

人类视觉生理学

The Physiology of Human Vision

2019.11.21

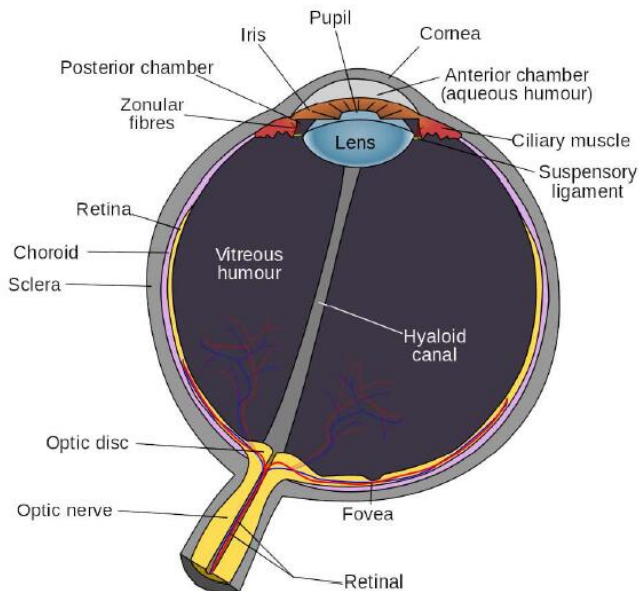
你所感知到的世界其实都是“你脑海中的世界”。

- 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是理想的。
- 视觉上近乎完美的图像呈现其实是一种错觉，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构建 VR 系统很重要。

Outline

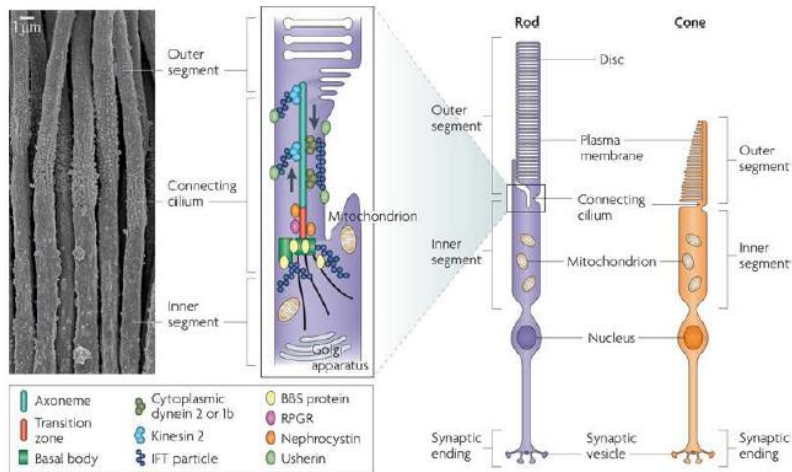
- ① 从角膜到感光体
- ② 从感光体到视觉皮层
- ③ 眼部运动
- ④ 对 VR 的影响

人眼的生理结构



感光体

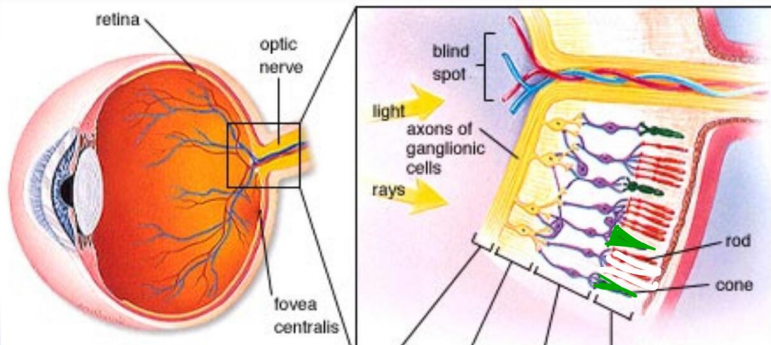
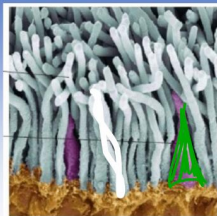
视网膜包含两种感光体：1) 视杆细胞，对弱光敏感；2) 视锥细胞，强光刺激才能引起兴奋，具有分辨颜色的能力。



