



## 交流图片




图像和视频编码课程

2014 年, Pages 317-360

## 第10章-测量和管理图片质量

大卫·R·布尔

显示更多 ∨

 大纲 |  共享  引用<https://doi.org/10.1016/B978-0-12-405906-1.00010-6>

获取权利和内容

## 摘要

本章涉及图像和视频质量评估的重要领域。首先描述了进行主观试验和分析结果的方法,并讨论了一些公开的[主观试验](#)数据库的性质。然后从视频质量的相关性和主观意见的角度对其进行回顾和比较。这些包括结构相似性、视频质量指标 (VQM)、视觉[信噪比](#) (VSNR)、空间和时间最明显失真 (MAD)、运动调谐时空质量评估 (又名电影) 和基于感知的视频指标 (PVM)。视频指标在图片压缩和交付过程中也发挥着重要作用,例如启用循环速率质量优化 (RQO)。讨论了这个问题,描述了一些最常见的技术,这些技术使我们能够为视频的每个时空区域选择最佳编码参数。最后,描述了速率控制方法。



## 关键词

主观测试; 视频数据库; 平均意见评分 (MOS); 视频质量指标; 峰值信噪比 (PSNR); 移动; 结构相似性 (SSIM); 速率失真优化 (RDO); 速率控制; 假设参考解码器 (HRD)

评估[感知质量](#)是图像和视频处理中最关键但最具挑战性的任务之一。视觉感知高度复杂,受诸多

混淆因素影响，未完全理解，难以建模。出于这些原因，[压缩算法](#)性能表征总是基于主观评估，其中一组观众被问及他们在一系列测试条件下对质量的看法。第10.2节描述了进行此类试验和分析结果的方法。第10.3节讨论了一些公开可用的[主观测试](#)数据库的性质。

传统的视频质量的客观衡量标准是使用图片的编码版本与其原始版本之间的绝对或平方差异来计算的。然而，众所周知，人类观众所经历的[感知扭曲](#)无法使用如此简单的数学差异来充分描述。由于基于失真的度量的局限性，基于感知的度量开始取代它们。这些提供了增强与主观意见相关性的潜力，从而能够更准确地估计视觉质量。第10.4节回顾了其中最有帮助的，包括结构相似性（SSIM）、视频质量指标（VQM）、视觉[信噪比](#)（VSNR）、空间和时间最明显的失真（MAD）和运动调谐时空质量评估（又名电影）。尽管与峰值信噪比（PSNR）等常规指标相比，这些方法的性能有所提高，但包括高计算复杂度、延迟和集成到编码过程的灵活性在内的一些问题仍然需要解决。在这方面，我们引入了一种替代方案，即基于感知的视频指标（PVM），它成功地解决了其中一些问题。

视频指标不只是用来比较不同的编码算法。它们在图片压缩和交付过程中也发挥着重要作用，例如启用循环速率-质量优化（RQO）。第10.5节讨论了这个问题，描述了一些最常见的技术，这些技术使我们能够为视频的每个时空区域选择最佳编码参数。最后，在第10.6节中描述了在时变信道条件（速率控制）的情况下控制编解码器生成的比特率的方法。

## 10.1。一般考虑和影响

首先，值得一提的是，评估受损图像或视频内容的质量，无论是由于传输丢失还是压缩[伪影](#)，都不是简单的。可以通过以下两种方式之一实现：

1. **主观：**需要许多观察者和许多有代表性的减值条件和内容类型的陈述。必须密切控制主观测试条件，对观察员进行适当筛选，并对结果进行后处理，以确保一致性和统计意义。它们既昂贵又耗时，但总体上是有效的。
2. **客观地：**使用试图捕捉[人类视觉系统](#)感知机制的指标。这里的主要问题是，简单的指标与HVS几乎没有相关性，通常与主观结果不相关，特别是在失真较高的较低比特率下。更复杂、更受感知启发的指标，虽然近年来有了显著改善，但在某些测试条件下仍然不一致。其结果是，基于[平均平方误差](#)（MSE）的指标仍然是最常用的评估方法，无论是用于循环优化还是外部性能比较。

### 10.1.1。我们想评估什么？

参考文献很好地概述了评估视觉质量的原因和过程。[\[1\]](#)，[\[2\]](#)。主要动机概述如下：

1. 比较不同视频编解码器在比特率和内容类型范围内的性能。
2. 比较不同视频编解码器在一系列[频道损伤](#)方面的性能。
3. 比较各种参数和编码选项对给定类型编解码器的影响。

后者与循环速率失真（或质量）优化RDO（RQO）特别相关，我们稍后将看到。

### 10.1.2。对感知质量的影响

## 人类视觉感知

一个多世纪前，视觉科学家开始关注我们对图像和视频失真的感知敏感性。我们在[第二章](#)中看到，这种灵敏度因屏幕亮度、局部[时空频率](#)特征、运动类型、眼球运动、各种类型的人工制品以及当然还有观看环境而异。为了确保[主观测试](#)的有效性和客观指标性能的一致性，这些敏感性的影响必须尽可能体现在用于评估的内容中，并体现在任何所用指标的结构中。

还应当指出，根据年龄、疾病、疲劳或视觉系统缺陷，HVS的性能因受试者而异。偏见也可能通过个人偏好甚至无聊来影响观众的意见。

## 观景环境

为了对图片质量进行一致的主观评估，必须记录观看条件的细节和参数，并在测试之间保持尽可能一致。在比较不同实验室的结果时，这一点尤为重要。浏览环境的几个因素会影响对画质的感知。这些包括：

- **显示器尺寸：**显示器填充外围视觉越多，它就越吸引人。
- **显示亮度和动态范围：**闪烁和时间CSF将取决于显示亮度，更高的动态范围显示会带来更大的深度感。
- **显示分辨率：**空间和时间更新率将影响感知的工件类型。
- **环境照明：**高环境照明水平和显示器反射将降低感知的对比度，并引入其他工件。
- **观察距离：**该参数与[时空分辨率](#)相互作用，将对感知质量产生重大影响。
- **音频质量：**最后，已经清楚地表明，音频质量的变化会影响对视频质量的感知。

## 内容类型

我们在[第二章](#)中看到了内容类型如何影响感知质量。例如，几乎没有运动的普通场景比具有复杂纹理和运动的时空繁忙序列更容易编码。如果编解码器A用于编写简单内容，而类似的编解码器B用于编写更困难的内容，那么编解码器A在速率-失真性能方面似乎优于编解码器B，但实际上并非如此。

因此，测试应基于具有代表性和合理挑战性的内容，提供一系列时空活动水平（见第[10.2.2](#)节）。为了确保[评估员](#)的重点仍然是画质，而不是叙事，通常选择短序列（通常为10秒），通常不伴有音频轨道。所有被测试的编解码器都应在所有序列和所有损伤级别上进行评估。

## 人工制品类型

我们不会在这里列出所有类型的编解码器生成的所有工件类型，但希望读者现在能从前几章中知道这些。一般来说，它们可以分为以下几种：

- **边缘：**由于从变换或运动估计的阻塞，或与[量化](#)相关的轮廓效应。
- **模糊：**由于在块变换或小波系数量化过程中高频细节丢失。
- **振铃：**滤库合成阶段的量化伪影可引起振铃效应。

- **不同**：例如，由于某些编解码器中平面区域的几何变换不一致，扭曲刚性纹理或合成动态纹理时出错（更多详细信息见第13章）。

## 10.2。主观测试

### 10.2.1。理由

尽管客观指标的性能最近有所进步，而且数量众多，但尚未被普遍接受为质量的明确指标。<sup>1</sup> 因此，有必要使用受控**主观测试**。主观评估方法被广泛用于描述、比较或验证视频**压缩算法**的性能。基于一组有代表性的测试内容和损伤条件，它们旨在提供可能查看测试系统的人的反应的有力迹象。

大多数主观测试实验是基于国际电联和其他组织多年来根据许多组织的集体经验提出的建议进行的。主要参考文件是ITU-R rec. BT.500[4]和ITU-T建议。P.910 [5]。以下讨论主要基于这些建议。

### 10.2.2.测试序列和条件

#### 测试材料

一般来说，应选择适合所解决的问题或应用领域的测试材料。它通常根据其固定参数来定义，这些参数包括：使用的序列数量、序列持续时间、序列内容、空间分辨率、时间分辨率和位深度。

对于大多数一般评估，测试材料将是“关键，但不是不适当的”[4]。这意味着它应该包含难以编码的内容，但应该仍然代表典型的观看。选择的序列应始终包括关键材料，因为通常无法推断非关键材料的结果。然而，如果测试旨在描述特定困难案例的性能，那么预计将选择使用的材料来反映这些案例。

一般来说，通常选择至少四种类型的序列进行测试。这个数字将提供活动级别的最低覆盖率，同时不会用冗长的测试会话来烦闷**评估者**。

#### 活动或信息级别

夹子中的时空活动量对可能的压缩程度和由此产生的重建的质量有很大影响。一般来说，测试序列应选择与感兴趣应用领域普遍存在的渠道条件范围一致。在最终选择测试材料之前，正式评估候选剪辑的时空活动水平非常有用，以确保它们涵盖适当的时空信息空间[3]，[5]。

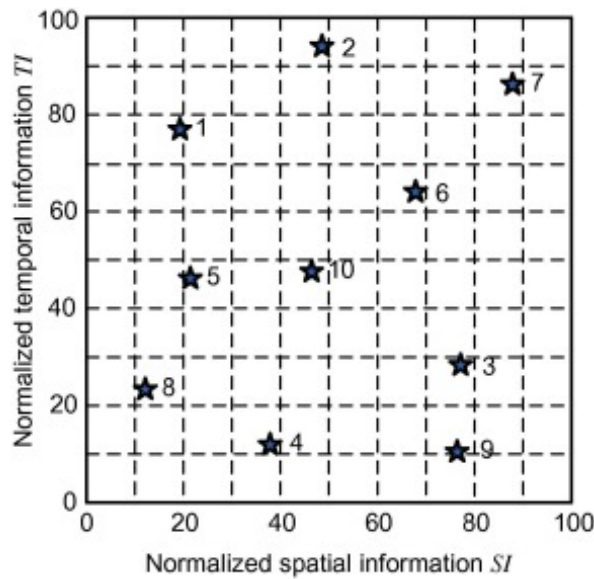
遵循建议ITU-T建议的方法。BT.910[5]，空间信息测量基于每帧的标准偏差 $S_z$ 在索贝尔滤波后的顺序中。所有帧的标准偏差的最大值被选择为国际单位制指标：

$$SI = \max_{V_z} \left\{ \sigma_{V_z} (Sobel(S_z(x,y))) \right\} \quad (10.1)$$

BT.910中使用的时间信息度量（TI）基于连续帧中同地流马像素值之间的差值的标准偏差。如上所述，TI测量计算为序列中所有帧的标准偏差的最大值。因此：

$$TI = \max_{\forall z} \left\{ \sigma_{\forall(x,y)} (S_z(x,y) - S_{z-1}(x,y)) \right\} \quad (10.2)$$

通常制作一个SI对TI图，以确保选择用于测试的序列为SI-TI空间提供足够的覆盖范围。图10.1显示了覆盖范围良好的指示性图。应当指出，温克勒[3]基于运动矢量，采用了与上述略有不同的方法。他还包括一个基于颜色的信息测量。



下载：下载全尺寸图像

图10.1。一系列测试序列的空间和时间信息覆盖范围。

## 测试条件

测试条件是指在测试期间将更改的参数。这些可能包括编解码器类型、编解码器参数和比特率，典型的测试将比较两个或多个编解码器或编解码器配置，以获得一系列序列和编码速率。

在一组有代表性的、分布良好的测试条件下评估编解码器性能是正常的。大多数评估方法对条件的范围和分布都很敏感。通过限制条件范围，也可以通过包含一些显式（直接锚定）极端情况或在整个测试中分发这些情况，而无需显式识别（间接锚定），来降低灵敏度。此外，测试条件越多，测试时间就越长。测试必须避免查看器疲劳，这对使用的序列数量和被评估的测试条件的数量施加了限制。

演讲次数等于  $N = N_s \times N_c \times N_r$  在哪里  $N_s$  是源序列的数量， $N_c$  是测试条件的数量（ $N_c = N_{c1} N_{c2}$  在哪里  $N_{c1}$  是测试的编解码器数量和  $N_{c2}$  是测试的比特率（通常包括未压缩）的数量，以及  $N_r$  是冗余因子（即每个条件重复的次数）。如果每次演示都  $T$  几秒钟，那么每次测试的时间是  $N \times T$  几秒钟，如果有  $K$  观察员，那么完成审判的总时间是  $N \times T \times K$  秒（不包括换机和休息的死时间）。

