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。）

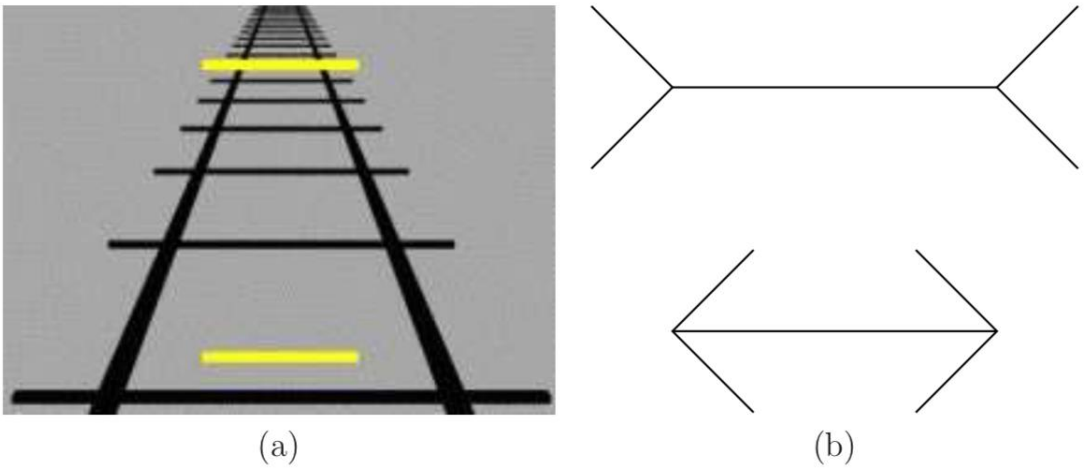


图 6.2：即使是简单的线条图也能提供重要的线索。 （a）蓬佐错觉：上面的黄色条看起来更长，但两者长度相同。 （b）米勒莱尔错觉：下面的水平段似乎比上面的短，但它们的长度相同。

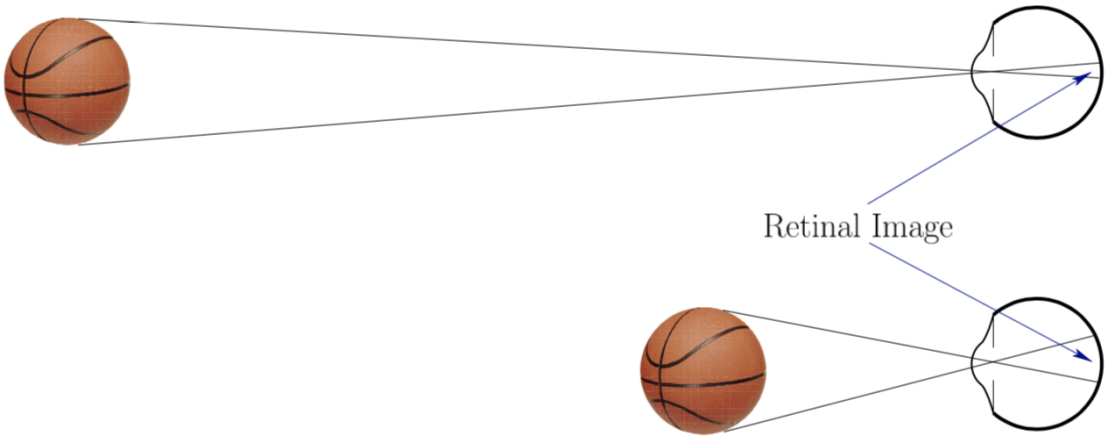


图 6.3：一个熟悉物体的视网膜图像大小是一个强大的单眼深度线索。较近的物体投射到覆盖视网膜较大部分的大量感光器上。

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。

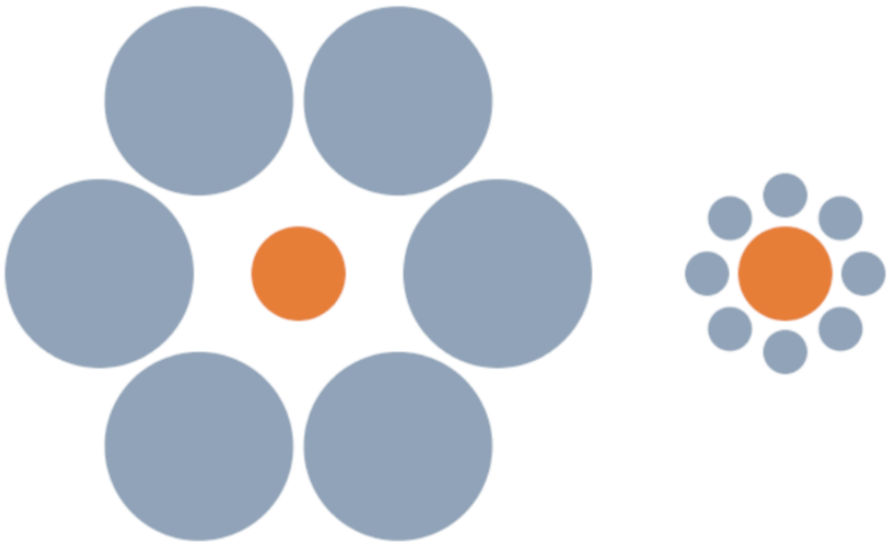
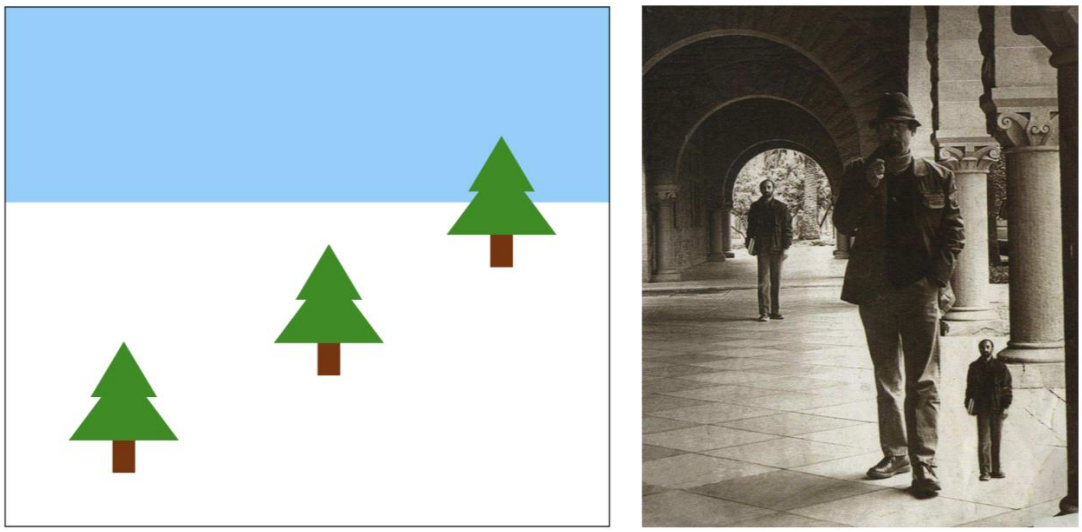


