



交流图片

图像和视频编码课程

2014 年, 第 99-132 页

第4章-数字图片格式和表示

大卫·R·布尔

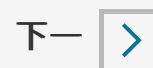
显示更多 ∨

 Outline |  共享  引用<https://doi.org/10.1016/B978-0-12-405906-1.00004-0>

获取权利和内容

摘要

本章介绍了支撑图像和视频编码过程的图像格式、处理技术和评估方法。它首先展示了样品如何在数字图像和视频信号中表示, 强调了颜色和时空表示的重要方面。然后, 它检查样品如何分组形成图片, 以及如何分组图片以形成图片或视频序列。考虑了将光水平映射到人类视觉系统感知到的亮度, 这表明必须如何使用非线性映射(伽马校正)转换来自相机的信号, 以确保最佳地使用可用位。接下来, 它介绍了色空间和色空间转换的区域, 展示了流色和色度信息的分离如何导致整体比特率较低。最后, 它介绍了图像和视频质量评估的重要主题——既作为比较不同压缩系统性能的手段, 也作为在压缩期间控制编码器模式的手段。通篇提供了示例。



关键词

视频格式; 视频扫描; 颜色空间; 相宽比; 图像质量; 伽马校正; 速率-失真优化

自数字视频通信的起源以来, 就一直需要了解人类视觉系统和图像表示之间的联系。本章以 and 第1章、第2, 章, 和第3, 章为基础, 第1章第, 2 and 章第3章, 描述了支撑编码过程的图片格式、处理技术和评估方法; 介绍了使压缩如此有效的处理方法和采样结构。

在第三章数学序的基础上，我们首先展示了样品如何在数字图像和视频信号中表示，强调了颜色和时空表示的重要方面。我们研究如何将样本分组形成图片，以及如何分组形成图片或视频序列。

第4.4节考虑了人类视觉系统感知到的光级到亮度的映射，展示了如何使用非线性映射（伽马校正）来转换来自相机的信号，以确保最佳地使用可用的位。接下来，进一步发展第二章中引入的三色视觉理论，我们在第4.5节中引入了颜色空间面积；这些区域很重要，因为我们只需花费相对较少的努力，就可以创建表示形式，将数字视频的比特率降低50%。最后，我们介绍了图像和视频质量评估的重要主题——既作为比较不同压缩系统性能的一种手段，也作为在压缩过程中控制编码器模式的手段。

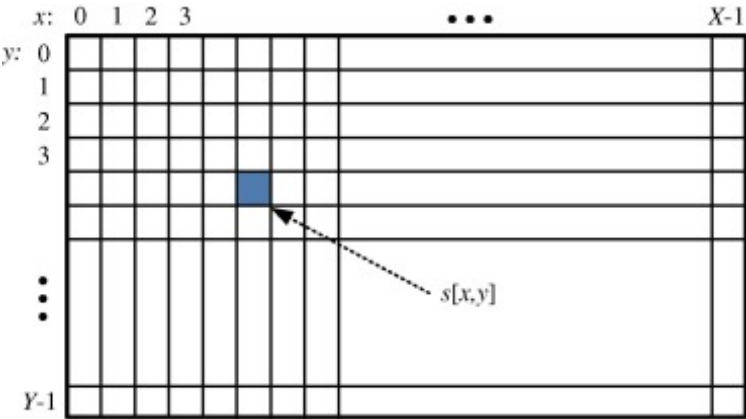
4.1。像素、方块、图片

4.1.1。像素、样品或皮

静止图像或图像是样本值的空间分布，随时间而恒定。这些通常被称为像素或pel。

单色图像

单色图像是具有空间维度（通常称为分辨率）的亮度样本 S 的二维矩阵 $X \times Y$ 。如图4.1所示。如果我们放大实际的图像，如图4.2所示，其中像素分辨率（以它在视网膜上夹角的角度而言）开始在HVS的分辨率范围内，我们开始看到像素化效果。当我们进一步放大时，我们失去了底层图像的结构，像素结构占主导地位。



[下载：下载全尺寸图像](#)

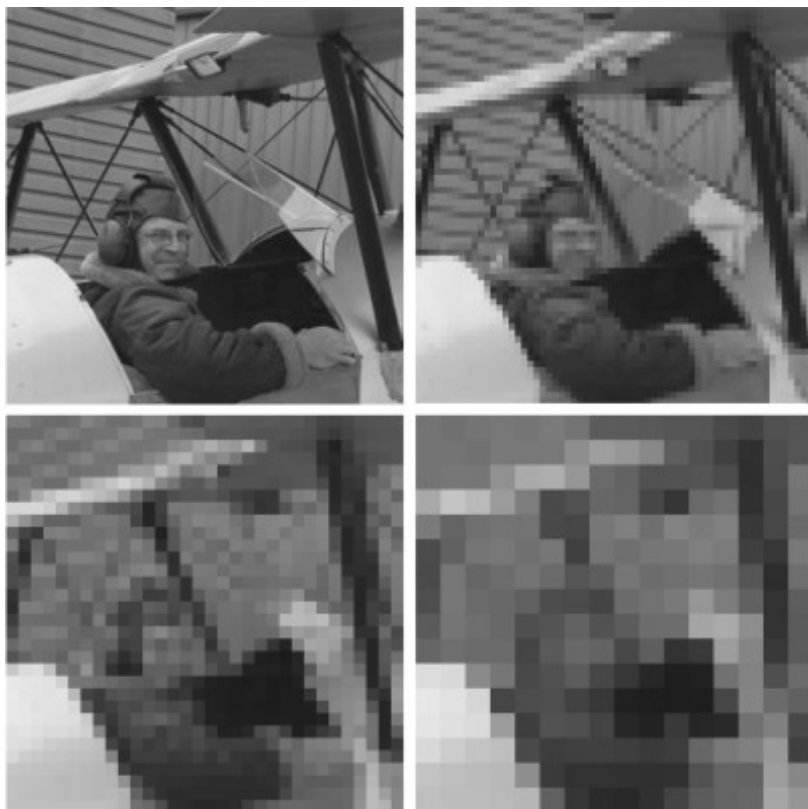
图4.1。图像样本阵列。



[下载：下载全尺寸图像](#)

图4.2。图像样本。

图4.3通过使用四种不同的**抽样率**以略微不同的方式呈现了这一点。如果你离开这个数字，你会看到（大约5米）前两张图片开始看起来很相似。如果你能走得更远（大约10米），所有四张图片的外观都将相似。这强调了这样一个事实，即分辨率不在于水平和垂直样本的数量，而在于每个像素在观众视网膜上所处的角度。



[下载：下载全尺寸图像](#)

图4.3。不同分辨率的像素化。左上角到右下角：256 × 256;64 × 64;32 × 32;16 × 16。

彩色图像

在彩色图片的情况下， s 的每个元素本身就是向量 s ，三维表示给定颜色空间中的强度（例如RGB、HSI或 YC_bC_r ）。

因此，我们在矩阵形式上有：

$$\mathbf{S} = \begin{bmatrix} s[0,0] & s[0,1] & \cdots & s[0,X-1] \\ s[1,0] & s[1,1] & \cdots & s[1,X-1] \\ \vdots & \vdots & & \vdots \\ s[Y-1,0] & s[Y-1,1] & \cdots & s[Y-1,X-1] \end{bmatrix} \quad (4.1)$$

其中（例如RGB信号）：

$$\mathbf{s}[x,y] = [s_R[x,y] \quad s_G[x,y] \quad s_B[x,y]]$$

4.1.2。移动图片

视频信号是随时间变化的空间强度模式。图片以恒定帧率进行抓拍和显示， $1/T$ 在哪里 T 是采样期。为了将**时间工件**（闪烁、抖动等）与可用带宽交换，通常选择24、25或30Hz的帧速率。最近，随着屏幕变大，空间分辨率增加，50和60Hz的帧率更高。新的UHDTV标准Rec.2020规定最大帧率为120赫兹。读者请参阅**第12章**，以更深入地讨论帧率要求。

在我们的离散信号表示方面，我们添加了一个额外的维度（时间或帧索引）， z ，对于与静止图像相关的两个空间维度，因此：

$$\mathbf{S} = \begin{bmatrix} s[0,0,z] & s[0,1,z] & \cdots & s[0,X-1,z] \\ s[1,0,z] & s[1,1,z] & \cdots & s[1,X-1,z] \\ \vdots & \vdots & & \vdots \\ s[Y-1,0,z] & s[Y-1,1,z] & \cdots & s[Y-1,X-1,z] \end{bmatrix} \quad (4.2)$$

其中（再次用于RGB信号）：

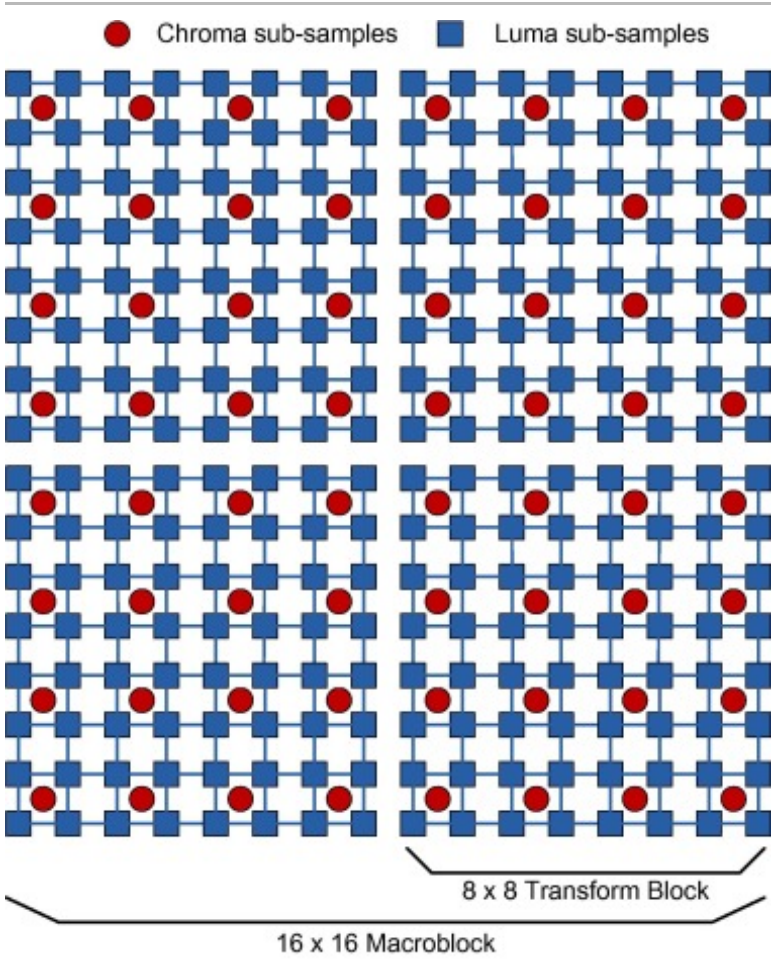
$$\mathbf{s}[x,y,z] = [s_R[x,y,z] \quad s_G[x,y,z] \quad s_B[x,y,z]]$$