感光体

Photoreceptors

There are two types:



感光体

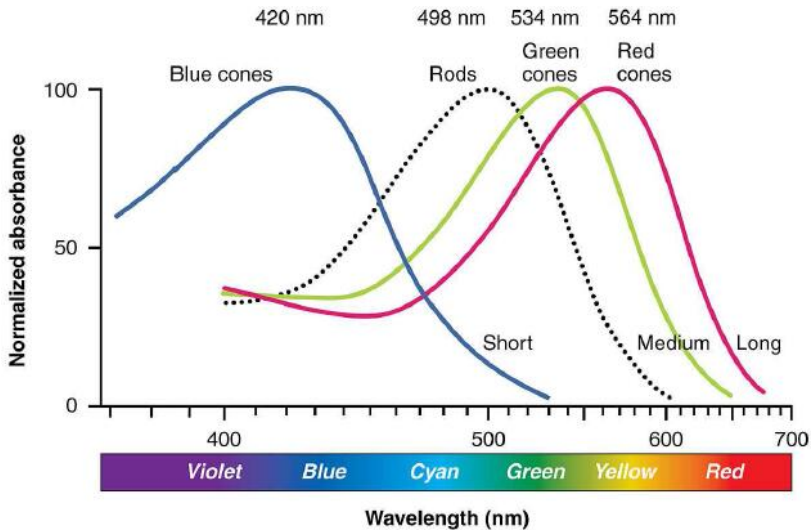


Figure 5.3: The sensitivity of rods and cones as a function of wavelength [29].

感光体

亮度 (luminance) 由国际单位坎德拉/平方米 (cd/m^2) 表示.

夜视觉: 单色模式

明视觉: 三色模式

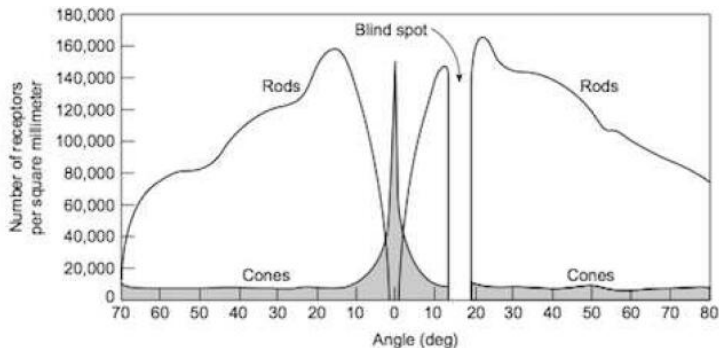
Light source	Luminance (cd/m^2)	Photons per receptor
Paper in starlight	0.0003	0.01
Paper in moonlight	0.2	1
Computer monitor	63	100
Room light	316	1000
Blue sky	2500	10,000
Paper in sunlight	40,000	100,000

Figure 5.4: Several familiar settings and the approximate number of photons per second hitting a photoreceptor. (Figure adapted from [163, 207].)

图片来自 [3]

感光体密度

- 中心凹: 最内部直径仅 0.5mm, 角度 $\pm 0.85^\circ$, 含几乎全部视锥细胞, 整个中心凹直径 1.5mm (角度范围 $\pm 2.6^\circ$), 外圈主要是视杆细胞。
- 周边视觉: 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。
- 盲点: 不包含任何感光体。



盲点测试



Figure 5.6: An experiment that reveals your blind spot. Close your right eye and look directly at the “X”. Vary the distance of the paper (or screen) from your eye. Over some range, the dot should appear to vanish. You can carry this experiment one step further by writing an “X” and dot on a textured surface, such as graph paper. In that case, the dot disappears and you might notice the surface texture perfectly repeating in the place where the dot once existed. This is caused by your brain filling in the expected texture over the blind spot!