图 6.4：对于艾宾豪斯错觉，当小圆盘包围时，内圆盘看起来更大。 无论哪种情况，内盘的尺寸都是相同的。 这可能是真实视角（或视网膜图像大小）与感知视角之间差异的证据。

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）。

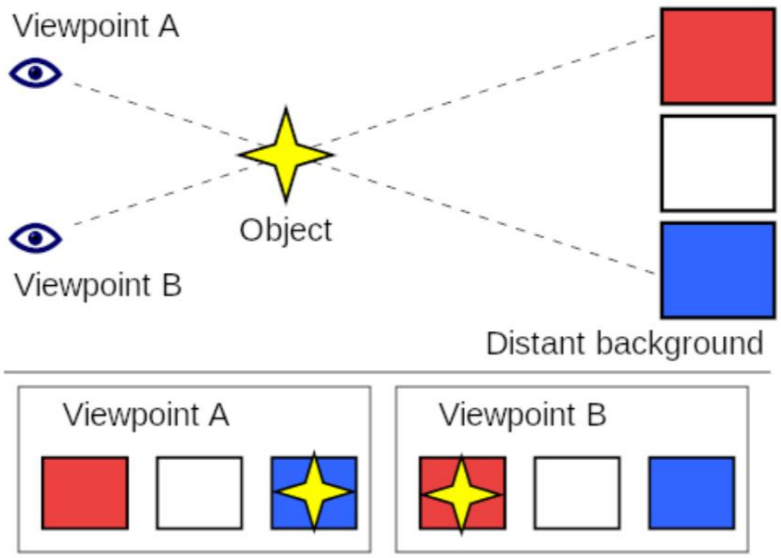


图 6.6：运动视差：当视角横向变化时，较近的物体比其他物体具有更大的图像位移。（图来自维基百科。）

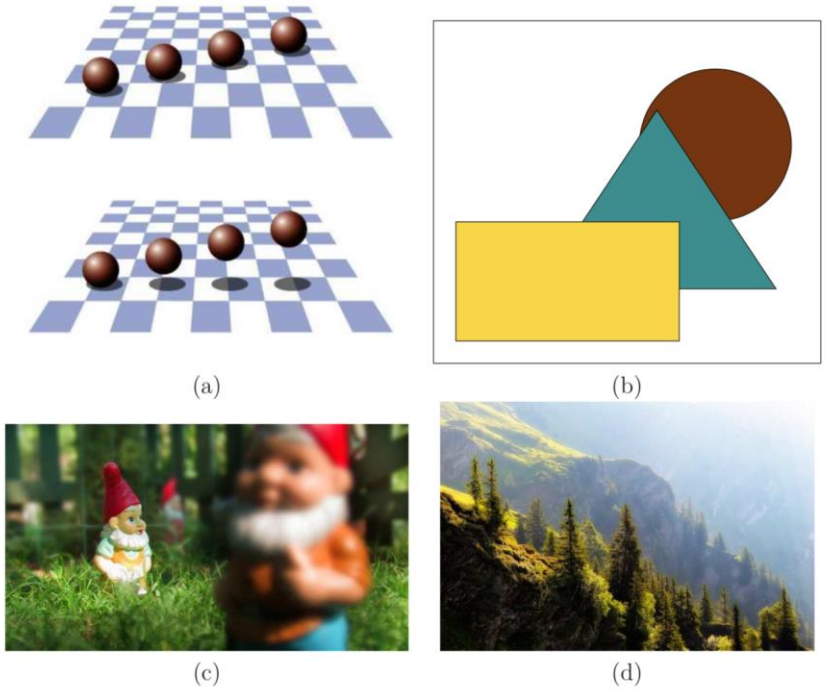


图 6.7：更多的单眼深度线索：（a）阴影解决球和幻影中不明确的深度。 （b）物体的插入提供了一个有序的深度线索。 （c）由于图像模糊，一个侏儒似乎比其他人更近。 （d）这个场景提供了一个大气线索：由于对比度较低，所以有些景观被认为距离较远。