4.1.3。编码单元和宏块

正如我们稍后将看到的，视频**压缩算法**很少以图片或像素的比例处理信息。相反，编码单元通常是正方形的像素块。在H.264标准水平和包括H.264的标准中，这采取了一个 16×16 块，由亮度和色度信息组成，称为**宏块**。我们将在**第4.5节**中检查颜色空间和色度子采样的性质。因此，现在让我们简单地检查一下这个宏块的高级组织。

宏观块

典型的宏块结构如图4.4所示。显示的宏块对应于所谓的4:2:0格式（见第4.5节），由一个 16×16 卢马样品阵列和两个子样本 8×8 色度（色差）样品阵列。这种宏块结构在编码时，必须包含重建空间细节所需的所有信息。例如，这可能包括变换系数、运动矢量、量化器信息以及与用于预测目的的进一步块分区相关的其他信息。 16×16 块大小通常是用于运动估计的基本大小；在此范围内，装饰相关变换通常应用于任一 8×8 或者 4×4 水平。



[下载：下载全尺寸图像](#)

图4.4。典型的宏块结构。

编码树单元

HEVC将宏块的大小扩展到 64×64 样品支持更高的空间分辨率，转换大小可达 32×32 。它还在块分区方面提供了更大的灵活性，以支持其各种预测模式。有关此的更多详细信息，请参阅第12章。

4.1.4。图片类型和图片组

框架类型

大多数视频编码格式支持三种类型的图片（或帧1），根据其预测属性进行分类。这些描述如下：

- 内码（I）图片：我的图片（或框架）是编码的，不参考任何其他图片，但可以用作其他图片的参考。
 - 它们提供主码图片，支持随机访问。
 - 它们对错误具有最高的弹性，因为它们不会传播以前图片中的错误。
 - 因为它们不利用时间相关性，因此在所有帧类型中，它们的压缩比最低。
- 预测（P）图片：P图片是相互编码的，即它们包含当前帧之间的压缩差异和基于之前的I或P帧的预测。
 - 像I帧一样，它们可以用作其他帧的参考。
 - 为了解码P帧，变换后的残差必须伴有一组运动矢量，这些矢量提供有关预测参考位置的信息。
 - P图片比我的图片效率更高，因为它们利用了空间和时间相关性。
 - 与I帧不同，它们会传播由于预测编码而产生的错误。
- 双预测（B）图片：B图片包括当前帧之间的压缩差和基于最多两个I、P或B帧的预测。
 - 在H.264/AVC之前的标准中，B图片不被用作预测其他图片的参考。然而，H.264/AVC和HEVC已经概括了它们的定义，在这种情况下，它们可以用作参考，尽管通常以等级的方式（见第12章）。
 - 它们需要伴随的运动矢量。
 - 由于利用时间和空间冗余，它们提供了高效的编码，如果不用作参考，可以更大量量化。
 - 如果不用作其他图片的参考，它们不会传播错误。

一组图片（GOP）

所有视频编码方法和标准都强制要求一些**比特流**语法，以便比特流可解码。这与图片的编码、解码和显示顺序以及不同类型的图片的组合方式有关。[图4.5](#)显示了典型的共和党结构（MPEG-2等标准的典型）。

示例4.1帧内编码与帧间编码——预测的好处

考虑一个 $512 \times 512@30$ fps 彩色视频，每个彩色通道有8位样本。如果压缩后每像素的平均比特数如下，则比较压缩内、压缩间压缩和未压缩的比特率：

- 帧内模式：0.2 bpp（假设这是一个适用于所有颜色通道的平均值）。
- 帧间模式：0.02 bpp（假设GOP长度为100张图片，此值包括运动矢量的任何开销）。

解决方案

(a) 未压缩

未压缩的比特率可以计算为：

$$512 \times 512 \times 30 \times 24 \simeq 189 \text{ Mb/s}$$

(b) 压缩模式内

在模式内，所有帧都独立压缩，以给出以下比特率：

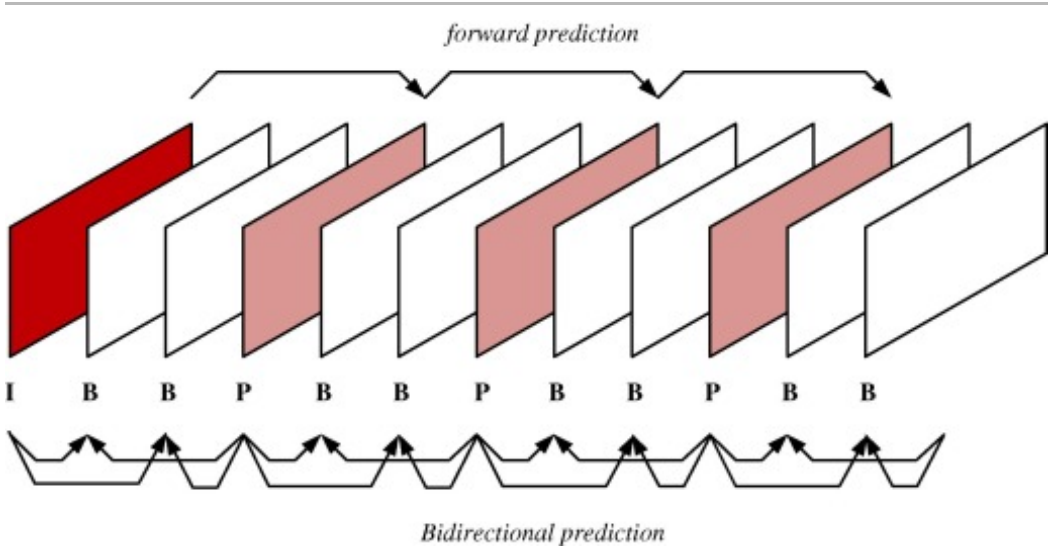
$$512 \times 512 \times 30 \times 0.2 \times 3 \simeq 4.7 \text{ Mb/s}$$

(c) 压缩模式间

在模式间，每100帧中就有1帧是帧内帧，其他99帧是帧间。因此，比特率是：

$$512 \times 512 \times 30 \times \left(\frac{0.6 + 99 \times 0.06}{100} \right) \simeq 514 \text{ kb/s}$$

这表明，通过使用帧间编码来利用时间和空间相关性，我们可以节省大量费用（此处给出的数字在实践中并非不现实）。



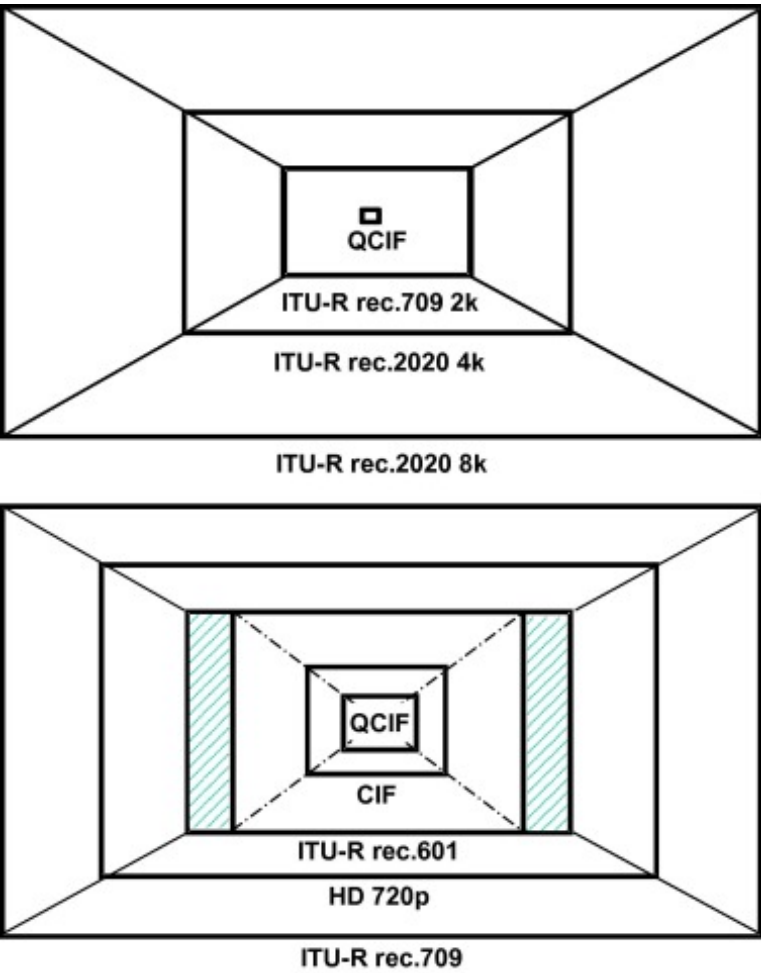
[下载：下载全尺寸图像](#)

图4.5。典型组图片结构及预测模式。

4.2。格式和长宽比

我们结合原始比特率和可用传输带宽的限制，研究了第一章中常用的视频格式的范围。视频行业自开始编码QCIF以来已经取得了长长的路（176 × 1441989 年左右用于低比特率会议应用程序的图像。随着HEVC的推出，我们现在准备能够获取、处理、存储、交付和显示超高清电视格式，最初以3840 × 2160，但很快就升到了8k(7680 × 4320)对于一些应用程序。这些格式的相对大小如图4.6所示，其中绘制的大小归一化为视网膜上单个像素的观看角度。顶部的子图形在显示QCIF

