$ 和 $\theta((k+1) \Delta t)$ 之间变化很大，那么知道 $\hat{\omega}[k]$ 对应什么是有帮助的。它可以是间隔 $\Delta t$ 开始时的角速度，间隔的结束时间或者间隔内的平均值。如果是开始或结束，那么积分的梯形近似可能会随着时间的推移产生更多的误差[133]。

## 配准

在式(9.8)中，假定初始方向  $\theta(0)$  是已知的。实际上，这与配准问题相对应，这是真实和

虚拟世界之间的初始对齐。为了理解这个问题，假设  $\theta$  表示 VR 头盔的偏航方向。一种可能性是分配  $\theta(0) = 0$ ，其对应于当追踪系统打开时头显设备面对的任何方向。这可能是系统启动时。如果头显设备具有“头戴式”传感器，则可以在用户将头显设备戴上时开始。很多时候，前进的方向可能会以不好的方式无意中设置。例如，如果一个人开始一个 VR 演示，并将头显设备交给面对另一个方向的其他人，那么在 VR 中，用户不会朝向预期的前进方向。这可以通过导致“向前”( $\theta(t)$ )被重新定义为用户目前面对的任何方向的简单选项来修复。

这个整体问题的一个替代方法是声明  $\theta(0) = 0$  以对应于物理世界中固定的方向。例如，如果用户坐在计算机监视器前面的桌子上，则向前方向可以被定义为用户和头盔面向监视器的偏航角。实施这个解决方案需要一个能够测量相对于周围物理世界的偏航方向的传感器。例如，在 Oculus Rift 中，用户面对一个固定的摄像头，它对应于前进方向。

## 漂移校正

为了建立一个有用的追踪系统，不能允许漂移误差(式 9.3)累积。即使陀螺仪被完美校准，由于诸如量化输出值，采样率限制和未建模噪声等其他因素，漂移误差仍会增加。第一个问题是估计漂移误差，通常用附加传感器完成漂移误差。这个实例将在 9.2 节中给出。对于简单的旋转木马的示例，想象一下架空摄像机偶尔会拍摄一张照片来测量方向。令  $\hat{\theta}_d[k]$  表示从该单个传感器测量结果到达阶段  $k$  的估计方位。

由于漂移误差，现在有两个相互矛盾的信息来源：1)通过积分陀螺仪估计的方向  $\hat{\theta}[k]$ ，以及 2)由相机(或其他一些独立的传感器)瞬时估计的方向  $\hat{\theta}_d[k]$ 。混合这两种来源的经典方法是互补过滤器，它在两种估计之间进行数学插值：

$$\hat{\theta}_c[k] = \alpha\hat{\theta}_d[k] + (1 - \alpha)\hat{\theta}[k], \quad (9.10)$$

其中  $\alpha$  是必须满足  $0 < \alpha < 1$  的增益参数。以上， $\hat{\theta}_c[k]$  表示阶段  $k$  处的校正估计。由于陀螺仪在短时间内通常是准确的，但会逐渐漂移，所以选择  $\alpha$  接近于零(例如  $\alpha = 0.0001$ )。这导致瞬时估计值  $\hat{\theta}_d[k]$  逐渐产生影响。另一方面，如果  $\alpha$  接近于 1，那么由于每个阶段  $\theta_d[k]$  的误差，估计的方向会大幅波动。另一个考虑因素是，如果传感器输出  $\hat{\theta}_d[k]$  的速率比陀螺仪采样速率低得多，则使用最近记录的输出。例如，相机图像可能产生 60Hz 的方位估计，而陀螺仪产生 1000Hz 的输出。在这种情况下， $\hat{\theta}_d[k]$  将保持 16 或 17 个阶段的相同值，直到新的相机图像变为可用。

选择增益  $\alpha$  足够高以便漂移得到校正是非常重要的，但要足够低以便用户不会感觉到校正。增益可以通过使用卡尔曼滤波器“最优地选择”[45,140,158];然而，如果我们有一个线性随机系统，最优性只能保持不变，而这种情况不适用于人体追踪。卡尔曼和补充滤波器之间的关系，对于本章中使用的确切模型，见[119]。

使用简单的代数，(9.10)中的互补过滤器公式可以重新生成以产生以下等效表达式：

$$\hat{\theta}_c[k] = \hat{\theta}[k] - \alpha\hat{d}[k] \quad (9.11)$$

其中，

$$\hat{d}[k] = \hat{\theta}_d[k] - \hat{\theta}[k]. \quad (9.12)$$

以上， $\hat{d}[k]$  只是阶段  $k$  中漂移误差的估计值。因此，互补滤波器可以被想象为以一个小的比例量  $\alpha$  来施加有符号的误差，以试图逐渐将其强制为零。

## 9.2 3D 方向追踪

### IMU

回忆 2.1 节(图 2.9), IMU 最近已经从大型机械系统转向廉价微型 MEMS 电路。这种进展是高质量定位追踪的关键推动力。陀螺仪测量沿着三个正交轴的角速度, 以获得  $\omega_x$ ,  $\omega_y$  和  $\omega_z$ 。对于每个轴, 感应元件位于垂直平面内, 非常类似于前庭器官的半规管(8.2 节)。每种情况下的传感元件都是微机械的机械元件, 像音叉一样振动和操作。如果传感器以其灵敏度方向旋转, 则元件会受到科里奥利力, 被转换成电信号。这些信号被校准以产生每秒度数或弧度的输出; 见图 9.1。

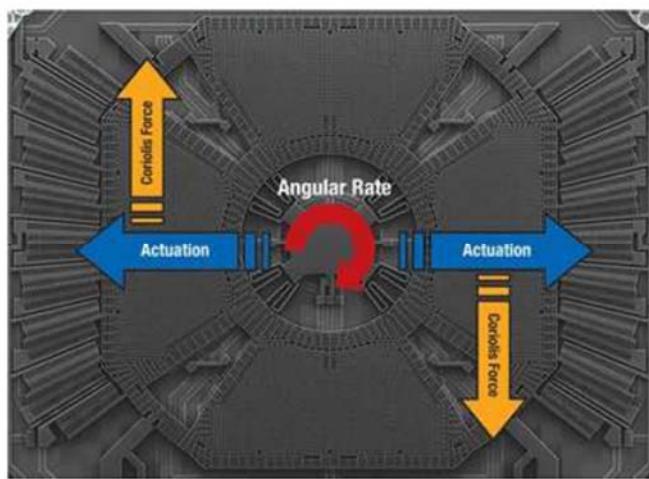


图 9.1: 振动 MEMS 元件在旋转过程中对科里奥利力做出响应, 将其转换为电信号。(图源自 Fabio Pasolini)

IMU 通常包含用于检测漂移误差的附加传感器。最常见的情况是, 加速度计沿着三个轴测量线性加速度以获得  $a_x$ ,  $a_y$  和  $a_z$ 。其操作原理如图 9.2 所示。MEMS 磁力计也出现在许多现代 IMU 上, 这些 IMU 沿着三个垂直轴测量磁场强度。这通常通过 MEMS 结构的机械运动来完成, 所述 MEMS 结构在磁场内传导时受到洛伦兹力的作用。

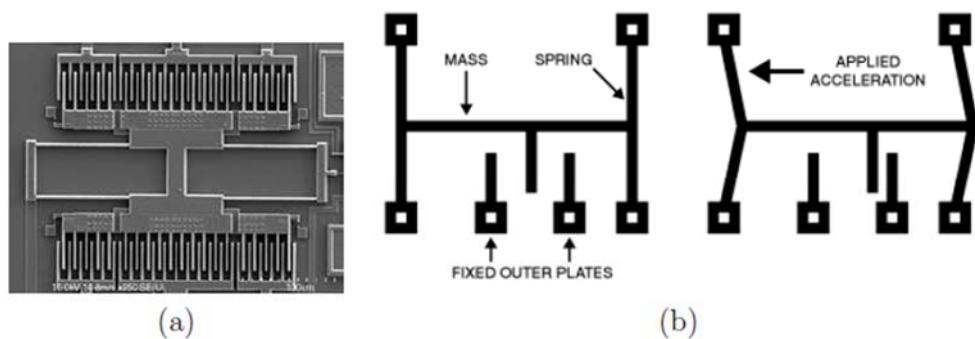


图 9.2: (a) 用于感测线性加速度的 MEMS 元件。 (b) 由于在一个方向上的线性加速度, 这些极板移动并导致在外板之间测量的电容变化。 (图源自 David Askew)

## 校准

回顾第 9.1 节传感器输出由于校准问题而失真。在一维角速度情况下，只有两个参数，出现在(9.1)中的比例尺和偏移量。在 3D 设置中，这自然会扩展到 3 个比例尺和 3 个偏移量参数；然而，情况由于 MEMS 元件的非正交性带来的误差会更糟。所有这些都可以由排列在齐次变换矩阵中的 12 个参数来解释：

$$\begin{bmatrix} \hat{\omega}_x \\ \hat{\omega}_y \\ \hat{\omega}_z \\ 1 \end{bmatrix} = \begin{bmatrix} a & b & c & j \\ d & e & f & k \\ g & h & i & \ell \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \omega_x \\ \omega_y \\ \omega_z \\ 1 \end{bmatrix} \quad (9.13)$$

共有 12 个而不是 6 个自由度，因为左上角的  $3 \times 3$  矩阵不被限制为旋转矩阵。 $j, k$  和 1 参数对应于偏移量，而其他所有参数都处理比例和非正交性。采用与 9.1 节相同的方法，可以通过最小平方误差相对于更好的传感器的输出来估计该变换的逆，其提供了真实情况。MEMS 传感器的输出通过应用估计的齐次变换被调整以提高性能(这是(9.7)到 12 参数情况的延伸)。这种通用方法适用于校准陀螺仪和加速度计。磁力计也可以用这种方式进行校准，但还会有其他问题，如软铁偏压。

MEMS 传感器面临的另一个挑战是处理其他微妙的依赖关系。例如，输出对 MEMS 元件的特定温度敏感。如果 VR 头显设备在使用过程中变热，那么在实践中可能出现的每个温度都需要校准参数。幸运的是，IMU 通常包含一个温度传感器，可用于将校准参数与相应的温度相关联。最后，MEMS 元件可能对作用在电路板上的力敏感，这可能会被改变，例如，由悬挂连接器改变。必须注意将 MEMS 电路的外部电路板力量隔离开来。

## 集成

现在考虑将陀螺仪输出序列转换为 3D 定向估计的问题。在每个阶段  $k$ ，一个向量

$$\hat{\omega}[k] = (\hat{\omega}_x[k], \hat{\omega}_y[k], \hat{\omega}_z[k]) \quad (9.14)$$

从传感器到达。在 9.1 节中，传感器输出  $\hat{\omega}[k]$  在方向上被转换为  $\Delta \theta[k]$ 。对于 3D 情况，方向的变化表示为四元数。

令  $q(v, \theta)$  为由轴角转换公式(3.30)得到的四元数。回顾第 8.1.2 节，瞬时旋转轴是角速度的大小。因此，如果  $\hat{\omega}[k]$  是阶段  $k$  中的传感器输出，则估计的旋转轴是

$$\hat{v}[k] = \hat{\omega}[k] / \|\hat{\omega}[k]\|. \quad (9.15)$$

此外，在时间  $\Delta t$  期间发生的估计旋转量是

$$\Delta \hat{\theta}[k] = \|\hat{\omega}[k]\| \Delta t. \quad (9.16)$$

使用估计的旋转轴(9.15)和数量(9.16)，估计方向随时间的变化  $\Delta t$

$$\Delta \hat{q}[k] = q(\hat{v}[k], \Delta \hat{\theta}[k]). \quad (9.17)$$

在每个阶段使用(9.17)，获得最新传感器输出后的估计方向  $q[k]$  从  $q[k-1]$  递增计算为

$$\hat{q}[k] = \Delta \hat{q}[k] * \hat{q}[k-1], \quad (9.18)$$

其中\*表示四元数乘法。这是(9.9)的 3D 泛化，其中可以使用简单的加法来结合 2D 情况下的旋转。在(9.18)中，需要四元数乘法来聚合方向的变化(简单加法是可交换的，这对于 3D

旋转是不合适的)。

## 配准

偏航组件的配准问题与第 9.2 节中的相同。正向可以从刚体的初始方向选择，或者可以相对于世界中的固定方向来确定。俯仰和滚转方向应该被确定以便它们与重力对齐。虚拟世界不应该出现相对于现实世界的倾斜(除非这是所期望的效果，这种情况很少)。

## 倾斜校正

(9.3)中的漂移误差  $d(t)$  是单个角度，可能是正值或负值。如果增加到估计值  $\hat{q}(t)$ ，则将获得真正的方向  $q(t)$ 。它与 3D 情况类似，但是具有四元数代数。三维漂移误差表示为

$$d(t) = q(t) * \hat{q}^{-1}(t), \quad (9.19)$$

如果  $q(t) = \hat{q}(t)$ ，它等于旋转。此外，请注意，将漂移误差应用于估计产生  $q(t) = d(t) * \hat{q}(t)$ 。

由于漂移误差是 3D 旋转，因此可以将其构造为偏航，俯仰和滚转的结果。让倾斜误差表示俯仰和滚转相对应的漂移误差部分。这将使用“向上”传感器进行检测。让偏航误差指的是漂移误差的剩余部分，将使用“指南针”进行检测。实际上，并不存在完美的“上”和“罗盘”传感器，这些传感器将在后面进行介绍。

假设连接到刚体的传感器总是返回与固定世界坐标系中的  $y$  轴平行的“向上”向量。换句话说，它将与重力平行。由于传感器安装在身体上，它会在身体的坐标框架中报告其值。例如，如果身体滚动 90 度，以使其  $x$  轴向上指向，则“向上”向量将被报告为  $(0, 0, 1)$ ，而不是  $(0, 1, 0)$ 。为了解决这个问题，将传感器输出转换为世界范围会很方便。这涉及通过  $q(t)$ (身体方向)旋转它。对于我们的例子，这个滚动旋转会将  $(0, 0, 1)$  变换为  $(0, 1, 0)$ 。图 9.3 显示了一个 2D 例子。

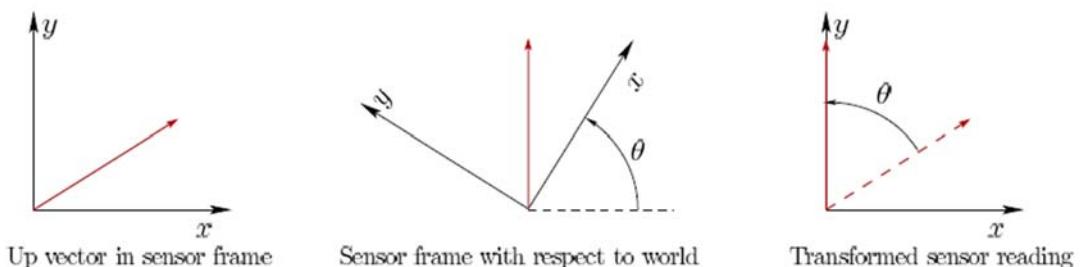


图 9.3：如果通过旋转  $\theta$  的加速度计完全检测到“向上”，则其输出需要旋转  $\theta$  以从世界框架中查看它。

现在假设发生了漂移误差，并且  $\hat{q}[k]$  是估计的方向。如果将这个变换应用于“向上”向量，那么由于漂移误差，它可能不会与  $y$  轴对齐，如图 9.4 所示。向上向量  $u$  被投影到  $xz$  平面上以获得  $(u_x, 0, u_z)$ 。倾斜轴位于  $xz$  平面内，并构造为投影向量的法向： $t = (u_z, 0, -u_x)$ 。围绕轴线  $t$  执行  $\phi$  的旋转将使向上向量与  $y$  轴对齐。因此，漂移误差的倾斜误差部分是四元数  $q(t, \phi)$ 。

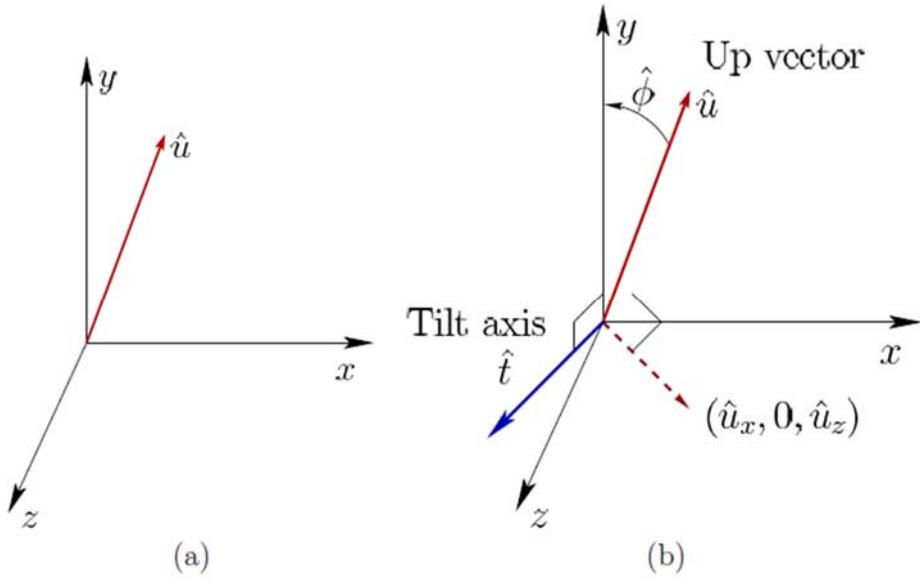


图 9.4: (a) 倾斜误差导致 y 轴与使用估计的  $\hat{q}[k]$  旋转感知测得的向上矢量之间的差异以获得  $\hat{u}$ 。 (b) 倾斜轴与  $\hat{u}$  垂直; 围绕倾斜轴的  $-\hat{\theta}$  旋转将使它们对齐, 从而消除了倾斜误差。

不幸的是, 没有传感器直接测量“向上”。实际上, 由于重力作用在传感器上, 所以加速计用于测量“向上”方向, 导致向上加速度的感觉大约为  $9.8\text{m/s}^2$ 。问题在于它也会响应刚体真正的线性加速度, 并且由于爱因斯坦等价原理的原因, 这不能与重力分开。它测量重力和真正的线性加速度的矢量和, 如图 9.5 所示。一个简单的启发式方法就是只有当它的幅度接近  $9.8\text{m}^2$  时[75], 才将加速度计输出信任为“向上”方向的估计值。这可以对应于刚体静止的常见情况。然而, 这种假设是不可靠的, 因为向下和横向线性加速度可以组合以提供接近  $9.8\text{m}^2$  的输出量值, 但是具有远离“向上”的方向。可以同时考虑其他传感器的输出或“上”似乎改变的速率来构建更好的启发式方法。

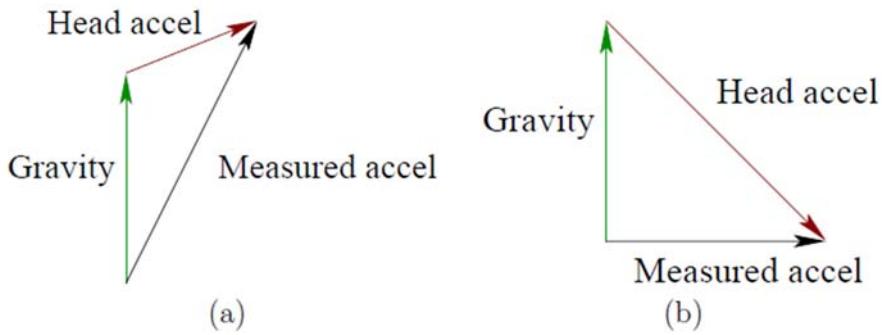


图 9.5: (a) 没有重力传感器; 加速度计测量由重力引起的视在加速度和身体的真实加速度的矢量和。 (b) 仅在幅度约为  $9.8\text{m}^2$  时接受读数作为重力的简单启发式算法在某些情况下会失败。

假定加速度计产生重力方向的可靠估计, 则通过使用(3.34)从加速度计输出  $\hat{a}$  计算向上矢量  $\hat{u}$  以获得

$$\hat{u} = \hat{q}[k] * \hat{a} * \hat{q}[k]^{-1}. \quad (9.20)$$

## 偏航校正

剩余漂移误差分量由“罗盘”检测，该“罗盘”输出位于世界 xz 平面的矢量，并始终指向“北”。假设这是  $\hat{n} = (0, 0, -1)$ 。传感器输出再一次出现在身体坐标系中，需要通过  $q[k]$  进行变换。 $\hat{n}$  和 z 轴之间的差异是导致的偏航漂移误差。

就像“up”传感器一样，现实世界中没有“指南针”。相反，有一个磁力计，它可以测量一个三维磁场矢量：( $m_x, m_y, m_z$ )。假设这是用来测量地球的磁场。事实证明，场向量不会“指向”北极。地球的磁场产生的 3D 矢量通常不在水平面上，从而导致倾斜的角度。因此，第一个问题是传感器输出必须投影到 xz 平面中。厄瓜多尔居民可以享受近乎水平的磁场矢量；然而，在芬兰，它们更接近于垂直方向；见图 9.6。如果磁场矢量接近垂直，那么水平分量可能变得太小而无用。

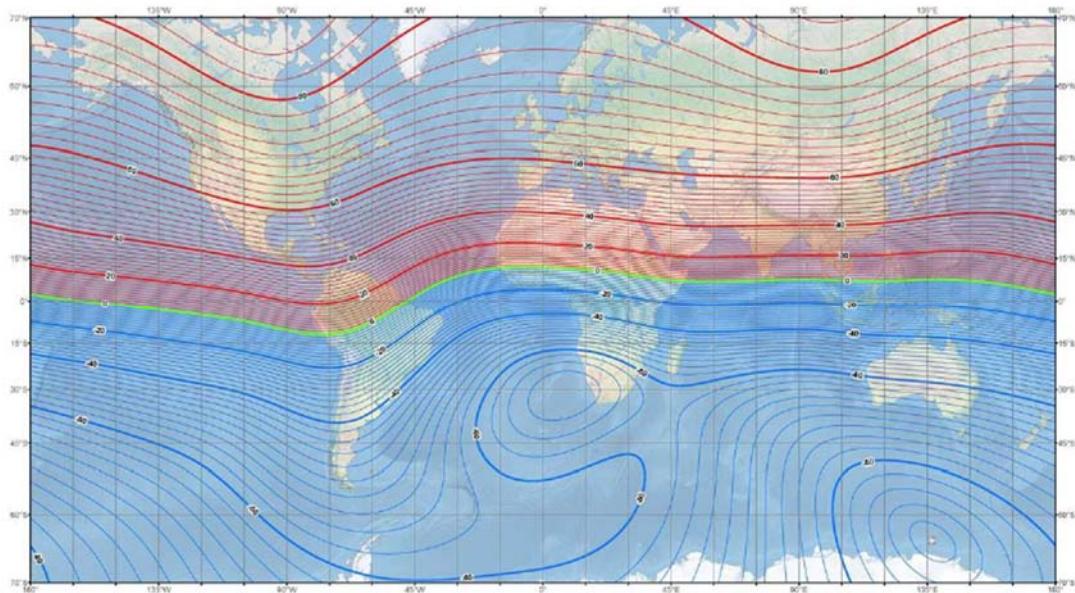


图 9.6：地球磁场矢量的倾角在地球上变化很大。（地图由 NOAA / NGDC 和 CIRES 开发）

另一个问题是水平面上的投影矢量不指向北，导致偏角；这是偏离北方。幸运的是，真正的北方的参考并不重要。只需要配准传感器输出记录在登记阶段提供一个固定的偏航基准。

最重要的问题是磁力计测量所有磁场源的矢量和。除了地球之外，建筑物由于铁磁金属而产生自己的场。此外，这种材料通常存在于包含传感器的电路板上。对于这种情况，场与传感器一起移动，产生恒定的矢量偏移。作为磁场源的材料被称为硬铁。其他材料扭曲穿过它们的磁场；这些被称为软铁。磁力计校准方法主要考虑由于硬铁偏置和由于软铁偏置引起的偏心造成的偏移[92,155]。

在这些磁力仪校准完成之后，偏航漂移误差可以从大多数位置以几个精度进行估算，这足以保持偏航误差逐渐积累。但仍然会存在问题。如果在传感器附近放置强场，则读数将取决于小的位置变化。这可能导致测量的方向随着刚体来回平移而改变。另一个问题是，在一些建筑物位置，地球磁场的矢量和由建筑物产生的场可能近似为零（如果它们具有相似的大小并指向相反的方向）。在这种不幸的情况下，磁强计不能产生用于偏航漂移误差检测的有用输出。

## 滤波

使用检测到的漂移误差，滤波的工作方式与第 9.1 节中所述的相同。补充过滤器(9.10)升级为与四元数一起使用。用  $\alpha$  来表示插值变得稍微复杂一些。令  $(v, \theta)$  表示方向  $d[k]$  的轴角表示，其是估计的漂移误差(四元数值)。令  $q(v, \alpha\theta)$  表示由  $v$  轴和角度  $\alpha\theta$  给出的四元数。对于较小的  $\alpha$  值，这可以被认为是一个“向” $\hat{d}[k]$  的一小步。

四元数互补滤波器变成

$$\hat{q}_c[k] = q(v, -\alpha\theta) * \hat{q}[k], \quad (9.21)$$

其形式与(9.12)相似。2D 情况下的简单减法已经在上面通过乘以左侧的反四元数来取代。估计的漂移误差 $\hat{d}[k]$  通过将估计的倾斜和偏航误差相乘而获得。或者，他们可以分别对互补滤波器做出贡献，每种都有不同的增益，甚至与来自更多来源的漂移误差估计相结合 [197]。

## 设置视点

为了预测或图像变形的目的，虽然可能需要调整以考虑替代时间，但使用估计方位 $\hat{q}[k]$  设置视点，如 7.4 节所述。令  $\hat{q}(t)$  表示时间  $t$  的估计方向。就 3.4 节的转换而言，我们刚估计了  $R_{eye}$ 。为了计算正确的观点，需要进行反变换。因此， $\hat{q}^{-1}(t)$  将正确转换模型以考虑估计的视点。

## 调试提示

将追踪的方向连接到视点时，程序员经常犯错误。图 9.7 显示了一个常见错误的表格。

Yaw	Pitch	Roll	Error
+	+	+	None
-	+	+	L/R mix, flipped $x$
+	-	+	L/R mix, flipped $y$
+	+	-	L/R mix, flipped $z$
+	-	-	Inverse and L/R mix, flipped $x$
-	+	-	Inverse and L/R mix, flipped $y$
-	-	+	Inverse and L/R mix, flipped $z$
-	-	-	Inverse

图 9.7：帮助调试常见视点转换错误的一览图。每个+表示当执行偏航，俯仰或滚动时，虚拟世界移动正确的方式。每个-意味着它以相反的方式移动。如图所示，第一种情况是正确的，其他的都是错误。“L/R 混合”意味着左手和右手坐标系混合。

为了确定变换是否被正确应用，人们应该戴上头戴式设备并尝试围绕三个典型轴旋转：纯偏航，纯俯仰和纯翻滚。让+表示世界正在相对于头部旋转正确移动。让-表示它似乎向相反的方向移动。图 9.7 显示了八个可能的结果以及每个问题最可能的原因。

## 头部模型

头部运动的转换部分尚未得到解决。理想情况下，虚拟世界中的头部应与现实世界中的高度相同。这可以通过转换  $T_{eye}$  矩阵来处理(3.36)。

我们还必须说明，随着头部转动，眼睛也会改变其位置。例如，在偏航头部运动中(点头“否”)，瞳孔在  $x$  方向上移动几厘米。更准确地说，它们沿水平面上的圆弧移动。为了更接近模仿现实世界，可以通过根据虚拟头部模型改变旋转中心来模拟眼睛在空间中的移动 [3]。这个技术到 9.3 节将被使用，其中位置是由更多的传感器估算的。

回想 3.5 节，首先考虑了 cyclopean 的观点，然后通过插入  $T_{left}$ (3.50) 和  $T_{right}$ (3.52) 来应用

水平偏移来修改处理左眼和右眼。以类似的方式，可以增加 y 和 z 方向上的偏移来考虑来自旋转头的位移。结果是在  $T_{right}$  和  $T_{left}$  之前或之后插入以下内容：

$$T_{head} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & h \\ 0 & 0 & 1 & p \\ 0 & 0 & 0 & 1 \end{bmatrix}, \quad (9.22)$$

其中  $h$  是高度参数， $p$  是突出参数。见图 9.8。这个想法是选择对应于头部旋转中心的  $h$  和  $p$ 。参数  $h$  是沿着  $y$  轴从旋转中心到眼睛高度的距离。典型值是  $h=0.15m$ 。突起  $p$  是从旋转中心到环眼的距离。一个典型值是  $p=-0.10m$ ，这是负值，因为它与  $z$  轴相反。当用户旋转他的头部时，使用假头部模型来近似眼睛的位置，然而这并不是完美的。如果躯干移动，那么这个模型完全失效，导致真实和虚拟世界头部运动之间的不匹配。尽管如此，这款头戴机型目前在三星 Gear VR 等流行设备中使用。

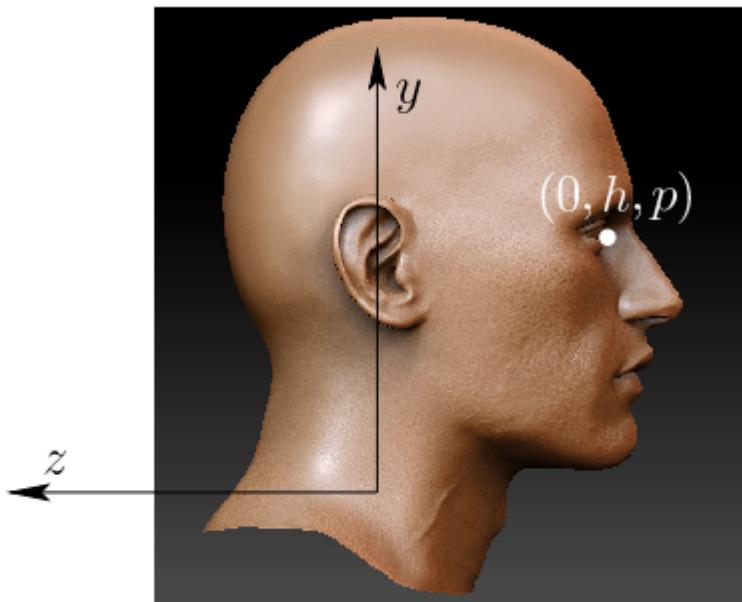


图 9.8：为了获得头部模型，移动旋转中心使得方向改变引起眼睛的合理平移。高度  $h$  沿着  $y$  轴，并且点  $p$  沿着  $z$  轴(这将导致结果为负值)

头部中心的高度也存在问题。用户可能坐在现实世界中，但站在虚拟世界中。这种不匹配可能会让用户体验不佳。不管 VR 提供的视觉刺激，大脑知道身体是由于本体感觉而坐着的。如果用户站立，则可以设定头部中心高度，以使眼睛处于与现实世界中相同的高度。对于全六自由度追踪的情况，甚至也会存在这个问题，接下来将介绍这个问题，那就是用户可能会坐着，可以添加一个垂直偏移量以使他看起来像站在 VR 中。

### 9.3 追踪位置和方向

本部分介绍追踪运动刚体的所有 6 个自由度，其中最重要的情况是头部追踪。为了方便起见，我们将把身体的位置和方向称为姿势。如(9.22)所述，六自由度追踪使得  $T_{eye}$  从 3.4 可以完全从传感器数据中获得，而不是从头部模型发现位置。通过估计位置，随着用户将头部左右移动，强大的视差深度提示变得更强。她甚至可以接近一个小物体，并从任何角度观察它，例如从上面，下面或侧面看。本节中的方法也可用于追踪空间中的手或在 VR 体验中操

作的对象。

## 为什么不整合加速度计？

试图用 IMU 单独完成 6 自由度追踪似乎非常的自然。回想图 9.5，加速度计测量真实线性加速度和重力加速度的矢量和。如果将重力分量从输出中减去，如启发式地完成倾斜校正，则看起来剩余部分是纯粹的身体加速度。为什么不简单地整合这个加速两次以获得位置估计？问题在于，漂移误差率比陀螺仪大得多。简单的校准误差会导致陀螺仪外壳中的漂移误差线性增长，因为它是一次积分的结果。经过双重积分后，校准误差会导致二次增长的漂移误差。在几分之一秒后，这个误差在实践中变得难以忍受。此外，真实的身体加速度无法准确提取，特别是当身体快速旋转时。最后，随着漂移的累积，可以使用什么传感器来估计位置漂移误差？单靠 IMU 无法帮助。请注意，它甚至不能以恒定速度区分运动，包括零运动，这与我们的前庭器官相同。尽管存在缺陷，现代 IMU 仍然是六自由度追踪系统的重要组成部分，因为它们具有高采样率和精确处理旋转部件的能力。

## 制造你自己的浪潮

基于 IMU 的追踪方法是被动的，因为它依赖于环境中已经存在的信息来源。相反，可以通过将波传播到环境中的方法。由于人类在相同的环境中工作，所以可以感知的波如光和声音都不是优选的。相反，主动追踪系统中的常见能源包括红外线，超声波和电磁场。

考虑从扬声器发射超声波脉冲(高于 20,000 赫兹)并使用麦克风收听其到达。这是一个发射探测器对的例子：扬声器是发射器，而麦克风是探测器。如果时间测量在源和目的地之间同步，则可以计算到达时间(TOA 或飞行时间)。这是脉冲通过发射器和检测器之间的距离  $d$  所花费的时间。根据介质中的已知传播速度(超声波 330 m / s)，估计距离  $d$ 。超声系统的一个不好的限制是表面之间的混响，导致在每个检测器处多次接收脉冲。

在正确运行时，探测器的位置可以缩小到以发射机为中心的半径为  $d$  的球体；见图 9.9(a)。

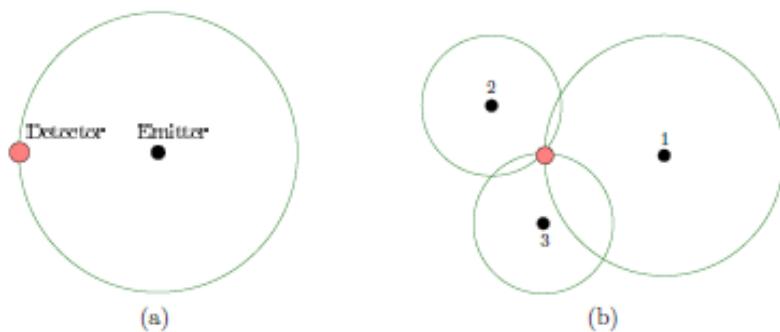


图 9.9：三边测量的原理使得探测器的位置可以通过估计到已知发射器的距离来确定。图示为一个 2D 示例：  
(a)来自单个发射器的探测器可以在沿着圆的任何地方；(b)使用三个发射器，该位置是唯一确定的。

通过使用两个发射器和一个麦克风，位置被缩小到两个球体的交点，产生一个圆圈(假设发射器位置已知)。使用三台发送器时，位置被缩小到两点，并且有四个或更多发送器，位置是唯一确定的。发送器和探测器的角色可以很容易地颠倒过来，以便被追踪的物体携带发射器，并将几个接收器放置在其周围。将这些测量结合起来以确定位置的方法称为三边测量。如果使用无线电，光线或红外线等电磁波代替超声波，则即使不可能直接测量传播时间，仍可以应用三边测量法。如果发射机幅度是已知的，则可以基于功率退化来估计距离，而不是 TOA。或者，可以发射时变信号，并且当接收到的信号叠加到发射信号上时可以估计其反射相移。

如果检测器不知道脉冲开始的准确时间，那么他们可以比较它们之间到达时间的差异；这被称为到达时间差(TDOA)。这组可能的位置是双曲面而不是球面。然而，双曲面片可以与多个发射器——检测器对交叉以获得多点定位的方法。这在第二次世界大战的迪卡导航系统中用于定位船只和飞机。我们的耳朵也使用这个原理来定位声音的来源，这将在第 11.3 节中介绍。

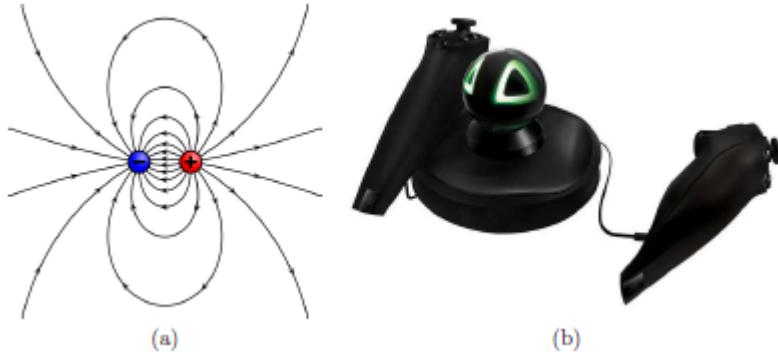


图 9.10：(a)磁偶极子提供一个随位置改变而改变其大小和方向的场。(b)Razer Hydra 游戏控制器系统，使用基站产生弱磁场，使其能够追踪控制器的位置。

最后，一些方法可以通过发射在追踪区域上变化的复杂场来追踪位置。例如，通过创建一个磁偶极子(可能用信号编码以区别于背景场)，可以在野外估计场中的身体位置和方向；见图 9.10(a)。Razer Hydra 追踪系统中的这种原理被用于产生磁场的基站中的视频游戏；见图 9.10(b)。一个缺点是该场可能在每个环境中变得不可预测地扭曲，导致直线运动被估计为弯曲。请注意，这些要求与 9.2 节中使用磁强计进行偏航校正所需要的相反；在这种情况下，场地需要在追踪区域保持不变。为了估算位置，该领域应该在不同地点有很大差异。

## 能见度

6 自由度追踪最强大的范例是可视性。这个想法是识别称为特征的物理世界的特殊部分，并计算它们沿着视线射线到达已知位置的位置。图 9.11 显示了一个受相机启发的例子，但也可以使用其他硬件。追踪的一个关键方面是区别。如果所有功能看起来都是相同的，那么在追踪过程中可能难以确定和维护“哪个是哪个”。随着时间的推移，每个特征都应该被分配一个独一无二的标签，因为世界上的刚体移动。相互混淆的特征可能会对身体姿势造成灾难性的糟糕估计。

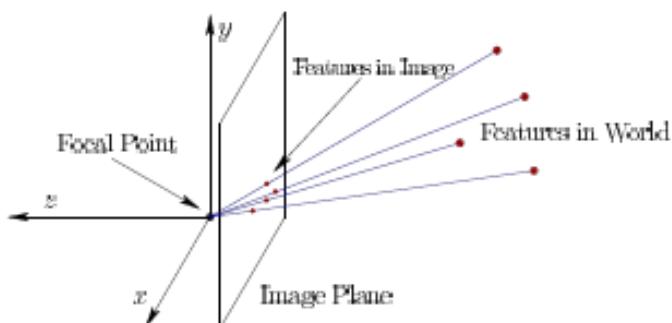


图 9.11：真实世界包含特殊的特征，这些特征被确定为位于通过透视投影连接到焦点的线段上。

用于检测功能的最常见的传感器是数码相机。检测，标记和追踪功能是计算机视觉或图像处理中的常见任务。有两个功能选项：

- 1.自然特征：在追踪过程中会自动发现这些功能，分配标签并进行维护。
- 2.人工特征：将功能设计并放置到环境中，以便可以轻松检测到这些功能，与预先分配的标签进行匹配并进行追踪。

自然特征是有利的，因为没有设置成本。环境不需要设计。不幸的是，它们也更加不可靠。使用相机，这被认为是一个硬计算机视觉问题，因为它可能像人类视觉系统一样具有挑战性。对于某些物体，纹理和光照条件，它可以很好地工作，但要使其可靠地用于所有可能的设置是非常困难的。想象一下，试图在空白的墙上找到并追踪功能。因此，人工特征在产品中更为常见。

对于人造特征，最简单的解决方案之一是将特殊标签打印到要追踪的物体上。例如，可以在对象上打印明亮的红色点，然后在图像中以红色斑点的形式扫描它们的外观。为了解决可区分性问题，可能需要多种颜色，例如红色，绿色，蓝色和黄色点。如果这些颜色自然存在于图像的其他部分，可能会出现问题。更可靠的方法是设计一个与图像其余部分明显不同的特定标签。这样的标签可以被编码以包含大量的信息，包括唯一的识别号码。最常见的编码标签之一是 QR 码，图 9.12 显示了它的一个例子。



图 9.12：示例 QR 码，可以打印并用作人工特征。 (来自维基百科的图片)

到目前为止描述的特征被称为被动，因为它们不释放能量。希望在这个世界上有足够的光线，以便足够反射该特征并进入相机传感器。更可靠的替代方案是设计发出自己的光的主动功能。例如，彩色 LED 可以安装在头盔或控制器的表面上。这是以需要电源并增加整体物体成本和重量为代价的。此外，其工业设计可能会因为它可能像圣诞树一样点亮而受损。



图 9.13：Oculus Rift 包含隐藏在 IR 透明塑料后面的 IR LED。(由 iFixit 拍摄)

## 红外线遮掩

幸运的是，所有这些技巧都可以转移到光谱的红外(IR)部分，以便相机可以看到这些特征，但对人不可见。图案可以绘制到高度反射红外能量的物体上，或者，IR LED 可以安装到设备上。Oculus Rift 就是这种情况，红外 LED 甚至隐藏在透明的红外能量塑料后面，但对人类来说却是黑色的，见图 9.13。

在某些情况下，可能很难将 LED 安装在物体上，例如追踪整个人体的细微运动。这称为 MOCAP 或动作捕捉，在第 9.4 节中进行了描述。在 MOCAP 系统中，强大的红外线发光二极管位于相机周围，以便它们照亮放置在场景中的反光标记。每个标记都可以被想象成光谱 IR 部分的球面镜。一个缺点是它的范围受到限制，因为红外能量必须从摄像机位置移动到目标并再次返回。由于能量作为距离的函数以二次方式消散，所以将距离加倍到达相机的四分之一的能量水平。

在这一点上，我们很自然地想知道，既然最终的图像处理问题微不足道，为什么要捕捉整个图像。主要原因是低成本数码相机和图像处理软件的普及。为什么不简单设计一个产生二进制读数的发射器-检测器对，指示可见光束是否被遮挡？这正是探测光束在自动车库门系统中的工作原理，以确保门不会关闭时夹到人：红外 LED 向检测光电二极管发射能量，该光电二极管本质上是一个开关，当它接收到足够的能量水平其目标波长(在这种情况下是 IR)。为了减少能量消耗的量，镜子或镜头可用于聚焦能量。

更好的是，红外激光可以直接瞄准探测器。接下来的任务是使用镜头和移动镜子，以便从某个固定位置可见的每个检测器都会在某个点照亮。光束可以使用透镜从一个点扩展到一条线，然后使用旋转镜在线移动。这是 HTC Vive 的灯塔追踪系统的基础，本节稍后会对此进行介绍。

## N 点透视(PnP)问题

使用  $n$  个观察特征需要“移动”刚体。这被称为  $n$  点透视(或 PnP)问题。我们可以借用第 3 章的大部分数学原理，然而，在这里我们考虑的是真实世界中的身体位置，而不是虚拟世界。此外，我们有一个反向问题，即根据图像中的点确定身体位置。到目前为止，已经考虑

了相反的问题。对于第7章中的视觉渲染，图像是基于(虚拟)世界中已知的身体位置生成的。这些功能可以放置在物体上或周围的世界中，具体取决于传感方法。现在假设他们在体上。每个特征对应于具有在身体的框架中定义的坐标的点  $p = (x, y, z)$ 。设  $T_{rb}$  是包含姿势参数的齐次变换矩阵，其假定是未知的。如(3.22)中所示，将变换  $T_{rb}$  应用于点  $p$  可以将其置于真实意义上的任何位置。回想一下变换链公式(3.41)，它进一步确定身体上的每个点在图像中出现的位置。矩阵  $T_{eye}$  保持相机姿态，而  $T_{vp}$  和  $T_{can}$  包含透视投影并将投影点转换为图像坐标。

现在假设已经观察到一个特征位于图像坐标中的位置  $(i, j)$ 。如果  $T_{rb}$  未知，但给出了所有其他变换，则将有 6 个独立参数进行估计，对应于 6 个 DOF。观察  $(i, j)$  为变换链(3.41)提供了两个独立的约束条件，其中一个为  $i$ ，另一个为  $j$ 。因此刚体失去 2 个自由度，如图 9.14 所示。这是 P1P 问题，因为  $n$  是特征的数量。



图 9.14：每个可见的特征消除了 2 个 DOF。在左侧，可以看到单个特征，而刚刚生成的刚体只剩下 4 个自由度。在右侧，两个特征是可见的，导致只有 2 个自由度。这可以如下所示。如果还施加了一些旋转，则接触两个分段的边缘可以在保持其长度的同时来回移动。围绕边缘共有的轴旋转提供第二个 DOF

P2P 问题对应于观察图像中的两个特征并导致四个约束。在这种情况下，每个约束消除了两个自由度，导致只剩下两个自由度，见图 9.14。继续下去，如果观察到三个特征，则对于 P3P 问题，保留零自由度(除了在身体上选择共线特征的情况)。看起来问题完全解决了；然

而，零自由度允许多种解决方案(它们是解决方案空间中的孤立点)。P3P 问题对应于试图将给定的三角形放入由光线形成的金字塔中，以便每个三角形顶点触及不同的光线。这通常可以通过四种方式完成，难以形象化。想象一下，试图将由奶酪制成的高而薄的金字塔(单面)切片，以使四个不同的切片具有完全相同的三角形尺寸和形状。P4P 和 P5P 的情况也会导致模糊的解决方案。最后，在 P6P 的情况下，如果没有四个特征共面，总会得到独特的解决方案。所有的数学细节都在[354]中计算出来。

PnP 问题已经被描述为理想的情况，即对身体上的特征点进行完美的坐标分配，并通过成像过程完美地观察这些特征点。在实践中，由于诸如传感器噪声，图像量化和制造容差等因素而产生小误差。这会导致估计姿态的模糊和错误，这可能与正确答案有很大差异[281]。因此，在实践中可以使用更多的特征来提高准确性。此外，可以在使用该设备之前应用诸如光束法平差[111,280,325]之类的校准过程，使得可以在执行姿态估计之前更精确地分配特征点位置。使用 RANSAC 可以提高鲁棒性[77]。