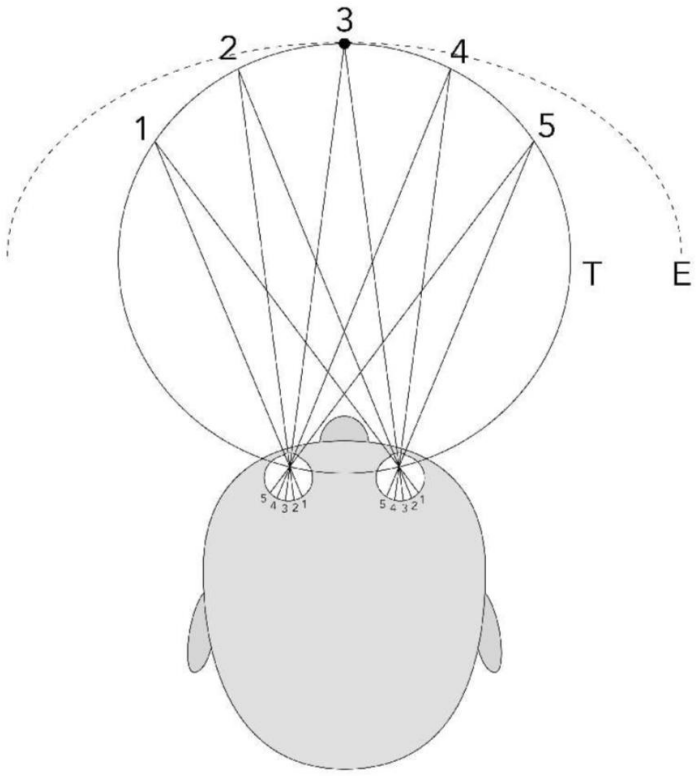


图 6.8：双眼视界是眼睛可以聚焦并聚焦在单个深度上的点的轨迹。 T 曲线显示了基于简单几何理论的双眼直角坐标。 E 曲线显示了经验性的双眼视界，其范围大得多并且对应于感知单个聚焦图像的区域。 （图

由 Rainer Zenz。）

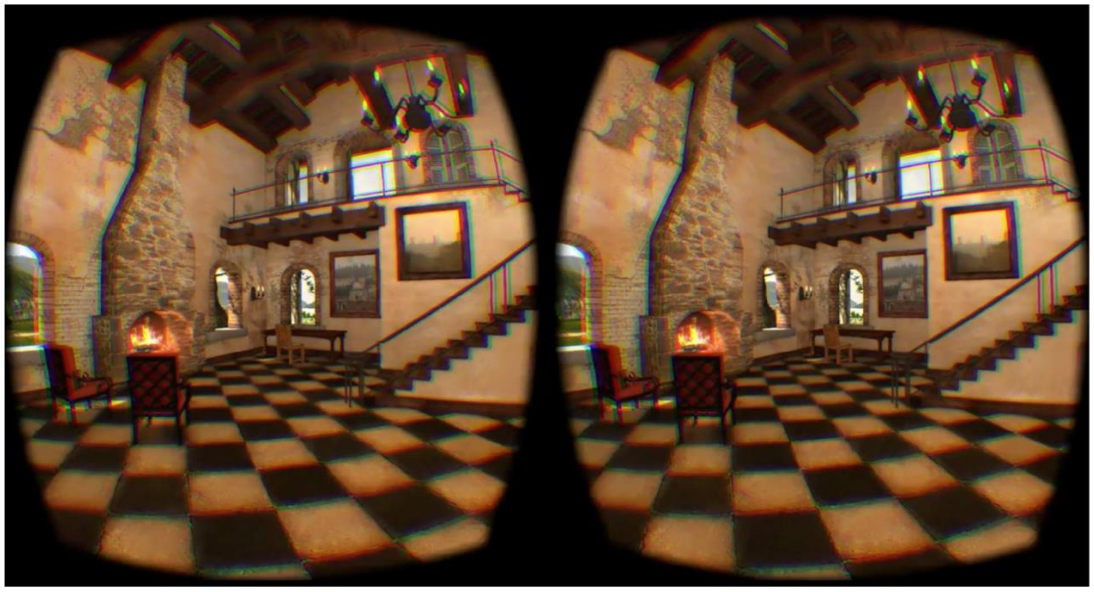


图 6.9：在 Oculus VR 的托斯卡纳演示中，没有足够熟悉的对象来精确解析深度和大小。 你曾经去过这样的别墅吗？ 地砖是否是熟悉的尺寸？ 桌子是否太低了？

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。 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，这会显着影响深度感知。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。

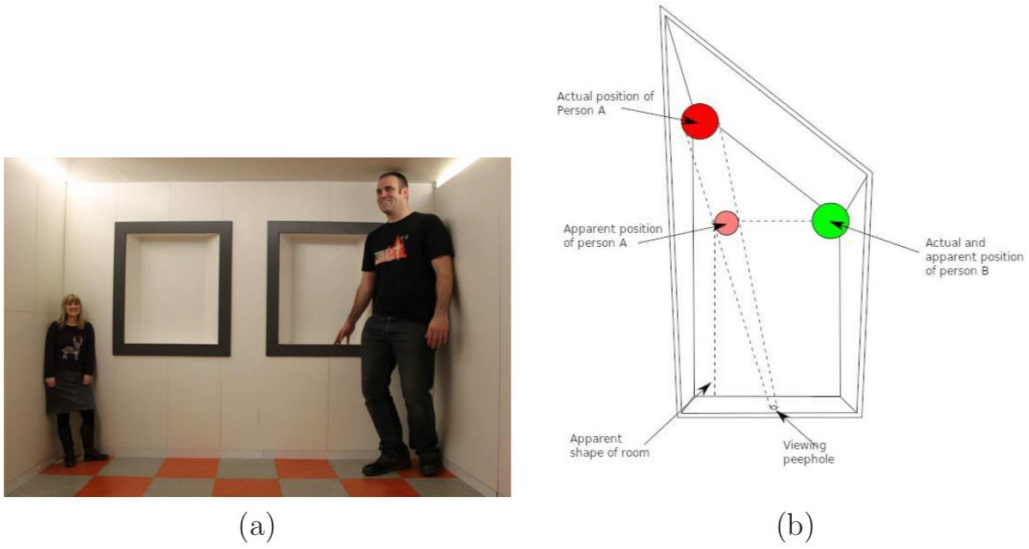


图 6.10：艾姆斯房间：（a）由于深度线索不正确，导致尺寸检测结果不正确。 （b）房间被设计成在透视投影后只呈现矩形。 一个人实际上远比另一个人远。 （图由亚历克斯 Valavanis。）

