

视觉感知 Visual Perception

2019.11.25

Outline

- ① 感知深度
- ② 运动感知
- ③ 对颜色的感知
- ④ 结合信息来源

感知深度

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
- 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

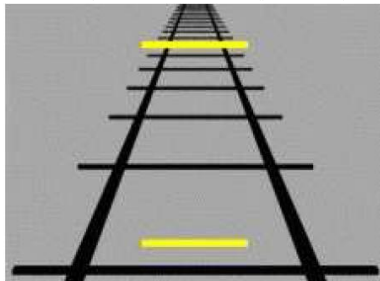
单眼与立体线索

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。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。

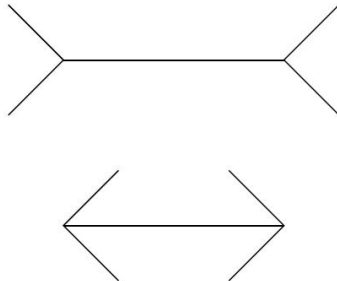


单眼与立体线索

简单的线条图也足以提供强有力的线索



(a)



(b)

Figure 6.2: Even simple line drawings provide significant cues. (a) The Ponzo illusion: The upper yellow bar appears to be longer, but both are the same length. (b) The Müller-Lyer illusion: The lower horizontal segment appears to be shorter than the one above, but they are the same length.

图片来自 [3]

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。

- 观众必须熟悉该物体，以便知道其实际尺寸
- 物体必须自然出现，以免与其他深度线索发生冲突

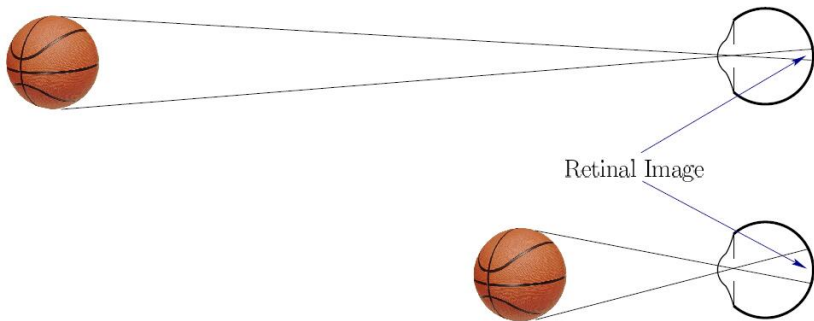


Figure 6.3: The retinal image size of a familiar object is a strong monocular depth cue. The closer object projects onto a larger number of photoreceptors, which cover a larger portion of the retina.

图片来自 [3]

单目深度线索

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。

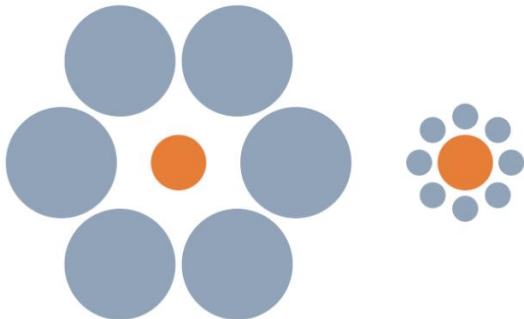
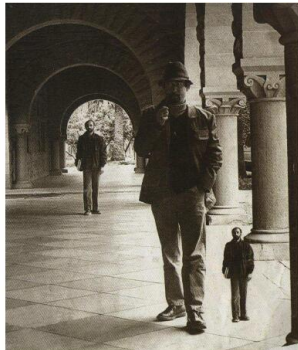
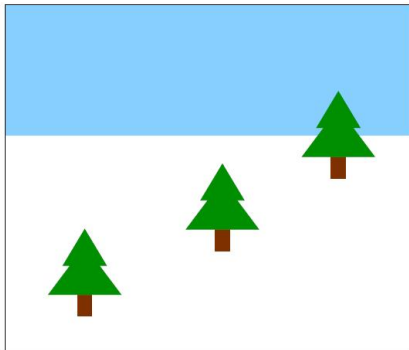


Figure 6.4: For the Ebbinghaus illusion, the inner disc appears larger when surrounded by smaller discs. The inner disc is the same size in either case. This may be evidence of discrepancy between the true visual angle (or retinal image size) and the *perceived* visual angle.