和UHDTV格式的相对尺寸方面尤其令人震惊。



[下载：下载全尺寸图像](#)

图4.6。常用格式的长宽比，按分辨率进行规范化。

4.2.1。宽比

表4.1说明了一些更常见的视频格式长宽比。长宽比的选择一直是许多争论的话题；选择一个最接近视网膜视角比的选择似乎是最明显的选择。这些在时间上约为95°，水平子午线上下约为70°。然而，不难看出这造成了一个问题——它认为纵横比应该是无限的！那么，让我们再尝试一次——凹陷视野呢？我们视网膜的这个大约圆形区域覆盖了我们视野的2°和3°，远离凹的视觉敏锐度迅速下降，通常在凹陷处20/20，在20°的视野角处为20/200。所以因为这大致是圆形的，我们应该假设正方形长宽比最好，还是屏幕应该是圆形的？答案显然介于1:1和无穷之间，也许，我们不需要矩形屏幕！一些人认为，1.61:1的*黄金比例*最美观，尽管原因并不特别清楚。

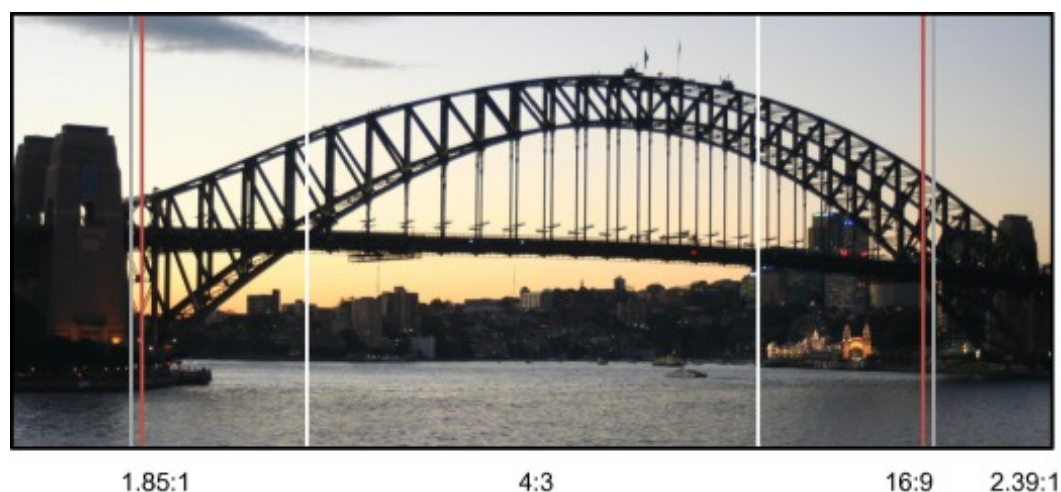
表4.1。电影和电视的长宽比。

格式	宽比
SDTV	1.33 (4:3)
WSTV/HDTV	1.78 (16:9)
宽屏35毫米胶片	1.85
70毫米胶片	2.10
影镜变形35毫米胶片	2.35
变形35毫米电影显微镜 (1970+)	2.39

对于更大更近的屏幕，当然是一个实用的解决方案，那就是创造屏幕充满我们外围视觉的印象，而不是完全这样做。这提供了更好的沉浸在内容中的感觉，同时不需要过宽的屏幕——因此典型的纵横比为1.77-2.39。然而，所有这些都无法解释大屏幕的观看模式的变化，我们表现出更大的跟踪眼睛和头部运动物体的倾向。这就介绍了我们将在[第13章](#)中再次触及的各种新问题。

在实践中，为大多数消费者格式选择的16:9格式是一种妥协，可以优化电影和4:3内容的裁剪。16:9比（1.78）非常接近1.33和2.39的几何平均值，因此所有格式都占据相同的图像圆。归根结底，最佳长宽比将在很大程度上取决于构图以及主体、框架边缘和无趣空间大小之间的关系。摄影师的艺术是以引人注目的方式创作内容，无论帧长宽比如何。

[图4.7](#)给出了场景上叠加的一些常见长宽比示例。当然，偏好也取决于观看条件。在舒适观看的约束下，根据相同的宽度、高度或面积比较4:3和16:9格式，大多数受试者更喜欢提供最大观看面积的格式。



[下载：下载全尺寸图像](#)

图4.7。宽屏格式。

用于传感器或显示器 N_p 像素和长宽比， r ，计算水平和垂直[像素计数](#)很简单， x 和 y 如下：

$$\begin{aligned} X &= \sqrt{rN_p} \\ Y &= \sqrt{\frac{N_p}{r}} \end{aligned} \quad (4.3)$$

视野比

让我们考虑视角比（或视野比）而不是纵横比。很容易显示这是与屏幕大小无关的， s ，但会随观看距离而非线性变化， d 。图4.8中以通常采用的电视和计算机显示器格式16:9为例，但其他长宽比也表现出类似的特征。这个图以屏幕高度的倍数绘制了观看角度比 H 。从这张图表中可以提取一些重要的东西。首先，作为 d 增加，观看角度比 $\theta_w/\theta_h \rightarrow r$ 在哪里 r 是屏幕的长宽比。其次，作为 θ_w/θ_h 减少以下 $3H$ 这个比率在团结方面迅速下降 $d = 0$ 。对于16:9的长宽比，观看距离 $1.5H$ （专为UHDTV指定）对应约1.65的比例，实际上非常接近上述黄金比例！示例4.2更详细地检查了这一点。

示例4.2视野比

导出大小屏幕的视野比表达式 s （对角线尺寸），具有长宽比， r ，从远处看， d 。

解决方案

通常，屏幕的尺寸是根据其对角线尺寸给出的， s 。计算屏幕高度是一件简单的事情 H 宽度 W ，使用方程（4.4）：

$$H = \frac{s}{\sqrt{1+r^2}}; \quad W = \frac{s}{\sqrt{1+1/r^2}} \quad (4.4)$$

纵横比屏幕对应的水平和垂直视角 $r(W \times H)$ 从远处观看 d 然后是：

$$\theta_w = 2 \arctan\left(\frac{W}{2d}\right) \quad (4.5)$$

$$\theta_h = 2 \arctan\left(\frac{H}{2d}\right) \quad (4.6)$$

及其比率，此处定义为视野比率， Ψ 因此：

$$\Psi = \frac{\arctan\left(\frac{W}{2d}\right)}{\arctan\left(\frac{H}{2d}\right)} \quad (4.7)$$

这是为图4.8中不同观看距离的16:9长宽比屏幕绘制的。



[下载：下载全尺寸图像](#)

图4.8。视场随观看距离的变化（此处的宽比=16:9）。

4.2.2。显示不同的格式

多种格式（更不用说三维、二维等）的存在为内容制作人，特别是电影制作人创造了重要的额外工作。电影或旧电视节目以某种方式裁剪或扭曲是相当常见的，总的来说，这对那些觉得自己没有充分利用屏幕的消费者来说很烦人！

例如，在16:9屏幕上显示2.39:1格式会产生信箱的外观。这可以通过扭曲原始安装或缩放来避免，但两者都有问题。另一方面，例如16:9屏幕上显示的4:3格式，会产生柱状盒的外观。同样，这可以拉伸以产生短而肥胖的受试者，也可以放大以失去边缘内容。

平扫和活动格式说明

主动格式描述（AFD）是数字广播中使用的一种机制，以避免上述问题。这是一种对图片的主动内容进行信号传递的方式，以便能够对接收端的图片进行智能处理，以便以最佳的方式将格式映射到显示器。例如，考虑到图4.7，如果我们想在16:9显示器上观看2.39:1版本，蓝色框（代表16:9图片）不需要停留在2.39:1图片的中心。相反，它可以向左或向右移动，以最适合较小屏幕的动作。

该功能在共和党的基础上作为MPEG-2的非规范功能引入，并得到更新标准的支持（例如通过H.264/AVC中的SEI数据）。AFD代码包括有关活动视频位置和必须显示的任何保护区的信息。该区域以外的图片边缘可以丢弃，而不会对观看体验产生重大影响。

4.3。图片扫描

4.3.1。交错与渐进扫描

交错扫描试图通过从两个连续字段构建每帧来权衡带宽、闪烁和分辨率。字段在不同时间采样，因此连续的行属于备用字段；美国采用60Hz格式，而欧洲采用50Hz格式。

早期的模拟电视系统需要将带宽降至最低，同时保持无闪烁的观看。交错是作为高时间更新率（50或60赫兹的现场速率）和较低的有效带宽（由于25或30赫兹帧率）之间的折衷方案，前者减少了大多数内容的闪烁。荧光粉在早期阴极射线管显示器中的持久性，加上人类视觉的持久性，这意味着，只要扫描图像的垂直维度足够快，HVS就不会看到交替线的效果，并体验到连续运动。**图4.9**说明了交错扫描和渐进扫描之间的区别。在现代数码相机和显示器中，如**图4.9**所示，扫描不再通过偏转电子束进行。然而，行扫描和字段扫描模式与图所示相同。

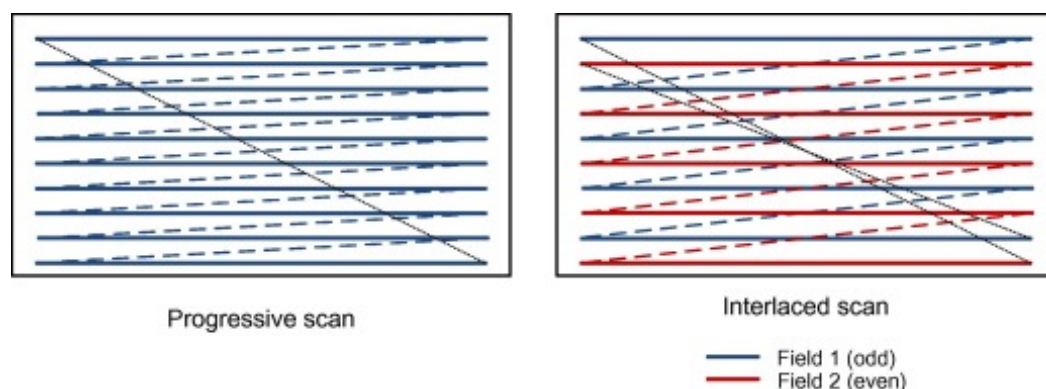
符号

我们将在这里采用的符号是欧洲广播联盟（EBU）和SMPTE使用的符号：

$$Y\pi/f$$

在哪里 Y = the number of vertical lines in the frame; π = p（渐进式）或i（交错）；以及 f = frame rate。

