

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。

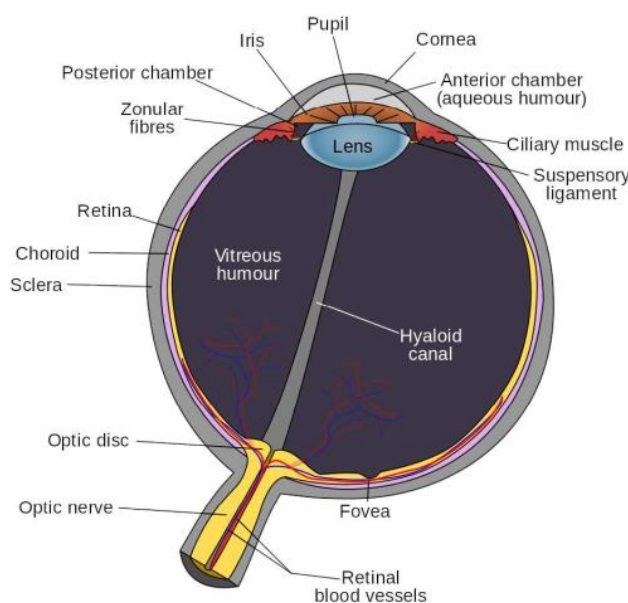


图 5.1：人眼的生理结构，图中所示的是右眼的水平横截面（鼻子在左边）（源自维基用户

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 (cd/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）