图片来自 [3]

Outline

- ① 从角膜到感光体
- ② 从感光体到视觉皮层
- ③ 眼部运动
- ④ 对 VR 的影响

单个感光体对神经网络的影响

感光体：将光能刺激转化为神经冲动电信号的传感器，从而将关于外界的信息传入到神经结构中。

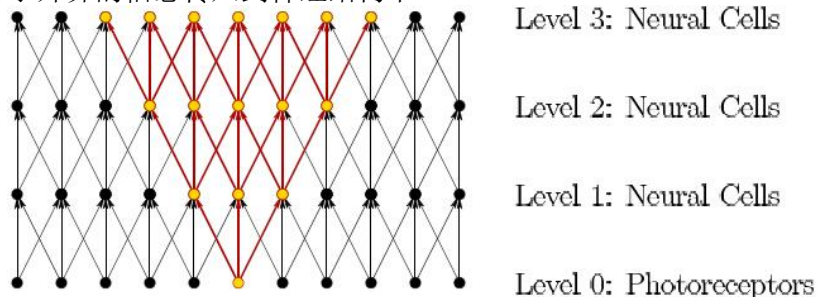


Figure 5.7: Four levels in a simple hierarchy are shown. Each disk corresponds to a neural cell or photoreceptor, and the arrows indicate the flow of information. Photoreceptors generate information at Level 0. In this extremely simplified and idealized view, each photoreceptor and neuron connects to exactly three others at the next level. The red and gold part highlights the growing zone of influence that a single photoreceptor can have as the levels increase.

图片来自 [3]

影响单个神经元的感光体

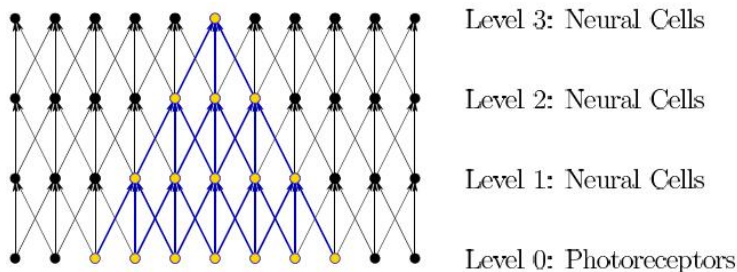


Figure 5.8: This diagram is the same as Figure 5.7 except that the information feeding into a single neuron is highlighted. Consider the set of photoreceptors involved in the reaction of a single neural cell. This is called the *receptive field*. As the level increases, the receptive field size grows dramatically. Due to the spatial arrangement of the photoreceptors, this will imply that each neuron responds to a growing patch in the image on the retina. The patch increases in size at higher levels.

图片来自 [3]

影响单个神经元的感光体

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。

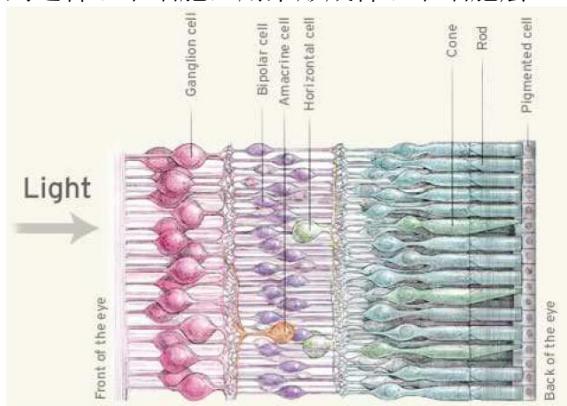


Figure 5.9: Light passes through a few neural layers before hitting the rods and cones. (Figure by the Institute for Dynamic Educational Advancement.)