### 10.2.3. 选择主题

根据测试的性质，观察员可以是专家或非专家。研究发现，进行类似测试的不同实验室之间可能会发生系统差异[4]。原因尚不完全理解，但很明显，专家观察员对材料的看法与非专家不同。



建议的其他解释包括性别、年龄和职业[4]。在大多数情况下，对于消费者应用程序，预计大多数观察员应该是非专家，并且没有人直接参与受测系统的开发。

在最终选择评估人员之前，应对所有候选人进行筛选，以确保他们具有正常的视力（无论是否带矫正镜）。这可以使用斯内伦图来检查视觉敏锐度和石原图来检查彩色视觉。使用的评估员人数取决于测试的规模以及所采用方法的敏感性和可靠性。通常建议至少雇用15门科目。作者建议数字稍高一点，以便在结果处理期间删除异常值。

10.2.4。测试环境

测试环境通常根据测试要求指定——通常意味着现实的消费环境或实验室条件：

- **实验室条件：**这些旨在提供测试条件，使损伤具有最大可检测性。
- **家庭条件：**这些旨在允许评估更接近典型消费者端环境的查看条件。

与这两个环境相关的选定参数列于表10.1。

表10.1。精选主观测试查看环境参数。来自裁判。 [4]， [5]。

| 查看条件         | 首页         | 实验室              |
|--------------|------------|------------------|
| 非活动与峰值屏幕亮度之比 | ≤ 0.02     | ≤ 0.02           |
| 显示器峰值亮度      | 200 cd /m² | 见建议BT.814和BT.815 |
| 背景色度         | –          | D65              |
| 最大观测角        | 30°        | 30°              |
| 观景距离         | 3-6H       | 3-6H             |
| 房间照明         | 200 lx     | 低的(≤ 20 lx)      |

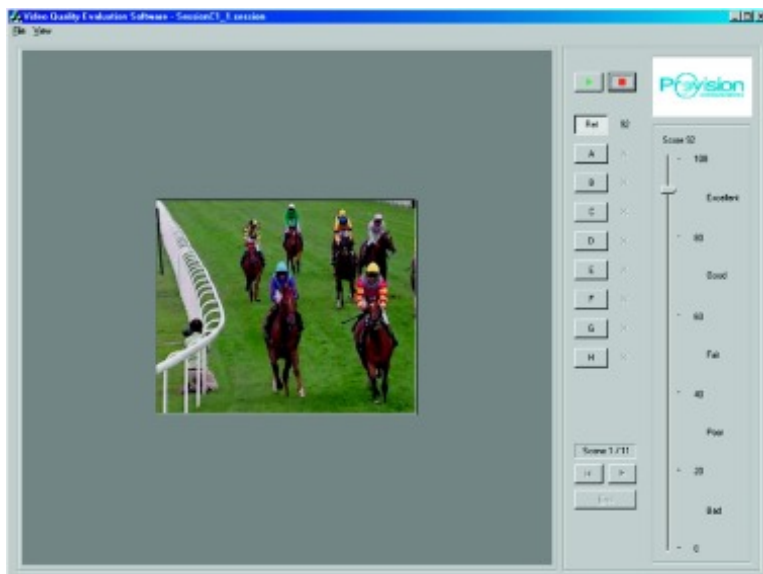
10.2.5。测试方法和记录结果

主观测试方法的通用图如图10.2所示。消息来源直接向受试者或通过受测系统提供演示片段，该系统根据当时设置的测试条件引入缺陷。受试者将同时或按顺序查看评估显示器上的内容，该显示器显示受损的视频剪辑，在某些情况下，还会显示未受损的剪辑。控制器将控制时间和在不同测试条件之间的切换。在某些情况下，用户可能能够影响剪辑的顺序、重复和时间。



[下载：下载全尺寸图像](#)

图10.3。DSCQS评估表的一部分。



[下载：下载全尺寸图像](#)

图10.4。用于SAMVIQ测试的接口。

礼遇：ProVision Communications。

- **记录测试条件：**由于主观测试之间可能存在不一致，因此记录有关测试的所有显著信息以及使用的参数和设备至关重要。易卜拉欣的团队在HEVC的上下文中提供了很好的例子——例如，请参阅参考文献。[6]，[7]。

## 双刺激方法

双刺激评估仍然是评估压缩视频质量的最流行手段。最常用的程序是ITU-R rec. BT.500虽然最初用于电视应用，但现在使用范围更广。ITU-T rec.P.910专门针对多媒体应用程序，但与BT.500有很多相似之处。ITU-R建议。BT.500描述了两种主要的双重刺激方法：

- **双刺激连续质量表（DSCQS）：**DSCQS方法最适合测试材料和原始材料质量相似的情况，目的是评估被测试系统相对于原始系统的性能如何。测试被安排为一组配对的夹子（A和B），其中一个原始（锚），另一个是被测试系统破坏的原始夹。由于订单是随机的，评估人员不知道哪个剪辑是原始剪辑。如图10.5所示。这对情侣被展示两次，第一次是为了熟悉，第二次是为了投票。





## 10.2.6. 统计分析与意义测试

必须以稳健和一致的方式分析和呈现收集的数据。大多数测试结果被缩放到0-100之间，这里假设。一般来说，按照BT.500[4]中描述的过程，让我们假设有 $N$ 演示文稿，每个演示文稿由一个 $N_c$ 测试条件。

### 平均分数的计算

每个演讲的所有观察员的平均分数由以下人员给出：

$$\bar{u}_{csr} = \frac{1}{K} \sum_{k=1}^K u_{kcsr} \quad (10.3)$$

在哪里 $u_{kcsr}$ 是观察者的分数 $k$ 响应序列 $s$ 在测试条件下 $c$ 重复 $r$ ，以及 $K$ 是观察员总数。在筛选观察员以制作后，进行类似的处理 $\bar{u}_{sc}$ ，给定测试条件下每个序列的均值意见得分（MOS）。

在许多测试中，相对于参考内容进行分数补偿是常见的。这对于单一刺激方法尤为重要，这种方法采用隐藏引用删除过程来产生差异平均意见评分（DMOS）。

### 置信区间

当研究结果被提出时，良好做法是也包括置信区间[9]。平均值的置信区间是包含总体平均一定比例时间的区间，通常为95%。置信区间基于每个样本的大小及其标准偏差，提供的信息比点估计更多。95%置信区间由以下方式给出：

$$[\bar{u}_{csr} - \delta_{csr}, \bar{u}_{csr} + \delta_{csr}] \quad (10.4)$$

### 算法10.1

#### 使用 $\beta_2$ 测试[4]

- 
1. REPEAT for each observer,  $k$ :
  2.  $k = k + 1$ ;
  3. FOR  $c, s, r = 1, 1, 1$  to  $C, S, R$  DO:
  4. IF  $2 \leq \beta_{2csr} \leq 4$  THEN: IF  $u_{kcsr} \geq \bar{u}_{csr} + 2\sigma_{csr}$  THEN  $P_k = P_k + 1$ ;  
IF  $u_{kcsr} \leq \bar{u}_{csr} - 2\sigma_{csr}$  THEN  $Q_k = Q_k + 1$ ;
  5. ELSE: IF  $u_{kcsr} \geq \bar{u}_{csr} + \sqrt{20}\sigma_{csr}$  THEN  $P_k = P_k + 1$ ; IF  $u_{kcsr} \leq \bar{u}_{csr} - \sqrt{20}\sigma_{csr}$  THEN  $Q_k = Q_k + 1$ ;
  6. END FOR;
  7. IF  $\frac{P_k + Q_k}{N_c N_s N_r} > 0.5$  and  $\left| \frac{P_k - Q_k}{P_k + Q_k} \right| < 0.3$  THEN reject observer  $k$ ;
  8. UNTIL  $k = K$ .
- 

[下载：下载全尺寸图像](#)

在哪里

$$\delta_{csr} = 1.96 \frac{\sigma_{csr}}{\sqrt{K}}$$

每个演示文稿的标准偏差由以下方式给出：

$$\sigma_{csr} = \sqrt{\sum_{k=1}^K \frac{(\bar{u}_{csr} - u_{kcsr})^2}{(K-1)}} \quad (10.5)$$

实验平均值和真实平均值之间的绝对差异（基于无限数量的观察者）小于概率为95%的方程置信区间（10.4）（前提是分数正常分布）。

## 筛选观察员

常规使用正常分布的分数集筛选（用于DSIS和DSCQS） $\beta_2$ 测试，基于库特系数。时间 $\beta_2$ 在2到4之间，分布可以假设为正常。屈曲率系数定义为四阶矩与二阶矩平方的比率。因此：

$$\beta_{2\,csr} = \frac{m_4}{(m_2)^2} \quad \text{with} \quad m_x = \frac{\sum_{k=1}^K (u_{kcsr} - \bar{u}_{csr})^x}{K} \quad (10.6)$$

我们现在计算 $P_k$ 和 $Q_k$ ，如算法10.1中给出的，以确定在形成分布时是否应拒绝任何观察者。

### 10.2.7. 完整的评估过程

前面几节对主观测试方法进行了概述。通过ITU-R rec等建议，双重和单一刺激方法都已正式化。BT.500-13，ITU-T建议。P.910和SAMVIQ [2], [4], [5]。

#### 算法 10.2 主观测试方法论

- 
1. Select the codecs under test and the parameter settings to be used (GOP structures, filtering, RDO modes, ME modes, etc.);
  2. Define the coding conditions to be evaluated (spatial and temporal resolutions, bit rates, etc.);
  3. Identify or acquire the test sequences that will be used, ensuring that they are sufficiently varied and challenging, representative of the applications of interest and of appropriate duration;
  4. Select an appropriate evaluation methodology (dependent on conditions being tested, impairment levels, etc.);
  5. Design the test session and environment (complying with recommendations of whichever methodology is being adopted);
  6. Dry run the session to remove any bugs or inconsistencies;
  7. Organize the full test, inviting sufficient subjects (assessors), with pre-screening to eliminate those with incompatible attributes;
  8. Run the tests (including assessor briefing and dummy runs) and collect the results;
  9. Perform post-processing to remove outliers and assess significance;
  10. Produce test report fully documenting laboratory conditions, test conditions, and analysis/screening methods employed.
- 

[下载：下载全尺寸图像](#)

---

建议读者在实施任何实际主观测试实验之前参考相关标准或程序文件。然而，作为指南，主观测试的一般方法在算法10.2中进行了描述。

## 10.3. 测试数据集以及如何使用它们

一个可靠的主观数据库有三个主要用途。首先，它将在使用的测试条件下对选定的[压缩方法](#)进行鲁棒比较。其次，它为基于既定统计方法验证客观指标提供了非常有用的基础。最后，它可以用于在压缩背景下描述HVS属性，从而为完善客观质量指标提供了有价值的工具。

### 10.3.1. 数据库

以下是主要公开数据库的简要概述。有关更多详细信息，请参阅与每次审判相关的适当报告。Winkler [\[3\]](#) 提供了对现有数据库的出色评论和比较。

#### VQEG FRTV

最早的客观视频质量评估的主要主观数据库是通过视频质量专家组（VQEG）FRTV第一阶段计划[\[10\]](#)（随后是2003年第二阶段[\[11\]](#)）生成的。第一阶段数据库于2000年建成，以解决与全球引入数字电视标准相关的质量问题。该数据库使用287名[评估员](#)（分为四组）和20个视频序列，其中包含16种失真类型，以产生超过26,000个主观意见分数。FRTV-I使用了两个决议， $720 \times 486$ 和 $720 \times 576$ ，一切都交织在一起。使用的测试方法是前面描述的[双刺激连续质量量表](#)（DSCQS）方法。

尽管FRTV-I数据库具有良好的覆盖率和统一性，但它在当代编码要求中确实有局限性：

1. 提供的[工件](#)没有充分反映视频压缩的最新进展，因为编码主要基于MPEG-2。
2. 其评估基于标准定义格式。不含高清素材,不含高清素材
3. 其内容是交错的，而当前的趋势是渐进式格式，特别是流式应用程序。