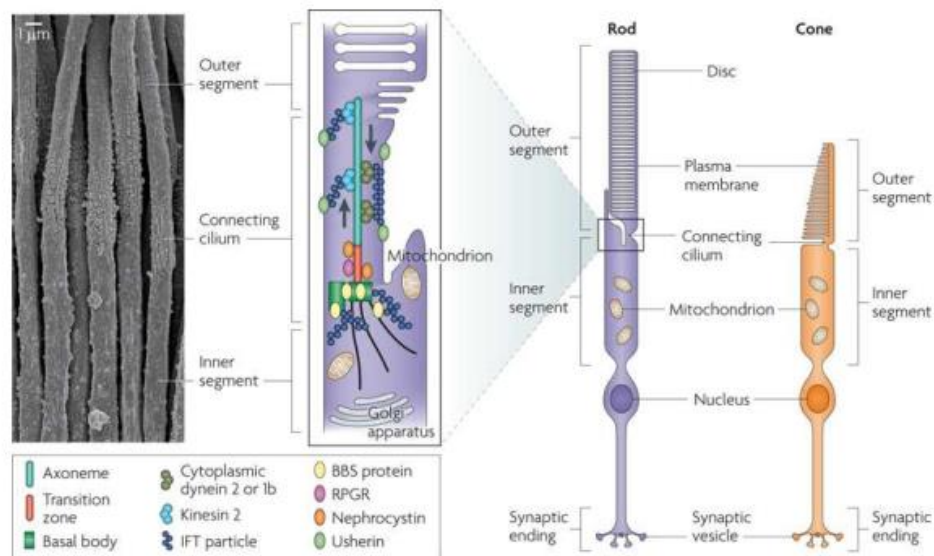


图 5.2 左图是感光体的电镜图；右图是视杆细胞和视锥细胞的组成和结构，外部片段含有光色素，在受到光子击中时可以产生电化学反应。

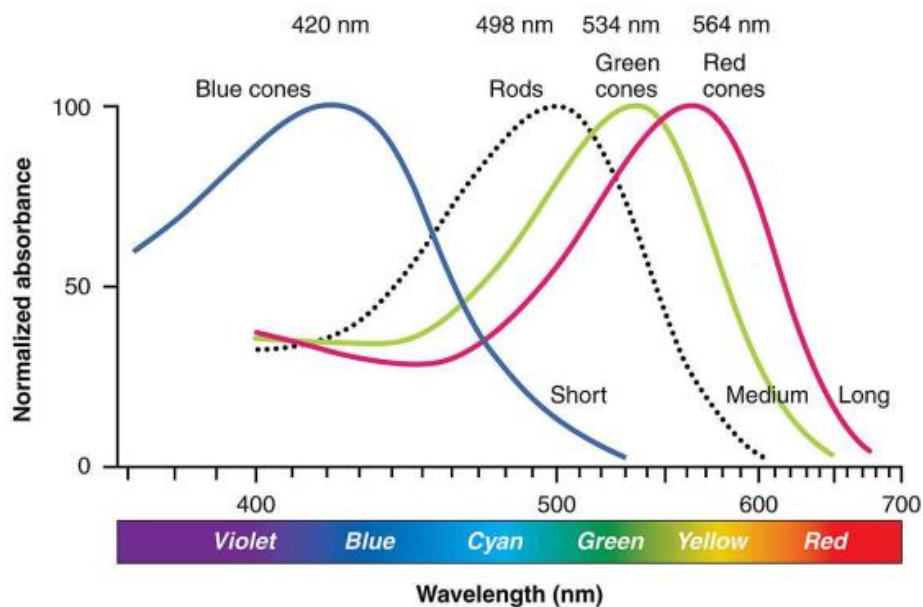


图 5.3 视杆细胞和视锥细胞对光波长的敏感度曲线（图由 OpenStax College 提供）

Light source	Luminance (cd/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）

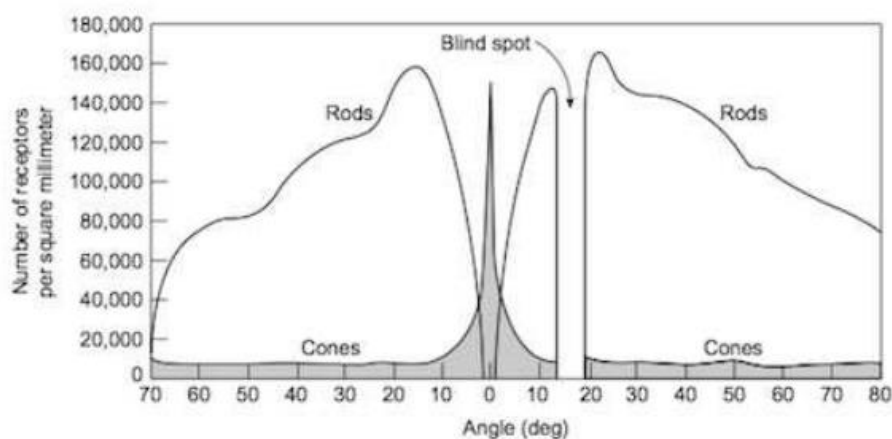


图 5.5 感光体密度和角度的关系函数，中间的空白是鼻侧（对应着从相反方向进入的光线）

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。最低层的神经对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。

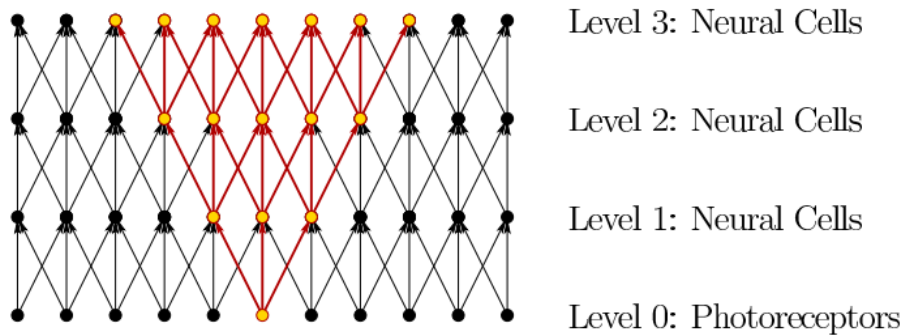


图 5.7 简单层次结构的四个层级。每个圆点对应一个神经细胞或感光体，箭头表示信息流。感光体在 0 级产生信息。在整个非常简化和理想化的视图中，每个感光体和神经元恰好连接到下一层的三个感光体或神经元。红色和金色部分突出显示了单个感光体随着层数增加而可能产生的增长区域。