图片来自 [2]

从感光体到视觉皮层

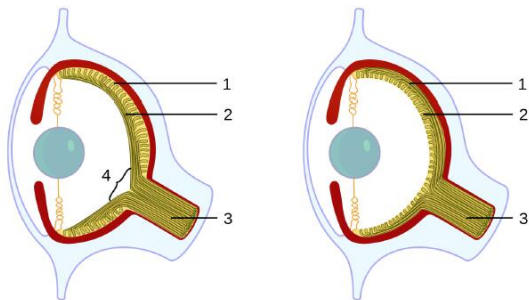


Figure 5.10: Vertebrates (including humans) have inside-out retinas, which lead to a blind spot and photoreceptors aimed away from the incoming light. The left shows a vertebrate eye, and the right shows a cephalopod eye, for which nature got it right: The photoreceptors face the light and there is no blind spot. (Figure by Jerry Crimson Mann.)

图片来自 [3]

内核层 (Inner nuclear layer)

- 双极细胞 (Bipolar cells): 从感光器连接到神经节细胞, 有 ON (光子吸收率增加时激活) 和 OFF (光子吸收率减少时激活) 两种类型。
- 水平细胞 (Horizontal cells): 连接 1mm 半径范围的感光器输入 (树突) 与双极细胞。
水平细胞的输出 (轴突) 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。
- 无长突细胞 (Amacrine cells): 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, 作用不明。

神经节层 (Ganglion cell layer)

- 神经节细胞的三种常见类型:midget, parasol, and bistratified。它们对其接收区域中感光器刺激中的空间，时间和光谱（颜色）变化，执行简单的滤波操作。
- 神经节细胞可视为微小图像处理单元，可检测和强调简单的图像特征，如边缘。当神经节轴突通过视神经离开眼睛时，已经执行了大量的图像处理以帮助视觉感知。

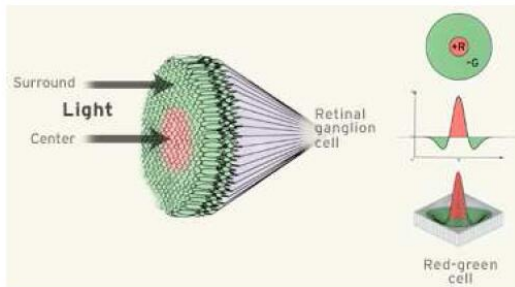


Figure 5.11: The receptive field of an ON-center ganglion cell. (Figure by the Institute for Dynamic Educational Advancement.)

从感光体到视觉皮层

- 视神经连接到称为外侧膝状核（lateral geniculate nucleus, LGN）的一部分丘脑（thalamus）。
- 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（primary visual cortex, V1）。

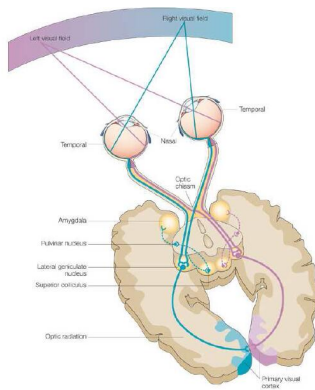


Figure 5.12: The visual pathway from the eyes to the LGN to the visual cortex. Note that information from the right and left sides of the visual field becomes swapped in the cortex. (Figure from Nature Reviews: Neuroscience)

图片来自 [3]