视野中的高度

另一个重要的线索：视野中物体的高度

- 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。
- 可将尺寸恒定标度与视野中的高度相结合。



适应（Accommodation）

人眼晶状体可通过适应过程改变其光焦度。

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。

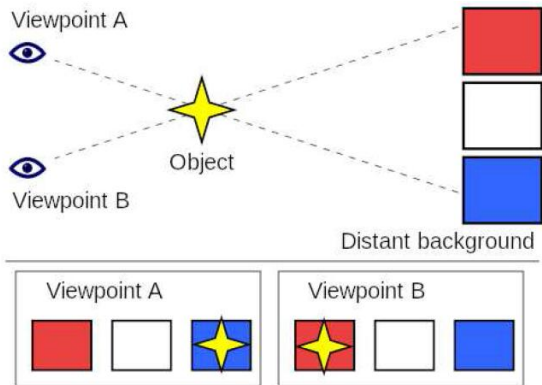
这是不依赖于光感受器产生的信号的第一个深度线索。

运动视差

速度的相对差异称为视差，是一个重要的深度线索。

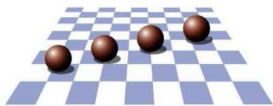
如从快速移动的车辆侧窗看过去，附近的物体比其他物体的速度快得多。

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。

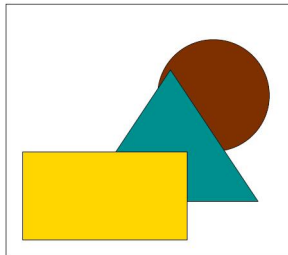


其他单眼线索

阴影；序数深度信息；图像模糊线索；大气线索



(a)



(b)



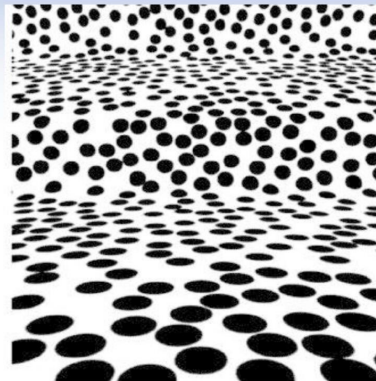
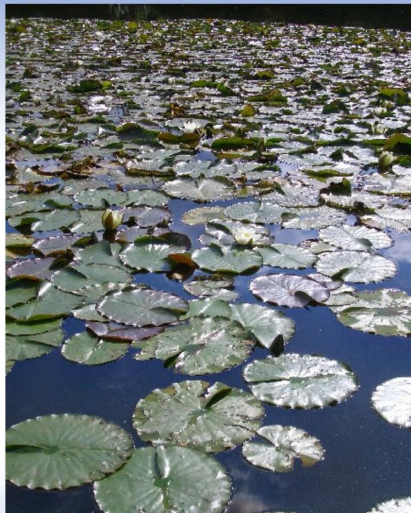
(c)



(d)

单眼线索

Depth Perception:



单眼线索

Depth Perception:



立体深度线索

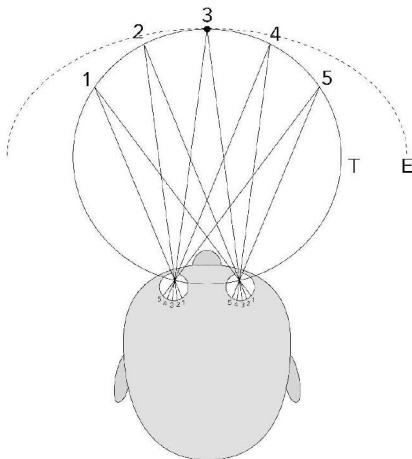


Figure 6.8: The *horopter* is the loci of points over which the eyes can converge and focus on a single depth. The T curve shows the theoretical horopter based on simple geometry. The E curve shows the empirical horopter, which is much larger and correspond to the region over which a single focused image is perceived. (Figure by Rainer Zenz.)