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。

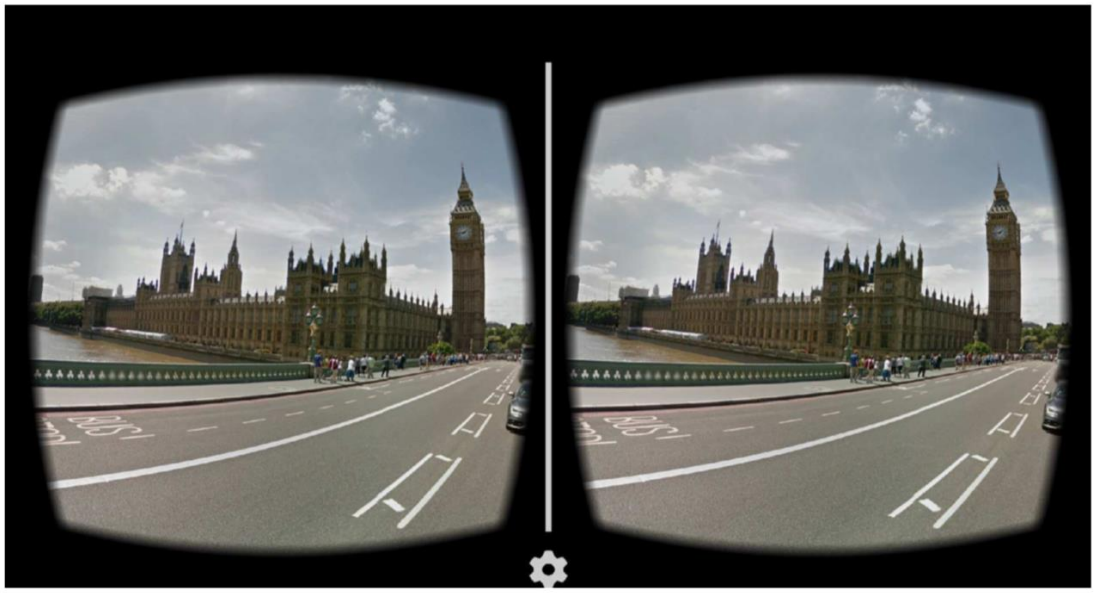


图 6.11：在 Google Cardboard 和其他 VR 头戴设备中，可以查看数亿个全景街景图像。 由于单视觉深度线索，即使将相同的图像呈现给双眼，也具有显着的深度感知。

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]。

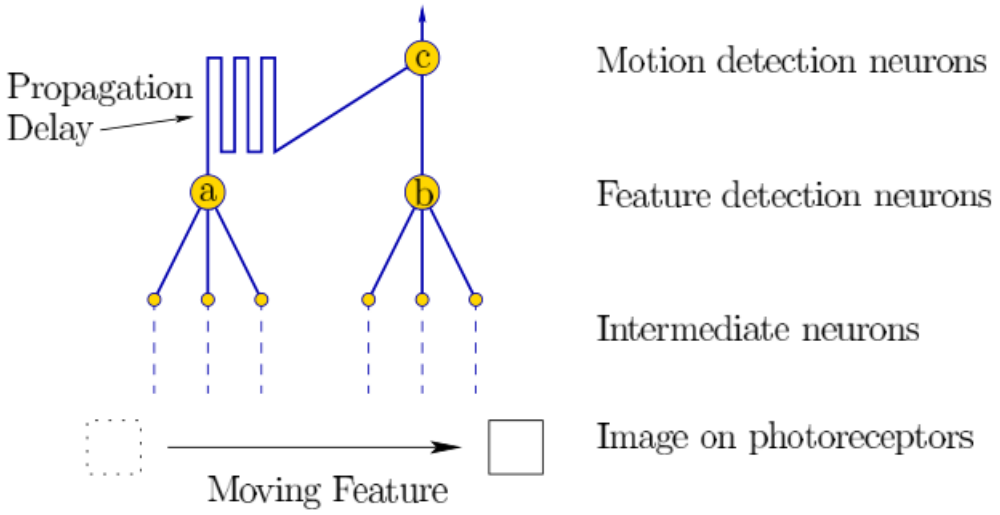


图 6.12：神经通路支持运动检测。 随着图像特征穿过视网膜，附近的特征检测神经元（标记为 a 和 b）连续激活。 它们的输出连接到运动检测神经元（标记为 c）。 由于从 a 和 b 到 c 的路径长度不同，激活信号在不同的时间到达。 因此，c 在 b 检测到到特征时已经激活。

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。详情请参阅第 8.4 节和第 10.2 节。

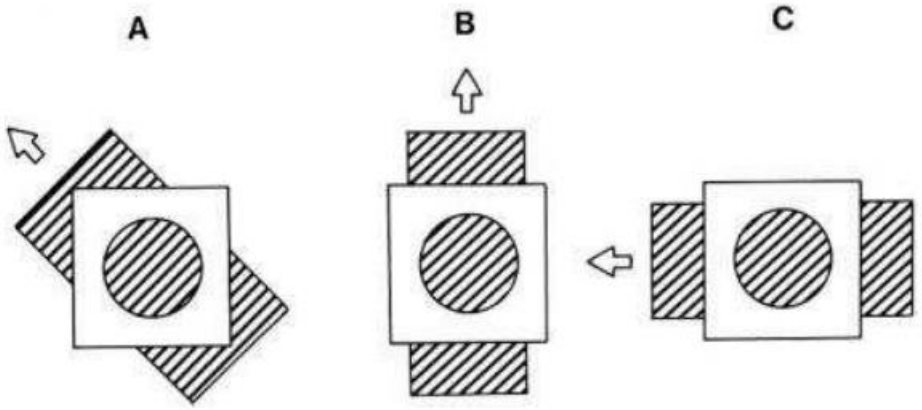


图 6.13：由于运动探测器的局部性质，会导致孔径问题。较大身体的运动在通过小孔感知时是模糊的，因为大量可能的运动在洞内会产生相同的效果，因而导致不正确的运动推断。