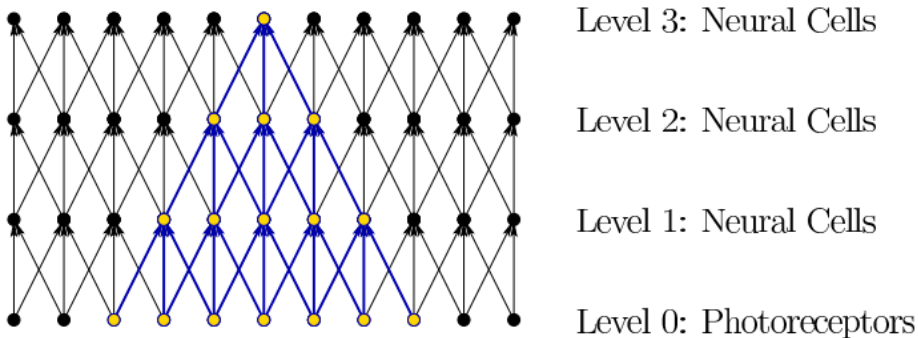


图 5.8 与图 5.7 相同，只是突出展示了进入单个神经元的信息。考虑参与单个神经细胞反应的一组感光

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。

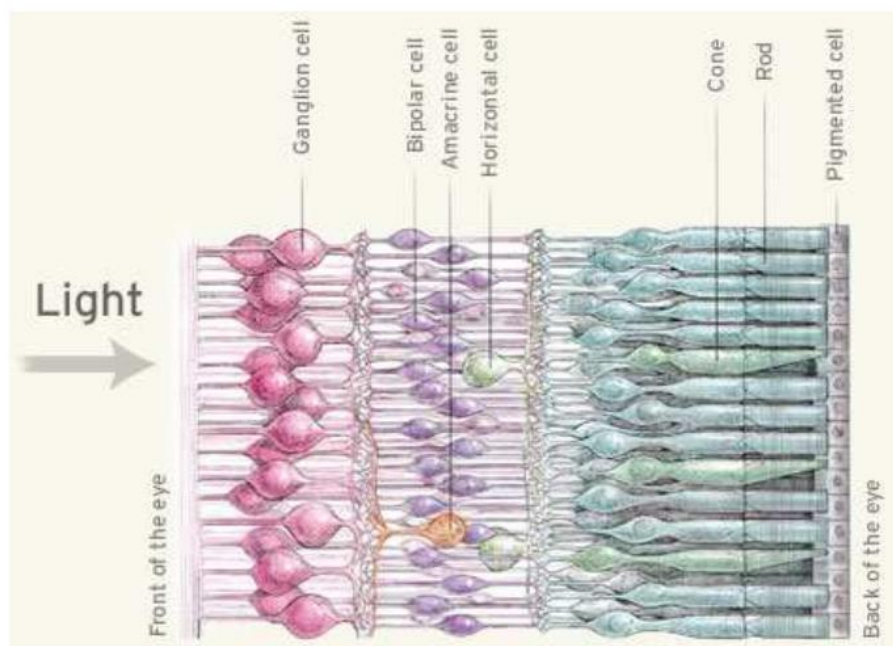


图 5.9： 光线经过视杆细胞和视锥细胞之前要先穿过一些神经层（Institute for Dynamic Educational Advancement 提供）

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、无长突细胞和水平细胞。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。

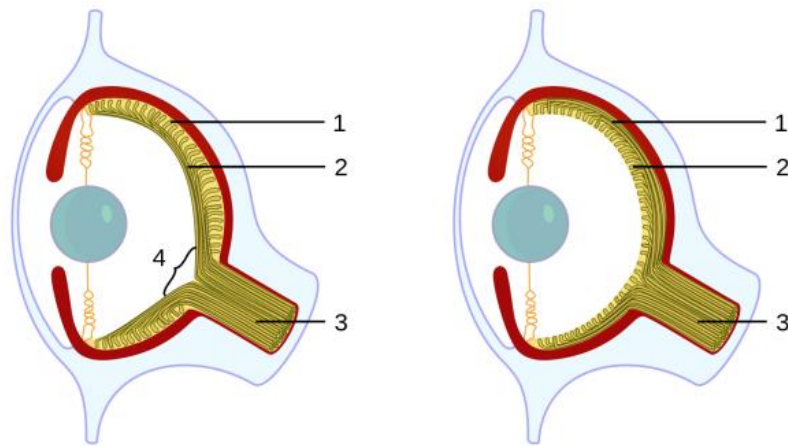


图 5.10: 脊椎动物（包括人类）具有内外视网膜，这导致出现盲区并且也导致光感受器远离入射光。左边为脊椎动物的眼睛，右边为头足类眼睛，可以解释：光感受器面向光线，没有盲点。（图由 Jerry Crimson Mann 提供。）

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，无长突细胞和双极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。