然而，尽管有这些限制，但由于序列数量众多、主题数量众多、覆盖范围良好且一致性良好 [\[3\]](#)，VQEG FRTV 第一阶段仍然是用于客观质量指标验证的更常用的数据库之一。

#### 现场表演

一个最近和同样重要的[主观测试](#)数据库是在图像和视频工程实验室（LIVE）[\[12\]](#)，[\[13\]](#)，[\[14\]](#)开发的数据库。LIVE数据库基于MPEG-2和H.264/AVC[压缩算法](#)，并包含模拟有线和无线传输错误的结果。十个参考序列用于分辨率为 $768 \times 432$ ，其中每个引用有15个扭曲的版本。该数据集采用单一刺激连续测试程序，最终分数再次在0到100之间进行缩放。与VQEG FRTV-I数据库相比，现场观看试验仅雇用了38名评估员。根据Winkler[\[3\]](#)的分析，LIVE还提供了更窄的源内容范围（以SI和TI为单位）和失真（以PSNR和MOS值为单位）。

#### 其他

其他可用于评估指标的数据库包括IRCyN/IVC[\[15\]](#)、VQEG HDTV数据库[\[16\]](#)和VQEG多媒体第一阶段数据库[\[17\]](#)。这些通常用于特定研究目的，通常不太适合全面验证图像和视频质量指标。参考文献中可以对这些进行比较。 [\[3\]](#)。

### 10.3.2. 平均意见得分与客观指标的关系

主观结果经常被用作基准，以确定DMOS（或MOS）分数与特定客观图像质量指标之间的关系。客观视频质量指标产生的分数必须以可预测和可重复的方式与观众分数相关。预测和DMOS

之间的关系不必是线性的，因为[主观测试](#)可以在测试范围的极端处进行非线性质量评级压缩。因此，关系的线性性不那么关键，而是关系的稳定性和确定预测性有用的数据集误差方差。

BT.500[4]描述了一种在 $\bar{u}$ （平均分数）和基于[逻辑函数](#)的度量。首先将平均分数值的范围归一化如下（经过筛选）：

$$p = \frac{(\bar{u} - u_{\min})}{(u_{\max} - u_{\min})} \quad (10.7)$$

典型的关系 $p$ 和给定的失真度量 $D$ 通常表现出斜对称的乙状体形式。因此函数 $p = f(D)$ 可以用形式的逻辑函数近似：

$$\tilde{p} = \frac{1}{1 + e^{(D - D_M)G}} \quad (10.8)$$

在哪里 $D_M$ 和 $G$ 是从实验数据中可以非常简单地导出的常量[4]。

### 10.3.3. 使用公共（或私人）数据库评估指标

对特定目标指标相对于与特定数据库关联的MOS值主体的性能的判断通常基于某些统计属性。这些通常与预测准确性、[单调性](#)和适合度的一致性有关。常用的措施如下：

#### 线性相关性

[皮尔逊线性相关系数](#)（LCC）用于衡量指标与主观分数的拟合准确性。它描述了被测指标在多大程度的预测主观质量评级。一般形式是为了一组 $N$ 测量-预测对 $(x_i, y_i)$ 作为：

$$r_p = \frac{\sum_{i=0}^{N-1} (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=0}^{N-1} (x_i - \bar{x})^2} \sqrt{\sum_{i=0}^{N-1} (y_i - \bar{y})^2}} \quad (10.9)$$

在哪里 $x_i$ 将代表实际的MOS（或DMOS）分数和 $y_i$ 将代表预测的MOS（或DMOS）分数。

#### 秩阶相关性

使用秩阶度量评估模型的预测与主观质量评级的相对大小的相关性程度。这标志着预测单调性，即不同测试之间的差异迹象在多大程度上与主观分数与指标对它们的预测相关。通常，[斯皮尔曼秩阶相关系数](#)用于此目的：

$$r_s = \frac{\sum_{i=0}^{N-1} (\mathcal{X}_i - \bar{\mathcal{X}})(\mathcal{Y}_i - \bar{\mathcal{Y}})}{\sqrt{\sum_{i=0}^{N-1} (\mathcal{X}_i - \bar{\mathcal{X}})^2} \sqrt{\sum_{i=0}^{N-1} (\mathcal{Y}_i - \bar{\mathcal{Y}})^2}} \quad (10.10)$$

在哪里 $\mathcal{X}_i$ 是排名顺序 $x_i$ ， $\mathcal{Y}_i$ 是排名顺序 $y_i$ ，以及 $\bar{\mathcal{X}}$ 和 $\bar{\mathcal{Y}}$ 是他们的中值。

#### 异常值比

异常值比， $r_o$ ，有效地衡量预测一致性——该指标在内容和减值范围内预测主观得分的能力。异常值通常归类为预测数据点，该数据点大于与相应MOS点的阈值距离。通常使用MOS值标准偏差误差两倍的阈值。因此，如果满足方程的数据点数（10.11）是 $N_o$ ，然后，异常比只需在方程（10.12）中给出。



$$|x_i - y_i| > 2\sigma_{y_i} \quad (10.11)$$

$$r_o = \frac{N_o}{N} \quad (10.12)$$

## 10.4. 客观质量指标

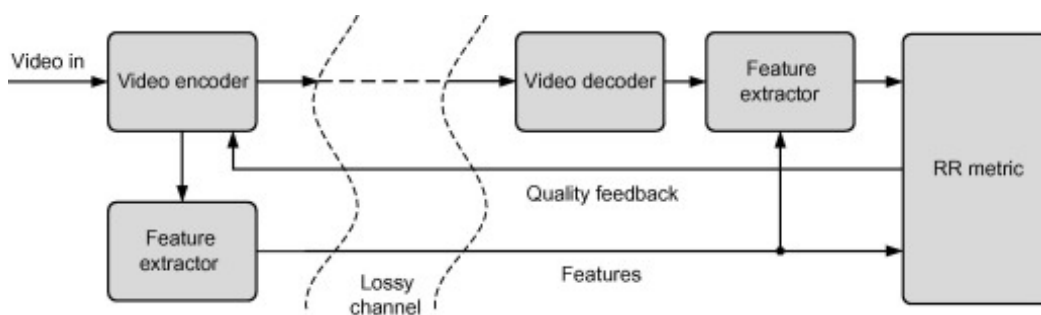
### 10.4.1. 为什么我们需要质量指标？

图像和视频质量评估在图像和视频处理的许多方面发挥着至关重要的作用，特别是与视频编码和通信相关的方面。在交流图片的语境中，它们有三个主要用途：

1. **算法开发和基准测试：**我们以前已经看到，主观评估虽然在描述竞争编码算法的优缺点方面非常有效，但代价高昂且耗时。可靠和一致的客观指标的存在提供了一种简单得多的比较手段。此外，他们可以就对给定编解码器进行各种算法修改的益处或其他好处提供指导（例如，添加循环过滤器或使用多个参考帧的好处）。
2. **速率-质量优化：**在编码循环中，越来越需要质量评估，以便在某些内容和速率约束的情况下，立即就使用哪种编码模式和参数设置来优化性能做出RQO决策。
3. **流控：**在网络传输视频内容的情况下，编码器和发射机在解码后，了解接收端信号的质量是有利的。这使得编码器能够了解流行的信道条件，从而在速率和错误控制方面做出适当决定。

客观质量评估方法通常分为全参考（FR）、减少参考（RR）或无参考（NR）。

- **FR方法**广泛用于原始材料可用的应用程序，例如评估图像和视频编码算法性能，或做出速率-质量优化决策时。
- **NR方法**通常只在没有参考内容的情况下使用[18]。情景可能包括在评估有损通信系统对接收器的影响时。产生“盲”指标极其困难，因此NR方法的使用通常仅限于特定的操作场景和失真类型。它们不能很好地推广，如果可能，最好减少参考指标。
- **RR方法**[19]仅使用有关来源的部分信息来预测质量。他们在有损通信网络中找到了应用，在那里，质量预测可以通过部分了解原始内容，也可能通过信道状态来提供信息。图10.6描述了RR度量的可能场景，即在编码器中提取稀疏特征，并通过受保护（低数据速率）的侧信道传输到解码器。在解码器中，从重建的信号中提取类似特征，并在RR度量中进行比较。然后，解码器重建质量的指示可以反馈给编码器，以便它可以根据流行的信道状态做出明智的编码决策。显然，任何额外的侧面信息都会对编码信息的比特率产生开销，这必须在质量收益的背景下进行评估。



[下载：下载全尺寸图像](#)

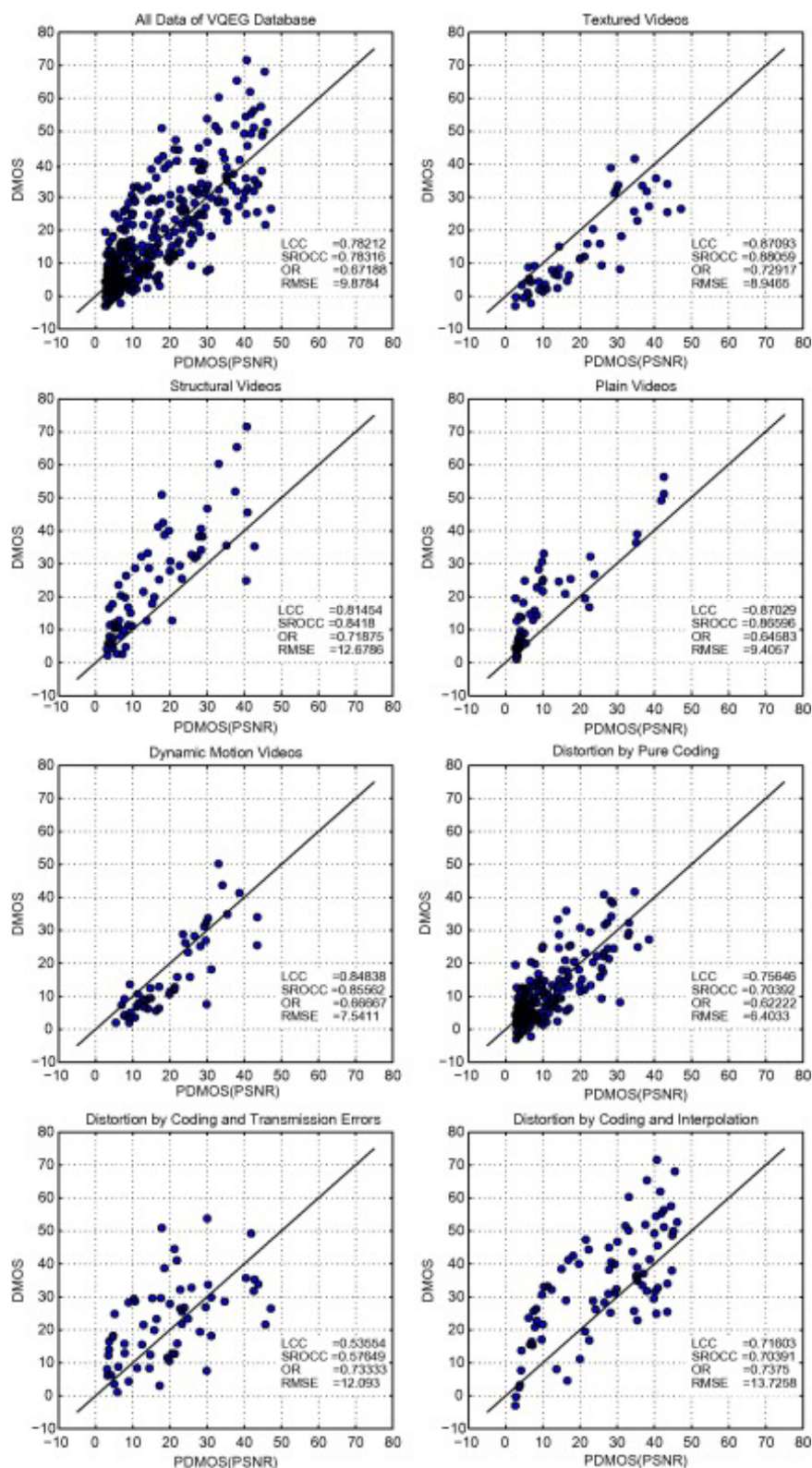
图10.6。减少参考视频质量评估。

#### 10.4.2。PSNR的特征

对于某些测试条件下的特定类型的视觉信号，PSNR可以提供一种简单有效的失真估计方法。例如，Huynh-Thu和Ghanbari[20]表明，当PSNR用于比较基于相同测试数据的类似编解码器或编解码器增强时，可以提供一致的结果。然而，同样明显的是，MSE措施对于某些实际上几乎没有感知影响的损害将严重失败。例如，空间或时间的微小变化、照明变化或上下文纹理的微小变化[21]都会导致MSE的大幅增加。Girod[22]全面概述了基于MSE的措施的局限性，王和博维克[21]也进行了出色的分析。他们很好地总结了情况，列出了MSE使用背后的假设：