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。

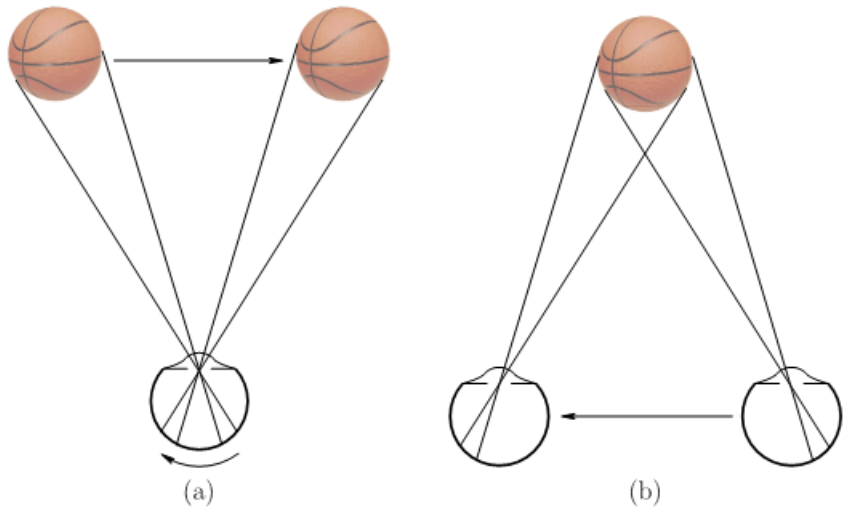


图 6.14：导致图像在视网膜上等效移动的两种运动：（a）眼睛被固定并且物体移动；（b）当物体固定时眼睛移动。

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。

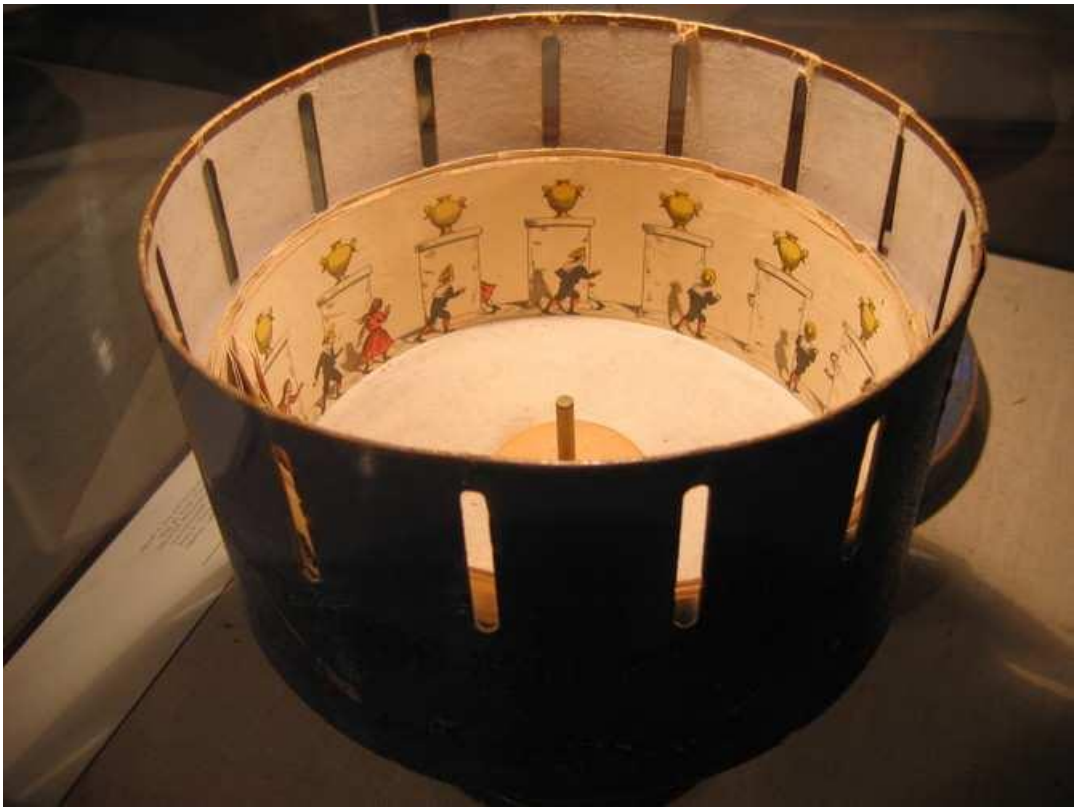


图 6.15：西洋镜在 1830 年代发展起来，并提供频闪视运动，通过旋转盘中的狭缝可以看到图像。

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]。

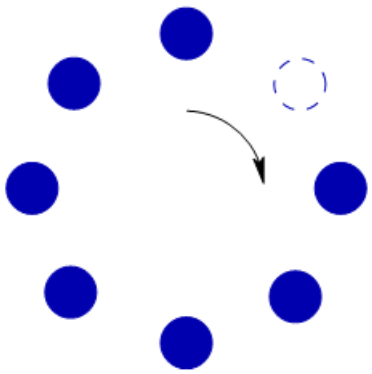


图 6.16：phi 现象和 beta 运动在生理学上是不同的，可以感知运动[347，307]。 在点序列中，任何时候都会关闭一个点。 按照顺时针方式，每帧中关闭一个不同的点。 在非常低的速度下（2 FPS），beta 运动会触发后一关闭点的运动感知。以更高的速度，例如 15 FPS，整体看上去就像是一个移动的洞，这对应于 phi 现象。