## 基于相机的实现

可见性问题可以使用相机，通过两种常规方式来解决，如图 9.15 所示。相机镜头是类似于第 3 章中图 3.14 的眼镜框。世界固定的相机通常是静止的，这意味着相机镜头相对于世界不会移动。可以使用单个变换来将从相机框架估计的对象姿态转换为简单的世界框架。例如，对于 Oculus Rift 头戴式设备，头部姿态可以转换为世界坐标系，其中-z 方向指向相机，y 为“上”，并且位置位于相机的中心追踪区域或基于用户初始头部位置的合适默认值。对于固定对象的摄像机，估计的姿态是从相机画面转换到世界画面的，是从世界上仍然固定的物体得如果将 QR 码放在墙上，就可以获得这种情况。

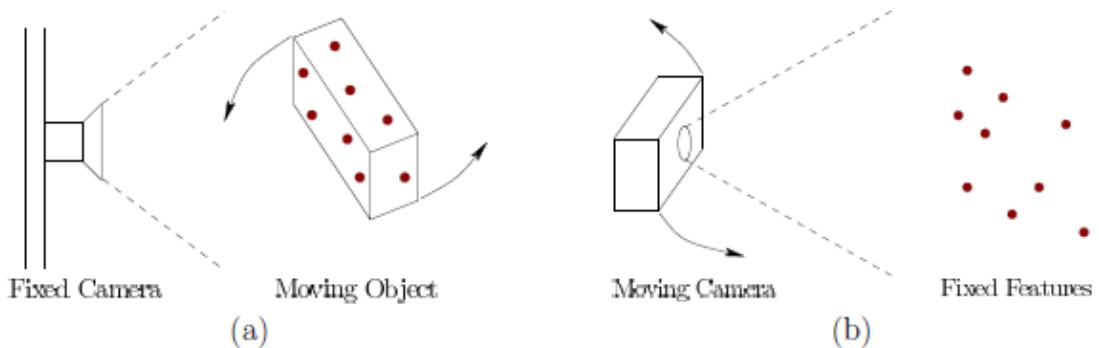


图 9.15：摄像机放置的两种情况：(a)世界固定的摄像机是静止的，并且物体相对于它的运动使用物体上的特征进行估计。(b)物体固定摄像机经常处于运动状态，理想状态下的特征固定在世界坐标系中。

就像 IMU 一样，校准对于提高传感精度非常重要。以下齐次变换矩阵可以应用于相机产生的图像：

$$\begin{bmatrix} \alpha_x & \gamma & u_0 \\ 0 & \alpha_y & v_0 \\ 0 & 0 & 1 \end{bmatrix} \quad (9.23)$$

矩阵中出现的五个变量称为相机的内部参数。 $\alpha_x$  和  $\alpha_y$  参数处理缩放， $\gamma$  处理剪切， $u_0$  和  $v_0$  处理光轴偏移。这些参数通常是通过拍摄所有尺寸和距离都经过仔细测量的物体的图像并进行最小二乘估计来选择减少平方和误差的参数(如 9.1 节所述)来估算的。对于广角镜头，可能需要进一步校准以克服光学失真(回忆第 7.3 节)。

现在假设在图像中观察到一个特征，也许使用某种形式的斑点检测从图像的其余部分提取与其对应的像素[280,318]。这对全局快门相机来说是最简单的，因为所有像素都将对应相同的时间。在卷帘快门的情况下，可能需要变换图像以消除运动的影响(回忆图 4.33)。观察到的特征的位置被计算为斑点像素位置的统计量。最常见的情况是，使用所有斑点像素的平均值，导致非整数图像坐标。许多问题会影响性能：1)由于每个像素的图像坐标都是整数，所以会出现量化误差；2)如果该特征没有覆盖足够的像素，则量化误差更差；3)照明条件的变化可能使得难以提取特征，特别是在自然特征的情况下；4)在某些角度，图像中两个或多个特征可能变得接近，使得难以分离它们相应的斑点；5)由于各种特征进入或离开摄像机视图，所得到的估计姿态可能会跳跃。此外，误差往往沿光轴方向变大。

## 基于激光的实现

通过设计一个特殊的发射器-探测器对，能够在很远的距离上准确地解决可见性问题。这是通过 2016 年 HTC Vive 的灯塔追踪系统和 1989 年的明尼苏达扫描仪[301]完成的。图 9.16 显示了 HTC Vive 的灯塔追踪硬件，图 9.17(a)有效地模拟了摄像机的操作。



图9.16：HTC Vive 头戴式设备中使用的基于激光的追踪方法：(a)基站包含旋转鼓，可发出水平和垂直的红外光片。左上方出现一排红外LED，提供同步闪光光。(b)头戴式头盔前面的口袋中的光电二极管检测入射的IR 光。

如果基站是摄像机，则扫描垂直带将对应于估计表征该特征的像素行，见图9.17(a)。同样，扫描的水平条纹对应于像素列。旋转鼓的旋转速率是已知的，并且与相机帧率类似。精确的时间记录为光束击中每个光电二极管。

考虑相对于基站的极坐标(距离和角度)。使用扫描的角度和相对定时差，可以容易地估计从基站“观察”的特征之间的角度。尽管特征之间的角度很容易确定，但它们相对于来自基站的某个固定方向的角度必须确定。这是通过一系列同时脉冲的红外 LED 实现的，以便所有光电二极管都能检测到闪光光(如图 9.16(a)所示)。例如，这可以对应于每个光束处于 0 方向的时刻。基于从闪光光直到光束撞击光电二极管的时间以及已知的角速度，确定所观察特征的角度。为了减少时间漂移误差，可以在操作期间周期性地使用闪光灯。

与照相机一样，从基站到特征的距离也是未知的，但可以通过解决 PnP 问题来确定。也可以使用多个基站，其方式与使用多个摄像头或多个眼睛来推断深度相当。结果是在大面积上进行精确追踪，如图 9.17(b)所示。

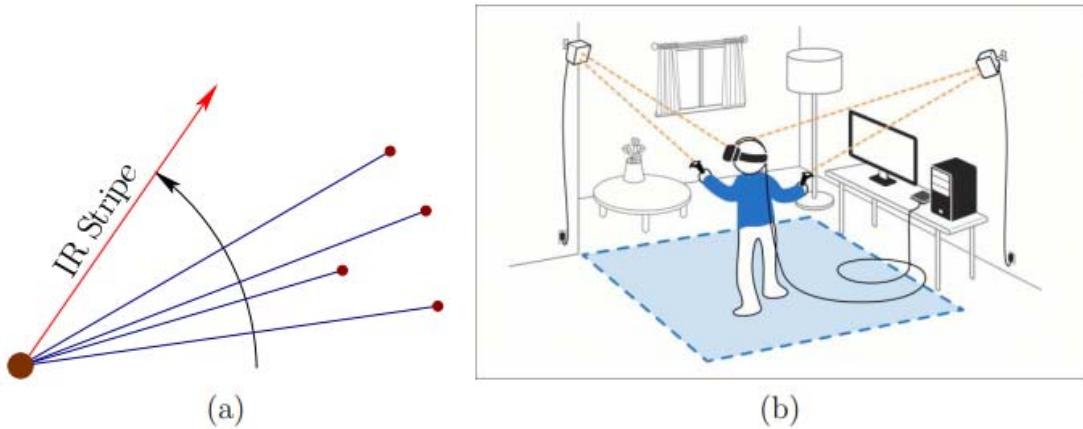


图 9.17: 这是基于激光追踪方法中红外条纹角度扫描的 2D 视图(如 HTC Vive)。对应于顶视图，其中垂直条纹围绕底座旋转偏航旋转。在这种情况下，观察水平方向上的角度位置，类似于相机图像的列坐标。这也对应一个方面在这种情况下，观察垂直条纹以俯仰旋转和垂直方向上的角度位置自旋。当光束击中光电二极管的特征时，由于自同步闪烁光以来的旋转速率和时间，所以方向是已知的。(b)通过将两个基站放置在追踪区域的角落处的极点之上，可以精确地追踪头盔和控制器的大区域。(由Chris Stobing 所画)

### 滤波

如第 9.2 节所述，来自传感器的输出通过滤波方法随时间合并以维持预测。在当前的设置中，可以通过组合可见性信息和 IMU 的输出来维持姿态。对于姿态的方向分量，可以使用(9.10)中的互补滤波器。相机为检测方向漂移误差提供了额外的来源。摄像机光轴是偏航误差估计检测的直接参考，这使得它成为磁力仪的一个合适的替代品。如果相机倾斜度已知，则相机还可以提供精确的倾斜误差估计。

由于陀螺仪提供的角速度的准确，高频估计，IMU 对获得高度精确的定向追踪至关重要。如果摄像机或灯塔系统的帧速率非常高，则可能存在足够的传感器数据以进行精确的位置追踪；但是，最好直接测量衍生工具。不幸的是，IMU 不能测量线速度。但是，线性加速度计的输出可以按照本节开始部分的方法进行使用。假设加速度计估计出的世界框架内的主体加速度为(假定重力分量已从加速度计输出中减去)。

$$\hat{a}[k] = (\hat{a}_x[k], \hat{a}_y[k], \hat{a}_z[k]) \quad (9.24)$$

通过数值积分，可以通过  $a[k]$  估计速度  $v[k]$ 。通过对速度估计进行积分来估计位置  $p[k]$ 。使用简单欧拉积分的更新方程如下

$$\begin{aligned} \hat{v}[k] &= \hat{a}[k]\Delta t + \hat{v}[k-1] \\ \hat{p}[k] &= \hat{v}[k]\Delta t + \hat{p}[k-1]. \end{aligned} \quad (9.25)$$

请注意，每个方程实际上同时处理三个分量  $x$ ,  $y$  和  $z$ 。通过在右侧添加  $\frac{1}{2}\hat{a}[k]\Delta t^2$ ，可以进一步提高第二个方程的精度。

如前所述，加速度的双重积分导致快速增长的位置漂移误差，用  $d_p[k]$  表示。PnP 解决方案检测到的误差提供了  $d_p[k]$  的估计值，但可能比 IMU 产生的观测值低很多。例如，相机可能以 60 FPS 拍摄照片，而 IMU 可能会以 1000 FPS 记录加速度。

从(9.10)的互补滤波器可以扩展到双重积分的情况下，我们得到

$$\begin{aligned} p_c[k] &= \hat{p}[k] - \alpha_p \hat{d}_p[k] \\ v_c[k] &= \hat{v}[k] - \alpha_v \hat{d}_p[k]. \end{aligned} \quad (9.26)$$

上式中,  $p_c[k]$  和  $v_c[k]$  分别是校正的位置和速度, 它们分别由互补滤波器计算。估计值  $p[k]$  和  $v[k]$  使用(9.25)计算。与 IMU 更新相比, 参数  $\alpha_p$  和  $\alpha_v$  控制给予漂移误差估计的重要程度。

方程(9.26)实际上等价于卡尔曼滤波器, 它是具有高斯噪声的线性动力系统的情况以及也受高斯噪声影响的传感器的最佳滤波器(提供可能的最精确的估计)。设  $\omega_d^2$  表示动态系统中估计的高斯噪声的方差, 并设  $\omega_s^2$  表示传感器噪声方差。如果参数选择为

$\alpha_p = \sqrt{2\omega_d/\omega_s}$  和  $\alpha_v = \omega_d/\omega_s$  [119], 互补滤波器(9.26)等效于卡尔曼滤波器。存在各种各样的替代过滤方法; 然而, 相对于校准, 传感器误差模型和特定于设置的动态系统模型, 使用不同滤波方法的影响通常较小。此外, 性能要求主要是基于感知的, 这可能与设计滤波方法的经典标准不同[167]。

一旦过滤器运行, 其姿态估计可用于帮助 PnP 问题。PnP 问题可以通过使用最新的加速度计输出扰动滤波器估计的姿态来逐步解决, 以便观察到的特征完美匹配。可以对姿势进行小的调整, 以便将平方和误差降低到可接受的水平。在大多数情况下, 当 PnP 问题具有模糊的解决方案时, 可见的特征很少, 这可以提高可靠性。如果没有逐步地确定姿态, 可能会灾难性的跳转到另一个 PnP 解决方案。

## 9.4 追踪附属体

许多追踪问题涉及估计一个物体相对于另一个附着的运动物体的运动, 例如眼睛在眼窝内旋转。虽然眼睛在空间被视为一个刚体时可能有六个自由度, 但是一旦给出头部姿态, 它的位置和方向就可以用两个或三个参数完全表征。其他例子包括相对于躯干的头部, 相对于手腕的手部, 以及手指相对于其中间骨头的尖端。基于骨架, 整个人体甚至可以排列成树形的附属体。此外, 也可以通过类似的方式发现其他生物体上的附着体, 例如狗或猴子, 以及机器, 例如机器人或汽车。在汽车的情况下, 车轮相对于车身旋转。在所有这些情况下, 结果都是多体系统。物体相对于彼此姿态的数学表征被称为多体运动学, 并且完全确定它们的速度和加速度被称为多体动力学。

### 人眼追踪

一个多世纪以来, 视觉科学家一直在使用眼动追踪系统来研究眼球运动。VR 的三个主要用途是: 1)如第 5.4 节所述, 完成中心渲染, 以便高分辨率渲染只需要在中心凹上的图像部分执行。2)通过记录追踪数据来研究人类行为, 从而帮助治疗 VR 疾病, 提高注意力和体验的有效性等。3)在 VR 中渲染视线的方向, 以便通过提供眼神接触和指示某人关注的焦点来改善社交互动; 参见 10.4 节。

现如今已经开发了三种通用类型的眼动追踪方法[63,338]。第一种是眼电图(EOG), 它从位于每只眼睛周围的面部皮肤上的多个电极获得测量结果, 电位对应于眼肌活动, 通过滤波来确定相对于头部的眼睛方向。第二种方法使用隐形眼镜, 隐形眼镜包含一个微小的电磁线圈, 导致周围电磁场的潜在变化。第三种方法称为视频眼动(VOG), 它将红外光照射到眼睛上, 并使用相机或光电二极管感测其角膜反射。反射基于 Purkinje 图, 如图 9.18 所示。由于其成本低, 侵入性小, 这是目前最常用的方法。隐形眼镜方法是最准确的; 然而, 也是最不舒服的。

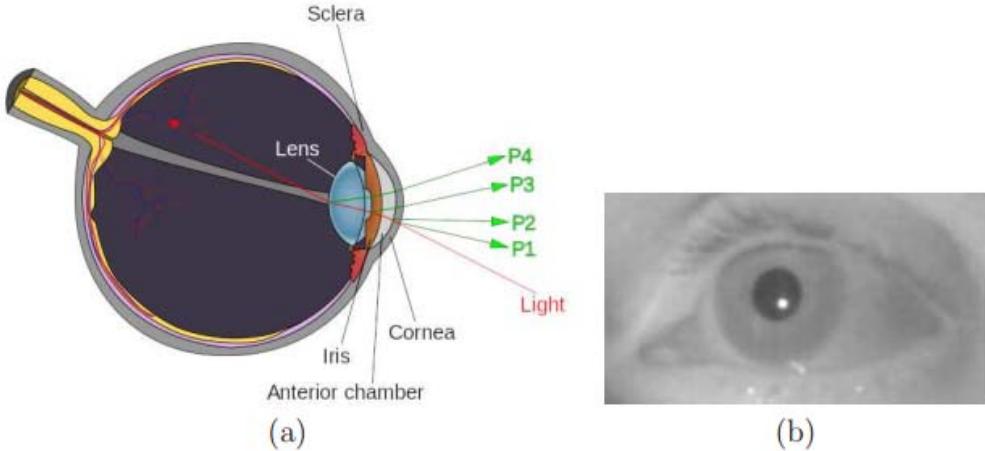


图 9.18: (a)红外光源的第一幅, 有时是第四幅 Purkinje 图像用于人眼追踪(图源于 Wikipedia)(b)第一幅 Purkinje 图像如图所示产生明亮的反射(图源于 Massimo Gneo, Maurizio Schmid, Silvia Conforto, Tomasso D'Alessio)

## 正向运动学

假设眼睛追踪方法估计了相对于人类头骨的眼睛方向, 然后需要将相应地放置在虚拟世界中。这种转变必须涉及头部和眼部变换的组合。对于一个更复杂的问题, 考虑每个关节处使用躯干的姿势以及骨骼之间形成的所有角度, 在虚拟世界中将手指的位置放准确。要理解这些和其他相关问题如何解决, 我们先介绍一些 2D 的例子。

多体系统的每个主体称为连杆, 并且一个主体连杆在一个关节处, 这允许它们之间存在一个或多个运动自由度。图 9.19 显示了一个平面物体在附着到另一个平面物体时常见的两种运动方式。旋转关节最为常见, 并表示人体肘部所允许的运动。

考虑定义  $m$  条连杆,  $B_1$  到  $B_m$ , 并确定两个连杆相交点的位置。每个连杆上的点是使用其自身的主体框架的坐标定义的。在这个框架中, 主体如图 9.20 中的  $B_{i-1}$  所示, 原点位于连接  $B_{i-1}$  到  $B_i$  的连接点, 而  $x_{i-1}$  轴指向连接  $B_{i-1}$  到  $B_i$  的连接点。将  $B_i$  上的点移动到  $B_{i-1}$  的主体框架中的适当位置, 应用齐次变换

$$T_i = \begin{pmatrix} \cos \theta_i & -\sin \theta_i & a_{i-1} \\ \sin \theta_i & \cos \theta_i & 0 \\ 0 & 0 & 1 \end{pmatrix}. \quad (9.27)$$

这将  $B_i$  旋转  $\theta_i$ , 然后沿  $x$  轴将其平移  $a_{i-1}$ 。对于旋转关节,  $\theta_i$  是一个变量,  $a_{i-1}$  是一个常数。对于棱形关节,  $\theta_i$  是常数,  $a_{i-1}$  是一个变量。

通过应用乘积  $T_2 \cdots T_i$  将  $B_i$  上的点移入  $B_1$  的主体框架中。图 9.21 显示了一个三连杆的例子。移动第一条连杆  $B_1$  进入世界框架, 可以应用一般的 2D 齐次变换:

$$T_1 = \begin{pmatrix} \cos \theta_i & -\sin \theta_i & x_t \\ \sin \theta_i & \cos \theta_i & y_t \\ 0 & 0 & 1 \end{pmatrix}. \quad (9.28)$$

这种变换只是通过将  $T_1 T_2 \cdots T_i$  添加到矩阵乘积中来移动每个  $B_i$ 。

3D 连杆链在概念上以相同的方式处理, 但代数运算变得更加复杂。有关更多详细信息, 请参见[163]的第 3.3 节。图 9.22 显示了通过允许一对 3D 连杆相互滑动而获得的六种不同

类型的关节。每个连杆都根据关节分配了一个方便的坐标系。每个齐次变换  $T_i$  包含常量和变量的混合，其中变量对应于关节允许的自由度。最常见的分配方案称为 Denavit-Hartenberg 参数[110]。在某些设置中，可能最好将每个  $T_i$  替换为旋转主体的参数化四元数，然后添加一个简单的加法对主体做平移。

连杆树也可以考虑；一个常见的例子是人体躯干作为根部，头部，两条手臂和两条腿是从它延伸出来的连杆。人的手是另一个例子。这种情况下坐标系通常使用 Kleinfinger-Khalil 参数[148]分配。

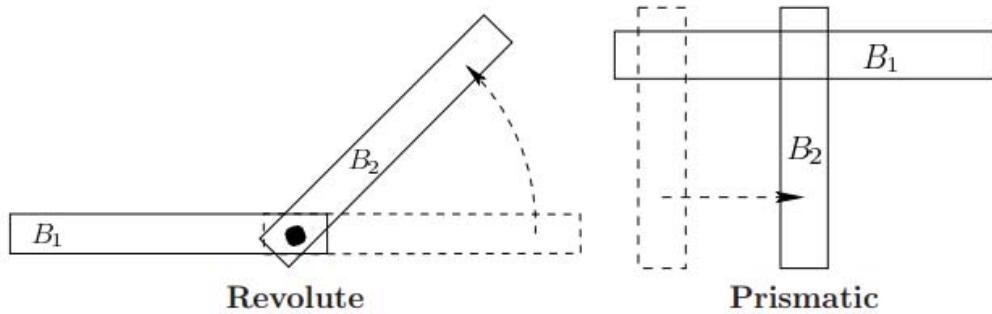


图 9.19：两种类型的 2D 关节：旋转关节允许一个连杆相对于另一个旋转，并且棱形关节允许一个连杆相对于另一个进行平移。

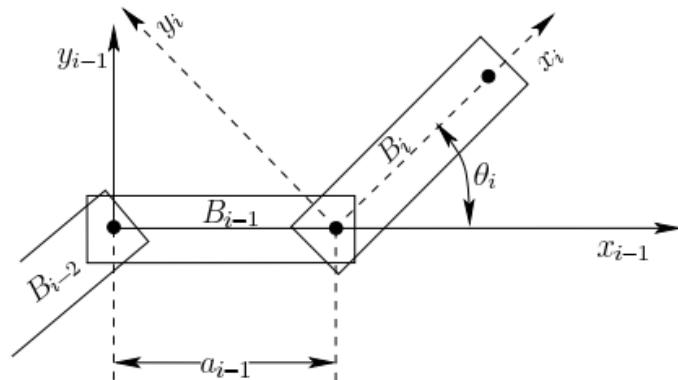


图 9.20：对于  $1 < i < m$ ，主体框架的每个  $B_i$  基于这些关节连接到  $B_{i-1}$  和  $B_{i+1}$ 。

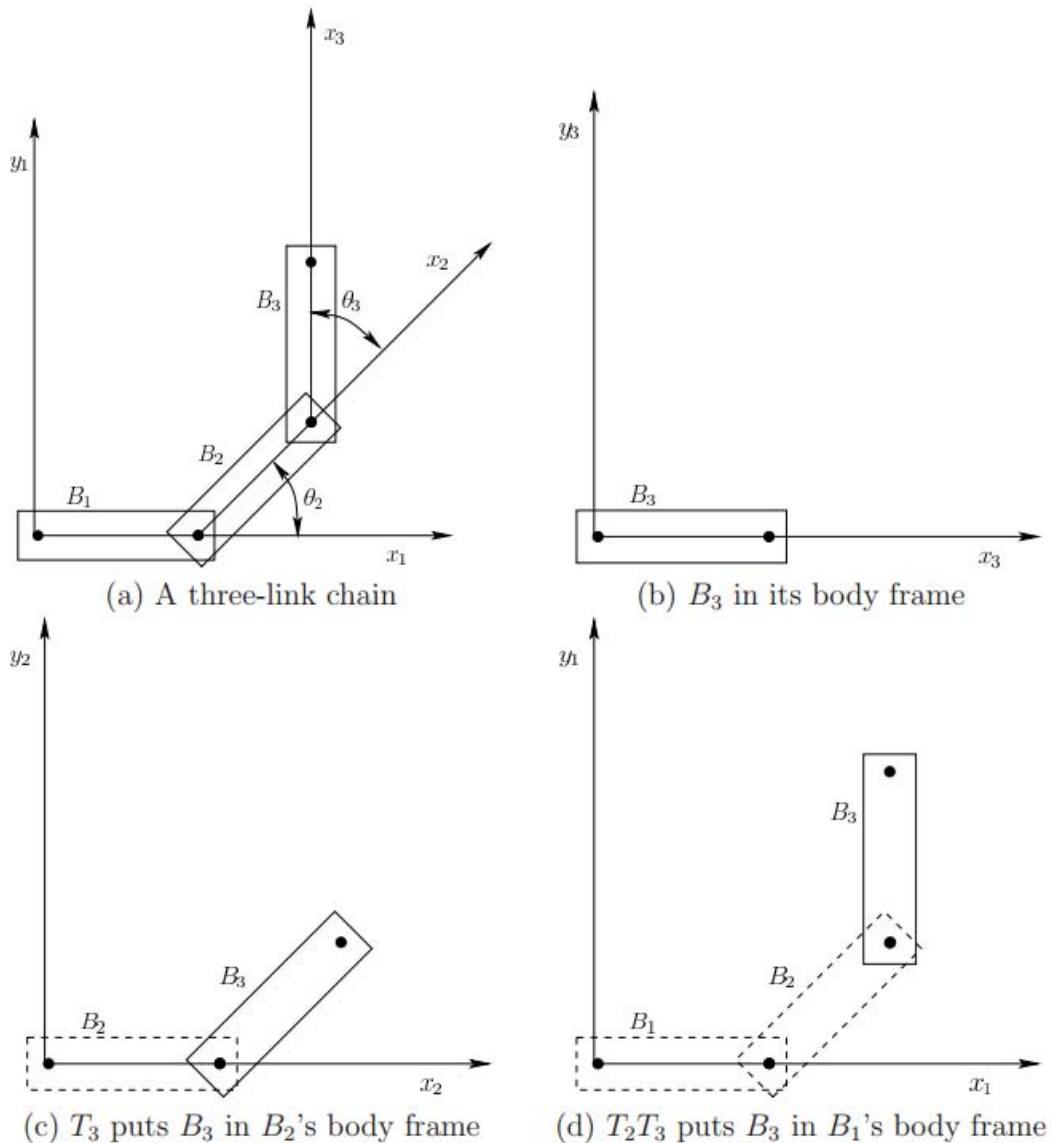


图 9.21：将变换  $T_2T_3$  应用于  $B_3$  的模型上。如果  $T_1$  是单位矩阵，那么这产生了  $B_3$  中虚拟世界中的点的位置。

### 约束及逆向运动学

回顾第 9.3 节中的 PnP 问题，该问题涉及到通过观察到的约束来计算物体的姿态。一个类似的问题是通过考虑对主体的约束来确定一连串主体的关节参数。一个常见的例子是仅使用手的姿势来计算手臂连杆的姿势，我们称之为逆向运动学问题（参见[8]和[163]的第 4.4 节）。就 PnP 而言，解的数量可以是无限的，有限的，一个或零个。一些 2D 例子如图 9.23 所示。通常，如果最后一个链路受到限制，则中间链路的运动自由度随着链路数量的增加而增加。假设连杆不是在一些特殊的单一配置中的情况，Chebychev-Grübler-Kutzbach 准则可以给出自由度的数量[9]。动画视频游戏角色的一个常见问题是保持运动学约束，例如手抓住门把手，即使躯干或门正在移动。在这种情况下，迭代优化通常用于执行每个联合参数，直到误差被充分减小。在我们的例子中，通过测量手和门把手之间的距离来计算误差。

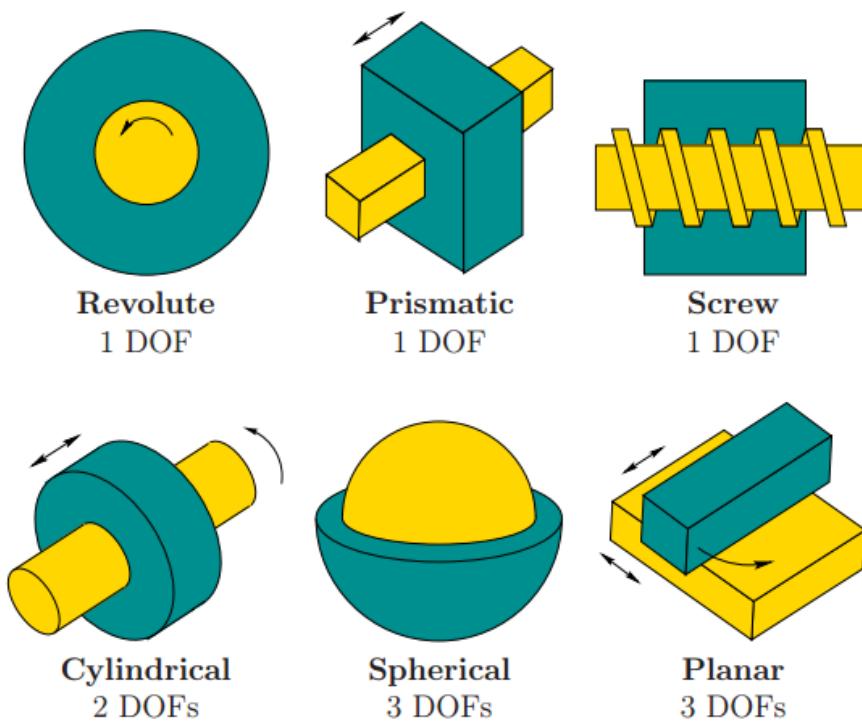


图 9.22：由两个物体之间的 2D 表面接触生成 3D 关节的类型。

## 运动捕捉系统

附属体的追踪系统使用运动学约束来提高准确性。最常见的应用是追踪人体，对于该人体而言，骨骼在连杆和关节方面是很好理解的[366]。这种动作捕捉系统一直是电影行业的重要技术，因为真人演员的动作被带入虚拟动画世界。操作如图 9.24 所示。第 9.3 节介绍的特征与人体相同，并可在捕捉工作室周围安装摄像头。对于可见度也存在同样的选择，过去十年中最常用的方法是使用带有周围红外 LED 的摄像头并将反光标记放置在演员身上。

为了获得每个身体部位的独特姿势，似乎需要六个特征(回顾第 9.3 节中的 P6P);然而，由于运动学约束，并不需要这么多。如果我们是为了在沿着骨架移动时捕捉皮肤运动，则可以使用额外的特征。这对于面部运动尤其重要。目前许多新的 MOCAP 技术正在开发中。例如，Noitom 开发的系统仅通过将 IMU 放置在身体上来捕捉人体运动。某些系统仅通过摄像头捕捉运动，如 Leap Motion(见图 9.25)用于手部追踪的情况，以及由微软和 8i 用于通过在绿色屏幕上提取轮廓来进行全身追踪的系统。基于现代深度传感器的解决方案在不久的将来也可能会流行起来。一个挑战是降低成本和安装高度准确可靠的系统。

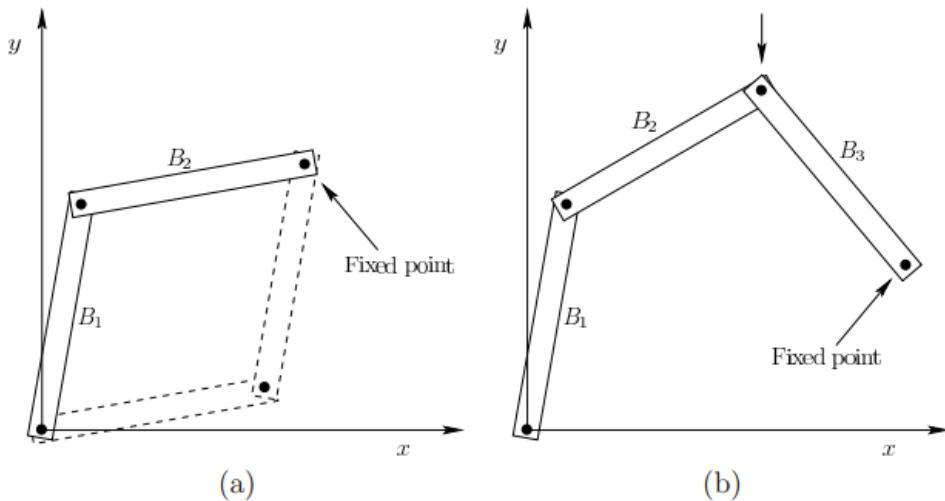


图 9.23: (a)两条连杆的方向可以从固定点的位置推断;然而,如果角度不受限制,还有第二种解决方案。(b)在三条连杆的情况下,当一端固定时,存在一维解决方案族。可以通过向下按顶部接头来观察,这会导致  $B_1$  逆时针旋转。这相当于传统的四杆机构,用于驱动蒸汽机的车轮(第四个“连杆”仅仅是固定的背景)。

## 9.5 环境三维扫描

目前为止,本章介绍了如何使用传感器来追踪一个或多个刚体的运动。在本节中,我们讲述如何使用传感器来构建刚体的几何模型,可能是可移动的,也可能是固定的,如 3.1 节所介绍的。可移动的模型通常对应于由用户操控的物体,如剑、锤子或咖啡杯等。这些模型通常由 3D 扫描器构建,以受控的方式从不同角度对物体成像。物体放置在被摄像机或其他传感器包围的平面上,或者放置在转台上,以便从不同的视点角度来观察物体。或者,也可以采取物体保持静止不动而传感器在周围移动,如图 9.26(a)。

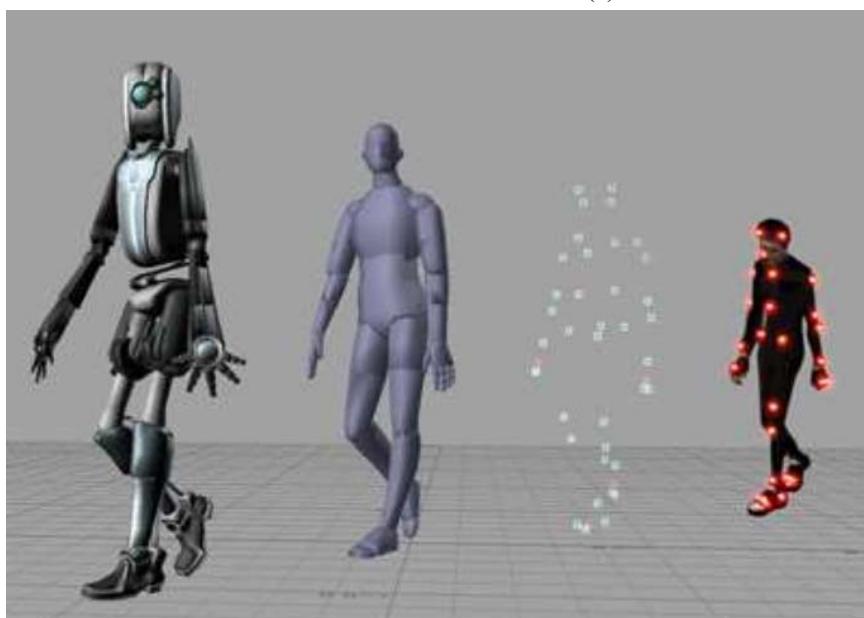


图 9.24: 通过运动捕捉系统(MOCAP),将特征点放置在演员的身体周围。提取其运动并与运动学模型进行匹配。模型中的每个刚体都有一个相关的几何模型,将该模型渲染成最终的动画角色。(图源于 Wikipedia 用户 Hipocrite)

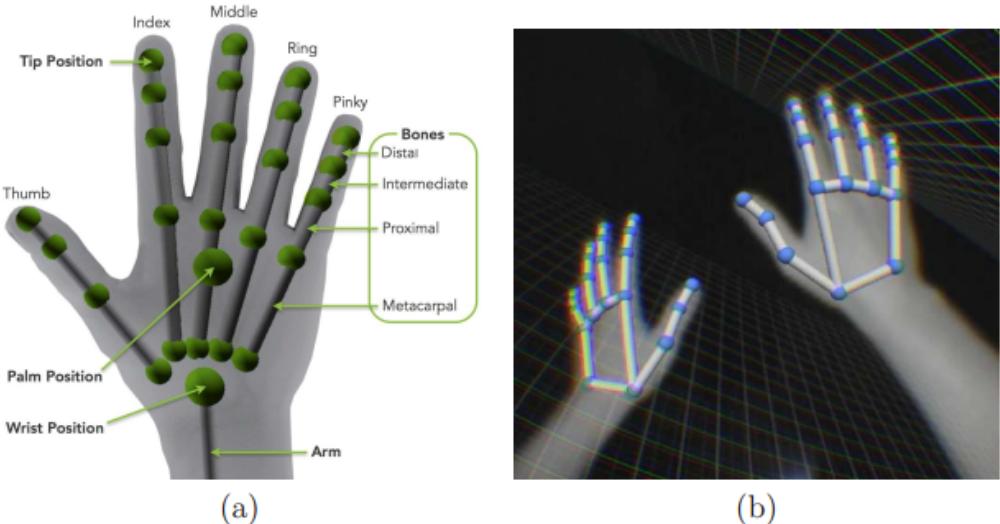


图 9.25: (a) 使用 Leap Motion 追踪得到的手部模型 (b) 将追踪模型叠加在实际手部的图像上

## SLAM

3D 扫描器适用于较小的物体，里面的传感器是面朝内部的。对于较大的物体和固定模型，传感器通常是面朝外部的，如图 9.26(b)。静态模型的一个常见例子是建筑物的内部。扫描这些模型对于调查和取证变得越来越重要。这也是经典的机器人领域中的地图构建问题，携带传感器的机器人以导航和避免碰撞为目的建立其环境的 2D 或 3D 表示。机器人通常需要根据传感器估算其位置，这就是所谓的定位问题。VR 的机器人定位和追踪机构基本上是同样的问题，主要的区别在于将已知的动作命令给予机器人，但并没有直接给出其对应的人类意图。机器人经常需要同时完成地图构建和定位的任务，这会导致即时定位与地图构建问题；常用缩写 SLAM 表示。由于定位，地图构建和 VR 追踪问题的相似性，机器人和 VR 之间存在着深层次的联系。因此，许多数学模型，算法和传感技术也有所重叠。

考虑到 VR 的大型固定模型的可能用途。可以将其捕获然后给予用户在当前时间或稍后时间的虚拟世界。图像数据可以与模型的三维坐标相结合，生成照逼真的模型(回忆 2.2 节的图 2.14)。这是通过将图像块贴图到模型三角形上来实现的。



图 9.26: (a)Afinia ES360 扫描仪，物体在转台上旋转时可以生成物体的 3D 模型 (b)FARO Focus3D X 330 时一部朝外的扫描仪，用于构建大型环境的精准 3D 模型；它包含一个 GPS 接收器，用于将独立的扫描结果

融合成一幅连贯的地图

## 当前场景的实时捕捉

传感器可以用来捕捉用户当前正在体验 VR 的物理世界，而不用去捕捉要传到用户眼中世界。这允许匹配区域中的障碍物在虚拟世界中呈现，这对于安全性或改善交互性可能是有用的。为了安全起见，可以对匹配区域的边界进行渲染，以提醒用户。像是一杯咖啡洒了或宠物穿过匹配区之类的危险情形也应当指示出来。通过将来自物理世界的固定物体带入虚拟世界可以提高交互性。例如，如果用户坐在桌子前面，则可以在虚拟世界中绘制一张桌子。如果她推一下虚拟桌子，她会感觉到真正的桌子在向后退。这是在 VR 中提供触碰反馈的相对简单的方法。

## 全景体验已经可以满足了吗？

在着手创建周围 3D 世界的大型详细地图之前，我们要考虑是否有必要。如第 7.5 节所述，现如今的全景图像和视频越来越容易捕捉。在一些应用中，建立用户在从彼此接近的多个位置捕捉的全景图之间进行传送的体验可能就足够了。

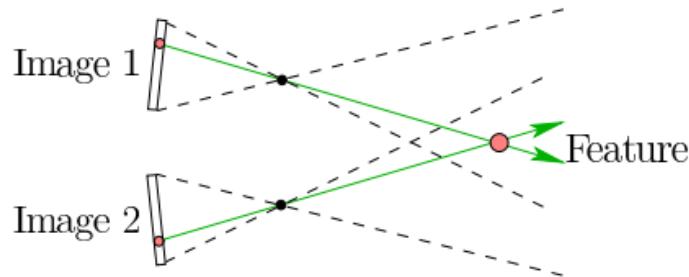


图 9.27：通过使用两个摄像头，立体视觉使得 3D 世界中的特征的位置能够通过与来自每个摄像头的相应可见光线相交来确定。为此，必须知道摄像机的校准参数和相对姿态。同样，一个照相机可以被照亮该特征的激光器替代，以使其对于其余照相机可见。在任何一种情况下，原理都是通过交叉两条可见光线来得到结果。

## 主要的要素

利用传感器数据建立 3D 模型的步骤如下：

1. 从固定位置提取 3D 点云
2. 将多个位置的点云结合起来
3. 将点云转换为三角形网格

第一步，在提取 3D 点的同时将传感器放置在固定的位置和方向，这可以通过多种方式来完成。理论上，6.1 节中的任何深度线索都可应用于相机图像以提取 3D 点。焦点，纹理和阴影的变化通常在计算机视觉中被用作单目线索。如果两台摄像机面对相同的场景并且它们的相对位置和方向已知，则使用双目线索来确定深度。通过识别两幅图像中相同的自然特征，可以通过每个图像相应的射线相交的位置来识别空间中的点，如图 9.27。

如第 9.3 节所述，存在自然和人为特征之间的选择。可以组合使用单台摄像机和红外投影仪或激光扫描仪，通过识别图像中亮点出现的位置来提取深度。这是 Microsoft Kinect 传感器的基础(请回顾第 2.1 节中的图 2.10)。由此产生的 3D 点集合通常称为点云。

在第二步中，问题在于如何合并来自多个位置的扫描数据。如果扫描的时候扫描器之间的相对位置和方向已知，那么问题很好解决。在图 9.26(a)所示的物体扫描仪的情况下，这是通过在转台上旋转物体来实现的，以便位置保持固定，并且每次扫描的方向都是精确已知的。

假设传感器是由安装在机器人上的，例如无人机。机器人通常保持自己的姿态估计，以避免冲突并确定其任务是否达到。这对于确定与执行扫描的时间相对应的姿势也是有用的。通常情况下，姿态估计不够精确，所以要解决估计姿态变化的优化问题，直到重叠扫描之间的数据很好地对齐。在这种情况下，通常使用 $\hat{\text{EM}}$ 算法，该算法在统计学意义上以产生数据的最大似然解释的方式递增调整姿态。如果传感器由人携带，则扫描设备可能包含额外的传感器，如图9.26(b)中扫描仪的GPS情况；否则，融合来自多个扫描的数据的问题会变得非常困难。

第三阶段，我们已经获得了大量的点云，问题是产生一个纯粹的几何模型。这个过程存在着很多的困难。首先，点密度可能差异很大，尤其是在进行两次或多次重叠扫描的情况下。在这种情况下，有些点可能会被丢弃。另一个问题是可能存在异常值，它们对应着远离其正确位置的孤立点。我们需要一些方法来检测和删除异常值。还有一个问题是数据中可能存在大的空洞或间隙。一旦数据已被充分清理，通常使用表面来拟合数据，从而形成三角网格。每一个问题本身都可以成为一个研究方向。想要更加熟悉整个过程，可以考虑试用开源的Point Cloud Library，它可以处理第二和第三阶段出现的操作。一旦获得三角形网格，如果图像数据也可用，就可以执行纹理贴图操作。VR面临最大的挑战之一是，相较于在计算机屏幕上显示，模型的许多缺陷通常在VR显示中观察的更为明显。

## 拓展阅读

除了[82]等学术论文外，IMU校准的一些最有用的报道出现在公司白皮书中，例如[244]。有关磁力计校准，请参见[92,155,165,329]。[166, 164, 165, 168]中介绍了Oculus Rift 3D方向追踪。要充分理解基于视觉的追踪方法，请参阅视觉书[111,191,318]。许多研究文献中的PnP appear方法，如[354,367]。对于VR/AR的追踪方法的一个很好但较为远古的方法是[344]。被引用最多的作品之一是[141]。关于集成IMU和可视化数据进行追踪可以参见[235]。

眼动追踪在[63,338]中进行了调查。人体追踪在[368]中介绍。要充分理解运动学约束和逆向运动问题的解决方案，请参阅[8,10,48]。从机器人的角度来看SLAM在[322]中有详细介绍。最近一项基于计算机视觉的SLAM调查出现在[86]中。在更大范围内的过滤或传感器融合可以用信息空间来表示（见[163]的第11章）。

# 第十章 交互

在 VR 中，用户应该如何与虚拟世界交互呢？他们应该如何移动？他们如何抓住和放置物体？他们应该如何与对方互动？他们如何使用互联网？下文则为上述问题提供了许多见解。

## 通用仿真原理

VR 可以模拟来自真实世界的任何交互机制。

例如，用户可以通过转动旋钮并拉动来打开门。此外，用户还可以坐在模拟驾驶舱内操作虚拟飞机（如图 1.15 所示）。人们甚至可以戴上 VR 显示器，从而获得与梦境相媲美的体验！

尽管有通用的模拟原则，但回忆第 1.1 节，VR 的目标不一定是要完全契合现实，而是追求更好的交互感。因此，本章将介绍在物理世界中没有的交互机制。

第 10.1 节介绍了一般的运动学习和控制概念。最重要的概念是重映射，其中现实世界中的运动可以被映射到虚拟世界中的另一种运动。这一概念为许多强大的交互机制的实现奠定了基础。这些机制的任务是开发易于学习，易于使用，快速执行并提供舒适用户体验的交互流程。第 10.2 节讨论了用户如何在虚拟世界中移动自己，同时保持现实世界中的固定状态。第 10.3 节介绍了用户与虚拟世界中其他对象进行交互的方式。第 10.4 节讨论了社交互动机制，它允许用户直接交流。10.5 节简要介绍了一些额外的交互机制，如文本编辑，3D 结构设计和 Web 浏览。

## 10.1 运动程式和重映射

### 动作程序

在生活中，我们会学习动作技能以完成很多特定任务，例如写文字，绑鞋带，扔球和骑自行车。这些通常被称为动作程序，通过反复的试验学习得到，随着练习量的增加，精确性和易用性逐渐提高[196]。最终，我们甚至可以无意识地产生动作。例如，大多数人可以在不注视方向盘，制动器和加速器的情况下驾驶汽车。

同样，我们大多数人都学会了如何使用计算机的接口设备，例如键盘，鼠标和游戏控制器。有些设备更容易学习。例如，学会使用鼠标不需要很长时间，但想利用键盘快速输入可能需要数年才能掌握。是什么原因让一个技能比另一个更难呢？这不容易预测，比如由 Destin Sandlin 利用逆向转向操作设计的逆向大脑自行车，其以相反的转向控制前轮[21]。Sandlin 花了六个月的时间学习如何骑行这种自行车，最后他却无法骑一辆普通的自行车。因此，他花费了大量精力学习新的骑行方式，缺忘记了如何骑一辆普通的自行车。

### 设计注意事项

在虚拟现实交互机制的发展中，主要考虑因素是：

1. 达到所需速度，精度和运动范围的有效性。
2. 学习新的动作程序的难度；理想情况下，不应期望用户花费数月时间掌握新机制。
3. 在认知负荷方面的易用性；换句话说，就是在一些实践之后，交互机制应该很少或根本不需用户集中注意力。
4. 在长时间使用期间需要得到放松；除非是进行一些体育锻炼，否则用户不应该出现肌肉疲劳。

为了设计和评估新的交互机制，首先了解获得动作技能和程序的生理学和心理学。第 5 章和第 6 章介绍了视觉感知，这是将感官输入转化为感知体验的过程。我们现在考虑在物理世界中以身体运动的形式产生输出的相应部分。在这种情况下，大脑将运动信号发送给肌肉，使其移动，同时利用感知过程引入感官反馈。

