



交流图片

图像和视频编码课程

2014 年, 第 17-61 页

第2章-人类视觉系统

大卫·R·布尔

显示更多 ∨

☰ Outline | 🔗 共享 🔖 引用

<https://doi.org/10.1016/B978-0-12-405906-1.00002-7>

获取权利和内容

摘要

本章旨在让读者对**人类视觉系统** (HVS) 有基本的了解, 并在可能的情况下将其特征与视觉冗余并最终与压缩图像和视频信号的方法联系起来。它首先考虑了HVS的物理架构, 以及它对我们看待世界的方式施加的限制。然后, 它回顾了视觉中与亮度、对比度、纹理、颜色和运动相关的感知方面, 指出关键视觉参数的物理限制。视觉遮罩是基于感知的压缩的一个关键组成部分, 本章以研究某些时空视觉特征如何被某些类型的内容遮盖而结束。



关键词

人眼; 人类视觉系统 (HVS); 视觉皮层; 视觉敏锐度; 色彩处理; 空间处理; 对比敏感性; 视觉感知; 眼睛运动; 视觉遮罩

对**人类视觉系统** (HVS) 的结构、功能和感知限制的研究为我们提供了线索, 说明如何利用视觉信息的冗余, 以最小的感知质量下降来压缩它。人类看待和感知数字图像和视频的方式与我们在日常生活中如何感知世界密切相关。虽然对HVS的完全理解还有很长的路要走, 但我们确实有足

够的知识，能够在图像序列中创造高质量的错觉，其中可能每200位中就有199位被丢弃。

本章旨在让读者对[人类视觉系统](#)有基本的了解，并在可能的情况下将其特征与视觉冗余并最终与压缩图像和视频信号的方法联系起来。我们首先考虑HVS的物理架构及其对我们看待世界的方式施加的限制。然后，我们回顾了视觉中与亮度、对比度、纹理、颜色和运动相关的感知方面，指出关键视觉参数的物理极限。视觉遮罩是基于感知的压缩的关键组成部分，因此我们最后研究某些时空视觉特征如何被某些类型的内容遮罩。

本章主要回顾了其他人的工作，他们比作者更擅长该主题的生物、神经和心理学方面。有关更多信息，读者可以参考许多关于这个主题的优秀文本，包括斯诺登等人的文本。[\[1\]](#)，Mather [\[2\]](#)，Wandell [\[3\]](#)和Marr[\[4\]](#)。

2.1.人类视觉原理和理论

视觉系统是我们感官中最强大和最复杂的，我们直到19世纪90年代亥姆霍兹的开创性工作开始理解它[\[5\]](#)。

据估计，人脑中40%至50%的神经元与视觉信息处理有关。相比之下，听力约为10%，其他感官约为10%，运动功能约为20%，其他一切为10-20%——比如下棋[\[1\]](#)。这并不完全令人惊讶，因为视觉活动与我们所做的几乎所有事情都相关联。基本上，我们解释颜色，对形状进行分类，检测和评估运动，估计距离，并很好地利用视网膜上的二维图像创建一个三维世界。

那么，我们所看到的和所感知之间的区别是什么？实际上相当多。例如，我们填补视神经穿过眼睛后部的盲点，纠正视网膜上扭曲的不完美图像，并忽略了阻碍我们鼻子和眼睛血管等阻碍的东西。在更高层次上，很明显，视觉系统的存在不是为了忠实地记录图像，而是为了解释世界并在其中生存——识别朋友，评估威胁，导航，觅食和寻找伴侣。人类还利用视觉刺激来娱乐、沟通和提供信息，这可能是在我们20万年的大部分历史中——从洞穴绘画到现代媒体。

典型的视网膜包含约1.3亿个光敏杆和锥细胞。它的神经处理器是视觉皮层的延伸，是从体外可见的大脑中唯一部分。每根视神经从眼睛到大脑传输信号，由大约100万根纤维组成，正是这些纤维将信号从视网膜传输到视觉皮层的其余部分。这与其他感官形成鲜明对比——例如，每根[听觉神经](#)由大约3万根纤维组成。

我们将在第2.2节中检查视网膜和视觉皮层的结构：[人眼](#)，[2.3视觉皮层](#)，但首先让我们简要地看看关于我们视觉系统如何运作的理论。

2.1.1.视觉理论

视觉感知高度专业化和复杂。历史上对视觉系统的生物和解剖学方面的许多研究都基于灵长类大脑，灵长类大脑似乎与HVS非常相似。然而，最近，fMRI等技术彻底改变了神经科学和神经心理学，为人类系统的工作提供了越来越多的洞察力。

直到20世纪70年代末，有两种主要的相互竞争的视觉理论——来自[经验主义者](#)和[理性主义者](#)。他们对视觉系统工作的看法大不相同。经验主义者相信先天、低级、自下而上的处理，对更高层次的功能分配的重要性较小。例如，朱尔斯[\[6\]](#)表明，深度感知可以产生于没有高层次意义的[随机点立体图](#)，因此推断这不能由自上而下的机制驱动。经验主义模型主要基于特征的提取，这有助于

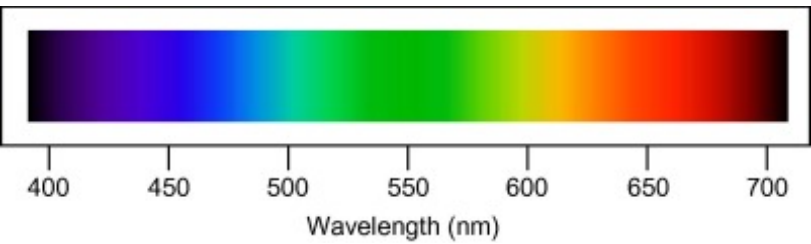
生成连接模式。胡贝尔和威塞尔[7]的开创性工作展示了低级视觉系统的方向和比例选择性，1961年，乌尔和沃斯勒展示了来自低级特征的字符识别。然而，经验主义者没有考虑背景。

相比之下，理性主义者专注于对视觉的高级控制和大脑感知的构建。他们假设用于影响我们寻找和验证较低水平视觉特征的心理模型[8]。该领域的早期工作开发了使用二维结构描述二维属性的图片语法[9]。然而，这些描述了场景语法，而不是场景语义，并且再次没有显示如何表示三维视觉输入。这反过来又导致了一种场景语法方法，试图将三维场景映射到二维描述符。这是通过高层形状（如立方体）或使用低层结构元素（如顶点和边缘）来实现的。这些模型仍然无法描述人类如何产生对陌生物体的感知，因为模型依赖于事先对“语法”的知识来以高水平的方式描述视觉输入。

从理性主义方法中出现了一种混合连接主义方法。直到1982年马尔的贡献[4]，这种方法才正式化。连接主义理论强调处理单元之间的相互联系，既体现了自下而上（数据驱动）的特征提取，又结合基于这种低级信息生成感知的自上而下（解释性）影响。这可以通过注意力等过程来强烈调节。这种连接主义分布式处理类型的方法仍然很有影响力，并大量利用生物证据，与实验数据密切相关。

2.2.收购：人眼

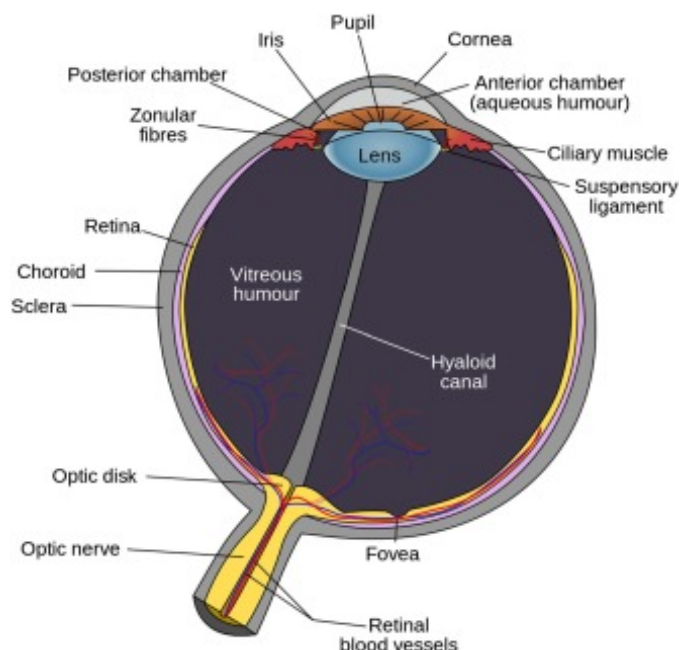
来自太阳或其他光源的光子反射物理物体，其中一些光子进入我们的眼睛。如图2.1所示，我们的视觉系统感知电磁波谱部分的能量辐射范围约为380至720纳米（420至790 THz）。这个范围没有什么特别的，只是我们的生理学已经适应了它，它适应了它，因为它对寻找食物和检测威胁和潜在伴侣等有用。一些动物可以看到其他范围，例如许多鸟类和昆虫在紫外线范围内看到，因为这使它们能够在花卉等中看到某些特征。人眼在光谱的绿色区域的峰值灵敏度约为555纳米——也许这不是巧合，因为由于树叶覆盖，地球表面主要反射绿色带的光从550纳米到560纳米[1]。



[下载：下载全尺寸图像](#)

图2.1。可见光谱。

人眼在广泛的亮度值上大致对数地感知亮度，还可以看到光谱中不存在的颜色。不饱和的颜色，如粉红色，不存在，由多个波长组合形成的品红色等颜色也不存在。横断面人眼的图表如图2.2所示。现在让我们更详细地看看它的主要组成部分。



下载：下载全尺寸图像

图2.2。人眼的横截面。（公共领

域：http://commons.wikimedia.org/wiki/File:Schematic_diagram_of_the_human_eye_en.svg。）

2.2.1。视网膜组织层

眼睛是一个大约球形的体积，充满了液体，由四层组织组成——巩膜、视网膜、脉络膜和睫状体。

巩膜

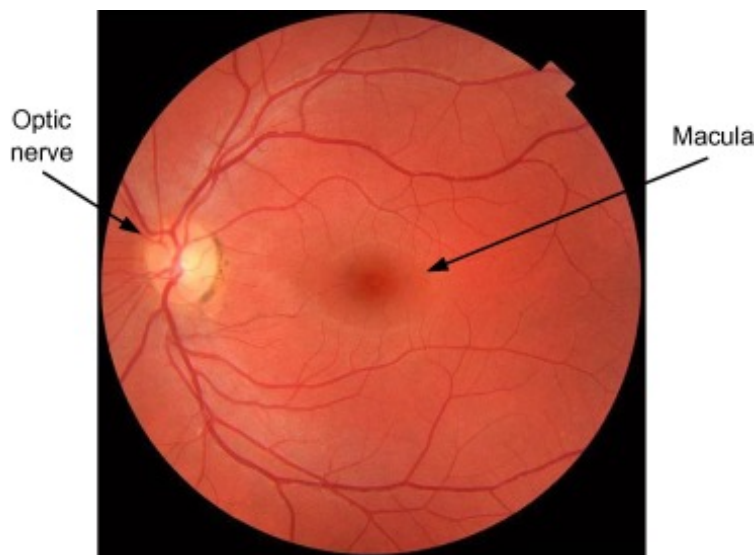
巩膜是提供眼睛结构主要强度的白色组织层。它在眼睛前部变得透明，光线穿过角膜。

睫状体

睫状体是围绕眼睛的组织环。它包含通过改变晶状体形状来调整晶状体屈光力的睫状肌。

视网膜

视网膜包含对光敏感的**光感受器**以及其他几种类型的神经元，这些神经元处理和组合来自光感受器的信号，并通过视神经将这些信号传输到大脑的主要区域。健康人类视网膜的眼底照片如图2.3所示。黄斑是中心较暗的区域，神经纤维穿过眼睛后部的视盘可以看作左侧较亮的区域。有趣的是，尽管我们的视觉系统很大，但它在这个盲点上很好地创造了视觉的印象，尽管那里没有感光器！主要神经通路被视为从视盘辐射的白色条纹图案，血管也可以清晰地看到。



下载：下载全尺寸图像

图2.3。视网膜健康眼底图像。（公共领域来自：http://commons.wikimedia.org/wiki/File:Fundus_photograph_of_normal_right_eye.jpg。）

脉络膜

脉络膜是毛细血管床，为感光器提供营养。它还含有光敏、吸收紫外线的颜料——黑色素。

现在让我们更详细地看看眼睛的前部，光线进入的地方。

2.2.2。光学处理

角膜

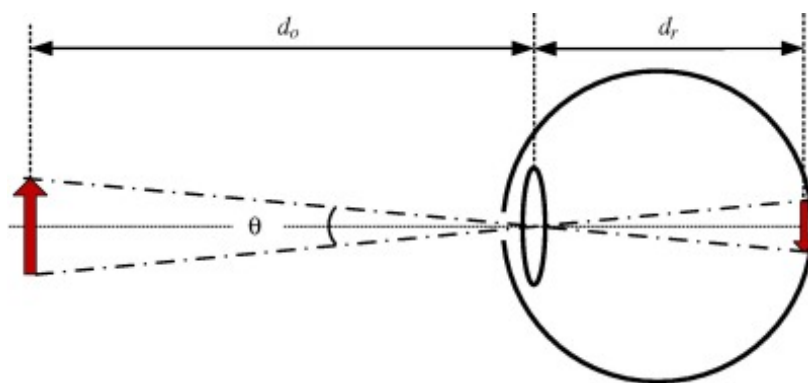
角膜是眼睛前部的敏感透明圆形区域，光线进入。它将光线折射到镜头上，然后聚焦到视网膜上。