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。着名的 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。因为在外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。

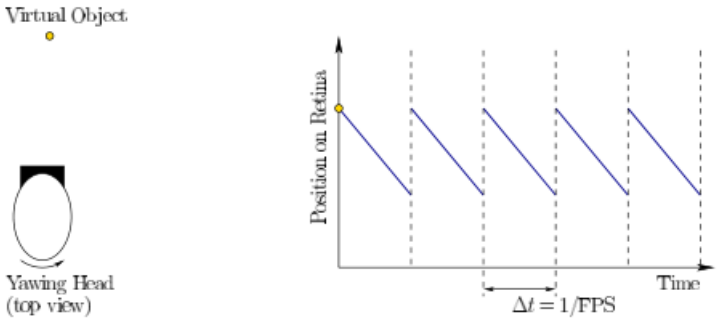


图 6.18：在频闪视运动下观察平稳性的问题：随着 VOR 的执行，特征的图像以重复模式滑过视网膜。

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。这种显示模式是低余辉的，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。

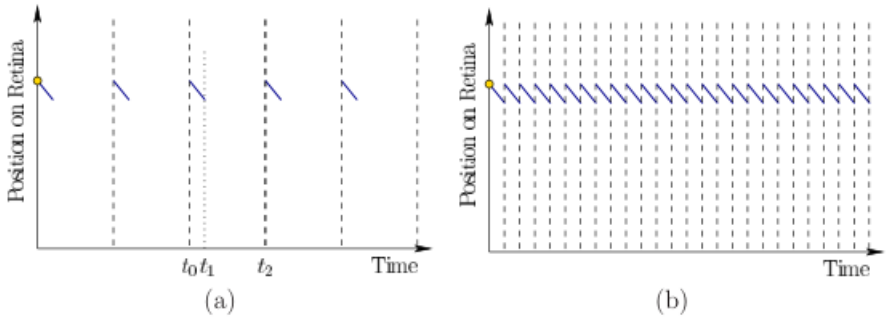


图 6.19：减少视网膜图像滑动的工程解决方案：（a）使用低余辉，显示器点亮足够短的时间以激活感光器（t1 - t0），然后在剩余时间（t2 - t1）内产生空白。 t1 -t0 通常为 1到 2 毫秒。（b）如果帧率非常大（至少 500 FPS），则不需要空白间隔。

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 RGB =(255,0,0)产生，绿色为 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 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。