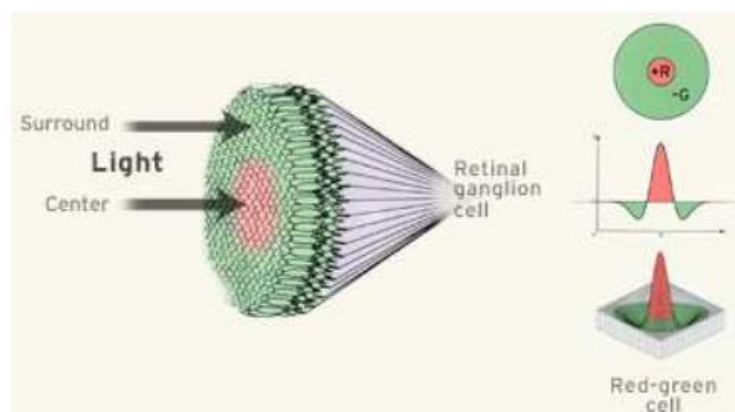


图 5.11: 中心神经节细胞的感受野 （图由动态教育发展研究提供）

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。

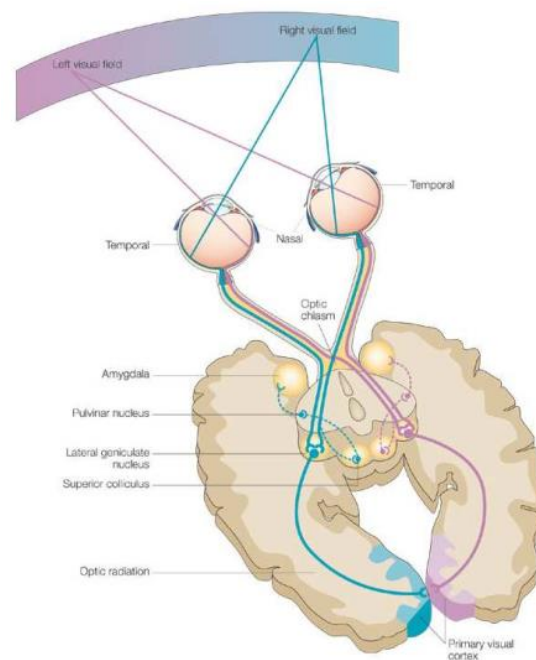


图 5.12：从眼睛到 LGN 到视觉皮层的视觉通路。 请注意，来自视野的左侧和右侧的信息在皮质中交换。

（图片来自 Nature Reviews: Neuroscience）

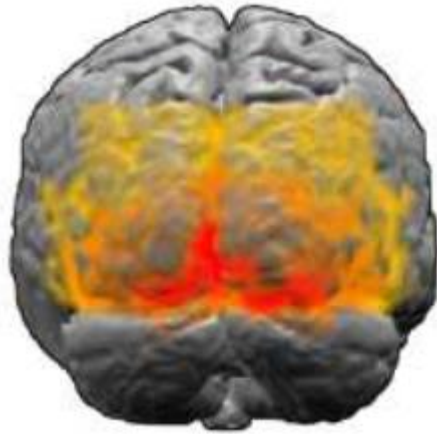


图 5.13：视觉皮层位于头部后方（图由华盛顿·欧文提供）

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。

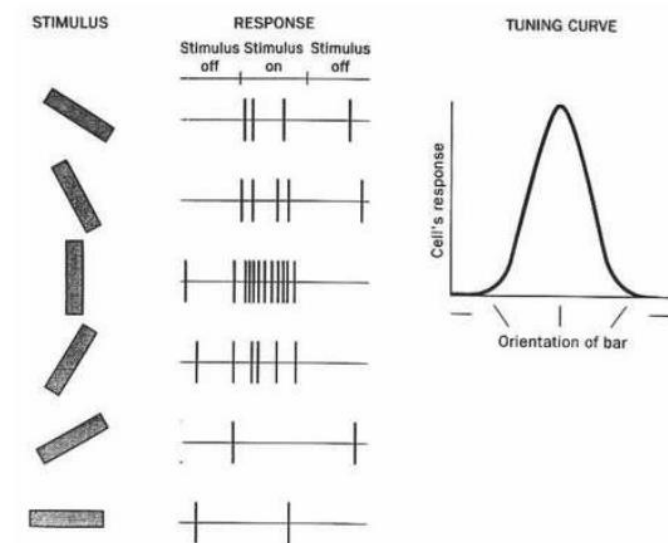


图 5.14：视觉皮层功能的一个示例是定向调整，其中单位记录由皮质中的单个神经元组成。 当一个长条形的物体在眼前旋转时，神经元的响应变化如图所示。它强烈支持一个特定的方向。