镜头

晶状体是位于瞳孔后面的透明结构，与角膜结合，折射入射光将其聚焦到视网膜上。对于遥远的物体，睫状肌会使它变薄，对于近距离物体，它会变胖。有趣的是，晶状体仅占眼睛光学折射功率的25%左右，其余部分由角膜提供。虽然镜头的屈光度比角膜小得多，但它可以改变焦距，使我们能够在不同距离容纳感兴趣的物体。

物体距离的关系 d_o ，视网膜距离 d_r ，焦距 f 如图2.4和方程（2.1）所示。人眼的焦距约为17毫米，光学功率约为60屈光度。

$$\frac{1}{f} = \frac{1}{d_o} + \frac{1}{d_r}$$



[下载：下载全尺寸图像](#)

图2.4。镜头的焦距。

虹膜

虹膜是眼睛前部的单独颜色圆形区域，包含控制瞳孔孔的睫状肌。

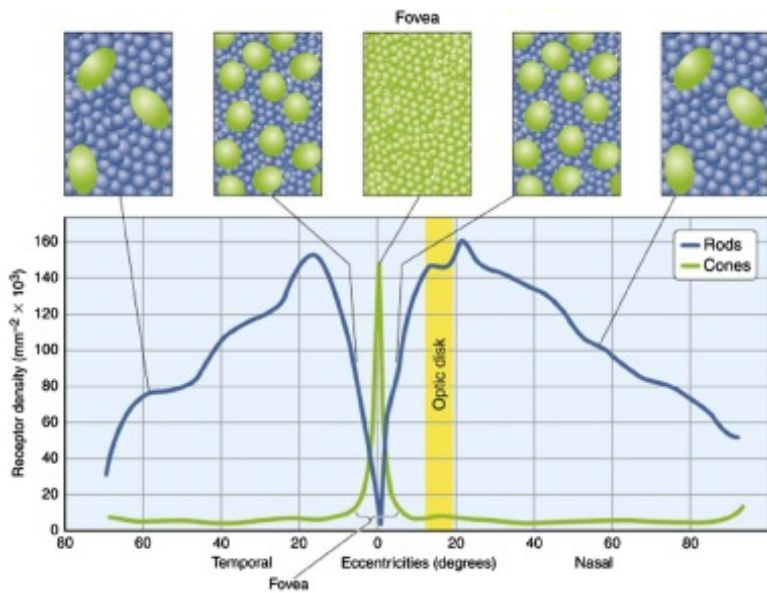
瞳孔

瞳孔是控制光线进入眼睛并在更明亮的环境中变小的孔径。然而，瞳孔只调整了大约16:1倍（从2毫米膨胀到8毫米）。允许我们适应大约8-9个亮度级的不同光级的大部分补偿是由视网膜和大脑其他区域进行的。

现在让我们转到眼睛后部的[传感器](#)——视网膜。

2.2.3。视网膜光感受器及其分布

当光击中视网膜时，它被感光器的反应转化为电脉冲。视网膜中通常有大约1.3亿个传感器细胞——1.25亿根杆子和600万到700万个锥体。这些分布如[图2.5\[12\]](#)所示。虽然锥体的填料在中央凹层密度最大，但外围的每平方毫米仍有数千个。



[下载：下载全尺寸图像](#)

图2.5。视网膜中的光感受器分布。

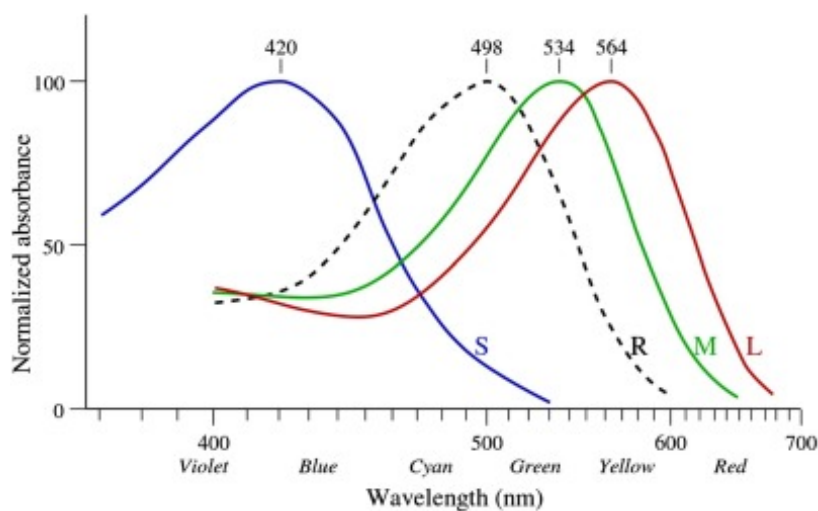
经许可转载自：Musafia等人。[\[12\]](#)。

杆状细胞

视网膜大约有1.25亿根棒。这些负责低光水平（视物视觉）的视觉，并且非常敏感，以至于在中光水平（中光视觉）时饱和或漂白。杆子不介导彩色视觉，提供比凹锥细胞更低的空间敏锐度。光色素红紫素对波长约498纳米的绿光最敏感。中央凹——视觉密度最高的区域——杆密度非常低，这就是为什么在晚上，我们可以通过稍微远离物体来更好地看到物体。

锥细胞

锥体的光照水平高于杆，并提供所谓的光视觉。与杆子相反，它们提供一系列不同的光谱灵敏度特征，并介导彩色视觉。锥体存在三种类型，它们通常称为短波长（S）或蓝色；中波长（M）或绿色；长波长（L）或红色。这些光谱响应具有广泛的重叠，峰值分别为420、534和564纳米左右。蓝色波长的光往往不聚焦，因为它比红光和绿光折射更多。杆和锥的归一化频率响应特性如图2.6所示。



下载：下载全尺寸图像

图2.6。人体视觉系统的归一化杆和锥体反应。

经以下网站许可转载：鲍梅克和达特纳

尔[33].//www.ncbi.nlm.nih.gov/pmc/articles/PMC1279132/。参考维基共享资源。

如图2.5所示，中央凹具有最高的锥密度，因此提供了最高视觉敏锐度的区域。

马库拉

黄斑是靠近视网膜中心周围直径5.5毫米的凹陷区域。正是在这个地区，锥细胞的数量开始急剧增加。它有两层或多层神经节细胞，其中心是凹陷。由于其黄色，黄斑吸收多余的蓝色和紫外线，并作为天然防晒霜。

福韦亚

凹陷是黄斑中心直径约1.5毫米的凹陷，负责中心、高分辨率的视觉。如图2.5所示，它具有很高的空间密度，杆子很少。它还与视网膜中血管空隙的区域重合，随着分散和损失的最小化，这进一步提高了敏锐度。凹陷的中心，即凹陷，直径约0.2毫米，完全由非常紧凑、六边形填充、结构呈棒状的薄锥体组成。凹陷主要包含M和L锥体，其中约5%为S锥体。

视网膜图像的采样分辨率高度不均匀，随着偏心率的增加而迅速脱落。这在外围视觉中创造了一个更模糊的图像，由于快速的眼球运动和更高级的视觉处理，我们无法感知到。在凹陷处的有效角约为 2° ，因此我们真的只“看到”一小部分视觉场景，但我们仍然认为一切都很清晰。凹陷内锥状受体的平均间距为 $2\text{--}3\text{ }\mu\text{m}$ （在外围，这增加到周围 $10\text{ }\mu\text{m}$ ），正是这一点有效地将我们的空间分辨率限制在大约1弧分。当我们评估高分辨率电视格式在观看距离方面的相对优点时，这一点很重要。间距还给了我们感知的最大空间频率。在实践中，眼睛不能解析每度约60个周期的空间频率（cpd）。

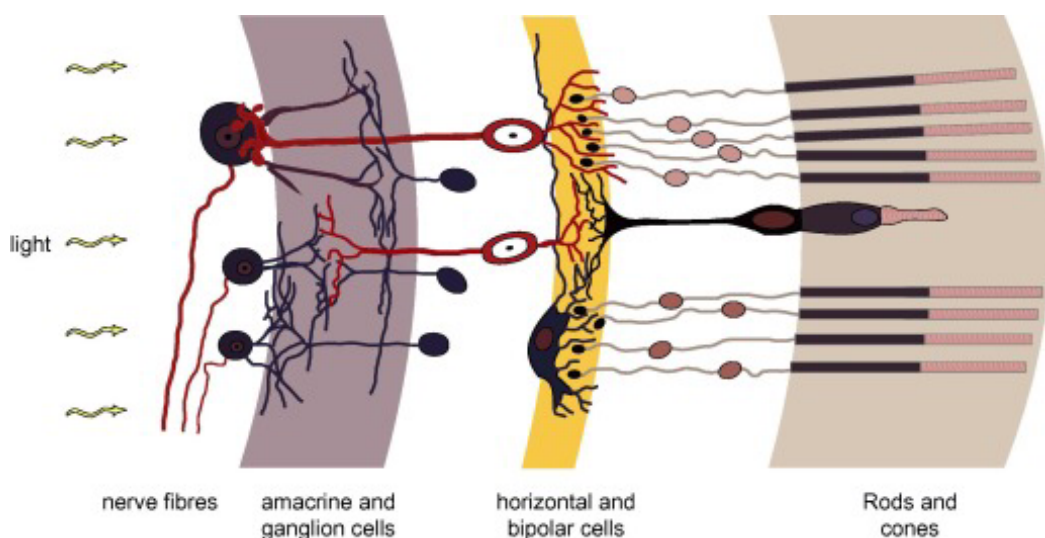
视盘和神经

视盘是视网膜上可见的视神经部分，其中来自视网膜神经节细胞的信号沿着轴突离开视网膜，进入视觉皮层。视神经中大约一半的神经纤维来自中枢，其余的携带来自视网膜其他部分的信息。

视网膜包含许多其他细胞类型，这些细胞类型提供早期视觉处理并减少信息量（HVS压缩！）传输到主视觉皮层。第2.2.4节简要解释了这些单元格。

2.2.4。视网膜中的视觉处理

来自视网膜光感受器的反应由复杂的（远非完全理解）神经网络处理。这个网络包括：双极细胞、水平细胞、无上腺细胞和视网膜神经节细胞。这些共同提供了早期的视觉预处理、特征提取和检测，从而显著减少了通过视神经的信息流。这个结构如图2.7所示。



下载：下载全尺寸图像

图2.7。视网膜细胞架构（公共域图像改编自<http://commons.wikimedia.org/wiki/File:Retina-diagram.svg>）。

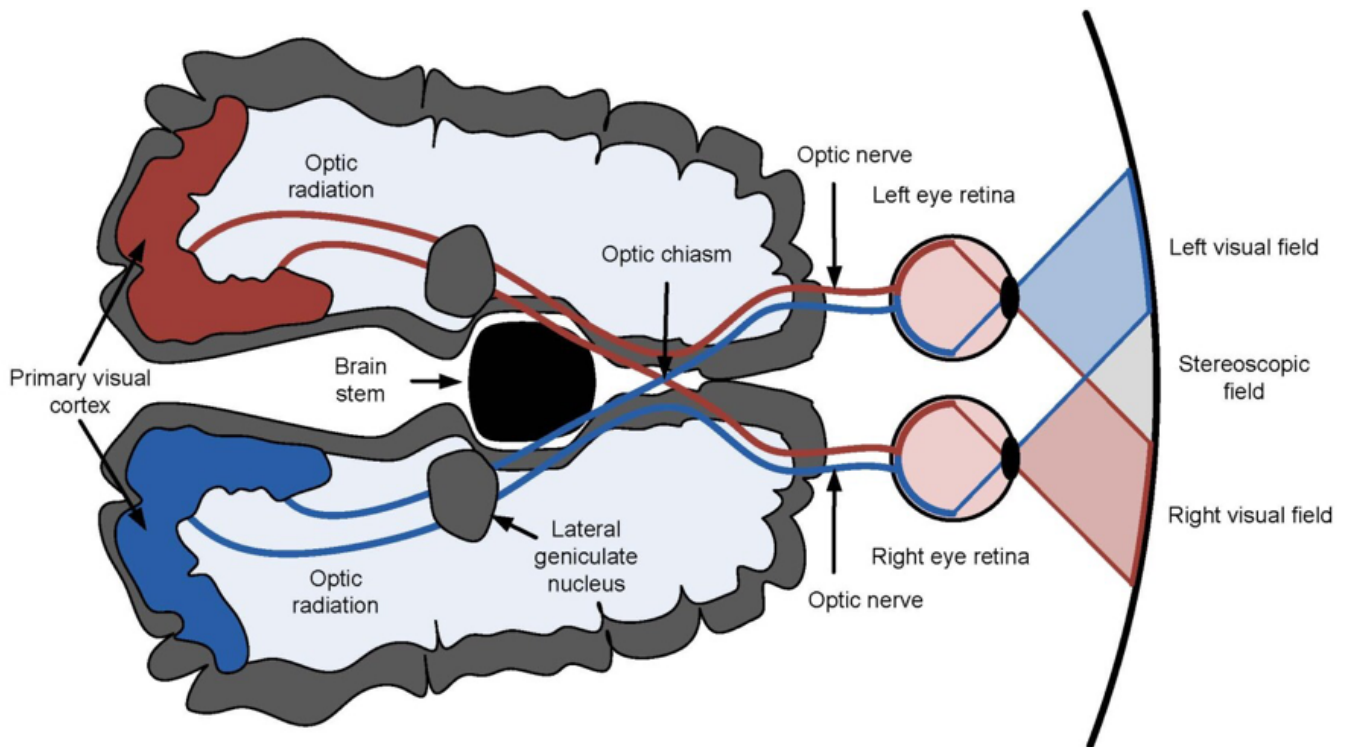
如图2.7所示，双极细胞连接到杆子或锥体，并通过相关感光子的光子增加而激活。水平细胞横向连接视网膜，空间更大，由于受光感受器附近的刺激，可以进行横向抑制。孤立的双极细胞的反应相当粗糙，而水平细胞的影响是添加一个对手信号，其中一个或多个光感受器的反应可以影响周围受体的反应，从而塑造受体场。水平细胞还已知在不同照明条件下调节光感受器信号 [10]。