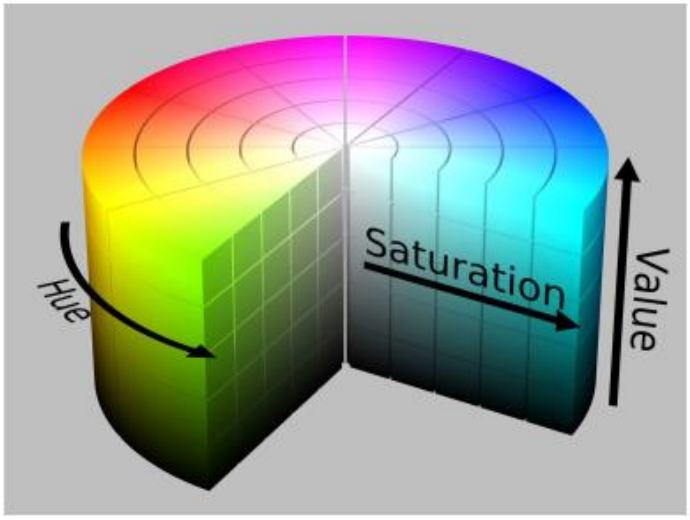


图 6.21：HSV 色彩空间的一种表示，它涉及三个参数：色调，饱和度和值（亮度）。（由维基百科用户 SharkD 提供）

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。

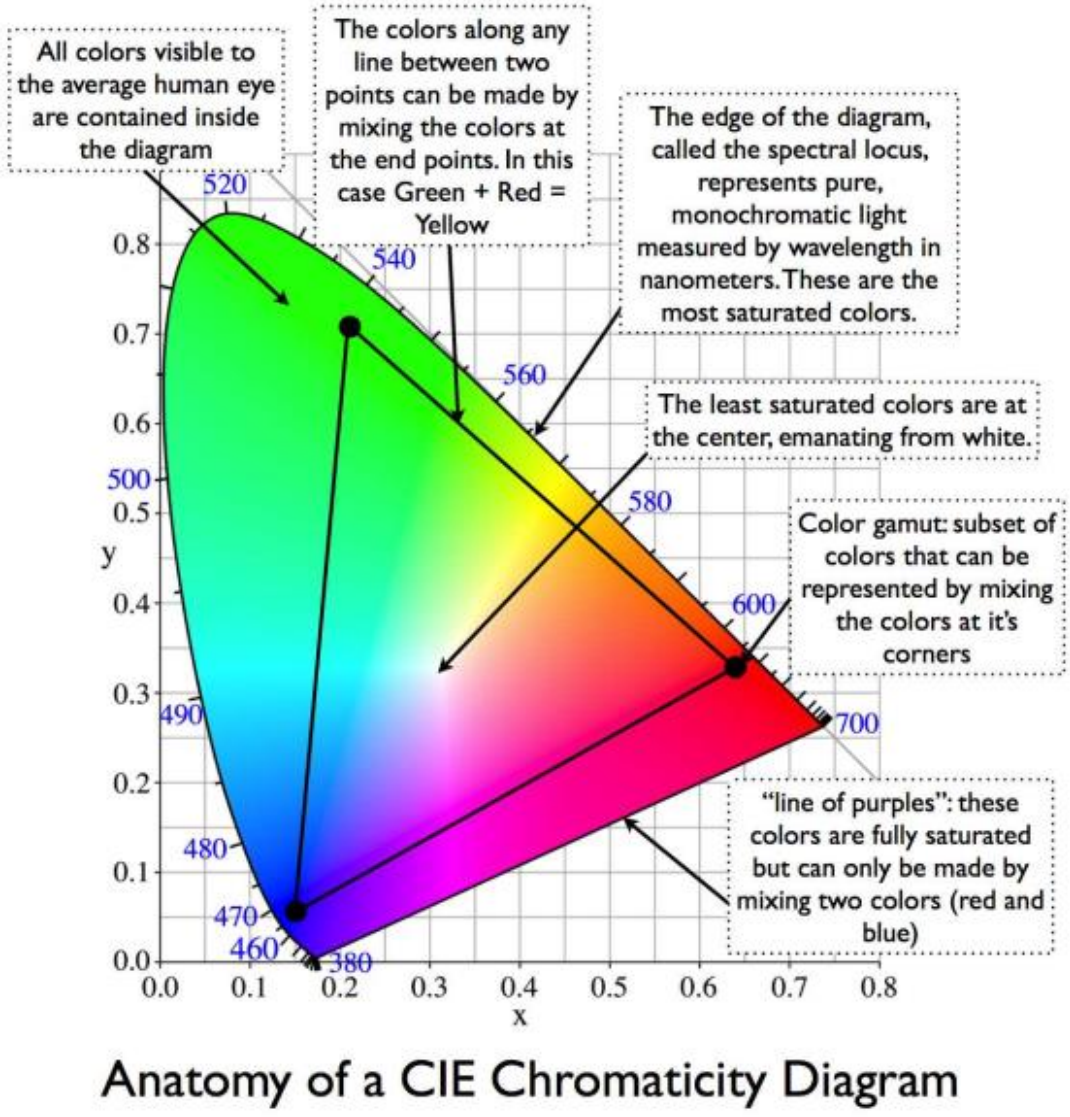


图 6.22：1931 年 CIE 颜色标准与 RGB 三角形。这种表示在感知颜色之间的距离方面是正确的。（由杰夫 Yurek 提供）

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。

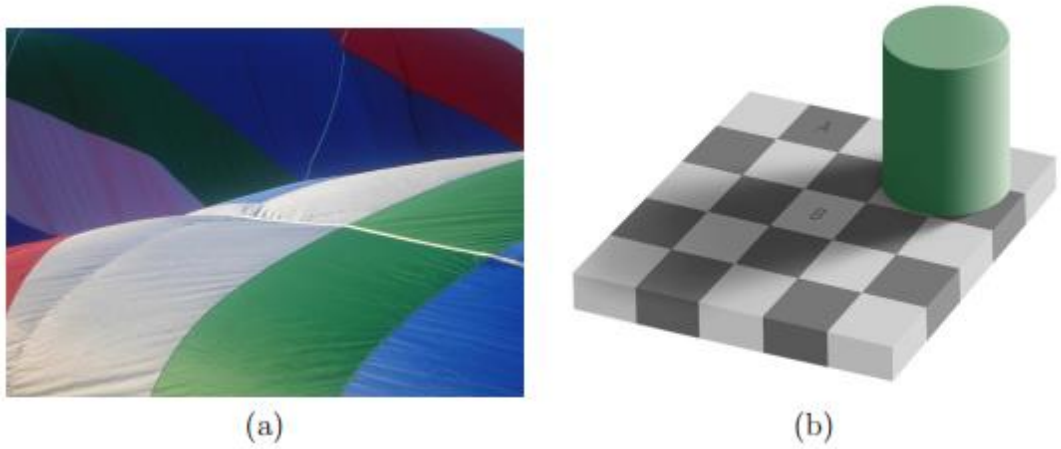


图 6.23：（a）无论在阳光直射或阴影中，感知到的热气球颜色都是相同的。（图由维基百科用户 Shanta 提供）（b）第 2.3 节中的阴影错觉由亮度常数原理解释，因为阴影会促使对感知亮度进行补偿。（图由 Adrian Pingstone 提供。）

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。

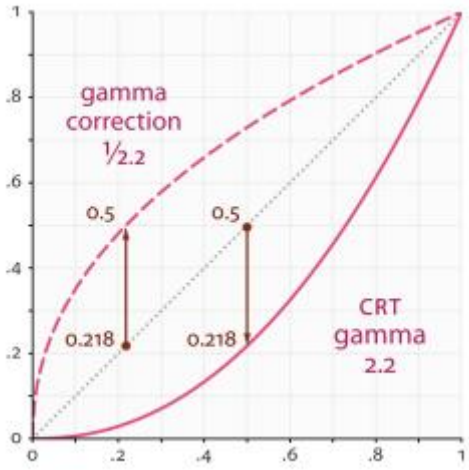


图 6.24：虽然位数有限，但伽马校正用于跨越更多数量级。变换是 v'=cv γ，其中 c 是常数（通常 c = 1），并且 γ 控制校正或失真的非线性。

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

$$H = \{h_1, h_2, \ldots, h_n\} \tag{6.1}$$

为一组假设（或解释）。同样，让

$$C = \{c_1, c_2, \ldots, c_m\} \tag{6.2}$$

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 = {棕色，蓝色，绿色，淡褐色}。 (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

$$P(h_1) + P(h_2) + \cdots + P(h_n) = 1, \tag{6.4}$$

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

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(奥巴马|布朗)，应该是大于 P(巴拉克・奥巴马|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：