例如：1080p/50表示具有空间分辨率的高清渐进格式1920 × 1080和50 fps的时间采样。同样，1080i/30代表相同的空间分辨率，但这次以30和60帧/秒交错。

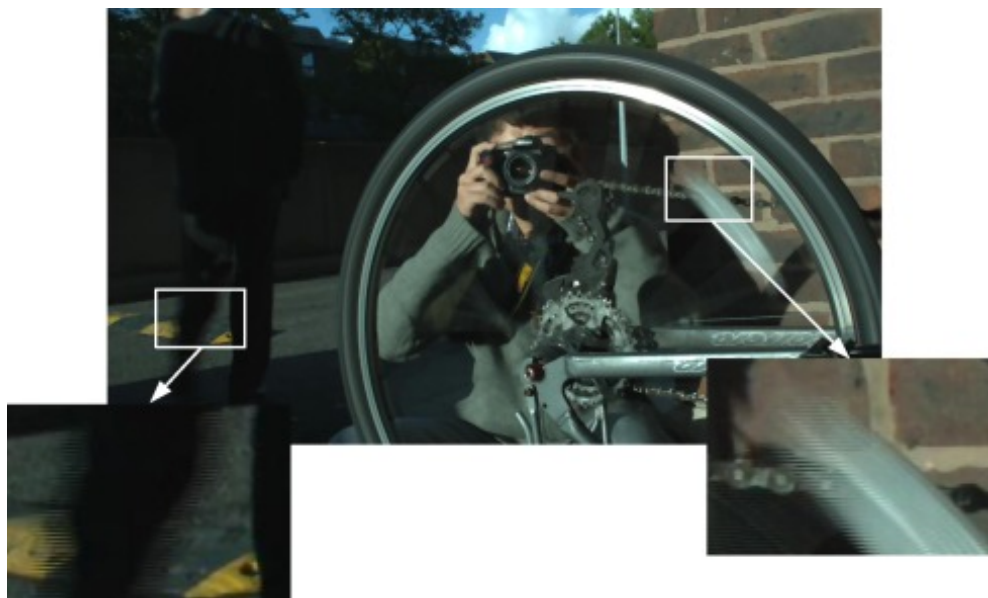


[下载：下载全尺寸图像](#)

图4.9。隔行帧扫描与渐进帧扫描。

交错引起的问题

如**图4.10**所示，在快速运动的情况下，交错的影响变得很明显。由于两个组成字段之间的时间偏移，插图显示了这种格式的**梳状工件**。交错还引入了第二个别名问题，通常称为**推特**。当信息以类似于现场扫描速率的速度垂直变化时，就会发生这种情况。当内容的垂直分辨率与格式分辨率相似时，这种情况变得尤为普遍。例如，穿着细水平条纹的人的衣服会在推特上出现。过滤经常用于减少这种效果。参考文献中可以观察到推特的一个例子。[\[2\]](#)。



[下载：下载全尺寸图像](#)

图4.10。隔行扫描与隔行不良的效果示例。

最后值得注意的是，在数字视频压缩的情况下，与模拟系统相关的交错的带宽优势会降低，特别是当具有两个字段的帧被处理为一帧时。MPEG-2和后续标准引入了自适应现场帧模式，试图改善这种情况。

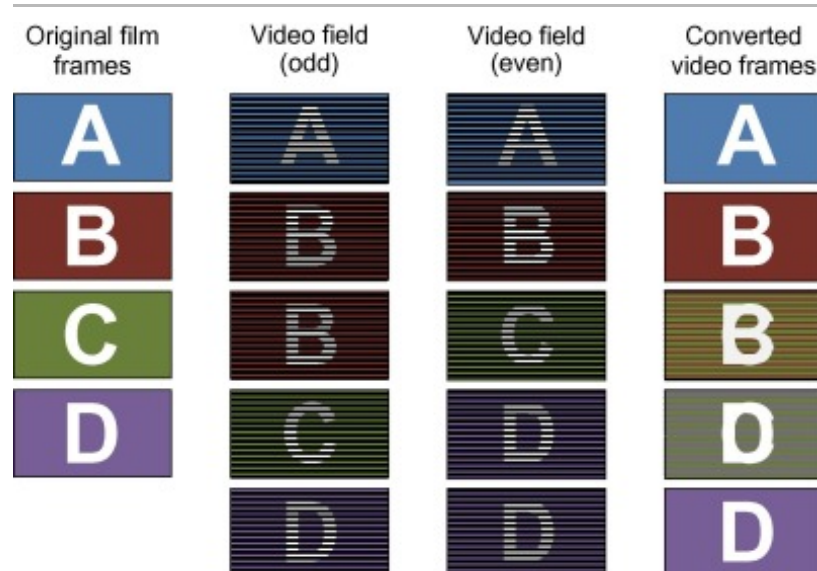
如果接收器和显示器采用良好的去插错方法，与隔行相关的工件可以最小化[5]。作为组合过程的一部分，这两个字段采用局部自适应过滤。

4.3.2。标准转换

3:2 下拉

自公共电视广播开始以来，需要在电影和电视格式之间转换一直是一项要求。电影和数字电影内容传统上使用渐进扫描以24帧/秒的速度捕获。这需要在美国根据60赫兹交错扫描转换为29.97 fps，在欧洲根据50赫兹转换为25 fps。对于25 fps的情况，胶片简单地加速25/24倍是正常的。这会产生一些视听变化，但这些变化通常不明显。然而，为了以29.97 fps精确呈现运动，必须使用远程电影设备来重新采样24 fps信号。通常使用一种被称为3:2下拉的技术。

3:2下拉使用算法4.1中描述的方法，如图4.11所示。



[下载：下载全尺寸图像](#)

图4.11：3:2下拉示例。

一些流媒体网站，如Netflix，对其内容有严格要求，并要求反向电视剪辑操作来检测和删除电视剪辑视频源中的3:2下拉，从而重建原始每秒24帧的电影格式。这提高了与非隔行显示器兼容性的质量，并消除了压缩前的一些冗余数据。

算法4.1 3:2下拉。

1. Play the film at 23.976 frames/s;
2. Split each cinema frame into video fields. At 23.976 frames/s, there are four frames of film for every five frames of 60 Hz video: i.e. $23.976/29.97 = 4/5$;
3. Convert these four frames to five by alternately placing the first film frame across two fields, the next across three, and so on.

[下载：下载全尺寸图像](#)

运动补偿标准转换

更通用的标准或格式转换由基于相位相关的运动补偿方法提供（见第6章）。这些是转换内容所必需的，以便在交错和渐进格式之间、高清和SD和其他格式之间转换内容，以及不同帧速率和颜色空间的格式之间。相位相关性是一种基于傅里叶的方法，它利用时域（或空间）位移和频域相移之间的对偶性。与传统的块匹配方法相比，它对噪声和亮度变化具有更好的免疫力。子像素精度可以很容易地实现，就像以目标格式生成新帧一样。

4.4。伽马校正

伽马校正[5]用于纠正相机捕获内容的方式、显示器显示内容的方式和视觉系统处理光线的方式之间的差异。我们的眼睛对光的反应不像相机捕捉光那样。基本来说，如果光子数达到数字图像传

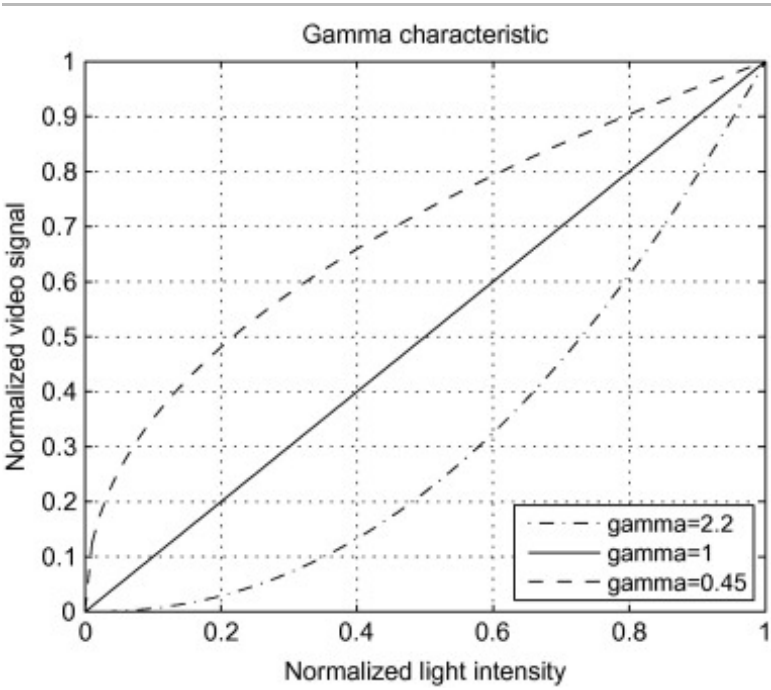
传感器的两倍，那么输出电压将是数字图像传感器的两倍。然而，对于基于CRT的旧相机来说，情况并非如此，光强和输出电压之间存在高度非线性的关系。HVS对亮度水平也有非常非线性的反应，对暗区域的微小变化更敏感，对光区域的敏感度低得多（回顾第二章中的韦伯定律）。如果不进行校正，为了避免由于编码级别之间可感知的跳跃而产生的带状效应，大约需要11位。然而，在大多数标尺的高端，编码将非常冗余，因为编码级别是难以察觉的。通过伽马预校正，我们只需8位即可获得可接受的表示形式。

光电人类视觉特征可以通过幂律函数近似。使用此类功能校正相机输出的信号，确保不会将太多比特分配给人类歧视程度较低的明亮区域，而更多分配给更黑暗区域，而这些区域则更具有辨别。因此，摄像机通常执行照明强度与输出值的非线性映射，为较暗的值提供更均匀的感知尺度和更精细的增量。这被称为设备的伽马特性，如传输函数所述：

$$V = c_1 \Phi^\gamma + c_2$$

(4.8)