对无上腺素细胞的作用知之甚少，除了它们可能拥有广泛的树突状树，并以类似于水平细胞的方式为双极细胞和神经节细胞的受体领域提供反馈。

大约有160万个视网膜神经节细胞，这些细胞创造了视觉系统中最早的感受场，为低水平的特征检测和对手的颜色编码提供了基础（见第2.5节）。这些轴突形成视神经，并将电化学信号传输到视觉皮层的更高功能。眼睛中的感光细胞大约是神经节细胞的100倍——这为视网膜压迫提供了证据。

细胞的感受场是刺激必须发生的区域，如果刺激要影响该细胞的反应。感受场是利用空间响应形成的，即来自感受场一部分的反应被感受场的其他部分对立或抑制。双极细胞和水平细胞之间的

相互作用为神经节细胞感受场提供了定向不敏感的中心周围组织，其中中心和周围区域对立。如图2.8所示，这可能包括兴奋中心和抑制性包围，反之亦然。神经节细胞充当变化探测器，仅对其感受场光强（或对比度）的差异做出反应，而不是绝对光水平。因此，它们对边缘很敏感，可以建模为高斯函数（DoG）的差异。



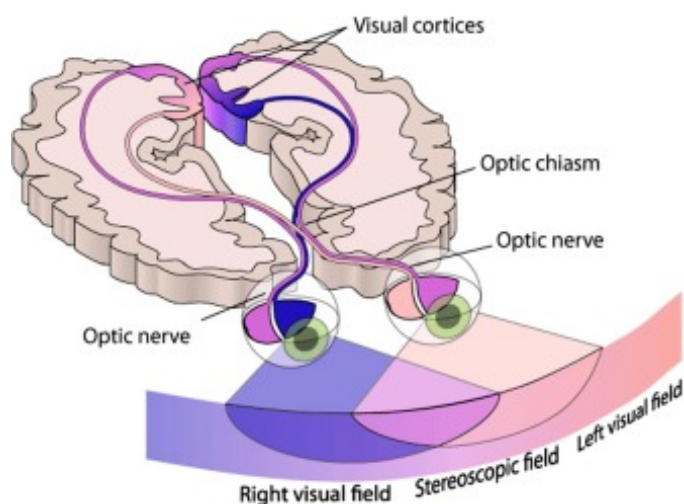
[下载：下载全尺寸图像](#)

图2.8。空间对立性，显示细胞的中心包围及其因激发而产生的发射模式。

因此，随着不同颜色通道的强度变化，场景被有效地编码。然而，很明显，我们看到的不仅仅是边缘，因此这提供了证据，表明更高水平的视觉系统在平均强度水平的接受场响应之间进行填充。这很有趣，因为它可以被视为一种简单的图像合成形式，为感知冗余去除提供了生物学类似物。一个特别引人注目的例子是视觉填充我们的盲点，它不含光感受器。

2.3.视觉皮层

图2.9显示了从视网膜到初级视觉皮层的主要视觉路径的简单图表。每只眼睛视网膜神经节细胞的视神经轴突在视裂处相遇，每只眼睛鼻侧的所有纤维交叉，与另一只眼睛??侧的纤维连接起来，形成两条视道。因此，左视野由右半球处理，反之亦然。视束中的纤维然后在侧膝状核（LGN）突触，在那里它们分布到视觉皮层的其他区域。LGN之后，信号传递到位于头骨后部的初级视觉皮层V1。然后，它们传递到V2，并扩展到至少20个其他中心，每个中心都提供专门功能，如检测运动、识别人脸或解释颜色和形状。



[下载](#): [下载全尺寸图像](#)

图2.9。视觉皮层。

HVS主要通过抑制、激励、有偏见的竞争、输入种族、对手编码和适应等组合过程来提供其功能。下面将更详细地讨论这些。

2.3.1。对立过程

如上所述，如图2.8所示，视觉系统中的许多神经功能似乎使用对手编码对数据进行编码。对立编码是神经元在其单个输出中编码相反或相反特征的过程。例如，运动敏感细胞的基本发射率在检测到一个方向的运动时会降低，并且随着相反方向的运动而增加。对手的抑制和兴奋过程似乎对功能至关重要，抑制反应通常用于控制正常的兴奋状态。

我们还通过对位处理来自锥体和杆子的信号来解释颜色。L、M和S锥光谱响应在波长方面重叠，因此视觉系统更有效地使用它们响应之间的差异，而不是单个响应。这表明有三个对口渠道：红色对绿色、蓝色对黄色、黑色对白色。我们将在第2.5节中进一步考虑这一点。

2.3.2。偏见竞争

有偏见的竞争是指信息在视觉路径和皮层中争夺有限资源的方式。来自视觉系统中特征检测器的输入竞争编码为短期视觉记忆。根据更高级别的反馈（称为关注）来权衡此类投入之间的竞争，以确保与当前任务相关的功能被优先排序。例如，在Stroop任务期间（要求主体声明一系列单词的颜色，而不是阅读单词本身），人们观察到，基于任务的关注会导致大脑相关颜色特征敏感区域的增强，同时抑制来自进行文字处理的区域的反应。

2.3.3。适应过程

适应是当单元的输入保持不变时，其响应随着时间的推移而减少的过程。不同程度的适应在大多数皮质细胞中很常见，当与对手编码过程相结合时，会造成我们经历的与亮度、颜色或运动相关的后遗症。当对手编码特征调谐的细胞适应特定刺激并移除该刺激后，该细胞就会做出反应——短暂地发出它适应的相反特征信号。这种后遗症的例子包括高速旅行后的速度缓慢的感觉，以及

静态物体向与之前看到的移动物体相反方向移动的感觉。戏剧性的亮度和色彩后遗症也经常被用作视觉错觉的基础。即使是长时间观察微笑，也会让毫无表情的脸看起来更悲伤！

2.3.4。V1——初级视觉皮层

主视觉或条纹（分层）皮层（也称为V1）可能是HVS最了解的区域。V1的作用是从场景中提取基本的视觉特征，并已证明V1包含对线条方向、颜色和[空间频率](#)敏感的接收字段。通过将视觉刺激抽象到更高层次的结构中，从而影响视觉数据压缩。大多数V1连接来自视网膜，但众所周知，它也会从更高功能区域接收反馈连接，这会影响其细胞的接收场。

胡贝尔和维塞尔[7]在V1中确定了三个主要探测器：

- **简单单元格**：调谐到边缘的特定方向。
- **复杂细胞**：简单细胞的相位不敏感版本，即无论它位于受体场的哪个位置，对适当定向刺激的反应都是相同的。
- **超复杂细胞**：显示刺激长度敏感性，即细胞反应随着刺激长度的增加而增加。

V1中的神经元通常被分组成对特定低水平特征（如线方向、空间频率或颜色）做出响应的列。众所周知，V1在其组织中保存了视网膜神经节细胞的视网膜位素映射。然而，这不是线性或均匀的映射，因为V1表现出皮层放大，其细胞中不成比例地用于处理凹陷信息。这强调了视觉领域高分辨率中心部分的重要性。

2.3.5。V2——前纹状皮层

V2通过[前馈](#)和反馈连接与V1紧密相连，并与V3、V4和V5紧密相连。尽管V2细胞不如V1了解，但似乎以类似的方式工作，结合了V1根据方向、空间频率和颜色调整的功能，并响应更复杂的模式。V2中的损坏会导致纹理和形状识别不良。还有证据表明，双目差异调节V2细胞。

2.3.6。背流和腹流

从V1到V2区域的连接通常被认为分为两条流——背部和腹部（也称为“哪里”和“什么”或“行动”和“感知”）。背流与运动和深度感知有关，起源于V2，通过大脑向视觉皮层的运动敏感部分。另一方面，腹流与形状和物体识别的感知有关，同样起源于V2，但在这种情况下，通过V4通过心室传递到下??皮层（IT）。

2.3.7。条纹外地区

V1以外的视觉皮层区域通常称为[星条外区域](#)。这些具有更大的接收场，通常计算更复杂和抽象的特征。它们通常也表现出很强的注意力调节。虽然这些领域的功能尚未被很好地理解，但它们的连通性是信息丰富的。

V3与V2连接牢固，V1提供了额外的输入。它同时处理背流和腹流，并已被证明对全局运动做出反应。

V4位于腹流上，已被证明对相对高度复杂的形状有反应。它还表现出很强的注意力调节和大量的色彩敏感性。对V4的损坏会导致形状和纹理识别的损伤。

V5包含对复杂运动模式做出反应的细胞，被认为是运动感知的主要领域。

内侧??上区 (MST) 主要来自内?? (MT) 区域，对运动也很敏感。有证据表明，MST计算光流，特别是在全局膨胀和旋转的情况下。

视觉短期内存 (VSTM) 可以被视为一个短期缓冲区，其中存储优先级数据以供进一步处理。VSTM已被证明在视觉中断（如拍子和眨眼）中持续存在，从而提供了中断后之前查看的场景和新场景之间的联系。

下??皮层 (IT) 通过从HVS的较低级别解析更简单的形状来提供复杂的形状识别。它包含响应特定形状和模式触发的单元格，并托管已知响应面孔的单元格。它融合了来自视觉场两半的信息（到目前为止，视觉场在大脑的半球之间是分开的）。在视觉处理的更高阶段，受体场的连接性变得更大、更复杂。例如，在IT中，一些可接受领域被认为覆盖了整个视觉领域。

视觉皮层的许多其他重要区域已经确定和调查，包括梭形面部区域 (FFA)，主要负责识别面部和其他复杂实体。额眼场 (FEF) 是控制自愿眼动的另一个重要领域。有人提出，这是基于某种形式的视觉显著性地图，该地图具有行为显著特征。

2.4。视野和敏锐度

2.4.1。视野

人眼的视野在??方向（远离鼻子）约为95°，在鼻部方向（朝向鼻子）约为60°，向下为75°，向上为60°，使水平视野超过180°。显然，当包括眼睛旋转时，水平视野甚至进一步增加，即使没有头部运动。

2.4.2。敏锐度

仅仅根据凹陷中锥体的分布来估计视网膜的空间敏锐度并不难。我们在示例2.1中说明了这一点。测量的HVS的奈奎斯特频率通常在40到60cpd之间。这个上值归因于六边形而不是矩形的感光器采样网格模式和其他局部增强技术。HVS通常不会产生混叠效应，这是因为眼睛的光学执行低通滤波，起到抗混叠滤波器的作用。

例2.1 视网膜的视觉敏锐度

假设图2.5所示的感光密度，并且从镜头到视网膜的距离为16.5毫米，估计人类视觉系统在凹中的敏锐度。

解决方案

从图2.5中可以看出，中央凹陷的锥体密度在周围达到峰值150,000 cones/mm²。如果我们假设一个非常简单的包装安排，那么每线性毫米大约有400个锥体。为了保守起，让我们把锥形中心之间的距离定为0.003毫米。眼睛的光学技术[2]规定，对于晶状体到视网膜的距离为16.5毫米，则有0.29毫米/度。因此，锥心之间的角分辨率为 $0.003/0.29 = 0.0103^\circ$ 。现在，由于我们每个周期至少需要两个样本才能满足Nyquist准则，这意味着每个周期有 0.0206° 。因此，最大可检测空间频率的估计为49 cpd。

2.4.3。光、亮度和亮度

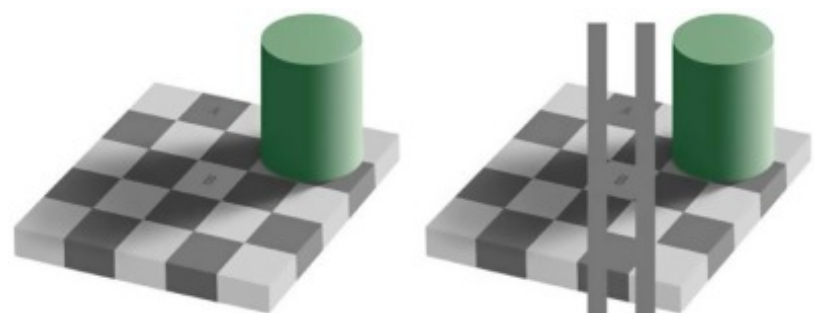
亮度感应不是光度强度（亮度）的**直接类似物**。进入眼睛的光子数量和对亮度的感知仅间接相关。例如，在等光条件下，消色差目标的表现亮度将取决于其周围反射的电磁辐射的频率含量。因此，相对黑暗背景上的灰色目标看起来比较浅色背景上的同一目标更亮。由于这种基于上下文的感知，人们普遍认为视觉系统使用场景中跨对比度边界的亮度比率计算亮度。颜色对亮度的影响也提供了证据，证明对亮度的感知与电磁频率有关[11]。

在图2.10所示的马赫带中，可以看到局部对比度的影响，结合横向抑制的影响。在这里，每个条的宽度的强度是均匀的，但视觉外观是，每个条的右侧都比左侧更暗。另一个引人注目的例子是阿德尔森网格，如图2.11所示，其中正方形A和B由于阴影而看起来非常不同，但实际上是相同的。



下载：下载全尺寸图像

图2.10。马赫乐队效应。



下载：下载全尺寸图像

图2.11。阿德尔森的网格。

经以下网站许可转载：http://web.mit.edu/persci/people/adelson/checkershadow_illusion.html。

与光辐射相关的术语存在重大混淆，因此让我们定义一些与光相关的术语以及我们如何感知光[13]。

辐射强度和辐射

辐射能（焦耳）是从源传播的能量和辐射通量 Φ 是单位时间（焦耳/秒或瓦特）的辐射能量。辐射强度 I 定义为给定方向的**辐射光**功率，以通过立体角（瓦特/立体）的光量来衡量。辐射 L 定义为