视觉皮层

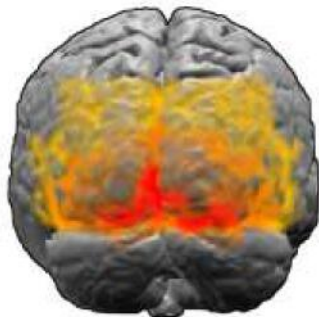
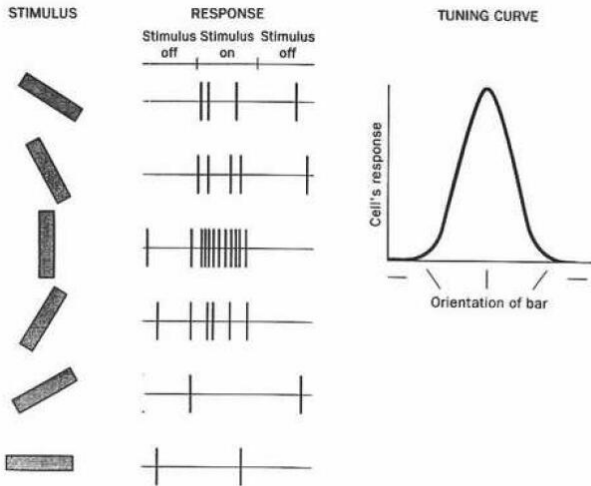


Figure 5.13: The visual cortex is located in the back of the head (Figure by Washington Irving).

图片来自 [3]

视觉皮层

视觉感知是视觉皮层处理的有意识的结果，基于神经回路，视网膜刺激，来自其他感官的信息以及基于先前经验的预期。



Outline

- ① 从角膜到感光体
- ② 从感光体到视觉皮层
- ③ 眼部运动
- ④ 对 VR 的影响

眼部运动

眼球旋转是人类视觉的复杂且不可分割的部分。眼球运动的原因：

- 将感兴趣的特征定位在中央凹上;
- 眼睛的光感受器刺激响应缓慢，需长达 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;
- 实验表明，当眼球运动完全被抑制时，视觉感知完全消失。

眼部运动



Figure 5.15: The trace of scanning a face using saccades.

图片来自 [3]

眼部运动

随着运动的结合以形成一致的观点，科学家很难预测和解释人们如何解释某种刺激。



Figure 5.16: The fractal appears to be moving until you carefully fixate on a single

眼睛肌肉

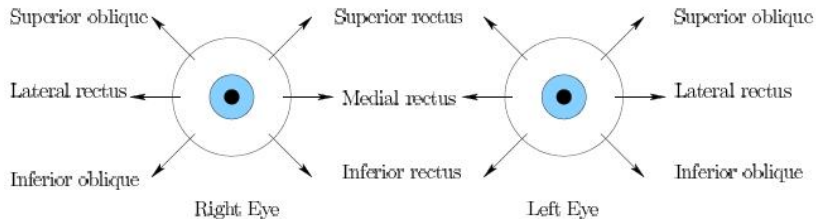


Figure 5.17: There are six muscles per eye, each of which is capable of pulling the pupil toward its location.

图片来自 [3]

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）。

肌腱以相反的方式牵拉眼睛。

在多数情况下，可合理将眼旋转估计为仅包含偏航和俯仰的 2D 集合。

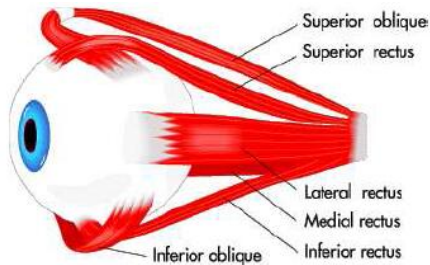


Figure 5.18: The six muscle tendons attach to the eye so that yaw, pitch, and a small amount of roll become possible.

图片来自 [3]

眼球运动类型

眼球运动共有六种类型：

- 快速扫视 (saccades)
- 平稳移动 (smooth pursuit)
- 前庭眼反射 (vestibulo-ocular reflex)
- 视动反射 (optokinetic reflex)
- 聚散 (vergence)
- 微跳动 (microsaccades)

睡觉的时候

- 快速眼动 (REMs)

快速扫视 (saccades)

- 眼睛可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。
- 大脑使用扫视掩蔽来隐藏从记忆中出现扫视的时间间隔。这会导致时间感知扭曲。



Figure 5.15: The trace of scanning a face using saccades.

