

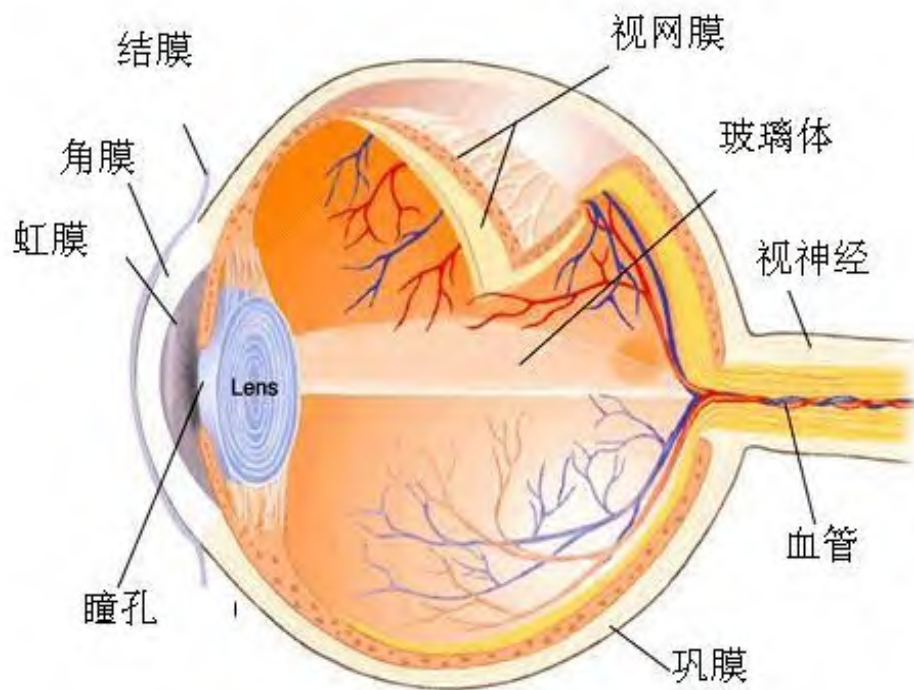
# 第1讲：数字图像和视频

---

- 人类视觉系统与色彩
  - 色觉，对比灵敏度，空-时频率响应
  - 立体/深度感知，适应与会聚（辐辏）
- 模拟视频：逐行扫描与隔行扫描
- 数字视频：表示与标准
  - 空间分辨率，帧率，色彩，动态范围，位深
- 3D视频：显示与表示
  - 立体，多视角，多视角+深度，超多视角，全息成像
- 数字视频的应用
  - 数字电视，数字影院，互联网中的视频流
- 数字图像与视频的质量
  - 视觉效果损伤，主观质量评价，客观质量评价

# 人眼的构造

**Lens（晶状体）：** 调节焦距



**视网膜：** 由大量的光敏细胞和神经纤维组成

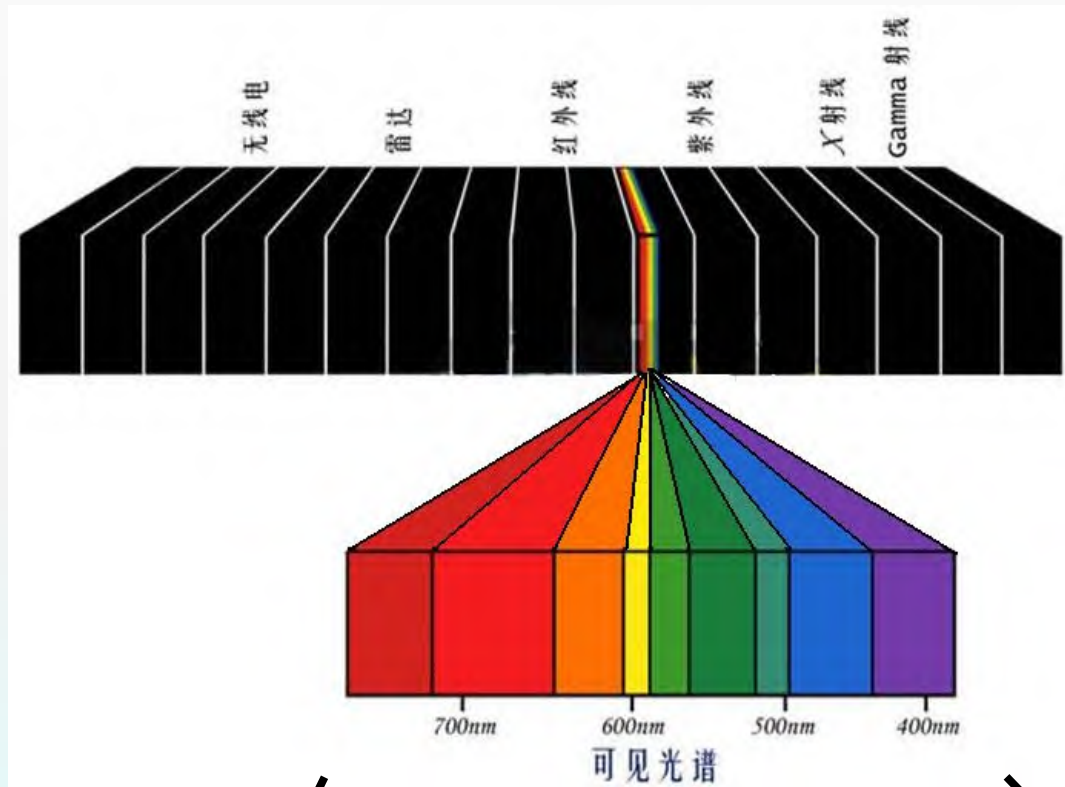
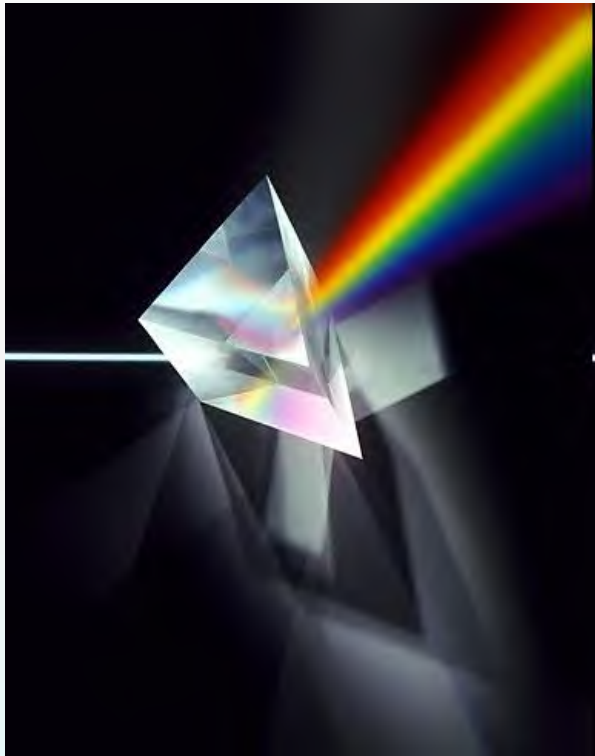
**光敏细胞**包括：