1. 信号质量与样本之间的时间或空间关系无关。
2. 信号质量与原始信号和错误信号之间的任何关系无关。
3. 信号质量与错误信号无关。
4. 所有样品对信号质量的贡献均等。

基于VQEG FRTV第一阶段数据库的结果，张和公牛[42]分析了PSNR质量预测（非线性回归后）和主观微分平均意见评分（DMOS）之间的相关性。主要结果如图10.7所示，可以观察到，一般来说，这些点只松散地适合，在较低的质量水平（较高的DMOS）下，适合度会恶化。



下载：下载全尺寸图像

图10.7。DMOS与PSNR的散点图预测了VQEG数据库中各种类别序列的DMOS[42]。

张和布尔研究了内容如何影响基于PSNR的预测，根据主要内容类型（空间纹理、结构元素、纯亮度、动态运动或混合）将FRTV-I数据集细分为五组。这种分组是主观决定的，因为常规空间信息（SI）和时间信息（TI）措施不能提供足够的粒度。这些内容类的相关性如图10.7所示。可

以观察到，对于具有显著纹理内容的视频，大多数主观DMOS散点低于PSNR预测。这说明了静态和动态纹理内容的HVS遮罩效果的存在。相反，对于结构含量显著的序列，PSNR预测的DMOS值往往低于主观分数，这表明（尽管边缘遮罩效应）HVS对结构元素周围的错误非常敏感。高频能量很少的普通内容的散射图同样表明遮罩保护很少。

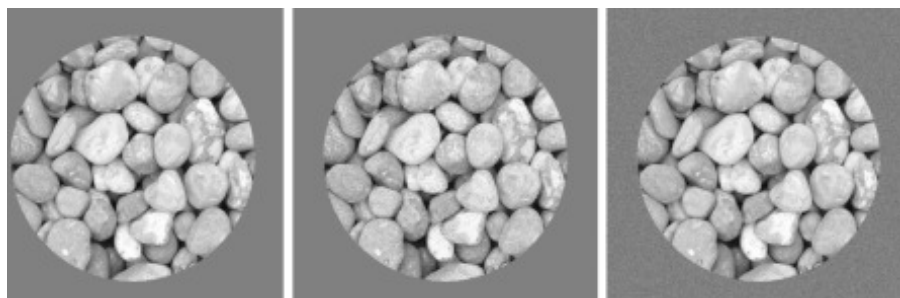
张和公牛[42]还根据不同类型的编码失真提出了PSNR预测的DMOS与DMOS散射图：压缩内容、传输错误压缩和内插值压缩。这些结果显示在图10.7的最后三个子情节中。PSNR在错误分布广泛的纯编码失真方面表现良好，但对插值错误和传输故障的情况预测要差得多。显然，在内容高度扭曲的情况下，PSNR无效，需要替代方法。

这一分析证实：

1. 视觉遮罩存在，在具有空间和时间纹理的内容中比普通亮度区域更明显。
2. 高可见的伪影，如平原地区或由于传输丢失的伪影，往往会导致HVS高估失真。
3. HVS使用两种不同的感知策略——近阈值和超阈值。显然，对于感知质量较低的案例，工件检测更重要。

### 示例10.1纹理遮罩和PSNR

考虑以下原始内容 $256 \times 256$ 图像（左）及其两个扭曲版本。在中间图像中，前景被方差0.003的加性高斯白噪声（AGWN）损坏。右手图像的背景损坏，AGWN的方差为0.003。两个损坏图像的PSNR值相同，前景和背景区域面积相等。



[下载：下载全尺寸图像](#)

直观地比较这些图像，并评论噪音的可见性（读者请参阅网站上的电子版）。

### 解决方案

失真图像的PSNR值为28分贝，因此在平均平方误差意义上，两者具有相同的失真。大多数观察者会发现背景的扭曲比前景的扭曲更明显，特别是在纹理内容与场景中上下文相关，并且在没有参考图像的情况下查看内容的情况。这表明具有不同感知影响的失真并不总是被PSNR措施区分开来。

#### 10.4.3. 度量发展的感知基础

任何视频质量指标的主要目的是在各种条件下和损伤下与视觉感知很好地关联。次要要求还可能包括公制复杂性和公制延迟，特别是与编解码器或传输系统内的在线决策相关。



在编码和质量评估中，对比度敏感度和视觉遮罩等HVS特性已被利用。例如，脑脊液加权已被用于反映空间和[时间频率](#)范围内的HVS敏感性，视觉遮罩——减少某些亮度和时空纹理的[感知畸变](#)——也被利用。遮罩效应在空间纹理区域更明显，在那里，HVS可以比在平滑的区域承受更大的扭曲。动态纹理也存在类似的现象。

当利用HVS属性时，通常发现与MSE等常规失真措施相比，该指标与主观判断提供了增强的相关性。有人指出，当扭曲的内容与原始内容相似时，这种改进会更显著[\[24\]](#)。

当失真变得显著时，视觉注意力更有可能被可见的工件吸引。在这种情况下，描述工件通常比纯粹计算失真更有效。基于这一概念，开发了质量指标，使用两步策略模拟感知过程：[近阈值](#)和[超阈值](#)。这类指标的例子包括Kayargadde等人报告的例子。[\[25\]](#)，卡鲁纳塞克拉等人。[\[26\]](#)，Carnec等人。[\[27\]](#)，钱德勒和赫马米（VSNR）[\[28\]](#)，拉尔森和钱德勒（MAD）[\[29\]](#)。

#### 10.4.4. 基于感知的图像和视频质量指标

在过去十年左右的时间里，提出了几个感知启发的客观质量评估指标。这些经常利用频率和空间领域的视觉统计和特征。其中一些在下面描述。有关更多详细信息，请参阅参考文献。[\[1\]](#)，[\[30\]](#)，[\[31\]](#)，[\[32\]](#)，[\[33\]](#)。

#### VQM

平森和沃尔夫的VQM[\[34\]](#)是一种客观的视频质量评估方法，可以密切预测主观质量评级。它基于损伤过滤器，结合了模糊、颠簸、全局噪声、块失真和颜色失真措施。VQM根据不同的质量特征计算七个参数。这些是通过过滤受损视频和参考视频来提取感兴趣的财产获得的。然后提取时空特征，并通过比较过滤的原始和受损视频区域的统计数据来获得该特征的质量参数。

VQM中使用的参数是 $si\_loss$ （模糊）、 $hv\_loss$ （边缘移位）、 $hv\_gain$ （阻塞）、 $si\_gain$ （边缘锐化）、 $chroma\_spread$ （颜色分布）、 $chroma\_extreme$ （局部颜色损伤）、 $ct\_ati\_gain$ （时空损伤）。VQM形成这些[线性组合](#)，以产生最终的质量评估。在2004年的VQEG测试中，VQM在预测MOS分数方面表现出了与所有其他被评估算法相比的优越性能。它随后被采纳为ANSI和国际电联的标准。

#### SSIM

图像或视频中结构信息的完整性是视觉感知的重要线索。Wang等人[\[35\]](#)开发了一种[图像质量评估](#)方法SSIM（结构相似性图像指标），该方法根据参考图像和失真图像之间的局部信息的统计特性来估计结构相似性退化。这是之前通用图像质量指数（UIQI）的改进版本，结合了基于亮度、对比度和结构的三种局部相似性测量。经过一些简化后，这三个项可以像方程 [\(10.13\)](#) 一样重新排列，以给出受损图像之间的结构相似性 $s$ 及其参考文献 $s_R$ 冒号：

$$SSIM = \frac{(2\mu_s\mu_{s_R} + C_1)(2\sigma_{ss_R} + C_2)}{(\mu_s^2 + \mu_{s_R}^2 + C_1)(\sigma_s^2 + \sigma_{s_R}^2 + C_2)} \quad (10.13)$$

在哪里 $\mu$ 和 $\sigma$ 是两张图像的局部平均值和标准偏差， $C_1$ 和 $C_2$ 是在弱分母存在下用于稳定方程的常数，以及 $\sigma_{ss_R}$ 是样本的交叉相关性（零平均值）。传统 $C_1 = (k_1 L)^2$ 和 $C_2 = (k_2 L)^2$ 在哪里 $k_1 = 0.01, k_2 = 0.03$ ，以及 $L$ 是像素值的动态范围。



SSIM值的范围在-1和+1之间扩展，如果两张图像相同，则仅等于1。通常使用滑动窗口计算SSIM是常规的，通常尺寸为 $11 \times 11$ ，使用圆对称的高斯加权函数。然后将整个图像的SSIM值计算为所有单个窗口结果的平均值。

SSIM的优势在于，在许多情况下，它比PSNR性能更好，而且实现起来相对简单。因此，在编写本报告时，它可能是最常用的非MSE指标。然而，人们已经认识到，SSIM存在一些问题，特别是它对相对缩放、翻译和旋转很敏感。为了解决这些问题，已经开发了一种复杂的小波方法，CW-SSIM[36]以及增强的多尺度版本（MS-SSIM）[37]。SSIM的进一步扩展称为V-SSIM，它也考虑了时间信息[38]加权所有帧的SSIM指数。与VQEG FRTV第一阶段数据库上的PSNR相比，该指标的性能有所提高。

有趣的是，在[示例10.1](#)中，尽管两种失真图像的PSNR值相同，但右手图像的SSIM分数为0.81，中间图像为0.95。这表明SSIM在纹理掩蔽的背景下更好地反映了主观意见。

## 电影

场景运动是影响视觉感知的关键组成部分，它可以提供相当程度的人工掩蔽。然而，感知质量也取决于观察行为；例如，观察者是否跟踪移动物体。Seshadrinathan等人[13]引入了运动调谐时空质量评估方法（MOVIE）。MOVIE使用时空Gabor过滤器家族分析失真内容和参考内容，质量指数由空间质量组件（受SSIM启发）和基于运动信息的时间质量组件组成。

基于MOVIE的预测的准确性已使用VQEG FRTV第一阶段数据库进行了评估，该数据库证明与PSNR和SSIM相比，该数据库提供了显著的相关改进。它还在LIVE视频数据库上展示了出色的性能。然而，MOVIE的一个问题是其高计算复杂性，因为使用了大量Gabor滤波器及其计算所需的时间范围。

## VSNR

钱德勒和赫马米在其视觉信噪比（VSNR）静止图像度量[28]中利用了HVS的近阈值和超阈值属性。这模拟了使用小波变换的HVS的皮层分解。然后采用两阶段方法来评估失真的可检测性，并确定视觉信噪比的最终测量值。VSNR已使用LIVE图像数据库进行了评估，结果非常好。

## MAD和STMAD

基于VSNR使用的方法，拉尔森和钱德勒[29]开发了最明显的失真模型（MAD）。MAD模型包括近阈值畸变和基于外观的畸变。它对高质量图像（接近阈值失真）和低质量图像（超阈值失真）采用不同的方法。这些使用非线性模型组合起来，以获得最终的质量指数。

MAD被扩展为应对参考文献[39]中的时空成分，其中根据空间MAD处理时空切片，然后使用运动信息对结果进行加权。时间MAD索引与空间MAD（从单个帧计算）相结合，以获得用于视频的时空MAD（ST-MAD）。基于LIVE视频数据库，报告了与主观结果的良好相关性。

## VDP和VDP-2

在VDP-2 Mantiuk等人[40]中，基于之前的VDP指标，创建一个与强度范围延长或缩小的内容的主观意见密切相关的指标。VDP-2指标预测误差可见性和质量（MOS），并基于新的对比度敏感

性测量。该模型使用LIVE（图像）和TID2008图像数据库进行了验证，并展示了与其前身HDR-VDP和VDP指标相比的性能更好的性能，特别是在低亮度条件下。它还与MS-SSIM指标进行了比较。

## 降低复杂性指标和循环评估

目前的图像和视频质量评估方法是根据各种HVS特征开发的，以实现接近主观意见的质量预测。这些方法提供了与主观分数的增强相关性，但往往因高复杂性或高延迟而处于不利地位，因此不适合实时操作。