## 运动的神经生理学

首先考虑参与学习，控制和执行自愿运动的神经组成。如图 10.1 (a) 所示，大脑皮层的某些部分为人体运动服务。运动皮层是控制运动神经信号的主要源，运动前区和辅助运动区也会参与运动的准备和规划。如图 10.1 (b) 所示，还有许多部分参与运动并通过神经信号进行通信。最有趣的部分是小脑，意思是位于头骨后部的“小脑袋”。它是一个特殊的处理单元，主要用于运动，但也涉及注意力和语言等功能。广泛认为小脑损伤会影响精细运动控制和学习新的运动项目。据估计，它包含大约 1010 亿个神经元[7]，远远超过整个大脑皮层，其中包含大约 200 亿。虽然小脑所占体积要小得多，但通过更小，更密集的细胞也能形成一个大的集合。除了协调动作外，它也是动作技能的储存中心。

小脑在 VR 中最相关的一个作用是学习感知模拟器的关系，这些关系被编码成一个动作程序。所有的身体运动都涉及某种感官反馈。最常见的例子是手眼协调；然而，即使闭着眼睛移动手臂，本体也会以动作信号的参考副本的形式提供信息。因而，在动作控制信号与感知信号之间建立紧密的联系对许多任务至关重要。这在工程系统中也广为人知，其中传感器反馈和运动控制结合在机器人和飞机稳定控制中有所应用；与此相关的主题被称为控制系统。众所周知，闭环系统是优选的，其中传感器信息在执行期间提供反馈，与开环相反，开环指定运动信号作为时间的函数。

学习某个动作模式所需要的时间是最重要的因素之一。在这一点上，不同人的差异很大。一个关键的概念是神经可塑性，这是大脑重组神经结构并形成适应新刺激途径的能力。幼儿具有高水平的神经可塑性，随着时间的推移，突触修剪过程会大大降低神经可塑性。这导致健康的成年人每个神经元的突触数量比两三岁的孩子多一半[99]。结果便是成年人获得新技能的难度更大，例如学习新语言或学习使用复杂的界面。除了随着年龄减少神经可塑性之外，它在同龄人中的差异也很大。

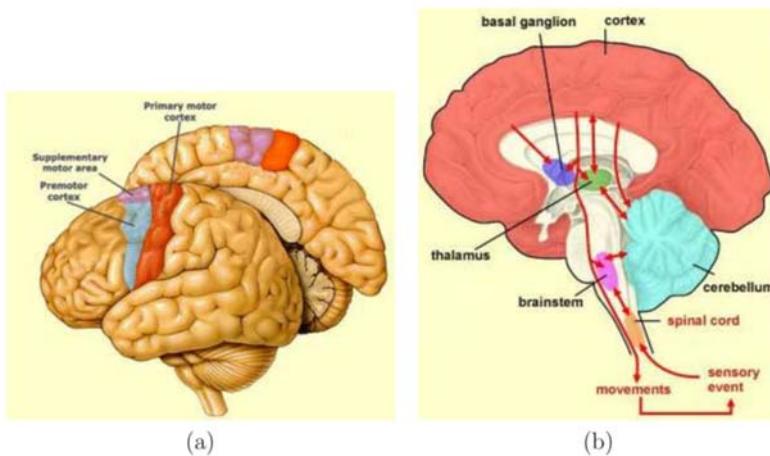


图 10.1：(a) 部分大脑皮层致力于运动。 (b) 许多其他部分与皮质相互作用以产生和执行运动，包括丘脑，脊髓，基底神经节，脑干和小脑。 (由 McGill University 的 The Brain from Top to Bottom 提供的数据)

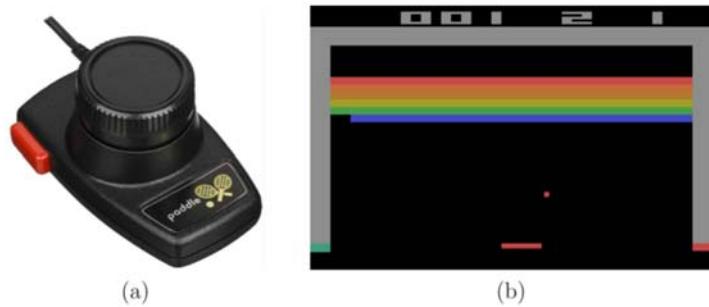


图 10.2: (a) Atari 2600 桨控制器。 (b) Atari Breakout 游戏，其底线部分是一个虚拟桨，允许球弹跳到顶部并消除接触后的块。

### 学习动作模式

现在考虑学习一个计算机界面呈现的动作程序。一个简单而经典的例子是 1976 年由 Atari 开发的视频游戏 Breakout。玩家转动一个旋钮，如图 10.2 所示。这会导致屏幕底部的线段水平移动。相应的设备包含一个电位计，通过校准可以准确地估计旋钮的方向。玩家观察位于屏幕底部的线段并快速调整旋钮方向。因此，学习过程涉及从视觉感知中获取信息，并通过转动旋钮并确定感知运动关系来获取本体信号。熟练的玩家便可以快速转动旋钮，以便更快地移动线段，而不用在现实世界中来回移动一个小托盘。因此，我们已经有了一个例子，其在虚拟世界的版本比现实中的表现更好。

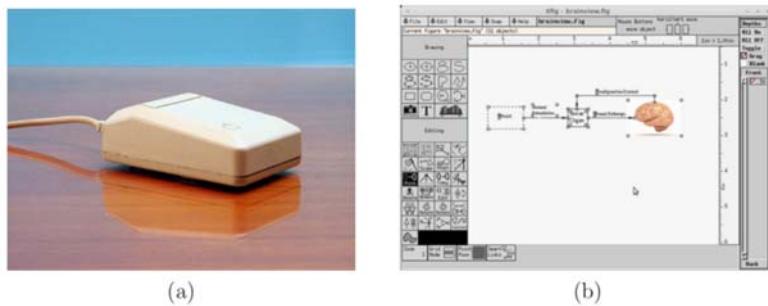


图 10.3: (a) Apple Macintosh 鼠标。 (b) 当鼠标在桌面上移动时，屏幕上的虚拟手指会相应地移动，但它旋转了 90 度并需要移动较长的距离。

在 Breakout 示例中，旋钮方向和线段位置之间有一个一维映射学习过程。实际上，这里还有许多替代控制方案；但是，它们并不能让用户感到满意。如果你发现一个模拟器存在可以完善和突破的地方，那它很可能涉及到通过按键达到移动效果的问题。在这种情况下，按下按键的时间量对应于位移量。相应物体的速度则由程序设定。使用现代硬件的合理替代方案需要手指在触摸屏上来回移动，而相应的控制物体直接出现在触摸屏上。由于额外的自由度，手指不会受到太大的约束，而手指的快速来回运动可能会导致肌肉疲劳，尤其是在屏幕较大的情况下。在手划完成操作的情况下，运动由手指完成，手指灵活性高，而前臂移动少得多。该映射提供了身体运动和虚拟物体放置之间的关联，实现了高精度，快速放置和长期舒适性的交互。

图 10.3 显示了一个比较熟悉的例子，它是电脑鼠标。当鼠标在桌子上被推动时，编码器确定位置，该位置被转换为屏幕上的指针位置。该运动感知的映射看起来比 Breakout 例子要复杂一些。儿童往往能立即学会如何使用鼠标，而老年人则需要一些练习。鼠标的 2D 位置被映射到屏幕上的 2D 位置，具有两个基本的失真：1.与桌子相比，屏幕旋转了 90 度

(水平到垂直运动)。2.运动被缩放，以小的物理运动产生了更大的屏幕运动。最初的施乐奥拓鼠标的优点在[39]中从人类技能学习和费茨定律[78,192]两方面进行了科学讨论，这从数学上将任务的难度指向了达到目标所需的时间。

最后假设通过按一个键，字母“h”立即以熟悉的字体出现在屏幕上。我们的视觉感知系统将“h”视为与纸面版本等效。因此，输入关键字会导致“h”的感知。这是一个相当舒适，快速和强大的操作。

## VR 的动作模式

迄今为止所给出的例子已经与 VR 很相似。感知体验由通过硬件设备感知的身体移动来控制。使用通用模拟原理，可以将这些以及更多的交互机制引入到 VR 系统中。物理交互部分大体上是相同的，或者可以通过另一种控制器进行模拟。

使用第 9 章的跟踪方法，可以可靠地估计身体各部位的位置和方向并将其引入 VR。对于头部追踪的情况，保持视点的准确性和零有效延迟至关重要；否则，VR 体验质量会显著降低。这是至关重要的，因为必须保持稳定的感觉，才能维持可信度和舒适度。感官的运动必须与跟踪系统相匹配。

## 重映射

对于其他身体部位的运动，这种完美匹配并不重要。我们的神经系统可以在舒适性方面产生更好的关联，就像鼠标和键盘在现实世界中的运行方式一样。因此，我们想要进行重映射，这涉及到学习一种感知运动映射，这种映射在虚拟世界中产生不同于现实世界所期望的结果。上面的键盘示例是重映射最常见的示例之一。按下一个按键取代了在纸上划动铅笔以产生一个字母的过程。

VR 的重映射是很自然的。例如，可以不用伸手抓住一个虚拟的门把手，而是按下一个按钮来打开门。对于更简单的情况，请考虑按照 HTC Vive 系统允许的方式，通过空间控制器追踪用户姿态。设置缩放参数可以使得真实世界中一厘米的手位移对应于虚拟世界中的两厘米位移。这与鼠标的缩放参数相似。第 10.2 节介绍了通过控制器实现从现实世界中的自然行走到虚拟世界中等价状态的重映射。10.3 节介绍了对象交互方法，它们通过重映射再次实现。可以预计在未来几年还会看到许多 VR 重映射方法。

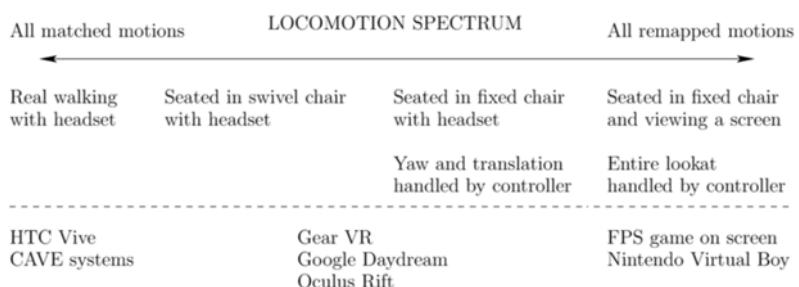


图 10.4：从左向右移动时，实际运动和虚拟运动之间的视点失配量增加。

## 10.2 运动

假设虚拟世界覆盖的面积远大于被追踪的真实世界的面积。换句话说，匹配区域相对于虚拟世界而言较小。在这种情况下，需要某种形式的互动机制来将用户移到虚拟世界中，同时用户仍然保持在现实世界的跟踪区域内。以这种方式移动用户的交互机制称为运动控制 (Locomotion)。就好像用户正在虚拟世界中驾驶虚拟车辆一样。

图 10.4 显示了一系列常见的运动场景。在左侧，用户在戴着 HMD 的同时四处走动。此情况下，除非虚拟世界大于开放空间，否则不需要运动控制。这种情况不涉及实际和虚拟运动之间的不匹配。

两种中间情况对应于佩戴 HMD 的坐在座位上的用户。在这些情况下，就需要使用交互机制来改变虚拟世界中匹配区域的位置。如果用户坐在旋转椅上，那么他可以通过旋转椅子来改变朝向（偏航方向）。这可以被认为是在虚拟世界中定位用户的躯干。如果用户坐在固定的椅子上，则通常使用控制器来改变虚拟躯干方向，这会导致更多的不匹配。极限情况显示在图 10.4 的右侧，其中甚至没有头部跟踪。如果用户正面对屏幕，如在屏幕上的第一人称射击游戏的情况下，则使用游戏控制器来改变用户在虚拟世界中的位置和方向。这是不匹配程度最大的情况，因为视点中的所有更改都是由控制器生成的。

### 重定向行走

如果用户在非常大的空间（例如每边至少 30 米的正方形区域）进行运动并被跟踪，重定向行走则有可能使她认为自己正在以直线行走数公里，而她实际上正在绕圈行走。这种技术被称为重定向行走[261]。在没有视觉线索的情况下长距离沿直线行走对于人类（和机器人）实际上是不可能的，因为在现实世界中不可能实现完美的对称。

想象一下虚拟城市包含长而直的街道的虚拟现实体验。当用户在街上走时，视点的偏航方向可以逐渐变化。这代表了真实和虚拟世界之间的一小部分不匹配，并且导致用户沿着圆弧行走。但这种技术的主要问题是用户有自由意志，并且可能使得用户走向现实世界中匹配区域的边缘，即使用户不能直接感受到这一点。在这种情况下，可能会出现警告，建议用户必须旋转以重置偏航方向。

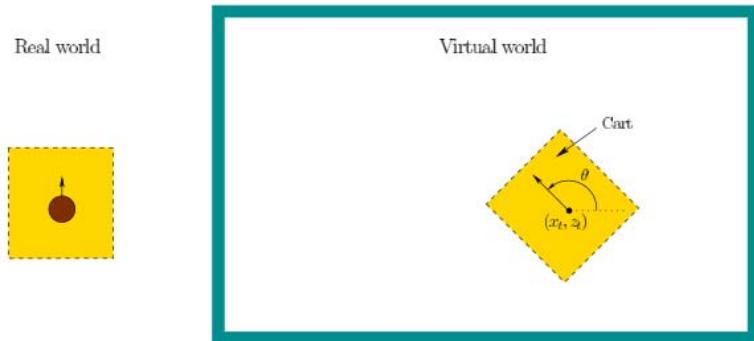


图 10.5：沿着水平地形的运动可以被模拟为在虚拟世界中的车辆。这里显示自上而下的视图。黄色区域是匹配的区域（回忆图 2.15），其跟踪用户视点。通过使用控制器来改变  $x_t$ ,  $z_t$  和  $\theta$  的值。

### 运动执行

现在考虑图 10.4 中坐着并戴着 HMD 的中间情况。运动可以通过控制器移动视点简单地实现。将匹配区域视为在虚拟环境中移动的可控车体是有帮助的；见图 10.5。首先考虑地面是水平面的简单情况。令  $T_{track}$  表示在物理世界中眼睛跟踪位置和方向的齐次变化矩阵。

9.3 节中描述的方法可用于为当前时间提供  $T_{track}$ 。

车体的位置和方向由控制器决定。齐次矩阵：

$$T_{cart} = \begin{bmatrix} \cos \theta & 0 & \sin \theta & x_t \\ 0 & 1 & 0 & 0 \\ -\sin \theta & 0 & \cos \theta & z_t \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad (10.1)$$

对位置 ( $xt$ ,  $zt$ ) 和方向  $\theta$  进行编码 (作为偏航旋转角, 从 (3.18) 中借用)。 (10.1) 中的高度设置为  $yt=0$ , 这样它就不会改变跟踪系统或其他系统所确定的高度 (回顾 9.2 节, 如果用户坐在真实世界中, 但是在虚拟世界中是站着的, 高度可以是人为设置的)。通过连接矩阵  $T_{track}$  和  $T_{cart}$  获得眼部变化状态:

$$T_{eye} = (T_{track} T_{cart})^{-1} = T_{cart}^{-1} T_{track}^{-1} \quad (10.2)$$

回顾第 3.4 节, 眼部变换是放置几何模型所需变换的逆变换。因此, (10.2) 对应于由于车体而产生的视角变化, 其次是跟踪的头部视角。

为了将视点移动到固定方向  $\theta$ ,  $xt$  和  $zt$  分量通过以下方程获得:

$$\begin{aligned}\dot{x}_t &= s \cos \theta \\ \dot{z}_t &= s \sin \theta.\end{aligned}\quad (10.3)$$

在时间步长  $\Delta t$  上积分 (10.3), 位置更新为

$$\begin{aligned}x_t[k+1] &= x_t[k] + \dot{x}_t \Delta t \\ z_t[k+1] &= z_t[k] + \dot{z}_t \Delta t.\end{aligned}\quad (10.4)$$

式子 (10.3) 中的变量  $s$  是前进速度。人类的平均步行速度约为每秒 1.4 米。通过按下设置  $s = 1.4$  的按钮或按键, 可以向前移动虚拟车体。另一个按钮可用于分配  $s = -1.4$ , 这会导致向后移动。如果没有按键或按钮被按下, 则  $s=0$ , 使得车体保持静止。另一种控制方案是使用两个按钮来增加或减少速度, 直到达到某个最大限制。在这种情况下, 无需按住某一个键就可以保持运动。

除了前进/后退运动之外, 按键还可用于提供侧向运动。这在视频游戏中被称为扫射。如果可能的话应该避免这种运动, 因为它会导致不必要的侧向牵引。

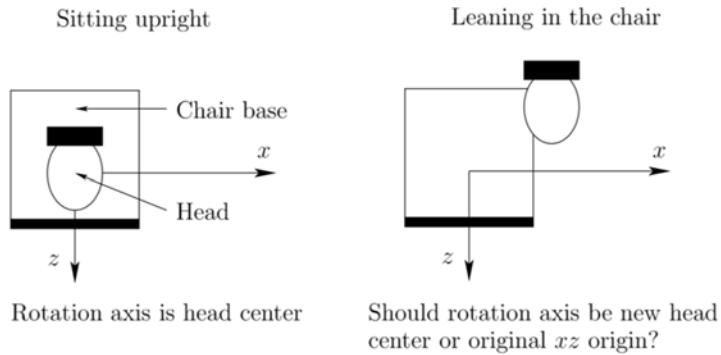


图 10.6: 对于右侧直坐在椅子上的用户, 偏航旋转轴线位于头部的中心。对于左侧, 用户靠在椅子上。这时, 旋转轴应该保持固定, 还是与用户一起移动?

### 改变方向的问题

现在考虑方向  $\theta$ 。为了向不同的方向移动,  $\theta$  需要再分配。可以根据用户的头部偏航方向进行分配。当用户坐在转椅上向前看时, 这种操作比较方便且令人舒适。通过转动转椅, 可以设定方向。(但是, 这可能会引起有线耳机的问题, 因为耳机线可能会缠绕用户。)

在固定的椅子上, 控制  $\theta$  可能没那么简单, 因为舒适的头部偏航范围仅限于每个方向 60

度（回忆图 5.21）。在这种情况下，可以使用按钮在顺时针或逆时针方向上以小增量改变  $\theta$ 。不幸的是，根据恒定的角速度改变  $\theta$  会导致偏航，这对很多人来说是不舒服的。一些用户喜欢点击一个按钮，每次立即偏航约 10 度。如果增量太小，则再次取向；如果增量太大，则用户会对其取向感到困惑。

另一个问题是如何定位旋转中心，如图 10.6 所示。当用户在真实世界中从椅子中心移开时会发生什么？旋转中心的位置应该是原来的头部中心还是新的头部中心？如果选择它作为原始中心，则当  $\theta$  改变时，用户将感知到大的平移。如果产生新的中心，那么变化量就会减少，但可能与现实不相符合。

对于另一个变化情况，可以使用 8.3.2 节中的车状运动模型（8.30），以便转换视点才能旋转。换句话说，虚拟形象会有最小的转弯半径。更一般来说，可以通过控制任何虚拟车辆模型来改变视点。第 1 章的图 1.1 显示了“车辆”是一只鸟的例子。

### 减少相对运动错觉的策略

运动的主要问题是产生相对运动错觉并导致 VR 晕眩。在第 8.4 节，我们讲述了六种不同的运动，每个自由度都对应着一种运动。此外，影响敏感度的因素很多。减少这些因素的权重应该会减少对身体的负面影响，并希望减轻 VR 晕眩。

减少基于相对运动错觉所产生的晕眩的几种策略是：

1. 如果光流的视场减小，则视野减弱。一个常见的例子是制造阻挡大部分光流的驾驶舱或汽车内部。
2. 如果视点太靠近地面，则运动特征的速度和加速度矢量的幅度更高。这就是为什么你可能会觉得，如果你乘坐的是一辆小型汽车，比起在卡车或小型货车上以相同的速度骑行时，要快得多。
3. 令人惊讶的是，短时间内较大的不匹配可能优于长时间较小的不匹配；见图 10.7。
4. 具有高空间频率将为人类视觉系统追踪提供更多特征。因此，如果通过的环境更加平滑，细节更少，则相对运动错觉将减少。考虑走楼梯的情况。如果这些步骤清晰可见，以至于它们表现为移动的水平条纹，那么用户可能会因强烈的垂直信号而迅速感到不舒服。
5. 减少对比度，例如在加速时让世界看起来朦胧或模糊，可能会有所帮助。
6. 提供其他感官线索，如吹风或移动音频源可能会提供更强的运动证据。包括隆隆或振动形式的前庭刺激也可能有助于降低前庭信号的信心。即使使用头部倾斜来诱发虚拟世界运动的变化也可能有所帮助，因为它会引起前庭信号分散注意力。
7. 如果世界应该移动，而不是用户，那么通过提示或特别指示来清楚说明可以提供帮助。
8. 提供具体的任务，例如对飞虫发射激光，可能会对前庭冲突提供足够的注意力。如果用户完全专注于动作，那么她可能会更快地晕眩。
9. 反复练习可能会降低运动的负面影响。经常在大屏幕前玩 FPS 游戏的人似乎已经降低了对 VR 中的感光度的敏感度。要求用户在晕眩减少之前进行练习对希望推出新产品的公司来说可能不是明智的策略。想象一下，尝试一些新的食物，让你在吃了前 20 次后感到恶心，但随后逐渐变得更加可以接受。谁会继续尝试呢？

最后的建议是尽可能避免运动！尝试设计不严重依赖于它的体验。

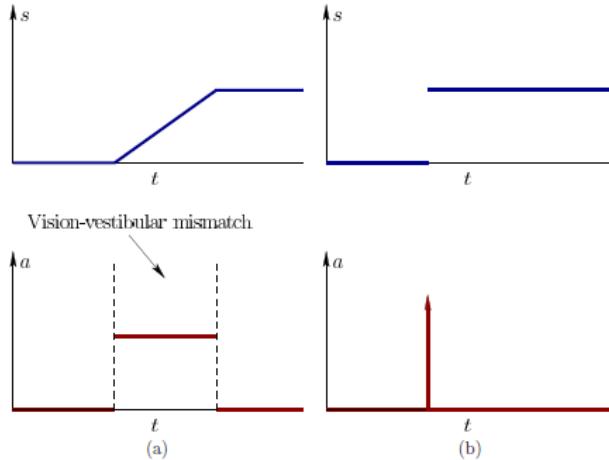


图 10.7: (a) 在一个时间间隔内应用恒定的加速度, 以使停止的化身达到速度限制。上图显示了随着时间的推移速度。下图显示加速度。非零加速的时间间隔对应于与前庭感觉的不匹配。(b) 在这种情况下, 施加加速脉冲, 导致立即达到所需的速度限制。在这种情况下, 不匹配发生在有效长度为零的时间间隔内。实际上, 感知的速度在一对连续的帧中变化。令人惊讶的是, 情况(b)比(a)更舒适。看起来大脑更喜欢在很短的时间间隔内出现异常值, 因为在较长的时间间隔内(如 5 秒)应该会有较小的持续不匹配。

### 非平面运动

现在考虑更复杂的运动案例。如果用户在地形上行走, 则可以简单地增加或减少 y 分量以反映高度的变化。这似乎是现实的, 但请记住, 它会增加真实世界与虚拟世界之间的不匹配数量, 因为垂直切入与正向切入相结合。

在移动通过 3D 媒体的情况下, 8.4 节的所有六种形式都可以使用。常见设置包括虚拟航天器, 飞机或潜水员, 他们可以轻松生成偏航, 俯仰和滚动视野。例如, 想象一下飞行虚拟航天器, 通过以滚动的方式运动, 当周围的星星以环形模式旋转时, 可能会引起滚动的相对运动错觉。如果开发人员必须以这种方式制作飞船, 那么应该遵循先前关于减少切割强度的建议。此外, 应该对人体进行仔细的实验, 以确定哪种形式的视力在特定的应用中更差, 详见第 12 章。为了避免奇点, 对于所有 3 个旋转运动自由度都是可能的系统, 虚拟车辆变换最好用四元数来维护(回顾在 3.3 节的问题)。

添加移动视点的特殊效果将导致更多的难度。例如, 使角色上下跳动将导致垂直视角。在走路时摇摆的头部动作也是一个影响视觉的情形, 因为它导致不匹配度的增加,。设想一个更糟糕的情况, 如果你是一个体操的虚拟角色, 在多次翻转过程中, 世界的观点可能变得无法忍受。

### 专门的硬件

现在已经开发了多种硬件来支持运动。最古老的例子之一是为飞机飞行模拟创建一个完整的驾驶舱(图 1.15)。图 10.8 (a) 显示了一个全方位跑步机, 可以在任何方向和任何距离上行走。诸如固定式自行车之类的练习机已经连接到 VR 系统, 以便用户可以踩踏和操纵引导自己穿越大型虚拟世界, 如图 10.8 (b) 所示。图 1.1 显示了一个像鸟一样虚拟飞行的机械平台。



图 10.8: (a) 美国陆军用于 CAVE 系统的全方位跑步机, 用于训练。(b) 由 Paul Dyan 开发的与 VR 设备相连的家庭自行车自行车系统。

### 瞬间移动

迄今为止涉及的运动方法主要集中于再现现实世界中熟悉的体验, 这些体验提供了通用模拟原理的实例。然而, 在虚拟现实中, 我们也可能采取物理上不合理的方式。最常见的是远距传物, 就像电视连续剧“星际迷航”中的运输车一样。用户立即被运送到另一个地点。

如何确定所需的位置? 一种简单的机制是虚拟激光指示器 (或 3D 鼠标), 通过用户持有与现实世界中的激光指示器形状相似的控制器来完成, 甚至智能手机都可以使用。用户旋转控制器来移动虚拟世界中的激光点。这需要执行光线投射操作 (第 7.1 节) 沿着对应于激光的光线寻找最近的可见三角形。

要选择用户喜欢站立的位置, 她可以简单地指向虚拟激光并按下一个键即刻传送。为了更容易指向地板, 实际上可以是一个跟随重力的抛物线弧, 类似于一股水流, 见图 10.9。不可见的地方可以使用弹出式地图进行选择, 甚至可以执行基于文本的搜索 (可以使用语音命令代替打字)。一种叫做微缩世界的方法可以向用户展示一个虚拟的小型环境版本[312], 这实际上是一张 3D 地图。



图 10.9: 一个虚拟的“激光指示器”, 它遵循抛物线弧线, 因此可以很容易地将传送目的地指定为地板上的一个点。 (图片来自 HTC Vive 平台上的 Budget Cuts 游戏。)

### 路径寻找

学习空间表示并将其用于导航的认知问题称为路径寻找。这是一个比低层运动机制更高层次的过程, 但这两者密切相关。运动系统在现实世界中并不熟悉的一个问题是, 用户可能无法了解周围世界的空间布局。如果你能够在不同的地方传送, 你的大脑是否仍然会形成真实世界环境的地方记忆细胞? 我们广泛观察到这种现象, 发现人们大多仅学习使用 GPS 或出租车服务导航城市道路, 而不是自己寻找路线。

远距传送机制减少了运动, 因此减少了 VR 晕眩, 然而, 这可能是以减少对环境空间布

局的学习为代价的。进行远距传物时，不要改变视点的偏航方向；否则，用户可能变得更加迷失方向。他可能不了解他现在在虚拟世界中相对于以前的位置定位和定位的位置。

请注意，通用仿真原理可以再次用于从现实世界借用任何有效的导航工具。如果虚拟建筑和城市以现实世界中常见的布置，那么他们应该更容易导航。标志和地标甚至可以放置在虚拟世界中以帮助导航。在现实世界中，标志通常会告诉我们出口的位置，街道的名称或区域的边界。诸如高楼，风车或塔楼等地标提供了可以长距离导航的视觉线索。许多这些想法在[31]的第7章中讨论。

### 10.3 操作

我们与真实世界中的物体进行交互的原因很多。你可以通过在碗和嘴之间移动勺子来吃一碗汤，你可能会拿起一块石头并尽可能地扔掉，你可以穿上一条裤子……这些例子属于操纵的主题。在现实世界中，操纵涉及复杂的感觉运动关系，通过进化和经验，我们可以在各种设置下操纵对象。物体的变化包括尺寸，重量，摩擦，柔韧性，温度，易碎性等方面差异，但不知何故，我们的身体都可以处理。让机器人以人类的方式进行操纵一直是一条漫长而令人沮丧的道路，只有有限的成功[203]。

由于现实世界中的操作复杂性，它是应用10.1节中的重映射概念在VR中尽可能简化操作的理想人选。虚拟世界不必遵循复杂的操纵物理学。相反，优选的是进行诸如选择，抓取，操纵，携带和放置物体的操作尽可能快和简单。此外，应避免广泛到达或其他形式的肌肉紧张，除非VR体验旨在提供锻炼。

#### 避免“大猩猩手臂”

在公众中最常见的误解之一是汤姆克鲁斯在电影“少数派报告”中使用的接口是可取的；见图10.10。事实上，它很快导致了“大猩猩手臂”这个众所周知的问题，使用者很快感觉到手臂长时间疲劳。你可以在自己面前直接握住手臂而不会疲劳吗？



图10.10：汤姆克鲁斯在2002年电影“少数派报告”中在全息显示中移动窗口。这对好莱坞来说是一个很好看的互动机制，但在现实中令人非常疲劳。用户会很快体验大猩猩的武器。

#### 选择

在虚拟世界中选择一个物体的最简单方法之一就是使用虚拟激光指示器，该指示器在10.2节中有描述。多种变化可能有助于改进选择过程。例如，用户可能会拿着照亮潜在选择的虚拟手电筒。手电筒的视野可以调节[83]。可以放置虚拟镜像，以便可以在角落进行选择。[31]的第5章提供了许多其他建议。

通过使用指针，用户只需照亮感兴趣的对象并按下一个按钮。如果目标是检索对象，那么它可以立即放置在用户的虚拟手中或库存中。如果目标是以标准的，重复的方式操作对象，

那么按下按钮可能会导致执行虚拟电机程序。例如，这可以用来转动门把手，从而打开一扇门。在这样的使用中，开发者可能希望设置激光指示器深度的限制，以便用户必须站得足够近才能进行交互。例如，把门把手从房间转过来似乎是不合适的！

如果物体很难看清楚，那么选择过程可能会很复杂。它可能在用户头后，这可能需要不舒服的转弯。该物体可能很小或很远，以至于它只能在屏幕上占据几个像素，因此难以精确选择它。如果感兴趣的物体周围存在大量杂物，特别是如果其他可选物体在附近，问题会变得更加严重。最后，该对象可能部分或完全被遮挡。

## 操纵

如果用户长时间携带物体，则不需要挤压或离合控制器，这会产生不必要的疲劳。在某些情况下，用户可能需要仔细检查物品，同时拥有它。例如，他可能想要将其移动到手中以确定其 3D 结构。可以将对象方向设置为完全遵循用户所持有的控制器的 3D 方向。用户甚至可以握住由外部相机跟踪的真实物体，但在虚拟世界中具有不同的外观。这使得熟悉的力量反馈给用户，这个概念在 13.1 节中重新讨论。请注意，一个对象甚至可以直接在虚拟世界的原始位置进行操作，而不会将其靠近用户的虚拟身体[30]。在这种情况下，虚拟手被带到物体上，而物理手保持在原位。还可以模拟比正常情况更长的手臂[254]，以便更长距离地检索和放置物体。

## 放置

现在考虑解开物体并将其放入世界。对于用户来说，简单的情况是按下一个按钮，让对象简单地落在正确的位置。这是通过吸引域来实现的，该吸引域是在目标姿态（位置和方向）的邻域中定义的吸引力潜在函数，见图 10.11。潜在函数的最小值在目标上。物体释放后，物体通过移动落入目标姿态，以使电位降至最小。在许多 2D 绘图程序中可以看到这种行为，以便线段的端点方便地相遇。一个方便的对象放置的例子是马库斯佩尔松（Notch）在 2011 年的 Minecraft 沙盒游戏中，其中的积木很容易落到位。孩子们用这种方式建立了数以百万计的虚拟世界。

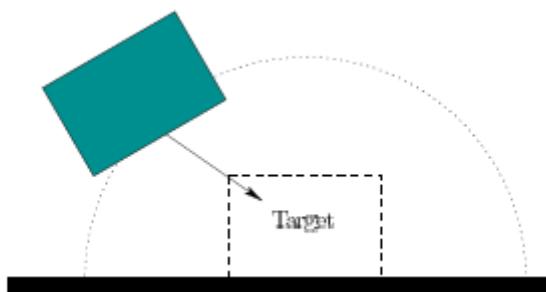


图 10.11：为了让使用者的生活更轻松，可以在一个物体周围定义一个吸引域，以便当物体进入时，落下的物体被直接吸引到目标位置。

或者，用户可能需要微妙地放置物体。也许应用程序涉及尽可能高的堆叠和平衡对象。在这种情况下，精度要求会非常高，给控制器跟踪系统和用户带来负担。

## 重映射

现在考虑重映射的功能，如 10.1 节所述。最简单的情况是使用按钮来选择，抓取和放置对象。除了按钮之外，用户可以生成连续的运动并由系统进行跟踪。示例包括转动旋钮，移动滑块，在手指上移动手指以及通过空间移动自由浮动身体。回想一下，重映射最重要的

方面之一是易学易用性。减少重新映射的自由度数通常会简化学习过程。为了避免“大猩猩的手臂”和相关问题，可以在跟踪设备上施加一个比例因子，以便控制器中的少量位置变化对应于虚拟世界中的大运动。这个问题可以再次使用 Fitts 定律来研究，就像电脑鼠标一样。请注意，这可能会对虚拟世界的精度产生不利影响。在某些设置中，定向缩放也可能是可取的。在这种情况下，三维角速度 ( $\omega_x$ ,  $\omega_y$ ,  $\omega_z$ ) 可以按比例缩放，以在虚拟世界诱发更多的旋转，而不是真实世界。

### 目前系统设计

操纵互动机制的发展仍然是 VR 面临的最大挑战之一。当前一代消费类虚拟现实头戴式设备或者利用现有的游戏控制器，比如 2016 年将 Xbox 360 控制器与 Oculus Rift 捆绑在一起，或者引入假设大型手部动作的系统，如 HTC Vive 控制器，如图 10.12。使用户在空间中移动手的控制器与图 10.10 中显示的少数报告交互机制似乎并不相同。其他人正在开发不涉及硬件的手势系统，就像 9.4 节中图 9.25 所示的 Leap Motion 系统一样。这些可能是 20 世纪 90 年代流行的“护目镜和手套”愿景的更新版本（回顾 1.3 节图 1.29 (c)）。在未来几年，操作方法和技术的快速发展可以预期，并且越来越重视用户的舒适度和易用性。



图 10.12：(a) 2016 年随 HTC Vive 一起推出的一对手持控制器；该设备包括侧按钮、触发器和用于拇指的触摸板。(b) 尝试控制器的用户（原型版本）。

## 10.4 社交互动

沟通和社交互动是很大的主题，远远超出了本书的范围。此外，虚拟现实或社交虚拟现实中的社交互动仍处于起步阶段，并在进行实质性的实验和对范式的反思。尽管如此，将人类连接在一起是 VR 技术最大的潜力之一。尽管在我们周围展示我们自己与世界之间的显示可能是孤立的，但我们也可以通过成功的交互机制将它们拉近在一起。本节重点介绍了几个有关社交互动的有趣问题，而不是提供完整的评论。

### 超越 Shannon-Weaver 的交流

一个重要的因素是有多少人会通过媒体进行互动。从两个人开始说起，有史以来最强大的数学模型之一是 Shannon-Weaver 通信模型，数十年来一直是工程通信系统设计的基础；见图 10.13。该模型涉及发件人和收件人。通信系统对来自发送者的消息进行编码，然后通过噪声信道发送。在另一端，系统解码该消息并且它到达收件人。收件人可以提供反馈，以指示邮件是否已完整收到。这种通信模式引发了信息理论领域，它为通信信道带来了明确的带宽概念，并揭示了数据压缩的局限性。

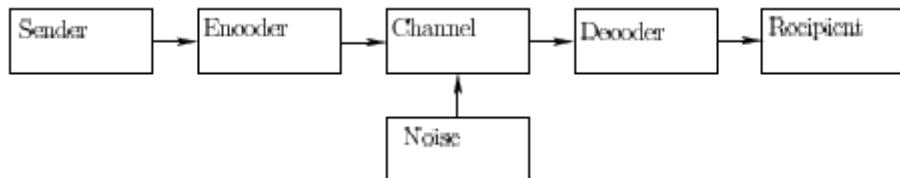


图 10.13：经典的 Shannon-Weaver 传播模型（从 1948 年开始）。发送者向编码器提供消息，编码器通过受噪声干扰的信道发送消息。另一方面，解码器将该消息转换成适合接收机的格式。这个模型作为工程领域的交流理论的基础。

这个模型的强大之处在于它能够在数学上量化人类的相互作用，但它也不足以涵盖 VR 中可能存在的各种相互作用。通过再次遵循通用模拟原理，现实世界中存在的任何类型的人际交互都可以引入 VR。Shannon-Weaver 模型受到诸如 19 世纪电报或 20 世纪手持式接收器（或对讲机）之类的交互机制的启发。在这些情况下，人类彼此完全隔离，并且该技术提供了类似于写信的突发信息。我们已经从文本到音频再到视频通信，并且可以通过将触摸和嗅觉等其他感官的显示结合起来，进一步延伸；同时，也有很多机会使用合成模型，可能与摄像头和麦克风捕获的实际信息相结合。简单的手势和行为举措可以提供交互的微妙但重要的组成部分，而这些组成部分并未被经典通信模式所捕捉。

尽管存在虚拟现实的缺点，但请记住，Shannon-Weaver 模型能够为计算机网络和系统提供强大的带宽和延迟分析，最终支持任何形式的社交互动。因此，它对 VR 系统中可以或不可以实现的内容具有深远的影响。发生这种情况是因为所有“通信”都会转换为通过电缆或网络连接发送的位流。一个关键的问题是确保有针对性的社交互动 VR 体验在计算机网络上舒适，可靠地得到支持。

### 从头像到视觉捕捉

其他人如何在 VR 中看到你？这是最有趣的问题之一，因为它取决于社会背景和技术限制。这存在一个明确的可能性范围。在某种极端情况下，用户可以通过虚拟形象代表自己，虚拟形象是一种 3D 表现形式，可能完全不符合用户可看到、可听到的形象以及行为特征；如见图 10.14。在另一个极端情况下，用户可能会使用成像技术进行拍摄，并在虚拟世界中以高度精确的 3D 表示进行复制；见图 10.15。在这种情况下，虚拟世界的人物形象像是直接从真实世界传送过来一样。这样的想法还存在许多其他可能性，值得考虑和权衡。



图 10.14：Second Life 提供的初始人物造型集合。

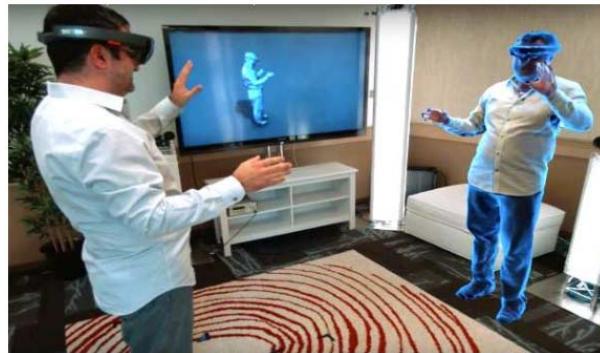


图 10.15：2016 年来自微软的全息通信研究。通过增强现实眼镜（Hololens）可以看到，人的 3D 表示被实时提取并叠加在世界中。



图 10.16：2009 年的数字艾米莉项目：(a) 真人成像。 (b) 几何模型与复杂的渲染技术融合以制作动画，产生逼真的面部运动。

VR 中的虚拟形象的一个主要吸引点在于匿名性，它提供了在社交环境中扮演不同角色或展现不同个性特征的机会。在一种称为变形人效应的现象中，人们已经观察到，一个人的行为会根据虚拟人物的虚拟特征而改变，这与人们在穿着不同制服或服装时会表现出不同的形象类似[360]。用户可能想要生活在幻想中，或尝试从不同的角度看待世界。例如，如果人们能够在一个从种族，性别，身高，体重，年龄等方面都不同的世界进行体验，便可能会产生同理心。

用户也可能想要尝试其他形式的体验。例如，一群孩子可能想要居住在动物的尸体内，并进行正常的生活。畅想一下，如果人们把你看作是外星人、昆虫、汽车，甚至是奶酪。1986 年，当皮克斯在动画短片 Luxo Jr 中拥有台式电脑的生活时，人们感到非常惊讶。好莱坞电影在过去的几十年里充满了动画角色，我们有机会体验其中的一些人，同时居住在虚拟世界！现在考虑这种方式的物理可行性。基于目前的技术，可以独立考虑三种主要的相似性：

1. **视觉外观：**在可见特征方面，虚拟形象与真实人物看起来有多接近？
2. **听觉外观：**来自虚拟形象的声音与人的声音，语言和言语模式相匹配多少？
3. **行为表象：**虚拟形象的动作与身体语言，步态，面部表情和其他动作的匹配程度如何？

第一种相似性可以通过在虚拟世界中创建一个运动学模型（回忆第 9.4 节）来进行与人的匹配，该运动学模型的大小和移动性与实际的人相对应。此外也可以执行其他比较简单的匹配，例如头发颜色，肤色和眼睛颜色。为了进一步提高真实感，可以使用纹理映射将皮肤和衣服映射到虚拟形象上。例如，用户脸部的图像可以被纹理映射到虚线形象脸部。也可以通过构建合成模型来进行高精度匹配。迄今为止进行的一些最佳合成匹配是由 USC 创新技术研究所的研究人员完成的；见图 10.16。如第 1.1 节所述，一个令人沮丧的问题是恐怖谷理论。人们常常会描述计算机生成的动画，比如看僵尸或谈论尸体等，这些动画往往趋近于人类自身的现实。因此，不完美的匹配通常比视觉外观上的“几乎相同”的匹配要好得多。对于听觉部分，Second Life 和类似系统的用户喜欢发短信。这种互动在 VR 中就好像是他们正在讨论。发短信有助于确保匿名。在 VR 中记录和再现语音很简单，使听觉外观与视觉外观相匹配更加简单。我们必须注意以适当的区域化来呈现音频，以使其他人看上去是从虚拟

人物的嘴巴里说出来的；具体细节请参阅第 11 章。目前，已可以通过使用实时语音更改软件（如 MorphVOX 或 Voxal Voice Changer）进行录音，轻松保留匿名消息，这可能比在某些设置中发短信的方式更受欢迎。

最后要注意的是，人的行为体验可以在 VR 中完美匹配，而虚拟人物会具有完全不同的视觉外观。这是动作捕捉系统产生的主要原因，其记录真实演员的动作，然后用动画制作虚拟人物。请注意，电影制作通常是一个漫长的离线过程。在低成本的虚拟现实系统中，目前还无法实现与人的视觉和行为外观实时准确匹配的功能。此外，尽管最近在研究在这方面取得了一些进展[180]，但如果人脸的一部分被头戴式耳机覆盖，其捕获将会很困难。

不过，当前的跟踪系统可以在某些情况下提供精确匹配的行为外观。例如，头部追踪可以直接与头像相关联，以便其他人可以知道其头部转向的位置。用户还可以理解头部运动或手势的意义，如“是”或“否”。图 10.17 显示了一个简单的 VR 体验，朋友可以一边看电影，一边由被跟踪的虚拟头像（他们也可以相互交谈）代替。在一些系统中，可以使用眼动追踪，以便用户可以看到虚拟人物正在看的位置；然而，在某些情况下，这可能会再次陷入恐怖谷理论。如果跟踪手部运动，可以使用如图 10.12 所示的控制器完成，这种情况仍然可以使用户沉浸于虚拟世界。



图 10.17：Oculus Social Alpha，这是三星 Gear VR 的应用程序。多个用户可以在虚拟世界中见面并进行社交。在这种情况下，他们一起在剧院看电影。他们的头部动作是使用头部追踪数据提供的。他们也可以用本地化的音频互相交谈。

### 从一对一到整个社会

现在考虑不同规模的社交互动。我们在现实世界中进行的绝大多数一对一互动是与我们认识的人进行的。同样，通过科技进行交互时也是如此，无论是通过短信，电话还是视频聊天。我们通过科技进行的大多数互动都是针对性的，因为参与有特定的目的。这表明，虚拟现实技术可以用于将视频聊天提升到一个新的水平，在这个阶段，两个人会感觉他们在虚拟世界中是面对面的。但要注意的是，在现实世界中，我们可能会有一些无意的互动，而不是只进行有针对性的社交。

一对一交流的一个重要方面在于两个人之间的关系是对称的还是互补的（来自 Paul Watzlawick 的交际公理）。在一个对称的关系中，两个人的地位是平等的，而在一种互补的关系中，一个人处于更高的地位，就像老板和员工或父母和孩子一样。这极大地影响了互动的风格，特别是在有针对性的活动中。

现在考虑现实世界中的一小群人之间的交互。可能是一个家庭，一帮同事共同吃饭。也可能是孩子们一起在操场上。也可能是朋友和家人聚集在一起度假或庆祝生日。这种互动类型的 VR 版本可以专注于有针对性的活动。也许你是一个不能亲自出席的人，而是通过一些 VR 界面与团队“出去”。也许有一次会议，一些人需要远程参加，这一点在目前通过被称为

“电话会议”的方式进行处理，由网络传输语音和视频。但最接近虚拟现实的常见情景是人们在网络视频游戏中发生的一些社交互动。他们可能会组队并通过短信或语音进行互动。随着人数增加到十几个，互补关系的情况导致人们需要进行自我介绍或采访。这有一些例子如老师向一个班的学生讲课，政治家在一群记者面前讲话等。在这些场景中，领导者已被明确指派与团队沟通。这种设定在 VR 中可通过拍摄全景视频来重现。或者，整个活动直接在虚拟世界中进行。在对称关系的情况下，人们可能会在大型聚会中并以小组形式进行对话。这也可以在 VR 中复制。