图 5.15：使用扫视法扫描人脸的痕迹。

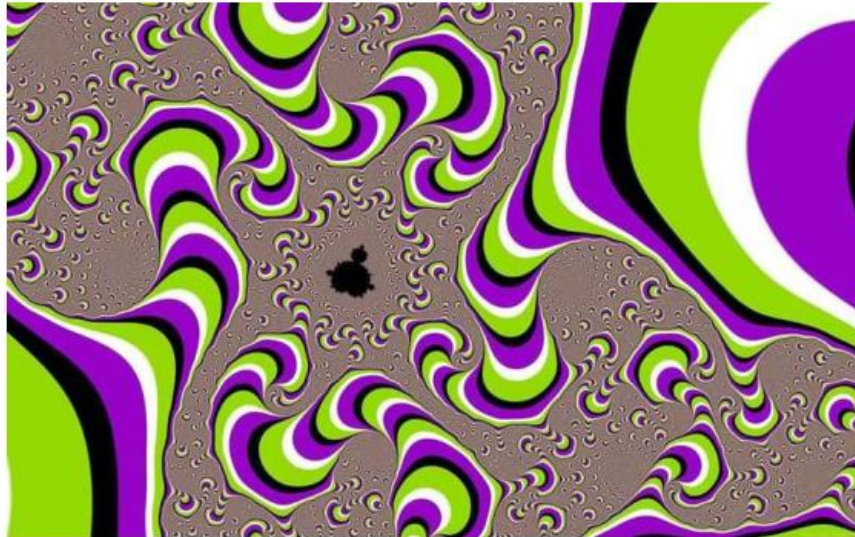


图 5.16: 分形图似乎在移动，直到您仔细注意到单个零件位置。

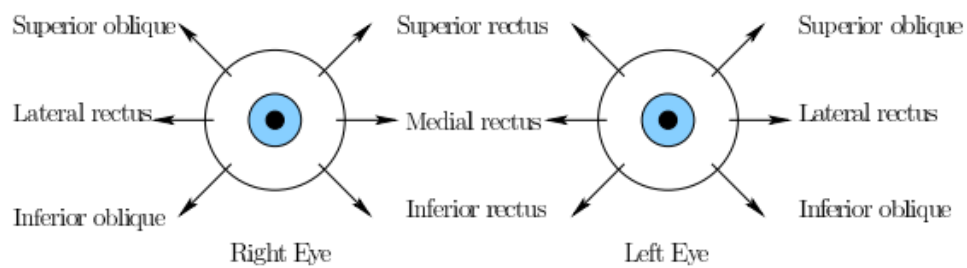


图 5.17: 每只眼睛有六块肌肉，每块肌肉都能将瞳孔拉向其位置。

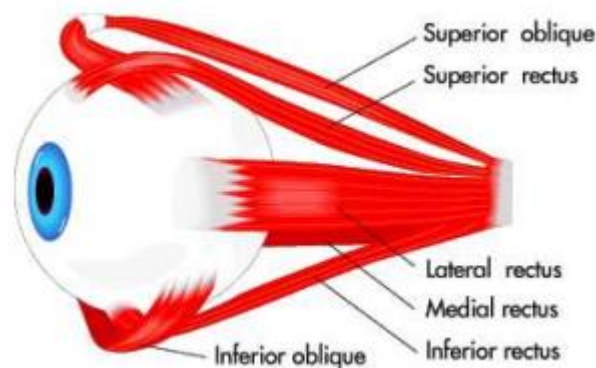


图 5.18: 六条肌肉肌腱附着在眼睛上，这样可以实现偏航，俯仰和少量滚动。

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。聚散运动产生的眼睛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。