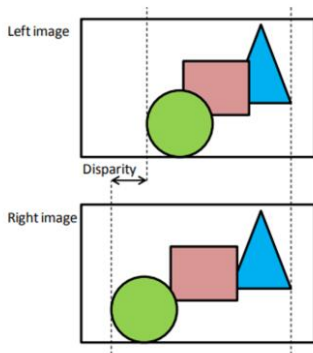
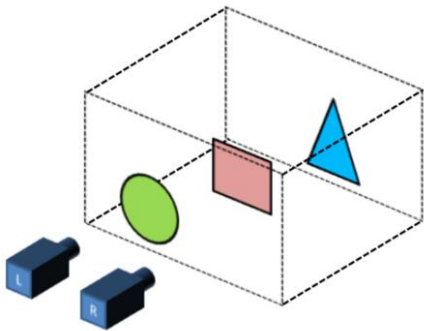
Optimal Focal Curve

Optimal Focal Curve



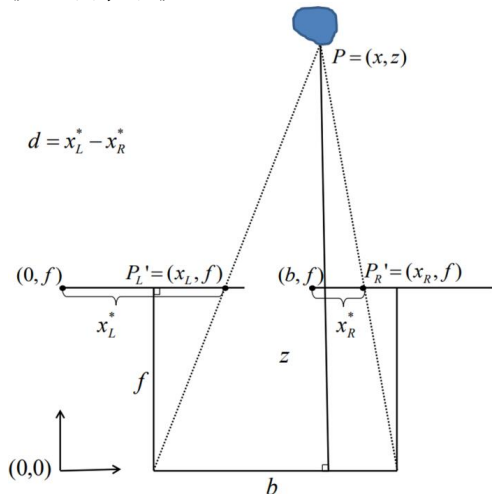
双目视差 disparity

简单的立体相机模型：左右摄像机是相同的（相同的镜头，相同的焦距等）。相机的光轴是平行的，指向 Z 方向即场景深度。



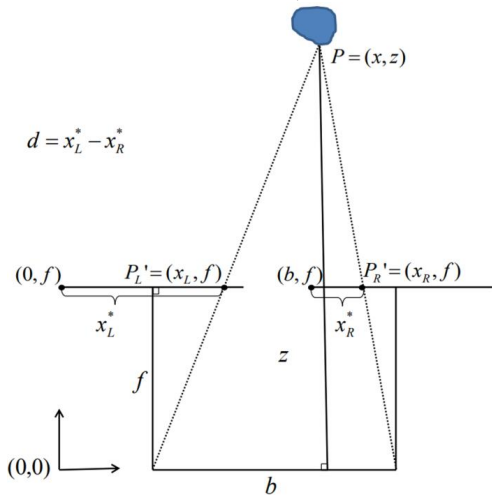
双目视差 disparity

简单的立体相机模型: f 是焦距, b 是基线的长度, $d = x_L^* - x_R^*$ 是视差或水平视差



双目视差 disparity

简单的立体相机模型: f 是焦距, b 是基线的长度, $d = x_L^* - x_R^*$ 是视差或水平视差: $d = fb/z$



双目视差 disparity

内旋立体相机模型：

- 类似于人类视觉系统
- 当人类将物体聚焦在特定距离处时，我们的眼睛围绕垂直轴沿相反方向旋转，以获取（或保持）单眼双目视觉-所谓的聚散。
- 焦点物体上以及与观察者的距离与焦点物体相同的位置处的视差 $d = 0$ 。
- 对于较近的物体（负视差）， $d < 0$ 。
- 对于比聚焦物体更远的物体（所谓的正视差）， $d > 0$ 。

复视问题 diplopia

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，可能与熟悉的物体相似，但其确切比例可能难以确定。

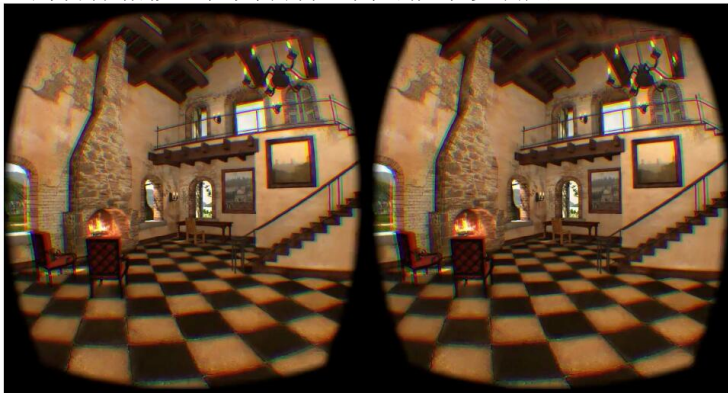


Figure 6.9: In the Tuscany demo from Oculus VR, there are not enough familiar objects to precisely resolve depth and size. Have you ever been to a villa like this? Are the floor tiles a familiar size? Is the desk too low?