最近，Soundararajan和Bovik提出了一种基于时空熵差异的降低参考视频质量指标[19]。这表现良好，通过使用简化时空内容来支持低复杂度操作。

在感知视频编码的背景下，张和布尔提出了一种基于人工的视频指标（AVM）[23]，使用DT-CWT作为评估常规压缩和合成内容的基础。AVM与主观VQEG分数关系良好，其优点是广泛的参数重用，其灵活性高，复杂性低，因此可以轻松集成到基于合成的框架中。

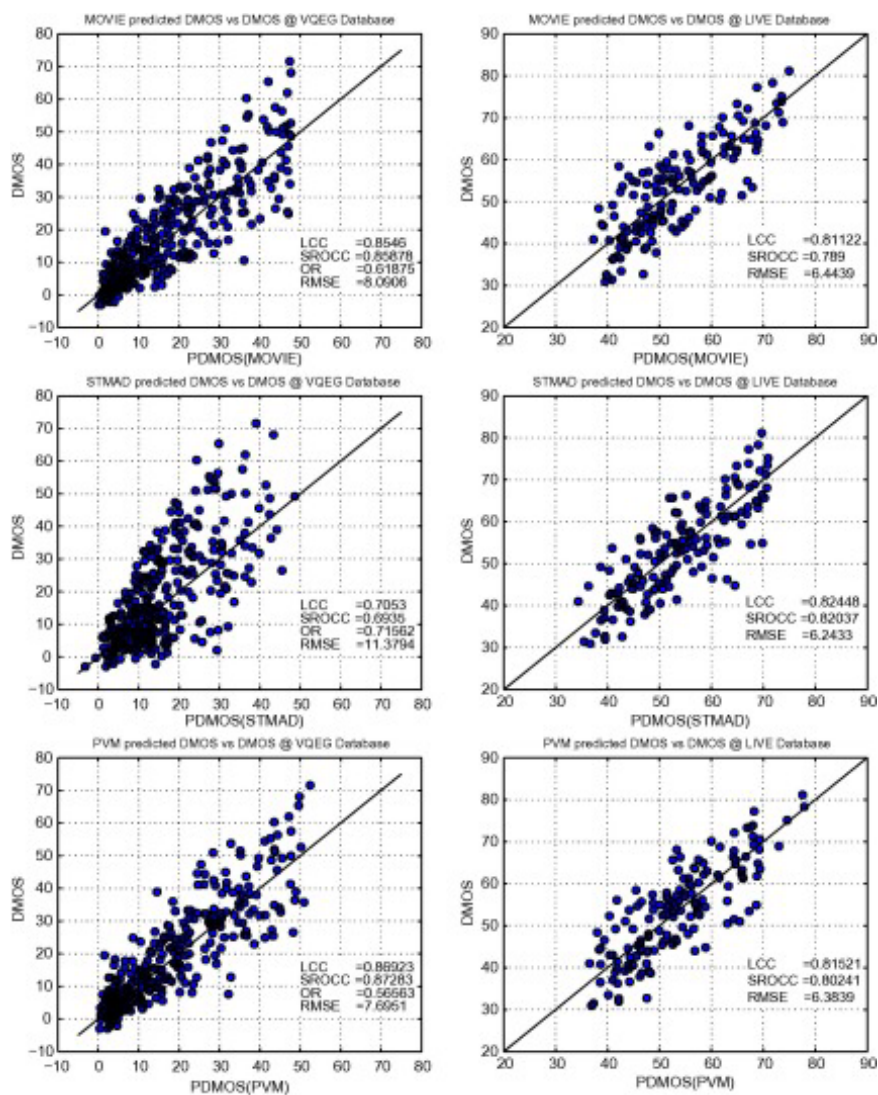
## PVM

张和公牛最近在他们的AVM模型的基础上提出了基于感知的视频质量指标（PVM）。PVM通过使用增强的非线性模型自适应地结合明显的失真和模糊伪影（两者都利用了双树复小波变换的移位不变性和定向选择性属性）来模拟HVS感知过程。显著的失真通过使用空间和时间掩码来定义绝对差异的，这些掩码是纹理掩码效果的特征，当失真视频与原始视频相似时，这对质量评估做出了重大贡献。通过计算高频能量变化估计并按运动速度加权的模糊工件被发现，当它与明显的失真相结合时，可以在低质量情况下提高整体公制性能。

重要的是，PVM与其前身AVM一样，旨在与合成和常规编码材料一起使用。早期结果表明，这些是唯一能够在这方面取得稳健性能的指标。

## 比较结果

将PVM、MOVIE和ST-MAD的性能作为图10.8中的散点图进行了比较，用于LIVE视频和VQEG FRTV第一阶段数据库。每个图表上都给出了相关系数、异常值和RMSE的值。可以看出，PVM在两个数据库中更加一致，并在VQEG数据库的上下文中提供了卓越的性能。



下载：下载全尺寸图像

图10.8。LIVE和VQEG FRTV数据库中MOVIE、ST-MAD和PVM的比较[42]。

#### 10.4.5. 指标的未来

虽然感知指标在过去十年左右显然取得了显著进展，但它们仍然没有被广泛接受为一般图片质量评估。这有很多复杂的原因。例如，它们主要测量视频保真度（受损版本与原始版本的接近程度），它们通常测量退化（即假设受损版本比原始版本更差），并且通常不考虑观看模式（观众倾向于关注感兴趣的区域）。

未来最具挑战性的问题之一是创建一个可靠但低复杂性的循环内质量评估措施，以准确评估主观质量并检测任何值得注意的编码伪影。MSE方法目前被广泛用于此，但随着基于合成的编码变得更加普遍，MSE方法将不合适。

对于实时或循环处理，质量指标应该能够：

1. 在不同空间级别（如共和党、图片、区域或编码单元）进行评估。
2. 提供可管理的计算复杂性。

3. 区分扩展视频参数空间的影响，包括更高的动态范围、帧率和分辨率，并考虑不同的环境条件。
4. 理想情况下，与新兴[感知编码](#)和分析-合成[压缩方法](#)兼容（见第13章）。

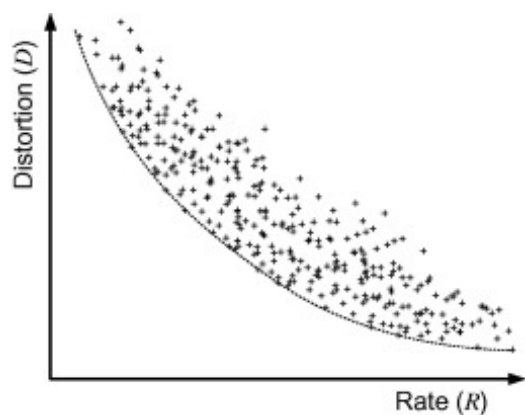
很少有现有指标符合这些标准。

## 10.5。速率-失真优化

在[第四章](#)对速率和失真的介绍的基础上，我们将在这里更详细地考虑压缩的这一重要领域。如前所述，任何压缩序列的比特率将取决于许多因素。这些包括：

- 视频内容（高时空活动通常需要更多的位来编码）。
- 使用的编码算法（例如，基于小波或DCT，仅内部或运动补偿）。
- 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、内部模式与模式间模式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和基于不良选择的编码之间的性能差异可能非常显著。

给定编解码器-序列组合的速率与失真（或速率与质量）特性为比较编解码器和评估参数选择如何影响性能提供了一种方便的方法。对于一系列不同的参数值，操作点图可能看起来像[图10.9](#)。目的是确保任何时候选择的参数是那些在任何给定编码速率下产生尽可能低失真的参数。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10.9。用于各种编码参数选择的速率失真图。关联的凸壳为给定的编解码器-源组合提供了操作的R-D特性。

为了进一步阅读该主题，奥尔特加和拉姆钱德兰[\[43\]](#)以及沙利文和维冈德[\[44\]](#)对速率失真理论和实践进行了出色的评论。Wang等人[\[45\]](#)也很好地区分了R-D界限。有关当前标准的最新 RDO 工作，请参阅 Refs. [\[46\]](#)，[\[47\]](#)。



### 10.5.1. 经典速率-失真理论

速率失真理论是源代码编码的核心。它指的是源保真率和编码率之间的权衡，可以简单地表述为：

1. 表示给定质量约束的最小比特数的源，或者，
2. 实现给定比特率的最佳质量。

香农的分离原则指出，对于通过噪声信道传输数据（在某些假设下），如果源和信道编码独立处理，带宽的最佳使用结果。换句话说，我们应该首先在无错误的环境中为源获得最有效的数据表示，然后添加适当的保护级别，以确保在有损通道上无错误地交付。正如我们在第九章中已经讨论过的，基本假设并不总是适用，但总的来说，这是一个不错的起点。

速率失真理论通常用于为给定类型的源建立性能界限（以比特率和质量），通常根据其信号统计来描述。在这方面面临的一个问题是，简单的统计特征通常不适合复杂的非固定来源，如视频。

即使在优化标量量化器的简单情况下，我们仍然需要为源选择合适的统计模型。只有这样，我们才能调查实用的算法，并评估它们与该分布的界限有多接近。这种方法有两个问题：

1. 如果模型很差，那么界限就毫无意义，实用或启发式设计可能远远超过它们。Ortega和Ramchandran[43]提供了一个低复杂度小波图像编码器优于基于i.i.d的编码器的例子。高斯模型超过3分贝。
2. 即使模型良好，实际解决方案也可能不切实际地复杂，或需要巨大的延迟。例如，参考文献[43]中给出，小波编码器的复杂度相对较低，而“R-D最佳”编码器的复杂度是无限的。

#### 失真措施

设编码源符号之间的失真 $\tilde{s}$ 及其原始版本 $s$ 表示为 $d(s, \tilde{s})$ 。在 $s = \tilde{s}$ 然后 $d(s, \tilde{s}) = 0$ 。源的平均失真由以下方式给出：

$$D = d(s, \tilde{s}) = E\{d(s, \tilde{s})\} = \sum_s \sum_{\tilde{s}} P(s, \tilde{s}) d(s, \tilde{s}) \quad (10.14)$$

通常使用的失真度量是平方误差， $d(s, \tilde{s}) = (s - \tilde{s})^2$ ，或绝对错误， $d(s, \tilde{s}) = |s - \tilde{s}|$ 。源的平均失真 $N$ 对于平方误差失真，样本是MSE：

$$d(s, \tilde{s}) = \frac{1}{N} \sum_{n=1}^N d_n(s, \tilde{s}) = \frac{1}{N} \sum_{n=1}^N (s[n] - \tilde{s}[n])^2 \quad (10.15)$$

更常见的是，在实际的RDO应用程序中，使用平方差分之（SSD），该和与方程（10.15）相同，但省略了 $\frac{1}{N}$ 学期。

#### 失忆的高斯源

正如我们上文所讨论的，R-D边界的闭形式解通常很难获得。然而，对于具有方差的高斯i.i.d.源 $\sigma^2$ ，边界可以计算（假设高比特率），并由以下方式给出：



$$\bar{R}(D) = \begin{cases} \frac{1}{2} \log_2 \left( \frac{\sigma^2}{D} \right) & 0 \leq D < \sigma^2 \\ 0 & D \geq \sigma^2 \end{cases} \quad (10.16)$$

在哪里 $R(D)$ 表示速率-失真函数（给出比特率） $R$ 需要造成扭曲 $D$ 对于给定的来源）和 $\bar{R}(D)$ 是速率-失真界限（即无限长度向量的所有可能编码器之间的最小速率）。

在实践中，对于更一般和复杂的来源，如自然视频，理论研发界限几乎没有作用，必须采取更**务实的方法**。接下来要考虑这些。

### 10.5.2. 实用的速率-失真优化

源编码器的操作控制是视频压缩中的一个主要问题，实用RDO的目的是为当前操作条件选择最佳编码模式和参数设置。因此，可用比特预算的分配方式必须根据给定的速率约束，将总体失真尽可能保持在低水平。现代编解码器的参数设置将涵盖广泛的编码器属性，如块大小、运动估计引用、模式间或模式内选择以及**量化步骤大小**。由于视频内容的非固定性，优化问题特别困难，其中最佳编码器设置可能因不同的时空区域而异。此外，并非所有编码单元的速率失真成本通常独立（例如，由于预测的影响）。