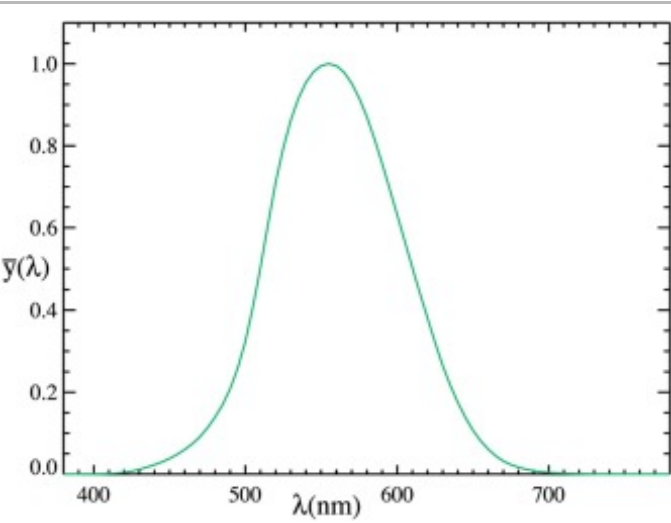
落在单位投影区域上的光强（watts/steridian/m²）。

亮度

上述措施都没有考虑到光的波长或波长的组成。亮度通过根据人类视觉反应加权辐射值来做到这一点。亮度， Y 因此，通常用方程 (2.2) 表示，其中 $L(\lambda)$ 是入射光强度与给定位置和时间波长的函数，以及 $\bar{y}(\lambda)$ 是一种波长相关加权（或相对发光效率），反映了入射通量的感知亮度。亮度以candelas/m² (cd/m²)冒号：

$$Y = K \int_{320}^{720} L(\lambda) \bar{y}(\lambda) d\lambda \tag{2.2}$$

常数 $K = 685$ 流明/流明/流明。功能 $\bar{y}(\lambda)$ 会因人而异，但如图2.12所示，CIE¹在1929年为标准观察者测量了这一点。如果在方程积分 (2.2) 中，则项 $L(\lambda)$ 被辐射通量取代 $\Phi(\lambda)$ 然后结果是称为**发光通量**，并具有腔的单位。



下载：下载全尺寸图像

图2.12。CIE**发光效率**曲线。（公有领域图像：http://en.wikipedia.org/wiki/File:CIE_1931_Luminosity.png。）

亮度

亮度是由视觉目标的亮度引发的感知，因此代表光生理感觉的非定量指示。它通常由观察者根据调整为与被观察的目标具有相同亮度的白色区域进行主观评估。

卢马

图像传感器通常产生与辐射成正比的输出值，入射光通常滤波段为R、G波段和B波段。然而，在这些从相机中出现之前，它们通常根据HVS的响应进行非线性映射，使用一种称为**伽马校正**的技术。这些输出通常标记为 R' ， G' ，以及 B' 。Luma是用于表示这些校正组件的组合，根据所用颜色空间的规格，产生伽马校正亮度信号。伽马校正**在第4章**中进一步考虑。

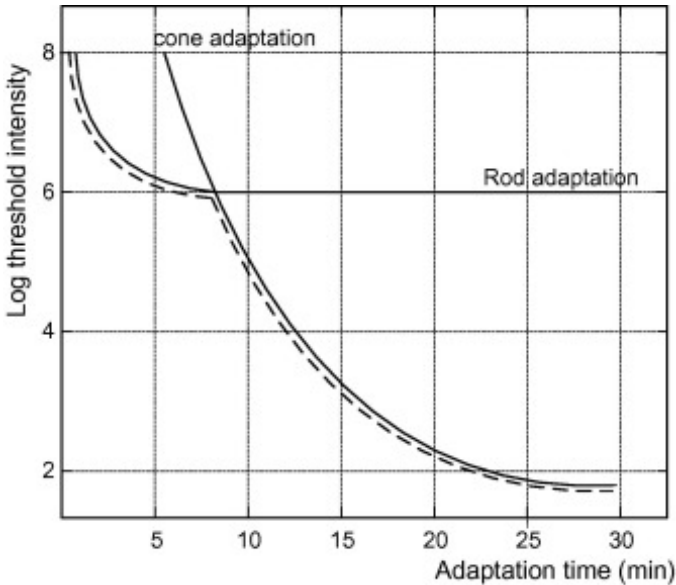
2.4.4。光级适应

我们适应各种光线水平的能力对功能和生存至关重要。随着环境光强的变化，感光器，实际上是整个视觉系统，必须保持敏感。这种变化非常戏剧性，正如我们能够在星光下看到的（ 10^{-3} cd/m^2 ），通过室内照明（ 10^2 cd/m^2 ）到明亮的阳光（ 10^5 cd/m^2 ）。因此，眼睛可以在亮度水平上（至少在某种程度上）在大约八个数量级上发挥作用——比例为100,000,000:1。

当然，眼睛无法立即应对这个动态范围。如果我们考虑我们眼睛的瞬时动态范围（瞳孔开口是固定的），那么大约是10个f-stops。如果我们把快速的学生适应情况考虑在内，这增加到大约14个f-stops。瞳孔直径通常在明亮条件下约为2毫米，在黑暗条件下约为8毫米，给出16:1的动态范围或4 f-stop。

杆和锥体也进行了适应。锥体比杆子适应得更快，这解释了为什么适应黑暗（黑暗适应）比适应亮度要长得多。

典型的黑暗适应特征如图2.13所示。这说明了个人在适应强光后阈值随时间的变化。这个图说明暗影适应过程相当缓慢，需要几分钟才能完全适应，但也说明适应分两个阶段进行。第一个是由于锥体适应，第二个是由于杆子。视网膜光感受器通过色素漂白进行适应，但也通过水平细胞的反馈进行适应。



[下载：下载全尺寸图像](#)

图2.13。杆和锥体的黑暗适应性。

2.5.颜色处理

如果你问某人是否更喜欢看彩色电视节目或电影，而不是单色电视节目或电影，他们总是会以更接近现实为由说是。这是显而易见的——想想图2.14，图右侧显示了澳大利亚卡卡杜的黄水彩色图像，左侧显示了同一图像的单色版本。大多数人都会同意，彩色图像更容易解释，改善了自然物体和场景特征之间的区分，并具有更好的深度和参与感。



[下载：下载全尺寸图像](#)

图2.14。增加彩色图像的浸入感。

然而，尽管很少有人看到黑白分明的世界，但我们欣然接受**单色图像**，并完全乐意为了娱乐或信息而暂停怀疑。在某些情况下（例如史蒂文·斯皮尔伯格的**辛德勒名单**），导演实际上选择使用单色图像来影响情绪和增加电影中的影响。这种对亮度信号的接受和偶尔偏好可能与我们处理三色传感器的方式有关。

这里值得强调的最后一点是，显示器中主要红色、绿色和蓝色的选择不是基于S、M和L锥体的敏感性。相反，它们被选择为光谱间隔良好的颜色，从而允许产生大色域。

根据第2.4节和方程 (2.2)，我们可以生成以下对入射辐射的平均锥体响应 R 、 G 、 B ，其中 $\bar{y}_r(\lambda)$ 、 $\bar{y}_g(\lambda)$ ，以及 $\bar{y}_b(\lambda)$ 红色、绿色和蓝色通道分别是**发光效率**：

$$\begin{aligned} R &= K_r \int_{380}^{720} L(\lambda) \bar{y}_r(\lambda) d\lambda \\ G &= K_g \int_{380}^{720} L(\lambda) \bar{y}_g(\lambda) d\lambda \\ B &= K_b \int_{380}^{720} L(\lambda) \bar{y}_b(\lambda) d\lambda \end{aligned} \quad (2.3)$$

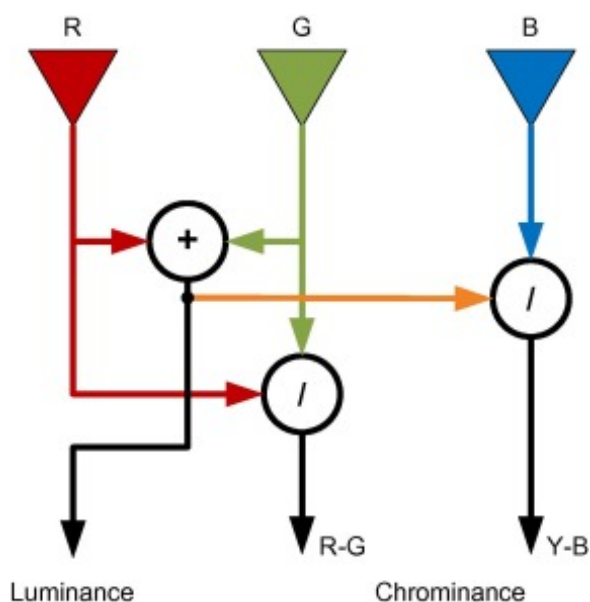
2.5.1. 颜色的对立理论

我们三个锥形。然而，大多数哺乳动物只有两种类型——从一个以550纳米为中心的**黄绿色锥体**进化而来（相当于叶绿素反射率最高值）。然后添加了一个以430纳米为中心的**蓝色锥体**，这有助于植物识别。人类和其他灵长类动物在光谱的红端添加了第三种**锥体类型**（实际上黄绿色锥体可能分为两种类型）。这使其所有者能够更有效地觅食（红色水果更容易区分）和信号更显著。作为一个有趣的地方，一些鸟拥有四条通道，螳螂虾有12条通道！

如前所述，HVS中的色彩感知采用基于对手的方法。如果我们考虑颜色恒定性要求，原因就变得很清楚了。显然，如果来自特定物体反射的光发生变化，我们理想地希望将其视为同一颜色的更

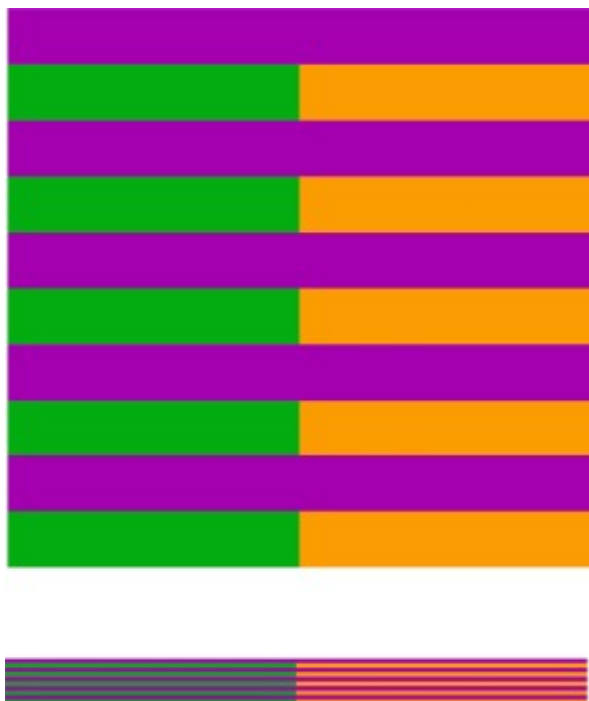
亮或更暗的版本。如果我们考虑**光感受器**的输出比例，而不是它们的绝对输出，这一点是可以实现的。赫林在1878年首次发现了这一点，他意识到某些色调不可能在单一的颜色感觉中共存。例如，我们可以体验红黄=橙和蓝绿=青色，但不能轻易体验红绿或蓝黄。因此，**红绿蓝黄**的对被称为**对手对**。鲑鱼认为，我们不能同时体验两种颜色，因为它们编码在同一路径中。这一观察已经从心理物理学实验中得到证实，很容易说服自己。例如，如果你盯着绿色看足够长，你会得到红色的后遗症（反之亦然），如果你盯着蓝色足够长，你会得到黄色后遗症（反之亦然）。

图2.15给出了一个显示可能对手模型的**图表**。这显示了由绿色和红色锥体产生的信号之和产生的亮度信号，以及由红色与绿色和黄色到蓝色的比例产生的两个**色度**通道。这里应该注意三件事。首先，专家对组合过程的性质有不同的意见，有一些证据表明色度通道是基于颜色差异（即R-G和Y-B）。然而，由于锥**光谱响应**的灵敏度重叠，大多数波长将至少刺激两种锥体。因此，一些三刺激值不能作为**加性颜色**空间存在，并暗示三个原色中至少有一个是负值。其次，众所周知，颜色感觉会随着上下文和**空间频率**而有很大差异（见图2.16，自己试试这个——如果你把页面移近，蓝色条会变成绿色）。沙普利和霍肯等人提出了解释这方面面的理论。最后，观察的读者会注意到，**亮度通道**中没有蓝色锥体的贡献。造成这种情况的主要原因是，中央凹陷的蓝色锥体相对较少，中心几乎没有蓝色锥体。这可能是因为**色差**的影响，**色差**会显著降低敏锐度。



[下载：下载全尺寸图像](#)