- 1) **锥状细胞：** 明亮环境
- 2) **杆状细胞：** 黑暗环境  
灰度感觉

**锥状细胞**具有3种类型，它们具有不同的光谱特性：

**红、绿、蓝**3个子频段——彩色视觉**三基色理论**的生理依据

# 可见光谱



$\lambda = 780 \text{ nm}$

$\lambda = 380 \text{ nm}$

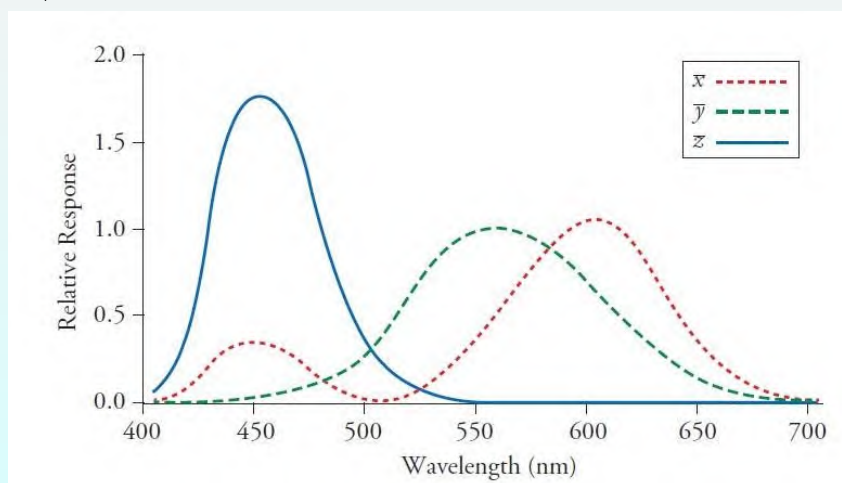


# 色觉：三基色组成

- **色觉**：可通过将波长 $\lambda$ 分成三段来仿真三种视网膜锥形细胞的色度感觉

$$f_k(x_1, x_2) = \int f(x_1, x_2, \lambda) m_k(\lambda) d\lambda \quad k = r, g, b$$

其中 $m_k(\lambda)$ 为第 $k$ 个锥形细胞类型或色度传感器的**波长灵敏度**函数(色彩匹配函数)



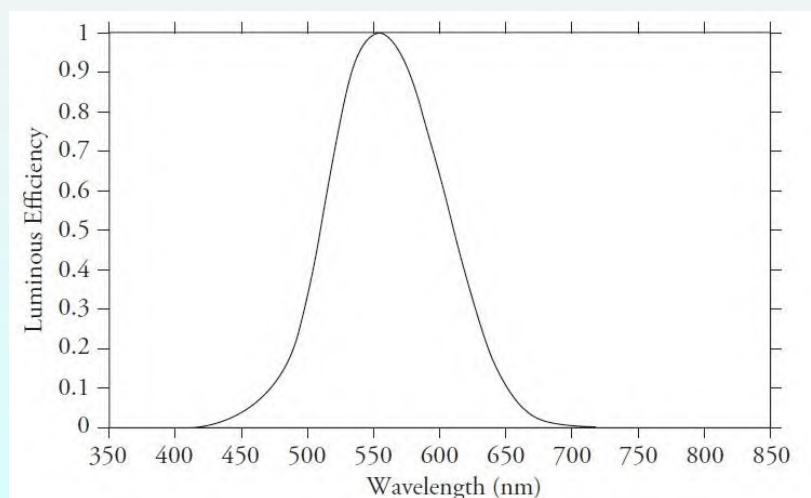
# 色觉：亮度效率

- **二次处理**：人眼具有二次处理阶段，锥状细胞所感觉到的R、G、B值被转换成一个**亮度值**和两个**色度（色差）值**。与光的感观亮度相关的亮度值Y可写成：

$$Y(x_1, x_2) = \int f(x_1, x_2, \lambda) l(\lambda) d\lambda$$

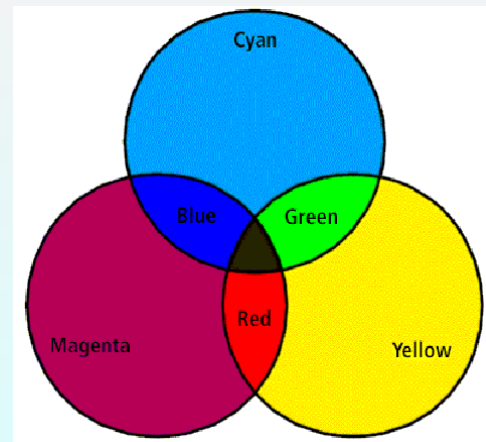
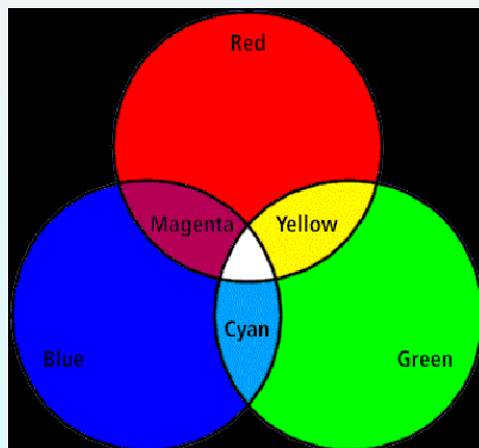
其中  $l(\lambda)$  为国际照明委员会所确定的**发光效率函数**，表明了对于人类亮度感觉每个波长的能量贡献。