### 改善的社交互动

本书中两个共同的主题是 VR 可以产生比现实更好的体验，并且我们的感知系统会适应新的刺激。因此我们会很自然地想知道如何通过 VR 改变或改善社交互动。转型社会交往的概念已经被引入 [16]。一个发人深思的例子如图 10.18 所示。在一个虚拟的世界里，老师可以在讲课时，注视着每个学生。这在现实世界中是不可能的，但在 VR 中很容易制作，因为每个学生都可以看到不同版本的虚拟世界。当然，学生可能会认为老师不可能观察所有人，但是她正在关注自己的可能行会对学习成果产生重大影响。虚拟世界的课堂上看起来可能只有少数学生，而实际上会有成千上万的学生参加。所以还有多少社交互动机制可以在现实世界中实现？我们的大脑能够多快适应它们？在什么样的环境中会更喜欢这种互动，以便在现实世界中相遇？这些问题可以在未来的虚拟世界中得以解决，同时其应会带来更多激动人心的社交新机制。

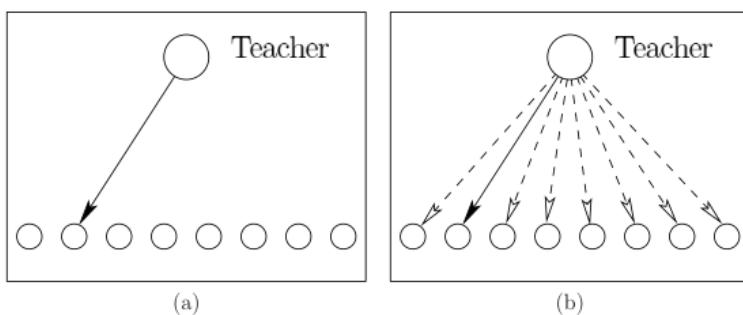


图 10.18：(a) 显示了一个普通教室的自上而下的描述，教师可以直接看一个学生。(b) 在一个 VR 教室里，至少从每个学生的角度来看，老师可以同时看每个学生。

### 10.5 其他交互机制

本章涵盖了三个互动机制的呈现方式：运动，操纵和社交。这些方式从数十年的研究和开发中涌现出来，但并未完全涵盖各种相互作用。许多系统要求构建不属于三种方式的自定义交互机制。此外，随着目前广泛使用低成本的 VR 系统，我们预计新的互动机制将会出现。这里介绍几个其他交互机制和相关的挑战。

#### 与信息和媒体的互动

在遵循通用模拟原则的条件下，互联网内容可以以多种方式进入 VR。第 1.2 节中的图 1.7 显示了虚拟影院中的电影屏幕。在这种情况下，可能需要简单的交互来暂停或更改电影。作为一个更复杂的例子，网络浏览器可以出现在虚拟世界中的公共显示器上或现实世界中用户熟悉的任何其他设备上。或者，虚拟屏幕可以直接浮在用户面前，同时提供稳定，熟悉的背景；见图 10.19。

数十年来，人们通常使用两个输入设备与计算机和浏览器交互，一个用于打字，另一个

用于其他的指向。在个人电脑的情况下，这已经具体为了键盘和鼠标的形式。使用现代智能手机时，人们则需要在小型触摸屏上键入内容，或者使用语音或轻扫式等替代方法。他们通过手指触摸进行指向，而另外一只手指进行屏幕缩放等操作。



图 10.19：在 HTC Vive 中查看 Valve Valve 游戏应用商店。

### 文本输入和编辑

智能手机上的打字选项足以让用户输入检索词或简短信息，但对于撰写小说而言，这些选项还不够。对于目前坐在电脑面前撰写报告、计算机程序、报纸文章的专业人员而言，需要什么类型的界面才能吸引他们在 VR 中工作？

一种选择是跟踪真实的键盘和鼠标。跟踪指尖可能需要提供视觉反馈。这就需要开发相应的系统，将桌面和周围环境神奇地转化为任何东西。就像在桌面系统上使用背景图像一样，令人轻松的全景图像或视频可以在工作时包围用户。对于实际工作部分而言，VR 应该在周围和不同深度出现多个屏幕或窗口，而不是仅在用户面前放置一个屏幕。

这很容易借用现有桌面窗口系统的界面概念进行类比，但仍有许多研究在设计和评估全新的界面，以提高写作时的生产率和舒适度。虚拟现实中的文字处理过程如何？用于编写和调试软件的集成开发环境（IDE）是什么样的？如果键盘和鼠标被其他接口取代，那么用户可能甚至不需要坐在办公桌上工作。这其中的一个挑战是让用户学习一种提供与使用键盘相媲美的文本输入速度的方法，但同时使他们能够更舒适地工作。

### 3D 设计和可视化

能够居住在 3D 虚拟世界的优势是什么？除了视频游戏外，其他几个领域也促进了计算机图形学的发展。在进行计算机辅助设计（CAD）之前，建筑师和工程师花费了很多时间用铅笔和纸张辛苦地在纸上画出精确的线条。计算机已被证明是设计不可或缺的工具。数据可视化在过去几年中一直是计算机领域的关键应用。例如医疗，科学和市场数据。有了这些用途，我们被迫通过在屏幕上操纵 2D 投影来查看设计和数据集。

VR 则可以提供设计或数据集 3D 版本交互和查看的能力。而且被观察事物可以从外向内看，也可以从内部看。如果仔细研究第 6 章中的感知概念，那么设计的对象或环境与真实环境之间的差异可能比以往任何时候都小。在构建物理原型之前，查看 VR 中的设计可以被看作是一种虚拟原型。这可以在产品开发周期中实现快速，低成本的突破。

交互机制是实现基于 VR 的设计和可视化的根本挑战。什么能让建筑师，艺术家，游戏开发人员，电影制作人员或工程师长时间舒适地构建 3D 世界？什么工具可以让人们在 3D 投影世界中准确地操纵高维数据集？

## 未来

即使仅应用通用的模拟原理，也可以设想更多的交互形式。视频游戏已经提供了许多通过标准游戏控制器进行交互的想法。除此之外，任天堂 Wii 遥控器在制作诸如保龄球或网球等运动的虚拟版本时尤其有效。什么样的新互动机制对 VR 来说是舒适和有效的？如果显示器可以呈现出视觉以外的感觉，则会出现更多的可能性。例如，如果一个人穿着一套能够对身体施加适当力量的西服，你能否在虚拟世界中感受到另一边某人的一个拥抱呢？

## 拓展阅读

有关人体运动控制和学习的概述，请参阅书籍[196,271]。[61]中介绍了 VR 环境中的本体感受问题。有关运动的更多信息，请参见[31]的[52]和第 6 章和第 7 章。为了掌握更多关于机器人的问题，请参见[203]。

以下创新性的论文研究了手眼协调的局限性：[56,70,346]。[229]中引入了幂法则，这表明反应时间的对数会随着练习量线性减少。将 Fitts 定律与指向设备操作联系起来的研究包括[72,193,194,302]。有关人机交互的广泛报道，请参阅[37,40]。有关通过头像进行社交互动的其他参考资料，请参[20,210,328]。

# 第十一章 音频

听觉是 VR 重要的一部分，但直到本章才开始被正式提及。VR 系统的开发人员往往主要关注视觉部分，因为这是我们最强烈的感觉；然而，VR 的音频组件功能强大，并且存在将高保真音频体验带入 VR 的技术。在现实世界中，音频对艺术、娱乐和口头交流至关重要。正如第 2.1 节所述，音频录制和复制本身也可以被视为 VR 体验，既有穴状的版本（环绕声），也有头戴式的版本（戴着耳机）。当其与视觉部分结合在一起时，音频有助于提供引人注目和令人舒适的 VR 体验。

本章的每一部分都是对第 4 章到第 7 章中的听觉（或音频）补充。整个过程从物理学到生理学，然后从感知到渲染。第 11.1 节以波、传播和频率分析来解释声音的物理特性。11.2 节描述了人耳的各个部分及其功能。这自然会引导到听觉部分，便是第 11.3 节的内容。第 11.4 节呈现了听觉渲染的过程，可以通过模型合成，产生声音或再现所采集的声音。在阅读这些章节时，务必记住每个主题的视觉部分。其中的相似之处使得它更容易被理解，差异则导致了不同寻常的工程解决方案。

## 11.1 声音的物理学

本节与第 4 章中的许多概念相似，后者涵盖了光的基本物理学。声波传播在许多方面与光相似，但有一些关键的差异，会导致重大的感知和工程后果。光是横波，在垂直于其传播的方向上振荡，声音是纵波，在平行于其传播的方向上振荡。图 11.1 给出了一个平行波前的例子。

声音对应于介质中的振动，通常是空气，但也可以是水或任何其他气体、液体或固体。真空中没有声音，这与光传播不同。对于声音来说，介质中的分子发生位移，引起压力变化，从极端压缩到极端稀疏解压缩。在空间的固定点，压力随时间而变化。最重要的是，这可能是人耳鼓膜上的压力变化，它会转化为感知体验。声压水平通常以分贝（简写为 dB）做单位，其定义如下

$$N_{db} = 20 * \log_{10}(p_e/p_r). \quad (11.1)$$

上式中， $p_e$  是峰值压缩的压力水平， $p_r$  是参考压力水平，通常取  $2 \times 10^{-7}$  牛顿/平方米。声波通常通过振动固体材料产生，特别是当它们相互碰撞或相互作用时。一个简单的例子是敲击一口大钟，使其振动数秒钟。通过足够的空气流动，材料也可能会被强制振动，例如吹长笛的情形。人体通过使用肺来强制空气通过声带产生声音，从而使声音振动。这使得我们会说话、唱歌、尖叫等等。

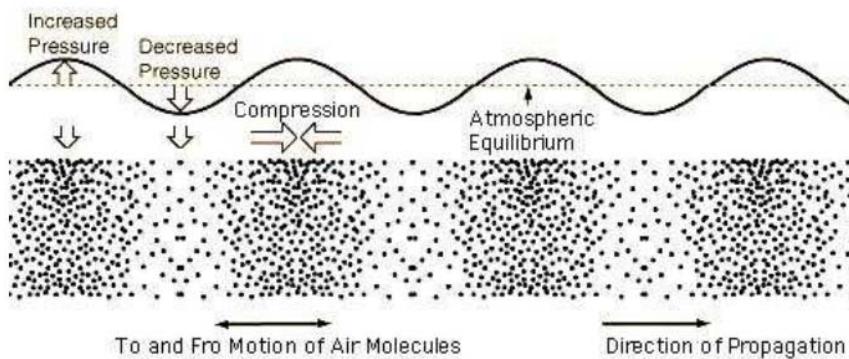


图 11.1 声音是空气分子的纵向压缩和稀疏波。这里显示纯音的情况，这生成了正弦函数。（图源于 Pond Science Institute）

## 声源和衰减

就光线而言，我们可以考虑每条声线垂直于声音传播波阵面的光线。可以定义一个点声源，它可以在各个方向产生相同功率的发射射线。这也导致功率关于离源距离的函数以二次方速率降低。这样的点源对建模很有用，但在现实世界中很难实现。平面波前可以通过振动大而平坦的平板来实现，其导致平行光的声学等效。然而，一个重要的区别是声音在媒介中传播时的衰减。由于分子振动中的能量损失，对于距离平面源的每个单位距离，声音强度以固定因子（或固定百分比）降低；这是一个指数衰减的例子。

传播速度：声波以每秒 343.2 米的速度在 20° C (68° F) 的空气中传播。作为比较，光传播速度快了约 874,000 倍。我们有飞机和汽车可以超过声速，但远不及以光速行驶。这可能是制作 VR 系统的声音和光线之间最重要的区别。其结果是人类感官和工程传感器可轻松测量声波到达时间的差异，从而更强调时间信息。

## 频率和波长

如 4.1 节所述，将波分解为频率分量变得重要。对于声音，频率是每秒压缩的次数，称为音频。通常认为该范围是从 20Hz 到 20,000Hz，这是基于人类听觉的，与光的频率范围基于人类视觉的方式大致相同。超过 20,000 赫兹的振动被称为超声波，可以被某些动物听到。低于 20 Hz 的振动称为次声波。

使用 4.1 节中的 (4.1) 和传播速度  $s = 343.2$ ，还可以确定声波的波长。在 20Hz 时，波长  $\lambda = 343.2 / 20 = 17.1\text{m}$ 。在 20,000 赫兹时，它变为  $\lambda = 17.1\text{mm}$ 。这大约是我们世界中物体的大小。所以，这会导致声音会以复杂的方式被一些物体所干扰，在尝试重现 VR 中的行为时很难进行建模。相比之下，光波的波长很小，范围从 400 纳米到 700 纳米。

## 多普勒效应

上述声压变化是针对固定的接收点。如果该点远离声源，则波前将以较低的频率到达。例如，如果接收器以  $43.2\text{m/s}$  的速度离开声源，那么波似乎仅以每秒  $343.2 - 43.2 = 300$  米的速度行进。接收到的频率由于声源和接收器之间的相对运动而发生变化。这就是所谓的多普勒效应，在接收器处测得的频率可以计算为

$$f_r = \left( \frac{s + v_r}{s + v_s} \right) f_s, \quad (11.2)$$

其中  $s$  是介质中的传播速度， $v_r$  是接收器的速度， $v_s$  是声源的速度， $f_s$  是声源的频率。在我们的例子中， $s = 343.2$ ， $v_r = -43.2$ ， $v_s = 0$ 。结果是，频率  $f_s = 1000\text{Hz}$  的声源将被接收器感知为具有频率  $f_r \approx 876.7$ 。这就是为什么警笛似乎在警车经过时改变声音的原因。多普勒效应也适用于光照，但在正常的 VR 环境下效果可以忽略不计（除非开发人员想要试验虚拟时间膨胀，太空旅行等等）。

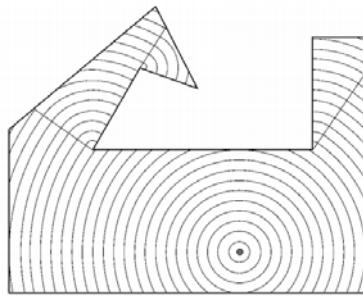


图 11.2：由于衍射，波甚至可以在角落弯曲。这里显示了房间的自顶向下视图。在三个内角处，传播的波前都在其周围扩大。

### 反射和传播

与光一样，波传播受到媒介传播的强烈影响。想象一下，当某人在房间里面大喊大叫的时候，声波就会冲击内墙。由于反射，大部分声音都会被弹起来，就好像墙壁是一面镜子一样。但是，一些声能会穿透墙壁。声音通过坚固的材料传播得更快，导致其穿透时弯曲。这是折射。有些声音从墙壁的另一边穿过并通过相邻房间的空气传播，导致进一步地传播。因此，邻近房间里的人可以听到大喊。声波在撞击到墙壁之前所含的总能量会被反射和折射分开，并且由于衰减而产生额外的损失。

### 衍射

波前也可以在拐角处弯曲，这就是所谓的衍射；见图 11.2。这将使人们能够听到建筑物角落的声音，而不依赖任何反射或传输。更长的波长会发生更多的衍射；因此，较低音频的声音更容易在角落附近弯曲。也解释了为什么相比光衍射，我们更关心房间的声衍射，尽管后者通常对透镜来说很重要（回想第 7.3 节的菲涅耳透镜的缺点）。

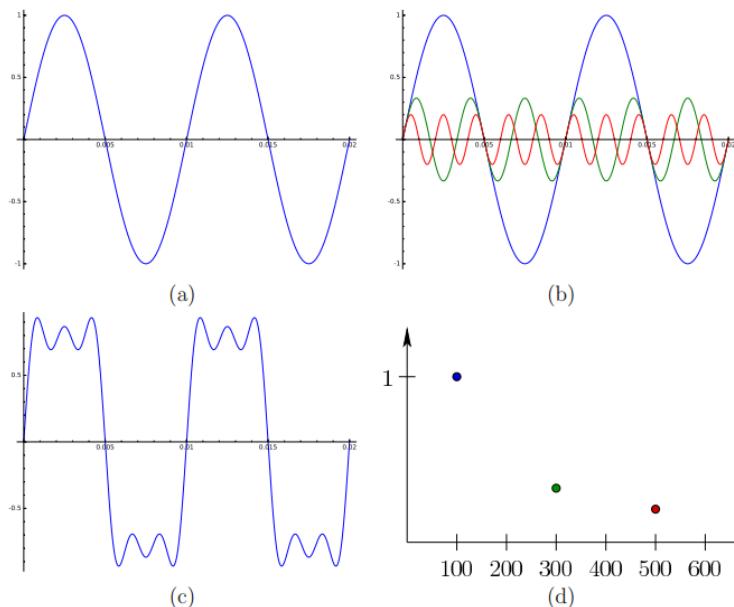


图 11.3：(a) 单位幅度和频率 100Hz 的纯音（正弦曲线）。(b) 三种纯音；除原始蓝色外，绿色正弦波幅度为  $1/3$ ，频率为 300 赫兹，红色波幅为  $1/5$ ，频率为 500 赫兹。(c) 直接添加三个纯音，会产生接近方形的波形。(d) 在频谱中，有三个非零点，每个纯音一个。

## 傅里叶分析

光谱分解对于表示 4.1 节中的光源和反射非常重要。在声音的情况下，它们甚至更重要。如图 11.3 (a) 所示，正弦波对应于纯音，其具有单个相关频率；这类似于来自光谱的颜色。更复杂的波形，例如钢琴音符的声音，可以由各种纯音的组合构成。图 11.3 (b) 至 11.3 (d) 提供了一个简单的例子。这个原理来自傅里叶分析，它可以通过简单地将它们相加来将任何周期函数分解为正弦曲线（在我们的例子中是纯色调）。每个纯音都有一个特定的频率，幅度或比例因子，以及一个可能的峰值定时，称为相位。通过简单地添加有限数量的纯音，实际上任何有用的波形都可以近似地逼近。较高频率的低幅度正弦曲线通常称为高次谐波；最大的幅度波被称为基频。作为频率函数的幅度和相位的曲线通过应用傅里叶变换获得，将在 11.4 节中简要介绍。

## 镜头在哪里？

在这一点上，与第 4 章相比最明显的遗漏是镜头的声学等价物。如上所述，折射发生在声音上。为什么人耳不像眼睛那样可以将声音聚焦在空间图像上呢？一个原因是与光相比较长的波长。回顾第 5.1 节，中心窝的感光密度接近可见光的波长。“耳朵凹陷”可能要跨越几米或更多，这种结构下，人类的头会非常大。另一个问题是低频声波以更复杂的方式与世界中的物体相互作用。因此，我们的耳朵不是形成图像，而是通过执行傅里叶分析来筛选出各种频率，振幅和相位的正弦波形式的声波结构。每个耳朵更像是以每秒数万帧操作的单像素相机，而不是以较慢的帧速率捕捉较大的图像。听觉是在时间上产生重要作用，而视觉主要从空间上产生作用。尽管如此，时间和空间对听力和视觉都很重要。

## 11.2 人类听觉的生理学

人耳将声压波转换为神经冲动，最终导致感知的体验。人耳的解剖结构如图 11.4 所示。基于声波的流动，耳朵分为外部，中部和内部。回顾第 5.3 节眼球运动的复杂性。猫和其他动物都可以旋转它们的耳朵，但人类不能，这简化了 VR 工程的相关问题。

### 外耳

从头部突出的耳朵的松软部分被称为耳廓。它主要作为收集声波并将其导入漏斗状的耳道内。它具有放大 1500 至 7500Hz 频率范围内的声音的效果[362]。它还会对声音进行微妙的过滤，导致高频范围内的某些变化取决于声源的射入方向。这为声源的方向提供了强有力的线索。

沿着耳道行进后，声波使耳膜振动。耳膜是一个锥形膜，将外耳与中耳分开。它只覆盖了 55 平方毫米的面积。如果这是一台相机，那么在这一点上它将具有一个像素的分辨率，因为除了可以从膜振动推断出的信息之外不存在额外的空间信息。

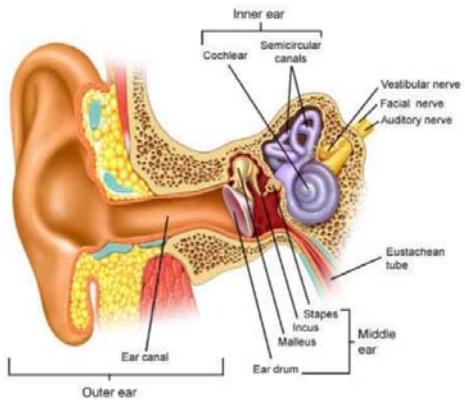


图 11.4: 人类听觉系统的生理学结构

## 中耳

中耳的主要功能是将外耳中的振动空气分子转换成内耳中的振动液体。这是通过将耳膜连接到内耳的骨骼完成的。内耳的空气和液体具有不同的阻抗，这是抗振动的。骨骼被称为锤骨（锤），砧骨（砧）和镫骨（镫），它们通过肌肉和韧带串联连接，允许相对运动。骨骼的目的是匹配阻抗，使压力波以尽可能小的功率损失传输到内耳。这避免了阻抗较高的材料反射声音的情况。这方面的一个例子是声音在湖面上发生的是反射，而不是传播到水中。

## 内耳

内耳包含 8.2 节所述的前庭器官和耳蜗，它是听觉的器官。耳蜗通过机械感受器将声音能量转化为神经冲动。这是以一种完美的方式完成的，该方法在该过程中执行谱分解，以便神经脉冲对频率分量的幅度和相位进行编码。

图 11.5 说明了它的过程。如图 11.5 (a) 所示，耳膜振动转换为耳蜗底部卵圆窗的振荡。含有被称为外淋巴液的液体的管从椭圆形窗口延伸到另一端的圆窗。基底膜是穿过耳蜗中心的结构，其大致使含有外淋巴管的长度加倍。管的第一部分称为前庭阶，第二部分称为鼓阶。随着椭圆形窗口的振动，波沿着管道传播，导致基底膜移位。在基部附近（靠近椭圆形和圆形的窗户），膜是薄的和刚性的，并且在最远的点（称为顶点）上逐渐变软和松软；见图 11.5 (b)。这导致膜上的每个点仅在特定的窄频率范围内振动。

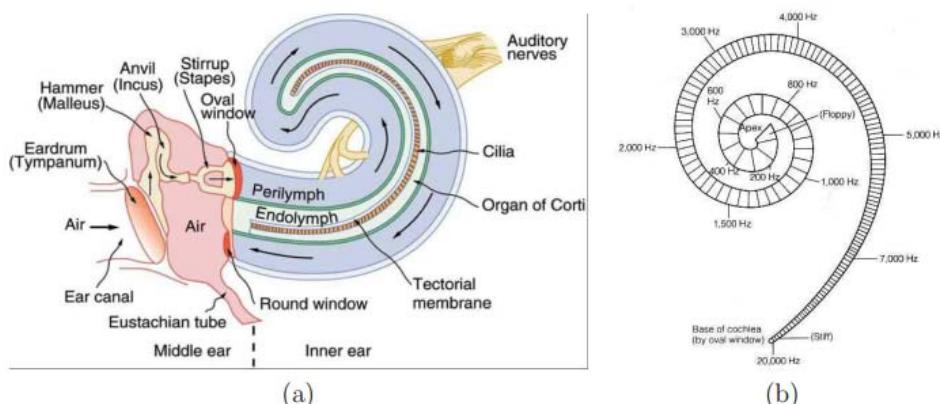


图 11.5: 耳蜗的运作: (a) 外淋巴传送由椭圆形窗口强迫形成的波，通过延伸耳蜗长度并返回到圆窗的管。(b) 由于厚度和硬度的变化，中心脊柱（基底膜）对特定频率的振动敏感；这触动了机械刺激感受器，并最终听觉感知，这是频率敏感的。

## 机械刺激感受器

基底膜被称为哥蒂氏器官的更大更复杂的结构所包围，它还含有类似于 8.2 节所示的机械刺激感受器。见图 11.6。机械刺激感受器将毛发的位移转化为神经冲动。当基底膜振动时毛发会移位，因为一些毛细血管的末端附着在盖膜上。基底膜和盖膜的相对运动引起剪毛动作，使毛发移动。每只耳朵含有约 20,000 个机械刺激感受器，这比眼睛中 1 亿个感光器少得多。

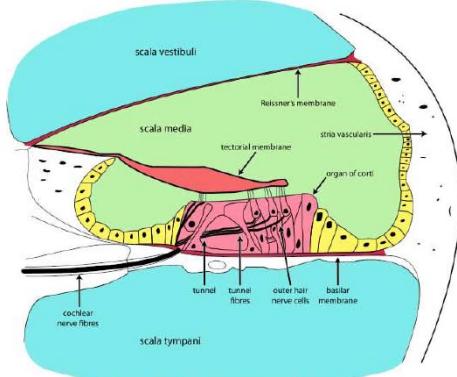


图 11.6：螺旋器器官的横截面。基底膜和盖膜相对于彼此移动，导致机械性感受器中的毛发弯曲。（来自多个维基百科用户。）

## 光谱分解

通过利用基底膜基于频率的灵敏度，大脑可以有效地获得入射声波的频谱分解。它与第 11.1 节讨论的傅立叶分解类似，但并不完全相同。在第 4 章的[204]提到了一些二者之间的不同之处，如果两个不同频率的单音同时出现在耳朵上，那么可以听到基底膜产生的第三种音调[149]。此外，由机械性感受器输出引起的神经冲动与频率幅度不成正比。此外，单音检测可能会导致检测到附近的音调（频率）被抑制[277]，非常类似于水平单元中的横向抑制（5.2 节）。第 11.4.1 节将阐明这些差异如何使人耳在滤波方面更加复杂。

## 听觉途径

神经脉冲从左耳蜗和右耳蜗传播到听觉的最高水平中枢，也就是大脑中的主要听觉皮层。像往常一样，当信号通过神经结构组合时发生分层处理，这使得可以分析多个频率和相移。上级橄榄的早期结构从两个耳朵接收信号，以便可以处理幅度和相位的差异。这在第 11.3 节中对于确定音频源的位置将变得重要。在最高水平上，初级听觉皮层是以 tonotopically（位置基于频率）绘制出来的，与视觉皮层的表面图非常相似。

### 11.3 听觉感知

现在我们已经了解了听觉器官，下一部分是知道我们如何感知声音。在视觉案例中，我们看到感知体验往往令人惊讶，因为它们基于适应，缺失数据，由神经结构填充的假设以及许多其他因素。听觉体验也是如此。此外，听觉幻象与光学幻觉一样存在。第 6.4 节的 McGurk 效应是一个使用视觉诱发不正确听觉的例子。

## 优先效应

更常见的听觉错觉是优先效应，如果两个几乎相同的声音在稍微不同的时间到达，则只有一个声音被感知，见图 11.7。声音经常从表面反射，引起混响，这是由于从反射，透射和

衍射获取的不同传播路径而引起的声音延迟“副本”到达耳朵造成的。但人们通常听到的仍是一个声音，而不是听到混乱的声音。这通常由具有最大振幅声波决定。如果定时差异大于回声阈值（在一项研究中，其范围为 3 至 61ms [358]），则会感知到回声。其他听觉幻觉会产生不正确的定位（Franssen 效应和 Glissando 幻觉[60]），幻觉的连续性[339]以及永久增加的音调（Shepard tone illusion [285]）。



图 11.7：由于优先效应，如果头部放置在立体声扬声器之间，则会出现听觉错觉，因此一个比另一个靠得更近。如果它们同时输出相同的声音，则人会感知到来自靠近扬声器的声音，而不是感知回声。

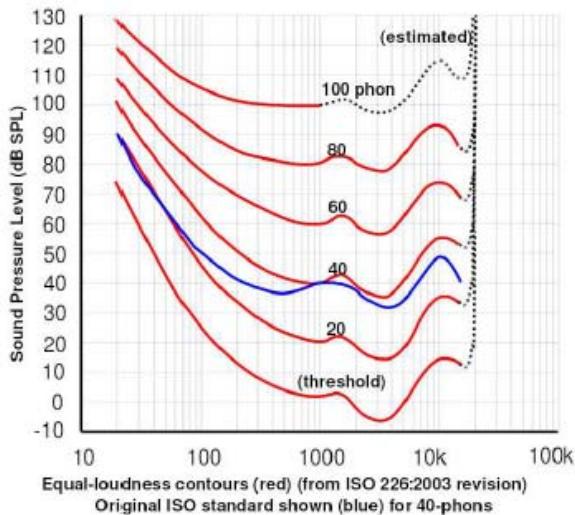


图 11.8：作为频率函数的等响度感知等高线。

### 心理声学和响度感知

第 2.3 节介绍的心理物理学领域专门用于听觉感知情况下的心理声学。史蒂文斯的感知刺激强度定律和韦伯定律的明显差异 (JNDs) 是重要的理论基础。例如，史蒂文斯定律的指数（回忆 (2.1)），对于 3000Hz 纯音的感知响度  $x = 0.67$  [311]。这大致意味着，如果声音增加到更高的压力水平，我们认为它只是更响亮一点。来自心理声学的一个更复杂的例子如图 11.8 所示，它对应于作为频率函数的等响度感知的等高线。换句话说，随着频率的变化，声音被认为有相同的响度时是在什么水平？这需要仔细设计人体对象的实验，这也是 VR 开发过程中常见的问题，见第 12.4 节。

### 音调感知

当考虑感知时，声波的频率被称为音调。知觉心理学家已经研究了人们检测目标音调的能力，尽管其他波长和相位的声音都有所混淆。一个基本的观察结果是，听觉感知系统执行临界带遮蔽以有效地阻挡具有特定感兴趣范围外的频率的波。另一个深入研究的问题是对音高（或频率）差异的认识。例如，对于 1000 赫兹的单音，有人可以将其与 1010 赫兹的音调区分开来吗？这是 JND 的一个例子。事实证明，对于低于 1000 赫兹的频率，人类可以检测到小于 1 赫兹的频率变化。随着频率的增加，鉴别能力下降。在 10,000 赫兹时，JND 约为 100 赫兹。就百分比而言，这意味着音调感知在低频率方面有优于 0.1% 的差异，但对于较高频率则增加到 1.0%。

另外关于音高感知，当基频从复杂波形中移除时，会出现令人惊讶的听觉错觉。回想一下图 11.3，方波可近似表示为增加幅度越来越小但幅度更高的正弦曲线。事实证明，人们可以感觉到基本频率的音调，即使它被移除，只有高次谐波仍然存在，在[204]中的第 5 章总结了几种理论。

## 定位

心理声学的主要领域之一是定位，这意味着通过听到声音来估计声源的位置。这对许多 VR 体验至关重要。例如，如果人们正在社交，那么他们的声音应该来自相应虚拟形象的嘴巴才会更真实。换句话说，听觉和视觉线索应该匹配。任何类型的声音效果，如汽车经过，都应该有相应的线索。

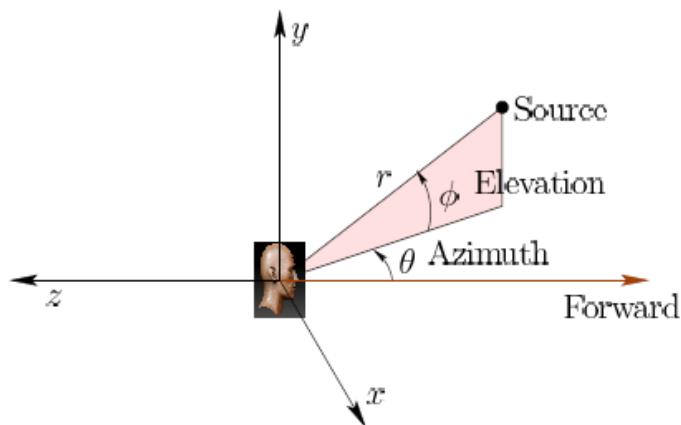


图 11.9：球形坐标用于听觉定位中的源点。假设头部以原点为中心并朝向-z 方向。方位角  $\theta$  是在将源投影到  $xz$  平面之后相对于正向的角度。仰角  $\phi$  是由一个垂直三角形构成的内角，该垂直三角形将原点连接到光源并将光源投影到平面中。半径  $r$  是从原点到源的距离。

将 JND 概念应用于定位以获得最小可听角度 (MAA)，其是可以由人类听觉检测到的角度变化的最小量。通常使用球坐标系进行定位，其中听者头部位于原点，见图 11.9。在正向和源之间的水平面内的角度被称为方位角，其从-180 度延伸到 180 度。角度对应于源与水平面的偏差称为高度，从-90 度延伸至 90 度。第三个坐标是从原点（头部中心）到源的半径或距离。MAA 取决于频率和源的方向。图 11.10 显示了 MAA 作为频率函数的曲线图，几个方位角值。变化量令人惊讶。在某些频率和地点，MAA 降至 1 度。但是，在其他组合中，定位非常糟糕。

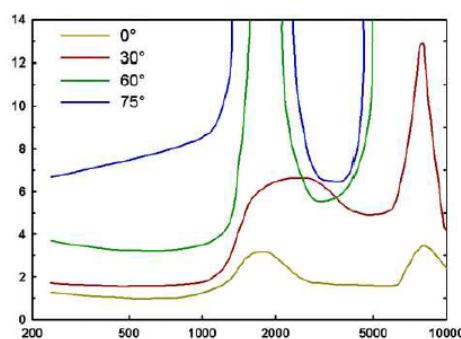


图 11.10：作为频率函数的最小可听角 (MAA) 图。每个绘图对应于不同的方位角。

## 单声道提示

听觉定位类似于视觉的深度和尺度感知，6.1 节已对此进行了介绍。由于人类有一对耳朵，因此可以将定位分为使用单耳和使用双耳的两种提示。这类似于单眼和双目视觉提示。单声道提示依靠到达单耳的声音来约束可能的声源集合。几个单声道提示[363]：

1.耳廓的形状是不对称的，所以传入的声音会以取决于它到来的方向，特别是仰角的方式变形。尽管人们没有意识到这种扭曲，但听觉系统将其用于定位。

2.声音的振幅随着距离的变化而呈二次曲线下降。如果它是一种熟悉的声音，那么它的距离可以从感知的幅度来估计。熟悉度影响这种线索的方式与熟悉物体使得深度和尺度感知被分开的方式相同。

3.对于远处的声音，会出现频谱失真，因为高频分量比低频分量衰减更快。例如，遥远的雷声被认为是深沉的隆隆声，但是附近的雷声包含了更高的爆音。

4.最后，随着声音反弹，进入耳朵的混响提供了强大的单声道提示，这在室内体现的尤其强大。尽管优先效应阻止我们感知这些混响，但大脑仍然使用这些信息进行定位。这种提示单独被称为回声定位，这是由一些动物，包括蝙蝠所使用的。有些人可以通过发出咔嗒声或其他尖锐的声音来执行此操作；这使得盲人可以进行声学寻路。

## 双耳定位

声源定位则使用双耳听到的线索作为结果。最简单的情况是耳间水平差异 (ILD)，这是每个耳朵听到的声音幅度的差异。例如，一只耳朵可能面对声源，而另一只耳朵则位于声影中（由声源前方的物体造成的阴影与来自光源的阴影相似）。较近的耳朵会比另一耳朵更强烈的振动。

另一种双耳线索是双耳时间差 (ITD)，与第 9.3 节所述的 TDOA 感知方法密切相关。两个耳朵之间的距离大约为 21.5 厘米，这会导致来自声源的声音到达时间不同。请注意，声音在大约 0.6ms 内传播 21.5 厘米，这意味着定位使用了令人惊讶的小差异。

假设大脑将到达时间的差异测量为 0.3ms。源可能产生的地方是什么？这可以通过设置代数方程来解决，这会形成锥面被称为双曲面。如果不知道哪个声音首先出现，那么这组可能的地方是两个不相交的双曲面。由于大脑知道哪一个先出现，这可能区域被缩小成一个双曲面片，这被称为混淆锥，见图 11.11（在大多数情况下，即使它是双曲面，它看起来大致像一个圆锥体）。然而，通过使用耳廓的变形，可以部分地解决该锥体内的不确定性。

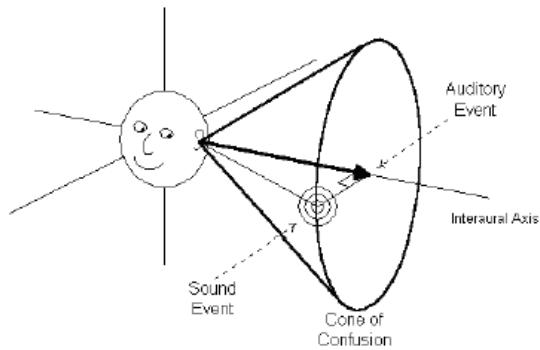


图 11.11：混淆锥是使用 ITD 双耳线索后点源可能存在的位置集合。它在实现上是一个双曲面，但大致看起来像一个锥体。

## 运动的力量

更重要的是，人类通过简单地移动头部就可以解决很多模糊问题。正如头部运动导致视

差带来的强大视觉深度提示一样，它也提供了更好的听觉定位。事实上，听觉视差甚至提供了另一个定位提示，因为附近的音频源会更快地改变它们的方位角和仰角。关于 ITD，想象在短时间内，每个头部姿势都有一个不同的混淆锥。通过整合其他感官，可以估计相对的头部姿势，这便可以大致形成多个混淆锥的交集，直到声源精确定位为止。最后请记住，源相对于接收器的运动会引起多普勒效应。相比于视觉而言，基于听觉的感知结果与物体的实际运动相关联。这对媒介过程是有帮助的（回顾第 8.2 节）。

## 11.4 听觉渲染

我们现在开始讨论为虚拟世界制作声音并将它们发送到听觉显示器（扬声器）的问题，以便用户感觉它们是为 VR 体验设计的。它们应该与视觉线索和现实世界中的听觉体验一致。无论是录制的声音，合成声音还是其组合，虚拟压力波及对演讲者的渲染，这些虚拟声音都应能欺骗到用户的大脑。

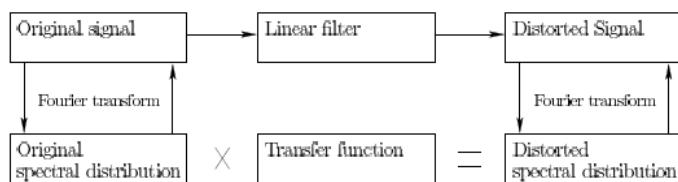


图 11.12：线性滤波器的概述及其与傅里叶分析的关系。顶部块对应于时域，而最下行是频率（或频谱）域。

### 11.4.1 基本信号处理

到目前为止，声波中频率分量的重要性应该是清楚的。这仍然适用于 VR 合成声音的工程问题，该问题属于信号处理领域。这里给出一个简要的概述，见[11, 189]以供拓展阅读。由于这个主题的核心是对信号进行转换或扭曲的滤波器的表征或设计。在我们的例子中，信号是可以完全合成的，如使用麦克风捕获的声波或某种组合。（回想一下，合成模型和捕获模型也存在于视觉案例中。）

图 11.12 显示了整个方案，将在本节中介绍。原始信号出现在左上角。首先，按照从左到右的路径。信号进入一个标有线性滤波器的黑盒并变为失真状态，如右图所示。什么是线性滤波器？在回答这个问题之前，需要一些背景概念。

### 采样率

信号处理公式既适用于连续时间，它可以提供很好的公式和数学证明，也可以用于离散时间，直接对应于计算机处理信号的方式。由于其实用价值，我们主要关注离散时间案例。

将信号作为时间的函数，其值表示为  $x(t)$ 。使用数字信号处理，它将定期进行采样。设  $t$  是采样间隔。采样率或（采样频率）大约为  $1/\Delta t$  Hz。例如，采样频率为 1000Hz 时， $\Delta t$  为 1 毫秒。根据奈奎斯特 - 香农采样定理，采样率应该至少是信号中最高频率分量的两倍。由于音频的最高频率分量为 20,000 Hz，这表明采样率应至少为 40,000 Hz。并非巧合，CD 和 DVD 的采样率分别为 44,100 Hz 和 48,000 Hz。

通过对信号进行采样，可以生成一个数组。在 1000Hz 条件下，数组每秒钟将包含一千个数值。使用索引变量  $k$ ，我们可以将第  $k$  个样本称为  $x[k]$ ，它对应于  $x(k\Delta t)$ 。例如，第一个样本是  $x[0] = x(0)$ 。

### 线性滤波器

在信号处理的背景下，滤波器的作用是将一个信号映射到另一个信号。每个信号都是时

间的函数，滤波器就像一个黑盒子，接收一个信号作为输入，并产生另一个信号作为输出。如果  $x$  代表一整个信号（全程），那么可以令  $F(x)$  代表通过滤波器运行后得到的信号。

线性滤波器是一种特殊的滤波器，它满足两个代数性质。第一个代数属性是加法性质，这意味着如果两个信号被添加并通过滤波器发送，则结果应该与它们各自独立地通过滤波器并加和后的信号相同。即对于任何两个信号  $x$  和  $x'$ ，有  $F(x+x') = F(x) + F(x')$ 。例如，如果两个不同的声音被发送到滤波器中，无论它们在过滤之前还是之后进行组合，结果都应该是相同的。随着多个正弦波通过滤波器，这个概念将变得很有用。

第二个代数属性是同质性的，这意味着如果信号在经过滤波器发送之前通过一个常数因子进行缩放，则结果将与之后由相同因子缩放的结果相同。使用符号表示，对于每个常数  $c$  和信号  $x$ ，有  $cF(x) = F(cx)$ 。这意味着如果我们将声音幅度加倍，那么来自滤波器的输出声音也将其幅度加倍。

线性滤波器通常采用这种形式

$$y[k] = c_0x[k] + c_1x[k-1] + c_2x[k-2] + c_3x[k-3] + \cdots + c_nx[k-n], \quad (11.3)$$

其中每个  $c_i$  都是一个常数，并且  $n+1$  是与滤波器相关的样本数量。可以考虑  $n$  趋于无穷的情况，但这里不会。毫不意外，(11.3) 是一个线性方程。这种特殊的形式也导致了这种滤波器是因果滤波器，因为等式右边的样本从时间上不会晚于样本  $y[k]$ 。非因果性滤波器需要依赖未来的样本，这对记录的信号是合理的，但不适用于现场采样（未来是不可预测的！）。

以下是线性滤波器的一些例子 ((11.3) 的特例)。它们取最后三个样本的移动平均值：

$$y[k] = \frac{1}{3}x[k] + \frac{1}{3}x[k-1] + \frac{1}{3}x[k-2]. \quad (11.4)$$

这是指数平滑（也称为指数加权移动平均）的示例：

$$y[k] = \frac{1}{2}x[k] + \frac{1}{4}x[k-1] + \frac{1}{8}x[k-2] + \frac{1}{16}x[k-3]. \quad (11.5)$$

### 有限脉冲响应

一个重要且有用的结果是，线性滤波器的行为可以用其有限脉冲响应 (FIR) 来充分表征。(11.3) 中的滤波器通常被称为 FIR 滤波器。有限脉冲是所有  $k > 0$  时  $x[0] = 1$  且  $x[k] = 0$  的信号。任何其他信号可以表示为时移有限脉冲的线性组合。如果有限脉冲发生偏移，例如  $x[2] = 1$ ，对于所有  $k \neq 2$  且  $x[k] = 0$  的情况下，线性滤波器会产生相同的结果，但它仅延迟两步。由于滤波器的线性，有限脉冲可以重新调整，输出只是重新调整比例。通过滤波器发送缩放和移位脉冲的结果也可以由线性直接获得。

### 非线性滤波器

不遵循式 (11.3) 的任何（因果）滤波器称为非线性滤波器。回忆第 11.2 节，人类听觉系统的操作几乎就是一个线性滤波器，但展现出非线性滤波器的特性。线性滤波器被优先选择，因为它们与信号的频谱分析或频率分量紧密连接。即使人类听觉系统包含一些非线性行为，但基于线性滤波器的分析仍然是有价值的。

### 返回到傅里叶分析

现在考虑图 11.12 的底层。线性滤波器的操作在频域中易于理解和计算。这是通过对信

号进行傅立叶变换而获得的函数，该信号为每个频率和相位的组合提供一个幅度。这个转换在 11.1 节简要介绍过，并在图 11.3 中说明。形式上，它被定义为离散时间系统

$$X(f) = \sum_{k=-\infty}^{\infty} x[k]e^{-i2\pi fk}, \quad (11.6)$$

其中  $X(f)$  是所得到的频谱分布，是频率  $f$  的函数。指数包含  $i = \sqrt{-1}$ ，并通过欧拉公式与正弦曲线相关：

$$e^{-i2\pi fk} = \cos(-2\pi fk) + i \sin(-2\pi fk). \quad (11.7)$$

单位复数被用作表示相位。傅里叶逆变换的形式类似，并将频谱分布转换回时域。这些计算在实践中通过使用快速傅立叶变换（FFT）可以快速执行[11,189]。

### 传递函数

在一些情况下，需要通过修改频谱分布来设计线性滤波器。它可以放大一些频率，同时抑制其他频率。在这种情况下，滤波器根据传递函数来定义，其应用如下：1) 使用傅立叶变换来变换原始信号，2) 将结果与传递函数相乘以获得失真的频谱分布，最后 3) 应用傅里叶逆变换以获得作为时间函数的结果。通过将离散拉普拉斯变换（称为  $z$  变换）应用于有限脉冲响应，可以从线性滤波器计算传递函数[11,189]。

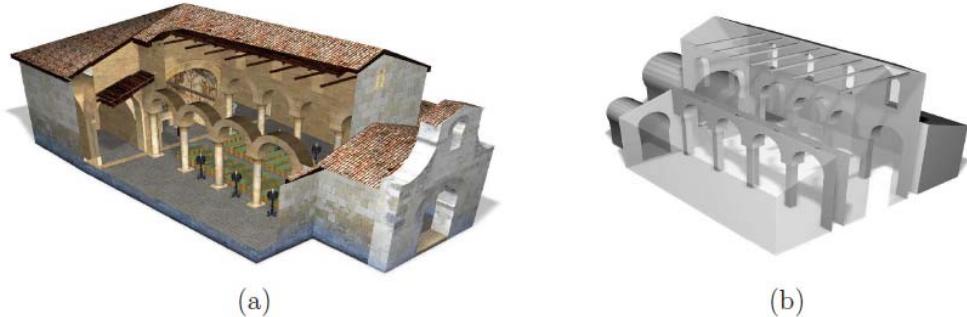


图 11.13：音频模型要简单得多。（来自 Pelzer, Aspock, Schroder 和 Vorlander, 2014, [248]）

### 11.4.2 声学建模

除了视觉方面之外，第 3.1 节中的几何建模概念适用于 VR 的听觉方面。事实上，两种都可以使用相同的模型。在虚拟世界中反射光的墙也反射声波。因此，两者都可以用相同的三角形网格表示。理论上这没问题，但细节的精细程度或空间分辨率对音频无关紧要。由于视觉敏锐度高，为视觉渲染设计的几何模型可能具有高度的细节。回想 5.4 节，人类可以根据不同的视角区分 30 条或更多的条纹。在声波的情况下，小的结构基本上不可感知。一个建议是，声学模型需要的空间分辨率只有  $0.5m$  [334]。图 11.13 显示了一个例子。因此，任何小波纹，门把手或其他精细结构都可以被简化。自动转换为视觉渲染而设计的 3D 模型到为针对听觉渲染优化的 3D 模型仍然是一个挑战。