在哪里 Φ 是发光通量（归一化）和 V 是信号的值（电压）， c_1 是相机灵敏度和 c_2 是一个偏移值。图4.12显示了三个伽马值的转移特征图， $\gamma = \{0.45, 1, 2.2\}$ 。这些价值是现代电视系统中使用的典型值。图4.13显示了不同伽马水平图像的一些更正示例。



[下载：下载全尺寸图像](#)

图4.12。伽马曲线 $\gamma = \{0.45, 1, 2.2\}$ 。



[下载：下载全尺寸图像](#)

图4.13。伽马校正的例子。

应当指出，Rec.709等现代格式使用从亮度到输出的分段传输特性。这在较低的区域是线性的，然后在范围的其余部分更改幂律。对于Rec.709，这是：

$$V = \begin{cases} 4.5\Phi & \Phi < 0.018 \\ 1.099\Phi^{0.45} - 0.099 & \Phi \geq 0.018 \end{cases} \quad (4.9)$$

虽然这是一个有点混乱的话题，但人们越来越接受将伽马校正的亮度信号称为亮度信号。

4.5。色彩空间和色彩转换

4.5.1。颜色描述和HVS

三色度理论

正如我们在第二章中所研究的，视网膜中的锥细胞是正常光条件下的主要光感受器，其灵敏度在短波长（S：420-440纳米）、中波长（M：530-540纳米）和长波长（L：560-580纳米）附近达到峰值。（低光传感器（杆）峰值约为490-495纳米。）人类的颜色感觉是400-700纳米范围内的电磁辐射的结果，它发生在视网膜上。

引入彩色电视的挑战之一是在向屏幕提供逼真的颜色与节省带宽之间权衡。简单地说，彩色信息的传输（假设使用三个彩色信道（RGB））占用的带宽是单色传输的三倍。模拟电视的解决方案（灵感来自三色视觉理论（见第2章））将颜色信息与亮度信息分开编码，并显著降低分辨率，以节省带宽。这样做的好处是，亮度信号向后兼容旧的单色电视机，而对于彩色电视机，高分辨率亮度和低分辨率彩色信号在HVS中结合，产生感知高分辨率彩色图像。自模拟彩色电视推出以来，随着技术的进步，人们投入了大量精力来开发适合现代数字编码、格式和显示器的新彩色空间表示形式。其中一些将在下文中进行审查。有关更多信息，请参阅参考文献[1]和Poynton[3]、[5]对颜色空间的很好介绍。

大部分比色法是基于颜色的三刺激理论——即任何颜色都可以由三个原色组成，只要它们是正交的，但也基于反映颜色的线性和加法规则的格拉斯曼定律。当两种刺激的强度由同一因素增加或减少时，任何两种刺激之间的颜色匹配是恒定的。例如，如果我们有两个来源， C_1 和 C_2 ，则在以下内容：

$$\begin{aligned} C_1 &= R_{C1}(R) + G_{C1}(G) + B_{C1}(B) \\ C_2 &= R_{C2}(R) + G_{C2}(G) + B_{C2}(B) \quad \text{and} \\ C_3 &= C_1 + C_2 \end{aligned} \tag{4.10}$$

然后：

$$\begin{aligned} C_3 &= [R_{C1} + R_{C2}](R) + [G_{C1} + G_{C2}](G) \\ &+ [B_{C1} + B_{C2}](B) \end{aligned} \tag{4.11}$$

颜色空间

我们在第二章中看到，CIE根据三色表示定义了颜色空间的坐标，该表示描述了人类观察者经历的一组颜色感觉。其中第一个，CIE 1931（开发于1931年），基于标准观察者的概念，将大量（正常）受试者的观测结果与受控刺激相匹配。这些刺激是基于初选：

$$[R_0 \ G_0 \ B_0] = [700 \ 546.1 \ 435.8] \text{ nm}$$

色彩空间变换

根据比色法的基本原则，可以推导出RGB刺激值的各种其他线性组合，这些组合可能比原始值提供更好的属性。CIE引入了另一种参考系统（或颜色空间），称为XYZ，这可以为表示三刺激值提供更灵活的基础。颜色空间根据三色加色模型将一系列物理产生的颜色（例如混合光）映射到这些颜色感觉的客观描述，尽管通常不是LMS空间。CIE XYZ可以代表普通人可以体验的所有颜色感觉，是 $[R_0, G_0, B_0]$ 空间。CIE将这种转换定义为：

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \begin{bmatrix} 2.365 & -0.515 & 0.005 \\ -0.897 & 1.426 & -0.014 \\ -0.468 & 0.089 & 1.009 \end{bmatrix} \begin{bmatrix} R_0 \\ G_0 \\ B_0 \end{bmatrix} \tag{4.12}$$

CIE定义XYZ以至于Y分量对应于亮度信号，归一化为等能量白色。此外，与一些RGB系统不同，RGB系统的贡献必须是负的才能匹配颜色，XYZ系统是完全积极的。然而，这个问题在于它导致XYZ值是复数，即它们不能通过实际刺激实现。

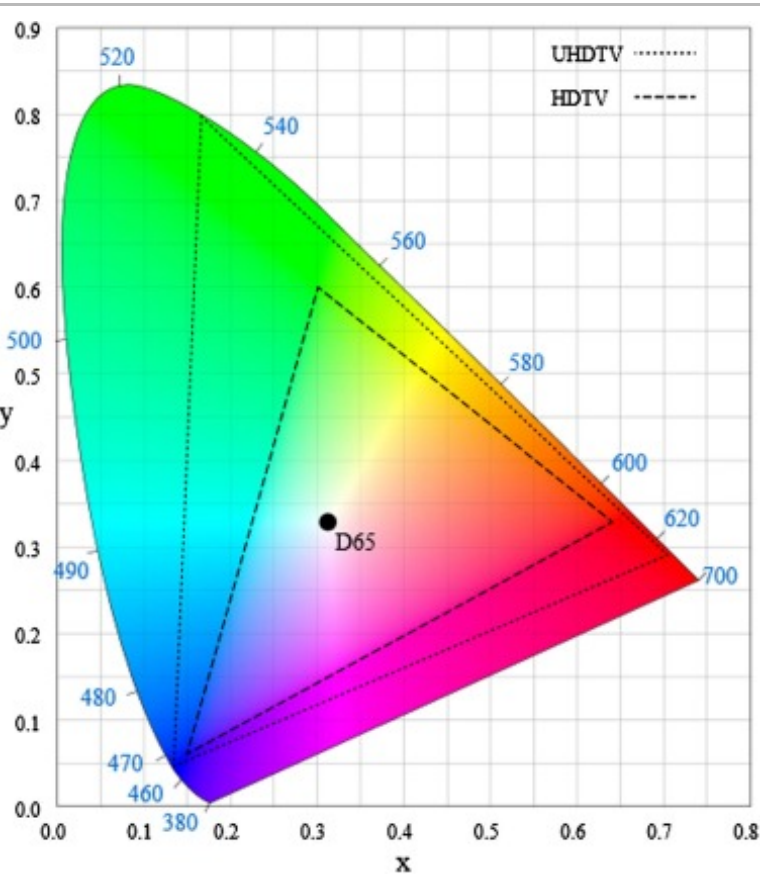
色度值也可以在XYZ系统如下：

$$\begin{aligned}x &= \frac{X}{X+Y+Z} \\ y &= \frac{Y}{X+Y+Z} \\ z &= \frac{Z}{X+Y+Z}\end{aligned}\tag{4.13}$$

大号 x 值对应红色或橙色， y 值对应绿色、蓝绿色或黄绿色，大 z 对应于蓝色、紫色或紫色的大色调。因为 $x + y + z = 1$ ，实际上只需要其中两个值，所以通常这表示为 Yxy 系统。

色度图

色度坐标可以在二维色度图上表示，如图4.14所示。这张图表显示了标准观察者可以感知到的各种色调(x,y)对并指示主要单频颜色的光谱波长。图上显示的D65点是白点，在这种情况下，它对应于6500K的颜色温度（代表平均日照时间）。



下载：下载全尺寸图像

图4.14。CIE 1931 色度图。

公有领域图像来
自：http://commons.wikimedia.org/wiki/File:CIExy1931_Rec_2020_and_Rec_709.svg。

CIE XYZ或CIE 1931是一个标准参考，它被用作定义大多数其他颜色空间的基础。这基于两个CIE参数映射了人类的颜色感知 x 和 y 。尽管在1960年和1976年进行了修订，但CIE 1931仍然是最常用的参考文献。

如果我们再次考虑图4.14中的CIE 1931颜色空间，这也显示了两个三角形，它们对应于Rec.709（HDTV）和Rec.2020（UHDTV）颜色空间表示的颜色。有趣的是，后者可以代表许多无法用

Rec.709显示的颜色。Rec.2020初选对应630纳米（R）、532纳米（G）和467纳米（B）。Rec.2020空间覆盖了CIE 1931空间的约76%，而Rec.709空间仅覆盖了36%，从而更好地覆盖了更饱和的颜色。

模拟电视的色彩空间

几乎在所有情况下，一种格式 YC_1C_2 用于模拟和数字格式，其中 Y 表示信号亮度。这样做有两个原因：

- 与单色显示器兼容。
- 色度信号可能带宽减少，不会失去感知质量。

例如，在美国彩色电视的NTSC系统中，初选是根据CIE定义的 xyz 坐标如下：

$$\mathbf{m} = \begin{matrix} & x & y & z \\ R: & 0.67 & 0.33 & 0.00 \\ G: & 0.21 & 0.71 & 0.08 \\ B: & 0.14 & 0.08 & 0.78 \end{matrix} \quad (4.14)$$

电视传输格式中使用的实际值是伽马校正后的值。因此，经过各种归一化和旋转[1]后，由此产生的颜色空间，定义为 YIQ ，由以下变换给出（其中 R', G', B' 表示伽马校正相机输出信号）：

$$\begin{bmatrix} Y \\ I \\ Q \end{bmatrix} = \begin{bmatrix} 0.299 & 0.587 & 0.114 \\ 0.596 & -0.274 & -0.322 \\ 0.211 & -0.523 & 0.311 \end{bmatrix} \begin{bmatrix} R' \\ G' \\ B' \end{bmatrix} \quad (4.15)$$