图片来自 [3]

平稳移动 (smooth pursuit)

- 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，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。

前庭眼反射 (vestibulo-ocular reflex, VOR)

- 眼球旋转抵消头部的旋转，运动控制绕过了更高的大脑功能。
- 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。

前庭眼反射 (vestibulo-ocular reflex, VOR)

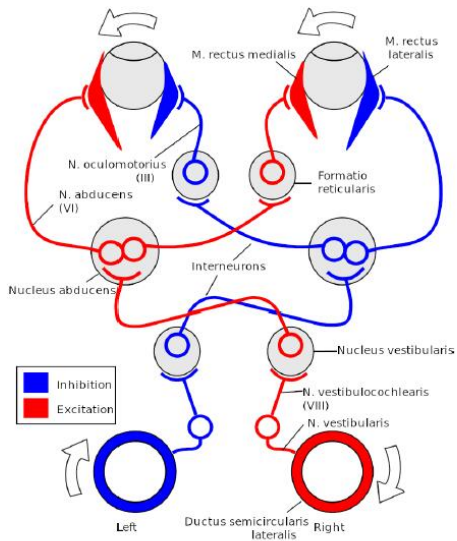


Figure 5.19: The vestibulo-ocular reflex (VOR). The eye muscles are wired to

视动反射 (optokinetic reflex)

- 在快速物体快速移动时发生，如站在附近的固定地面上在观看快速行驶的列车。
- 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 (vergence)

聚散运动来将眼睛与物体对齐，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。

- 如果物体比先前的固定更近，则会发生收敛运动。眼睛旋转，瞳孔越来越近了。
- 如果物体更远，则发生发散运动，这导致瞳孔进一步分开。

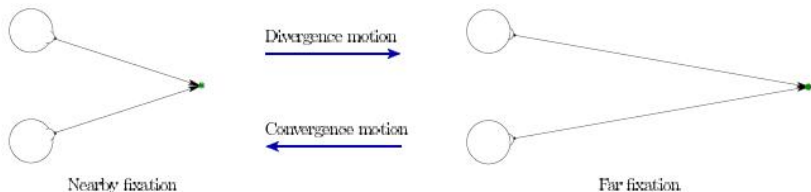


Figure 5.20: In the process of stereopsis, both eyes are fixated on the same feature in the world. To transition from a close to far feature, a divergence motion occurs. A convergence motion happens for the opposite transition.

图片来自 [3]

微跳动 (microsaccades)

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。

- 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。
- 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，行为非常复杂，并且没有完全理解。

眼睛和头部运动结合

大部分时间眼睛和头部在一起移动

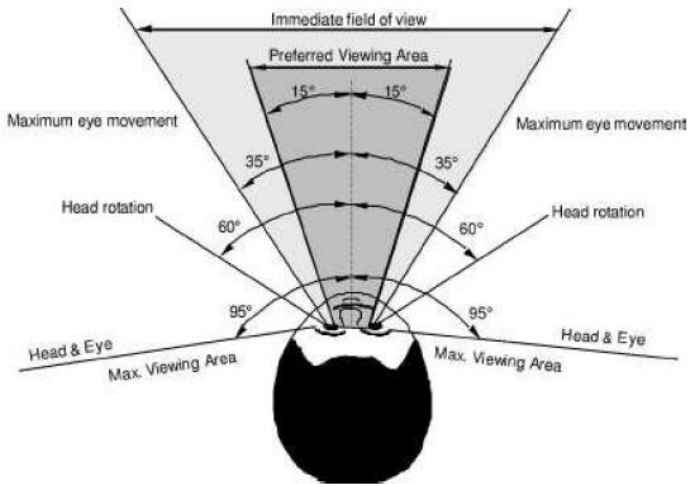


Figure 5.21: The head and eyes rotate together to fixate on new or moving targets.
(Figure from MSC/Cine 082.20 December 2000.)