现在考虑虚拟环境中的声源。例如，这可能是发出声波或振动平面表面的点。白光的等价物称为白噪声，理论上它包含可听频谱中所有频率的相等权重。来自模拟电视或收音机的纯净电就是一个近似的例子。在实际环境中，感兴趣的声音在特定频率中集中度很高，而不是均匀分布。

声音如何与表面相互作用？这与 7.1 节中的阴影问题类似。对于光线情况，漫反射和镜面反射依赖颜色发生。而在声音的情况下，同样存在两种可能性，同样依赖于波长（或者等同于频率）。对于大而平滑的表面，会发生声波的镜面反射，出射角度等于入射角度。反射的声音通常具有不同的幅度和相位。由于部分声音会被吸收到材料中，所以振幅可能会减小一个常量。该因数通常取决于波长（或频率）。[334]包含许多常见材料的吸收系数。

对于较小的物体或具有重复结构的表面（如砖块或波纹），声波可能会以难以表征的方式散射。这与光的漫反射相似，但声音的散射模式可能难以模拟和计算。一个不幸的问题是散射行为取决于波长。如果波长比结构（整个物体或波纹）的尺寸小得多或比其大的多，那么表面主要反射声波。如果波长接近结构尺寸，则可能发生显著且复杂的散射。

在建模重任的极端情况下，可以构建双向散射分布函数（BSDF）。BSDF 可以通过放置在不同位置的扬声器和麦克风阵列的组合来估计真实世界中的等效材料，以测量特定方向上的散射。对于波长较大的平面材料来说，这种等效的结果可能很好，但它仍不能处理可能出现在表面上的各种复杂结构和图案。

### 捕捉声音

声音也可以使用麦克风在现实世界中被捕捉，然后将其带入物理世界。例如，匹配区域可能包含麦克风，这些麦克风会成为现实世界中等效的扬声器。就视频采集而言，制作一个完全捕捉声场的系统是具有挑战性的。在[256]中提出了基于多个麦克风捕获的声音插值的简单而有效的技术。

#### 11.4.3 听觉化

虚拟世界中的声音传播与视觉渲染一样，处理波的传播主要有两种方式。最昂贵的方法是尽可能准确地模拟物理世界，包括计算精确模拟波传播的偏微分方程的数值解。廉价的方法是拍摄可见光线并表征声源，表面和耳朵之间的主要相互作用。两种方法之间的选择还取决于特定的设置；一些系统便涉及这两种计算[208,334]。如果波与环境中的物体相关程度较大，换句话说，频率低并且几何模型具有高度的细节时，则优先数值方法。。在高频或更大，更简单的模型中，基于可见性的方法将被优先。

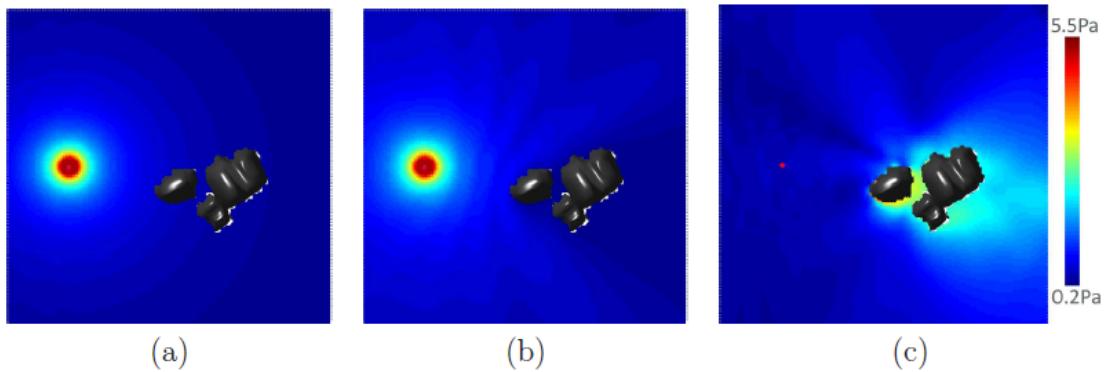


图 11.14：通过数值求解 Helmholtz 波动方程计算声传播结果（取自[208]）：(a) 考虑障碍物相互作用前的压力大小。(b) 考虑到散射后的压力。(c) 散射分量，即 (b) 的压力减去 (a) 的压力。

### 数值波传播

Helmholtz 波动方程根据压力函数的偏导数来表示  $R^3$  中每个点的约束。它基于频率的形式是

$$\nabla^2 p + \frac{\omega^2}{s^2} p = 0, \quad (11.8)$$

其中  $p$  是声压， $\nabla^2$  是微积分的拉普拉斯算子， $\omega$  与频率  $f$  有关， $\omega=2\pi f$ 。

(11.8) 封闭形式的解决方案不存在，除了微不足道的情况。因此，数值计算是通过迭代更新空间上的值来执行的；[208]中对听觉渲染的方法进行了简要的调查。波动方程是在虚拟世界的无障碍部分上定义的。这个空间的边缘变得复杂，导致边界条件。边界的一个或多个部分与声源相对应，声源可被视为振动物体或将能量带到世界的障碍物。在这些位置，(11.8) 中的 0 被替换为强制函数。在其他边界，波可能经历吸收，反射，散射和衍射的某些组合。这些情况极难建模；详情请参见[264]。在一些渲染应用中，这些边界的相互作用可以用简单的 Dirichlet 边界条件和 Neumann 边界条件进行简化和处理[361]。如果虚拟世界是无界的，那么需要额外的 Sommerfeld 辐射条件。有关各种设置中声音传播的详细模型和方程式，请参见[264]。图 11.14 显示了一个计算声场的例子。

### 基于可见性的波传播

数值计算的替代方法是逐步将压力数传播到空间中，这是基于可见性的方法，它考虑从源发出并在障碍物之间反弹的声线路路径。这些方法涉及确定射线与几何模型基元的交点，这类似于 7.1 节的射线追踪操作。

了解虚拟世界中声源的脉冲响应是非常具有远见性的。如果环境被认为是线性滤波器，那么脉冲响应为任何其他声音信号提供一个完整的表征[209,248,258]。图 11.15 显示了矩形房间内反射脉冲响应的简单情况。基于可见性的方法在模拟混响方面特别好，这对于感知原因重现非常重要。更一般地说，基于可见性的方法可以考虑与所有反射，吸收，散射和衍射情况相对应的射线。由于表征所有射线的高计算成本，随机射线追踪通过对射线及其与材料的相互作用进行随机采样提供了一种实用的替代方案[334]。这属于蒙特卡洛方法，例如，这些方法用于近似高维集成和优化问题的解决方案。

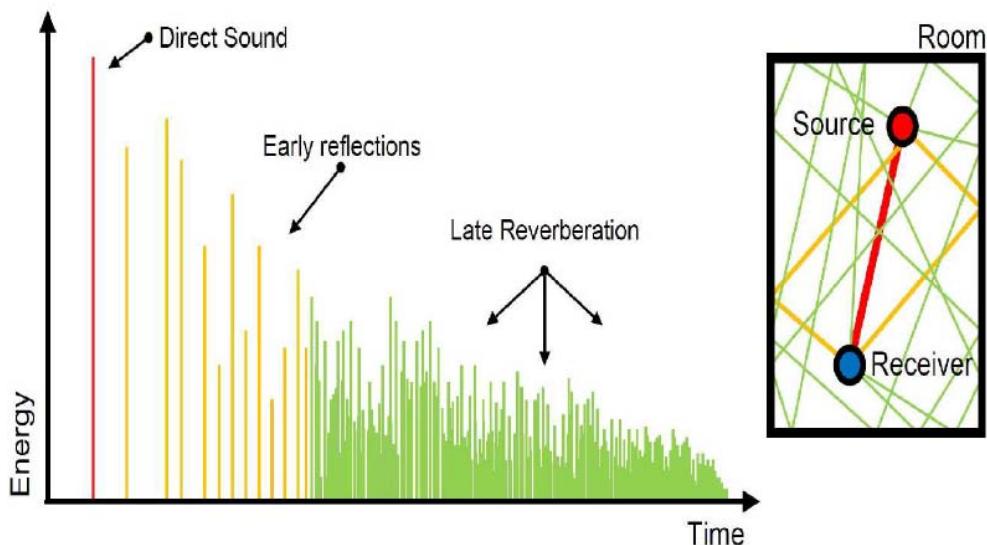


图 11.15：混响。（来自 Pelzer, Aspock, Schroder 和 Vorlander, 2014, [248]）

### 耳朵输入

在虚拟世界中产生的声音必须传输到物理世界中的每个耳朵。仿佛置于虚拟世界中的虚拟麦克风捕捉模拟声波。然后通过位于耳前的扬声器将其转换为音频输出。回想第 11.3 节，

人类能够从听觉线索本地化声源。如果所有的声音都来自固定扬声器，这在 VR 上会如何发生？实际上，可以通过确保每只耳朵接收到合适的声音幅度和相位来提供 ILD 和 ITD 提示，以确定幅度和时间的差异是否正确。这意味着物理头部必须在虚拟世界的某个细节层次上进行复制，以便正确计算这些差异。例如，耳朵之间的距离和头部的大小可能变得尤为重要。

## HRTF

这种解决方案仍然不足以解决混淆的问题。回想第 11.3 节，耳廓的形状会以一种与方向有关的方式扭曲声音。为了充分考虑可能会使输入声音失真的耳廓和头部的其他部分，解决方案是开发一种头部相关传输函数（HRTF）。这个想法是把这种失真看作一个线性滤波器，它可以用传递函数来表征（回忆图 11.12）。这是通过将人体放入消声室并将声源置于头部周围空间的不同位置来完成的。在每个位置，在扬声器上产生脉冲，并且利用放置在人或人体耳道内的小型麦克风记录脉冲响应。通过递增地改变距离，方位角和仰角来选择位置；回忆图 11.10 中的定位坐标。在很多情况下，远场近似是合理的，在这种情况下，距离值较大且固定。这导致 HRTF 仅取决于方位角和仰角。

当然，为每个用户构建一个 HRTF 是不切实际的。使用具有代表性的单个 HRTF 是比较合理的作法；然而，困难在于它在某些应用中可能不够用，因为它不是为个人用户设计的（参见[334]的第 6.3.2 节）。一种妥协方式可能是向用户提供一小部分 HRTF，以解释人群之间的差异，但他们可能无法挑选出最适合其特定的耳廓和头部模型。另一个问题是传递函数可能取决于经常变化的因素，例如戴帽子，穿上带帽或大衣领的外套，或理发。回想一下，适应性几乎发生于整个人类感知和 VR 的所有方面。如果人们适应现实世界中他们头部附近的频繁变化的状况，那么他们也许也会适应并不完美的 HRTF。这个领域仍然存在重要的研究问题。

## 跟踪问题

最后的挑战是确保物理和虚拟耳朵在区域中相匹配。如果用户转动头部，则应该相应地调整声音。如果声音来自固定源，则在转动头部时应该将其视为固定。这是稳定感的另一个例子。因此，需要跟踪耳朵姿势（位置和方向）以确定适当的“视点”。这相当于用右耳和左耳简单的位置和方向偏移进行头部跟踪。可以跟踪头部朝向，每个耳朵的完整姿态由头部模型确定（回忆图 9.8）。或者，可以跟踪完整的头部姿势，通过偏移变换直接提供每个耳朵的姿态。为了优化性能，用户特定参数可以提供完美匹配：沿着 z 轴从眼睛到耳朵的距离以及耳朵之间的距离。后者类似于 IPD。

## 拓展阅读

有关声学的数学和计算基础，请参见[264, 321]。[204]的[218,362]和第 4 章和第 5 章介绍了生理学和心理声学。本地化在[23]中有详细的介绍。[290]中讨论了混淆的锥体。回声阈值涵盖于[268, 358]。

一些基本的信号处理文本是[11,189]。有关听觉显示的概述，请参阅[335]。[256]提供了从心理物理角度简单地放置音频声源的方法。听觉渲染在书[334]中有详细介绍。一些关于听觉渲染的重要文章包括[87,209,248,257,258]

## 第十二章 VR 系统和体验评估

哪个头戴式设备更好？哪些 VR 体验在长时间内更舒适？多少体验的视野是足够大的？什么是最合适的交互机制？工程师和开发人员想知道这些问题的答案；然而，需要明确的是，由于人类生理学和感知在工程系统中操作和交互的方式，这些问题还难以回答。相比之下，纯粹的工程问题，比如“电池寿命估计值是多少？”或“车辆在平地上的最高速度是多少？”，这些问题更容易理解和解答。

回想第 1.1 节涉及生物体的 VR 定义。当科学家应用 VR 来研究老鼠的神经结构和感知时，老鼠和科学家之间是有明确界限的。但是，对于 VR 来说，开发人员会经常尝试自己的想法。在这种情况下，开发人员应在科学家和老鼠的角色之间切换。这就带来了很多问题，尤其是当开发人员对感知问题还不够熟悉的情况下。

类似的适应问题进一步复杂化，会发生在各个方面。例如，几个星期内多次评估 VR 体验的人可能在最初觉得不舒服，但后来会习惯。当然，这并不意味着使新用户患病的可能性变低。此外，人与人之间的差异也很大。任何一个人，包括开发人员，都只提供一个或一组数据。对晕眩免疫的人在使用这种系统时则很有可能没有任何阻碍。

另一个因素是，大多数搭建系统的人都更偏爱他们自己创造的东西。此外，正如 8.4 节所讨论的那样，只要知道体验的具体内容就会影响到最终的观点。这些几十年来一直在研究的问题统称为人为因素。一个与之密切相关的领域是人机交互 (HCI)，它也使用本节讨论的方法。然而，由于 VR 通过破坏我们日常所信赖的感官系统的低级操作来工作，所以由 VR 产生的最低级别副作用到最高级别认知效果的并发症似乎是前所未有的。

从硬件，低级软件到内容创建引擎的各个层面，都存在结果不佳的可能性。随着硬件和低级软件的迅速发展，引擎和 VR 体验的开发人员的负担也越来越重。本章将介绍几个有助于工程师和开发人员寻求建立更好的 VR 系统和体验的主题。第 12.1 节介绍了指导他们提高区分能力的方法。开发者可以培养自己对问题的敏感性，而不是因为适应性而忽略一些重要问题。第 12.2 节应用本书的基础知识为 VR 开发人员提供简单的建议。第 12.3 节介绍了 VR 晕眩，包括主要症状和原因，以便改善 VR 系统和体验。第 12.4 节介绍了设计关于人类受试者实验的一般方法，并包括一些来自心理物理学的具体方法。本章中的所有概念都是为了取得实质性的反馈，并避免在 VR 开发过程中掉入陷阱。

### 12.1 感知训练

大部分第一次尝试 VR 的人都不会知道经验丰富的工程师和开发人员会有很明显的技术上的缺陷。虚拟现实体验的最佳情况应该如此：用户被主导的视觉刺激所淹没，并感觉自己仿佛置身于虚拟世界。可以存在一些微小的缺陷，因为注意力主要集中在目标体验上（如 1.1 节中 VR 的定义所考虑的那样）。有些部分可能没有按照设计运行，或者一些感知问题可能被忽略。但这样的问题也可以被改善，并产生更好的体验。更糟的情况是，这些缺陷可能会导致用户疲劳或生病。最后，这些用户也无法意识到产生了什么问题。他们可能会追究到任何一点，比如特定的视觉刺激，特殊的体验，耳机硬件，甚至 VR 的整体概念。

通过培训特定的用户和开发人员来关注常见的缺陷类型，可以一定程度上缓解此问题。通过开发感知训练计划，可以要求用户寻找特定的缺点，或者反复练习来执行一些任务。在整本书中，我们已经看到适应在人类感知过程中的重要性。例如，如果长时间产生持续刺激，则感知强度会减弱。

通过对特定的 VR 系统和体验进行反复和引导性的训练，用户可以调整他们的感知系统。这是感知学习的一种形式，它是感知心理学的一个分支，它研究生物体因环境而引起的感知系统的长期变化。由于虚拟现实成为一种新环境，学习的机会和局限性在很大程度上仍然未被探索透彻。通过主动训练，可以控制用户适应的方式，从而增加他们的感知能力和辨别能力。这也可用于在开发过程中提供频繁反馈的评估人员。另一种方法是开发一个可以在没有人为干预的情况下自动检测缺陷的系统。人为和自动评估的结合在未来几年中很可能非常重要。

### 感知学习的例子

在日常生活中，我们会遇到许多关于感官学习的例子。关于视力，医生和医疗技术人员接受培训，从图像中提取相关信息，这些信息对未经训练的人来说非常难以捕捉。癌症专家可以在 CT 和 MRI 扫描中发现肿瘤。产科医生可以毫不费力地从手持式超声扫描仪中确定胎儿结构是否正常发育。关于听力，音乐家在广泛的练习之后学会区分和分类各种音符。由于录音，压缩，扬声器和室内声学问题，发烧音乐者会注意到音乐再现中的特殊缺陷。关于味道和嗅觉，侍酒师学习辨别葡萄酒之间的微妙差异。关于触摸，盲人学会阅读盲文，这是用指尖感觉到的凸起点的微小图案完成的。所有这些例子对于新手来说似乎都是不可能的，因为看起来我们甚至没有完成它的神经条件。尽管如此，通过建立感知训练计划和反复练习，人们可以获得惊人的感知能力。所以，为什么不可以对 VR 进行评估呢？

### 感知学习因素和机制

当这些形式的学习发生时，人类的观察系统会发生什么？一个重要的因素是神经可塑性，它使人类的大脑能够开发专门的神经结构作为对环境刺激的适应。虽然这对小孩子来说要强得多，正如学习母语所展示的那样，但神经可塑性仍然在成年人中存在；人与人之间最终的神经可塑性结果可能不同。

另一个因素是学习的方式。随意观察或专注于刺激的针对性策略均可能会导致适应。学习发生的时间和重复性可能会有很大不同，具体取决于任务，性能要求，刺激和人员。此外，还有可能会给他/她提供有监督的培训，在他/她试图提高自身表现时直接提供反馈。又或者，可能会出现无监督的训练，训练员已经在学习环境中放置了足够的刺激，但不会干扰学习过程。

四种基本机制已经被开发出来解释感知学习。

1. **注意力权衡：**与任务相关的功能的注意力增加，同时减少对其他功能的注意力。
2. **刺激印迹：**开发专门的受体来识别部分或全部相关刺激。这些可能是神经结构的功能。
3. **差异化：**曾经融合在一起的不同刺激共同激活了知觉。因而细微的差异需要被放大。
4. **单元化：**这个过程将许多不同的刺激结合或压缩成一个响应。这与差异化形成鲜明对比，且因与单元内部差异关联不大，对分类很有帮助。



图 12.1：蝴蝶出现在左眼的图像中，但相应右眼图像中没有蝴蝶。

本节的其余部分将在 VR 的背景下提供示例和有用的建议。该领域远离标准的感知训练课程，类似于医学影像或音乐训练。相反，我们提供关于如何移动以及在何处关注焦点的建议，同时尝试发现 VR 体验中的错误认知。这就要求人们继续意识到人为刺激造成的干扰，这与第 1.1 节所述的 VR 的定义并不完全相符。

### 立体声问题

图 12.1 显示了一个简单的错误，其中蝴蝶出现在场景中，而另一个中并没有。虚拟世界的其余部分被正确渲染。对于未经训练的人眼，这可能完全不会被注意到。解决方案则是：闭左眼，同时保持右眼张开；之后，切换到左眼张开，右眼闭合。通过在打开一只眼睛之间来回切换，不匹配现象应该变得清晰。这种方法被称为闭眼法。

另一个常见错误是将右眼和左眼图像颠倒过来。（3.50）中出现符号错误后，很容易出现这个问题，或者误解了视点需要为每只眼睛转换的方式。这种现象被称为视觉幻觉，其中物体的感知凹面可能被颠倒过来。但是，在很多情况下，这种现象很难直观地检测到。解决方案：接近物体的边缘，使其一侧仅对一只眼睛可见。这可以通过使用闭眼法来验证。根据物体的几何形状，确保侧面被正确可见。

最后，立体视觉可能在虚拟瞳孔之间具有不正确的距离（（3.50）中的  $t$  参数）。如果  $t=0$ ，则可以使用闭眼技巧来检测两个图像看起来是相同的。如果  $t$  太大或太小，则深度和尺度感知（第 6.1 节）会受到影响。更大的分离  $t$  会导致整个环境变得更小；较小的  $t$  会导致相反的结果。

### 典型的头部运动

现在考虑涉及到移动的错误，这可能是由于头部跟踪错误，渲染角度或某种组合引起的。做出谨慎的，可重复的运动是有益的，这被称为典型的头部运动。如果单独跟踪旋转，则有三个旋转自由度。要发现各种运动或视点错误，应对评估者进行培训，以仔细执行各自的基本旋转。纯偏航可以通过“否”来执行。纯俯仰表现为“是”。纯翻滚状态则更难以完成，这涉及到来回转动头部以使一只眼睛在极端情况下高于另一只眼睛。在任何这些运动中，尽可能少地转换环视点（双眼中心点）或采用 9.1 节头部模型引起的转换是有帮助的。

对于每一种基本旋转，评估人员都应该练习以各种恒定的角速度和幅度来执行它们。例如，评估人应该试着以恒定的速度慢慢地摇头，每个方向 45°。或者，她应该尝试以很快的速度旋转，每个方向可以达到 10 度，频率为 2 Hz。使用规范的头部运动，可以确定图 9.7 中给出的常见错误。其他问题，例如跟踪中的不连续性，倾斜误差，等待时间以及视点的不正确深度也可以通过这种方式被更容易地检测到。

如果对位置进行跟踪，那么三种甚至更多典型的头部运动会变得重要，每个位置都会有一个 DOF。因此，可以执行水平，垂直和深度变化运动来辨别存在的问题。例如，对于水平的左右运动，可以确定运动视差是否提供了正确的提示。

### VOR vs 平滑移动

回顾 5.3 节，5.4 节和 6.2 节，眼动在视觉感知中起着重要作用。评估人员在评估虚拟世界中的物体是否应该是静止时，应考虑特定的眼球运动模式。如果在眼睛注视对象时进行典型的偏航运动，则利用到前庭眼球反射（VOR）进行跟踪。在这种情况下，评估者可以确定物体图像在视网膜上固定时是否出现移动或扭曲的现象。同样，如果一个物体缓慢移动并且头部固定，则评估者执行平滑追踪以将物体保持在视网膜上。如第 5.4 节所述，物体逐行扫描显示出现扭曲的样式取决于运动是由于 VOR 还是由于平滑追踪。如果物体移动得非常快并且眼睛不能将其固定在视网膜上，那么可能会感觉到拉链效应。

## 外设问题

目前的一代 VR 头设具有明显的光学像差问题；根据 4.3 节，随着离光轴的距离增加，这会变得更加糟糕。区分以下两种情况很重要：1) 在检测周边变形的同时通过镜头中心观看，以及 2) 旋转眼睛直接观看镜头边缘。在第一种情况下，由于外围光感受器密度较低，失真可能不太明显；然而，不匹配可能会对舒适度和晕眩产生影响。光流信号在外围较强，而不匹配值可能被视为不正确的运动。

在第二种情况下，直接通过镜片观察可能会发现由球面像差引起的边缘缺乏焦点的现象。此外，色差可能变得可见，特别是对于黑色背景下的清晰的白线。此外，随着直线看起来变弯曲，枕形失真校正中的误差可能变得更明显。这些问题不能通过单个失真校正函数来修复（如 7.3 节所述），因为当眼睛旋转时，瞳孔会转换离开光轴。每个眼睛方向都需要一个不同的不对称校正函数，这就需要眼部跟踪来确定在各个时间点使用哪个校正函数。

为了观察枕形或桶形变形，评估人员应当在眼睛注视在物体上的同时，尽可能大幅度地施加典型的偏航运动。在这种情况下，VOR 将导致眼睛在大范围内旋转，同时从一侧到另一侧地扫过镜头，如图 12.2 所示。如果虚拟世界包含具有显着纹理或空间频率的大而平坦的墙壁，那么当运动期间墙壁看起来像在“呼吸”时，变形便清晰可见。如果墙壁上有规则的网格图案，效果可能会更明显。

最后，许多用户甚至没有注意到镜头的有限视场。回想一下第 5.4 节，放在眼前的任何平面屏幕只覆盖眼睛的一部分视野。因此，外围的感光器将不会接收到来自显示器的任何光线。在大多数情况下，它在 HMD 内部很暗，导致在显示器的可见部分周围出现黑带。一旦向用户看向这部分，就很难忽视掉这种现象。

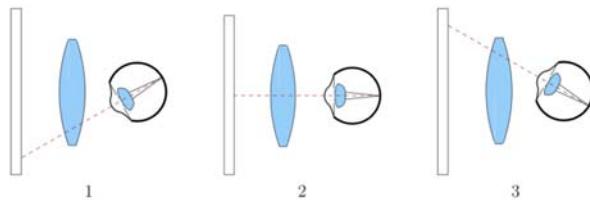


图 12.2：自上而下的视图，显示了当虚拟世界中的一个静止物体被眼睛注视时眼睛如何旋转，同时头部逆时针方向偏转（面向右侧）。周边的镜头扭曲干扰了对平稳性的看法。

## 延迟感知

人们对延迟的直接感受各不相同。即使不能察觉，它也是 VR 晕眩的主要原因之一[171]。适应性在其中产生了巨大的难题，因为人们可以通过长时间训练来适应固定的潜伏期；但在这种情况下回归到现实世界可能会很困难。一段时间内，现实世界中的大部分物体可能都不是静止的！

在我自己的 Oculus VR 工作中，当我在 2012 年开始使用 Oculus Rift 原型时，我可以检测到大约 40 毫秒的延迟。到 2014 年，通过以下过程，我可以把检测延迟降低至 2 毫秒。第一步是面对虚拟世界中的垂直边缘，例如门框。评估人员应保持舒适的距离，如两米。当视点固定在边缘上时，典型的偏航运动应该以非常低的幅度（例如几度）和大约 2Hz 的频率来执行。运动的幅度和频率很重要。如果幅度太大，则光学失真可能会产生干扰。如果速度太高，那么 HMD 可能会开始相对于头部翻转。如果速度太低，延迟可能不容易察觉。执行此动作时，如果存在严重的延迟，则边缘会与头部移动不同步。

回想一下，现在许多 VR 系统可以实现零有效延迟，正如第 7.4 节所述；尽管如此，由于硬件，软件和 VR 内容的特定组合，在许多系统上仍可能会出现可感知的延迟。通过使用预测，甚至可以获得负的有效延迟。配合增加或减少预测间隔的按键，我可以通过应用上述

方法将有效延迟调整到 2 毫秒。该方法与心理物理学调整方法密切相关，将后面的 12.4 节中介绍。之后我能够立即将延迟时间降至 10 毫秒，而无需进行任何其他调整或比较。虽然这不是一个科学的结论（参见第 12.4 节），但看起来在花费近两年时间调试 Oculus VR 的跟踪和渲染系统以将有效延迟降至零后，我经历了一次感知学习过程。

## 结论

本节提供了一些培训人员发现 VR 系统问题的建议。预计未来还会出现更多。例如，为了评估虚拟世界中的听觉定位，评估人员应该闭上眼睛，以标准动作移动头部。为了在使用菲涅耳透镜的系统中检测透镜眩光，他们应该寻找在黑暗背景下由亮光形成的图案。为了检测显示闪烁（从 6.2 节回忆），特别是低至 60Hz 时，评估人员应进入一个明亮的虚拟世界，最好是白色的，并放松眼睛，直至感到明显外围振动。为了发生视觉辐辏调节冲突（回忆 5.4 节）的相关问题，可以将虚拟物体放置得非常靠近眼睛。

此外，还需要有正式的培训机制或课程，工程师和开发人员可以使用这些机制或课程来提高他们的洞察力。在这种情况下，评估人员可以通过反复练习来提高他们的技能。想象一下把 VR 体验看作一种竞争性游戏，旨在增强发现虚拟现实缺陷的洞察力。

## 12.2 针对开发人员的建议

随着虚拟现实耳机的广泛应用和可承受性，近年来开发虚拟现实体验的人数急剧增长。迄今为止，大多数开发人员都来自视频游戏行业，他们在开发游戏和游戏引擎方面的技能和经验被“移植”到 VR。在某些情况下，简单的改编就足够，但是游戏开发者一直对一个非常成功和流行的游戏无法直接转换为舒适，甚至是有趣的 VR 体验感到意外和困惑。大部分困惑都是由于缺乏对人类生理和感知的理解。随着该领域的发展，开发人员来自于各种不同的背景，包括电影，广播，通信，社交网络，可视化工程等。艺术家和爱好者也加入并创造了一些最有创意的经历。

本节将提供一些有用的建议，这些建议基于本书所涵盖原则的组合以及其他开发人员指南（特别是[359]）的建议。这份建议当然还不完全，随着新的硬件和经验的发展，未来几年会有所增长。到目前为止，绝大多数 VR 体验都是基于成功的 3D 视频游戏。下面的大部分建议都与本书的前几部分有关，这些部分提供了科学的目的或进一步的解释。

### 虚拟世界

- 在虚拟世界中设置与现实世界相匹配的单位，以便更简单地匹配两者比例。例如，一个单位等于虚拟世界中的一米 这有助于深度和场景感知（第 6.1 节）。
- 确保对象完全建模，以便在用户查看屏幕上图形出现意想不到的情况时，缺少现象不再明显。
  - 由于视点不同，非常薄的物体（如树上的树叶）可能在 VR 中看起来不真实。
  - 设计好环境，减少运动量；例如，虚拟电梯比虚拟楼梯更舒适（8.4 节和 10.2 节）。
  - 考虑视觉和听觉渲染性能问题，根据需要简化几何模型，以便在目标硬件上保持适当的帧速率（7.4 节和 11.4 节）。

### 视觉渲染

- 左视图和右视图之间的唯一区别是视点，而不是模型，纹理，颜色等（3.5 节和 12.1 节）。
- 切勿将文字，对象或图像固定到屏幕的一部分；所有内容应该嵌入在虚拟世界中。回

想一下 2.1 节，屏幕上的静止不同于在虚拟世界中被认为是静止。

•调整渲染视场或任何影响镜头失真的参数时要谨慎，以免造成进一步的失配（7.3 节和 12.1 节）。

•重新评估常见的图形学技巧，如纹理映射和法线贴图，以确保它们在 VR 中有效，因为用户可以立体观看，并且能够快速更改观点（见第 7.2 节）。

•由于视角和立体视角不同，抗锯齿技术对 VR 来说更为关键（7.2 节）。

•渲染系统应该进行优化，以便所需的虚拟世界可以以至少与硬件要求一样高的帧率进行更新（例如，Oculus Rift 和 HTC Vive 的 90 FPS）；否则，帧率可能会下降并变化，导致不适（7.4 节）。

•避免导致大视点变化的物体移动；否则，用户可能会觉得自己正在移动（8.4 节）。

•确定如何挑选太靠近用户脸部的几何图形的方法；否则，会出现视觉辐辏调节冲突（第 5.4 节）。

•与游戏和摄影不同的是，视点不应以与头部追踪不匹配的方式改变，除非使用者感觉自己好像正在虚拟世界中移动。

•为了获得适当的深度和尺度感知，用户在现实世界中的瞳孔间距离应与虚拟世界中眼睛之间的相应视点距离相匹配（6.1 节）。

•与屏幕上的图形相比，降低模型的亮度和对比度可提高 VR 舒适度。

## 追踪和匹配区域

•切勿让头部追踪停止或延迟；否则，用户可能会立即感受到自我运动（8.4 节）。

•考虑立体声偏移（3.5 节），头部模型（9.1 节）和运动（10.2 节），确保眼睛视点正确定位。

•注意现实世界中存在而虚拟世界中不存在的障碍；当用户靠近障碍物时可能需要警告（见第 8.3.1 节）。

•同样，要小心虚拟世界中存在而现实世界中不存在的障碍。例如，如果用户决定戳穿墙壁（见第 8.3.1 节），可能会产生不可预料的后果。

•随着到达跟踪区域的边缘，逐渐降低对比度和亮度比仅保持位置固定更舒适（8.4 节）。

## 交互性

•通过赋予人类“超能力”而不是应用通用模拟原则的交互机制，来到达比现实更好的效果（第 10 章）。

•至于运动性，请遵循第 10.2 节中的建议以减少运动副作用。

•对于在虚拟世界中的操纵，系统应要求用户在物理世界中尽可能少移动（10.3 节）。

•关于与社会的互动性，由于恐怖谷理论，高现实度并不一定会更好（第 10.4 节）。

## 用户界面

•如果出现浮动菜单，如网络浏览器或其他类型的虚拟显示器，则应将其呈现在距离用户视角至少两米远的地方，以尽量减少视觉辐辏调节冲突现象的发生（第 5.4 节）。

•这种虚拟显示器应居中，并具有相对较窄的视野，大约为总观看区域的三分之一，以最大限度地减少眼睛和头部的移动。（第 5.3 节）。

•如果将菜单、选项、游戏状态和其他信息嵌入虚拟世界，则应以人们熟悉的方式嵌入其中；这遵循通用模拟原理（第 10 章）。

## 音频

- 注意外部扬声器与耳机之间的区别；如果没有跟踪，声源本地化将无法正确实现（第 2.1 节）。
- 运动的位置和方向跟踪都应该考虑到听觉化（第 11.4 节）。
- 多普勒效应提供了强大的运动信息提示（第 11.1 节）。
- 与视觉渲染相比，音频实现过程的几何模型可以大大简化；0.5 米的空间分辨率通常来讲就已经足够（第 11.4 节）。

## 用户自我外观

- 如果用户能够看到他/她的身体，则其在虚拟世界中的感觉以及判断能力会得到增强。
- 简单的虚拟身体比没有虚拟身体要好得多。
- 虚拟身体和真实身体之间的差异可能令人担忧。他们可能有不同的性别，体型或物种。这可能会导致用户得到前所未有的新鲜体验，但反之可能会发生不专心的现象。
- 如果只执行头部追踪，那么虚拟身体应该满足一些基本的运动学约束，否则会让用户在虚拟世界中感到绝望（第 9.4 节）。
- 用户的自我表现会影响他们的社交行为，以及他们身边的人对他们的反应（10.4 节）。

### 12.3 舒适度和虚拟现实病

作为使用 VR 系统的一个副作用，体验不适是过去几十年来该技术广泛采用的一个最大的威胁。它被认为是其未能在 20 世纪 90 年代初期达到预期的主要原因。很少有人需要此技术，因为 VR 被使用时往往会导致人们的身体受到损害，且这种现象会在很多情况下并且长时间的发生。由于许多因素，例如用户之间的差异，重复使用的适应性，测量症状的困难程度，快速变化的技术以及依赖于内容的敏感度等因素，研究人员对此也感到沮丧。显示器，传感器和计算机技术的进步已经导致硬件造成的副作用减少。随着硬件产生的副作用减少，相应的负担一直在向软件工程师和内容开发人员转移。发生这种情况是因为 VR 体验本身有可能让人感到恶心，即使硬件被认为是非常舒适的。事实上，最好的 HMD 可能比以往任何时候都更让人感到恶心！出于这些原因，VR 系统的工程师和开发人员必须了解这些副作用产生的原因，以便他们确定如何为绝大多数用户减少或消除这些副作用。

#### 晕眩或综合症

在本书中，我们讨论到 VR 晕眩中任何意想不到的副作用。这包括很多通常与晕眩无关的症状，例如疲劳。因此，更准确的短语可能是虚拟现实适应不良综合症，其中适应不良是有害而不是有益的，综合征是指与活动相关的一组症状。

#### 晕动症变种

了解与 VR 晕眩密切相关的术语是有帮助的，因为它们与类似的活动，症状集和潜在原因相关。这有助于进行相关的研究。最广泛的相关领域是晕动症，与真实世界中的运动症状有关。其通常涉及前庭器官（8.2 节），这意味着它们涉及感觉输入或加速冲突；事实上，没有功能性前庭器官的人不会遇到晕动的现象[145]。由于真实运动引起的运动症是由于经历了异常的力量而发生的。这可能是由于身体的运动，导致头晕和恶心。同样，这些症状是由可能产生极端力量的车辆运输而发生的。此外，各种游乐设施可以产生更极端的体验和副作用。

不幸的是，晕车远远超出娱乐范围，因为许多人在乘坐专为运输而设计的车辆时仍患有

晕车病。人们分别从汽车，船只和飞机中经历晕车，晕船和晕机症状。据估计，只有约 10% 的人在运输过程中从未经历过明显的恶心[171]。由于士兵们花费了长时间进行海上巡逻和高速战斗机飞行，军队进行了最大的晕动病研究。大约 70% 的海军人员会遇到晕船问题，其中约 80% 的士兵会降低其工作效率或动力[249]。最后，一个非常规的例子是太空旅行，其中遇到微重力的宇航员也会有恶心或其他症状，这就是所谓的太空病。

### 视觉诱发晕动症

迄今为止的晕动症都涉及到真实的动作。相反，运动病可能因为刺激反应过激而发生，这些刺激会说服大脑发生加速，尽管事实并非如此。这就是所谓的视觉运动。最常见的研究案例是视觉诱发的视觉运动，也称为视觉运动，见 8.4 和 10.2 节。与此相关的症状是视觉诱发晕动症的一部分。

媒介过程可以通过多种方式产生。回想一下 2.3 节的图 2.20，极端情况是由于人们在摇摆的房间引起。科学家们通过运动视觉刺激来围绕受试者进行运动视觉和感知运动病的对照实验。在涉及特定移动视野的各种研究中，只有少数受试者不受副作用的影响。大约 50% 至 100% 的人会出现头晕，约 20% 至 60% 的人会出现胃部症状；更确切的情况取决于具体的实验结果[171]。

此外，可以使用显示器来产生动感。回想一下第 6.2 节，由于帧的快速连续性，在这种情况下感知的光流是频闪表观运动。然而，使用视功力鼓的晕眩研究仍然有重要的参照意义。它们揭示了视觉诱发晕动症的严重程度，即使在没有数字伪像（如显示分辨率和帧率）的限制下也是如此。

### 模拟器晕眩和网络病

一旦使用显示器，2.1 节中设计的选择问题则又在此重现：定屏幕（如 CAVE VR 系统）还是需要跟踪的头戴式显示器。汽车模拟器也许是 VR 的第一个重要应用，最常见的例子是驾驶汽车，驾驶飞机或直升机。用户可以坐在固定的基座上，也可以坐在一个机动的基座上对应控制。后一种情况提供了前庭刺激，运动和视觉信息的时间同步，这对于减轻晕眩至关重要。通常，整个驾驶舱都在现实世界中重建，视觉刺激主要出现在窗户外。头部可以被追踪以提供立体视觉和不同的视点，但是大多数情况下并不会这样做，主要是为了舒适度最大化并且技术副作用被最小化。这种活动导致的视觉诱发晕动的分支症状地称为模拟病，美国军方已经对此进行了充分研究。

术语 cybersickness [206] 被提议用于涵盖与 VR（或虚拟环境）相关的任何晕眩，其中包括模拟器晕眩。不幸的是，这个词的含义近来已经扩大到包括花费太多时间与智能手机或电脑进行交互导致的晕眩。此外，网络这个术语在过去的几十年中已经积累了许多奇怪的内涵。因此，我们将视觉诱发晕动病以及由 VR 引起的任何其他形式的不舒适症状共称为 VR 晕眩。

### VR 病的常见症状

目前而言，已有各种术语用于指向 VR 运动和 VR 晕眩研究中的症状。最常见的是（基于[143,145,169,171,305]）：

- **恶心：**症状开始时，使用者可能有与胃，上腹部，食道或喉咙有关的不适感。随着强度的增加，它逐渐导致呕吐感。这是 VR 晕眩最消极和最吓人的症状。
- **头晕：**即使在刺激消除后，用户可能会感觉到运动的感觉，如旋转或摇摆。这也可能包括眩晕，它与前庭器官功能障碍相似且经常发生。
- **嗜睡：**用户可能会变得不那么警觉，打哈欠，并最终开始入睡。

- 流涎增加：口中唾液量增加，导致通常吞咽量增加。
- 冷汗：用户开始出汗或增加汗水，但不会因环境温度升高而出汗。
- 苍白：用户可能在脸上、耳朵、脖子和胸部都会出现皮肤苍白甚至脱落的现象。
- 脸色潮红：这与感觉到的温暖突然增加相似，类似于发烧。
- 头痛：用户头痛，可能会逐渐增加强度，并在使用后很长时间持续头痛。
- 疲劳：长时间使用后，用户可能会感到疲倦，筋疲力尽。
- 眼睛疲劳：用户可能会觉得他们的眼睛疲劳甚至疼痛。
- 聚焦问题：用户可能视力模糊或难以聚焦。

以上是部分体验者为某些 VR 测试所报告的症状。研究这些症状的目标是让 VR 系统和体验尽可能地消除这些症状。此外，通过重复训练和适应，大多数症状发生的可能性会被大大降低。

## 其他副作用

除了刚刚列出的直接症状外，其他一些现象也与运动和 VR 晕眩密切相关，并且会持续很长时间。其中一个是 Sopite 综合征[102]，与嗜睡密切相关，但可能包括其他症状，如懒惰，缺乏社会参与，情绪变化，冷漠和睡眠障碍。即使在适应上述系统后，这些症状仍可能持续存在。另一个现象是姿势不平衡，对平衡和协调产生不利影响[171]。最后，另一种现象是在头部或身体运动过程中丧失视力[171]，这似乎是 VOR (5.3 节) 适应虚拟现实系统缺陷的自然结果。这是因为尽管存在分辨率，等待时间，帧速率，光学畸变等问题，但仍然会对平稳性的感知造成影响。

## 体验后效应

VR 晕眩最令人担忧的一个方面是症状在使用 VR 后可能持续数小时甚至数天[304]。大多数在从 VR 体验中退出后立即出现症状的用户仍然会在后续的一段时间内表现出一些晕眩。有极少数的用户可能会持续数小时或数天出现症状。同样，一些遇过海难的人在返回地面后长期患有土地晕眩症。这对应于视觉世界的姿势不稳定性和感知不稳定性；对具有这种症状的人来讲，世界似乎在摇摆[171]。

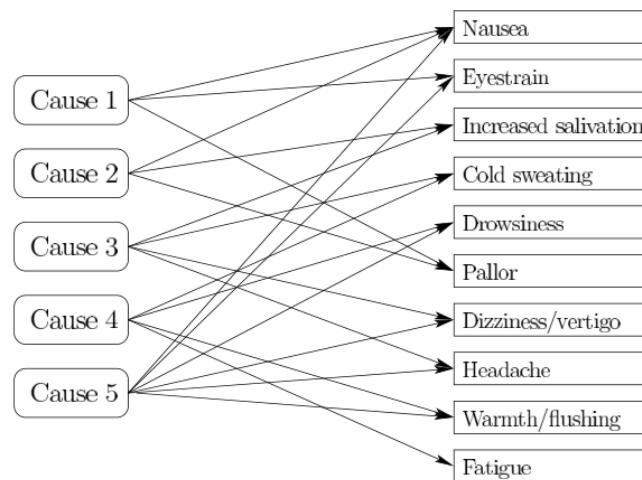


图 12.3：症状可以观察到，但原因并不能直接被测量到。研究人员面临逆向问题，根据观察到的症状推测原因。但目前的问题在于每种症状可能有很多可能的原因，其中一些可能与 VR 体验无关。

## 从症状到原因

症状是最终的结果，但它们的原因是什么？见图 12.3。VR 系统的科学家或评估者目前面临的问题在于只有症状可观察到。任何症状都有任何可能的原因。其中一些原因是已知的，而其他可能无法确定。例如，假设用户在 VR 体验 5 分钟后出现了轻微的恶心。除了因为他骑着自行车参加考试并忘了吃早餐，他还会有什么产生恶心的可能性呢？如果他前一天晚上喝醉怎么办？也许像这样的一些用户可能被作为离群值抛弃，但是如果前一天晚上举行了一个大型节日，在实验之前疲劳了的测试者应该怎么对待呢？其中一些问题可以通过将一些用户分进预计变异性较小的组来处理；见第 12.4 节。如果他们感到恶心，至少应该事先询问他们；然而，这样的行为可能导致他们更加关注虚拟内容而恶心。

通过在接受 VR 体验之前，期间和之后获得的数据中找到的统计相关性，可以回答有关 VR 痘原因的问题。原因的分析过程并不是通过直接见证其原因及其效果，而是通过观察明显的效果而确定的。消除不相关的原因是实验设计的重要组成部分，其中涉及仔细选择用户的过程并提前收集适当的数据。确定更具体的原因需要更多的试验性实验。这种情况很复杂，因为不同的试验不能轻易应用于同一用户。一旦人们生病了，他们将无法参与，或者至少会给出难以弥补的偏差结果。

## 用户之间的差异

另一个复杂因素是人们对 VR 晕眩易感性的变化性。在实验设计中必须考虑到群体间的个体差异；参见第 12.4 节。大多数研究人员认为女性比男性更容易患晕动病[143,241]；然而，这个结论在[171]中有争议。关于年龄，12 岁以下儿童的易感性似乎最高，然后随着年龄增长而迅速下降，同时在其一生中逐渐下降[262]。一项研究甚至认为中国人比其他一些组别更容易受到影响[309]。预测个体易晕车的最佳预测因素是确定他/她是否曾患有晕车症。最后，要注意的是，各组之间也可能存在差异，如症状的严重程度，发病的速度，实验后的持续时间以及用户适应 VR 的速度。

## 感知冲突理论

除了根据违规刺激来确定因果关系之外，我们还应该尝试理解人体为什么因 VR 产生负面影响。该过程涉及哪些生理和心理机制？为什么一个人不能快速适应某些刺激，而其他人很好？什么是特别不好的刺激？虽然目前没有广泛包含和公认的解释晕动病或 VR 晕眩的理论，但存在一些有用和可接受的理论。

理解 VR 晕眩必须具备的相关理论和强大理论之一是感知冲突理论[132,145]。回想一下第 2.1 节图 2.1 中 VR 系统的高级描述，VR 中存在两种不匹配：