- **两个色度值**描述了对光的**感知色度**



# 色彩空间

- **色彩空间**：定义了确切的关于色度的基色（例如：红、绿、蓝）。 诸如：CIERGB, CIEXYZ, sRGB。
- **色彩范围**：由于显示设备只能产生非负量，且需要有足够数量的亮度值，因此在一个给定设备上所能产生的颜色范围（色谱）是有限的。设备的颜色特征由国际色彩联盟(ICC) 配置文件来指定。
- **相加色彩空间**
  - 发射光
  - LCD, LED, 等
- **相减色彩空间**
  - 反射光
  - 打印文档，绘画等



# 色彩管理

- **色彩管理：** 不同设备产生一致的颜色，需将设备相关颜色值映射到一个与设备无关的标准色彩空间中。
- **ICC配置文件：** 从一个设备的颜色空间映射到一个标准的设备无关的颜色空间，称作配置关联空间(PCS)，如：CIEXYZ，然后再从 PCS 映射到一个设备的颜色空间。
- **RGB设备转换到CMYK设备：**
  - 使用RGB设备的ICC配置文件将每个(R, G, B)三基色转换成PCS。
  - 使用第二个设备的ICC配置文件将PCS转换成C, M, Y, K值。



# 色彩：亮度-色度表示

- **Y-Cr-Cb 色彩模型**被广泛应用于图像处理和压缩之中。
- **R-G-B 到 Y-Cr-Cb 的转换** (ITU BT. 601)

$$Y = 0.299 R + 0.587 G + 0.114 B$$

$$Cr = 0.499 R - 0.418 G - 0.0813 B + 128$$

$$Cb = -0.169 R - 0.331 G + 0.499 B + 128$$

- **Y-Cr-Cb 到 R-G-B 的转换**

$$R = Y + 1.402 (Cr - 128)$$

$$G = Y - 0.714 (Cr - 128) - 0.344 (Cb - 128)$$

$$B = Y + 1.772 (Cb - 128)$$



# 色彩：感知表示

- **归一化 rgb:** rgb值的定义是为了减少颜色成分对图像亮度的依赖性。

$$r=R/(R+G+B)$$

$$g=G/(R+G+B)$$

$$b=B/(R+G+B)$$

这些值的范围是 0 到 1, 且有:  $r + g + b = 1$ ;

因此, 其中只要有两个分量就可确定, 因此可以使用  $(r, g)$ 。

- **色调、饱和度、强度(HSI)**

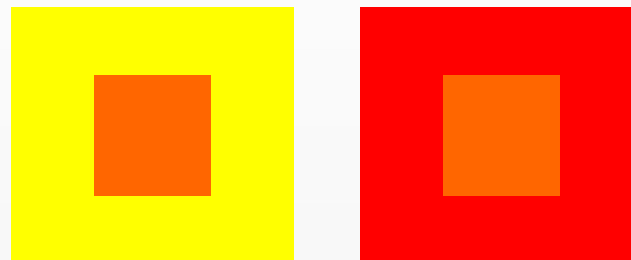
$$H = \begin{cases} \theta & \text{if } B \leq G \\ 360 - \theta & \text{if } B > G \end{cases} \quad \text{其中: } \theta = \arccos \left\{ \frac{1/2[(R-G)+(R-B)d]}{\sqrt{(R-G)^2+(R-B)(G-B)}} \right\}$$

$$S = 1 - \frac{3 \min\{R, G, B\}}{R + G + B}$$

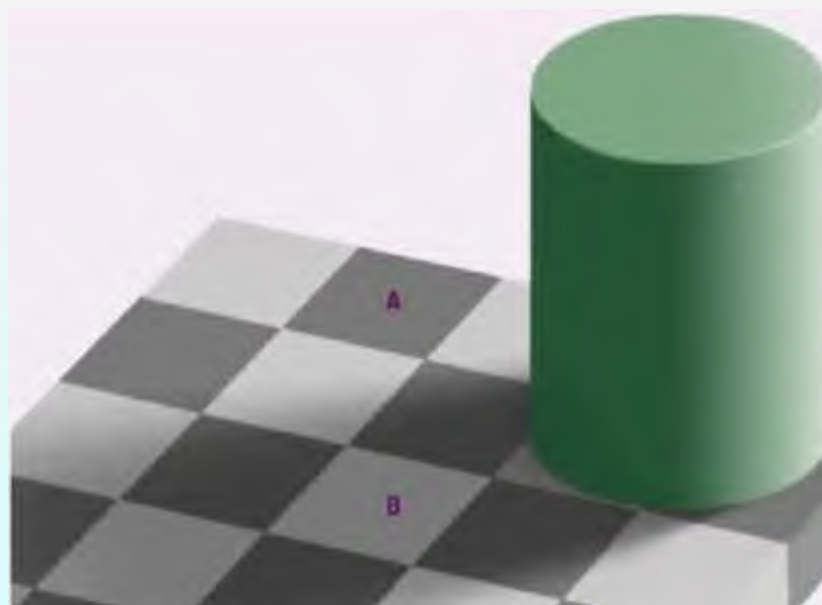
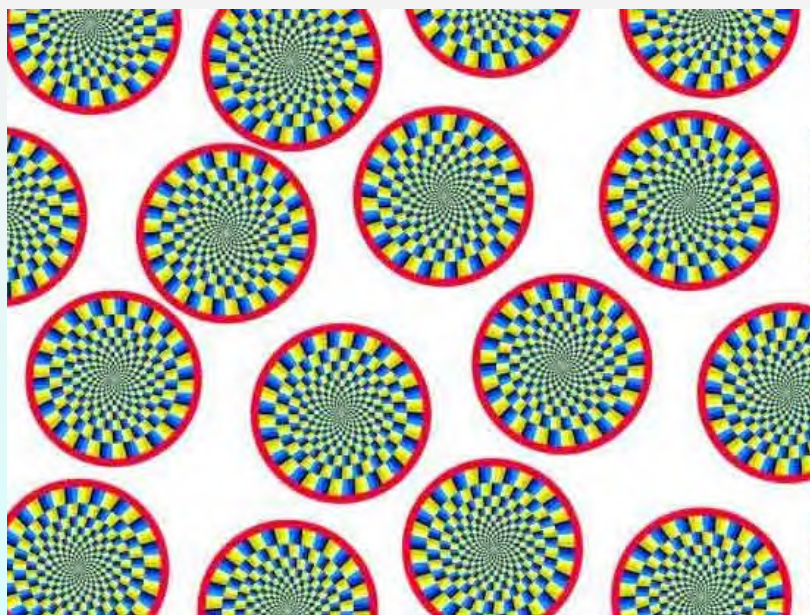
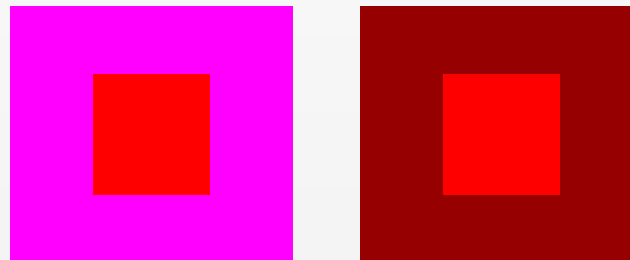
$$I = (R + G + B)/3$$