Outline

- ① 从角膜到感光体
- ② 从感光体到视觉皮层
- ③ 眼部运动
- ④ 对 VR 的影响

VR 视觉显示器性能需要有多好？

三个关键因素是：

- 空间分辨率：每平方面积需要多少像素？
- 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？
- 时间分辨率：显示器改变其像素需要多快？

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

Figure 5.4: Several familiar settings and the approximate number of photons per second hitting a photoreceptor. (Figure adapted from [163, 207].)

图片来自 [3]

多少像素密度是足够的？

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。
这是否合理？它与 VR 又有什么关系？



(a)



(b)

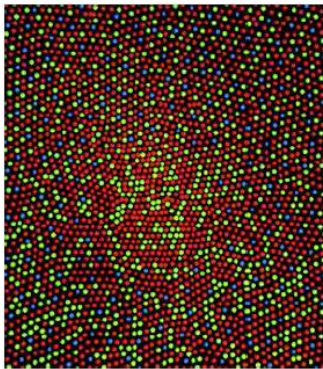
Figure 5.22: (a) Due to pixels, we obtain a bad case of the *jaggies* (more formally known as *aliasing*) instead of sharp, straight lines. (Figure from Wikipedia user Jmf145.) (b) In the *screen-door effect*, a black grid is visible around the pixels.

图片来自 [3]

多少像素密度是足够的？

假设中央凹直接指向显示器以提供最佳感测。第一个问题是红色，绿色和蓝色圆锥体分布像马赛克似的。

视觉科学家和神经生物学家已经通过视敏度的方法研究了有效或可感知的输入分辨率。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。



多少像素密度是足够的？

最为广泛使用的概念之一是每度的周期数 (cycle per degree)，它大致对应于可以沿着观察弧线看到的条纹（或正弦波峰）数。

正常视力：刚好能分辨出每度 30 个周期。 $s = d \tan \theta$

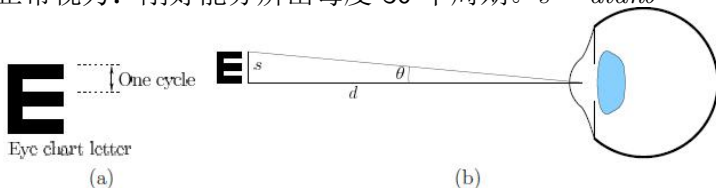


Figure 5.24: (a) A single letter on an eye chart. (b) The size s of the letter (or other feature of interest), the distance d of the viewer, and the viewing angle θ are related as $s = d \tan \theta$.

图片来自 [3]

Snellen 视力表

Snellen 视力表：当眼睛距离 20 英尺时，20/20 行当字母高度对应于每度 30 个循环。“E” 的总高度为 1/12 度。

E	1	20/200
F P	2	20/100
T O Z	3	20/70
L P E D	4	20/50
P E C F D	5	20/40
E D F C Z P	6	20/30
F E L O P Z D	7	20/25
D E F F O T E C	8	20/20
L E F O D P C T	9	
F D F L T C E O	10	
F E R O L C F T S	11	

多少像素密度是足够的？

对于离眼睛距离为 d 的观看视角 θ ，特征大小 s 应为： $s = d \tan \theta$
为了产生每度 30 个周期，每度必须至少有 60 个像素。

- 假设正常视力的人观看距离 20 英尺（6.096 米）的大屏幕，要求屏幕至少有 $60 / 4.189 = 14.32\text{PPI}$
- 假设智能手机屏幕距使用者眼睛 12 英寸，要求屏幕至少有 $60 / 0.209 = 286.4\text{PPI}$
- VR 头显设备距离只有 1.5 英寸，显示器必须至少有 2291.6 PPI 以实现每度 60 个像素



Figure: Sony Xperia Z5 Premium (801 PPI)