图2.15。颜色的对立处理。

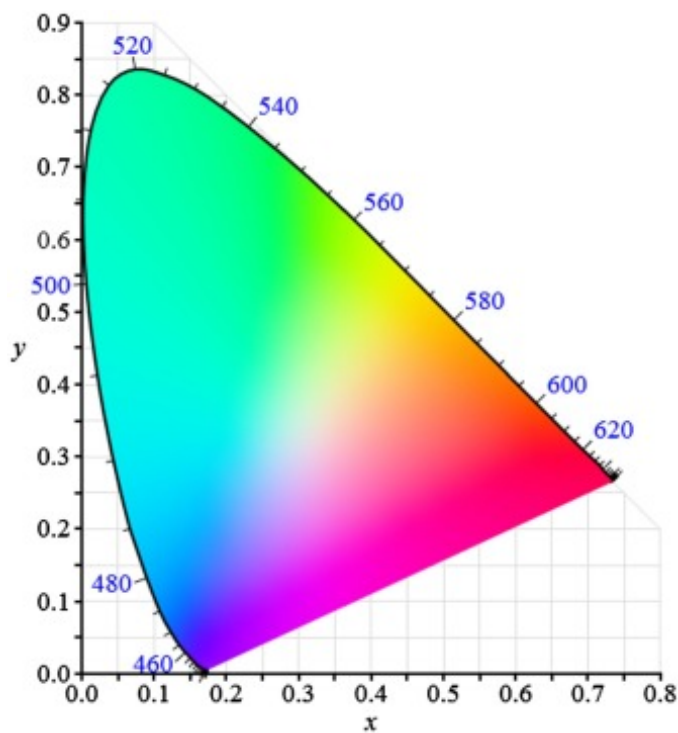


[下载：下载全尺寸图像](#)

图2.16。颜色依赖于上下文。底部图片只是顶部的压扁版本，但绿色条纹变成了蓝色。（为了解释这个人物传说中对颜色的引用，读者请参阅本书的网络版。）

2.5.2。CIE 1931 色度图

国际照明委员会CIE于1931年定义了一个基于标准观察者的颜色映射函数，该函数表示人类在 2° 弧内对原始物体的平均色度反应 $R = 435.8 \text{ nm}$, $G = 546.1 \text{ nm}$ ，以及 $B = 700 \text{ nm}$ 。图2.17显示了CIE 1931色度图。边界表示光谱颜色的最大饱和度，图表构成所有可感知色调的边界。通过任何三种原色（如RGB）的组合可以创建的颜色可以在色度图上通过连接这三种颜色坐标的三角形表示。第4章更详细地讨论了颜色空间。



下载: 下载全尺寸图像

图2.17。CIE 1931 色度图。

经裁判许可转载。[21]。

2.6.空间处理

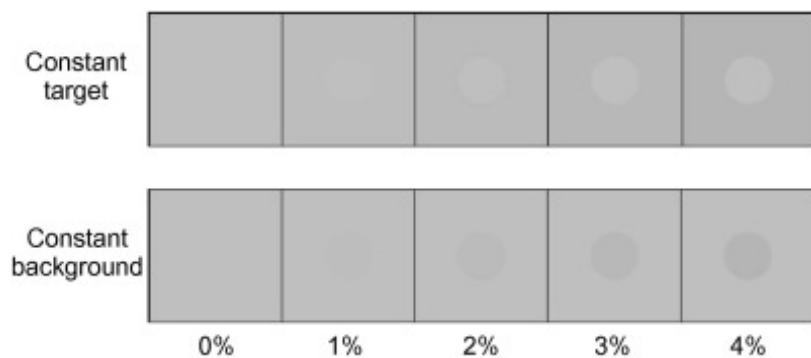
空间分辨率很重要，因为它会影响我们对物体的观察程度。如第2.4节所述，关键参数不是显示器每行或每列的像素数，而是角度的减小， θ ，通过观众视网膜上的每个像素。因此，我们在这里使用“空间”一词来表示不移动的东西，而不是暗示无论观看距离如何，都有一定的空间恒定性感。

2.6.1。只是明显的差异、对比度和韦伯定律

对比度是使对象与其他对象和背景区分开来的视觉属性。它与同一视野内物体和其他物体的颜色和亮度差异有关。由于HVS比绝对亮度对对比更敏感，因此无论照明的变化，我们都会以类似的方式感知世界。众所周知，对照明持续增量的感知并不统一，而是随照明水平而变化。这被韦伯定律所捕获，该定律确定，在我们可见范围内的合理范围内，以下表达式用于对比， C ，成立，对于一个目标 T 和背景 B 冒号：

$$\Delta C = \frac{\Delta Y}{Y} = \frac{\text{JND}(Y_T - Y_B)}{Y_B} = \text{constant} \quad (2.4)$$

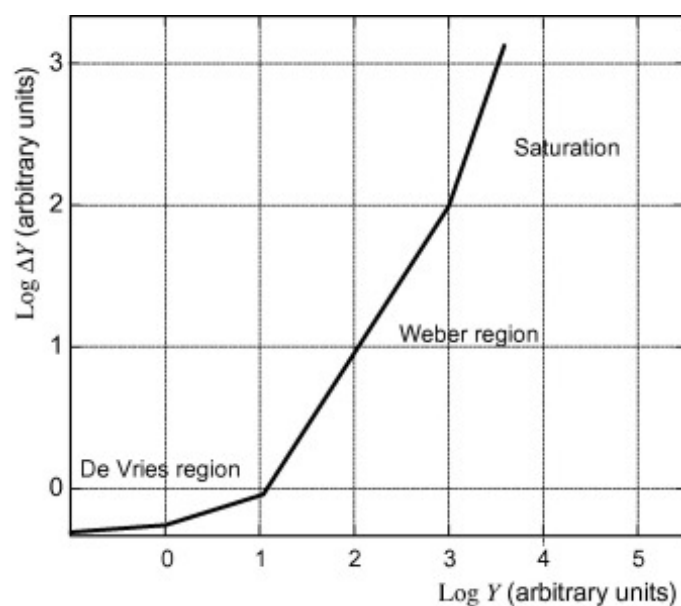
韦伯定律意味着两种刺激之间的显著差异（JND）与刺激的大小成正比。这有时与费希纳定律相混淆，费希纳定律指出主观感觉与刺激大小的对数成正比。图2.18显示了JND测试的示例测试图表。对大多数观察者来说，JND值在0.01到0.03之间。



[下载: 下载全尺寸图像](#)

图2.18。不同对比度增量的显著差异。

韦伯定律非常适合230多年的历史 $\log(Y)$ 如图2.19所示。当本底照明高于或低于该区域时，斜率将发生变化。正如我们将在后面的章节中所看到的，这种对亮度灵敏度的主观评估在图像和视频量化中可能很重要。



[下载: 下载全尺寸图像](#)

图2.19。人类视觉的JND曲线。

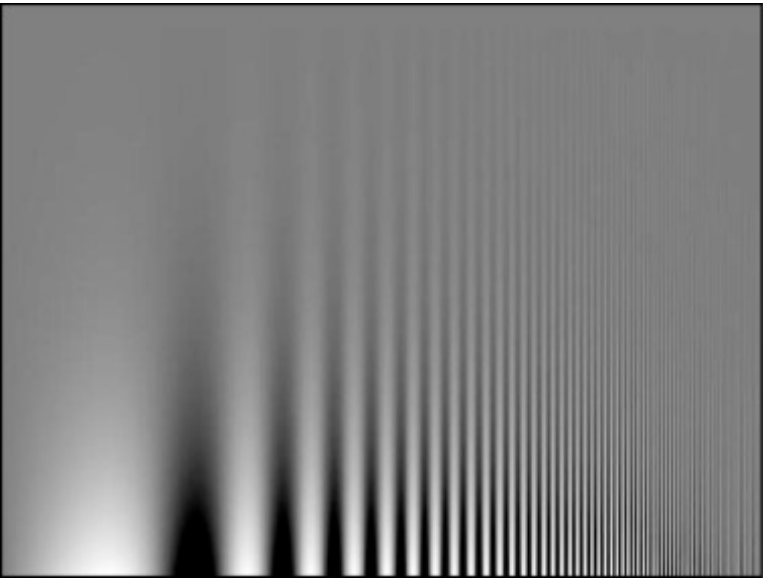
2.6.2.频率相关对比度敏感度

HVS模型最重要的问题之一是对比度敏感度和空间频率之间的关系。这种现象由对比度灵敏度函数（CSF）描述。CSF模型描述了HVS识别亮度和色度差异的能力，这是对比度和空间频率的函数。

空间对比度灵敏度图或正弦波光栅如图2.20所示。亮度和色度的典型对比度敏感反应如图2.21所示，其中对比度敏感性定义为：

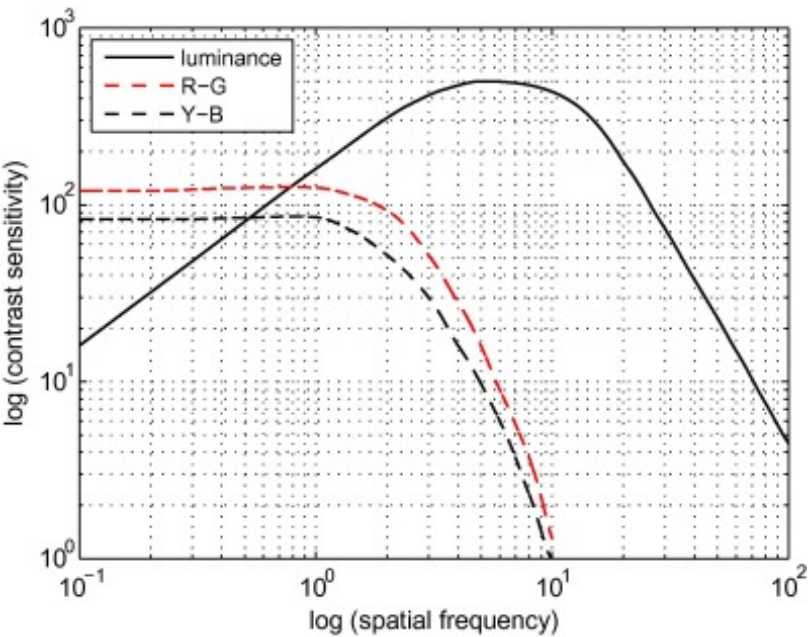
$$C(f) = \frac{Y_B(f)}{JND(Y_T(f)-Y_B(f))} \tag{2.5}$$

这说明了视觉系统在不同频率和对比度水平上对空间正弦模式的敏感性。带通特性主要是由于眼睛的光转移功能以及视网膜（侧向抑制）处理和V1区的一些特性。其形状不仅受到感光器间距和排列的影响，还可能受到V1中可表示的空间模式数量的限制。



[下载：下载全尺寸图像](#)

图2.20。对比灵敏度图表。



[下载：下载全尺寸图像](#)

图2.21。亮度和色度脑脊液反应。

本质上，HVS对较低的空间频率更敏感，对高空间频率更不敏感。曼诺斯和萨克里森[22]将CSF模型的灰度图像作为非线性变换，然后是调制传递函数（方程 (2.6) ）：

$$C(f)=2.6(0.0192+0.114f)e^{-(0.114f)^{1.1}} \tag{2.6}$$

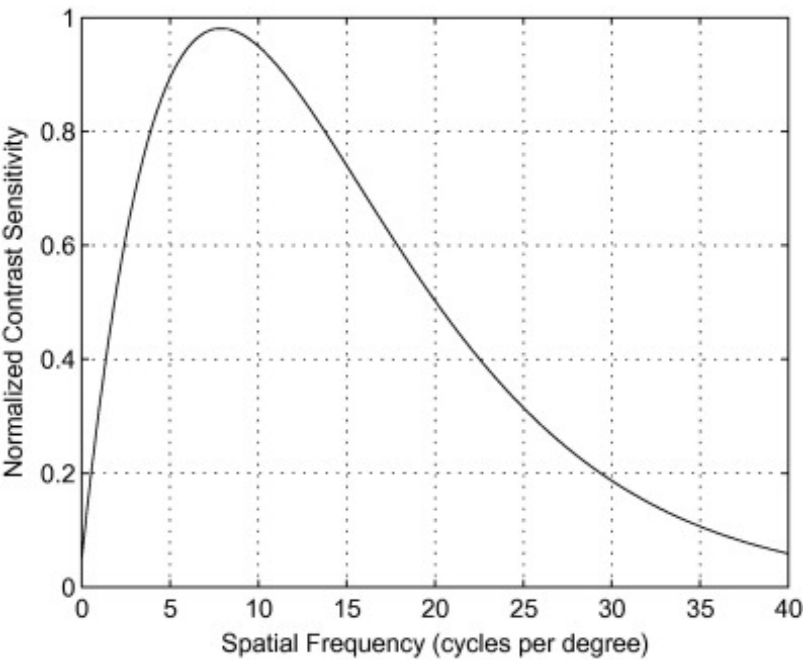
其中空间频率为：

$$f=\sqrt{f_h^2+f_v^2}$$

f_h 是水平频率分量和 f_v 是垂直频率组件。通常最好以每个像素的周期为单位，以规范化的空间频率来工作，以显示特性为单位，因此：

$$f_d=\frac{f}{f_e}$$

在哪里 f_e 是正在进行的实验的每个度像素数，这取决于显示器的分辨率、显示器的尺寸和观看距离。这与测量结果非常吻合，其CSF响应函数（归一化）如图2.22所示。



[下载：下载全尺寸图像](#)

图2.22。亮度对比灵敏度功能。

这里根据空间频率绘制了人类视觉的对比灵敏度。我们可以观察到：

1. 对亮度信息的敏感性在大约5-8个周期/度上达到峰值。这相当于一个条纹宽度为1.8毫米的对比网格，距离为1米。

2. 亮度灵敏度从峰值的两侧下降，灵敏度在50个周期/度以上几乎没有。
3. 色度敏感曲线的峰值发生在比亮度较低的空间频率，响应迅速下降超过大约2个周期/度。还应当指出，我们对亮度信息的敏感性约为R-G的三倍，而R-G的灵敏度约为B-Y的两倍。

HVS的对比敏感性引导我们获得压缩图像的第一个基本手段——感知的CSF模型可用于减少不可感知的信息，如果我们使用亮度和色差信号作为表示的基础，那么我们可以以亮度信号的一半左右的速度采样色度信号，而不会失去感知质量。此外，由于对比灵敏度降低，亮度和色度信号都可以在更高的空间频率上更粗量化。[第四章](#)将探索实现这一点的这些机制。