$$P(h \mid c) = \frac{P(c \mid h)P(h)}{P(c)}. \tag{6.5}$$

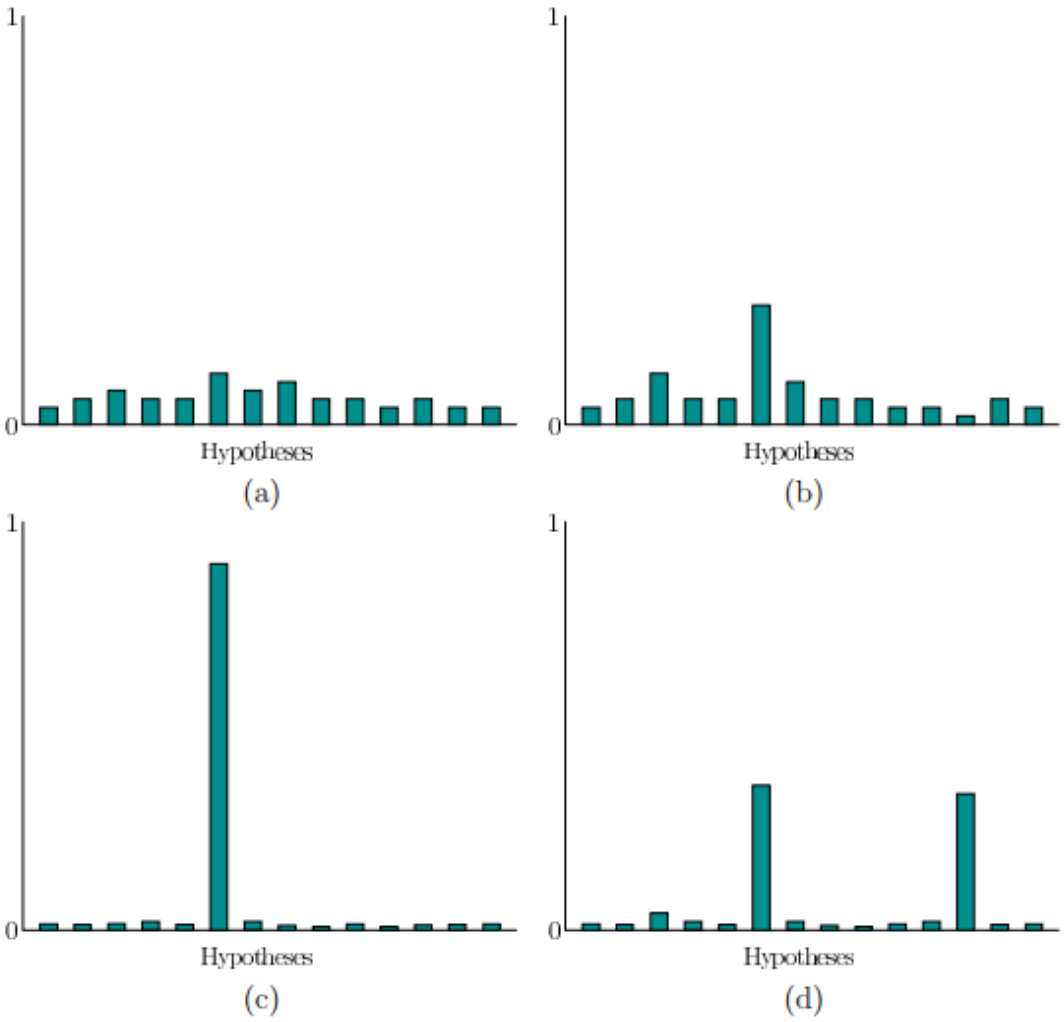


图 6.25：示例概率分布：（a）可能的先验分布。 （b）在条件后开始出现一种假设。 （c）强大一致的线索导致峰值分布。 （d）模糊性可能会导致两个（或更多）假设比其他假设更受欢迎；这是多重感知的基础。

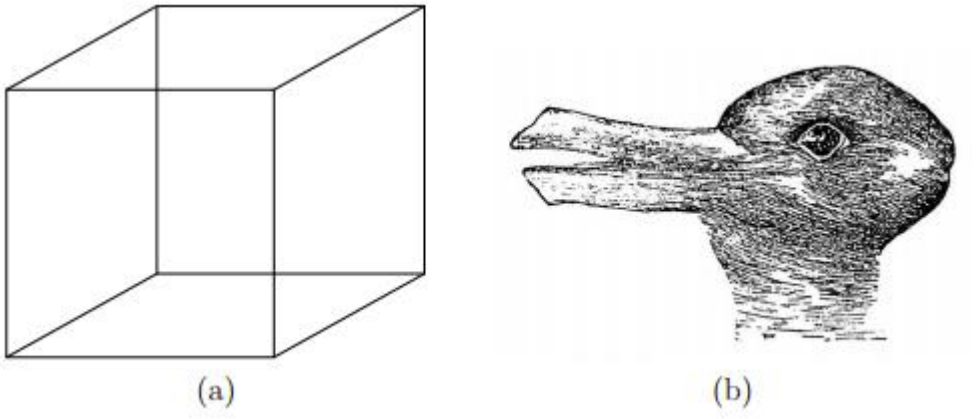


图 6.26：（a）1832 年由瑞士晶体学家路易斯·阿尔伯特·内克研究的 Necker 立方体。 （b）1892 年 10 月 23 日 Fliegende Blatter 的兔子鸭幻觉。

分母可以表示为

$$P(c) = P(c \mid h_1)P(h_1) + P(c \mid h_2)P(h_2) + \cdots + P(c \mid h_n)P(h_n), \qquad (6.6)$$

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，...，d_k}表示新检测器的可能输出。贝叶斯公式变成了

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

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。

进一步阅读

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。