1.与自然刺激相比，工程刺激与中枢神经系统和大脑预期的刺激程度并不相近。示例包括由于显示分辨率，帧率，光学失真，有限色彩，合成照明模型和延迟而引起的伪影。

2.一些感官系统不接受工程刺激。它们继续以自然的方式感知周围的物理世界，并相应地发送神经信号。例子包括来自前庭和本体感受系统的信号。前庭器官继续感知到真实世界的加速度，并且可以从运动信号粗略地估计身体部位的姿势。

毫不奇怪的是，VR 中最重要的冲突涉及到加速。人类视觉系统提供与运动一致的光流读数，但来自前庭器官的信号与之不匹配。请注意，这是一种常见的晕动病现象，它在现实中往往出现在行驶中的车辆中行驶而人们不向外看的情况。例如，想象一下当一名乘客在汽车里读书时的情况。在这种情况下，前庭系统报告汽车的加速度，但没有相应的光流量。

## 强制融合和疲劳

回想一下第 6.4 节，我们的感知系统整合了来自不同来源，不同感知模式的线索，以获

得连贯的感知解释。在线索之间存在微小差异的情况下，由此产生的解释可以被认为是强制融合[120]，尽管其中会有偏差，感知系统似乎难以形成匹配。执行传感器融合或视觉场景解释的工程系统下的情况与之类似；在存在更多噪音或不完整信息的情况下，优化或寻找合理解释的可能性会更大。强迫融合似乎直接导致疲劳和眼睛疲劳。因此，尽管存在工程上的缺陷，但人类在学习解释虚拟世界时被迫加大机体运行力度。幸运的是，反复的虚拟现实体验会导致感知系统的学习或适应，最终可能会减少疲劳。

### 毒药假说

感觉冲突似乎足以解释为何会产生额外负担，但似乎并不意味着会导致恶心。科学家们正在调查什么是造成这一现象和相关症状的原因。请注意，人类有能力自然地使自己摆脱不涉及技术的旋转运动。间接毒物假说认为与晕动病相关的恶心是一种在人体内进化机制带来的副产物，导致人们会呕吐意外摄入的毒素[323]。这种毒素频率的症状包括视觉和前庭线索之间的冲突。另一种解释是直接毒物假说，该假说认为恶心与毒素有关，因为它们在整个进化过程中与涉及增加或延长加速度的活动有关。[171]的第 23.9 节给出了这些替代假说及其不完整的详细评估。

### 虚拟现实晕眩的级别

为了改善虚拟现实系统和体验，我们必须对副作用进行适当的比较。因此，所产生的症状需要量化。对于每种症状，并不是简单的是/否回应，而是获得与相对严重程度对应的数值。对于一个特定的症状，几个重要的参量包括：

- 症状的强度。
- 刺激呈现时，症状发作或强度增加的速度。
- 刺激消除后，症状衰减或强度下降的速度。
- 在固定级别或更高级别遇到症状的用户的百分比。

前三个可以被视为随时间变化的强度图。最后一个统计属性；许多其他统计数据可以从原始数据中计算出来。

### 问卷调查

收集定量数据最常用的方法是让用户填写问卷。研究人员多年来设计了许多问卷[170]；最广为人知的是模拟病人问卷（SSQ）[144]。它被设计用于美国军方的模拟病症研究，已被广泛使用。其要求用户以四个等级评分 16 个标准症状中的每一个：0 无，1 轻微，2 中等和 3 严重。通常通过总计分数来汇总结果。为了确定起病率或衰退率，必须多次给予 SSQ，例如实验之前，实验后，10 分钟后，30 分钟后，甚至 60 分钟后。

但问卷调查主要有四个缺点。首先是答案是主观的。没有明确的方法来校准在用户感觉在等级“1”与等级“2”之间恶心的含义。单个用户甚至可能根据情绪或甚至其他症状的发作给出不同的评分。第二个缺点是要求用户注意他们的症状，这可能会影响他们的发病倾向。第三个缺点是用户必须中断，以便他们可以在会话期间提供分数。最后的缺点是，随着时间的推移，强度必须粗略地采样，因为每次感兴趣的时刻都必须填写新的问卷。

### 生理测量

另一种方法是将传感器连接到用户，以便在 VR 体验之前，之中和之后自动获取生理测量结果。这些数据可以在不中断用户或要求他们注意症状的情况下持续获得。然而，在身体上放置传感器可能会有一些不适或恐惧。研究人员通常会购买标准传感系统，如 Biopac MP150，其中包含一包传感器，记录数据并将其传输到计算机进行分析。

已经用于研究 VR 晕眩的一些生理学措施是：

- 心电图 (ECG): 该传感器通过将电极放置在皮肤上来记录心脏的电活动。VR 体验期间心率通常会增加。

- 胃电图 (EGG): 这与心电图相似，但是电极放置在胃部附近，以便估计肠胃的不适情况。

- 眼电图 (EOG): 将电极放在眼睛周围，以便估计眼球运动。或者，可以使用基于相机的眼动追踪系统（第 9.4 节）以确定眼旋转和眨眼率。

- 光电容积描记图 (PPG): 这提供了关于心脏运动的附加数据，并通过使用脉搏血氧仪获得。通常这种设备夹在指尖上来监测血液的氧饱和度。

- 电流皮肤反应 (GSR): 该传感器测量整个皮肤表面的电阻。当人流汗时，皮肤表面的水分会增加导电性。这提供了一种测量冷汗的方法。

- 呼吸运作: 呼吸频率和振幅是从胸腔上的一个贴片测量的，它对压差或膨胀产生反应。VR 体验期间呼吸速率可能会增加。

- 皮肤苍白: 可以使用相机和图像处理来测量。在最简单的情况下，IR LED 和光电二极管可用作测量皮肤反射率的发射器-检测器对。

- 头部运动: 头部跟踪系统是运动数据的丰富来源，它可以帮助估计疲劳或姿势不稳定性，而无需额外的成本或分散用户的注意力。

近期，[59]对生理测量和问卷调查的方法进行了比较，甚至得出了结论：可以从独立的生理数据中确定一个人是否正在经历 VR。

### 减少晕眩不适的策略

通过实验研究来确定用户的 VR 晕眩频率和强度，工程师和开发人员可以迭代并生成更舒适的 VR 体验。回顾之前讲过的“平稳感”的内容，我们发现在各个环节都需要改进。大多数现实世界都被认为是静止的，它应该和虚拟世界一样。视觉显示，渲染和跟踪方面的改进应该通过提高对平稳性的认识来帮助减轻晕眩不适，还应该减少或消除光学失真，混叠，延迟和其他伪影。当它们不能被消除时，应该找到适当的折衷方案。此外，采用新的显示技术，以降低视觉辐辏调节冲突的问题，当传统屏幕和镜头组合的头戴式设备上出现紧密物体时会引起严重不适（请参阅第 5.4 节）。

即使对于理想的头戴式设备，运动也会因为视力而导致晕眩。遵循 10.2 节中提出的策略可以减少晕眩症状。一个更好的想法是设计需要很少或不需要移动的 VR 体验。

作为最后的措施，还有两种策略可能有助于缓解 VR 晕眩[145]。首先是定期练习，这会导致适应。随着身体对不寻常的刺激组合的调整，预计强制融合导致的疲劳量会减少。当然，如果 VR 体验让大多数人感到不舒服，那么让他们“通过”十几次甚至更多的练习不是一个好的主意。最后，用户可以服用降低易感性的药物，就像有些人在登机前服用空气药丸一样。这些药丸通常是抗组胺药或抗胆碱药，它们具有一些副作用，例如疲劳，嗜睡，认知能力受损，以及在某些情况下成瘾的可能性。

## 12.4 人类受试者的实验

想象一下，你已经开发出一种新的运动方法，希望能减少 VR 晕眩。你和几个朋友可能会尝试它，并相信它比默认方法更好。你如何说服那些怀疑它的人，尤其是那些不太可能偏向于新方法的人？你可以争辩说它更好，因为它尊重来自人体生理学和知觉的已知问题，这将是一个合理的开始。这将为首先尝试该方法提供良好的动力；然而，它本身也还不够，因为人体与技术的交互方式存在很多不确定性。解决方案是设计一个实验，科学地确定你的方

法是否更好。这导致了许多挑战，比如确定有多少人应该尝试，应该做什么，应该做多长时间，谁应该分配给哪种方法，以及之后如何测量他们的晕眩程度。其中，一些困难讲述于 12.3 节。如果实验设计得很好，那么科学家会支持你的结果。如果有些人还不相信，那么至少你会得到那些相信科学方法的人的支持！幸运的是，这包括心理学家和神经科学家，甚至包括人机交互相关领域的研究人员[37,40]。

## 科学的方法

科学方法自古以来就已存在，并在特定的科学领域不断完善和改进。图 12.4 描述了它如何在 VR 开发中出现。想象一下，试图爬上梯子，第一步是通过研究适当的文献或获得必要的背景来设计一种可能有所改进的新方法。这将减少从梯子上坠落的可能性。第二步是设计和实施新方法。这一步可能包括对少数用户进行一些简单的评估，以确保值得进一步处理。

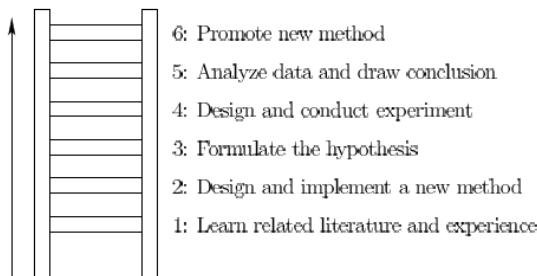


图 12.4：科学过程就像爬梯子一样。每次失败时都要小心，不要跌得太低！

第三步是精确地制定假设，关于它是如何改进的。例如：1) 减少不良症状，2) 提高舒适度，3) 提高解决任务的效率，4) 更强烈地相信虚拟世界是真实的，5) 更好地享受这项活动。评估多个标准通常是有意义的，但结果可能是新方法在某些方面更好，而其他方面更糟。这是一个普遍的结果，但这总比不以任何方式改善要好！这个假设甚至可能涉及改进未来的实验程序，在[59]中举了一个例子，在这个例子中，研究人员确定了哪些生理学指标比问卷更好地指示了 VR 晕眩。最后，该假设应简化实验的第四步，同时保持结果的可靠性。

对于第四步，应该设计和实施实验来测试假设。第五步是分析数据并得出结论。如果结果是关于感兴趣的标准的“更好”的方法，那么就到达了六步，此时新方法应该呈现给世界。

任何步骤都可能产生偏差。例如，在进行实验后，可能会意识到，项目池过于有偏向性。这时就需要退一步，重新设计或重新实施实验。如果第五步的结论是该方法不是明显的改进，或者更糟，这可能需要返回到第二步甚至第一步。所以设计关键是每个步骤都要小心谨慎，避免每次失败都会导致实验被重置得太多！

## 人类—实验研究对象

做关于人的实验很困难，特别是如果他们是科学实验的研究对象。他们之前的虚拟现实经验，对晕动病的敏感性，对技术的怀疑，喜怒无常以及倾向性可能会有很大差异。由于好奇心，经济补偿，感兴趣程度或学位要求（心理学学生常常被迫参加实验），他们可能会同意成为实验对象。一位科学家或许能够根据诸如性别，年龄或职业等因素猜测一些人在实验中的表现。应用科学方法来制定和评估关于人群（或动物）的假设的主题被称为行为科学[152]。

最大的挑战之一是他们是否在不经意间被观察到（甚至不知道他们是否是实验的一部分），或者实验是否呈现了在现实世界中永远不会遇到的刺激或情况。人为设置有时会导致科学家反对实验的生态效度。幸运的是，虚拟现实是我们想要评估的一个特别设计的设置。因此，关于 VR 使用的结论更可能在生态学上是有效的，特别是即使没有用户意识到实验也

可以获得实验数据。头部追踪数据可以在服务器上收集，有数百万人尝试 VR 体验。

## 道德标准

这导致了下一个挑战，那就是人类的权利，人们可能拥有比动物更多的权利。必须避免影响其隐私或健康的实验。涉及人类受试者的科学实验必须坚持高标准的伦理道德，这是从纳粹医学实验和 20 世纪中叶的 Tuskegee 梅毒实验中学到的一课。纳粹战争罪法庭的结果产生了纽伦堡法典，其中规定了一套关于人体实验的道德原则。今天，世界各地正在认真对待人类研究的伦理标准，正在进行的辩论或细节上的差异[236]。在美国，涉及人类受试者的实验需要由法律机构审查委员会（IRB）批准。通常情况下，IRB 这个术语也用于指由审核委员会批准的实验或一组实验的提案，如声明“需要 IRB”。涉及虚拟现实的实验通常不存在争议，并且与数十年来广泛认可的模拟病症实验类似。

## 变量

行为科学家总是关心变量。每个变量都有一个集合中的值，可以是数字，也可以是实数，也可以是符号，如颜色，标签或名称。从他们的角度来看，最重要的三类变量是：

- 依赖性：**这些是假设所关注的主要对象。
- 独立：**这些值是由科学家直接改变或操纵的值。
- 滋扰：**由于这些因素各不相同，它们的价值观可能会影响因变量的价值，但科学家对它们的控制力较弱，而且它们不是主要研究的对象。

高级任务是制定一个假设，可以根据独立变量和因变量之间的关系进行评估，然后设计一个可以控制滋扰变量的实验，并且可以在有限的时间，资源内完成。

用于制定变量如何行为并预测其行为的模型的基础数学是 6.4 节介绍的概率论。不幸的是，我们正面临着一个相反的问题，如图 12.3 所示。大多数行为都不是直接可观察的，这意味着我们必须收集数据并对基础模型做出推论，并尽可能获得更多的信服。因此，解决这一假设是应用统计学中的一个问题，它是概率论的自然补充或逆转。

## 提出假设

在最简单的情况下，科学家想要确定一个关于假设的二元结果：真或假。在更复杂的情况下，可能有许多相互排斥的假设，科学家们想确定哪一个是真的。例如，17 种不同的运动方法中哪一种最舒适？从简单的案例出发，假设已经根据 VR 晕眩确定了一种潜在更好的运动方法。令  $x_1$  表示使用原始方法，并让  $x_2$  表示使用新方法。

集合  $x = \{x_1, x_2\}$  是自变量。每个  $x_i$  被称为一次实验。接受原始方法的受试者被认为是对照组。如果正在评估一种药物，对照组不使用药物的话，那么他们将接受安慰剂。

回顾第 12.3 节，可以通过各种方式评估 VR 晕眩的水平。为了举例，假设选择一段时间内平均的 EGG 电压测量值作为因变量  $y$ ，对应于  $x_1$  或  $x_2$  条件下的胃肠不适程度。

该假设是将  $x$  与  $y$  联系起来的逻辑真/假陈述。例如，它可能是

$$H_0 : \mu_1 - \mu_2 = 0, \quad (12.1)$$

其中，通过对世界上所有人施加实验  $x_i$  可以得到每个  $\mu_i$ ，即在实验中同一点上  $y$  的“真实”平均值。假设  $H_0$  意味着新方法对  $y$  没有影响，通常称为零假设。 $H_0$  的负面被称为替代假设。在上述情况下，替代假设为

$$H_1 : \mu_1 - \mu_2 \neq 0, \quad (12.2)$$

这意味着新方法对胃肠不适有影响，然而，它可能会更好或更糟。

## 测试假设

不幸的是，科学家无法同时对所有人进行同样的实验。他们必须从人群中抽出一小部分人，并确定该假设是否属实。让指数  $j$  指向一个特定的选择主题，并让  $y[j]$  作为受试者对实验的回应；每个受试者的反应是一个因变量。两个统计数据对于结合因变量的信息很重要：

平均数：

$$\hat{\mu} = \frac{1}{n} \sum_{i=1}^n y[i], \quad (12.3)$$

这只是  $y[j]$  在主体上的平均值，以及方差：

$$\hat{\sigma}^2 = \frac{1}{n} \sum_{j=1}^n (y[j] - \hat{\mu})^2. \quad (12.4)$$

方差估计（12.4）被认为是“真实”方差的有偏估计量，因此，有时会应用贝塞尔的修正，即将  $n - 1$  放入分母而不是  $n$ ，从而得到无偏估计量。

为了检验这个假设，Student 的 t 分布（“学生”是威廉·西利戈斯特）被广泛使用，这是一个概率分布，它捕获了任意  $n$  个随机样本被选择及其响应  $y[j]$  被平均后均值的分布情况，见图 12.5。假定每个个体  $j$  的响应  $y[j]$  是一个正态分布（工程中称为高斯分布），它是最基本和最常见的概率分布。它的平均  $\mu$  和标准偏差  $\sigma$  完全给出。这些分布的确切表达式在这里没有给出，采用通用形式，具体请参阅[124]以及其他有关这些以及更多数学统计的书籍。

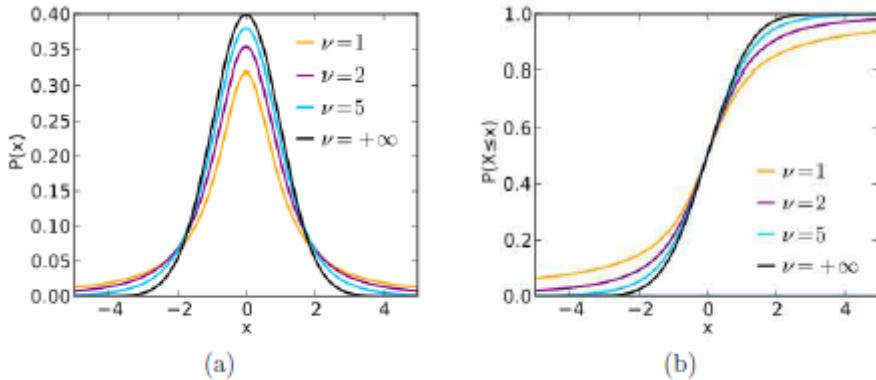


图 12.5：Student 的 t 分布：(a) 概率密度函数 (pdf)；(b) 累积分配函数 (cdf)。在这些图中， $v$  被称为自由度，并且对于被摄体数  $n$ ， $v = n-1$ 。当  $v$  很小时，pdf 的尾部比正态分布更大；然而，当  $v$  接近  $\infty$  时，Student 的 t 分布收敛于正态分布。（维基百科用户 skbkekas 提供）

Student 的 t 检验[314]涉及计算以下内容：

$$t = \frac{\hat{\mu}_1 - \hat{\mu}_2}{\hat{\sigma}_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}, \quad (12.5)$$

其中：

$$\hat{\sigma}_p = \sqrt{\frac{(n_1 - 1)\hat{\sigma}_1^2 + (n_2 - 1)\hat{\sigma}_2^2}{n_1 + n_2 - 2}} \quad (12.6)$$

其中， $n_i$  是接受治疗  $x_i$  的受试者数量。表达式中减 1 和 2 是由于贝塞尔的修正。基于 t

的值，通过查看 Student's t cdf 的表格来确定零假设  $H_0$  中的置信度  $\alpha$  (图 12 (b))。通常， $\alpha=0.05$  或更低就足以说明  $H_1$  是真实的 (相当于 95% 的置信度)。这样的表格通常是这样排列的：对于给定的  $v$  和  $\alpha$ ，提供确认  $H_1$  以置信度  $1-\alpha$  所需的最小  $t$  值。请注意，如果  $t$  是负数，那么  $x$  在  $y$  上的效果将朝相反的方向运行，并且用  $-t$  替换查表。

二元结果可能不够令人满意，但这不是问题，因为平均值  $\mu_1 - \mu_2$  中的差异是应用  $x_2$  与  $x_1$  相比的变化量的估计值。这被称为平均实验效果。因此，除了通过  $t$  检验确定  $H_1$  是否为真，我们还可以估计它对结果的影响程度。

学生的  $t$  检验假定每个组内的方差是相同的。如果不是，那么使用 Welch 的  $t$  检验[343]。请注意，在这两种情况下均未预先给出差异。它们是根据实验数据“实时”估算的。如果方差恰好相同，则 Welch 的  $t$  检验给出与 Student's  $t$  检验相同的结果，因此，如果存疑，最好应用 Welch 的  $t$  检验。科学家们则可以使用许多其他测试并在特定情况下进行辩论；见[124]。

## 相关系数

在许多情况下，自变量  $x$  和因变量  $y$  都是连续的（取实值）。这便引出了另一个重要的参数称为皮尔逊相关系数（或皮尔逊的  $r$ ）。这估计了两个变量之间的线性相关量。对于每个受试者，处理条件  $x[i]$  被应用，并且响应是  $y[i]$ 。请注意，在这种情况下，没有组的概念（或每个主题都是唯一组）。此外，任何治疗都可能适用于任何人；索引  $i$  仅表示特定的受试者。

当作为随机变量处理时， $r$  值被计算为  $x$  和  $y$  之间的估计协方差：

$$r = \frac{\sum_{i=1}^n (x[i] - \hat{\mu}_x)(y[i] - \hat{\mu}_y)}{\sqrt{\sum_{i=1}^n (x[i] - \hat{\mu}_x)^2} \sqrt{\sum_{i=1}^n (y[i] - \hat{\mu}_y)^2}}, \quad (12.7)$$

其中  $\mu_x$  和  $\mu_y$  分别是所有组的  $x[i]$  和  $y[i]$  的平均值。分母只是估计的标准偏差的乘积： $\sigma_{xy}$ 。

$r$  值介于 -1 和 1 之间。可能发生三种定性不同的结果：

- $r > 0$ : 这意味着  $x$  和  $y$  正相关。随着  $x$  的增加， $y$  趋于增加。 $r$  值越大意味着效果越强。
- $r = 0$ : 这意味着  $x$  和  $y$  是不相关的，这在理论上相当于零假设。
- $r < 0$ : 这意味着  $x$  和  $y$  是负相关的。随着  $x$  增加， $y$  趋于减小。 $r$  值越小意味着效果越强。

在实践中，从实验数据中很难获得  $r = 0$ ；因此，绝对值  $|r|$  给出了  $y$  取决于  $x$  的可能性的重要指示。与零假设 ( $r = 0$ ) 的理论等价性只会在受试者的数量趋于无穷大时才会发生。

## 处理滋扰变量

我们已经考虑了依赖性和自变量，但忽略了这些滋扰变量。这是实验设计中最具挑战性的部分。这里只给出了总体思路，请参阅[152,195]了解详尽的介绍。假设在查看数据时因变量  $y$  在很大程度上取决于对象的可识别属性，如性别。这个属性将成为一个令人麻烦的变量  $z$ 。我们可以想象设计一个实验来确定  $z$  是否影响  $y$  以及  $z$  对于  $y$  有多大影响。

对  $z$  的依赖性会导致主体间的差异很大；然而，如果它们被分成每个组内具有相同  $z$  值的组，则方差可能会相当低。例如，如果性别是滋扰变量，那么我们会将受试者分成男性和女性小组，并发现每个小组的方差都较小。这种技术称为分块，每个组称为一个块。在块内部，如果自变量  $x$  保持不变， $y$  的方差应该很小。

接下来的问题是要确定实验适用的对象。继续这个例子，对女性给予  $x_1$  实验，对男性给予  $x_2$  实验会是一个可怕的想法。这完全混淆了因变量  $y$  的滋扰变量  $z$  和自变量  $x$  的依赖关系。与此相反，将  $x_1$  应用于一半的女性和男性， $x_2$  应用于另一半，这显然更好。一个简单的选择是使用随机设计，其中受试者随机分配  $x_1$  或  $x_2$ 。这可以安全地消除意外偏差，并且实验者可以轻松开展实验。

如果有多个滋扰变量，那么分配过程变得更加复杂，这往往导致对随机化的依赖性。如果受试者参加了不同时间应用不同疗法的多阶段实验，则必须认真分配疗法。处理它的一种方法是根据拉丁方阵分配治疗方法，拉丁方阵是  $m \times m$  矩阵，其中每行和每列都是  $m$  个标签（在这种情况下为实验方案）的置换。

## 方差分析

剩下的主要挑战是找出对方差产生重大影响的滋扰变量。这称为方差分析（或 ANOVA，发音为“ay nova”），考虑到这一点的方法称为 ANOVA 设计。性别是一个容易想象的因素，但其他因素可能更为微妙，例如受试者玩过的 FPS 游戏的数量，或受试者已玩的时间。这个设计过于复杂，不能完全覆盖所有情况（见[152]），但重要的点是必须在主体之间发现低方差的群集，作为将它们分块的基础。这与无监督聚类问题（或无监督学习）密切相关。方差分析也被认为是对三个或更多变量的 t 检验的推广。

## 更多变量

除了独立、依赖和滋扰之外的变量有时在实验中也会起重要作用。控制的变量本质上也是一个变量，通过选择主体或实验性试验来保持固定不变。例如，通过控制对象选择，可以使方差保持较低，如在实验中仅使用年龄在 18 岁至 21 岁之间的男性。该方法有助于提高对实验结论的可信度，但可能会阻止将其发现推广到控制之外的情况。

混杂变量是导致独立变量和因变量相关的无关变量，但是一旦给出了混杂变量的值，它们就变得不相关。例如，拥有较大的鞋号可能与更好的语言能力相关。在这种情况下，混杂变量就是人的年龄。一旦知道了年龄，我们就会意识到老年人有比小孩子更大的鞋码，表达能力也更好。这说明了从统计相关性推断因果关系的冒险性。

## 心理物理学方法

回想第 2.3 节，心理物理学将感知现象与原始刺激联系起来，这对理解 VR 来说至关重要。史蒂文斯的幂律(2.1)将感知的刺激幅度与实际幅度相关联。JND 涉及确定差异阈值，这是可检测的最小刺激变化量。这种情况的一个特例是绝对阈值，它是可检测的最小幅度刺激（与零相比）。

心理物理学的规律或关系是通过对人体的特定实验获得的。术语心理物理学和研究领域由 Gustav Fechner [76]介绍，他提出了三种基本的实验方法，下面将对此进行介绍。假设  $x$  代表刺激幅度。其任务是确定  $x$  可以变得多小，这样受试者就会感受到差异。

传统的方法是：

- 不断刺激的方法：在这种情况下，连同参考刺激一起呈现不同大小的刺激。该方法要求对于每个刺激，受试者可以感知到刺激之间的差异。幅度通常以随机顺序呈现以抑制适应。基于对许多试验的反应，计算出最佳拟合心理测量函数，如图 2.21 所示。

- 限制方法：实验者以较小或较高的限制开始以较小的增量改变刺激幅度。在每种情况下询问受试者新的刺激是否比参考刺激小，相同或更大。

- 调整方法：允许对象在短时间内上下调整刺激强度，同时还可以与参考刺激进行比较。当他报告调整后的刺激有与参考刺激相同的幅度时，实验停止。

尽管这些方法是有效的并且被广泛使用，但存在一些问题。所有受试人在实验时都可能产生偏见。对于倒数两种方法，适应性可能会干扰结果。对于最后一个方法，没有办法控制主体如何做出决定。另一个问题是效率，因为考虑到远离参考刺激的刺激，多次刺激的更新会导致时间的浪费。

### 自适应方法

由于这些缺点，研究人员在过去几十年中找到了许多改进的实验方法。其中很大一部分在[324]中进行了调查和比较，属于适应性心理物理学方法的领域。大多数改进的方法执行阶梯程序，其中刺激量级从测试对象容易察觉的情况开始，并且逐渐减小梯度直至受试者出错[90]。到了这一点，再进行相反的过程，直到产生另一个错误。出现错误和改变方向的过程一直持续到在很短的多次实验中出现许多错误为止。步长必须仔细选择，甚至可以在实验过程中逐渐减小。当每次迭代收集数据时，方向（增加或减少）可以使用贝叶斯或最大似然算法来决定，该方法可以提供阈值的估计值[113,154,342]。这些方法通常属于随机逼近方法的范畴[266]。

### 刺激幅度估计

回想一下，史蒂文斯的幂律不是关于检测阈值，而是关于刺激的感知幅度。例如，一个盘子可能会感到两倍热。在这种情况下，可以要求受试者估计刺激之间幅度的相对差异。在足够数量的试验中，史蒂文斯幂指数（2.1）的指数可以通过选择使最小平方误差最小的值（指数）来估计（从9.1节回忆）。

### 拓展阅读

有关感知学习的调查，请参见[94,98,109,253]。感知学习的过度行为在[101,253]中进行了研究。在[283]中已经确定，感知学习可以在没有集中注意力的情况下发生。

[62,67,200,357]分析了人类对VR和计算机接口延迟的敏感性。在[289]中研究了立体显示器的舒适问题。关于姿势摇摆和晕眩之间的联系，请参见[294,313]。

有关与VR晕眩有关的一些重要研究，请参阅[13,146,147,153,223,263]。[143,169,305]给出了VR疾病的一般概述。[262]中调查了晕动病。参见[120,138,55,255]可以了解强制融合的更多信息。

有关人类受试者实验的数学方法和统计数据的介绍，请参见[152]。[195]因其在心理学背景下对假设检验的报道而受到欢迎。有关心理物理学方法的治疗，请参阅[176,324,349]和[93]的第3章。

# 第十三章 未来前景

我们来到最后一章，该章将介绍一些可能影响未来广泛使用 VR 的主题，但目前还处于研究和开发阶段。第 13.1 和 13.2 节涵盖了我们忽略掉的感官。在前面章节，我们介绍了视觉，听觉和平衡（前庭）感觉，还剩下触觉，嗅觉和味觉。第 13.1 节涵盖触觉，或者更一般地说，体感系统。这包括刺激体感系统的生理学，感知和工程技术。第 13.2 节涵盖了两种化学感官，气味和味道，并试图为它们设计“呈现方式”。第 13.3 节讨论了如何使用机器人进行远程监控，以及它们最终如何成为我们的替代者，通过它们可以在 VR 界面探索现实世界。就像虚拟世界中的化身（10.4 节）一样，机器人在现实世界中变成一种物理化身。最后，第 13.4 节讨论了人类增强和交互的最终级别的步骤：脑机接口。

## 13.1 触觉和本体感受

视觉和听觉是 VR 系统的主要焦点，因为它们相对更容易使用现有技术进行捕捉。相应器官集中在头部的一个小地方，头部跟踪技术便宜而准确。不幸的是，这忽略了触觉和本体感受的强大感觉以及与我们周围世界密切相关的系统。我们的眼睛和耳朵使我们能够远距离感知世界，但触摸似乎让我们能够直接感受到它。此外，本体感受使身体感觉到世界上任何地方的重力以及肢体和其他可以被肌肉移动的结构的相对位置或构型。因此，我们将考虑这些被忽视的感官，从它们的受体到感知，然后再考虑试图超越它们的工程系统。

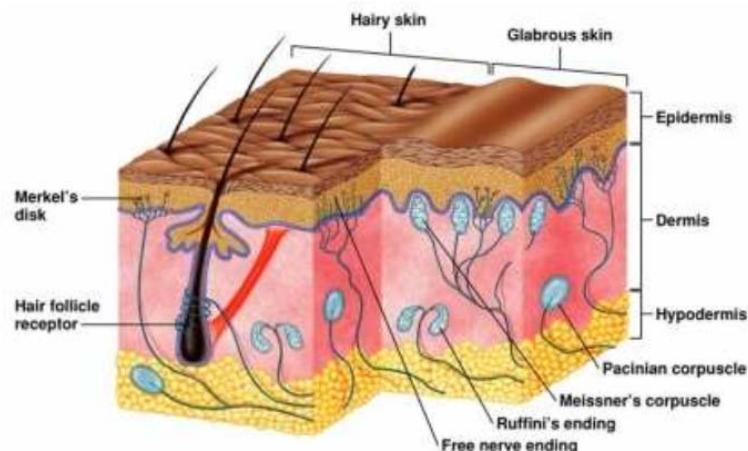


图 13.1 皮肤上的六种主要受体（图源于 Pearson Education）

### 躯体感觉系统

身体感觉向大脑提供关于人体本身的信号，包括与皮肤的直接接触，身体的构造和世界中的运动以及环境温度。在这一类别中，前庭系统（8.2 节）处理平衡，而躯体感觉系统处理触觉，本体感觉和动觉。考虑人体及其所有可移动部分，如腿，手臂，手指，舌头，嘴巴和嘴唇。本体感觉对应于每个部分相对于其他部分的姿势的意识，而动觉是运动本身的对应部分。换句话说，动觉提供速度，加速度和力量的信息。

躯体感觉系统至少有九种主要受体，其中六种用于触觉，其余三种用于本体感觉和动觉。图 13.1 描绘了嵌入皮肤（真皮）的六个主要触觉感受器。它们的名字，结构和功能是：

**游离神经末梢：**这些是没有专门结构的神经元。它们具有延伸到外皮（表皮）的轴突，

其具有感测极端温度（热和冷）的主要功能以及来自组织损伤的疼痛。这些神经元是特殊的（称为假单极），因为轴突在典型的神经细胞中起到树突和轴突的作用。

**鲁菲尼氏小体：**它们深深地嵌入皮肤中，并指示随时发生的拉伸量。它们有一个缓慢的时间响应。

**帕西尼安小球：**这些小体充满液体并对压力做出反应。它们的响应速度很快，可以感应高达 250 至 350 Hz 的振动（压力变化）。

**梅克尔氏小盘：**这些结构出现在表皮的正下方，并对静态压力作出反应（随时间变化很小或没有变化），并具有缓慢的时间响应。

**迈斯纳氏小体：**这些也只是在表皮下方，并且对轻微的触感作出反应。它们的反应速度比梅克尔氏小盘和鲁菲尼氏小体更快，可以感应到高达 30 到 50 赫兹的振动；这并不像帕西尼安小球那样高。

**毛囊受体：**这些对应于紧密缠绕在毛根周围的神经末梢；它们有助于轻微的触觉，并且如果毛发被移除也会造成疼痛。

这些受体的前四种出现在全身皮肤上。迈斯纳氏小体只在没有毛囊（无毛的皮肤）的部分，并且毛囊受体显然只出现在有毛的地方。在一些重要的地方，例如眼睑，嘴唇和舌头，称为 endbulbs of Krause 的热感受器也出现在皮肤中。另一类是痛觉感受器，它出现在关节组织中，并引起过度伸展，损伤或炎症引起的疼痛感。

触摸具有空间和时间分辨率。空间分辨率或敏锐度对应于密度或每平方米面积的受体，其在身体上变化。指尖密度高，背部非常低。这对触摸感知有影响。时间分辨率与听觉不同，延伸至 20,000 Hz；帕西尼安小球可以将高达几百赫兹的振动与静态压力区分开来。

关于本体感觉（和动觉），有三种受体：

**肌梭：**顾名思义，它们嵌入每个肌肉的内部，因此它们长度的变化可以传递给中枢神经系统（包括大脑）。

**高尔基腱器官：**它们嵌入肌腱中，每个肌腱都是纤维组织的强韧带，通常将肌肉与骨骼连接起来。器官报告肌肉紧张度的变化。

**关节感受器：**它们位于骨骼之间的关节处，协助肌肉运动，同时还向中枢神经系统提供关于骨骼相对位置的信息。

通过这些受体，身体意识到其各种运动部件的相对位置，方向和速度。

体感系统的神经通路以类似于 5.2 节的视觉路径的方式工作。信号通过丘脑传送，相关信息最终到达大脑中主要的体感皮层，在那里进行更高级别的处理。早在丘脑之前，一些信号也通过脊髓传送到控制肌肉的运动神经元。这使得快速的运动反应能够迅速从疼痛刺激中退出，并且可以用于膝跳反射。在主体感皮层内部，神经元以与其身体位置相对应的空间布置进行发射（地形图）。一些神经元也具有与皮肤上的局部斑块相对应的感受野，这与感受野对视觉的作用非常相似（回忆 5.2 节的图 5.8）。再一次，侧向抑制和空间对立性存在并形成检测器，估计沿着皮肤表面的尖锐压力特征。

## 体感知觉

我们现在从生理学转变为体感知觉。来自心理物理学的概念（2.3 节和 12.4 节）再次出现，导致确定检测阈值，感知的刺激强度，以及沿着时间轴和空间轴的敏锐度或分辨率。例如，在[26]中研究了以不同频率和温度呈现的振动存在的检测能力。

## 两点敏锐度

通过两点敏锐度测试研究空间分辨率，其中通过一对尖锐卡尺在附近两个地方戳皮肤。询问受试者是否在同一时间在不同的地方看到一次戳或两次戳。检测阈值然后由身体上的位

置排列以理解空间分辨率如何变化。最敏锐的敏锐度在舌头和手上，如果它们接近 2 或 3 毫米，可以区分出点。舌头和手指的尖端具有最高的敏锐度。对于额头，门槛大约 20 毫米。背部的敏锐度最低，导致约 60mm 的阈值。这些结果也被证明与躯体感觉皮层中感受野的大小直接对应。例如，对应于背部的神经元比指尖具有更大的场（在皮肤面积方面）。

### 纹理感知

通过滑动手指，可以感应到表面的纹理感知结果。表面突出或凹面的小元素的大小，形状，排列和密度影响最终感知的纹理。双重理论指出，较粗糙的纹理（较大的元素）主要由空间线索感知，而较精细的纹理主要通过时间线索感知[125,142]。通过空间提示，这意味着可以通过将手指按压在表面上来推断结构。通过时间提示，手指滑过表面，导致帕西尼娅和迈斯纳小体可以感知到的压力振动。对于更细的纹理，可能需要较慢的运动，以使振动频率保持在 250 至 350 Hz 以下。回想一下第 12.1 节，人们在阅读盲文时可以学会改善纹理感和敏锐度。因此，可以应用感知学习来改善触觉（触觉）感知。

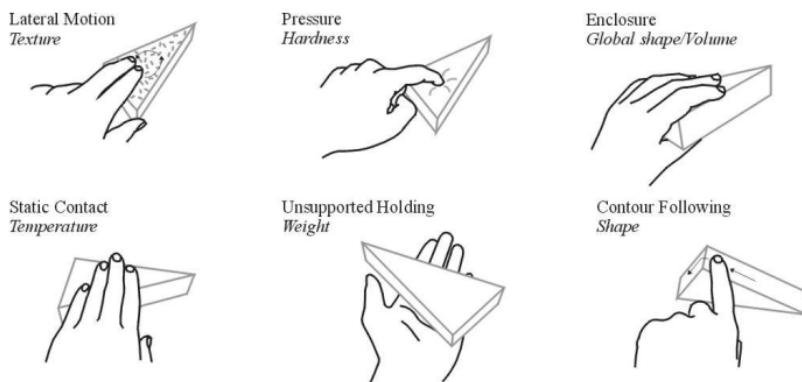


图 13.2 触觉探索涉及手和对象之间的几种不同类型的相互作用，以学习对象属性，例如大小，形状，重量，牢固度和表面纹理。

### 触觉感知

对于较大的物体，可以通过触觉探测来推断其整体几何形状，其涉及处理该物体。想象一下，有人给你一个未知的物体，并且你必须在蒙上眼睛时确定它的形状。图 13.2 显示了触觉探测的六种不同定性类型，其中每种都涉及不同类型的受体以及空间和时间信息的组合。通过整合这种手中操纵的体感信号，学习物体的几何模型。

### 体感幻觉

回想 6.4 节，大脑将多种感应模式中的信号相结合，以提供感知体验。正如麦格克效应使用视觉和听觉线索之间的不匹配性一样，幻象也被视觉和体感系统之间的线索不匹配所发现。橡胶手错觉是最广为人知的[66]之一。在这种情况下，科学家们进行了一项实验，让受试者坐在一张桌子上，双臂搁在桌子上。受试者的左臂被覆盖，但替代橡胶前臂被放置在桌子附近，并保持可见，使得它看起来好像是他们自己的左臂。实验者用油漆刷抚摸真实和假前臂，以帮助建立与假前臂的视觉和触觉关联。科学家使用功能性核磁共振成像扫描仪确定，无论是真实还是假前臂，大脑的相同部位都被激活。此外，他们甚至知道，用针做一个刺伤的手势会引起疼痛的预感和撤回没有受到威胁的真正左臂的倾向[66,293]，并且甚至可以通过一定的关联感知到冷热感觉[291]。

更一般地说，这被称为身体转移错觉[251,293]。在 1.2 节的图 1.14 中显示了一个 VR 系统的例子，其中男性和女性确信他们正在交换身体，而来自摄像机的视觉信息与协调的手部

运动相结合以提供触觉感官刺激。这种现象的应用包括移情并帮助截肢者克服幻肢感。这种幻觉还可以通过控制肌肉，同时从 VR 获得视觉反馈，来深入了解可能学习的运动项目类型，如 10.1 节和 10.3 节所讨论的。它进一步影响了对 VR 中自我的感知，这在 10.4 节和 12.2 节中已经讨论过。



图 13.3 橡胶手错觉，其中一个人对假手作出反应，仿佛假手是她自己的。

触觉界面通过工程设备产生的触觉是通过许多不同的系统提供的。第 1.1 节的图 1.1 显示了一个系统，其中通过用户推动机械机翼飞行来提供力反馈。此外，风扇模拟强度与真实飞行人员速度成比例的风。整个身体也倾斜，以便提供适当的前庭刺激。

图 13.4 显示了几个更多的例子。图 13.4 (a) 显示了一个带有滚轮的 PC 鼠标。当用中指旋转轮子时，会感觉到离散的凸起，因此可以产生更仔细的校准运动。图 13.4 (b) 显示了一种游戏控制器附件，其在体验期间的关键点处提供振动，例如爆炸或身体接触。

许多触觉系统涉及使用机械臂在小区域内的精确位置和方向上施加力或压力。图 13.4 (c) 显示了这样一个系统，用户拿着一支连接在机器人手臂上的笔。力量从机器人传达到手指。当笔触击虚拟表面时，机器人通过阻止其运动向用户提供力反馈。笔可以拖过虚拟表面以感受任何类型的纹理[237]; [50]中介绍了各种模拟纹理。在医疗设备的开发中提供这样的力反馈是重要的，使得医生能够通过连接到真实设备的接口执行手术程序。如果没有准确和及时的触觉反馈，医生很难执行很多程序。

图 13.4 (d) 显示了一个类似于视觉显示的触觉显示。矩形区域通过行和列索引，并且在每个位置处可以迫使小销向外。这使得图形出现在表面，同时也允许各种水平的压力和频率的振动。

所有这些例子都涉及应用于手部的触觉反馈；然而，触觉受体出现在人体各处。为了提供对较大部分受体的刺激，可能需要触觉套装，其在套装上的各个点处提供力，振动或甚至电刺激。这些系统的缺点是在每次会话时穿上和脱下衣服都很麻烦。



图 13.4 (a) 具有滚轮的罗技 M325 无线鼠标, 当轮子执行完整的转动时, 提供 72 个凹凸形式的触觉反馈。 (b) 世嘉 Dreamcast Jump Pack (1999), 附加到游戏控制器并在游戏过程中提供振动。 (c) 触觉全指向笔触觉装置, 通过笔将手指压力和振动传达给手指。 (d) KGS 点视模型 DV-2, 它是一个触觉引脚阵列。当指尖在其表面上扫描时, 引脚被强制向上以模拟各种纹理。

通过增强现实触摸反馈鉴于工程触觉显示器的困难, 另一种方法是依靠匹配区域中的真实物体向体感系统提供反馈。这有时被称为有形的用户界面[327]。如第 8.3.3 节所述, 通过调整真实世界和虚拟世界可以获得强大的体验。在极端情况下, 透明显示器 (如 1.2 节中显示的 Microsoft Hololens) 使用户能够查看周围的物理世界并与其进行交互。显示器只是将虚拟物体添加到现实世界中, 或者直观地增强真实物体。这些系统通常包含在增强现实或混合现实中。

## 13.2 嗅觉和味觉

到目前为止, 没有考虑到的人类感觉就是嗅觉和味觉[65]。它们通常被归为化学感觉, 因为它们的受体通过与到达它们的分子发生化学相互作用而起作用。所产生的化学感受器对特定物质和足够高的浓度水平作出反应。与其他感官相比, 关于它们的研究少得多, 并且能“展示”鼻子和舌头刺激的电子设备也很少。尽管如此, 这些感觉是非常重要的。人造气味的设计是一项庞大的业务, 其中包括香水, 除臭剂, 空气清新剂, 清洁剂和熏香。同样, 设计品味是现代食品工业的基础 (无论好坏)。

### 嗅觉生理学和感知

气味对于多种生物学的目的来说很重要, 包括寻找猎物, 选择潜在的配偶, 并判断食物是否能安全食用。嗅觉受体神经元位于鼻腔的顶部, 覆盖 2 至 4 平方厘米的面积。大约有 600 万个受体, 据说它们可以感受 500 到 1000 种不同的气味类型, 这取决于它们对特定化学成分的响应[204]。空气中的分子溶解到嗅觉粘液中, 便触发受感染纤毛(小毛发)的检测。嗅觉受体不断再生, 平均寿命约为 60 天。除了受体之外, 一些游离的神经末梢也位于嗅粘液中。感觉通路是不寻常的, 因为它们在达到它们的最高级别目的地之前不经过丘脑, 在嗅觉中被称为初级嗅觉皮层。从受体到杏仁核也有直接的路径, 这与情绪反应有关。这可能有助于解释嗅觉和情绪反应之间的密切联系。

就感知而言, 人类可以识别成千上万种不同的气味[286], 女性通常比男性表现得更好

[36]。鉴别能力取决于气味的浓度（以每立方米面积的分子计）。如果浓度较弱，则能力的差异程度下降。此外，当一种气味被认为是高浓度的气味时，对于另一种气味几乎不可检测到。因此，检测阈值根据物质而变化千分之一或更多。适应气味也很重要。人们不断地适应周围的气味，尤其是他们自己身体或家中的气味，以使得它们变得不再明显。吸烟者也会采取适应措施，以避免他们自己像非吸烟者那样感受到被污染的空气。

似乎人类可以识别比嗅觉受体数量更多的气味。因为编码的组合性，这其实是可能的。任何单一的气味（或化合物）都可能引发多种受体。同样，每个受体都可能被多种气味触发。因此，气味与受体之间存在多对多映射。这便能够根据被激活的受体类型的不同亚组区分出更多的气味。

### 嗅觉相关问题

给电影“添加香味”可以追溯到 20 世纪初。从 1960 年开始，一种系统被称为“气味 O 型视觉”开始使用，其在电影中的不同时间点将 30 种不同的气味注入电影院座位。1.3 节图 1.28 (c) 中提到的 Sensorama 系统也包含了气味。此外，军队几十年来一直使用气味作为模拟器的一部分。

曾经的一次嗅觉呈现调查显示在[136]中，其中也提到了当前的挑战和问题。人们普遍认为，嗅觉能够增加 VR 中的沉浸感。它还在某些涉及情绪反应的医疗治疗中有一定的优势。更令人惊讶的是，最近有证据表明宜人的气味有助于减少视觉诱发的晕动症[147]。