# 视觉特性—错觉

◆ 色调对比效应



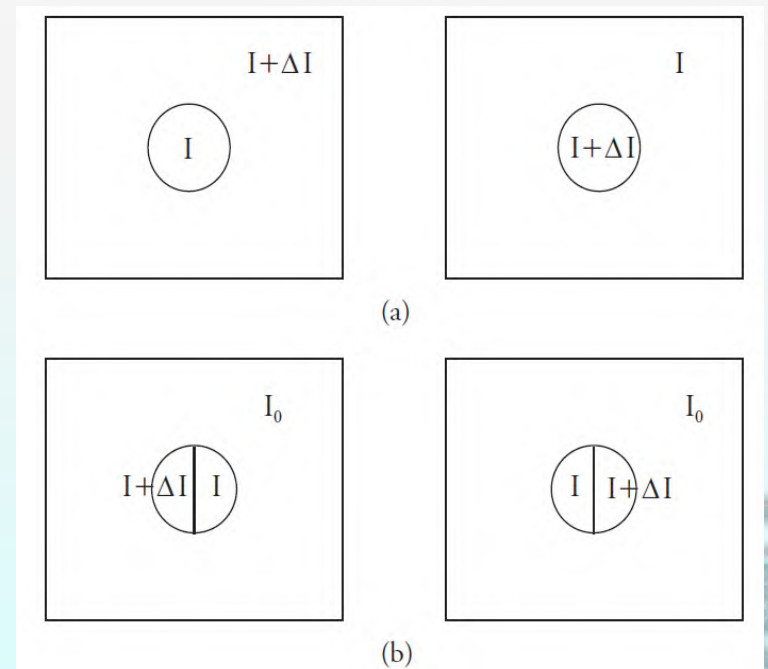
◆ 饱和度对比效应



# 对比灵敏度

- **韦伯定律**: 小强度差异在较暗背景上更具有可见性, 即:  
$$\frac{\Delta I}{I} = c \text{ (常数)}, \quad I > 0$$
其中  $\Delta I$  为 **最小可觉差(JND)**。

- **亮度适应**: 当观察者已适应的背景的亮度强度级别不同于前景亮度  $I$  时, 则观察者的亮度分辨能力会下降, 即: 如图所示, 如果  $I_0$  不同于  $I$ , 则 JND  $\Delta I$  将会提高。



# 对比灵敏度 (cont'd)

---

- 马赫带效应

- 人类视觉系统在亮度出现阶跃转换的周围会产生负脉冲和正脉冲信号

- 视觉掩蔽

- **空域掩蔽**：当目标细节与背景图像叠加在一起时会出现。当背景的空间复杂度较高时，目标细节的可见性就会较低。因此，噪声或损伤的可见性会随着背景的空间复杂度发生改变，空域平滑区域较大，而空域复杂区域则会降低。
- **时域掩蔽**：当两个激励按时间顺序依次呈现时会出现。快速运动物体和场景切换会引发时域掩蔽效应。



# 空间频率

- 低频
- 高频
  - 水平, 垂直
  - 角度方向



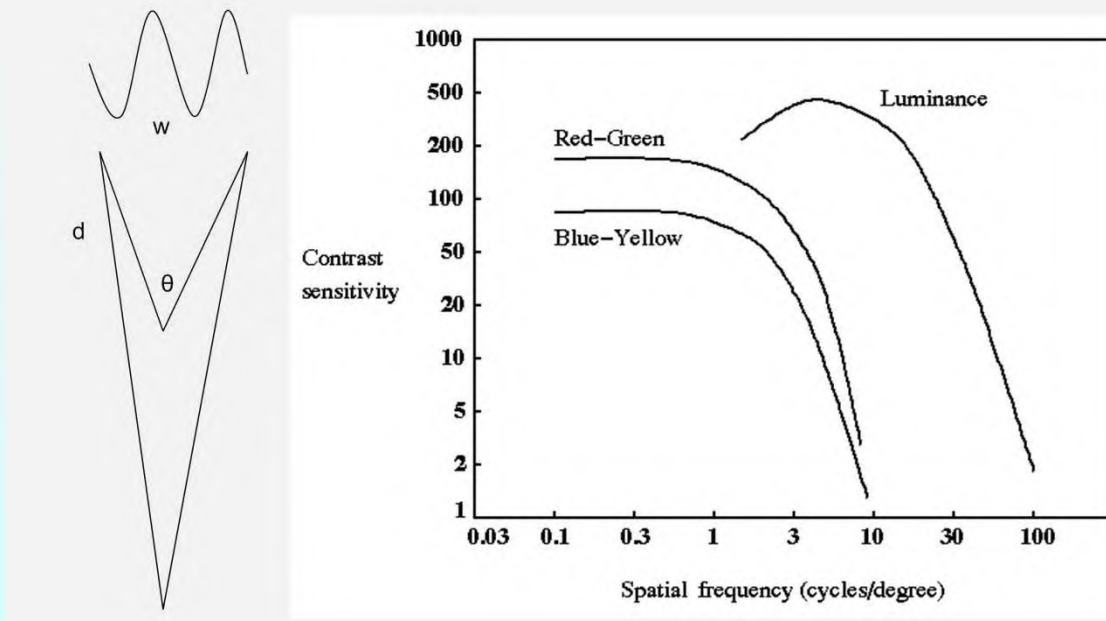
# 空间频率响应

- 设 **w** 表示**图像宽度**，考虑**观察者位置**同**图像的端点**以及和**图像的中间点**所形成的直角三角形，如果  $w/2 \ll d$ ，则：