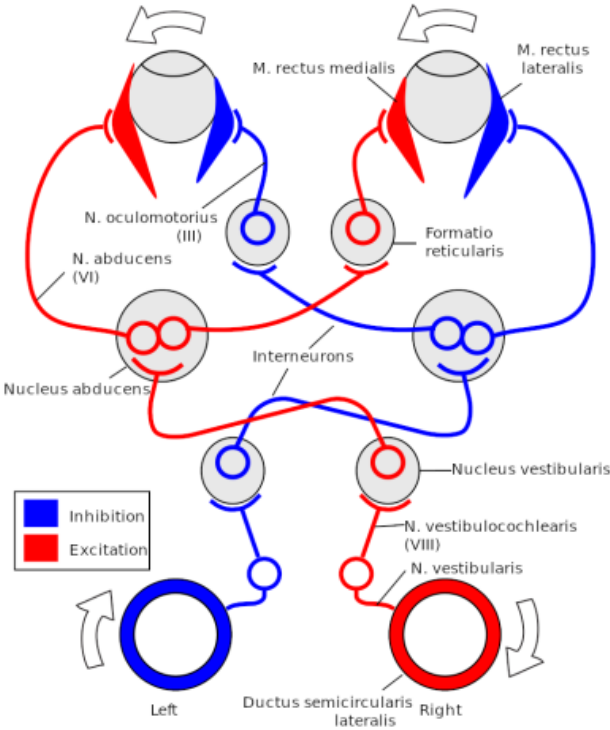


图 5.19: 前庭眼反射 (VOR)。 眼球肌肉与前庭器官内的角加速度计相连，以反向的眼球运动来抵抗头部运动，延迟小于 10 毫秒。 眼睛和前庭器官之间的连接由专门的前庭和眼外运动核提供，从而绕过更高的脑功能。

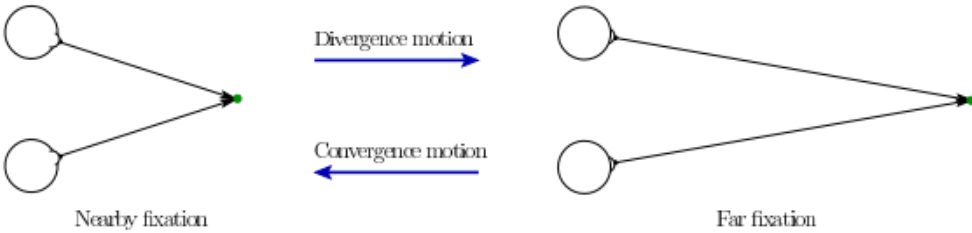


图 5.20: 在立体视觉的过程中，双眼都凝视着世界上的同一个特征。 要从近距离过渡到远距离特征，会发生背离运动。收敛运动则反之。

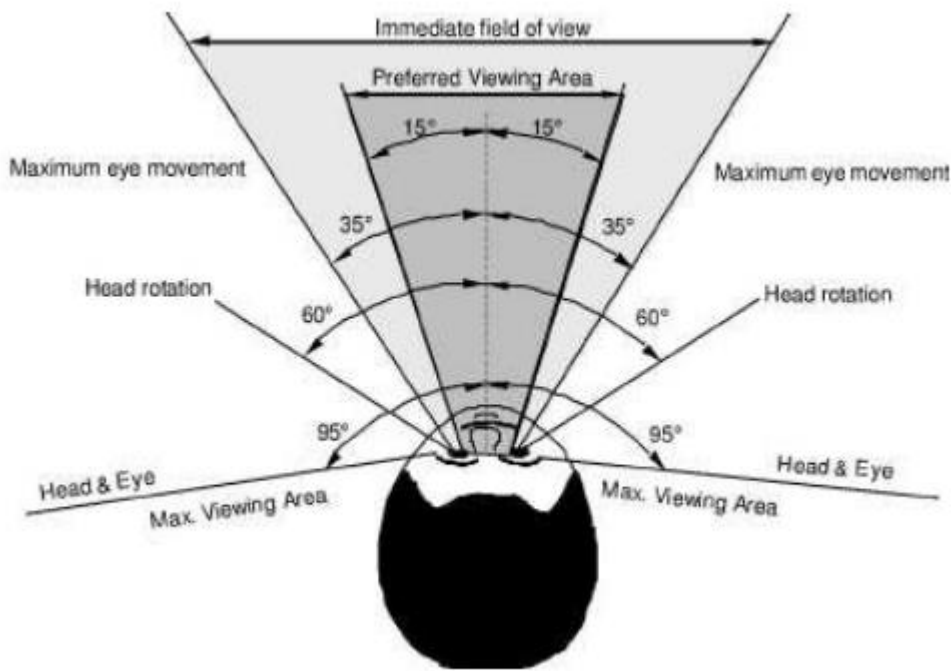


图 5.21: 头部和眼睛一起旋转以固定新的或移动的目标。

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° 左右对称，但双眼倾斜时却不是。人眼可以俯仰 20° 和俯仰 25° ，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。