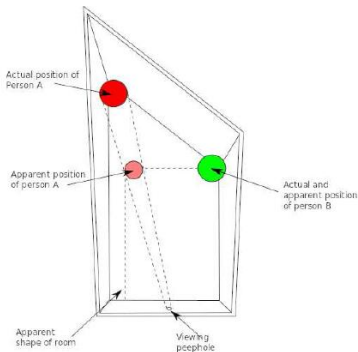
图片来自 [3]

对 VR 的影响-失谐

在 VR 系统中，很容易造成不匹配，如辐辏调节不匹配、不完美的头部追踪。



(a)



(b)

Figure 6.10: The Ames room: (a) Due to incorrect depth cues, incorrect size perception results. (b) The room is designed so that it only appears to be rectangular

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

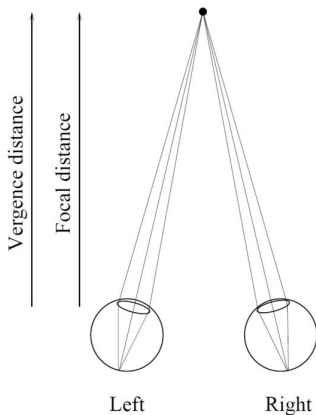
调节 (焦距) 当物体的距离改变时, 通过睫状肌控制晶状体形状在物体上保持清晰 (聚焦) 的图像。

辐辏 (收敛距离) - 当物体的距离改变时, 眼睛的聚散运动将眼睛和物体对齐。

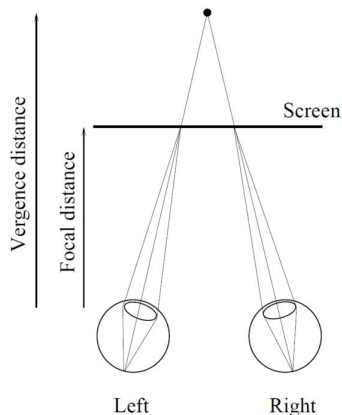
- 在人类视觉中, 通常焦距 = 收敛距离。
- 大多数 3D 视频/电影/电视观看要求焦距 \neq 收敛距离。

这造成了 Vergence-Accommodation 冲突-观看 3D 视频/电视/电影时眼睛疲劳和劳损的主要原因之一。

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



(a) Real World



(b) 3D Display

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。
单眼深度线索比立体线索更多。

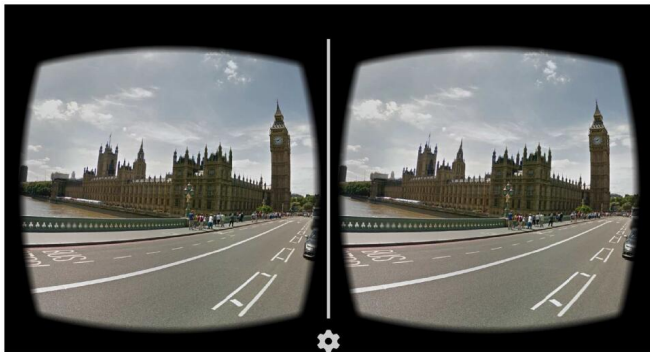


Figure 6.11: In Google Cardboard and other VR headsets, hundreds of millions of panoramic Street View images can be viewed. There is significant depth perception, even when the same image is presented to both eyes, because of monoscopic depth cues.