多少视野是足够的？

基于感光体密度图和眼旋转限制中，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，但镜头外围光学失真可能会限制有效的视野。

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。

Foveated rendering

目前显示设备没有利用光感受器密度远离中心凹处减少的特点，在整个区域保持高像素密度。

如果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，需要的像素就会变少，将大大减少图形渲染系统的计算负担。

Foveated 渲染：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上述过程。

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。

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）。

- 头部俯仰和左右摇晃的情况下，VOR 增益接近 1.0。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。

适应是我们感官系统的一个普遍特征，VOR 增益也不例外。

- 戴眼镜的人，镜片影响视野并且感知物体的大小和距离，VOR 通过改变增益来适应。
- 遭受缺陷的 VR 头显设备，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，平稳感觉在现实世界中会变得扭曲，直到你重新适应。

显示扫描

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

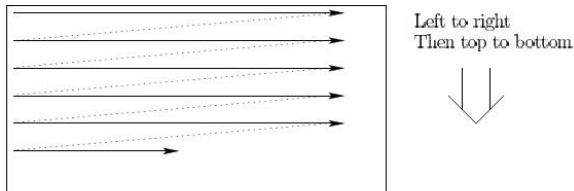


Figure 5.25: Most displays still work in the way as old TV sets and CRT monitors: By updating pixels line-by-line. For a display that has 60 FPS (frames per second), this could take up to 16.67ms.

图片来自 [3]

显示扫描

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。

一种可能性是通过渲染一个失真的图像来解决这个问题，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。

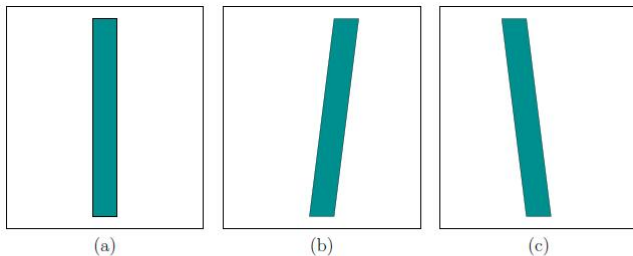


Figure 5.26: Artifacts due to display scanout: (a) A vertical rectangle in the scene. (b) How it may distort during smooth pursuit while the rectangle moves to the right in the virtual world. (c) How a stationary rectangle may distort when

辐辏调节不匹配 (Vergence-accommodation mismatch)

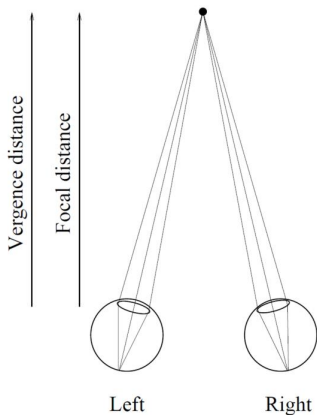
调节是改变眼睛镜片的光学能力的过程，以便可以将近距离物体聚焦。

- 距离较近时，在眼睛强烈聚焦时，眼睛会尝试增加镜头的力量。
- 距离合适放在焦距处，眼睛放松、且总聚焦。

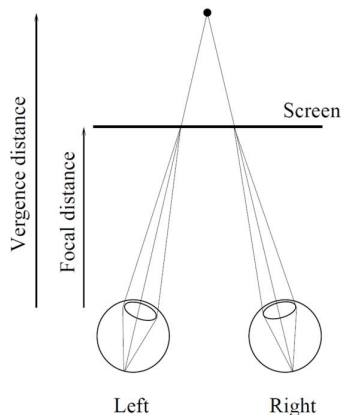
VR 中，物体被渲染到屏幕上让它看起来只有 10 厘米的距离，眼睛强烈地聚焦，但不需要改变眼睛镜头的光焦度。

眼睛可能会尝试调节，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。

辐辏调节不匹配 (Vergence-accommodation mismatch)



(a) Real World



(b) 3D Display

Any Questions?