需要注意的是，特别是对于较低的比特率，独立考虑**量化**和熵编码是不切实际的。与较高的比特率和质量设置不同，在较高的字母符号数量上，例如，对于较低的比特率的情况下，选择一个具有略高失真的量化点可能是有利的，因为它会导致显著较低的比特率。

### 从源统计到可参数化编解码器

实用的速率-失真优化方法通常不会从源模型开始。相反，它们基于给定参数化编码-解码组合的假设，如果选择正确的参数，已知该组合性能良好。在这种情况下，参数选择不是基于源统计，而是基于源数据的某些转换（例如DCT系数或小波子带系数）。这样做的好处是使建模更健壮，因为转换数据的统计信息往往更容易建模。例如，子带系数可以大致近似于用i.i.d建模的**无内存源**。拉普拉斯过程。

然而，如果子带系数确实是无记忆的，我们可以使用一阶熵根据这个假设对其进行编码。在实践中，该假设不够好，因为系数在量化后扫描时，通常呈现长链的零。因此，我们需要一些技巧，例如我们在**第7章**中探索的之字形运行值编码。

RDO方法必须根据最优**参数向量**找到最佳操作点， $\mathbf{p}_{\text{opt}}$ 。如果我们在速率范围内为源（例如图像）的给定实现计算所有参数组合，那么我们可以获得该编码器-序列组合的最佳速率-失真曲线。如**图10.9**所示，凸壳表示任何实用的RDO启发式算法都可以测量的操作界限。

### RDO复杂性

值得快速提及的是，RDO过程的复杂性。在离线情况下，单个编码将被多次重用，值命题将证明更复杂或多通编码是合理的（一个很好的例子是蓝光磁盘的编码），而在其他情况下，它可能不会。

### 拉格朗日优化

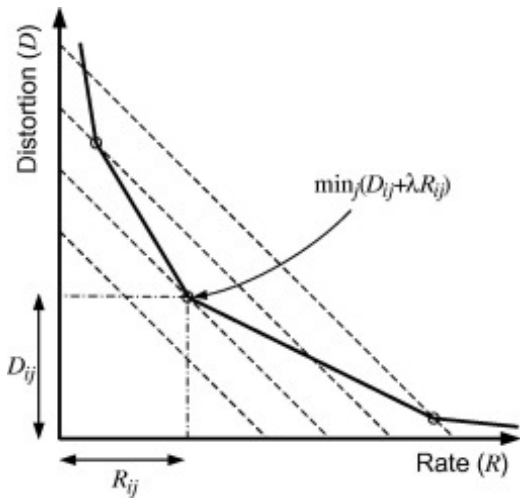
**拉格朗日方法**在最近的标准中被广泛接受为RDO的首选方法，主要是因为它们的有效性和简单

性[46], [48]。最初, 我们将考虑每个编码单元的情况,  $i$  (基本优化构建块), 可以独立于所有其他优化, 并且唯一优化参数是量化索引,  $j$ 。虽然在调用上下文依赖性预测的情况下, 这个假设显然会崩溃, 但它可以提供有用且易于处理的解决方案。

通常, 对于RDO, 采用离散拉格朗日方法, 如肖姆和格肖[43], [49]所引入的, 以便在不存在封闭形式解的情况下, 在外部约束下, 找到函数的最大值或最小值。编码单元的拉格朗日成本函数 $J_{ij}$ 由方程 (10.17) 给出:

$$J_{ij}(\lambda) = D_{ij} + \lambda R_{ij} \quad (10.17)$$

其中量化指数 $j$ 规定速率和失真与拉格朗日乘数之间的权衡 $\lambda$ 控制R-D平面上与R-D特性相交的线的斜率, 以选择特定的操作点。这如图10.10所示。



[下载: 下载全尺寸图像](#)

图10.10。编码单元的拉格朗日优化,  $i$ 。

如果我们最小化方程 (10.17), 那么我们得到:

$$\frac{\partial(J_{ij}(\lambda))}{\partial R_{ij}} = \frac{\partial(D_{ij} + \lambda R_{ij})}{\partial R_{ij}} = \frac{\partial D_{ij}}{\partial R_{ij}} + \lambda_{\text{opt}} = 0 \quad (10.18)$$

这意味着最佳值 $\lambda$ 由失真函数的负斜率给出:

$$\lambda_{\text{opt}} = -\frac{\partial D_{ij}}{\partial R_{ij}} \quad (10.19)$$

显然, 如果我们最小化 $J$ 对于 $\lambda = 0$ 我们正在尽量减少失真, 如果我们这样做, 当 $\lambda = \infty$ 我们把费率降到最低。然而, 一般来说, 斜率通常不为人所知, 因此我们必须使用下文所述的操作技术对其进行估计。应当指出, 拉格朗日方法被限制在R-D特征的凸壳上选择操作点。在某些情况下, 可能存在更可取的非最优点。这可以通过动态规划等替代但可能更复杂的技术来解决[43]。

### 10.5.3. 附加编码模式和参数的影响

现代编码标准提供了高编码效率，这不仅是因为它们的基础架构，还因为它们对不同内容类型和信道约束的适应性。我们将在[第12章](#)中更详细地研究这一点，但改编通常基于以下部分或全部决定：

- 量化器控制。
- 何时使用块或图片跳过模式。
- [运动矢量](#)和多个参考框架指标的确定。
- 使用整数、半像素或四分之一像素运动估计。
- 模式内部或模式间和相关块分区的确定。例如，在H.264/AVC中，该标准使用 $16 \times 16$  lumamacroblock具有以下选项：
  - 内部模式（适用于所有帧）：九种 $4 \times 4$ 模式和四种 $16 \times 16$ 模式。
  - 跨模式（仅适用于P帧和B帧）： $16 \times 16, 16 \times 8, 8 \times 16, 8 \times 8, 8 \times 4, 4 \times 8, 4 \times 4$ 。

每个模式都有相关的速率-失真成本，编码器必须选择成本最低的模式。当然，由于可选模式对不同内容类型和不同比特率将具有不同的效率，这当然变得复杂。优化问题更加复杂，因为并非所有编码单元的速率失真成本通常都是独立的；例如，因为通过内部和相互预测引入的空间和/或时间依赖关系。这意味着方程的拉格朗日优化 ([10.17](#)) 理论上应该对视频序列中的所有编码单元联合执行，在大多数情况下，这非常复杂。因此，提出了一些简化技术。

也许最明显的参数选择或模式决策方法是使用[详尽搜索](#)。通过比较所有可能的选项和参数组合，可以选择最佳选项。这会非常昂贵，所以通常使用替代品。基于指示性阈值可以制定更简单的决策策略。例如，关于是否应选择内部模式或模式间模式的决定可以基于对 $16 \times 16$ 关于其平均值的宏块，以及整数运动搜索后的最小SAD。其他决定也报告了类似的策略。然而，RDO的主要选择方法是使用拉格朗日乘数方法，这些方法通常分别应用于模式选择和运动估计问题。

## 重新访问拉格朗日乘数

按照上述方法和参考文献[\[46\]](#)所述，目的是解决受限问题：

$$\min_{\mathbf{p}} D(\mathbf{p}) \text{ s. t. } R(\mathbf{p}) \leq R_T \quad (10.20)$$

在哪里 $R_T$ 是[目标比特率](#)， $\mathbf{p}$ 是编码参数的矢量。和以前一样，这可以表示为基于以下成本函数的无约束[拉格朗日公式](#)：

$$\mathbf{p}_{\text{opt}} = \arg \min_{\mathbf{p}} \{D(\mathbf{p}) + \lambda R(\mathbf{p})\} \quad (10.21)$$

**模内：**我们先考虑编码内模的模式选择问题。如果我们假设量化器值 $Q$ 已知，拉格朗日参数 $\lambda_{\text{MODE}}$ 给定参数（或编码模式）选择由以下方式给出：

$$\begin{aligned} J_{\text{MODE}}(\mathbf{S}_k, \mathbf{p}_k \mid Q, \lambda_{\text{MODE}}) \\ = D(\mathbf{S}_k, \mathbf{p}_k \mid Q, \lambda_{\text{MODE}}) + \lambda_{\text{MODE}} \\ \cdot R(\mathbf{S}_k, \mathbf{p}_k \mid Q, \lambda_{\text{MODE}}) \end{aligned} \quad (10.22)$$

参数向量 $\mathbf{p}$ 具有给定编码单元（或子图像）的所有可能编码模式的变化， $\mathbf{S}_k$ 。如前所述，失真度量 $D$ 通常基于原始块与其重建版本之间的平方和差异以及速率 $R$ 通常通过熵编码后测量。

**跳过模式：**对于计算跳过模式的情况，失真和速率计算不取决于当前量化器值，而只是取决于当前编码单元与使用推断运动矢量进行预测产生的SSD。在H.264中，跳过块的速率约为每个编码单元1位。

**模式间：**对于模式/间的情况，让我们假设我们知道拉格朗日参数 $\lambda_{\text{MOTION}}$ 和解码的参考图片 $\tilde{\mathbf{S}}$ 。速率约束的运动估计可以通过最小化成本函数来执行如下：

$$\mathbf{p}_{\text{opt}} = \arg \min_{\mathbf{p}} \{ D_{\text{DFD}}(\mathbf{S}_k, \mathbf{p}_k) + \lambda_{\text{MOTION}} \cdot R_{\text{MOTION}}(\mathbf{S}_k, \mathbf{p}_k) \} \quad (10.23)$$

在这种情况下使用的失真度量通常是SAD或SSD，由方程 (10.24) 定义 $l = 1$ 或者 $l = 2$ 分别：

$$D_{\text{DFD}}(\mathbf{S}_k, \mathbf{p}_k) = \sum |s[x, y, z] - \tilde{s}[x - d_x, y - d_y, z - d_z]|^l \quad (10.24)$$

在哪里 $\mathbf{d}$ 是当前参数集的运动矢量。运动速率计算必须包括编码运动信息所需的所有位，通常基于熵编码预测。

Wiegand等人[44], [46]展示了拉格朗日乘数方法在块大小选择、运动估计和量化器控制方面的强大作用。在量化器控制的语境中，他们证明了量化器步长与 $\lambda$ 。他们表明，对于一阶近似值来说，这种关系与内容无关，并且（在H.263的上下文中）由以下给出：

$$\lambda_{\text{MODE}} = 0.85 Q_{\text{H.263}}^2 \quad (10.25)$$

拉格朗日方法自H.263引入以来一直被广泛使用，并在H.264/AVC和HEVC的测试模型中采用。对于H.264/AVC，对于I和P帧，根据经验确定了不同的表达式：

$$\lambda_{\text{MODE(I,P)}} = 0.85 \cdot 2^{(QP_{\text{H.264}} - 12)/3} \quad (10.26)$$

对于B帧：

$$\lambda_{\text{MODE(B)}} = \max\left(2, \min\left(4, \frac{QP - 12}{6}\right)\right) \cdot \lambda_{\text{MODE(I,P)}} \quad (10.27)$$

两者之间的关系 $\lambda_{\text{MODE}}$ 和 $\lambda_{\text{MOTION}}$ 由以下方式给出：

$$\begin{cases} \lambda_{\text{MOTION}} = \lambda_{\text{MODE}} : & \text{for SSD} \\ \lambda_{\text{MOTION}} = \sqrt{\lambda_{\text{MODE}}} : & \text{for SAD} \end{cases} \quad (10.28)$$

#### 10.5.4. 从速率-失真到速率-质量优化

本节中描述的技术基于两个假设：（i）编码过程的目的是尽可能紧密地表示图片样本值，以及（ii）MSE（或其他 $L_p$ 规范）准确反映了被编码时空区域的感知质量和重要性。在实践中，我们看到，尽管第二个假设明显被违反，但MSE可以为给定内容的给定编码策略中的参数变化提供单调和一致的指标[20]。

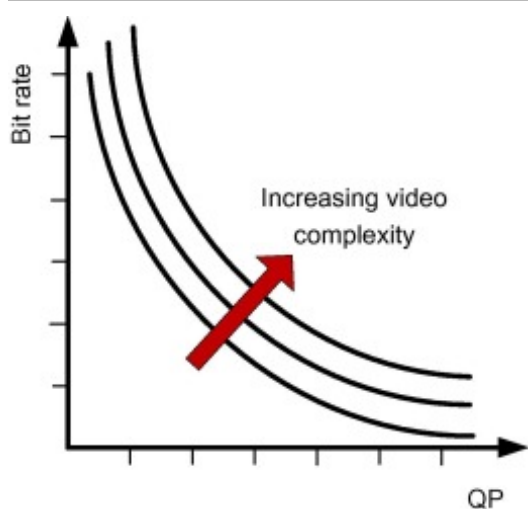
在第一个假设的背景下，基于纹理分析和合成模型的新感知启发编码策略正在出现（见第13