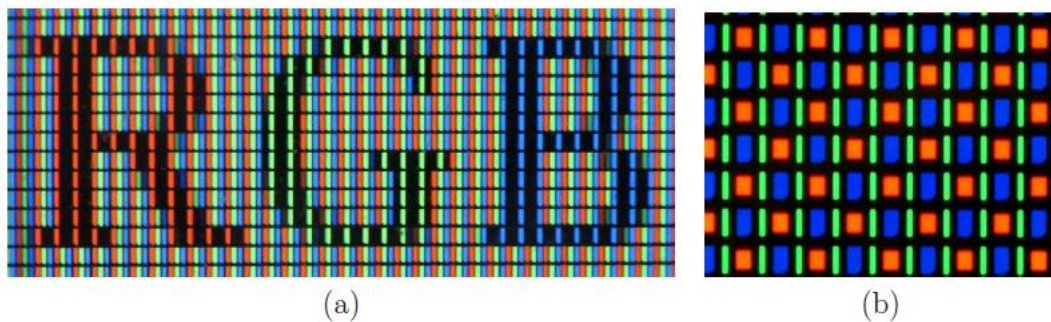


图 5.22: 显示过程中，像素以与光感受器分解成红、蓝、绿色成分相同的方式分解成子像素。 (a) 一个

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？

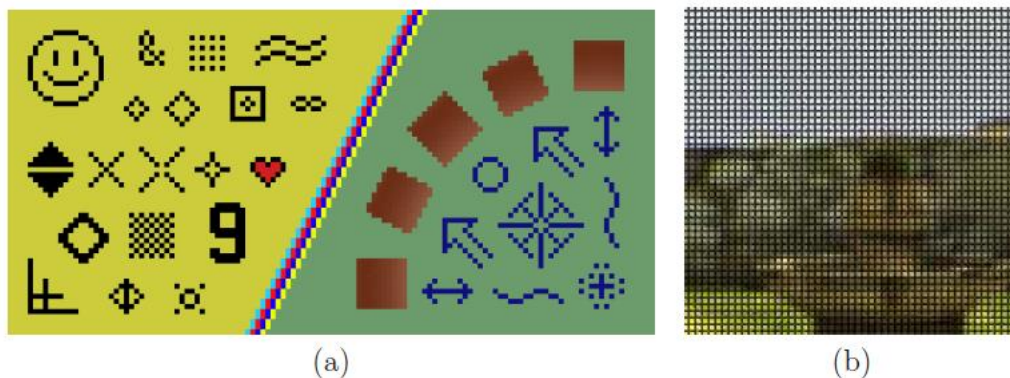
三个关键因素是：

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。另一个现象是纱窗效应，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**¹。这是否合理？它与 VR 又有什么关系？