图片来自 [3]

Outline

- ① 感知深度
- ② 运动感知
- ③ 对颜色的感知
- ④ 结合信息来源

运动感知

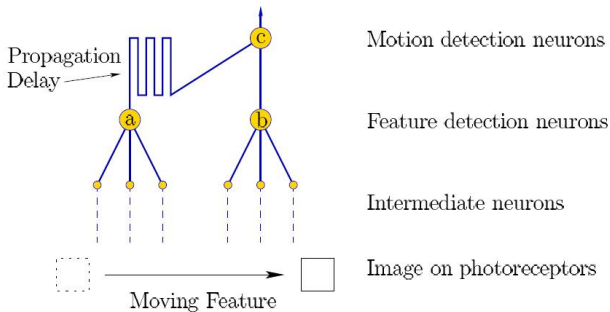
我们往往依靠视力来感知运动。

- 将移动的图形与静止的背景分开。
- 帮助人们评估物体的三维结构。
- 指导人类行动。
- 平稳感知，非运动的感知。

检测机制

Reichardt 探测器：对人类视觉系统中的定向运动产生响应。

- 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。



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。

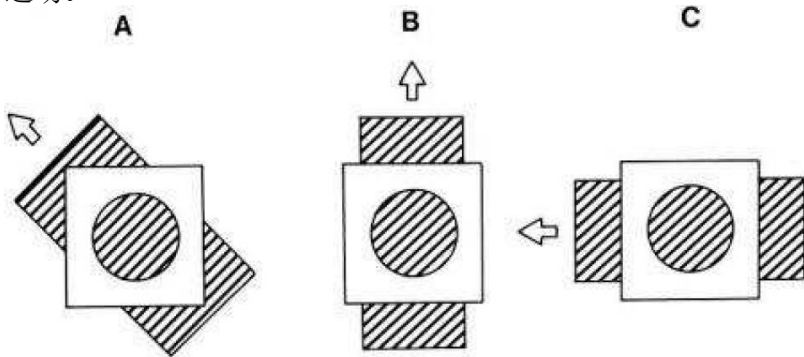


Figure 6.13: Due to local nature of motion detectors, the *aperture problem* results. The motion of the larger body is ambiguous when perceived through a small hole because a wide range of possible body motions could produce the same effect inside of the hole. An incorrect motion inference usually results.

区分物体运动和观察者运动

两种情况下，视网膜图像变化相同

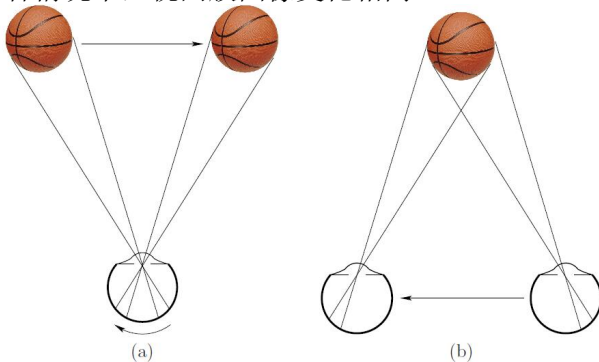


Figure 6.14: Two motions that cause equivalent movement of the image on the retina: (a) The eye is fixed and the object moves; (b) the eye moves while the object is fixed. Both of these are hard to achieve in practice due to eye rotations (smooth pursuit and VOR).

线索：Saccadic 抑制、眼部肌肉命令、大范围运动信息。

图片来自 [3]

光流 Optical Flow

光流

- 跟踪视网膜 (或图像平面, 或屏幕) 上特征的运动
- 图像平面 (球平面) 上的矢量场

频闪视运动

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。



Figure 6.15: The *zoetrope* was developed in the 1830s and provided stroboscopic apparent motion as images became visible through slits in a rotating disc.

频闪视运动

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。频闪视运动起作用的最可能原因是它触发了神经运动检测电路。

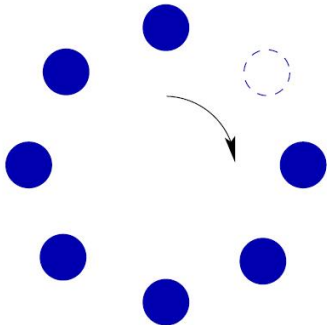


Figure 6.16: The *phi phenomenon* and *beta movement* are physiologically distinct effects in which motion is perceived [355, 313]. In the sequence of dots, one is turned *off* at any give time. A different dot is turned *off* in each frame, following a

帧率

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

Figure 6.17: Various frame rates and comments on the corresponding stroboscopic apparent motion. Units are in Frames Per Second (FPS).