章)。这些的目的不是最小化原始版本和编码版本之间的距离，而是获得一个感知上可信的解决方案。在这种情况下，MSE不太可能提供任何有意义的感知质量指标，也不再是一个有效的客观功能。因此，重点必须从速率-失真优化转向速率-质量优化，要求新的嵌入式感知驱动但低复杂性的质量指标[23]。

## 10.6。费率控制

速率控制是实用视频编解码器的重要组成部分，因为它确保以与当前内容复杂性和频道容量兼容的速度将编码视频交付到频道。目前，将特定编码模式的影响放在一边，影响编码视频比特率的主要问题是：

1. 对I-picture和P-或B-picture的不同要求（I-picture可以消耗P-或B-picture比特数的5-10倍）。
2. 源复杂性的影响。与低活动内容相比，时空活动量高的内容通常需要更多比特。图10.11描述了这一点。



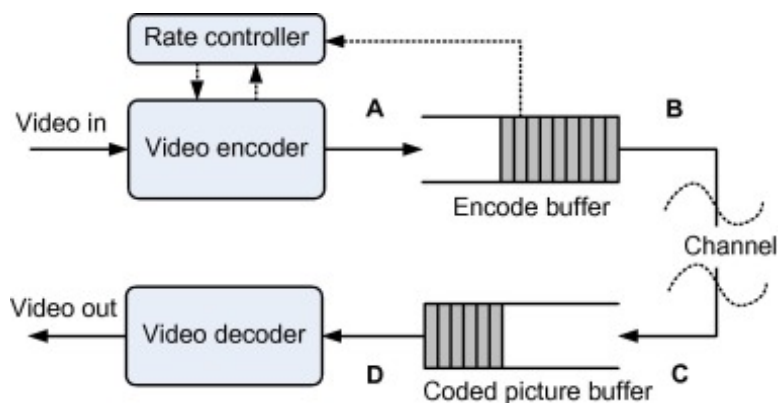
[下载：下载全尺寸图像](#)

图10.11。在源复杂性变化的背景下管理比特率。

因此，如果量化器保持不变，那么输出比特流的峰值平均比可能会过高，从而有效地使用带宽。峰值输出比特率也可能超过最大信道容量。

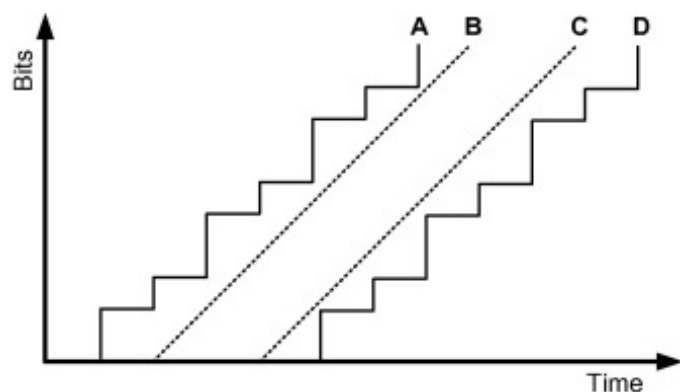
即使使用速率控制，比特率波动也是不可避免的，缓冲总是用于平滑这些变化。通过缓冲区对编码器与信道、信道和解码器之间的接口的简单描述如图10.12所示。此图显式包括一个速率控制器，用于缓和编码数据流到编码缓冲区。这通常通过速率量化模型来实现，该模型根据缓冲区占用率和内容复杂性的动态测量向编码器提供量化器参数（QP）。如果我们考虑缓冲操作和通道导致的延迟，我们可以在图10.13中表示此类配置的播放时间表。





[下载：下载全尺寸图像](#)

图10.12。编码器和解码器的传输缓冲。



[下载：下载全尺寸图像](#)

图10.13。根据图片逐张流式播放时间表，对应图10.12中标记的点。

在大多数标准中，上述缓冲区和相关速率控制机制是非规范的。然而，为了确保编码器正确执行速率控制，最近的标准纳入了假设参考解码器（HRD）的概念。这用于在缓冲和位消耗方面模拟理想化解码器行为。没有速率控制，解码器缓冲区将定期下溢或溢出，导致播放抖动或丢失。

速率控制和RDO共同呈现一个困境：为了对编码单元执行RDO，必须知道量化参数，这通常基于SAD或方差度。然而，编码单元的实际SAD要到RDO过程完成后才能获得；因此有必要估计当前编码单元的SAD。除此之外，我们还需要计算当前帧的目标位，这很复杂，因为编码单元模式和MV信息等头信息在RDO进程完成之前不可用。

### 10.6.1. 缓冲和人力资源开发

比特率分配通常与缓冲区模型相关联，缓冲区模型在视频编码标准中指定为假设参考解码器（HRD）的一部分。HRD通常是标准的规范组成部分，因为它在规定对兼容比特流的要求方面

很重要。

HRD是解码器的理想表示，在概念上与编码器的输出相关。它包括解码器缓冲区以及解码器本身。缓冲区通常使用泄漏的桶方法建模，该方法仅提供进出通道的**固定比特流**。位被假定以恒定的速度流入解码器缓冲区（有时称为**编码图片缓冲区**（CPB）），并假定解码器以图片大小的块立即从中提取。这是图10.12，图10.13中描述的场景，在信道中没有丢失的情况下，接收端的缓冲区内容只是反映了编码器的内容。因此，符合HRD的比特流必须由CPB处理，不得溢出和下溢。这一要求必须通过编码器中实现的速度控制来满足。

HRD缓冲模型通常由三个参数描述，（i）输出速率 $R_o$ （假设为常数），初始占用率， $F_i$ （即假定开始解码的占用级别）和缓冲区大小， $B$ 。费率控制器必须跟踪占用情况 $F$ 假设缓冲区并在其范围内调整编码器参数，以避免缓冲区下溢和溢出（见图10.14）。

**示例10.2**缓冲和人力资源开发参数

考虑一个24 fps的恒码率视频传输系统的运行，参数如下：

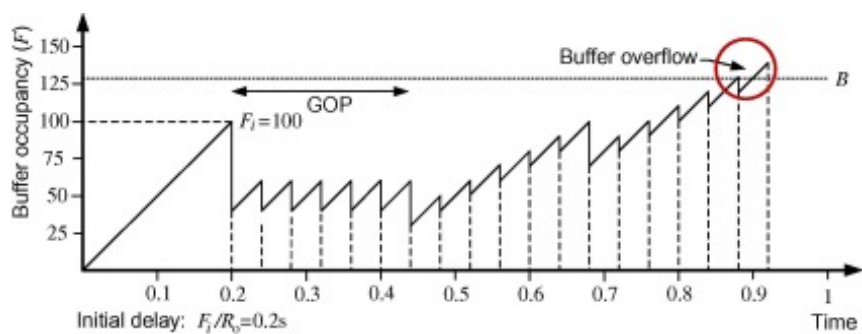
$R_o = 500 \text{ kbps}$ ;  $B = 128 \text{ kb}$ ;  $F_i = 100 \text{ kb}$ ;  $\text{GOP} = 6 \text{ frames}$

如果传输的初始图片有以下大小，计算解码器缓冲区的占用情况，并确定是否发生溢出或溢出。

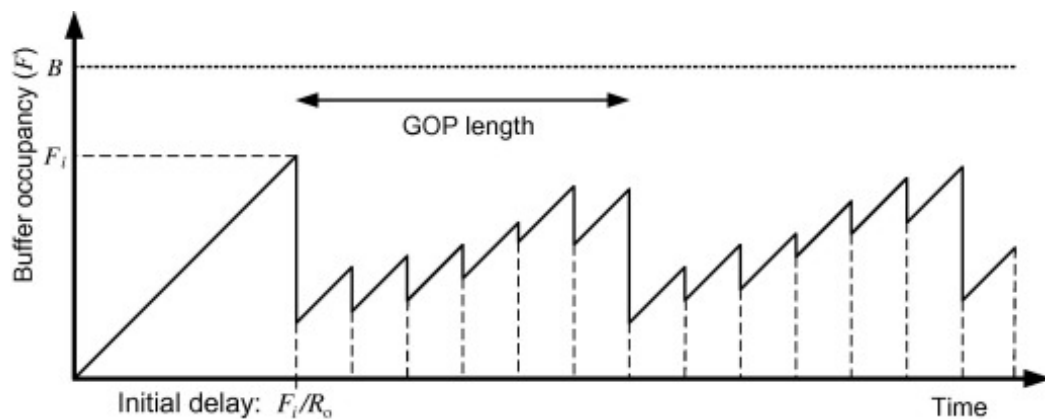
| 帧不。   | 图片大小（kbits） |
|-------|-------------|
| 1     | 60          |
| 2-6   | 20          |
| 7     | 30          |
| 8-12  | 10          |
| 13    | 30          |
| 14-18 | 10          |
| 19    | 50          |
| 20-24 | 30          |

**解决方案**

缓冲区的提升时间是 $F_i/R_o = 0.2 \text{ s}$ 。绘制这些初始图片1s周期的缓冲区占用情况，我们可以观察到当图片18和19传输时解码器缓冲区溢出：



下载：下载全尺寸图像

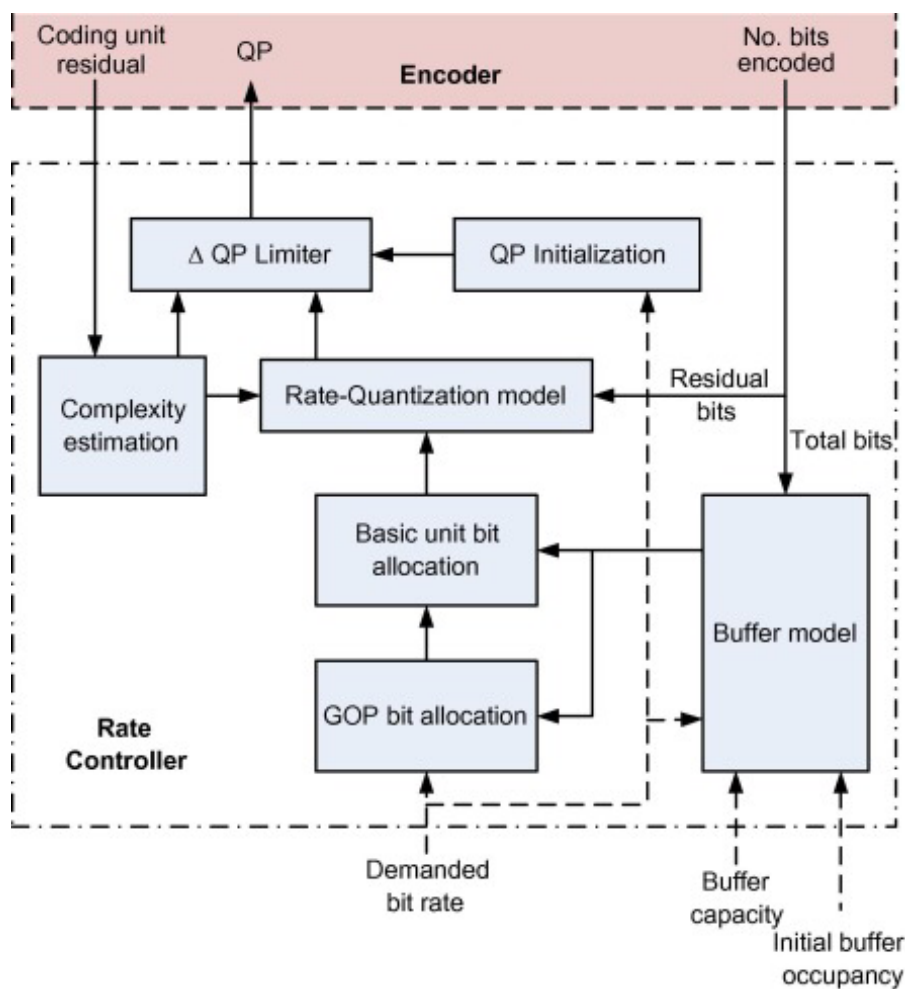


下载：下载全尺寸图像

图10.14。示例解码器缓冲定时，指示HRD参数。

## 10.6.2。实践中的费率控制

图10.15显示了通用速率控制架构。虽然简化了，但这与MPEG-2、H.264/AVC和HEVC中使用的有很多相似之处。基本要素如下。



[下载：下载全尺寸图像](#)