欧洲PAL系统的类似映射由以下方式给出：

$$\begin{bmatrix} Y \\ U_t \\ V_t \end{bmatrix} = \begin{bmatrix} 0.299 & 0.587 & 0.114 \\ -0.147 & -0.289 & -0.436 \\ 0.615 & -0.515 & -0.100 \end{bmatrix} \begin{bmatrix} R' \\ G' \\ B' \end{bmatrix} \quad (4.16)$$

数字格式的色彩空间

随着数字格式的引入，从SDTV（Rec.601）开始，需要调整颜色空间转换，以应对这些数字格式和有限的字长。这些组件再次根据伽马校正相机输出进行引用，并量化为8位。ITU-R Rec.601系统的表示形式称为 YC_bC_r 并通过以下从0到255范围内的修正8位 RGB 信号的转换给出。Studio R'G'B'信号在17到235之间，需要修改转换矩阵：

$$\begin{bmatrix} Y \\ C_b \\ C_r \end{bmatrix} = \begin{bmatrix} 0.257 & 0.504 & 0.098 \\ -0.148 & -0.291 & 0.439 \\ 0.439 & -0.368 & -0.071 \end{bmatrix} \begin{bmatrix} R' \\ G' \\ B' \end{bmatrix} + \begin{bmatrix} 16 \\ 128 \\ 128 \end{bmatrix} \quad (4.17)$$

为了提供编码空间，不使用256个值的完整范围。相反，使用了220个级别，从16到235个级别不等。应当指出，一些作者在 Y （luma）上面的组件，以区别于亮度信号，因为它是在基于 γ 可识别的非线性编码后产生的 RGB 初选。 C_b 和 C_r 是蓝差和红差色度成分。

同样，用于高清电视格式的ITU-R Rec.709系统由以下方式提供：

$$\begin{bmatrix} Y \\ C_b \\ C_r \end{bmatrix} = \begin{bmatrix} 0.183 & 0.614 & 0.062 \\ -0.101 & -0.338 & 0.439 \\ 0.439 & -0.399 & -0.040 \end{bmatrix} \begin{bmatrix} R' \\ G' \\ B' \end{bmatrix} + \begin{bmatrix} 16 \\ 128 \\ 128 \end{bmatrix} \quad (4.18)$$

以图4.15所示的塔拉海滩图像为例。这显示了 Y, C_b ，以及 C_r 组件。从中可以观察到几件事——例如，白云在两个色度成分中都表示为中间值，蓝天在蓝色通道中很强；树的沙子和棕色方面在红色通道中表现出很强。此外，相对于流光通道，这两种色度成分在人眼中似乎模糊不清，这提供了一些证据，表明我们可以在不失去整体感知细节的情况下对其进行下采样。



[下载](#): [下载全尺寸图像](#)

图4.15。使用 YC_bC_r 分解。左上角：原始。对：卢马频道。左下角： C_b 频道。对： C_r 频道。

4.5.2。子采样颜色空间

大多数捕获设备产生RGB输出，每个组件都需要高精度和高分辨率才能提供感知到的高质量图像。如上所述，这种格式很少用于传输，因为它需要单色信道的三倍带宽。因此，正如我们所看到的，颜色空间变换被用来产生亮度信号 Y 和两个色差信号， C_b 和 C_r 。这样做的动机是：

1. 色度子采样：正如我们在第二章中所讨论的那样，[人类视觉系统](#)对颜色信息的反应相对较差。我们彩色视觉的有效分辨率约为亮度信息的一半。因此，将亮度和色度成分分离意味着我们可以在不影响亮度分辨率或感知图像质量的情况下将色度进行2倍的次采样。这提供了即

时的压缩优势，因为我们在开始压缩之前将比特率降低50%！

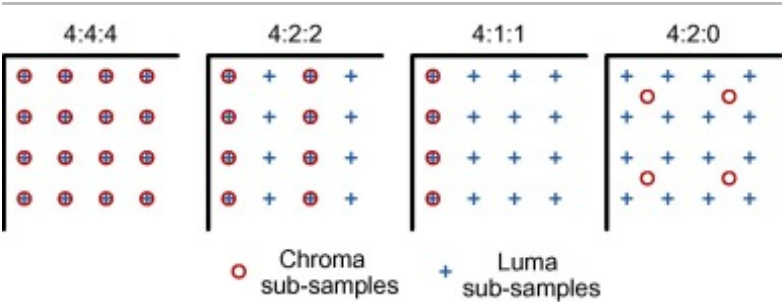
2. 亮度处理：除了与单色显示器兼容外，在压缩期间，独立亮度信号的可用性也很有用。例如，我们通常只对亮度分量进行运动估计，并使用生成的矢量来补偿亮度和色度分量。我们还经常只根据亮度组件进行客观的质量评估。

色度子采样

人眼对色度空间细节的反应相对较差。因此，色度组件通常进行子采样（通常为2倍），以减少带宽要求。最常用的子采样模式如图4.16所示。这显示了一系列常见子采样格式的亮度和色度样本之间的关系。这些是：

- 4:4:4——这是所有像素位置都有亮度和色度样本的全分辨率图像。当需要最高质量的工作室作品时，它会使用。
- 4:2:2——这是工作室制作的常规格式。大多数专业相机以4:2:2格式输出视频。它提供水平而不是垂直的子采样。这在交错扫描的情况下是有利的，因为它消除了颜色退化的问题。
- 4:2:0——该系统是广播交付、互联网流媒体和消费设备最常见的格式。它在水平和垂直方向上对色度信号进行2倍的子采样，最大限度地利用感知色冗余。
- 4:1:1——该系统用于DV格式相机。它水平采样4倍，但不是垂直采样。质量不如4:2:0格式，所以现在很少使用。
- 4:0:0——图中没有显示，但仅表示亮度或单色图像的简单情况。

问题在于，色度信号是如何下采样的？答案是使用线性相位FIR滤波器进行预处理，以避免混叠。同样，对图像进行采样，使用适当的线性相位FIR插值滤波器在图像上进行重建。应当指出，子采样过滤器将决定色度样品相对于亮度样品的地点。



[下载：下载全尺寸图像](#)

图4.16。常见的色度子采样格式。

符号4: X: Y。这种符号让大多数人感到困惑，并曾多次尝试拟合解释其含义的规则，但在某些情况下，所有这些规则都失败了。原因是，从历史上看，4:X:Y格式是在PAL和NTSC系统考虑组件视频时设计的，当时正在考虑子载波锁定采样。第一个数字表示卢马水平采样参考值为

$3\frac{3}{8}$ MHz。第二个数字 X 表示色度信息相对于第一个数字的水平采样因子。第三个数字 Y 与 X 相似，但呈垂直方向。

无需花太多思索就能意识到4:2:0和4:1:1格式不符合这个定义。因此，人们开始想出其他适合的规则。笔者的观点是，我们只用它们作为相关格式的标签。然而，如果您需要一个规则，那么最好的规则如下（不是很优雅，但它适用于大多数格式）：

1. 假设 $a4 \times 2$ 像素块，顶行有4个亮度样本。
2. X 是顶行的色度样本数。
3. Y 是第二行中额外的色度样本数量。

当然，一个问题是，在一些4:2:0格式（如图4.16所示）下，色度样本实际上与亮度样本并不共同选址。

示例4.3 ITU-R Rec.601中的色彩空间转换

考虑 YC_bC_r 与ITU-R Rec.601一起使用的格式。为以下情况计算组件值：

(a) $R'G'B' = [0 \ 0 \ 0]$ 。

(b) $R'G'B' = [220 \ 0 \ 0]$ 。

(c) $R'G'B' = [0 \ 220 \ 0]$ 。

解决方案

使用方程 (4.17)，我们可以计算这些值如下：

(a) $YC_bC_r = [16 \ 128 \ 128]$

在亮度和中心色度和等色度水平上，基底黑电平的信号：

(b) $YC_bC_r = [73 \ 95 \ 224]$

红色色差通道中亮度和蓝色通道较高的信号：

(c) $YC_bC_r = [127 \ 64 \ 47]$

注意这个和 (b) 之间的区别。绿色通道的重量更大，因为它对我们的亮度感贡献更大。

示例4.4色度子采样和压缩

如果使用ITU-R.601表示持续时间为120分钟的彩色电影（ $720 \times 576@30 \text{ fps}@8$ 位，4:2:0格式）：

(a) 存储整部电影需要多少硬盘容量？

(b) 如果电影是MPEG-2编码，压缩比 $CR = 50:1$ ，并通过卫星链路传输，开销50%的信道编码，视频信号所需的总比特率是多少？

解决方案

(a) 比特率由以下方式给出：

$$720 \times 576 \times 30 \times 8 \times \frac{3}{2} = 149,299,200 \text{ b/s} = 149 \text{ Mb/s}$$

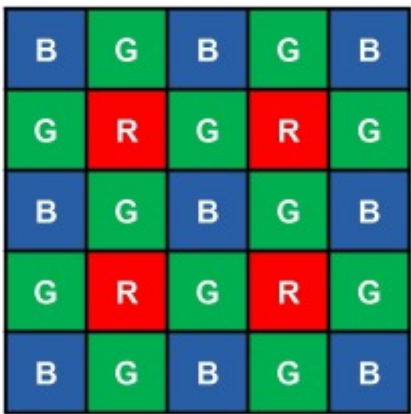
对于一部120分钟的电影，所需的总存储空间是：

$$120 \times 60 \times 149,299,200 = 1,074,954,240,000 \text{ bits} = 134 \text{ GB}$$

(b) 压缩比为50:1，比特率约为3 Mbps。编码开销为50%，这意味着总比特率为4.5 Mbps。

4.5.3。拜耳滤波

专业相机通常有三个CCD或CMOS传感器，每个颜色通道一个。另一种方法是将单个传感器与彩色滤波器阵列结合使用。彩色滤波器阵列（通常称为拜耳滤波器或拜耳马赛克）是一种应用于单个图像传感器的彩色滤波器模式，以使其能够产生R、G和B样品。这种方法用于大多数单片机数字图像传感器，并越来越多地用于专业设备。[拜耳模式阵列](#)[\[4\]](#)如[图4.17](#)所示。一半的滤芯是绿色的，其余的分为蓝色和红色。这近似于人类的光视觉，其中M和L锥体结合在绿色[光谱区域](#)产生偏置。