2.6.3. 多尺度边缘

早期视觉的边缘[定位](#)模型通常具有在多个图像尺度上进行一些调谐[空间过滤器](#)的排列，然后是特征提取过程，并最终将这些特征集成到地图中（例如马尔的原始草图[\[4\]](#)）。视觉系统的方向和[频率选择性](#)在许多方面类似于多尺度变换，如小波或[复小波](#)变换，这些都用于各种HVS模型。正如我们将在本书的后面看到的那样，这些工具也广泛用于图像和视频压缩，其中可以利用频率、方向和比例转换的图像数据来删除隐形内容，例如，通过具有CSF特性的过滤。

2.6.4. 对纹理的感知

纹理是我们在边缘定位、深度感知和一般对象识别等任务中利用的重要视觉线索。视觉系统似乎在场景中高熵的假设下运行。例如，鹅卵石在海滩上的分布提供了良好的深度线索，因为它们似乎离观众越远就越小。有人提出，纹理感知采用滤波器-整流器（FRF）过程，该过程首先对初始多尺度空间滤波器的输出应用非线性变换，然后是第二个滤波阶段，以提供一个低频表面进行边缘定位。该模型意味着，与亮度梯度相关的边缘相比，人类在确定纹理定义的边缘时应该不那么准确，这一点已在实验中得到证实。

这也表明，HVS将经历与纹理相关的一定程度的变化盲，这再次得到了主观试验的证实。考虑[图2.23](#)中的两张图像。它们看起来一模一样，但仔细检查后，它们实际上明显不同。这一观察再次强化了纹理分析-合成压缩方法可能提供潜力的概念。



[下载：下载全尺寸图像](#)

图2.23。纹理改变盲。（图片由汤姆·特罗相亚科提供。）

2.6.5。形状和物体识别

病变研究有证据表明，视觉处理可以分为三大阶段：

1. 提取本地特征。
2. 从这些特征中构建形状和表面。
3. 创建场景中对象的表示形式。

还有一些证据表明，HVS使用场景中的其他属性来帮助分组特征，例如接近、颜色和大小相似性、共同命运（具有相似方向和速度属性的区域可能是同一对象的一部分）和连续性。Marr[4]提出，特征按平均局部强度、平均大小、局部密度、局部方向、相邻对相似物之间的局部距离以及连接相邻物对的线的局部方向进行分组。

为了实际识别物体，必须在更大的空间距离上进一步整合特征。对此存在一些解释，包括来自视觉皮层更高功能区域的反馈，这些功能区域调节边缘检测器响应，以提高轮廓[2]。表面也可能是重要的分组线索，因为同一表面的特征可能是同一物体的一部分。

因此，对象表示可能由低级图像分析和通过层和区域之间的显著反馈耦合的更具符号性的高级抽象表示组成。对象识别之所以复杂，是因为对象既有内在属性又有外在属性。外部因素，如观点变化、遮挡或照明变化，可能会对可识别性产生重大影响。这些外部属性可以提供上下文，但也可能混淆识别。

2.6.6。阶段信息的重要性

据广泛报道，相位光谱在确定图像的视觉外观方面非常重要；相位信息的失真，特别是在较大区域，可能会导致图像质量不佳。例如，见图2.24。同样，已经证明，从局部星等信息中间接获得相位是可能的。有证据表明，人类视觉皮层V1中的复杂细胞对局部大小信息而不是相位反应。因此，有人建议，HVS可能会使用局部星等信息和局部相位来确定图像的外观。Vilankar等人[14]进行了一项实验，使用使用双树复合小波变换失真的图像，量化局部大小和局部相位对图像外观的贡献，作为空间频率的函数。他们证实，局部震级和局部相确实发挥着同样重要的作用，在某些情况下，局部相位可以主导图像的外观。虽然我们仍处于理解这些过程的早期阶段，但毫无疑问，它们将来可能会更好地理解图像压缩对**人类视觉系统**的影响。



[下载：下载全尺寸图像](#)

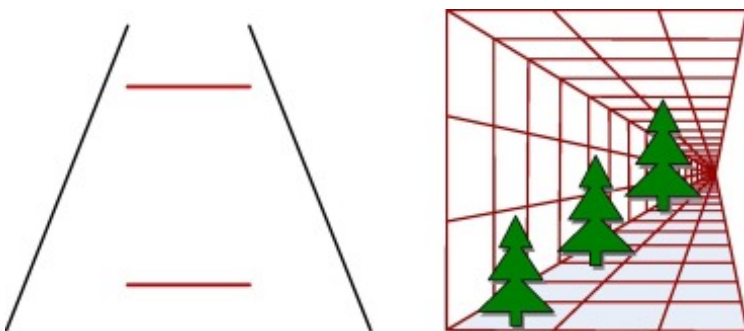
图2.24。相位信息在视觉感知中的重要性。左：原创。右：使用[复杂小波](#)变换的相位失真版本（经Vilankar等人许可转载）。[14]。

基于上述原因，人们普遍认为，图像上的信号处理操作应保留相位信息。因此，[线性相位FIR数字滤波器](#)（几乎）仅用于压缩操作，包括预处理和后处理以及插值。

2.7。对规模和深度的感知

2.7.1。尺寸或比例

大小和规模是视觉中的复杂主题。当以越来越远的距离观察空间模式时，它投射在视网膜上的图像会离视网膜更远；它投射的图像会刺激反映更高[空间频率](#)的反应。因此，在这个层面上，规模的变化与空间频率的变化有关。已经证明，在一次距离适应某个频率确实会影响在不同距离上对不同频率的感知，即使两者在视网膜上投影相同的空间频率。因此，即使距离没有集成到来自场景的频率信息中，深度信息也必须在更高的级别上集成，以补偿并允许人类在不同距离评估物体的大小。这种判断物体大小的能力，无论距离如何，称为大小恒定，并可能导致误导和娱乐结果，特别是当与其他深度线索（如透视）相结合时（[图2.25](#)）。



[下载：下载全尺寸图像](#)

图2.25。基于透视的深度线索可能非常引人注目和误导人。

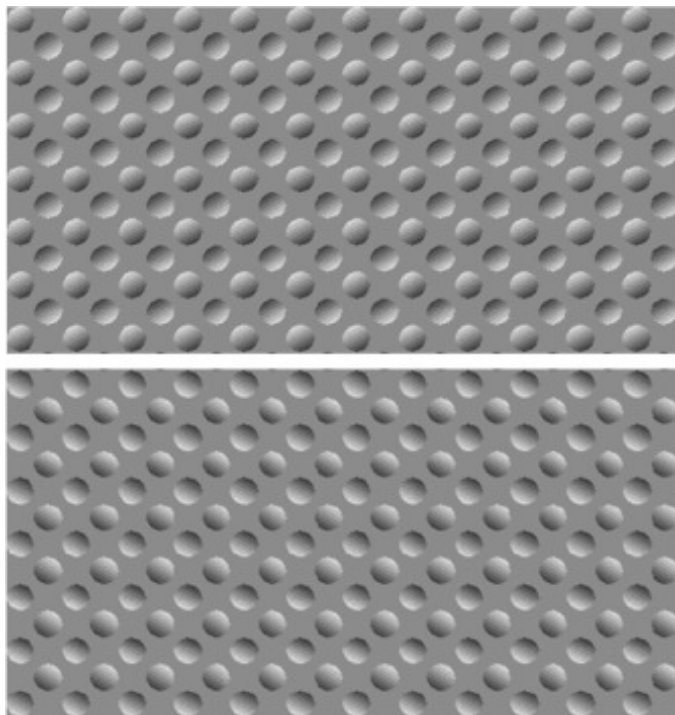
2.7.2。深度线索

毫无疑问，深度评估是我们解释场景能力的主导因素；它还有助于增强我们对显示图像和视频内容的参与感。通过双目人类视觉创造的立体视觉通常被认为是主要的深度线索，近年来在创造新的娱乐形式方面确实得到了不同程度的成功。然而，这只是人类使用的众多深度线索之一——可以说它不是最强的。[人类视觉系统](#)中使用的深度线索列表如下：

- **我们的三维世界模型：**自上而下地熟悉我们的环境使我们能够对深度做出相对判断。
- **运动视差：**当观察者横向移动时，较近的物体比较远的物体移动得更快。
- **运动：**刚性物体在离开或向观察者移动时会改变大小。
- **透视：**平行线将在无穷远处收敛——使我们能够评估定向平面的相对深度。
- **遮挡：**如果一个对象部分阻止另一个对象的视图，则显示更近。
- **立体化：**人类有两只眼睛和双目差异信息，这些信息来自物体在每个视网膜上的不同投影，使我们能够判断深度。
- **照明、阴影和阴影：**光线落在物体或场景上的方式告诉我们很多关于深度和方向的信息。见[图2.26](#)。
- **海拔：**我们感知到离地平线更近的物体。
- **纹理梯度：**当物体向后退到远处时，一致的纹理细节将显得更精细，并最终因视觉敏锐度的限制而消失。
- **适应：**当我们专注于物体时，我们的睫状肌要么伸展（远方物体的薄透镜）或放松（更近的透镜）。这为深度感知提供了眼球运动线索。
- **收敛：**对于较近的物体，我们的眼睛会随着它们聚焦而收敛。这拉伸了眼外肌肉，产生了深度感知。

应当指出，某些深度线索可能会令人困惑，并可能导致有趣的错觉。例如，我们有一个内置模型，它告诉我们光来自上方（即太阳在地平线上），我们对阴影的期望反映了这种自上而下的知识。[图2.26](#)正好显示了这一点。顶部图表清楚地显示了交替的凸起和凹坑，从顶部的一排凸起。底部图类似，只是从顶部的一排坑开始。实际上，与这些图表的唯一区别是底部是顶部翻转180°。通过将页面倒过来试试这个。这提供了一个绝佳的例子，说明某些视觉线索有多硬，以及这些线索如何导致有趣的错觉。由自上而下过程驱动的错觉的另一个例子是空心面具。在[图2.27](#)中，我们可以看到阿尔伯特·爱因斯坦的正常照片。实际上，这是一张凹面面具的照片。我们的视觉系统非常适应脸部，即使我们的差异线索与之冲突，我们也忍不住认为这是一张[凸脸](#)。右图显示，即使我们把遮罩旋转到特征扭曲的角度，并且你可以看到它明显凹陷，它仍然看起来像一

个凸脸！



[下载：下载全尺寸图像](#)

图2.26。坑和凹凸起——来自照明的欺骗性深度。



[下载：下载全尺寸图像](#)

图2.27。空心面具幻觉。

2.7.3. 深度线索和三维娱乐

三维视频技术利用了通过双目视觉产生的立体视觉。立体化是使用两个在互显距离间隔的相机产生的，以创建两个视图，这些视图可以以这种方式显示，左视图仅输入左眼，右视图输入右眼。虽然这确实会给人一种深度感，如果对某些类型的内容使用得当，效果会有效，但有几个问题会造成问题。这些与适应和目光的融合之间的冲突有关。

当我们看三维电影时，我们会在二维屏幕上观看，这样我们自然地关注它。然而，从屏幕上突出的三维物体会导致我们的眼睛收敛，因为它们看起来更近。这可能会导致令人困惑的深度线索，导致疲劳，在某些情况下，恶心。此外，据报道，多达20%的人口不看立体。其余部分，很大一部分人不喜欢看三维电影，而是选择二维看电影。对此给出了一系列原因，包括戴眼镜、当前三维眼镜导致的光照水平下降（高达 50%）或强烈、强调过重或不一致的深度线索的一般感觉。

同样，导演、摄影师和技术人员才逐渐开始理解他们领域之间的复杂互动。例如，导演和摄影师使用镜头来支持叙事，电影中的平均镜头长度惊人地短（约4s）。现在人们知道，HVS适应三维镜头之间的过渡需要比二维镜头更长的时间，因此三维情况下的镜头长度应该相应地更长（稍后会更多）。为了应对三维，[摄影](#)的规则不得不改变。

2.8. 时空反应

HVS的时间响应取决于许多内部和外部因素。本质上，[感光器](#)的脉冲响应定义了[视网膜图像](#)的最大时间采样。然而，环境照明、显示亮度和观看距离等外部影响也会导致变化。如果一个物体正在穿过视野而没有被跟踪，其图像将模糊（运动模糊），因为感光器需要大约100毫秒才能达到峰值响应。这些观察不仅对相机和显示器设计，而且对[摄影](#)有重大影响（例如，通常规定限制相机速度）[16]，[17]。

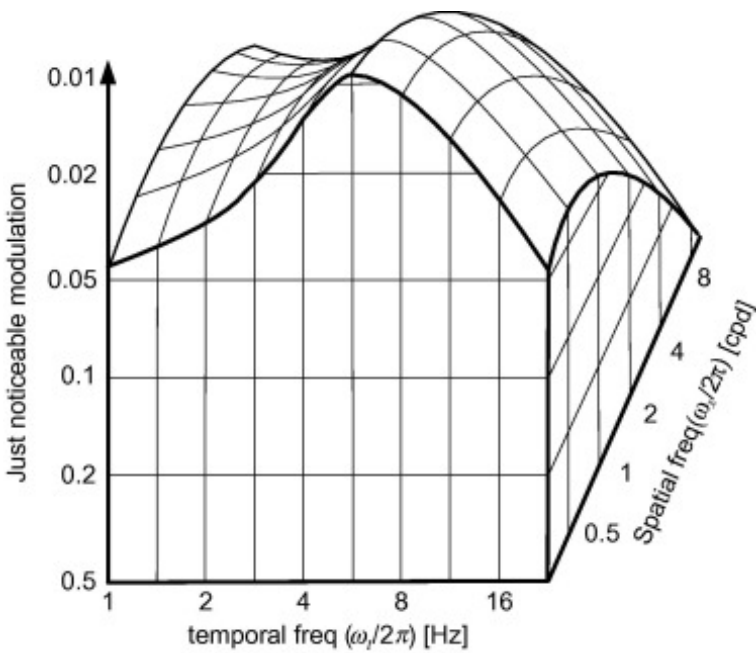
2.8.1. 时间脑脊液

视觉系统对照明的时间变化很敏感。这一特性在相机和显示器的设计中众所周知，并利用了它们来确定其时间更新率。时间限制由光感受器的反应速率、视网膜[回路](#)和视觉皮层的响应时间[2]而施加。眼睛在事件图像被移除后（视觉持久）后会保留短暂的感觉。直到最近，人们一直认为60赫兹的更新率足以传达平滑运动的感知。然而，最近的调查在更高的空间分辨率和更大的显示器的背景下对此提出了质疑。[第13章](#)对此进行了更详细的讨论。