图片来自 [3]

对 VR 的影响

VR 系统需要比平常更高的显示性能。

戴着 VR 头戴式显示器并注视虚拟物体时转动头部，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，其结果是一种抖动，物体会高频率但小幅度地左右晃动

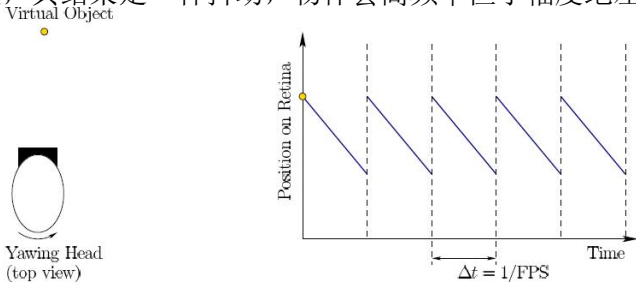


Figure 6.18: A problem with *perception of stationarity* under stroboscopic apparent motion: The image of a feature slips across the retina in a repeating pattern as the VOR is performed.

图片来自 [3]

对 VR 的影响

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。

低持久（low persistence）显示：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

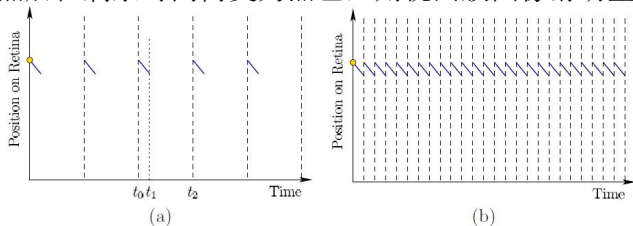


Figure 6.19: An engineering solution to reduce retinal image slip: (a) Using *low persistence*, the display is lit for a short enough time to trigger photoreceptors ($t_1 - t_0$) and then blanked for the remaining time ($t_2 - t_1$). Typically, $t_1 - t_0$ is around one to two milliseconds. (b) If the frame rate were extremely fast (at least

少。 500 FPS), then the blank interval would not be needed.

图片来自 [3]

Outline

- ① 感知深度
- ② 运动感知
- ③ 对颜色的感知
- ④ 结合信息来源

颜色感知

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。



Figure 6.20: In 2014, this dress photo became an Internet sensation as people were unable to agree upon whether it was “blue and black” or “white and gold”, which are strikingly different perceptions of color.

图片来自 [3]

降维

光能是具有光谱能量分布的不同波长或波强度的集合
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

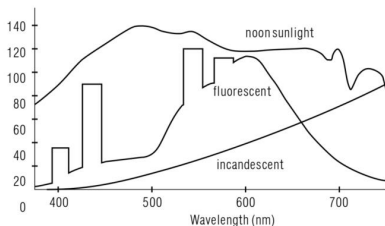


Figure 4.6: The *spectral power distribution* for some common light sources. (Figure from [226]).

图片来自 [3]

降维

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，

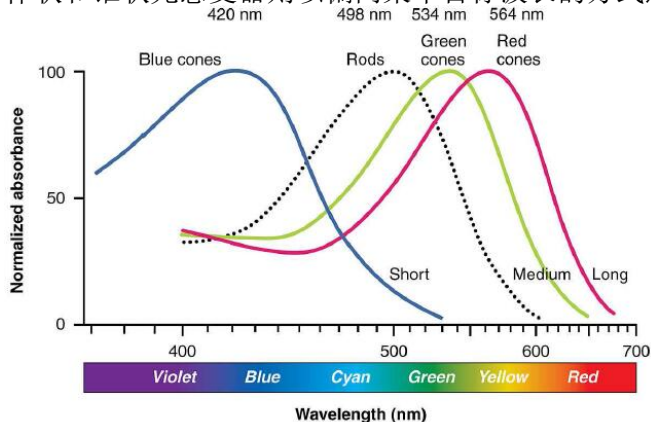


Figure 5.3: The sensitivity of rods and cones as a function of wavelength [29].
(Figure adapted by OpenStax College.)