[下载：下载全尺寸图像](#)

图4.17。拜耳滤波器马赛克。

显然，与三或四传感器阵列相比，拜耳滤波阵列对彩色信号的有效传感器分辨率有所下降，并且降解程度在很大程度上受到所使用的脱色算法的影响。根据[经验](#)，根据实际分辨率的测试，你可以大致将数组中的像素数除以2，这样一千万像素数组在解除后，实际上就变成了一个500万像素数组。

解冻算法有很多类，这里就不详细考虑了。这些范围从相对简单的[双线性插值](#)到双立方体或样条插值。然而，这些方法容易[出现伪影](#)，更复杂的方法，例如基于利用本地像素统计或梯度估计的方法，可以提供改进。还提出了基于超分辨率技术的更先进的方法来解决混叠问题。

4.6。测量和比较图像质量

我们在这里介绍一些图像和视频质量评估的基本主题。Winkler [6] 提供了视频质量指标和主观评估的绝佳介绍。我们还在第10章中更详细地讨论了这些主题。

4.6.1。压缩比和比特率

压缩比是衡量压缩性能的一个非常基本的指标。这简单地定义为：

$$C_1 = \frac{\text{No. bits in original video}}{\text{No. bits in compressed video}} \quad (\text{unitless}) \quad (4.19)$$

这提供了压缩性能的绝对比例，只有在类似内容、类似大小的图像和类似目标质量的情况下，才能用作比较器——否则不应该使用。一个常用的替代方案是归一化比特率，用于静止图像：

$$C_2 = \frac{\text{No. bits in compressed video}}{\text{No. pixels in original video}} \quad (\text{bits per pixel}) \quad (4.20)$$

这同样会因内容类型和图像分辨率而异，但也因原始信号字长而异。它只是一个费率指标，不提供质量评估。视频常用的一个参数是实际比特率：

$$C_3 = \frac{\text{No. bits in compressed video}}{\text{No. frames in original video}} \times \text{frame rate} \quad (\text{bits per second}) \quad (4.21)$$

同样，这没有提供压缩信号质量的信息，但它确实为我们提供了以给定方式编码的信号占用多大信道容量的绝对值。

为了使所有这些措施都有用，它们需要与相关的图像或视频质量测量相结合。这些将在下面讨论。

4.6.2。客观失真和质量指标

平均平方误差（MSE）

考虑一个图像 $S = s[x, y] : x = 0, \dots, X-1; y = 0, \dots, Y-1$ 及其重建版本 \tilde{S} 。如果图像的面积（或等价的单色图像，其中的样本数）是 $A = XY$ 然后MSE由以下方式给出：

$$\begin{aligned} \text{MSE} &= \frac{1}{A} \left(\sum_{x=0}^{X-1} \sum_{y=0}^{Y-1} (s[x, y] - \tilde{s}[x, y])^2 \right) \\ &= \frac{1}{A} \left(\sum_{x=0}^{X-1} \sum_{y=0}^{Y-1} (e[x, y])^2 \right) \end{aligned} \quad 4.22$$

信噪比在许多通信应用中也很常见：

$$\text{SNR} = 10 \cdot \log \left(\frac{E\{(s[x, y] - \mu_s)^2\}}{E\{(s[x, y] - \tilde{s}[x, y])^2\}} \right) = 20 \cdot \log \left(\frac{\sigma_s}{\sigma_e} \right) \text{ dB} \quad (4.23)$$

如果我们将MSE计算扩展到视频序列 K 帧，然后MSE由以下方式给出：

$$\text{MSE} = \frac{1}{KA} \left(\sum_{z=0}^{K-1} \sum_{x=0}^{X-1} \sum_{y=0}^{Y-1} (s[x, y, z] - \tilde{s}[x, y, z])^2 \right) \quad (4.24)$$

峰值信噪比 (PSNR)

基于平均平方误差的度量通常用作客观的失真度量，主要是因为它们易于计算。MSE较高的图像通常比MSE低的图像表达更明显的失真。然而，在图像和视频压缩中，使用峰值信号噪声比 (PSNR) 而不是MSE来描述重建的图像质量是常见的做法。如果图像与不同的动态范围进行比较，并且出于多种不同原因使用它，PSNR将特别有用：

1. 如果信号样本的字长发生变化，MSE值将具有不同的含义。
2. 与许多自然信号不同，图像或视频帧的平均数通常不是零，实际上会因帧而异。
3. PSNR规范了MSE相对于峰值信号值而不是信号方差，从而可以直接比较来自不同编解码器或系统的结果。
4. PSNR永远不能小于零。

我们图像S的PSNR由以下方式给出：

$$\text{PSNR} = 10 \cdot \log_{10} \left(\frac{As_{\max}^2}{\sum_{x=0}^{X-1} \sum_{y=0}^{Y-1} (e[x,y])^2} \right) \quad (4.25)$$

字长在哪里 B ： $s_{\max} = 2^B - 1$ 。我们也可以包括一个二进制掩码， $b[x,y]$ ，以便计算特定任意图像区域的PSNR值。因此：

$$\text{PSNR} = 10 \cdot \log_{10} \left(\frac{s_{\max}^2 \sum_{x=0}^{X-1} \sum_{y=0}^{Y-1} b[x,y]}{\sum_{x=0}^{X-1} \sum_{y=0}^{Y-1} (e[x,y])^2} \right) \quad (4.26)$$

虽然PSNR没有感知基础，但它确实与主观评估相当吻合，特别是在比较产生类似类型工件的算法的情况下。它仍然是最常用的客观失真度量，因为它的数学可处理性，并且缺乏任何被广泛接受的替代方法。然而，值得注意的是，PSNR可能会具有欺骗性，例如：

1. 重建信号存在相移。即使人类观察者不会注意到微小的相移也会产生重大变化。
2. 在编码过程中，存在视觉遮罩，通过在不太明显的区域隐藏失真来提供感知的高质量。
3. 错误会随着时间的推移而持续存在。单个帧中的单个小错误可能不明显，但如果它持续到多个帧上，可能会令人讨厌。
4. 正在比较不同的编码策略（例如基于合成和基于块变换（见第10章和第13, 章第10 and 章第13章））。

考虑图4.18中的图像。所有这些都有相同的PSNR值（16.5分贝），但大多数人会同意他们没有表现出相同的感知品质。特别是，右下角的空间移动很小，实际上与原始图像无法区分。在这种情况下，基于感知的指标可以提供与主观意见更紧密的相关性，这些指标在第10章中将得到更详细的考虑。

示例4.5计算PSNR

考虑一个 3×38 位值的图像S及其压缩重建后的近似值， \tilde{S} 。计算重建块的MSE和PSNR。

1	2	3
4	5	6
7	8	9

s

1	2	2
4	4	8
7	8	8

\tilde{s}

[下载：下载全尺寸图像](#)

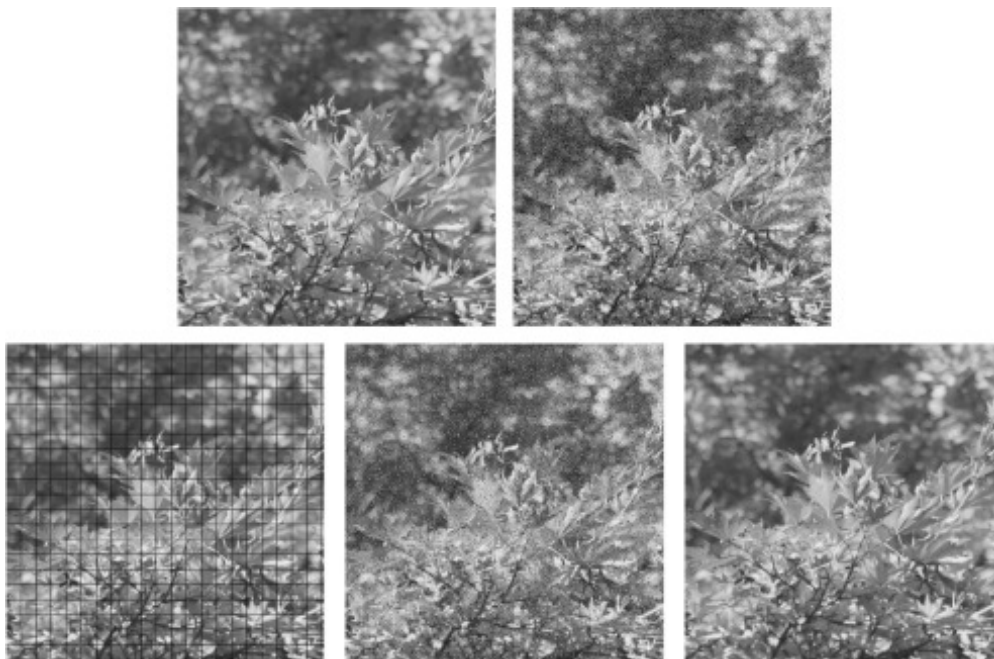
解决方案

我们可以看到 $X = 3, Y = 3, s_{\max} = 255$ ，以及 $A = 9$ 。使用方程 (4.22) 给出MSE：

$$\text{MSE} = \frac{7}{9}$$

从方程 (4.25) 可以看出，该块的PSNR由以下方式给出：

$$\text{PSNR} = 10 \cdot \log_{10} \left(\frac{9 \times 255^2}{7} \right) = 49.2 \text{ dB}$$



[下载：下载全尺寸图像](#)

图4.18。相同PSNR（16.5分贝）的质量比较。左上角到右下角：原始；AWGN（差0.24）；网格线；盐和胡椒噪声；垂直和水平移动5像素的空间。

用于彩色图像和视频的PSNR

对于每个像素有三个RGB值的彩色图像，PSNR可以如上所述计算，除了MSE现在是所有平方值差异除以图像大小和除以3的和。或者，可以为每个颜色组件单独计算PSNR。然而，由于人眼对所有信道的敏感度并不相同，因此通常只需计算亮度PSNR。当使用诸如 YC_bC_r ，其中亮度组件表示颜色通道的加权平均值。