嗅觉显示设备通常涉及可将化合物喷洒到空气中的气泵。这种气味的呈现可以通过靠近人的鼻子来完成。在这种情况下，设备可以穿在身上[356]。最新研发的一个系统如图 13.5 所示。又或者，这些气味可以根据房间的比例进行释放和散播。对于 CAVE 设置来说，这将是优选的，但通常很难控制气味的强度和均匀性，特别是从打开的窗户和通气孔中产生的气流。此外，可能还需要调整大面积上的气味浓度以便进行定位，但这又难以准确实现。

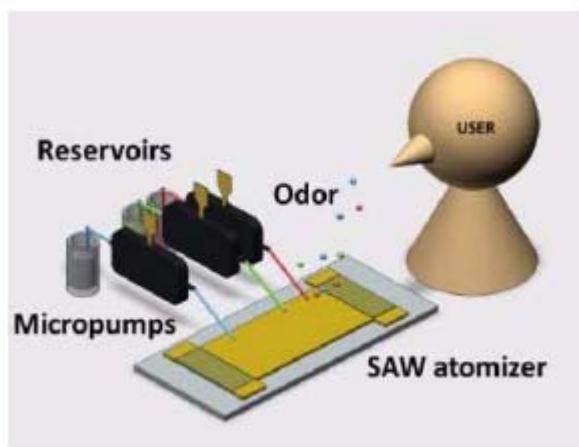


图 13.5：来自[114]的可穿戴嗅觉呈现设备的描述。微型泵迫使小型储液器排出液体。SAW 雾化器是一种将液滴转化为雾化气味的表面声波装置。

### 味觉和感知

我们现在从嗅觉转换到味觉。在人类的舌头上，大约有 10,000 个味蕾，每个味蕾包含一组约 50 至 150 个味觉受体[295]。受体平均存活 10 天，并不断再生。目前，已经确定的五种基本类型的味觉受体为：

- 鲜味 (Umami)：这种味道对氨基酸敏感，如味精（味精），承担对总体味道感知的工作。这使得食品制造商能够以低成本地添加使食物味道更好的化学品。生物学原因可能是因

为氨基酸是蛋白质的重要组成部分。

- 甜味：这对于根据糖含量来确定食物来源非常有用。
- 咸味：这对于确定食物来源是否具有足够的盐含量是非常有用的，对于正常的神经功能是必需的。

•酸味：这对确定食物中的酸量很有用，包含有益的维生素，未成熟，甚至变质食物中的细菌。

•苦味：这通常与有毒植物有关，可能会引发人们对它们的厌恶。

所有这些工作通过溶解食物并基于化学分解产生响应。感觉通路通过丘脑连接到味觉皮层和杏仁核，从而影响情绪反应。

味觉感受与味觉受体类型密切相关。最广为人知的模型之一是 1927 年开始的 Henning 四面体，它是一个具有三个维度的味道，它是使用重心坐标（第 7.2 节）在四个极端顶点上生成的，每个顶点代表纯甜，咸，酸或苦。因此，每个味道都是一个具有四元素的线性插值。这其中忽略了鲜味，它是最近才被添加到受体类型列表中的[43,228]。适应性在口味上也会产生，包括对可能导致恶心的食物。滋味的概念是一种将味觉，嗅觉，温度，触觉，视觉和声音提示相结合的感性体验。因此，理解创造滋味体验的机制是具有挑战性的[57]。

### 味觉刺激



图 13.6：电子棒棒糖是在新加坡国立大学开发的[260]。

迄今为止，以电子方式模拟味道的做法相对较少。图 13.6 显示了一个最近的例子，其中电极放置在舌下以模拟主要味道类型的刺激。在另一项工作中，品尝幻觉是通过伴随不正确的视觉和嗅觉线索进食而形成的[226]。VR 的味觉界面通常很难开发，因为实际上并没有让人们在体验中吃东西。除此之外，显然还有健康和卫生问题。

### 13.3 机器人接口

机器人是可编程设备，涉及传感器，执行器（电机）和计算设备的结合。它们通常被期望来解释高层命令，使用传感器了解周围的世界，并计划并执行使他们安全的行动，以实现其控制者制定的目标。它们的组件从多种形式上都模仿了人体的结构组成。机器人有传感器，人类有感官。对于某些特定的对应关系，机器人具有摄像头，IMU 和关节编码器，而人类则通过视觉，前庭和本体感受来测量相似的物理量。大多数机器人都有电动机，人类都有肌肉，两者都有相同的用途。机器人执行计算以将高级别目标与低级别电机命令相关联，同时解释来自传感器的数据。人类对高层次目标的处理过程也是如此，同时向肌肉发送运动信号并将感官刺激转化为感知现象。在对机器人和人类进行了很多比较之后，便产生一个问题就是：为什么不使用 VR 技术让人类能够栖身于机器人的身体？我们可以使用机器人作为我们自己的替身。

## 远距离操作

实现这一愿景的第一步是与机器人进行远距离互动。一个多世纪以来，车辆已经可以由遥控操作。最早的例子之一是 1898 年在尼古拉特斯拉在纽约公开展示的无线电控制船。在 20 世纪，许多远程操控机器人被开发用于远程或危险情况下的导航，例如处理放射性材料，太空旅行，海底探险。航天机构（如美国航天局）和军方对遥控车进行了广泛的研究和开发。遥控操作的另一个有趣的例子是 1995 年的 TeleGarden，它是一个机器人手臂，悬停在南加州大学的一个真正的花园上，与互联网连接。远程访客可以种植种子，或照顾花园。2001 年，遥控机器人被部署到世界贸易中心的轰炸现场搜寻遇难者。目前，各种遥控车辆已被广泛应用，包括汽车，固定翼飞机，四旋翼飞机（无人驾驶飞机），船只和潜艇。遥控操作通常是比较困难的，因为用户必须在处理控制器时从第三人称视角控制车辆。因此，许多车辆已经配备有无线摄像机，使得用户在屏幕上获得第一人称视角（FPV）。这是迈向网真的重要一步。遥控操作不仅限于车辆。卫生保健是遥操作领域中最大且不断增长的领域之一，其通常涉及操纵医疗仪器的固定式机器人手臂。有关网络机器人的调查，请参见[300]。

## 现代机器人

目前已经有成千上万种不同的机器人被设计和制造，其中一些具有特殊用途，例如在建筑物外面清洁窗户，另一些则用于更普通的目的，例如辅助生活。图 13.7 显示了具有拟人化或“类人”外观的人形机器人。图 13.8 显示了其他类型机器人的样品。1.2 节中的图 1.11 展示了另外两个例子，它们是立体云台模块和视频流无人机。



图 13.7：由日本国立先进工业科学技术研究所（AIST）和 Kawada Industries 在日本生产的 HRP-4 仿人机器人。

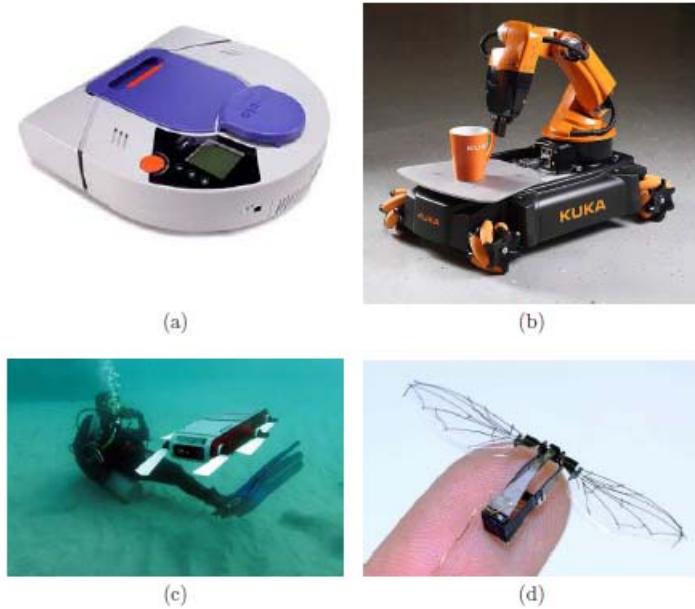


图 13.8：商业和大学机器人样本：(a) Neato XV 真空清洁机器人。 (b) Kuka YouBot，这是一个全方位的移动模块，顶部有一个机械臂。 (c) Aqua，一种来自 McGill 大学的水下机器人[64]。 (d) 哈佛大学飞行的微型机器人[190]。

除了硬件之外，还有大量的软件基础设施可以帮助开发人员，如 ROS（机器人操作系统）和 Gazebo。几乎所有的机器人都有可以连接遥控机器人 VR 接口的候选平台。摄像机和麦克风充当用户的替代眼睛和耳朵。在可实现并且对于具体应用来说比较重要的条件下，夹具（也称为末端执行器）可以用作远程手。用户可以通过键盘，控制器，声音或身体动作来指挥机器人的动作。对于类人机器人，甚至可以使用动作捕捉（第 9.4 节）跟踪人体，并直接映射到人形机器人的运动。更一般地说，机器人的任何拟人化方面都可能成为匹配区域的一部分。另一方面，机器人允许许多非人类的体验，例如变成小昆虫，在房间里飞行，或者像海中的鱼一样游泳。

### 远程呈现

远程呈现的术语和概念归功于人工智能的先驱 Marvin Minsky [213];另见[265,287,310]。对于最理想化的情况（用现有技术远远无法实现），它可以完全消除实际旅行的需要。它也可能会导致因残疾或高龄流动性受限的人的生活发生革命性变化。在技术挑战方面，网真涉及两部分的整合：遥控操作和 VR。

图 13.9 显示了一个市售的远程呈现系统，并作为一个有用的参考点。远程呈现研究中出现了类似的机器人[137,172,243,319]。机器人由用户通过平板电脑或智能手机控制，而在远程地点，机器人携带平板电脑，提供广角相机和屏幕以显示用户的面部。底座设计用于在典型的办公室空间中滚动，并且平板电脑高度可调，以实现面对面的互动。这个设备是“头重脚轻”的，这便可以通过倒立摆来稳定平板电脑。



图 13.9: Double telepresence 机器人是一个手柄上的屏幕和摄像头。该机器人的成本约为 2500 美元，屏幕是平板电脑，如 iPad。高度甚至可以远程调整，以使其看起来像坐着或站着。

关于远程呈现机器人，有几个方面可以想到：

**感官输入:** 来自远程物理世界的感觉是什么？对于视觉输入，它可能包含直接映射眼睛视点并且甚至与用户头部运动匹配的相机（如图 1.11 (a) 所示）。或者，它可以捕捉并传输整个全景图。更进一步，这可以延伸到深度和光场。使用一个或多个麦克风捕获听觉输入，具体取决于本地化的重要性。远程呈现的其他一些可能的输入是温度，接触力，湿度，气味以及机器人的剩余电池寿命。

**移动性:** 机器人在哪里？在没有移动性的情况下，网真被简化为固定的摄像头和麦克风。如果任务是与人互动，那么它应该能够移动到人们能够进入的相同位置。在其他情况下，可能需要许多移动模式，例如飞行，游泳，甚至通过管道爬行。

**视听输出:** 在某种极端情况下，网真系统可能看起来像是“在墙上飞”，并且不会干扰远程地点的生活。更常见的是，它被设计为与人互动，这可以通过屏幕和扬声器来完成。如果机器人具有一些拟人化特征，那么它也可以制作与其他人交流情感或意图的手势。

**操纵:** 图 13.9 所示的远程呈现系统面向面对面交互，因此忽略了能够操纵远程站点上的对象。远程呈现机器人如果能抓住，操纵，携带和取消对象，则功能更加强大。然后，它可以打开门，操作电梯，购物，等等。

本节的其余部分涵盖了开发更好的远程呈现系统方面的持续挑战。

## 机器代理问题

想象一下，人们会如何对自己的机器人做出反应。他们很可能不会像对待正常人一样对待机器人。回想第 10.4 节，VR 中的社交互动取决于人们选择代表自己的化身。使用远程呈现，你就会被视为机器人化身，这会导致相同类型的社交问题[242]。这个不寻常的形象可能看起来有些残缺或尴尬，导致他人回避互动。不幸的是，与纯虚拟世界中的交互相比，选择自己想要看到的方式的自由性要少得多。如果是在公共平台上，你可能不得不被大家认为是尴尬的。社交机器人和人机交互研究可能有助于通过这种机器人化身来改善社交互动 [81,129,274]。

## 远程控制 vs 自动化

假设机器人可能在比匹配区域更大的区域漫游，那它就需要一种移动方法。这意味着

用户应通过一个接口控制机器人的运动。在 10.2 节中，运动被提出用于在大型虚拟世界中进行导航，并被解释为控制一辆大车的行为（图 10.5）。现实世界中机器人的行为几乎与纯虚拟世界中的大车类似；然而，仍存在一些区别：1) 机器人不能轻松地传送到另一个地点。但是，如果有很多可用的机器人，则可以连接到不同的机器人，这会感觉像传送给其他用户一样。2) 机器人根据其物理设计和环境受到限制。它可能有滚轮或步行腿，可能或不可能轻易地穿越部分环境。它也具有有限的行驶速度，转弯速度和电池寿命。3) 机器人撞击到人或障碍物通常会消耗大量成本。

对于远程操控机器人的用户，存在一系列选择。在一个极端情况下，用户可以连续控制运动，就像使用遥控器驱动无线电遥控车一样。延迟在一些应用中变得非常重要，特别是远程手术[188,355]。另一方面，用户可以简单地在地图上指出位置或使用虚拟激光指示器（第 10.2 节）指向可见位置。在这种情况下，机器人可以并带着用户一起自主执行所有的动作。这要求机器人具有更高的自主性，因为它必须规划自己的路线，以实现目标而不会遇到障碍；这在机器人学中被称为运动规划[163]。这使用户可以不用将注意力集中在机器人移动上，但是对于机器人平台和环境的某些组合下，可能难以获得可靠的性能。

### 虚拟现实不适问题

由于与运动的关系，虚拟现实中给人带来移动感觉的认知因素(vection)再次出现（8.4 节）。第 10.2 节中提到的许多建议可用于减少这种因素，例如在机器人移动时降低对比度或视野。现在考虑一些特定于机器人的建议。用户可能会觉得自己控制机器人更舒适，而不是让机器人提高自主性，尽管它涉及到长时间乏味的专注过程。此外，通过运动规划算法确定的路径本身可以通过缩短发生加速的时间或通过避免靠近具有高空间频率和对比度的墙壁或对象来优化以减少不适。另一个想法是在机器人移动时，从第三人称视角在 2D 或 3D 地图上显示动作。当机器人在移动时，任何东西可以想象性地被呈现给用户，例如新闻馈送。就虚拟世界的运动而言，必须注意不要因为未能提供相对于旧的位置和方向足够的信息，来在用户抵达前简单推断新的位置，而让用户感到迷惑。

### 延迟问题

如预期的那样，时间延迟威胁到网真系统的性能和舒适度。这种延迟已经在视觉渲染（7.4 节）和虚拟世界模拟（8.3.2 节）方面进行了讨论。网络系统会导致新的延迟被添加到 VR 系统，因为信息必须从客户端传输到服务器并再次返回。此外，带宽（位每秒）有所限制，这可能会导致进一步的延迟或质量下降。作为参考，2012 年谷歌的平均返程时间为 100 毫秒（美国为 50 至 60 毫秒）[219]。请注意，通过向用户传输整个全景视图，网络延迟不应影响头部跟踪和渲染延迟。

然而，延迟对交互性有重大影响，这对网络游戏玩家来说是一个众所周知的问题。另一方面，人们发现，在抱怨交谈困难之前，人们通常可以容忍长达 200 毫秒的电话延迟；然而，如果他们期望机器人立即回应他们的移动命令，就有可能会感到沮丧。由于手眼协调的延迟，完成一个操控任务将会更加困难。在某些情况下，假设延迟在试验期间没有显著变化[68]，人们可以通过适应来克服高延迟。延迟对远程医疗应用提出了相当大的挑战。想象一下，如果你是一位通过网真系统推动手术刀的医生，但在 500 毫秒期间内都看不到或感觉不到是否该停止切割。这可能太晚了！

## 13.4 人脑-机器接口

人类和机器之间的最终界面可以通过直接感知和刺激神经元实现。向这个方向迈出的

一步是提取第 12.3 节介绍的生理学测量。我们可以利用诸如心率，电流皮肤反应和苍白程度来动态调整 VR 体验，而不是用它们来研究 VR 不适。各种目标将得到优化，例如兴奋，恐惧，舒适或放松。进一步来看，我们可以应用读取神经元激活的技术，以便 VR 系统通过改变视觉和听觉显示来响应它。用户可以了解到某些想法在虚拟现实中具有相应的效果，从而进行精神控制。神经可塑性和知觉学习的能力（第 12.1 节）可以使他们在虚拟世界中舒适有效地移动他们的虚拟身体。这可能听起来像科幻小说，但已取得实质性进展。例如，最近，杜克大学的神经科学家们开始训练猴子驾驶轮椅，仅使用他们的想法 [259]。在人脑-机器接口领域（或者 BMI，人脑-电脑接口或 BCI），已经进行了许多其他的实验，他们通过自己的想法将人和动物连接到机械系统和 VR 体验[173,175,185]。这方面的调查包括[89,232,351]。

## 测量方法

测量神经活动的设备的目标是破译用户的意图和决定。它们通常分为两类：非侵入性（允许将传感器接触皮肤）和侵入性（允许钻入颅骨）。

首先考虑非侵入性情况，这是迄今为止最适合人类的情况。迄今为止测量全脑活动最准确的方法是通过功能磁共振成像（fMRI），如图 13.10 所示。这与大多数人熟悉的 MRI 作为常用的医学扫描方法有关。普通 MRI 的不同之处在于它提供了静态结构的图像来识别异常，而功能磁共振成像提供的图像显示了部分大脑随时间的活动。不幸的是，fMRI 作为 VR 接口的日常使用速度太慢，昂贵且麻烦[173]。此外，用户通常会摄入一种染料，这种染料会增加由于血流量变化造成的对比度，并且必须保持固定。

因此，测量 BMI 的大脑活动的最常用方法是通过脑电图（EEG），其包括沿着头皮放置电极以测量源于神经活动的电场波动；见图 13.11。不幸的是，由于大脑组织，骨骼和皮肤有效地进行低通滤波，破坏大部分信号导致信噪比很低，对其他神经结构也有显着的衰减和干扰。通过 EEG 的信息传输速率在每秒 5 到 25 比特之间[173,351]。这大约相当于每秒几个字符，比平均打字速度慢两个数量级。从 EEG 信号中提取信息涉及困难的信号处理[278]；已经存在开源库，例如 INRIA Rennes 的 OpenVibe。

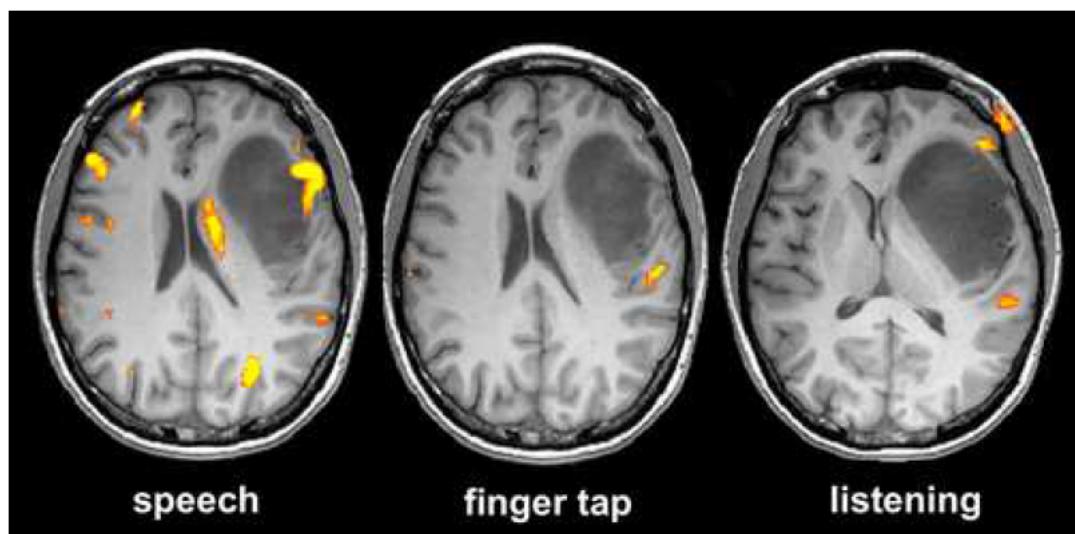


图 13.10：基于各种活动的 fMRI 扫描。（图源 Mayfield 的脑和脊椎）

对于侵入性病例，电极植入颅内（骨骼内）。这为科学家提供了更多的信息，但仅限于研究动物（以及一些患有帕金森病等神经疾病的人）。因此，侵入性方法不适合绝大多数人

作为 VR 接口。最简单的情况是为一个特定的神经元执行一个单位的记录；然而，这往往增加所需试验的次数，因为神经反应通常在不同试验中的不同神经元之间切换。随着神经元数量的增加，破译想法的问题变得更加可靠。许多记录可能来自执行已知功能的单个站点，或可能来自多个帮助理解大脑执行的分布式处理的站点 [173]。



图 13.11：EEG 系统将电极放置在颅骨周围：(a)一个颅骨帽，可以测量几十个信号。(b)无线情绪式 EEG 设备。

## 医疗目的

了解 VR 用户与 BMI 的主要群体之间的差异非常重要。BMI 领域发展迅速，因为它可能给患有神经肌肉残疾的人带来可移动性[351]。例如控制轮椅和仅利用思想移动假肢。第一个精神控制系统是由 Jacques Vidal 在 20 世纪 70 年代[331,332]建立的，从那时起许多系统都使用几种神经信号建立起来了。在所有情况下，操作这些接口都需要大量的训练和技能。作为治疗过程的一部分，运动障碍患者可能会有很强烈的动机利用日常时间进行训练，但大多数 VR 用户不会这样做。训练中一个有趣的问题是用户需要反馈，这才是 VR 的完美应用。VR 系统中的控制器基本上被分析神经信号的处理系统的输出所取代。因此，用户可以在接收来自 VR 系统的视觉反馈的同时练习移动虚拟轮椅或假肢。这可以防止他们在练习时伤害自己或损坏设备。

## 学习新的身体模式

当控制假肢时，人类对自己身体感知时会发生什么？身体内部的大脑表征被称为身体模式。这在一个世纪以前被提出[115]，当人们巧妙地使用工具时，身体模式会相应地适应，以便大脑感到有新的扩展身体在运作。这导致工具和手的感知同化，这已经从[131]中的神经信号得到证实。这为 VR 研究提出了一个耐人寻味的问题：我们的大脑通过不同的视觉身体表征（虚拟身体）以及运动和操纵的相互作用机制可以学习什么样的身体模式？

## VR 中的 BMI

在 VR 的背景下，大多数系统使用三种不同类型的 EEG 信号中的一种 [84,174,185,186,270]：1) 运动图像，2) SSVEP，3) P300。最常见的是运动图像，这是一个人在执行动作之前发生的心理过程。在此期间，人们会排练或模拟大脑中的运动，从而导致主要运动皮质中可测量的激活。用户可以想象旋转到位或在脚步中实现虚拟世界中的移动。不幸的是，大多数成功的系统仅限于一些简单的命令，例如开始和停止步行。尽管如此，用

户已经可以通过简单地想象动作去探索迷宫般的环境。

运动图像的一个优点是它确实需要来自系统的任何干扰或特殊刺激，从而允许用户在没有中断或特定计时的情况下继续进行。不幸的是另外两种信号需要产生刺激，然后通过 EEG 测量响应。其中之一是 SSVEP（稳态视觉诱发电位），当闪烁的视觉刺激呈现在 3.5 至 75Hz 范围内时发生。SSVEP 的信噪比非常高，用户可以通过对闪光的关注程度来影响结果。是否专注的决定被用作命令的基础。另一个信号是 P300，在出现罕见且相关的刺激后约 300ms 出现。这里再重复一次，它基于用户对刺激的关注程度来测量响应。

## 研究挑战

尽管 BMI 正在迅速成熟，但在能够广泛使用之前仍存在一些挑战：

- 保持非侵入性的测量神经信号的更好技术。理想情况下，人们希望测量数千个具有高信噪比的神经元的输出。功能磁共振成像近年来备受关注的一种替代方法是功能近红外光谱 (fNIRS)。这些信号可以与 EEG 结合使用以增强测量[134,227]。
- 用户可以指定的每秒多比特改进带宽，因此与利用身体动作或控制器相比，具有明显的优势。具有非侵入性 BMI 的 VR 系统通常每秒只提供一位，这是非常不足的[175]。
- 更好的分类技术，可以以更高的准确性和细节识别用户的意图和决定。现代机器学习方法可能有助于推动这一点。
- 在使用接口之前，大幅减少所需的训练量。如果相比于学习打字，它需要更多的工作量，那么就不太可能得到广泛应用。
- 更好地理解通过 VR 系统提供的反馈可以学习什么样的身体模式，以便大脑自然地接受虚拟身体。

因此，除了帮助患有运动障碍的人外，BMI 在达到科幻小说预期的精神控制水平之前还有很长的路要走。

## 大脑中的又一个“大脑”

为了建立一个广泛的，网络化的 VR 社会，在遥远的未来很容易考虑侵入性 BMI 的可能性。在继续推进之前，回顾第 12.4 节中对道德标准的讨论，并考虑这样的未来是否可取。假设除了测量神经输出，直接神经刺激也被使用。这将不需要在感官面前放置显示器。对于眼睛，信号可以直接发送到光感受器。这种技术被称为视网膜植入，并且已经存在来帮助盲人。同样，人工耳蜗可以帮助聋人听到东西。神经科学家们，例如斯坦福大学的 David Eagleman，甚至提议我们可以学习开发全新的感官。一个例子是通过将它们的频率，振幅和空间安排重新映射到身体上的其他集合体（如背部）来感知红外或无线电信号。神经可塑性的限制尚未完全被这样理解。

不考虑刺激受体，工程刺激甚至可以放在更高的神经水平。例如，如果可以直接刺激视神经，为什么还要刺激光感受器呢？这涉及到模拟神经节细胞进行的处理，这是具有挑战性的，但与刺激杆和锥体相比也会降低带宽要求。最终，直接的神经测量和刺激可能导致大脑中的又一个“大脑”，这在 1.1 节中已经提到。

大脑中是否还有一个“大脑”呢，我们不得而知。实际上，在这本 VR 书时，还有许多谜底等待着被揭晓。

## 参考文献

- [1] M. Abrash. Raster scan displays: More than meets the eye. Blog post. Retrieved from <http://blogs.valvesoftware.com/abrash/raster-scan-displays-morethan-meets-the-eye/>, January 2013. Last retrieved on Jan 10, 2016.
- [2] Z. M. Aghajan, L. Acharya, J. J. Moore, J. D. Cushman, C. Vuong, and M. R. Mehta. Impaired spatial selectivity and intact phase precession in two-dimensional virtual reality. *Nature Neuroscience*, 18(1):121–128, 2015.
- [3] Y. Akatsuka and G. A. Bekey. Compensation for end to end delays in a VR system. In *Proceedings IEEE Virtual Reality Annual International Symposium*, pages 156–159, 1998.
- [4] K. Akeley, S. J. Watt, A. Reza Girschick, and M. S. Banks. A stereo display prototype with multiple focal distances. *ACM Transactions on Graphics*, 23(3), 2004.
- [5] T. Akenine-Möller, E. Haines, and N. Hoffman. *Real-Time Rendering*. CRC Press, Boca Raton, FL, 2008.
- [6] D. Alais, C. Morrone, and D. Burr. Separate attentional resources for vision and audition. *Proceedings of the Royal Society B: Biological Sciences*, 273(1592):1339–1345, 2006.
- [7] B. B. Andersen, L. Korbo, and B. Pakkenberg. A quantitative study of the human cerebellum with unbiased stereological techniques. *Journal of Comparative Neurology*, 326(4):549–560, 1992.
- [8] J. Angeles. *Spatial Kinematic Chains. Analysis, Synthesis, and Optimisation*. Springer-Verlag, Berlin, 1982.
- [9] J. Angeles. *Rotational Kinematics*. Springer-Verlag, Berlin, 1989.
- [10] J. Angeles. *Fundamentals of Robotic Mechanical Systems: Theory, Methods, and Algorithms*. Springer-Verlag, Berlin, 2003.
- [11] A. Antoniou. *Digital Signal Processing: Signals, Systems, and Filters*. McGrawHill Education, Columbus, OH, 2005.
- [12] D. K. Arrowsmith and C. M. Place. *Dynamical Systems: Differential Equations, Maps, and Chaotic Behaviour*. Chapman & Hall/CRC, New York, 1992.
- [13] K. W. Arthur. Effects of Field of View on Performance with Head-Mounted Displays. PhD thesis, University of North Carolina at Chapel Hill, 2000.
- [14] K. J. Astrom and R. Murray. *Feedback Systems: An Introduction for Scientists and Engineers*. Princeton University Press, Princeton, NJ, 2008.
- [15] F. A. Azevedo, L. R. Carvalho, L. T. Grinberg, J. M. Farfel, R. E. Ferretti, R. E. Leite, W. Jacob Filho, R. Lent, and S. Herculano-Houzel. Equal numbers of neuronal and nonneuronal cells make the human brain an isometrically scaled-up primate brain. *Journal of Computational Neurology*, 513:532–541, 2009.
- [16] J. N. Bailenson, A. C. Beall, J. Loomis, J. Blascovich, and M. Turk. Transformed social interaction: Decoupling representation from behavior and form in collaborative virtual environments. *PRESSENCE: Teleoperators and Virtual Environments*, 13(4):428–441, 2004.
- [17] M. S. Banks, J. Kim, and T. Shibata. Insight into vergence-accommodation mismatch. In *Proceedings of SPIE*, 2013.
- [18] H. B. Barlow and R. M. Hill. Evidence for a physiological explanation of the waterfall illusion. *Nature*, 200:1345–1347, 1963.
- [19] H. H. Barrett and K. J. Myers. *Foundations of Image Science*. Wiley, Hoboken, NJ, 2004.

- [20] E. P. Becerra and M. A. Stutts. Ugly duckling by day, super model by night: The influence of body image on the use of virtual worlds. *Journal of Virtual Worlds Research*, 1(2):1–19, 2008.
- [21] C. Bergland. The wacky neuroscience of forgetting how to ride a bicycle. *Psychology Today*, May 2015. Posted online.
- [22] J. Birn. *Digital Lighting and Rendering*, 3rd Ed. New Riders, San Francisco, CA, 2013.
- [23] J. Blauert. *Spatial Hearing: Psychophysics of Human Sound Localization*. MIT Press, Boston, MA, 1996.
- [24] J. F. Blinn. Models of light reflection for computer synthesized pictures. In *Proceedings Annual Conference on Computer Graphics and Interactive Techniques*, 1977.
- [25] I. Bogost and N. Monfort. *Racing the Beam: The Atari Video Computer System*. MIT Press, Cambridge, MA, 2009.
- [26] S. J. Bolanowski, G. A. Gescheider, R. T. Verillo, and C. M. Checkosky. Four channels mediate the aspects of touch. *Journal of the Acoustical Society of America*, 84(5):1680–1694, 1988.
- [27] W. M. Boothby. *An Introduction to Differentiable Manifolds and Riemannian Geometry*. Revised 2nd Ed. Academic, New York, 2003.
- [28] D. Bordwell and K. Thompson. *Film History: An Introduction*, 3rd Ed. McGrawHill, New York, NY, 2010.
- [29] J. K. Bowmaker and H. J. A. Dartnall. Visual pigment of rods and cones in a human retina. *Journal of Physiology*, 298:501–511, 1980.
- [30] D. Bowman and L. Hodges. An evaluation of techniques for grabbing and manipulating remote objects in immersive virtual environments. In *Proceedings ACM Symposium on Interactive 3D Graphics*, pages 35–38, 1997.
- [31] D. A. Bowman, E. Kruijff, J. J. LaViola, and I. Poupyrev. *3D User Interfaces*. Addison-Wesley, Boston, MA, 2005.
- [32] K. Brown. Silent films: What was the right speed? *Sight and Sound*, 49(3):164–167, 1980.
- [33] M. Brown and D. G. Lowe. Automatic panoramic image stitching using invariant features. *International Journal of Computer Vision*, 74(1):59–73, 2007.
- [34] N. C. Burbules. Rethinking the virtual. In J. Weiss, J. Nolan, and P. Trifonas, editors, *The International Handbook of Virtual Learning Environments*, pages 3–24. Kluwer Publishers, Dordrecht, 2005.
- [35] D. C. Burr, M. C. Morrone, and L. M. Vaina. Large receptive fields for optic flow detection in humans. *Vision Research*, 38(12):1731–1743, 1998.
- [36] W. S. Cain. Odor identification by males and females: predictions vs performance. *Chemical Senses*, 7(2):129–142, 1994.
- [37] P. Cairns and A. L. Cox. *Research Methods for Human-Computer Interaction*. Cambridge University Press, Cambridge, U.K., 2008.
- [38] F. W. Campbell and D. G. Green. Optical and retinal factors affecting visual resolution. *Journal of Physiology*, 181:576–593, 1965.
- [39] S. K. Card, W. K. English, and B. J. Burr. Evaluation of mouse, rate-controlled isometric joystick, step keys, and text keys for text selection on a CRT. *Ergonomics*, 20:601–613, 1978.
- [40] J. M. Carroll. *HCI Models, Theories, and Frameworks: Toward a Multidisciplinary Science*. Morgan Kaufmann, San Francisco, CA, 2003.
- [41] E. Catmull. A subdivision algorithm for computer display of curved surfaces. PhD thesis, University of Utah, 1974.

- [42] A. Y. Chang. A survey of geometric data structures for ray tracing. Technical Report TR-CIS-2001-06, Brooklyn Polytechnic University, 2001.
- [43] N. Chaudhari, A. M. Landin, and S. D. Roper. A metabotropic glutamate receptor variant functions as a taste receptor. *Nature Neuroscience*, 3(3):113–119, 2000.
- [44] G. Chen, J. A. King, N. Burgess, and J. O’Keefe. How vision and movement combine in the hippocampal place code. *Proceedings of the National Academy of Science USA*, 110(1):378–383, 2013.
- [45] C. K. Chui and G. Chen. *Kalman Filtering*. Springer-Verlag, Berlin, 1991.
- [46] D. Claus and A. W. Fitzgibbon. A rational function lens distortion model for general cameras. In *Proc. Computer Vision and Pattern Recognition*, pages 213–219, 2005.
- [47] E. Cline. *Ready Player One*. Random House, 2011.
- [48] D. Cox, J. Little, and D. O’Shea. *Ideals, Varieties, and Algorithms*. SpringerVerlag, Berlin, 1992.
- [49] C. Cruz-Neira, D. J. SAndin, T. A. DeFanti, R. V. Kenyon, and J. C. Hart. The CAVE: Audio visual experience automatic virtual environment. *Communications of the ACM*, 35(6):64–72, 1992.
- [50] H. Culbertson, J. J. Lopez Delgado, and K. J. Kuchenbecker. One hundred datadriven haptic texture models and open-source methods for rendering on 3D objects. In *Proceedings IEEE Haptics Symposium*, pages 319–325, 2014.
- [51] C. A. Curcio, K. R. Sloan, R. E. Kalina, and A. E. Hendrickson. Human photoreceptor topography. *Journal of Comparative Neurobiology*, 292:497–523, 1990.
- [52] R. P. Darken and B. Peterson. Spatial orientation, wayfinding, and representation. In K. S. Hale and K. M. Stanney, editors, *Handbook of Virtual Environments*, 2<sup>nd</sup> Edition, pages 131–161. CRC Press, Boca Raton, FL, 2015.
- [53] R. Darwin. New experiments on the ocular spectra of light and colours. *Philosophical Transactions of the Royal Society of London*, 76:313–348, 1786.
- [54] M. de Berg, M. van Kreveld, M. Overmars, and O. Schwarzkopf. *Computational Geometry: Algorithms and Applications*, 2nd Ed. Springer-Verlag, Berlin, 2000.
- [55] K. N. de Winkel, M. KAtliar, and H. H. B”ulthoff. Forced fusion in multisensory heading estimation. *PloS ONE*, 10(5), 2015.
- [56] J. R. DeJong. The effects of increasing skill on cycle time and its consequences for time standards. *Ergonomics*, 1(1):51–60, 1957.
- [57] J. Delwiche. The impact of perceptual interactions on perceived flavor. *Food Quality and Perferences*, 15, 137–146.
- [58] J. L. Demer, J. Goldberg, H. A. Jenkins, and F. I. Porter. Vestibulo-ocular reflex during magnified vision: Adaptation to reduce visual-vestibular conflict. *Aviation, Space, and Environmental Medicine*, 58(9 Pt 2):A175–A179, 1987.
- [59] M. Dennison, Z. Wisti, and M. D’Zmura. Use of physiological signals to predict cybersickness. *Displays*, 44:52–52, 2016.
- [60] D. Deutsch, T. Hamaoui, and T. Henthorn. The glissando illusion and handedness. *Neuropsychologia*, 45:2981–2988, 2007.
- [61] P. DiZio, J. R. Lackner, and R. K. Champney. Proprioceptive adaptation and aftereffects. In K. S. Hale and K. M. Stanney, editors, *Handbook of Virtual Environments*, 2nd Edition. CRC Press, Boca Raton, FL, 2015.
- [62] M. H. Draper, E. S. Viire, and T. A. Furness amd V. J. Gawron. Effects of image scale and

- system time delay on simulator sickness with head-coupled virtual environments. *Human Factors*, 43(1):129–146, 2001.
- [63] A. T. Duchowski. *Eye Tracking Methodology: Theory and Practice*, 2nd Ed. Springer-Verlag, Berlin, 2007.
- [64] G. Dudek, P. Giguere, C. Prahacs, S. Saunderson, J. Sattar, L.-A. Torres-Mendez, M. Jenkin, A. German, A. Hogue, A. Ripsman, J. Zacher, E. Milios, H. Liu, P. Zhang, M. Buehler, and C. Georgiades. Aqua: An amphibious autonomous robot. *IEEE Computer Magazine*, 40(1):46–53, 2007.
- [65] R. L. Doty (Ed.). *Handbook of Olfaction and Gustation*, 3rd Ed. Wiley-Blackwell, Hoboken, NJ, 2015.
- [66] H. H. Ehrsson, C. Spence, and R. E. Passingham. That's my hand! Activity in premotor cortex reflects feeling of ownership of a limb. *Science*, 305(5685):875–877, 2004.
- [67] S. R. Ellis, K. Mania, B. D. Adelstein, and M. I. Hill. Generalizeability of latency detection in a variety of virtual environments. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, pages 2632–2636, 2004.
- [68] S. R. Ellis, M. J. Young, B. D. Adelstein, and S. M. Ehrlich. Discrimination of changes in latency during head movement. In *Proceedings Computer Human Interfaces*, pages 1129–1133, 1999.
- [69] M. Emoto, K. Masaoka, M. Sugawara, and F. Okano. Viewing angle effects from wide field video projection images on the human equilibrium. *Displays*, 26(1):9–14, 2005.
- [70] D. J. Encross. Control of skilled movement. *Psychological Bulletin*, 84:14–29, 1977.
- [71] R. Engbert and K. Mergenthaler. Mircosaccades are triggered by low retinal image slip. *Proceedings of the National Academy of Sciences of the United States of America*, 103(18):7192–7197, 2008.
- [72] B. W. Epps. Comparison of six cursor control devices based on Fitts' law models. In *Proceedings of the 30th Annual Meeting of the Human Factors Society*, pages 327–331, 1986.
- [73] C. J. Erkelens. Coordination of smooth pursuit and saccades. *Vision Research*, 46(1–2):163–170, 2006.
- [74] D. Fattal, Z. Peng, T. Tran, S. Vo, M. Fiorentino, J. Brug, and R. G. Beausoleil. A multi-directional backlight for a wide-angle, glasses-free three-dimensional display. *Nature*, 495:348–351, 2013.
- [75] J. Favre, B. M. Jolles, O. Siegrist, and K. Aminian. Quaternion-based fusion of gyroscopes and accelerometers to improve 3D angle measurement. *Electronics Letters*, 32(11):612–614, 2006.
- [76] G. T. Fechner. *Elements of Psychophysics* (in German). Breitkopf and H̄artel, Leipzig, 1860.
- [77] M. A. Fischler and R. C. Bolles. Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography. *Communications of the ACM*, 24(6):381–395, 1981.
- [78] P. M. Fitts. The information capacity of the human motor system in controlling the amplitude of movement. *Journal of Experimental Psychology*, 47(6):381–391, 1956.
- [79] R. C. Fitzpatrick and B. L. Day. Probing the human vestibular system with galvanic stimulation. *Journal of Applied Physiology*, 96(6):2301–2316, 2004.
- [80] R. C. Fitzpatrick, J. Marsden, S. R. Lord, and B. L. Day. Galvanic vestibular stimulation evokes sensations of body rotation. *NeuroReport*, 13(18):2379–2383, 2002.
- [81] T. Fong, I. Nourbakhsh, and K. Dautenhahn. A survey of socially interactive robots: Concepts,

- design, and applications. *Robotics and Autonomous Systems*, 42(3-4):143–166, 2003.
- [82] W. T. Fong, S. K. Ong, and A. Y. C. Nee. Methods for in-field user calibration of an inertial measurement unit without external equipment. *Measurement Science and Technology*, 19(8), 2008.
- [83] A. K. Forsberg, K. Herndon, and R. Zelznik. Aperture based selection for immersive virtual environments. In Proceedings ACM Symposium on User Interface Software and Technology, pages 95–96, 1996.
- [84] D. Friedman, R. Leeb, C. Guger, A. Steed, G. Pfurtscheller, and M. Slater. Navigating virtual reality by thought: What is it like? *Presence: Teleoperators and Virtual Environments*, 16(1):100–110, 2007.
- [85] H. Fuchs, Z. M. Kedem, and B. F. Naylor. On visible surface generation by a priori tree structures. In Proceedings ACM SIGGRAPH, pages 124–133, 1980.
- [86] J. Fuentes-Pacheco, J. Ruiz-Ascencio, and J. M. Rendon-Mancha. Visual simultaneous localization and mapping: a survey. *Journal Artificial Intelligence Review*, 43(1):55–81, 2015.
- [87] T. Funkhouser, I. Carlstrom, G. Elko, G. Pingali, M. Sondhi, and J. West. A beam tracing approach to acoustic modeling for interactive virtual environments. In Proceedings ACM Annual Conference on Computer Graphics and Interactive Techniques, pages 21–32, 1998.
- [88] J. Gallier. Curves and Surfaces in Geometric Modeling. Morgan Kaufmann, San Francisco, CA, 2000.
- [89] S. Gao, Y. Wang, X. Gao, and B. Hong. Visual and auditory brain-computer interfaces. *IEEE Transactions on Biomedical Engineering*, 61(5):1436–1447, 2014.
- [90] M. A. Garcia-Perez. Forced-choice staircases with fixed step sizes: asymptotic and small-sample properties. *Vision Research*, 38(12):1861–81, 1998.
- [91] G. M. Gauthier and D. A. Robinson. Adaptation of the human vestibuloocular reflex to magnifying lenses. *Brain Research*, 92(2):331–335, 1975.
- [92] D. Gebre-Egziabher, G. Elkaim, J. David Powell, and B. Parkinson. Calibration of strapdown magnetometers in magnetic field domain. *Journal of Aerospace Engineering*, 19(2):87–102, 2006.
- [93] G. Gescheider. Psychophysics: The Fundamentals, 3rd Ed. Lawrence Erlbaum Associates, Mahwah, NJ, 2015.
- [94] E. Gibson. Principles of Perceptual Learning and Development. Appleton-CenturyCrofts, New York, 1969.
- [95] W. Gibson. Neuromancer. Ace Books, 1984.
- [96] W. C. Gogel. An analysis of perceptions from changes in optical size. *Perception and Psychophysics*, 60(5):805–820, 1998.
- [97] E. B. Goldstein. Sensation and Perception, 9th Ed. Wadsworth, Belmont, CA, 2014.
- [98] R. L. Goldstone. Perceptual learning. *Annual Review of Psychology*, 49:585–612, 1998.
- [99] A. Gopnik, A. N. Meltzoff, and P. K. Kuhl. The Scientist in the Crib: What Early Learning Tells Us About the Mind. HarperCollins, New York, NY, 2000.
- [100] S. Gottschalk, M. C. Lin, and D. Manocha. Obbtree: A hierarchical structure for rapid interference detection. In Proceedings ACM SIGGRAPH, 1996.
- [101] A. C. Grant, M. C. Thiagarajah, and K. Sathian. Tactile perception in blind Braille readers: A psychophysical study of acuity and hyperacuity using gratings and dot patterns. *Perception and Psychophysics*, 62(2):301–312, 2000.
- [102] A. Graybiel and J. Knepton. Sopite syndrome - sometimes sole manifestation of motion sickness. *Aviation, Space, and Environmental Medicine*, 47(8):873–882, 1976.