$$\frac{\theta}{2} \approx \sin \frac{\theta}{2} = \frac{w/2}{d}$$

因此，

$$\theta \approx \frac{w}{d} (\text{弧度}) = \frac{180w}{\pi d} (\text{度})$$



## 空间频率响应 (cont'd)

- 设  $f_w$  表示每个图像宽度的周期数，则 **归一化水平空间频率**（即：每观察角度的周期数）为：

$$f_{\theta} = \frac{f_w}{\theta} = \frac{f_w d}{w} \text{ (周期/弧度)} = \frac{\pi d f_w}{180 w} \text{ (周期/度)}$$

- 当  $d$  增加时，同样的  $f_w$  会导致更大的  $f_{\theta}$ 。由于人眼对于高频成分具有较低的对比敏感度，因此从 **更远距离  $d$  观察** 相同的物体细节时 **将更加困难**。
- 人眼的空间频率响应具有 **低通/带通** 特性，对于 **亮度成分的高频细节** 比 **色度成分的高频细节** 更加敏感。

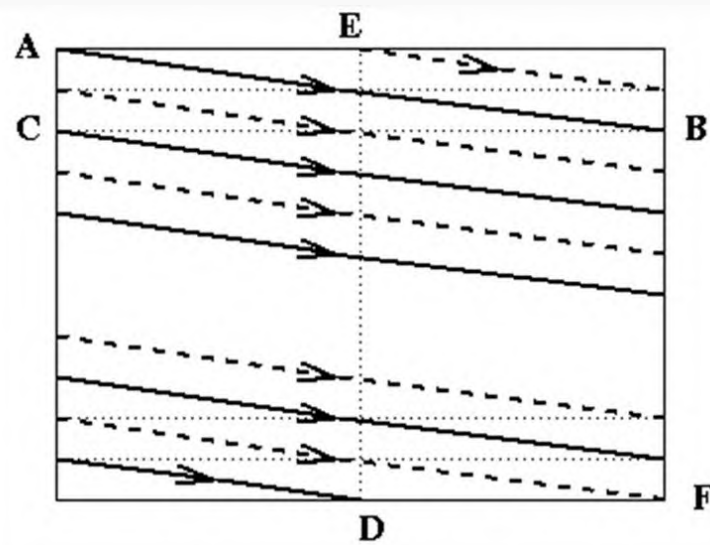
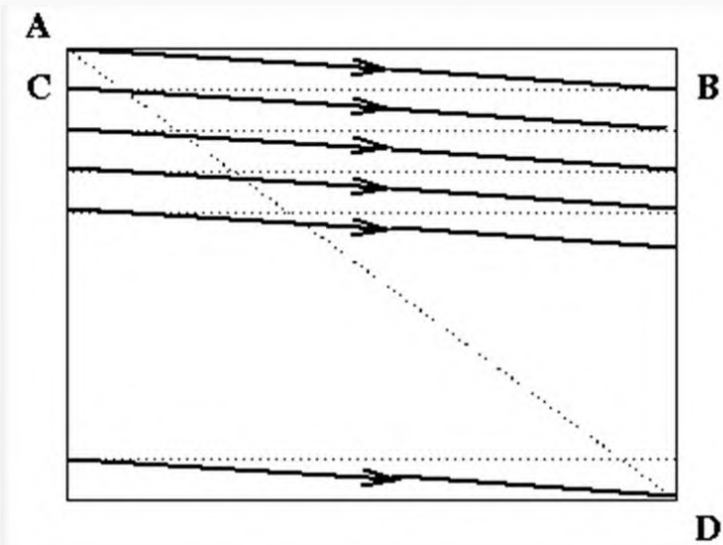
# 空-时频率响应

- **空间和时间频率响应**之间的相互作用
  - 我们可以交换空间的视频分辨率和时间分辨率。当视频具有**高运动**（快速的动作）时，眼睛不能感知高空间频率（细节）。
- **眼球运动**：人眼类似于一个可以在眼窝中自由运动的球体。有两种主要的目动类型，即“**眼跳**”和“**平稳跟踪**”，人眼对这两种运动类型的空频和时空频率响应会有所不同。
  - **眼跳**是在扫描场景时眼球的快速运动。“眼跳目动”使我们可以使用高分辨率视网膜中央凹对较大的场景区域进行扫描。
  - **平稳跟踪**是指当跟踪一个移动物体时的眼球运动。可使移动的图像保持在高分辨率视网膜中央凹上。