凯利[19]表明，在不同频率上，对时间变化模式的视觉敏感性响应本质上是带通的，高度依赖于显示器亮度，调光显示器的峰值约为7或8Hz，在24Hz左右变得越来越小。然而，对于更亮的显示器，峰值上升到大约20赫兹，以大约80赫兹消失。时间变化变得不明显的频率很重要，因为它与我们因刷新显示器而经历的闪烁有关。显示亮度的变化解释了为什么我们可以（几乎）为电影院提供24帧/秒的采样，但电视屏幕需要更高的速率，计算机显示器需要更高的速率（因为它们更近）。

2.8.2. 时空脑脊液

图2.28[18]显示了HVS[时空特征](#)的[图表](#)。这显示了上述低[空间频率](#)的带通时间特征。然而，随着空间频率的增加，时间响应在性质上变得更低通，同样，随着[时间频率](#)的提高，[空间响应](#)趋向于低通特性。因此，对于移动速度更快的物体，我们无法轻松地吸收其空间纹理。正如我们将在[第](#)



下载：下载全尺寸图像

图2.28。时空脑脊液。

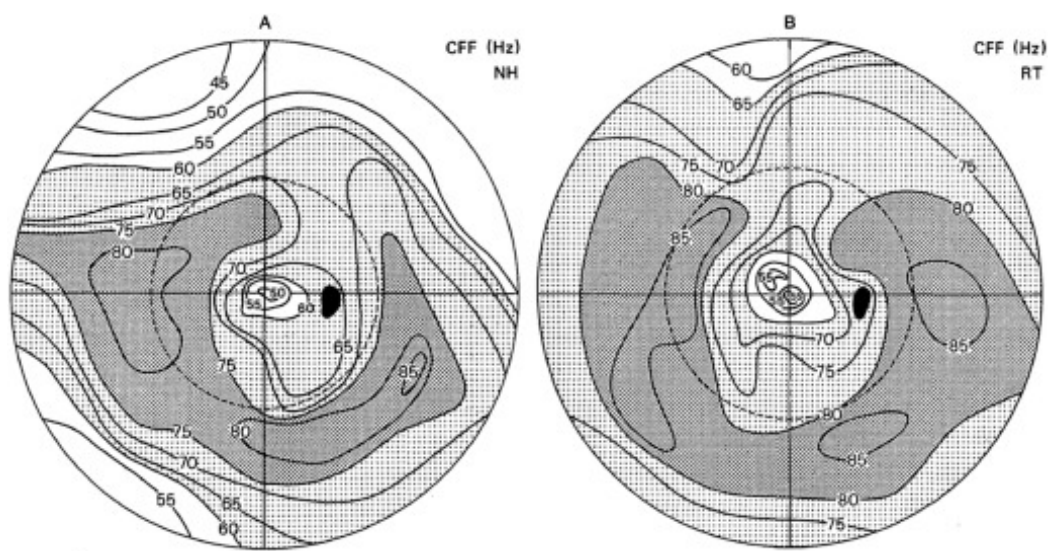
改编自凯利[18]。

2.8.3。闪烁和外围视觉

闪烁聚变阈值是间歇性光刺激在观察者看来稳定的频率。这取决于一系列因素，包括调制的频率和深度、照明的波长、刺激发生地在视网膜上的位置以及光或暗的适应程度。其他因素，如疲劳，也会产生影响。

泰勒[23]证明，闪烁阈值（或临界闪烁频率（CFF））随亮度（较亮的源较高）和视网膜位置（杆子响应快于锥体）而变化。因此，在外围视觉中，闪烁可以以比凹陷视觉更高的频率感知到闪烁。这很重要，因为我们关于闪烁的所有假设都基于中央视觉和小型实验屏幕。这解释了为什么随着屏幕变得更亮，尺寸变大，闪烁变得更加明显。这就是为什么电视制造商不得不将样本升级到300甚至600赫兹。

图2.29[23]显示了用于方波调制的视野CFF轮廓（以每秒周期为单位），作为偏心率和子午线的函数，场大小缩放以刺激每个偏心度的恒定数量的锥体，供两名观察者使用。这显示了外围高达90赫兹的闪烁频率。



[下载：下载全尺寸图像](#)

图2.29。临界闪烁频率的变化（经泰勒许可转载[23]）。

还必须注意，CFF通常指非结构化刺激。如果与空间特征的相互作用令人感兴趣，时间对比灵敏度函数（CSF）的信息量更大。有趣的是，人们观察到??脑脊液与鼻水平经的离心率无关，最高可达30°[24]。当然，在自然场景的现实中，复杂的交互存在于场景中不同的交互模式和动作中。

2.9.注意力和眼球运动

HVS为我们创造了一种完全感知场景的错觉，而实际上只有相对较少的信息传递给更高的大脑水平。这样，我们只感知绝对必要的内容，而忽略所有其他信息，以压缩不断进入视觉系统的大量视觉数据。实现这一目标的过程被称为注意力，并在所有皮层水平上得到支持，但可能最重要的是通过凹陷和眼球运动。

2.9.1。洒洒和关注

注意调节视觉系统其他部分的响应，以优化观看体验，并尽量减少实现这一目标所需的额外信息量。注意可以是自下而上（外来）或自上而下（内生）或两者兼而有之。在第一种情况下，当低级反应吸引注意力（例如通过检测外围的运动）时，就会发生这种情况。这个过程通常被称为显著性，基于低水平HVS对特定特征的反应强度。自上而下或内生注意力（更多）有意识地被手头的任务引导。

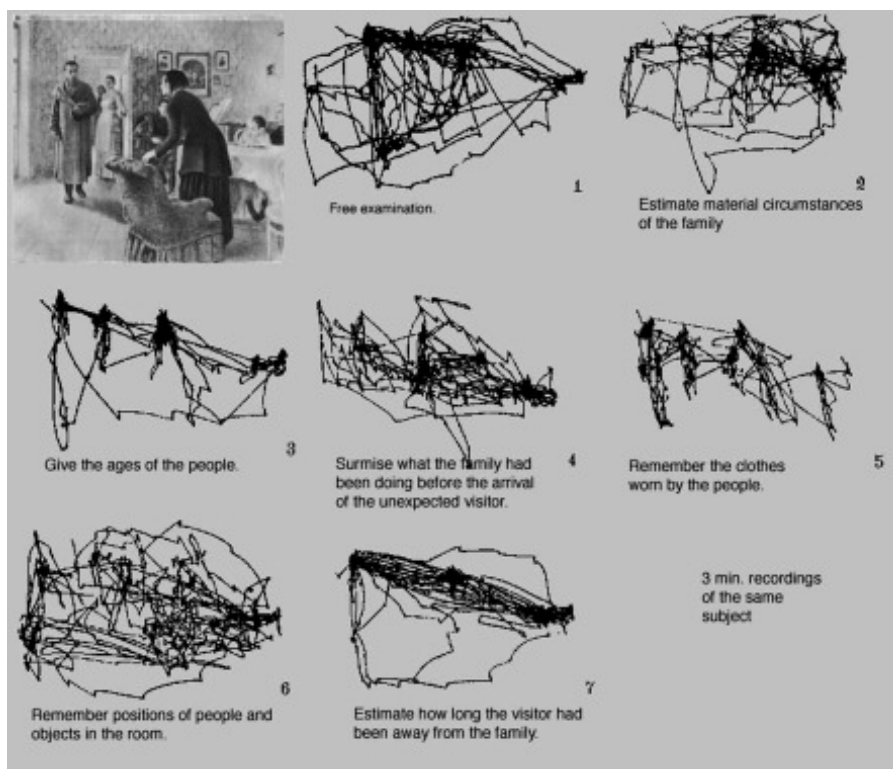
对显著性模型进行了大量研究，其中最值得注意的是Itti等人[25]自上而下的注意力调节自下而上的特征。眼球运动被用来让我们关注快速变化的感兴趣领域，我们接下来将简要考虑这些领域。在某些视频压缩场景中，注意力和焦点已被提出作为根据观看情况优化分配位的方法[26]、[27]。

2.9.2。眼球运动

除了补偿头部运动外，我们的眼睛还以两种不同的方式移动：

1. **萨卡德斯**：作为**固定点**之间的快速非自愿移动。视觉系统在囊形过程中是空白的，因此我们不会因扫描而体验到快速运动效果。萨卡德有两种形式：
 - a. **微囊**：用于刷新因适应而可能脱落的细胞反应的小型非自愿运动。
 - b. **焦点**：眼睛在光视觉中也快速移动，以探索场景的新区域，给人一种在整个视野中敏锐度增加的印象。
2. **平滑追击**：这些是注意力引导的更平滑的自愿运动，例如，在跟踪移动的物体时会发生。这个过程使感兴趣的对象以凹陷为中心，以保持高空间敏锐度并减少**运动模糊**。在平稳追击的条件下，之前讨论的时空界限发生了巨大变化。吉罗德[29]重新计算了这些特征，并证明这种类型的眼球运动具有延长我们视觉系统时间极限的效果，表明可以感知数百赫兹的频率。

眼动跟踪在视觉研究[32]中被广泛使用，以评估观察者对刺激的反应和固定。一些最早和最著名的研究是由亚布斯发表的[28]。雅布斯的实验结果如图2.30所示，该实验为观察者提供了一系列与游客绘画相关的观看任务。观察囊形模式与指定任务的关系很有趣。



下载：下载全尺寸图像

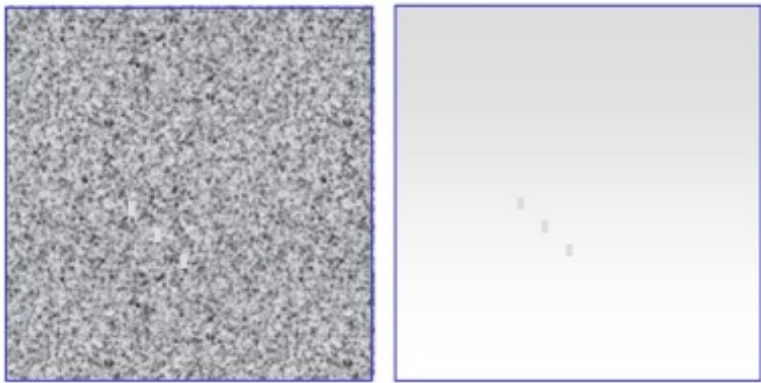
图2.30。响应任务的眼球运动（公共领域图像来自：http://commons.wikimedia.org/wiki/File:Yarbus_The_Visitor.jpg）。

2.10。视觉遮罩

视觉遮罩是通过呈现称为“遮罩”的第二个刺激来降低或消除一种被称为“目标”的短暂刺激的可见性。Girod [30a] 概述了遮罩和其他心理视觉效果的影响。

2.10.1。纹理遮罩

当背景是纹理而不是普通时，目标的可见性阈值会增加。当存在高对比度掩码时，空间（或纹理）掩码会导致给定空间频率的对比阈值上升。交互式感受场的抑制性导致当在其他模式的背景下查看HVS敏感性降低到某些空间模式。掩码的影响与掩码的空间频率含量与目标的相似性有关。这种效果可以通过识别掩码模式在目标测试模式刺激的同一空间过滤器中引发响应来解释，使其在后期视觉处理阶段无法区分[2]。纹理遮罩的例子如图2.31所示。这显示了两个子图形（三个矩形）中相同的目标。目标在右侧非常明显，背景平坦，但左侧很难检测到。

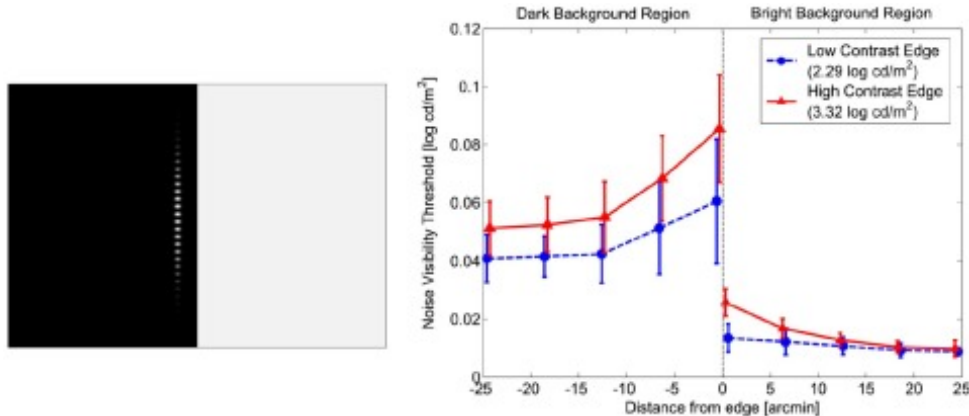


下载： [下载全尺寸图像](#)