此外，值得注意的是，基于单个帧的平均PSNR结果的视频信号的PSNR计算与计算序列中所有帧的平均MSE，然后计算PSNR不同。前者偏向产生信噪比波动的算法，因此通常使用后者。

在有损图像和视频压缩中，PSNR的典型值在30到50分贝之间，越高越好。超过40分贝的值通常被认为是非常好的，低于20分贝的值通常不可接受的。当两张图像相同时，MSE将为零，导致无限的PSNR。

平均绝对差（MAD）

基于误差的平均平方度量虽然相对简单，但仍然需要每个样本进行一次乘法和两次加法。为了进一步降低复杂性，特别是在一些高度搜索密集型的操作中，如运动估计，还使用了其他更简单的指标。其中一种指标是平均绝对差（MAD）。两个图像或区域之间的MAD， s_1 和 s_2 ，定义为：²

$$MAD = \frac{1}{A} \sum_{x=0}^{X-1} \sum_{y=0}^{Y-1} |s_1[x,y] - s_2[x,y]|$$

(4.27)

在哪里A定义与之前相同。

4.6.3。主观评估

基于数学、基于失真的指标，如MSE和PSNR在某些情况下可能是主观质量的不良指标。因此，有必要制定基于感知的标准，以实现更好的（主观意义）质量评估。基于感知的质量评估工具的一个例子是平均意见得分（MOS）。通过MOS测量，一些观察者查看图像，并以从坏到优的五点比例评估其质量。然后，通过对代表数量的受试者的评级进行统计分析来描述图像（MOS）的主观质量。

也许最常见的测试程序仍然是双刺激持续质量量表（DSCQS）方法，该方法在ITU-R Rec.BT.500[7]中描述，尽管一些多媒体质量评估将基于单一刺激测试。DSCQS测试采用五点减值表，如表4.2所示。ITU-R Rec.BT.500规定了广泛的测试标准和方法，包括观看条件、测试材料的选择、观察者组合的特点、进行测试的方法和结果分析的方法。更多详情见第10章。

表4.2。ITU-R Rec.BT.500主观评价减值表。

分数	损害	质量评级
5	察觉不到	优秀
4	可感知但不烦人	好
3	有点烦人	集市

2	令人讨厌	穷
1	非常烦人	坏

主观测试的问题在于它们很耗时，运行起来可能很昂贵。因此，这为开发强大的感知质量指标提供了额外的动力，这些指标可以充满信心地用于它们。

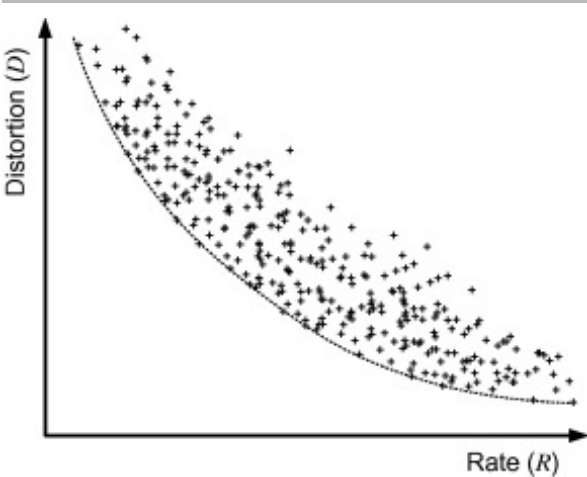
4.7。速率和失真

4.7.1。速率-失真特性

任何压缩序列的比特率将取决于：

- 使用的编码算法（帧内和帧间、整数或子像素运动估计、编码模式或块大小可用）。
- 内容（高时空活动通常需要更多的位来编码）。
- 选择的编码参数。在最粗糙的层面上，这包括空间分辨率和帧率等。在更细粒度下，量化器控制、模式内与模式间以及块大小选择等问题将是关键。基于良好参数选择的编码与基于糟糕选择的编码之间存在巨大差异。

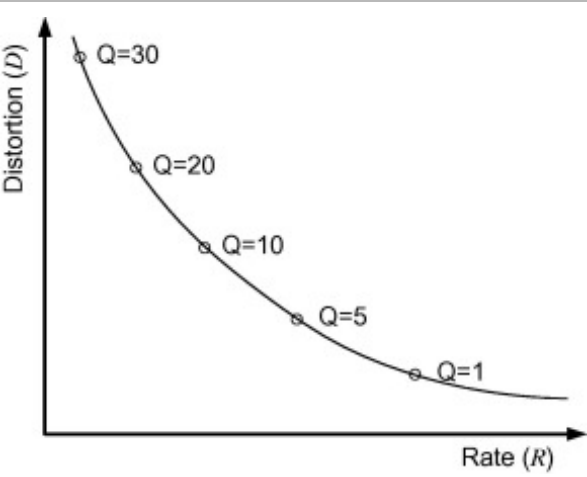
根据第4.6.1节的论点，我们需要一个编码器性能的描述，以捕捉它如何以速率与失真进行交易。这是使用绘制给定编解码器和给定视频文件速率与失真（或速率与质量）的图表来实现的。这使我们能够比较编解码器的性能，并评估单个编解码器中的参数选择如何影响性能。但是，如果我们随机选择编码参数，这些组合中可能会有一些产生好的结果，有些必然会产生不好的结果。对于许多不同的参数选择，点数的情节可能与图4.19相似。当然，目的是确保编解码器在实践中产生的操作点是虚点曲线上的操作点——帕累托曲线，该曲线在任何给定编码速率下产生尽可能低的失真。为了实现这一目标，我们必须执行速率失真（RDO）或速率质量优化（RQO）。



[下载：下载全尺寸图像](#)

图4.19。用于各种编码参数选择的速率失真图。

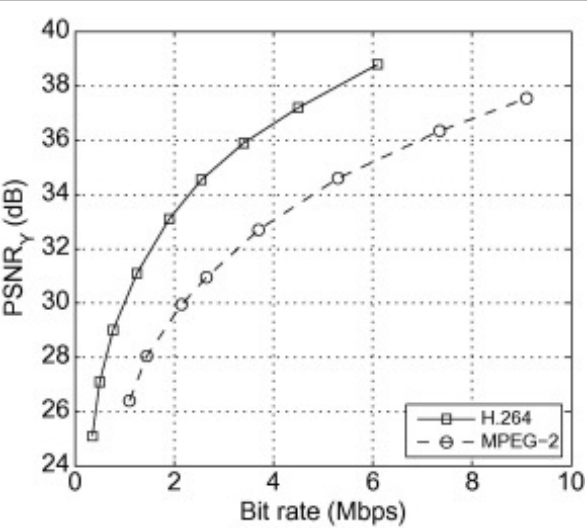
例如，一个简单的编解码器，其中对RD性能的唯一控制是量化器步长。在这种情况下，调整量化器将产生不同的编解码器操作点，如图4.20所示。



[下载：下载全尺寸图像](#)

图4.20。速率-失真图显示量化器控制的操作点。

图4.21显示了速率-质量曲线（基于PSNR而不是失真）示例。这比较了两个编解码器（MPEG-2和H.264/AVC）的性能。



[下载：下载全尺寸图像](#)

图4.21。典型标准定义娱乐内容的示例速率-质量比较曲线。

4.7.2。速率-失真优化

速率-失真优化（RDO）旨在最大限度地提高图像或视频质量，但须遵守比特率限制。RDO要求和方法在所有视频编码标准中都是非规范的，但就编码器性能而言，它们仍然是键分器。现代标

准（如H.264/AVC和HEVC）中包含的许多高级功能（例如块大小选择）如果不作为编码器控制过程的一部分进行优化，将无法节省。RDO必须为视频的每个区域选择最佳模式和参数集。通常采用拉格朗日优化方法，其中失真度量， D ，基于平方差和和比率， R ，包括与决策相关的所有位，包括标头、运动、侧面信息和转换数据。

对于所有可能的参数向量 \mathbf{p} ，目的是解决约束问题：

$$\min_{\mathbf{p}} D(\mathbf{p}) \quad \text{s. t.} \quad R(\mathbf{p}) \leq R_T \quad (4.28)$$

在哪里 R_T 是目标比特率，或求解无约束的拉格朗日公式：

$$\mathbf{p}_{opt} = \arg \min_{\mathbf{p}} \{D(\mathbf{p}) + \lambda R(\mathbf{p})\} \quad (4.29)$$

在哪里 λ 控制速率-失真权衡。有关 RDO 方法的更多详细信息，请参阅第 10 章。

4.8。摘要

本章介绍了支撑编码过程的数字表示、格式、处理技术和评估方法。伽马校正和色彩空间转换等技术已被证明对于提供最适合进一步压缩的视频信号的数字描述至关重要。事实上，通过利用HVS的颜色处理和感知属性，我们已经证明，即使在我们应用以下章节中描述的传统压缩方法之前，也可以设计适当的亮度/色度格式，将数字视频的比特率降低50%。最后，我们强调，评估方法是视频编解码器设计中必不可少的元素，既是比较不同压缩系统性能的一种手段，也是优化编解码器速率失真性能的基础。

Recommended articles

Citing articles (0)

参考文献

- [1] 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
- [2] <http://en.wikipedia.org/wiki/Interlaced_video>，2013年9月。
谷歌学术
- [3] C.Poynton，彩色空间导游，2013年8月。
<http://www.poynton.com/PDFs/Guided_tour.pdf>。
谷歌学术
- [4] B.拜耳，彩色成像阵列，美国专利3,971,065，1976年。
谷歌学术
- [5] C.波因顿
数字视频和高清

(第二版), 摩根·考夫曼 (2012)
[谷歌学术](#)

[6] 美国。温克勒
数字视频质量
约翰·威利 (2005)
[谷歌学术](#)

[7] 建议ITU-R BT.500-13, 电视图像质量主观评估方法, ITU-R, 2012年。
[谷歌学术](#)

☆ 有关[图4.10](#)、[图4.11](#)、[图4.14](#)、[图4.15](#)、[图4.17](#)、[图4.18](#)的彩色版本, 请参阅电子版或网站。

¹ 这些术语通常可以互换使用。

² 注意: 在搜索应用程序中, 使用绝对差分之 (SAD) 是正常的, 省略除法操作来降低复杂性。

版权所有?2014爱思唯尔有限公司。保留一切权利。



[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。