# 模拟视频

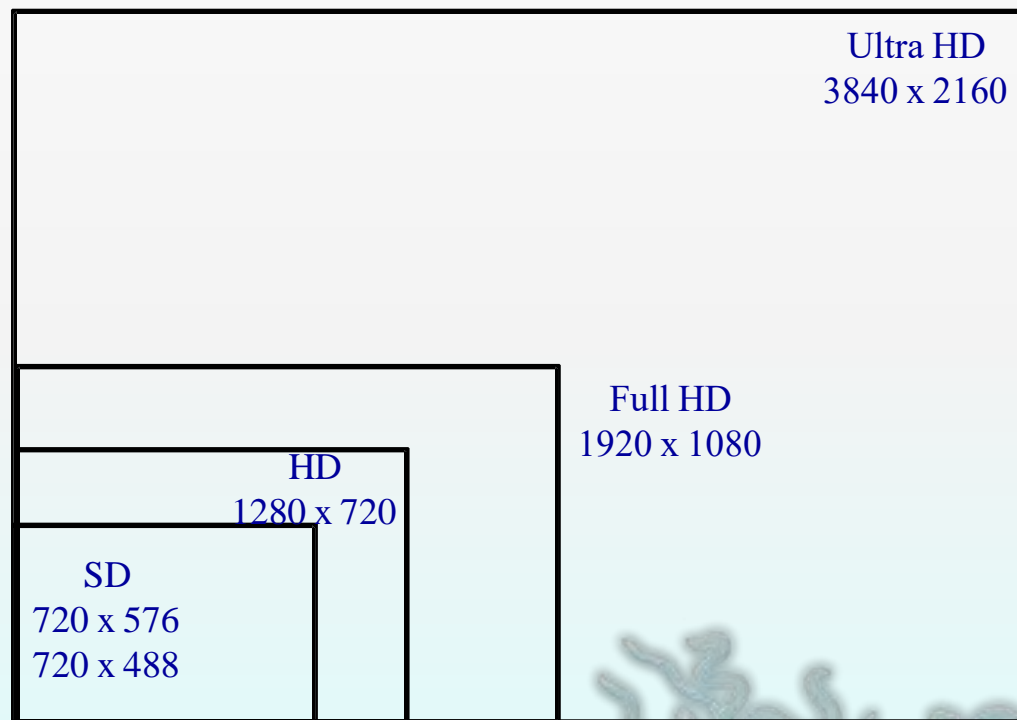
- 逐行 vs. 隔行扫描



- 模拟视频信号的参数包括：垂直分辨率、屏幕宽高比以及帧/场率。如果屏幕刷新频率高于 50 Hz，则人眼不会感觉到闪烁。逐行扫描方式具有较高的垂直分辨率，但同时需要更大的传输带宽。隔行扫描则是在同样带宽情况下牺牲垂直分辨率以换取更高的图像刷新频率。

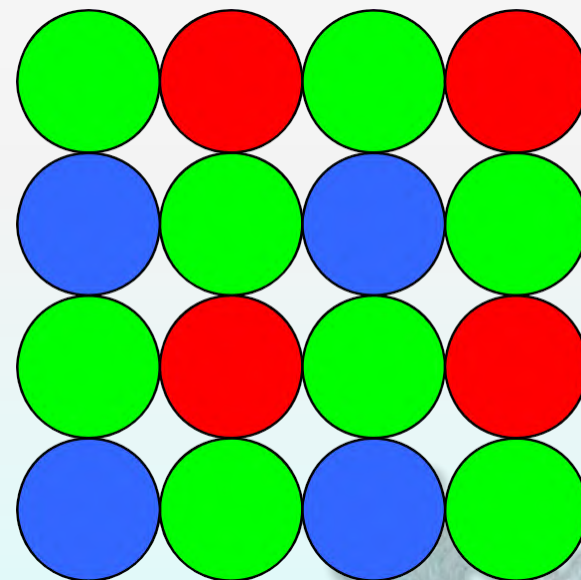
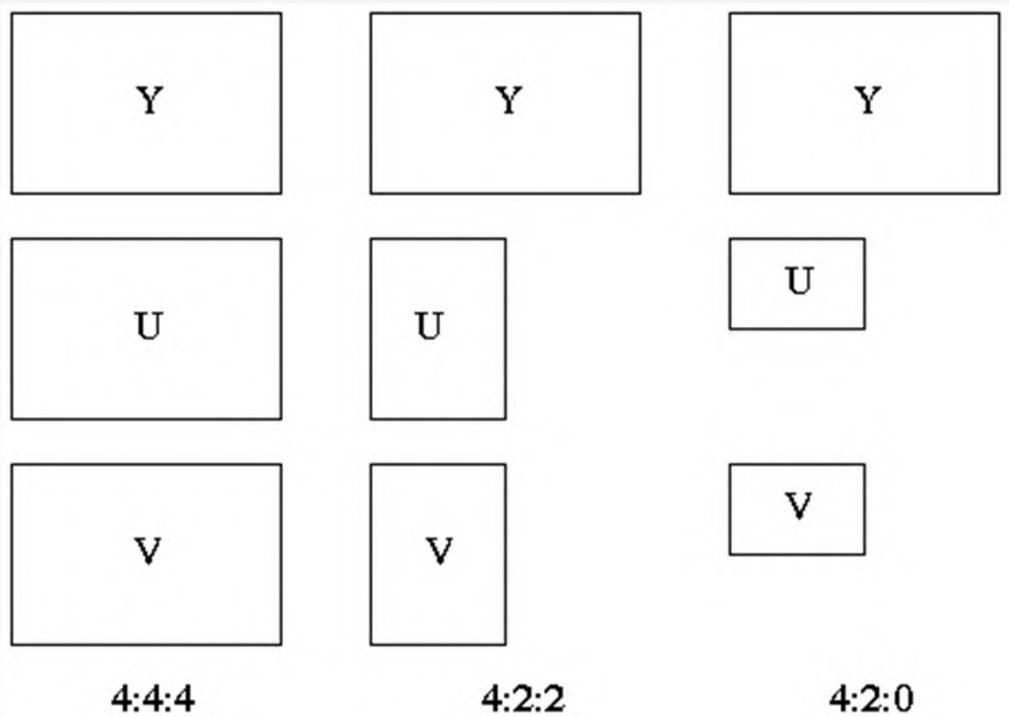
# 数字视频参数：分辨率

- 空间分辨率
- 帧率
- 逐行 vs. 隔行，例如：  
**Full-HD 1920×1080×50i**
- 图像宽高比，  
如： 16:9, 4:3



# 数字视频参数：色彩

- 色彩滤镜阵列 (CFA)，单传感器照相机采集、平板显示器呈现
- 压缩/传输所用的色度欠取样



Bayer CFA

# 数字视频参数：动态范围

- **照相机或显示器的动态范围** 是其能够表示或呈现的最大与最小亮度值之间的比率。可观察亮度级别的范围可从  $-4 \log \text{cd/m}^2$  (星光) 到  $6 \log \text{cd/m}^2$  (太阳光); 即: 动态范围是10 个对数单位。人眼具有分别处理5 个对数单位动态范围的快和慢适应方案。典型的照相机或显示器具有最大动态范围 300:1 (2.5 个对数单位)。
- **高动态范围(HDR) 图像采集:** 在不同的曝光水平下采集一系列图像, 然后将原始像素曝光数据 (与曝光时间呈线性关系) 进行合并, 以获取一张单独的HDR图像。
- **HDR 显示:** 新的显示技术可以显示50,000:1 或 4.7个对数单位动态范围, 最大亮度强度为  $8500 \text{cd/m}^2$ 。
- **位深:** 在8 - 16 bits/pixel/color之间。