图2.31。纹理遮罩示例。

2.10.2。边缘遮罩

据观察，在空间边缘附近存在明显的遮罩效应[30a]。张等人[31]在动态范围内容较高的情况下也证明了这一点。图2.32显示了目标和边缘的实验设置以及根据边缘对比度和目标与边缘距离的噪声可见性变化。

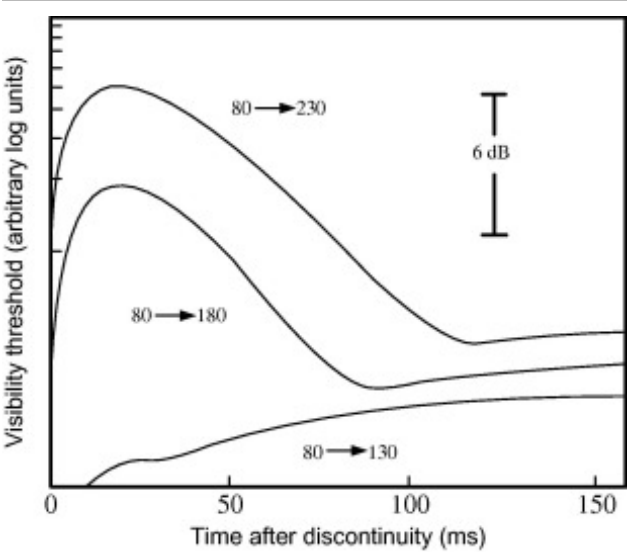


下载： [下载全尺寸图像](#)

图2.32。用于高动态范围和低动态范围内容的边缘遮罩。

2.10.3。时间遮罩

Girod [30a] 展示了视频在时间不连续的情况下的遮罩效果（例如电影中的镜头剪辑或场景变化）。对于伽马预失真视频信号，此效果如图2.33所示。可以观察到，随着时间不连续性程度的增加（此处显示为8位边缘），时间遮罩效应的持续时间超过100毫秒。



下载：下载全尺寸图像

图2.33。各种边缘台阶尺寸的时间遮罩效果。

经吉罗德许可转载[30b]。

2.11。摘要：图像和视频压缩的感知基础

2.11.1.影响因素

正如我们在本章中所看到的，许多因素影响着我们看待事物的方式：环境光、场景或显示亮度、场景的时空内容、观看距离、注意力、任务、期望、受试者之间的生理差异和许多其他微妙的环境因素。

许多研究人员试图通过各种抽象心理物理实验来理解这些对感知的影响。虽然这些为我们视觉反应的局限性和变化提供了非常有价值的见解，但很少有人涉及真实场景及其内部的复杂互动。然而，我们可以取消几个重要的准则，其中一些准则已用于采集、显示设备的设计，以及优化信号表示和压缩方法。其他人拥有未来的潜力。这些将在下一小节中总结。

2.11.2。我们学到了什么？

下面概述了我们视觉系统的观测特征，以及如何在设计基于视觉的技术中利用这些特征。

颜色处理

HVS特征	含义
-------	----

HVS使用电磁频谱可见部分的三个颜色通道感知颜色	用于传输和存储的采集和显示技术和可视化信号表示可以基于三刺激框架
基于对手理论，HVS将视觉信号处理为亮度通道和两个色度通道	仅限亮度的内容很容易被接受为自然内容。视觉信号表示可以基于亮度和色度分裂的颜色空间
HVS的亮度通道比其色度通道高	亮度通道可以分配比亮度通道低的带宽

空间响应

HVS特征	含义
HVS对空间频率的反应通常在5-8cpd时达到峰值，然后脱落，超过40-60cpd变得越来越小	超出上限的频率不需要处理。更高的频率可以更粗地编码，对量化产生影响。低频亮度滚出一般不被利用
边缘是人类视觉系统处理的一个重要方面。灵敏度随对比度增加而增加	边缘应该保留。量化引入的人工边缘工件非常引人注目，应该避免
HVS在平原地区比在纹理地区更容易检测到噪音	量化噪声的空间（对比）掩蔽可以在纹理区域利用

时间反应

HVS特征	含义
HVS时间对比度灵敏度峰值约为5-20赫兹，并降低至20-80赫兹以上（亮度取决于）	视觉处理技术应能在视觉阈值的两倍处采样，并应考虑到观看条件
我们以大约50赫兹的速度将单个帧视为连续体（没有闪烁）的能力	获取和显示的帧速率通常应至少为50赫兹
临界闪烁频率在外围高于凹区	屏幕更大、观看更近需要更高的帧刷新率

时空反应

HVS特征	含义
在较低的时间频率下，眼睛可以更好地解析空间频率，而在较高的空间频率下，时间响应会变成低通量	在交错扫描中，可以利用时间和空间敏锐度之间的这种权衡，以提供帧速率增加的感知

处理纹理

HVS特征	含义
有证据表明，HVS中会发生纹理填充	图像和视频编码的分析-合成方法保持潜力
纹理变化和变化比形状变化更难检测	如上所述

深度线索

HVS特征	含义
HVS在处理的所有阶段都利用了许多不同的深度线索	立体三维内容可能不是实现更多沉浸式内容的唯一方法

动态范围

HVS特征	含义
我们的视觉响应取决于显示器的平均亮度	电影帧率可能低于电视或计算机显示器。更高的动态范围显示器可以更好地感知现实和深度
在宽广的亮度水平上，JND与背景亮度之间存在线性关系，在低值和高值时进一步增加	这可以在应用强度依赖量化时加以利用，允许低电平和高电平信号的步长更快增加
亮度与感知亮度之间存在非线性关系	对信号编码的影响，因为在编码之前，非线性（伽马）函数应用于相机输出

注意力、凹陷和眼球运动

HVS特征	含义
我们的高视力仅限于2度弧度	位分配策略可以针对图像中注意力集中的区域（如果已知！）
HVS的时空频率响应因眼球运动而显著变化。在平稳的追击条件下，时间敏锐度可能高得多	特别是屏幕更大更亮，显示器必须能够应对数百赫兹的时间频率

主体和环境之间的生理差异

HVS特征	含义
个人的反应会自然而变化，也会随着年龄而变	主观质量评价实验应基于大量受试者的结果（通常>20）

这本书的其余部分将更详细地探讨这些特征的含义和利用。

[Recommended articles](#)

[Citing articles \(0\)](#)

参考文献

- [1] R.斯诺登, P.汤普森, T.特罗西亚科
基本愿景
牛津大学出版社 (2006)
[谷歌学术](#)
- [2] G.马瑟
感知和感知的基础
(第二版), 心理学出版社 (2009年)
[谷歌学术](#)
- [3] B.旺德尔
愿景基金会, Sinauer Assoc (1995)
[谷歌学术](#)
- [4] D.马尔
视觉
弗里曼 (1982)
转载 麻省理工学院出版社 (2010)
[谷歌学术](#)
- [5] H.冯·亥姆霍兹
生理光学手册
(第一版), 汉堡和莱比锡, 沃斯 (1896)
[谷歌学术](#)
- [6] B.朱雷斯
计算机生成模式的双目深度感知
贝尔系统技术期刊, 39 (1960年), 第1125-1162
[CrossRef](#) [在Scopus中查看记录](#) [谷歌学术](#)
- [7] D.胡贝尔, T.维塞尔
猫的纹状皮层中单个神经元的感受场
生理学杂志, 148 (1959年), p.574-591

- [8] R.格雷戈里
眼睛和大脑
(第五版)，普林斯顿大学出版社（1997年）
[谷歌学术](#)
- [9] M.克劳斯
转化语法与图片整理
A.格拉塞利（编辑），图像的自动解释和分类，学术出版社（1969）
[谷歌学术](#)
- [10] H. 科尔布
视网膜是如何工作的
美国科学家，91（2003），p.28-35
[在Scopus中查看记录](#) [谷歌学术](#)
- [11] R.彩票，D. 珀维斯
颜色对亮度的影响
自然神经科学，2（11）（1999），p.1010-1014
[在Scopus中查看记录](#) [谷歌学术](#)
- [12] D.穆斯塔菲亚，A. 恩格拉，K. 帕尔切夫斯基亚
锥形感光器的结构
视网膜和眼睛研究进展，28（4）（2009年），第289-302页
[谷歌学术](#)
- [13] A.内拉瓦利，B. 哈斯卡尔
数字图片：表示、压缩和标准
(第二版)，全体会议出版社（1995年）
[谷歌学术](#)
- [14] K. 维兰卡，L.瓦苏，D. 钱德勒
论自然场景中带限相位失真感
人类视觉和电子成像十六。SPIE会议记录，第7865卷（2011年），第78650C
[CrossRef](#) [在Scopus中查看记录](#) [谷歌学术](#)
- [15] L. 科马克
早期人类视觉的计算模型
A.博维奇（编辑），图像和视频处理手册，学术出版社（2000年）
[谷歌学术](#)
- [16] A.沃森（编辑），数字图像和人类视觉，麻省理工学院出版社（1998年）
[谷歌学术](#)
- [17] A.沃森，A. Allhumada，J.法雷尔

可见性窗口：时间采样视觉运动显示器保真心理物理理论

美国光学学会杂志A, 3 (3) (1986年), p.300-307

[CrossRef](#) [在Scopus中查看记录](#) [谷歌学术](#)

[18] D.凯莉

时空正弦波阈值的适应效应

视觉研究, 12 (1972), p.89-101

[在Scopus中查看记录](#) [谷歌学术](#)

[19] D.凯莉

对时间依赖性刺激的视觉反应

美国光学学会杂志, 51 (1961), 第422-429页

[CrossRef](#) [在Scopus中查看记录](#) [谷歌学术](#)

[20] R.沙普利, M.霍肯

原发性视觉皮层颜色感知的神经机制

神经生物学的当前观点, 12 (4) (2002), 第426-432页

[文章](#)  [下载PDF](#) [在Scopus中查看记录](#) [谷歌学术](#)

[21] <http://en.wikipedia.org/wiki/File:CIE1931xy_blank.svg>。

[谷歌学术](#)

[22] J.曼诺斯, D.萨克里森

视觉误差标准对图像编码的影响

IEEE 信息理论交易, IT-20 (1974年), p.525-536

[在Scopus中查看记录](#) [谷歌学术](#)

[23] C.泰勒

视觉调制灵敏度分析。三。外周闪烁灵敏度的子午变化

美国光学学会杂志, A4 (8) (1987年), p.1612-1619

[CrossRef](#) [在Scopus中查看记录](#) [谷歌学术](#)

[24] 五.维尔苏, J.罗瓦莫, P.劳里宁, R.纳萨宁

时间对比灵敏度和皮质放大

视觉研究, 22 (1982), p.1211-1217

[文章](#)  [下载PDF](#) [在Scopus中查看记录](#) [谷歌学术](#)

[25] L.伊蒂, C.科赫, E.尼布尔

基于显著性的视觉注意力模型, 用于快速场景分析

IEEE 模式分析和机器智能交易, 20 (11) (1998), p.1254-1259

[在Scopus中查看记录](#) [谷歌学术](#)

[26] D.Agrafiotis, C.卡纳加拉贾, D.公牛, J.凯尔, H.塞尔斯, M.Dye

用于低比特率手语通信的感知优化视频编码系统

信号处理: 图像通信, 21 (7) (2006年), p.531-549

[文章](#)  [下载PDF](#) [在Scopus中查看记录](#) [谷歌学术](#)

- [27] 美国。戴维斯, D. Agrafiotis, C. 卡纳加拉贾, D. 公牛
开放签名视频中用于凝视预测的多线索贝叶斯状态估计器
IEEE 多媒体交易, 11 (1)(2009), p.39-48
[CrossRef](#) [在Scopus中查看记录](#) [谷歌学术](#)
- [28] A.亚尔布斯
眼球运动和视觉
纽约全体会议 (1967)
[谷歌学术](#)
- [29] B.吉罗德
运动补偿; 视觉方面、准确性和基本极限
M.塞赞, R. Legendijk (编辑), 运动分析和图像序列处理, Kluwer (1993年), p.126-152
[谷歌学术](#)
- [30a] B.吉罗德
图像传播的心理视觉方面
信号处理, 28 (3) (1992年), 第239-251
[文章](#)  [下载PDF](#) [在Scopus中查看记录](#) [谷歌学术](#)
- [30b] B.吉罗德
视频信号中时空掩蔽的信息理论意义
会议记录SPIE/SPSE人类视觉、视觉处理和数字显示会议, 美国加利福尼亚州洛杉矶 (1989年), p.178-187
[CrossRef](#) [谷歌学术](#)
- [31] Y. 张, D. Agrafiotis, M.纳卡里, M. 姆拉克, D. 公牛
高动态范围内容的视觉掩码现象
IEEE国际图像处理会议记录 (2013年), p.2284-2288
[CrossRef](#) [在Scopus中查看记录](#) [谷歌学术](#)
- [32] 美国。肝刺, 我。吉尔克里斯特, S. 埃弗林
牛津眼动手册
牛津大学出版社 (2011)

 [Download](#)

人视网膜中杆状和锥体的视觉色素

生理学杂志, 298 (1980), 第501-511

[CrossRef](#) [在Scopus中查看记录](#) [谷歌学术](#)

☆ 有关图2.1、图2.11、图2.14、图2.16、图2.17的彩色版本, 请参阅电子版或网站。

1 国际照明委员会。



[关于ScienceDirect](#)

[远程访问](#)

[购物车](#)

[广告](#)

[联系和支持](#)

[条款和条件](#)

[隐私政策](#)



我们使用cookie来帮助提供和增强我们的服务，并定制内容和广告。继续即表示您同意**使用cookie**。

版权所有?2021爱思唯尔B.V.或其许可方或贡献者。ScienceDirect ®是爱思唯尔B.V.的注册商标。

ScienceDirect ®是爱思唯尔B.V.的注册商标。