- [103] J. Gregory. Game Engine Architecture, 2nd Ed. CRC Press, Boca Raton, FL, 2014.
- [104] J. E. Greivenkamp. Field Guide to Geometrical Optics. SPIE Press, Bellingham, WA, 2004.
- [105] B. Guentner, M. Finch, S. Drucker, D. Tan, and J. Snyder. Foveated 3D graphics. Technical report, Microsoft Research, 2012. Available at <http://research.microsoft.com/>.
- [106] P. Guigue and O. Devillers. Fast and robust triangle-triangle overlap test using orientation predicates. *Journal of Graphics Tools*, 8(1):25–32, 2003.
- [107] A. Guterstam, V. I. Petkova, and H. H. Ehrsson. The illusion of owning a third arm. *PloS ONE*, 6(2), 2011.
- [108] K. S. Hale and K. M. Stanney. Handbook of Virtual Environments, 2nd Edition. CRC Press, Boca Raton, FL, 2015.
- [109] G. Hall. Perceptual and Associative Learning. Oxford University Press, Oxford, UK, 1991.
- [110] R. S. Hartenberg and J. Denavit. A kinematic notation for lower pair mechanisms based on matrices. *Journal of Applied Mechanics*, 77:215–221, 1955.
- [111] R. I. Hartley and A. Zisserman. Multiple View Geometry in Computer Vision, 2nd Ed. Cambridge University Press, Cambridge, U.K., 2004.
- [112] C. D. Harvey, F. Collman, D. A. Dombeck, and D. W. Tank. Intracellular dynamics of hippocampal place cells during virtual navigation. *Nature*, 461:941–946, 2009.
- [113] J. O. Harvey. Efficient estimation of sensory thresholds with ml-pest. *Spatial Vision*, 11(1):121–128, 1997.
- [114] K. Hashimoto, Y. Maruno, and T. Nakamoto. Brief demonstration of olfactory and visual presentation using wearable olfactory display and head mounted display. In Proceedings IEEE Virtual Reality Conference, page Abstract, 2016.
- [115] H. Head and G. Holmes. Sensory disturbances from cerebral lesion. *Brain*, 34(2-3):102–254, 1911.
- [116] E. G. Heckenmueller. Stabilization of the retinal image: A review of method, effects, and theory. *Psychological Bulletin*, 63:157–169, 1965.
- [117] J. Heikkilä. Geometric camera calibration using circular control points. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(10):1066–1077, 2000.
- [118] J. Heikkilä and O. Silvén. A four-step camera calibration procedure with implicit image correction. In Proc. Computer Vision and Pattern Recognition, pages 1106–1112, 1997.
- [119] W. T. Higgins. A comparison of complementary and Kalman filtering. *IEEE Transactions on Aerospace and Electronic Systems*, 11(3):321–325, 1975.
- [120] J. M. Hillis, M. O. Ernst, M. S. Banks, and M. S. Landy. Combining sensory information: mandatory fusion within, but not between, senses. *Science*, 298(5098):1627–30, 2002.
- [121] P. Hoberman, D. M. Krum, E. A. Suma, and M. Bolas. Immersive training games for smartphone-based head mounted displays. In IEEE Virtual Reality Short Papers and Posters, 2012.
- [122] J. G. Hocking and G. S. Young. Topology. Dover, New York, 1988.
- [123] C. M. Hoffmann. Geometric and Solid Modeling. Morgan Kaufmann, San Francisco, CA, 1989.
- [124] R. V. Hogg, J. McKean, and A. T. Craig. Introduction to Mathematical Statistics, 7th Ed. Pearson, New York, NY, 2012.
- [125] M. Hollins, M. H. Buonocore, and G. R. Mangun. The neural mechanisms of top-down attentional control. *Nature Neuroscience*, 3(3):284–291, 2002.

- [126] G. C. Holst and T. S. Lomheim. CMOS/CCD Sensors and Camera Systems. SPIE Press, Bellingham, WA, 2011.
- [127] X. Hu and H. Hua. Design and assessment of a depth-fused multi-focal-plane display prototype. *Journal of Display Technology*, 10(4):308–316, 2014.
- [128] A. S. Huang, A. Bachrach, P. Henry, M. Krainin, D. Maturana, D. Fox, and N. Roy. Visual odometry and mapping for autonomous flight using an RGB-D camera. In *Proceedings International Symposium on Robotics Research*, 2011.
- [129] C.-M. Huang and B. Mutlu. The repertoire of robot behavior: Enabling robots to achieve interaction goals through social behavior. *Journal of Human-Robot Interaction*, 2(2), 2013.
- [130] W. Hugemann. Correcting lens distortions in digital photographs. In *European Association for Accident Research and Analysis (EVU) Conference*, 2010.
- [131] A. Iriki, M. Tanaka, and Y. Iwamura. Coding of modified body schema during tool use by macaque postcentral neurones. *Neuroreport*, 7(14):2325–2330, 1996.
- [132] J. A. Irwin. The pathology of sea-sickness. *The Lancet*, 118(3039):907–909, 1878.
- [133] A. Iserles. *A First Course in the Numerical Analysis of Differential Equations*, 2nd Ed. Cambridge University Press, Cambridge, U.K., 2008.
- [134] M. Izzetoglu, K. Izzetoglu, S. Bunce, H. Ayaz, A. Devaraj, B. Onaral, and K. Pourrazaei. Functional near-infrared neuroimaging. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 13(2):153–159, 2005.
- [135] J. Jerald. *The VR Book*. Association of Computer Machinery and Morgan & Claypool Publishers, 2015.
- [136] D. L. Jones, S. Dechmerowski, R. Oden, V. Lugo, J. Wang-Costello, and W. Pike. Olfactory interfaces. In K. S. Hale and K. M. Stanney, editors, *Handbook of Virtual Environments*, 2nd Edition, pages 131–161. CRC Press, Boca Raton, FL, 2015.
- [137] N. P. Jouppi and S. Thomas. Telepresence systems with automatic preservation of user head height, local rotation, and remote translation. In Proc. IEEE International Conference on Robotics and Automation, pages 62–68, 2005.
- [138] M. Kaluzhna, M. Prsa, S. Gale, S. J. Lee, and O. Blanke. Learning to integrate contradictory multisensory self-motion cue pairings. *Journal of Vision*, 15(10), 2015.
- [139] M. Kalloniatis and C. Luu. Visual acuity. In H. Kolb, R. Nelson, E. Fernandez, and B. Jones, editors, *Webvision: The Organization of the Retina and Visual System*. 2007. Last retrieved on October 18, 2016.
- [140] R. Kalman. A new approach to linear filtering and prediction problems. *Transactions of the ASME, Journal of Basic Engineering*, 82:35–45, 1960.
- [141] H. Kato and M. Billinghurst. Marker tracking and hmd calibration for a videobased augmented reality conferencing system. In *Proceedings of IEEE and ACM International Workshop on Augmented Reality*, 1999.
- [142] D. Katz. Der aufbau der tastwelt. *Zeitschrift für Psychologie, Ergänzungsband* 11, 1925.
- [143] R. S. Kennedy and L. H. Frank. A review of motion sickness with special reference to simulator sickness. Technical Report NAVTRAEEQUIPCEN 81-C-0105-16, United States Navy, 1985.
- [144] R. S. Kennedy, N. E. Lane, K. S. Berbaum, and M. G. Lilienthal. Simulator sickness questionnaire: An enhanced method for quantifying simulator sickness. *International Journal of Aviation Psychology*, 3(3):203–220, 1993.

- [145] B. Keshavarz, H. Hecht, and B. D. Lawson. Visually induced motion sickness: Causes, characteristics, and countermeasures. In K. S. Hale and K. M. Stanney, editors, *Handbook of Virtual Environments*, 2nd Edition, pages 647–698. CRC Press, Boca Raton, FL, 2015.
- [146] B. Keshavarz, B. E. Riecke, L. J. Hettinger, and J. L. Campos. Vection and visually induced motion sickness: how are they related? *Frontiers in Psychology*, 6(472), 2015.
- [147] B. Keshavarz, D. Stelzmann, A. Paillard, and H. Hecht. Visually induced motion sickness can be alleviated by pleasant odors. *Experimental Brain Research*, 233:1353–1364, 2015.
- [148] W. Khalil and J. F. Kleinfinger. A new geometric notation for open and closed-loop robots. In *Proceedings IEEE International Conference on Robotics & Automation*, volume 3, pages 1174–1179, 1986.
- [149] D. O. Kim, C. E. Molnar, and J. W. Matthews. Cochlear mechanics: Nonlinear behaviour in two-tone responses as reflected in cochlear-new-fibre responses and in ear-canal sound pressure. *Journal of the Acoustical Society of America*, 67(5):1704–1721, 1980.
- [150] H. Kingma and M. Janssen. Biophysics of the vestibular system. In A. M. Bronstein, editor, *Oxford Textbook of Vertigo and Imbalance*. Oxford University Press, Oxford, UK, 2013.
- [151] C. L. Kinsey. *Topology of Surfaces*. Springer-Verlag, Berlin, 1993.
- [152] R. E. Kirk. *Experimental Design*, 4th Ed. Sage, Thousand Oaks, CA, 2013.
- [153] E. M. Kolasinski. Simulator sickness in virtual environments. Technical Report 2017, U.S. Army Research Institute, 1995.
- [154] L. L. Kontsevich and C. W. Tyler. Bayesian adaptive estimation of psychometric slope and threshold. *Vision Research*, 39(16):2729–2737, 1999.
- [155] C. Konvalin. Compensating for tilt, hard-iron, and soft-iron effects. Available at <http://www.sensorsmag.com/sensors/motion-velocitydisplacement/compensating-tilt-hard-iron-and-soft-iron-effects-6475>, December 2009. Last retrieved on May 30, 2016.
- [156] B. C. Kress and P. Meyrueis. *Applied Digital Optics: From Micro-optics to Nanophotonics*. Wiley, Hoboken, NJ, 2009.
- [157] J. B. Kuipers. *Quaternions and Rotation Sequences*. Princeton University Press, Princeton, NJ, 1999.
- [158] P. R. Kumar and P. Varaiya. *Stochastic Systems*. Prentice-Hall, Englewood Cliffs, NJ, 1986.
- [159] R. Lafer-Sousa, K. L. Hermann, and B. R. Conway. Striking individual differences in color perception uncovered by the dress photograph. *Current Biology*, 25(13):R545–R546, 2015.
- [160] M. F. Land and S.-E. Nilsson. *Animal Eyes*. Oxford University Press, Oxford, UK, 2002.
- [161] D. Lanman and D. Luebke. Near-eye light field displays. *ACM Transactions on Graphics*, 32(6), 2013.
- [162] J. Lanman, E. Bizzi, and J. Allum. The coordination of eye and head movement during smooth pursuit. *Brain Research*, 153(1):39–53, 1978.
- [163] S. M. LaValle. *Planning Algorithms*. Cambridge University Press, Cambridge, U.K., 2006. Available at <http://planning.cs.uiuc.edu/>.
- [164] S. M. LaValle. Help! My cockpit is drifting away. Oculus blog post. Retrieved from <https://developer.oculus.com/blog/magnetometer/>, December 2013. Last retrieved on Jan 10, 2016.
- [165] S. M. LaValle. The latent power of prediction. Oculus blog post. Retrieved from <https://developer.oculus.com/blog/the-latent-power-of-prediction/>, July 2013. Last retrieved on Jan 10, 2016.
- [166] S. M. LaValle. Sensor fusion: Keeping it simple. Oculus blog post. Retrieved from

<https://developer.oculus.com/blog/sensor-fusion-keeping-it-simple/>, May 2013. Last retrieved on Jan 10, 2016.

- [167] S. M. LaValle and P. Giokaris. Perception based predictive tracking for head mounted displays. US Patent 20140354515A1, December 2014.
- [168] S. M. LaValle, A. Yershova, M. Katsev, and M. Antonov. Head tracking for the Oculus Rift. In Proc. IEEE International Conference on Robotics and Automation, pages 187–194, 2014.
- [169] J. J. LaViola. A discussion of cybersickness in virtual environments. ACM SIGCHI Bulletin, 32:47–56, 2000.
- [170] B. D. Lawson. Motion sickness scaling. In K. S. Hale and K. M. Stanney, editors, Handbook of Virtual Environments, 2nd Edition, pages 601–626. CRC Press, Boca Raton, FL, 2015.
- [171] B. D. Lawson. Motion sickness symptomatology and origins. In K. S. Hale and K. M. Stanney, editors, Handbook of Virtual Environments, 2nd Edition, pages 531–600. CRC Press, Boca Raton, FL, 2015.
- [172] D. Lazewatsky and W. Smart. An inexpensive robot platform for teleoperation and experimentation. In Proc. IEEE International Conference on Robotics and Automation, pages 1211–1216, 2011.
- [173] M. A. Lebedev and M. A. L. Nicolelis. Brain-machine interfaces: Past, present, and future. TRENDS in Neurosciences, 29(9):536–546, 2006.
- [174] A. Lecuyer, L. George, and M. Marchal. Toward adaptive vr simulators combining visual, haptic, and brain-computer interfaces. In IEEE Computer Graphics and Applications, pages 3318–3323, 2013.
- [175] A. L'ecuyer, F. Lotte, R. B. Reilly, R. Leeb, M. Hirose, and N. Slater. Braincomputer interfaces, virtual reality, and videogames. IEEE Computer, 41(10):66–72, 2008.
- [176] M. R. Leek. Adaptive procedures in psychophysical research. Perception and Psychophysics, 63(8):1279–1292, 2001.
- [177] R. J. Leigh and D. S. Zee. The Neurology of Eye Movements, 5th Ed. Oxford University Press, 2015.
- [178] J.-C. Lepecq, I. Giannopulu, and P.-M. Baudonnier. Cognitive effects on visually induced body motion in children. Perception, 24(4):435–449, 1995.
- [179] J.-C. Lepecq, I. Giannopulu, S. Mertz, and P.-M. Baudonnier. Vestibular sensitivity andvection chronometry along the spinal axis in erect man. Perception, 28(1):63–72, 1999.
- [180] H. Li, L. Trutoiu, K. Olszewski, L. Wei, T. Trutna, P.-L. Hsieh, A. Nicholls, and C. Ma. Facial performance sensing head mounted display. In Proceedings ACM SIGGRAPH, 2015.
- [181] M. C. Lin and J. F. Canny. Efficient algorithms for incremental distance computation. In Proceedings IEEE International Conference on Robotics & Automation, 1991.
- [182] M. C. Lin and D. Manocha. Collision and proximity queries. In J. E. Goodman and J. O'Rourke, editors, Handbook of Discrete and Computational Geometry, 2<sup>nd</sup> Ed., pages 787–807. Chapman and Hall/CRC Press, New York, 2004.
- [183] J. Linowes. Unity Virtual Reality Projects. Packt, Birmingham, UK, 2015.
- [184] S. Liversedge, I. Gilchrist, and S. Everling (eds). Oxford Handbook of Eye Movements. Oxford University Press, 2011.
- [185] F. Lotte, J. Faller, C. Guger, Y. Renard, G. Pfurtscheller, A. L'ecuyer, and R. Leeb. Combining BCI with virtual reality: Towards new applications and improved BCI. In B. Z. Allison, S. Dunne, R. Leeb, J. Del R. Mill'an, and A. Nijholt, editors, Towards Practical Brain-Computer Interfaces,

- pages 197–220. SpringerVerlag, Berlin, 2012.
- [186] F. Lotte, A. van Langenhove, F. Lamarche, T. Ernest, Y. Renard, B. Arnaldi, and A. L’ecuyer. Exploring large virtual environments by thoughts using a braincomputer interface based on motor imagery and high-level commands. *Presence: Teleoperators and Virtual Environments*, 19(1):154–170, 2010.
- [187] G. D. Love, D. M. Hoffman, P. J. H. Hands, J. Gao, A. K. Kirby, and M. S. Banks. High-speed switchable lens enables the development of a volumetric stereoscopic display. *Optics Express*, 17(18):15716–15725, 2009.
- [188] M. J. Lum, J. Rosen, H. King, D. C. Friedman, T. S. Lendvay, A. S. Wright, M. N. Sinanan, and B. Hannaford. Telepresence systems with automatic preservation of user head height, local rotation, and remote translation. In Proc. IEEE Conference on Engineering in MEdicine and Biology Society, pages 6860–6863, 2009.
- [189] R. G. Lyons. Understanding Digital Signal Processing, 3rd Ed. Prentice-Hall, Englewood Cliffs, NJ, 2010.
- [190] K. Y. Ma, P. Chirarattananon, and R. J. Wood. Design and fabrication of an insect-scale flying robot for control autonomy. In Proc. IEEE/RSJ International Conference on Intelligent Robots and Systems, pages 1133–1140, 2012.
- [191] Y. Ma, S. Soatto, J. Kosecka, and S. S. Sastry. An Invitation to 3-D Vision. Springer-Verlag, Berlin, 2003.
- [192] I. S. MacKenzie. Fitts’ law as a research and design tool in human-computer interaction. *Human-Computer Interaction*, 7(1):91–139, 1992.
- [193] I. S. Mackenzie. Movement time prediction in human-computer interfaces. In R. M. Baecker, J. Grudin, W. A. S. Buxton, and S. Greenberg, editors, *Readings in Human-Computer Interaction*, pages 483–492. Morgan Kaufmann, San Francisco, 1995.
- [194] I. S. MacKenzie and W. Buxton. Extending Fitts Law to 2D tasks. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pages 219–226, 1992.
- [195] N. A. Macmillan and C. D. Creelman. Dection Theory: A User’s Guide, 2nd Ed. Lawrence Erlbaum Associates, Mahwah, NJ, 2005.
- [196] R. Magill and 10th Ed. D. Anderson. Motor Learning and Control: Concepts and Applications. McGraw-Hill, New York, NY, 2013.
- [197] R. Mahoney, T. Hamel, and J.-M. Pfimlin. Nonlinear complementary filters on the special orthogonal group. *IEEE Transactions on Automatic Control*, 53(5):1203–1218, 2008.
- [198] A. Maimone, D. Lanman, K. Rathinavel, K. Keller, D. Luebke, and H. Fuchs. Pinlight displays: Wide field of view augmented-reality eyeglasses using defocused point light sources. *ACM Transactions on Graphics*, 33(4), 2014.
- [199] K. Mallon and P. F. Whelan. Precise radial un-distortion of images. In Proc. Computer Vision and Pattern Recognition, pages 18–21, 2004.
- [200] K. Mania, B. D. Adelstein, S. R. Ellis, and M. I. Hill. Perceptual sensitivity to head tracking latency in virtual environments with varying degrees of scene complexity. In Proceedings of Symposium on Applied Perception in Graphics and Visualization, pages 39–47, 2004.
- [201] W. R. Mark, L. McMillan, and G. Bishop. Post-rendering 3D warping. In Proceedings of the Symposium on Interactive 3D Graphics, pages 7–16, 1997.
- [202] S. Marschner and P. Shirley. Fundamentals of Computer Graphics, 4th Ed. CRC Press, Boca Raton, FL, 2015.

- [203] M. T. Mason. *Mechanics of Robotic Manipulation*. MIT Press, Cambridge, MA, 2001.
- [204] G. Mather. *Foundations of Sensation and Perception*. Psychology Press, Hove, UK, 2008.
- [205] G. Mather, F. Verstraten, and S. Anstis. The motion aftereffect: A modern perspective. MIT Press, Boston, MA, 1998.
- [206] M. E. McCauley and T. J. Sharkey. Cybersickness: Perception of self-motion in virtual environments. *Presence*, 1(3):311–318, 1992.
- [207] H. McGurk and J. MacDonald. Hearing lips and seeing voices. *Nature*, 264:746–748, 1976.
- [208] R. Mehra, N. Raghuvanshi, L. Antani, A. Chandak, S. Curtis, and D. Manocha. Wave-based sound propagation in large open scenes using an equivalent source formulation. *ACM Transactions on Graphics*, 32(2), 2013.
- [209] J. Merimaa and V. Pulkki. Spatial impulse response rendering I: Analysis and synthesis. *Journal of the Audio Engineering Society*, 53(12):1115–1127, 2005.
- [210] P. R. Messinger, E. Stroulia, K. Lyons, M. Bone, R. H. Niu, K. Smirnov, and S. Perelgut. Virtual worldspast, present, and future: New directions in social computing. *Decision Support Systems*, 47(3):204–228, 2009.
- [211] A. Mikami, W. T. Newsome, and R. H. Wurtz. Motion selectivity in macaque visual cortex. II. Spatiotemporal range of directional interactions in MT and V1. *Journal of Neurophysiology*, 55:1328–1339, 1986.
- [212] M. Mine and G. Bishop. Just-in-time pixels. Technical Report TR93-005, University of North Carolina, Chapel Hill, NC, 1993.
- [213] M. Minsky. Telepresence. *Omni magazine*, pages 44–52, June 1980.
- [214] B. Mirtich. V-Clip: Fast and robust polyhedral collision detection. Technical Report TR97-05, Mitsubishi Electronics Research Laboratory, 1997.
- [215] B. Mirtich. Efficient algorithms for two-phase collision detection. In K. Gupta and A.P. del Pobil, editors, *Practical Motion Planning in Robotics: Current Approaches and Future Directions*, pages 203–223. Wiley, Hoboken, NJ, 1998.
- [216] T. Moller. A fast triangle-triangle intersection test. *Journal of Graphics Tools*, 2(2):25–30, 1997.
- [217] T. Moller and N. Trumbore. Fast, minimum storage ray/triangle intersection. *Journal of Graphics Tools*, 2(1):21–28, 1997.
- [218] B. Moore. *An Introduction to the Psychology of Hearing*, 6th Ed. Brill, Somerville, MA, 2012.
- [219] G. Morrison. Input lag: How important is it? CNET, June 2013. Posted online at <https://www.cnet.com/news/input-lag-how-important-is-it/>.
- [220] H. S. Mortensen, B. Pakkenberg, M. Dam, R. Dietz, C. Sonne, B. Mikkelsen, and N. Eriksen. Quantitative relationships in delphinid neocortex. *Frontiers in Neuroanatomy*, 8, 2014.
- [221] M. E. Mortenson. *Geometric Modeling*, 2nd Ed. Wiley, Hoboken, NJ, 1997.
- [222] E. I. Moser, E. Kropff, and M.-B. Moser. Place cells, grid cells, and the brain’s spatial representation system. *Annual Reviews of Neuroscience*, 31:69–89, 2008.
- [223] J. D. Moss and E. R. Muth. Characteristics of head-mounted displays and their effects on simulator sickness. *Human Factors*, 53(3):308–319, 2011.
- [224] D. E. Muller and F. P. Preparata. Finding the intersection of two convex polyhedra. *Theoretical Computer Science*, 7:217–236, 1978.
- [225] D. Mustafi, A. H. Engel, and Palczewski. Structure of cone photoreceptors. *Progress in Retinal and Eye Research*, 28:289–302, 2009.

- [226] T. Narumi, S. Nishizaka, T. Kajinami, T. Tanikawa, and M. Hirose. Augmented reality flavors: gustatory display based on edible marker and cross-modal interaction. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pages 93–102, 2011.
- [227] N. Naseer and K.-S. Hong. fNIRS-based brain-computer interfaces: a review. *Frontiers in Human Neuroscience*, 9(3), 2015.
- [228] G. Nelson, J. Chandrashekhar, M. A. Hoon, L. Feng, G. Zhao, N. J. P. Ryba, and C. S. Zuker. An amino-acid taste receptor. *Nature*, 416:199–202, 2002.
- [229] A. Newell and P. S. Rosenbloom. Mechanisms of skill acquisition and the law of practice. In J. R. Anderson, editor, *Cognitive skills and their acquisition*, pages 1–55. Erlbaum, Hillsdale, NJ, 1981.
- [230] Y. M. H. Ng and C. P. Kwong. Correcting the chromatic aberration in barrel distortion of endoscopic images. *Journal of Systemics, Cybernetics, and Informatics*, 2003.
- [231] F. Nicodemus. Directional reflectance and emissivity of an opaque surface. *Applied Optics*, 4(7):767–775, 1965.
- [232] L. F. Nicolas-Alonso and J. Gomez-Gil. Brain computer interfaces, a review. *Sensors*, 12(2):1211–1279, 2012.
- [233] J. Ninio. *The Science of Illusions*. Cornell University Press, Ithaca, NY, 2001.
- [234] D. Nitz. A place for motion in mapping. *Nature Neuroscience*, 18:6–7, 2010.
- [235] G. N'utzi, S. Weiss, D. Scaramuzza, and R. Siegwart. Fusion of IMU and vision for absolute scale estimation in monocular SLAM. *Journal of Intelligent and Robotic Systems*, 61(1):287–299, 2011.
- [236] Office for Human Research Protections. International compilation of human research standards. Technical report, U.S. Department of Health and Human Services, 2016. Available at <http://www.hhs.gov/ohrp/international/compilationhuman-research-standards>.
- [237] A. M. Okamura, J. T. Dennerlein, and R. D. Howe. Vibration feedback models for virtual environments. In Proc. IEEE International Conference on Robotics and Automation, volume 1, pages 674–679, 1998.
- [238] J. O'Keefe and J. Dosytovsky. The hippocampus as a spatial map. Preliminary evidence from unit activity in the freely-moving rat. *Brain Research*, 34(1):171–175, 1971.
- [239] J. L. Olson, D. M. Krum, E. A. Suma, and M. Bolas. A design for a smartphonebased head mounted display. In Proceedings IEEE Virtual Reality Conference, pages 233–234, 2011.
- [240] G. Osterberg. Topography of the layer of rods and cones in the human retina. *Acta Ophthalmologica, Supplement*, 6:1–103, 1935.
- [241] G. D. Park, R. W. Allen, D. Fiorentino, T. J. Rosenthal, and M. L. Cook. Simulator sickness scores according to symptom susceptibility, age, and gender for an older driver assessment study. In Proceedings of the Human Factors and Ergonomics Society Annual Meeting, pages 2702–2706, 2006.
- [242] E. Paulos and J. Canny. Prop: Personal roving presence. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pages 296–303, 1995.
- [243] E. Paulos and J. Canny. Social tele-embodiment: Understanding presence. *Autonomous Robots*, 11(1):87–95, 2000.
- [244] M. Pedley. High-precision calibration of a three-axis accelerometer. Technical report, Freescale Semiconductor, 2015. Available at <http://cache.freescale.com/files/sensors/doc/appnote/AN4399.pdf>.

- [245] E. Peli. The visual effects of head-mounted display (HMD) are not distinguishable from those of desk-top computer display. *Vision Research*, 38(13):2053–2066, 1998.
- [246] E. Peli. Optometric and perceptual issues with head-mounted displays. In P. Mouroulis, editor, *Visual instrumentation : optical design and engineering principles*. McGraw-Hill, New York, NY, 1999.
- [247] J. Pelz, M. Hayhoe, and R. Loeber. The coordination of eye, head, and hand movements in a natural task. *Experimental Brain Research*, 139(3):266–277, 2001.
- [248] Sönke Pelzer, Lukas Aspöck, Dirk Schröder, and Michael Vorlander. Integrating real-time room acoustics simulation into a cad modeling software to enhance the architectural design process. *Buildings*, 2:1103–1138, 2014.
- [249] R. J. Pethybridge. Sea sickness incidence in royal navy ships. Technical Report 37/82, Institute of Naval Medicine, Gosport, Hants, UK, 1982.
- [250] S. Petitjean, D. Kriegman, and J. Ponce. Computing exact aspect graphs of curved objects: algebraic surfaces. *International Journal of Computer Vision*, 9:231–255, December 1992.
- [251] V. I. Petkova and H. H. Ehrsson. If I Were You: Perceptual Illusion of Body Swapping. *PloS ONE*, 3(12), 2008.
- [252] M. Pocchiola and G. Vegter. The visibility complex. *International Journal Computational Geometry & Applications*, 6(3):279–308, 1996.
- [253] T. Poggio, M. Fahle, and S. Edelman. Fast perceptual learning in visual hyperacuity. *Science*, 256(5059):1018–1021, 1992.
- [254] I. Poupyrev, M. Billinghurst, S. Weghorst, and T. Ichikawa. The go-go interaction technique: non-linear mapping for direct manipulation in VR. In *Proceedings ACM Symposium on User Interface Software and Technology*, pages 79–80, 1996.
- [255] M. Prsa, S. Gale, and O. Blanke. Self-motion leads to mandatory cue fusion across sensory modalities. *Journal of Neurophysiology*, 108(8):2282–2291, 2012.
- [256] V. Pulkki. Virtual sound source positioning using vector base amplitude panning. *Journal of the Audio Engineering Society*, 45(6):456–466, 1997.
- [257] V. Pulkki. Virtual sound source positioning using vector base amplitude panning. *Journal of the Audio Engineering Society*, 55(6):503–516, 2007.
- [258] V. Pulkki and J. Merimaa. Spatial impulse response rendering II: Reproduction of diffuse sound and listening tests. *Journal of the Audio Engineering Society*, 54(1/2):3–20, 2006.
- [259] S. Rajangam, P. H. Tseng, A. Yin, G. Lehew, D. Schwarz, M. A. Lebedev, and M. A. Nicolelis. Wireless cortical brain-machine interface for whole-body navigation in primates. *Scientific Reports*, 2016.
- [260] N. Ranasinghe, R. Nakatsu, N. Hieaki, and P. Gopalakrishnakone. Tongue mounted interface for digitally actuating the sense of taste. In *Proceedings IEEE International Symposium on Wearable Computers*, pages 80–87, 2012.
- [261] S. Razzaque, Z. Kohn, and M C. Whitton. Redirected walking. In *Proceedings of Eurographics*, pages 289–294, 2001.
- [262] J. T. Reason and J. J. Brand. Motion Sickness. Academic, New York, 1975.
- [263] M. F. Reschke, J. T. Somers, and G. Ford. Stroboscopic vision as a treatment for motion sickness: strobe lighting vs. shutter glasses. *Aviation, Space, and Environmental Medicine*, 77(1):2–7, 2006.
- [264] S. W. Rienstra and A. Hirschberg. An Introduction to Acoustics. Endhoven University of

- Technology, 2016. Available at <http://www.win.tue.nl/~sjoerdr/papers/boek.pdf>.
- [265] K. J. Ritchey. Panoramic image based virtual reality/telepresence audio-visual system and method. US Patent 5495576A, February 1996.
- [266] H. Robbins and S. Monro. Stochastic iteration: A Stochastic approximation method. *Annals of Mathematical Statistics*, 22(3):400–407, 1951.
- [267] C. P. Robert. *The Bayesian Choice*, 2nd. Ed. Springer-Verlag, Berlin, 2001.
- [268] P. Robinson, A. Walther, C. Faller, and J. Braasch. Echo thresholds for reflections from acoustically diffusive architectural surfaces. *Journal of the Acoustical Society of America*, 134(4):2755–2764, 2013.
- [269] M. Rolfs. Microsaccades: Small steps on a long way. *Psychological Bulletin*, 49(20):2415–2441, 2009.
- [270] R. Ron-Angevin and A. Diaz-Estrella. Braincomputer interface: Changes in performance using virtual reality techniques. *Neuroscience Letters*, 449(2):123–127, 2009.
- [271] D. Rosenbaum. *Human Motor Control*, 2nd Ed. Elsevier, Amsterdam, 2009.
- [272] S. Ross. *A First Course in Probability*, 9th Ed. Pearson, New York, NY, 2012.
- [273] G. Roth and U. Dicke. Evolution of the brain and intelligence. *Trends in Cognitive Sciences*, 9:250–257, 2005.
- [274] K. Ruhland, C. E. Peters, S. Andrist, J. B. Badler, N. I. Badler, M. Gleicher, B. Mutlu, and R. McDonnell. A review of eye gaze in virtual agents, social robotics and hci: Behaviour generation, user interaction and perception. *Computer Graphics Forum*, 34(6):299–326, 2015.
- [275] A. Ruina and R. Pratap. *Introduction to Statics and Dynamics*. Oxford University Press, Oxford, UK, 2015. Available at <http://ruina.tam.cornell.edu/Book/>.
- [276] W. Rushton. Effect of humming on vision. *Nature*, 216:1173–1175, 2009.
- [277] M. B. Sachs and N. Y. S. Kiang. Two-tone inhibition in auditory nerve fibres. *Journal of the Acoustical Society of America*, 43:1120–1128, 1968.
- [278] S. Sanei and J. A. Chambers. *EEG Signal Processing*. Wiley, Hoboken, NJ, 2007.
- [279] X. M. Sauvan and C. Bonnet. Spatiotemporal boundaries of linearvection. *Perception and Psychophysics*, 57(6):898–904, 1995.
- [280] D. Schmalsteig and T. Höllerer. *Augmented Reality: Principles and Practice*. Mendeley Ltd., London, 2015.
- [281] G. Schweighofer and A. Pinz. Robust pose estimation from a planar target. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 28(12):2024–2030, 2006.
- [282] A. R. Seitz, J. E. Nanez, S. R. Halloway, and T. Watanabe. Perceptual learning of motion leads to faster-flicker perception. *Journal of Vision*, 6(6):158, 2015.
- [283] A. R. Seitz and T. Watanabe. The phenomenon of task-irrelevant perceptual learning. *Vision Research*, 49(21):2604–2610, 2009.
- [284] M. Shelhamer, D. A. Robinson, and H. S. Tan. Context-specific adaptation of the gain of the vestibulo-ocular reflex in humans. *Journal of Vestibular Research: Equilibrium and Orientation*, 2(1):89–96, 1992.
- [285] R. N. Shepard. Circularity in judgements of relative pitch. *Journal of the Acoustical Society of America*, 36(12):2346–2453, 1964.
- [286] G. M. Shepherd. Discrimination of molecular signals by the olfactory receptor neuron. *Neuron*, 13(4):771–790, 1994.
- [287] T. B. Sheridan. Musings on telepresence and virtual presence. *Presence: Teleoperators and*

- Virtual Environments, 1(1):120–126, 1992.
- [288] W. R. Sherman and A. B. Craig. Understanding Virtual Reality: Interface, Application, and Design. Morgan Kaufmann, San Francisco, CA, 2002.
- [289] T. Shibata, J. Kim, D. M. Hoffman, and M. S. Banks. The zone of comfort: predicting visual discomfort with stereo displays. *Journal of Vision*, 11(8):1–29, 2011.
- [290] B. G. Shinn-Cunningham, S. Santarelli, and N. Kopco. Tori of confusion: Binaural localization cues for sources within reach of a listener. *Journal of the Acoustical Society of America*, 107(3):1627–1636, 2002.
- [291] M. Siedlecka, A. Klumza, M. Lukowska, and M. Wierzchon. Rubber hand illusion reduces discomfort caused by cold stimulus. *PloS ONE*, 9(10), 2014.
- [292] P. Signell. Predicting and specifying the perceived colors of reflective objects. Technical Report MISN-0-270, Michigan State University, East Lansing, MI, 2000. Available at <http://www.physnet.org/>.
- [293] M. Slater, B. Spanlang, M. V. Sanchez-Vives, and O. Blanke. Experience of body transfer in virtual reality. *PloS ONE*, 5(5), 2010.
- [294] L. J. Smart, T. A. Stoffregen, and B. G. Bardy. Visually induced motion sickness predicted by postural instability. *Human Factors*, 44(3):451–465, 2002.
- [295] C. U. M. Smith. Biology of Sensory Systems, 2nd Ed. Wiley, Hoboken, NJ, 2008.
- [296] G. Smith and D. A. Atchison. The Eye and Visual Optical Instruments. Cambridge University Press, Cambridge, U.K., 1997.
- [297] R. Sawdon Smith and A. Fox. Langford’s Basic Photography, 10th Ed. Focal Press, Oxford, U.K., 2016.
- [298] W. J. Smith. Modern Optical Engineering, 4th Ed. SPIE Press, Bellingham, WA, 2008.
- [299] N. Snavely, S. M. Seitz, and R. Szeliski. Photo tourism: exploring photo collections in 3D. *ACM Transactions on Graphics*, 25(3):835–846, 2006.
- [300] D. Song, K. Goldberg, and N. Y. Chong. Networked telerobots. In O. Khatib and B. Siciliano, editors, Springer Handbook of Robotics, pages 759–771. SpringerVerlag, Berlin, 2008.
- [301] B. R. Sorensen, M. Donath, G.-B. Yanf, and R. C. Starr. The minnesota scanner: A prototype sensor for three-dimensional tracking of moving body segments. *IEEE Transactions on Robotics*, 5(4):499–509, 1989.
- [302] R. W. Soukoreff and I. S. MacKenzie. Towards a standard for pointing device evaluation, perspectives on 27 years of Fitts law research in HCI. *International Journal of Human-Computer Studies*, 61:751–759, 2004.
- [303] M. W. Spong, S. Hutchinson, and M. Vidyasagar. Robot Modeling and Control. Wiley, Hoboken, NJ, 2005.
- [304] K. M. Stanney and R. S. Kennedy. Aftereffects from virtual environment explore: How long do they last? In Proceedings of the Human Factors and Ergonomics Society Annual Meeting, pages 48(2): 1476–1480, 1998.
- [305] K. M. Stanney and R. S. Kennedy. Simulation sickness. In D. A. Vincenzi, J. A. Wise, M. Mouloua, and P. A. Hancock, editors, Human Factors in Simulation and Training, pages 117–127. CRC Press, Boca Raton, FL, 2009.
- [306] A. Steed and S. Julier. Design and implementation of an immersive virtual reality system based on a smartphone platform. In Proceedings IEEE Symposium on 3D User Interfaces, 2013.
- [307] R. M. Steinman, Z. Pizlo, and F. J. Pizlo. Phi is not beta, and why Wertheimer’s discovery

- launched the Gestalt revolution. *Vision Research*, 40(17):2257–2264, 2000.
- [308] N. Stephenson. *Snow Crash*. Bantam Books, 1996.
- [309] R. M. Stern, S. Hu, R. LeBlanc, and K. L. Koch. Chinese hyper-susceptibility tovection-induced motion sickness. *Aviation, Space, and Environmental Medicine*, 64(9 Pt 1):827–830, 1993.
- [310] J. Steuer. Defining virtual reality: Dimensions determining telepresence. *Journal of Communication*, 42(4):73–93, 1992.
- [311] S. S. Stevenson. On the psychophysical law. *Psychological Review*, 64(3):153–181, 1957.
- [312] R. Stoakley, M. J. Conway, and R. Pausch. Virtual reality on a WIM: interative worlds in minature. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pages 265–272, 1995.
- [313] T. A. Stoffregen, E. Faugloire, K. Yoshida, M. B. Flanagan, and O. Merhi. Motion sickness and postural sway in console video games. *human factors. Human Factors*, 50(2):322–331, 2008.
- [314] Student. The probable error of a mean. *Biometrika*, 6(1):1–25, 1908.
- [315] I. E. Sutherland. The ultimate display. In Proceedings of the IFIP Congress, pages 506–508, 1965.
- [316] I. E. Sutherland. A head-mounted three dimensional display. In Proceedings of AFIPS, pages 757–764, 1968.
- [317] R. Szeliski. Image alignment and stitching: A tutorial. Technical Report MSR-TR-2004-92, Microsoft Research, 2004. Available at <http://research.microsoft.com/>.
- [318] R. Szeliski. Computer Vision: Algorithms and Applications. Springer-Verlag, Berlin, 2010.
- [319] L. Takayama, E. Marder-Eppstein, H. Harris, and J. Beer. Assisted driving of a mobile remote presence system: System design and controlled user evaluation. In Proc. IEEE International Conference on Robotics and Automation, pages 1883–1889, 2011.
- [320] Thomas and Finney. Calculus and Analytic Geomtry, 9th Ed. Addison-Wesley, Boston, MA, 1995.
- [321] L. L. Thompson and P. M. Pinsky. Acoustics. *Encyclopedia of Computational Mechanics*, 2(22), 2004.
- [322] S. Thrun, W. Burgard, and D. Fox. Probabilistic Robotics. MIT Press, Cambridge, MA, 2005.
- [323] A. Treisman. Focused attention in the perception and retrieval of multidimensional stimuli. *Attention, Perception, and Psychophysics*, 22(1):1–11, 1977.
- [324] B. Treutwein. Minireview: Adaptive psychophysical procedures. *Vision Research*, 35(17):2503–2522, 1995.
- [325] B. Triggs, P. McLauchlan, R. Hartley, and A. Fitzbiggon. Bundle adjustment - a modern synthesis. In Proceedings IEEE International Workshop on Vision Algorithms, pages 298–372, 1999.
- [326] R. Y. Tsai. A versatile camera calibration technique for high-accuracy 3D machine vision metrology using off-the-shelf TV cameras and lenses. *IEEE Journal of Robotics and Automation*, 3(4):323–344, 1987.
- [327] B. Ullmer and H. Ishii. Emerging frameworks for tangible user interfaces. In J. M. Carroll, editor, *Human-Computer Interaction for Tangible User Interfaces*, pages 579–601. Addison-Wesley, Boston, MA, 2001.
- [328] A. Vasalou and A. Joinson. Me, myself and I: The role of interactional context on self-presentation through avatars. *Computers in Human Behavior*, 25(2):510–520, 2009.
- [329] J. F. Vasconcelos, G. Elkaim, C. Silvestre, P. Oliveira, and B. Cardeira. Geometric approach to strapdown magnetometer calibration in sensor frame. *Transactions on Aerospace and Electronic*

- Systems, 47(2):1293–1306, 2011.
- [330] G. Vass and T. Perlaki. Applying and removing lens distortion in post production. Technical report, Colorfont, Ltd., Budapest, 2003.
- [331] J. Vidal. Toward direct brain - computer communication. Annual Review of Biophysics and Bioengineering, 2:157–180, 1973.
- [332] J. J. Vidal. Real-time detection of brain events in EEG. Proceedings of the IEEE, 65(5):633–664, 1977.
- [333] S. T. von Soemmerring. Über das Organ der Seele “. K“onigsberg, 1796. With afterword by Immanuel Kant.
- [334] M. Vorl“ander. Auralization. Springer-Verlag, Berlin, 2010.
- [335] M. Vorl“ander and B. Shinn-Cunningham. Virtual auditory displays. In K. S. Hale and K. M. Stanney, editors, Handbook of Virtual Environments, 2nd Edition. CRC Press, Boca Raton, FL, 2015.
- [336] C. W“achter and A. Keller. Instant ray tracing: The bounding interval hierarchy. In T. Akenine-M“oller and W. Heidrich, editors, Eurographics Symposium on Rendering, pages 139–149. 2006.
- [337] B. A. Wandell. Foundations of Vision. Sinauer Associates, 1995. Available at <https://foundationsofvision.stanford.edu/>.
- [338] X. Wang and B. Winslow. Eye tracking in virtual environments. In K. S. Hale and K. M. Stanney, editors, Handbook of Virtual Environments, 2nd Edition. CRC Press, Boca Raton, FL, 2015.
- [339] R. M. Warren, J. M. Wrightson, and J. Puretz. Illusory continuity of tonal and infratonal periodic sounds. Journal of the Acoustical Society of America, 84(4):1338–1142, 1964.
- [340] W. H. Warren and K. J. Kurtz. The role of central and peripheral vision in perceiving the direction of self-motion. Perception and Psychophysics, 51(5):443–454, 1992.
- [341] D. S. Watkins. Fundamentals of Matrix Computations. Wiley, Hoboken, NJ, 2002.
- [342] A. B. Watson and D. G. Pelli. QUEST: A Bayesian adaptive psychometric method. Perception and Psychophysics, 33(2):113–120, 1983.
- [343] B. L. Welch. The generalization of "Student's" problem when several different population variances are involved. Biometrika, 34(1-2):28–35, 1947.
- [344] G. Welch and E. Foxlin. Motion tracking: no silver bullet, but a respectable arsenal. IEEE Computer Graphics and Applications, 22(6):24–28, 2002.
- [345] R. B. Welch and B. J. Mohler. Adapting to virtual environments. In K. S. Hale and K. M. Stanney, editors, Handbook of Virtual Environments, 2nd Edition. CRC Press, Boca Raton, FL, 2015.
- [346] A. T. Welford. Fundamentals of Skill. Methuen Publishing, London, 1968.
- [347] M. Wertheimer. Experimentelle Studien “uber das Sehen von Bewegung (Experimental Studies on the Perception of Motion). Zeitschrift f“ur Psychologie, 61:161–265, 1912.
- [348] J. Westerhoff. Reality: A Very Short Introduction. Oxford University Press, Oxford, UK, 2011.
- [349] F. A. Wichman and N. J. Hill. The psychometric function: I. fitting, sampling, and goodness of fit. Perception and Psychophysics, 63(8):1293–1313, 2001.
- [350] J. M. Wolfe, K. R. Kluender, and D. M. Levi. Sensation and Perception, 4th Ed. Sinauer, Sunderland, MA, 2015.
- [351] J. R. Wolpaw, N. Birbaumer, D. J. McFarland, G. Pfurtscheller, and T. M. Vaughan. Braincomputer interfaces for communication and control. Clinical Neurophysiology, 113(6):767–

791, 2002.

- [352] A. F. Wright, C. F. Chakarova, M. M. Abd El-Aziz, and S. S. Bhattacharya. Photoreceptor degeneration: genetic and mechanistic dissection of a complex trait. *Nature Reviews Genetics*, 11:273–284, 2010.
- [353] F. E. Wright. The Methods of Petrographic-microscopic Research. Carnegie Institution of Washington, 1911.
- [354] Y. Wu and Z. Hu. PnP problem revisited. *Journal of Mathematical Imaging and Vision*, 24(1):131–141, 2006.
- [355] S. Xu, M. Perez, K. Yang, C. Perrenot, J. Felblinger, and J. Hubert. Determination of the latency effects on surgical performance and the acceptable latency levels in telesurgery using the dV-Trainer simulator. *Surgical Endoscopy*, 28(9):2569–2576, 2014.
- [356] T. Yamada, S. Yokoyama, T. Tanikawa, K. Hirota, and M. Hirose. Wearable olfactory display: Using odor in outdoor environment. In *Proceedings IEEE Virtual Reality Conference*, pages 199–206, 2006.
- [357] E. Yang and M. Dorneich. The effect of time delay on emotion, arousal, and satisfaction in human-robot interaction. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, pages 443–447, 2015.
- [358] X. Yang and W. Grantham. Effects of center frequency and bandwidth on echo threshold and buildup of echo suppression. *Journal of the Acoustical Society of America*, 95(5):2917, 1994.
- [359] R. Yao, T. Heath, A. Davies, T. Forsyth, N. Mitchell, and P. Hoberman. Oculus VR Best Practices Guide. Retrieved from <http://brianschrank.com/vrgames/resources/OculusBestPractices.pdf>, March 2014. Last retrieved on July 10, 2016.
- [360] N. Yee and J. Bailenson. The Proteus effect: The effect of transformed selfrepresentation on behavior. *Human Communication Research*, 33:271–290, 2007.
- [361] H. Yeh, R. Mehra, Z. Ren, L. Antani, M. C. Lin, and D. Manocha. Wave-ray coupling for interactive sound propagation in large complex scenes. *ACM Transactions on Graphics*, 32(6), 2013.
- [362] W. A. Yost. Fundamentals of Hearing: An Introduction, 5th Ed. Emerald Group, Somerville, MA, 2006.
- [363] P. Zahorik. Assessing auditory distance perception using virtual acoustics. *Journal of the Acoustical Society of America*, 111(4):1832–1846, 2002.
- [364] V. M. Zatsiorsky. Kinematics of Human Motion. Human Kinetics, Champaign, IL, 1997.
- [365] V. M. Zatsiorsky. Kinetics of Human Motion. Human Kinetics, Champaign, IL, 2002.
- [366] V. M. Zatsiorsky and B. I. Prilutsky. Biomechanics of Skeletal Muscles. Human Kinetics, Champaign, IL, 2012.
- [367] Y. Zheng, Y. Kuang, S. Sugimoto, and K. Aström. Revisiting the PnP problem: A fast, general and optimal solution. In *Proceedings IEEE International Conference on Computer Vision*, pages 2344–2351, 2013.
- [368] H. Zhou and H. Hu. Human motion tracking for rehabilitation - A survey. *Biomedical Signal Processing and Control*, 3(1):1–18, 2007.