# 数字视频标准

- ITU-R 电视广播标准

Standard	Pixels	Lines	Interlace/Progressive, Picture rate	Aspect ratio
BT.601-7 480i	720	486	2:1 Interlace, 30 Hz (60 fields/s)	4:3, 16:9
BT.601-7 576i	720	576	2:1 Interlace, 25 Hz (50 fields/s)	4:3, 16:9
BT.709-5 720p	1280	720	Progressive, 50 Hz, 60 Hz	16:9
BT.709-5 1080i	1920	1080	2:1 Interlace, 25 Hz, 30 Hz	16:9
BT.709-5 1080p	1920	1080	Progressive	16:9
BT.2020 2160p	3840	2160	Progressive	16:9
BT.2020 4320p	7680	4320	Progressive	16:9

- 显示/监视器标准(VGA, XGA)
- 数字视频接口标准 (DVI, HDMI)
- 数字视频压缩标准(MPEG, H.26x)

# 体视术: 3D 感觉

---

- **双目深度**—双目视差提供了相对深度暗示。3D显示仅提供两个视图。立体电视和数字影院只能提供双眼深度暗示。
- **头部运动视差**—观众希望当他们移动头部时会看到一个从略微不同的角度能够观察到的3D场景或对象。多视点或体显示技术可以提供有限的头部运动视差，如只在某个角度范围内的水平视差。
- **双眼竞争/抑制理论**—双眼竞争是一种当不同影像呈现给左、右眼时的视觉感知现象，如双眼分辨率与质量差异。

# 3D感觉：视觉不适

- **适应-会聚冲突**：人类立体视觉中有两种眼球运动机制：**调节（聚焦于何处）**和**会聚（观察于何处）**。调节是当物体离眼睛的距离发生变化时，使眼睛改变光学焦距以保持一幅清晰画面的过程。会聚则是指移动两只眼睛以确保物体的影像刚好落在两个视网膜相对应的点上。
  - 在**真实的3D视觉**中，**调节和辐辏的距离**是一样的。
  - 在平板显示器上，人眼聚焦于平面屏幕上但看到的物体却在屏幕前面。
- **串扰/泄漏**：是由于偏振滤波器（被动型眼镜）的缺陷或同步错误（主动式快门眼镜）导致的。在自动立体显示器中会更严重。会出现虚影和模糊。