图片来自 [3]

色彩空间

计算机图形学中最常见的一种色彩空间称为 HSV

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

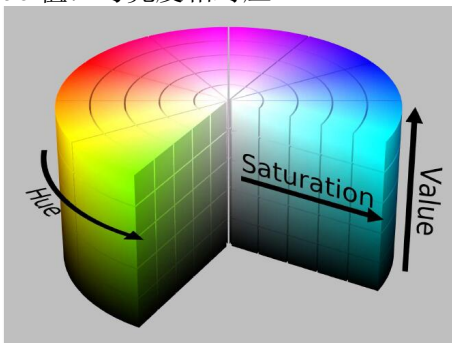
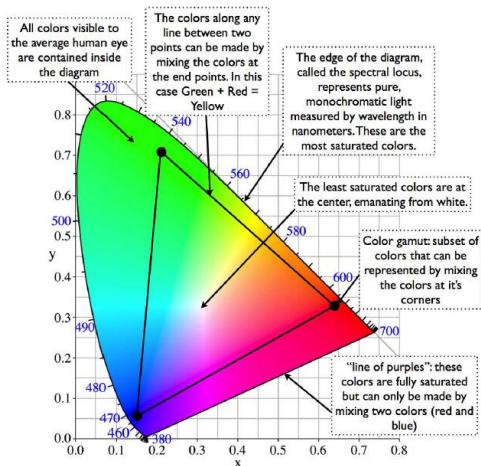


Figure 6.21: One representation of the HSV color space, which involves three parameters: hue, saturation, and value (brightness). (Figure by Wikipedia user

混合颜色



Anatomy of a CIE Chromaticity Diagram

Figure 6.22: 1931 CIE color standard with RGB triangle. This representation is correct in terms of distances between perceived colors. (Figure by Jeff Yurek, Nanosys.)

恒定性

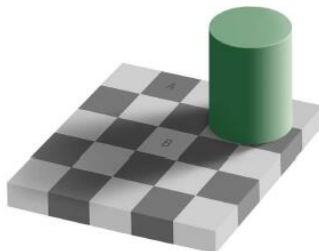
人的色彩感觉对于颜色来源有惊人的鲁棒性

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

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著变化之后



(a)



(b)

Figure 6.23: (a) The perceived hot air balloon colors are perceived the same regardless of the portions that are in direct sunlight or in a shadow. (Figure by Wikipedia user Shanta.) (b) The *checker shadow illusion* from Section 2.3 is ex-

Outline

- ① 感知深度
- ② 运动感知
- ③ 对颜色的感知
- ④ 结合信息来源

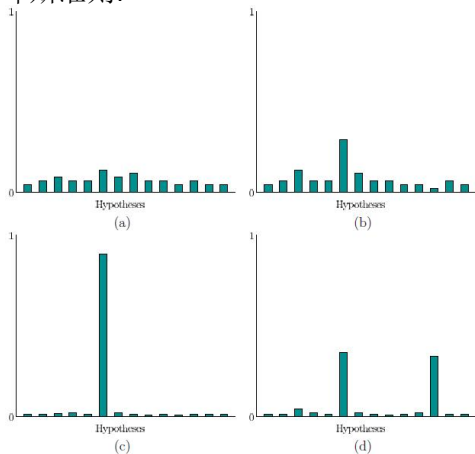
结合信息来源

感知过程结合多种来源信息。

- 可以结合同种感觉器官的多个线索，如用于判断深度的众多单眼线索
- 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
- 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。

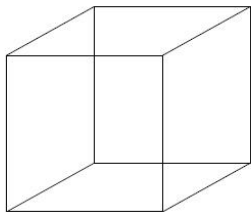
结合信息来源

信息是从先前的期望以及接受许多暗示中综合而来的, 统计决策理论提供了一个有用且直接的数学模型. 一个关键组成部分是贝叶斯准则.

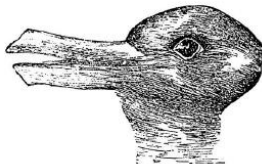


多重感知

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。



(a)



(b)

Figure 6.26: (a) The Necker cube, studied in 1832 by Swiss crystallographer Louis Albert Necker. (b) The *rabbit duck illusion*, from the 23 October 1892 issue of *Fliegende Blätter*.

McGurk 效应

McGurk 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

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

- 如果听到“ba”并且显示“ga”，则大多数会感知到“da”
- 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

- **Vection:** 视觉和前庭信息之间引起晕眩的矛盾
- 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。

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。

Any Questions?