¹ 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。

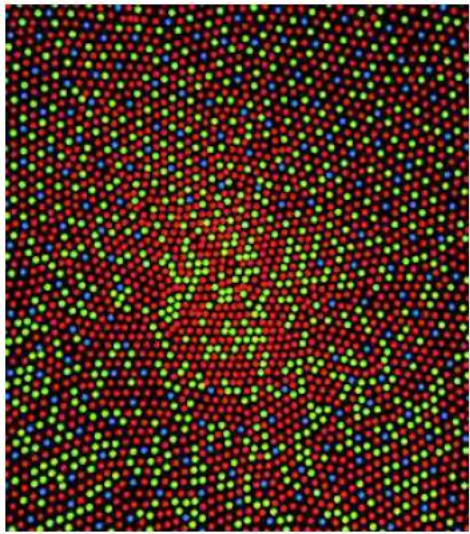


图 5.24：红色，绿色和蓝色光锥感光器分布在中心凹处复杂的马赛克中。

（图来自 Mark Fairchild）

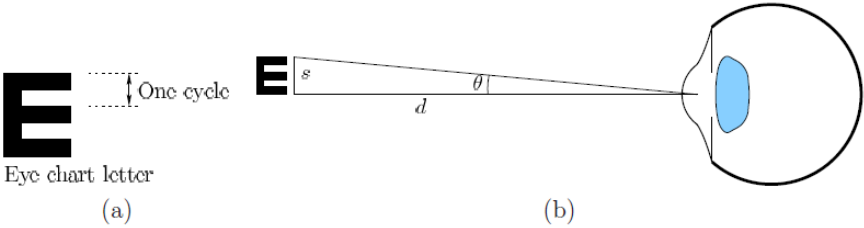


图 5.25：（a）视力表上的单个字母。（b）字母（或其他感兴趣特征）的大小 s ，观看者的距离 d 和视

角 θ 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 θ ，某个特征的大小 s 应该是多少。对于非常小的 θ ， $\tan \theta \approx \theta$ （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 θ 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，有优于 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° 左右，这比平面屏幕（小于 180°）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²。

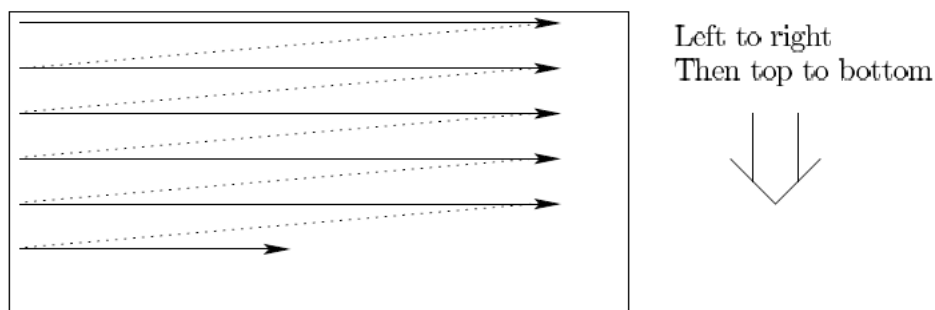


图 5.26：大多数显示器仍像旧电视机和 CRT 显示器一样工作：逐行更新像素。对于具有 60 FPS（每秒帧数）的显示器，这可能需要高达 16.67 毫秒。

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），屏幕上有荧光体元件。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。如果我們可以在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。

² 这在作者开发和测试 Oculus Rift 时经常发生。

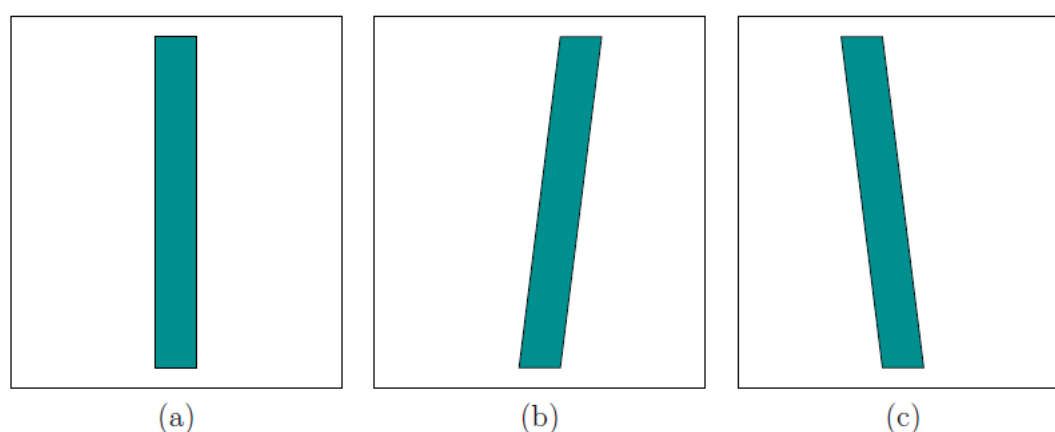


图 5.27：由于显示扫描造成的伪影：（a）场景中的垂直矩形。（b）矩形在虚拟世界中向右移动时，如何在平滑追踪期间扭曲。（c）在使用 VOR 进行补偿时，如果在将头旋转到右侧时固定矩形会如何变形。如果在每种情况下运动的方向相反，则（b）与（c）的情况交换。

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。结果被称为辐辏调节不匹配，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。