# 3D 数字视频显示

- 立体视法

- 被动式(极化)眼镜, 主动式(快门)眼镜

- 自动立体显示

- 各种光学技术, 如: 视差光栅、柱状透镜等

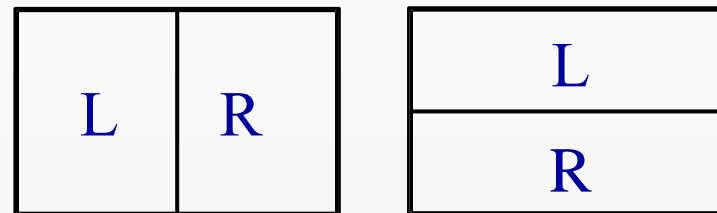




# 3D数字视频表示

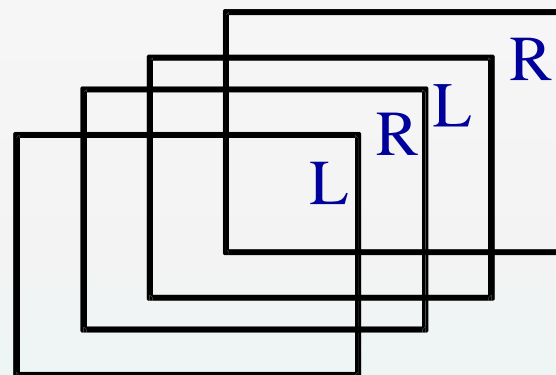
- 立体

- 帧兼容格式
- 全分辨率时序格式



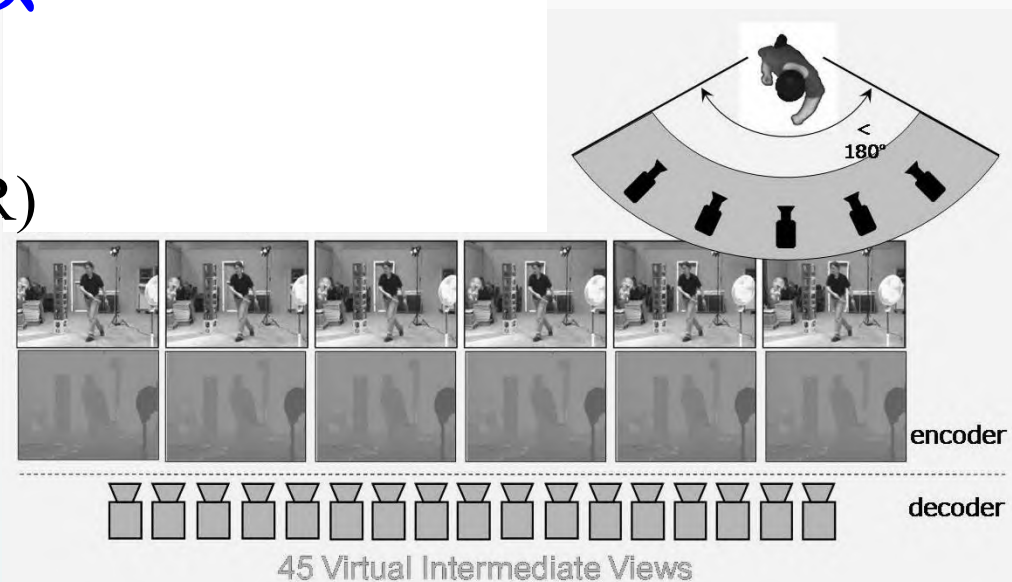
- 多视角

- 直线
- 圆弧
- 阵列

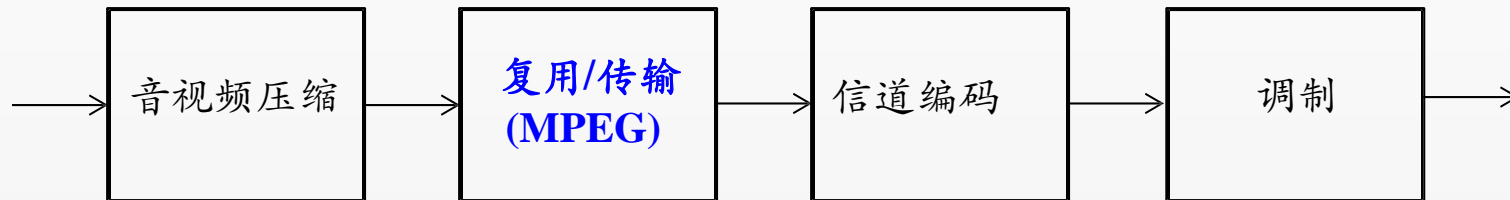


# 自由视角与3D视频表示

- 自由视角2D vs. 3D 视频
- 多视角 + 深度
  - 基于深度图像的渲染 (DIBR)
- 光场显示
  - 超多视角
  - 全息显示



# 数字视频应用：数字电视



- **ATSC (美国)**
  - A/53, 1995: MPEG video, Dolby AC3 audio
- **DVB (欧洲)**
  - DVB-S (1995), DVB-S2 (卫星)
  - DVB-T (1997), DVB-T2 (地面)
  - DVB-C2 (电缆)
- **日本: ISDB**
- **中国: CMMB, DMB**



# 数字视频应用：数字影院

---

- 数字影院格式

- 数字电影联盟(DCI) 兼容格式, 2005
- 支持 2K ( $2048 \times 1080$  or 2.2 M像素), 帧率为 24 或 48 fps; 4K ( $4096 \times 2160$  or 8.8 M像素), 帧率为 24 fps。
- JPEG2000 压缩
- 较强的知识产权保护

- 数字电影放映机

- DLP (Texas 仪器) 技术(2K 与 4K)
- SXRD (索尼) 技术—仅支持 4K

- 3D 数字影院

- RealD, MasterImage, Dolby3D (滤波器阵列)
- XpanD (左右顺序闪烁)
- IMAX (双投影)



# 数字视频应用：视频流

---

- 服务器-用户流：一对一传递协议
  - 实时流协议(RTSP) – IETF 标准, RTP/UDP
  - HTTP 流，基于TCP
    - HTTP 直播流(Apple)
    - 平滑流 (Microsoft)
    - HTTP 动态流(Adobe)
    - DASH (MPEG)
- 组播流：一对多传递协议
- **P2P流**：在P2P覆盖网中，为了降低服务器的负载，在一个直播流任务中，每个点（peer）将会发送数据包给其他点。存在一些协议（追踪器）帮助各个点为某个特定的流找到中继点。

# 数字图像/视频质量：损伤

- 噪声
- 空域降质
  - 模糊
  - 振铃效应
  - 块效应
- 时域降质
  - 跳帧
  - 冻结帧



- 损伤的可见性会随着观察条件的不同而变化，另外，由于空域和时域的掩盖效应，也会随着图像/视频内容的不同而变化。

# 数字图像/视频质量：主观

- **视频质量的主观评估** 具有很大的挑战性，因为：
  - 配置和观察条件，例如：房间的照明情况，显示类型，光亮度，对比度，分辨率，观察距离。
  - 专家（观察者）的年龄和受教育程度
  - 视频内容的选取以及视频的持续时间都会影响最后的评估结果。
- **典型的主观视频质量评估过程** 包括：
  - 选择视频序列用来测试
  - 选择系统用来测试的配置和设置进行评估
  - 选择测试方法(如何将序列呈现给专家以及如何收集他们的观点：DSIS, DSCQS, SSCQE, DSCS)
  - 邀请足够数量的专家(一般建议18岁以上)
  - 对每个测试设置，计算平均的专家意见参数(MOS)

# 数字图像/视频质量：客观

- 按**度量方法**，可分为：**全参考 (FR)**、**部分参考 (RR)** 和**无参考 (NR)**。
- 峰值信噪比 (PSNR)** 是一种FR度量方法，计算的是**最大信号功率与均方误差 (MSE)** 之间比值的对数，即：

$$\text{PSNR} = 10 \log_{10} \left( \frac{255^2}{\text{MSE}} \right)$$

其中测试视频 $\hat{s}[n_1, n_2, k]$  (每帧 $N_1 \times N_2$ 像素，共 $N_3$ 帧) 与相同大小的参考视频 $s[n_1, n_2, k]$ 之间**MSE的计算公式**如下：

$$\text{MSE} = \frac{1}{N_1 N_2 N_3} \sum_{n_1=0}^{N_1} \sum_{n_2=0}^{N_2} \sum_{k=0}^{N_3} (s[n_1, n_2, k] - \hat{s}[n_1, n_2, k])^2$$



## 数字图像/视频质量：客观(cont'd)

- **结构相似性索引(SSIM)** 是一种FR度量方法，旨在测量两个  $N \times N$  亮度块  $x$  和  $y$ （均值为  $\mu_x$  和  $\mu_y$ ，方差为  $\sigma_x^2$  和  $\sigma_y^2$ ）之间结构信息的感观变化，其计算公式如下：

$$\text{SSIM}(x, y) = \frac{(2\mu_x\mu_y + c_1)(2\sigma_{xy} + c_2)}{(\mu_x^2 + \mu_y^2 + c_1)(\sigma_x^2 + \sigma_y^2 + c_2)}$$

其中： $\sigma_{xy}$  为  $x$  和  $y$  之间的协方差， $c_1$  和  $c_2$  为小常数，用来避免除以非常小的数值。

- 其他例子：**PEVQ** 是一种FR度量方法，**VQM** 是一种RR度量方法。
- ITU 视频质量专家组 (VQEG) 对某些度量方法进行了标准化，其中包括：**PEVQ**、**SSIM**、**VQM**。