图10.15。通用费率控制器。

## 缓冲模型

缓冲区模型表示第10.6.1节中描述的假设参考解码器。它根据其大小和初始满度参数化，并根据信道的比特率进行进一步输入。它提供了当前占用率的输出， $F$ ，到下面描述的位分配块。

## 复杂性估计

为了控制位分配过程，必须估计编码剩余物的复杂性。源复杂性的估计通常基于剩余数据（因为这是量化器的影响）。平均绝对差分（MAD）、绝对值和（SAD）、平方差和（SSD）或绝对变换差分（SATD）[50]等措施通常用于此目的。

## 速率量化模型

速率量化模型反映了图10.11所示的可变性，描述了当编码不同复杂性的内容时，可用位与量化器值的关系。

## ΔQP限制器

为了避免QP值的过度变化，这可能会在速率控制器中引入明显的工件或导致振荡行为，QP值的变化通常仅限于一个小增量，通常 $\pm 2$ 。

## QP初始化

在编码之前，QP值必须设置为某个值。根据所需的比特率和视频格式，每个像素的比特率的平均值可以很容易地计算出来，因此：

$$\text{bpp} = \frac{R_i(1)}{f \cdot X \cdot Y} \quad (10.29)$$

在哪里 $R_i(k)$ 是编码帧时的瞬时可用比特率 $k$ 共和党 $i$ ， $f$ 是帧率，以及 $X$ 和 $Y$ 是图片中的水平和垂直像素数。然后，这可以通过查找表与QP值相关联，例如参考文献中推荐的H.264/AVC查找表。[51]。

## GOP位分配

根据所需的比特率和缓冲区占用率，计算了共和党的目标比特分配。GOP级别的位分配通常初始化IDR图片和第一个P图片。H.264/AVC中用于此的方案如下[52]：

$$\text{QP}_i(1) = \max \left\{ \text{QP}_{i-1}(1) - 2, \min \left\{ \text{QP}_{i-1}(1) + 2, \frac{\sum_{k=2}^{N_{i-1}} \text{QP}_{i-1}(k)}{N_{i-1}-1}, -\min \left\{ 2, \frac{N_{i-1}}{15} \right\} \right\} \right\} \quad (10.30)$$

在哪里 $\text{QP}_i(k)$ QP值是分配给图片的 $k$ 共和党 $i$ 和 $N_i$ 是共和党的图片总数 $i$ 。如果 $\text{QP}_i(1) > \text{QP}_{i-1}(N_{i-1}) - 2$ ，进一步调整如下：

$$\text{QP}_i(1) \leftarrow \text{QP}_i(1) - 1 \quad (10.31)$$

在哪里 $\text{QP}_{i-1}(N_{i-1})$ 表示上一个P图的QP值。

## 编码单元位分配

这指的是将比特分配给较小的编码元素的过程，例如单个图片、切片或较小的编码单元。对每张图片的位分配基于目标缓冲区级、实际缓冲区占用率、所需的比特率，当然还有帧率。因此，每个剩余的P和B帧的QP值是根据共和党可用的剩余比特数分配的。分配过程还应该考虑当前图片是P型还是B型，因为这些图片可能有不同的位分配。H.264/AVC中使用的方案基于假设拉普拉斯源分布的二次模型，如下[52]：

首先，估计当前存储图片的MAD， $\tilde{\xi}_i(k)$ ，基于上一张图片的相应值，如方程 (10.32) 所示：

$$\tilde{\xi}_i(k) = a_1 \cdot \xi_i(k-1) + a_2 \quad (10.32)$$

在哪里 $a_1$ 和 $a_2$ 是预测模型的系数，初始化为1和0，并在每个单元编码后更新。接下来，使用方程 (10.33) 计算其余每张P图的量化器步长。

$$T_i(k) = c_1 \cdot \frac{\tilde{\xi}_i(k)}{Q_{\text{step},i}(k)} + c_2 \cdot \frac{\tilde{\xi}_i(k)}{Q_{\text{step},i}^2(k)} + m_{h,i}(k) \quad (10.33)$$

在哪里 $m_{h,i}(k)$ 是头位和运动矢量位的总数， $c_1$ 和 $c_2$ 是二次模型的系数，在处理每张图片后更新，



以及 $Q_{\text{step},i}(k)$ 是图片的量化器步长吗 $k$ ，对应目标比特数。为了计算方程 (10.33) 中的量化器步长，我们首先需要计算帧的目标比特数的值 $k, T_i(k)$ 。这通常根据缓冲区占用率和共和党其余部分剩余比特数的加权组合计算。有关此方案的更多详细信息，请参阅参考文献。[52]。

还提出了其他几种速率控制方案，以提高纹理和标头位的预测精度；例如，权等人的贡献。[50]。在HEVC的背景下，对上述方法进行了几项修改，读者请参阅Sanz-Rodriguez和Schierl[53]以及Li等人的贡献。[54]。

### 10.6.3. 利率和利率控制区域

在某些应用程序中，对于某些类型的内容，可以利用有关场景或其中区域的先验信息来提高编码效率或更好地管理比特率。例如，在许多监控视频中，可能有一个或多个感兴趣的区域（ROI）将受益于优先位分配。萨德卡[55]很好地讨论了其中涉及的权衡。

Agrafiotis等人[56]提出了一种改进的速率控制方法，该方法允许根据投资回报率特征和目标比特率定义多个优先区域，其质量各不相同。结果表明，编码灵活性增强，可以显著提高投资回报率的质量，并在不影响目标比特率的情况下，在不影响帧其余部分的质量上发生明显愉悦的变化。这项工作是在移动终端的低比特率手语编码背景下提出的，但它也可能在包括监控和医疗保健在内的许多其他应用中受益。

## 10.7. 摘要

本章涵盖了一系列测量和控制视觉质量的方法。我们首先描述了最常见的主观测试方法，并展示了这些方法如何提供一致的编码器性能比较方法。然后，我们讨论了一些客观措施，在某些情况下可以有效地替代主观评估。简要回顾了基于MSE的方法，并强调了其优缺点。我们演示了人类观众经历的感知扭曲，但不能用如此简单的数学差异来描述。因此，出现了基于感知的视频指标，其中一些指标进行了审查和比较，表明与主观分数的相关性有所改善。

本章第二部分讨论了速率-失真优化问题，描述了一些最常见的技术，这些技术使我们能够为视频的每个时空区域选择最佳编码参数。在回顾了传统的速率失真理论后，我们专注于实用解决方案，例如使用拉格朗日乘数的解决方案，这表明这些解决方案可以为现代编解码器提供可处理的解决方案。与此相关，本章最后一部分研究了速率控制的要求和方法，该速率控制是根据给定信道的容量调整编码器输出比特率的一种手段。

Recommended articles

Citing articles (0)

## 参考文献

- [1] 美国。温克勒  
数字视频质量：视觉模型和指标  
威利（2005）  
谷歌学术
- [2] F. 科扎梅尔尼克，P. Sunna，E. 威肯，D. 彼得森  
互联网视频编解码器的主观质量评估——使用SAMVIQ进行第二阶段评估

EBU技术评论（2005年）， p.1-22

[谷歌学术](#)

- [3] 美国。温克勒  
**质量评估公共图像和视频数据库分析**  
IEEE信号处理精选主题期刊, 6 (6) (2012), p.1-10  
[CrossRef](#) [谷歌学术](#)
- [4] 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](#)
- [5] 建议ITU-T P.910, 多媒体应用的主观视频质量评估方法, ITU-T, 1999年。  
[谷歌学术](#)
- [6] F. 德西蒙, L. 戈德曼, J.-S.李, T. 易卜拉霍米  
**迈向高效视频编码: 潜在编码方法的主观评价**  
视觉传播和图像表示杂志, 22 (2011年), p.734-748  
[文章](#)  [下载PDF](#) [在Scopus中查看记录](#) [谷歌学术](#)
- [7] P.汉哈特, M.丽贝卡, F. 德西蒙, T. 易卜拉欣  
**即将出台的HEVC视频压缩标准进行主观质量评估**  
数字图像处理的应用XXXV, 卷。8499 (2012)  
[谷歌学术](#)
- [8] H. 霍夫曼, T. 板垣, D. 伍德, T. 欣茨, T. 维冈德  
**一种高清电视格式主观画质评价新方法及进一步研究**  
IEEE 广播交易, 54 (1) (2008), p.1-13  
[在Scopus中查看记录](#) [谷歌学术](#)
- [9] D.莱恩, 统计学导论。<<http://onlinestatbook.com/>>, 赖斯大学。  
[谷歌学术](#)
- [10] 视频质量专家组, 视频质量专家组关于验证视频质量评估客观质量指标的最终报告, VQEG, 技术报告, 2000年。<<http://www.its.bldrdoc.gov/vqeg/projects/frtv-phase-i/frtv-phase-i.aspx>>第一阶段。  
[谷歌学术](#)
- [11] 视频质量专家组, VQEG关于视频质量评估客观模型验证的最终报告, VQEG, 技术报告, 2003年。<<http://www.its.bldrdoc.gov/vqeg/projects/frtv>>第二阶段。  
[谷歌学术](#)
- [12] H. 谢赫, Z. 王, L. 科马克, A. Bovik, 实时图像质量评估数据库发布2.  
<<http://live.ece.utexas.edu/research/quality>>。  
[谷歌学术](#)
- [13] K. 塞沙德里纳坦, A. 博维克

## 自然视频的动态调音时空质量评估

IEEE图像处理交易, 19 (2010年), 第335-350页

[在Scopus中查看记录](#) [谷歌学术](#)

- [14] K. 塞沙德里纳森, R. 南卡罗来纳州桑塔拉詹博维克, L.K.科马克  
**视频主观客观质量评价研究**  
IEEE图像处理交易, 19 (2010年), 第335-350页  
[在Scopus中查看记录](#) [谷歌学术](#)
- [15] P.勒卡莱, F. Autrusseau, 主观质量评估IRCyN/IVC数据库, 2005年。  
<<http://www.irccyn.ec-nantes.fr/ivcdb/>>。  
[谷歌学术](#)
- [16] 视频质量专家组, 高清视频内容视频质量模型验证报告, VQEG, 技术报告, 2010年。  
<<http://www.its.bldrdoc.gov/vqeg/projects/hdtv/hdtv.aspx>>。  
[谷歌学术](#)
- [17] 视频质量专家组, 视频质量专家组关于验证多媒体质量评估客观模型的最终报告, VQEG, 技术报告, 2008年。  
<<http://www.its.bldrdoc.gov/vqeg/projects/multimedia-phase-i/multimedia-phase-i.aspx>>。  
[谷歌学术](#)
- [18] P.马尔齐利亚诺, F. 杜福, S. 温克勒, T. 易卜拉欣  
**无引用感知模糊度量**  
IEEE国际图像处理会议记录, 第3卷 (2002年), 第3页。57-60  
[在Scopus中查看记录](#) [谷歌学术](#)
- [19] R.Soundararajan, A. 博维奇  
**通过减少参考时空熵差异来评估视频质量**  
IEEE 视频技术电路和系统交易, 23 (4) (2013), p.684-694  
[CrossRef](#) [在Scopus中查看记录](#) [谷歌学术](#)
- [20] D.胡恩图, M.甘巴里  
**PSNR在图像/视频质量评估中的有效性范围**  
电子信件, 44 (13) (2008), p.800-801  
[CrossRef](#) [在Scopus中查看记录](#) [谷歌学术](#)
- [21] Z. 王, A. 博维克  
**平均平方误差: 爱还是留它?**  
IEEE信号处理杂志, 26 (1) (2009年), p.98-117  
[在Scopus中查看记录](#) [谷歌学术](#)
- [22] B.吉罗德  
**均方误差是怎么回事?**  
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](#)

- [23] F. 张, D. 公牛  
**使用基于区域的纹理模型进行视频压缩的参数框架**  
IEEE信号处理精选主题期刊, 5 (7) (2011), p.1378-1392  
[CrossRef](#) [在Scopus中查看记录](#) [谷歌学术](#)
- [24] T. 帕帕斯, T. 米歇尔, R. 后人  
**超阈值感知图像编码**  
IEEE国际图像处理会议记录 (1996年), p.234-240  
[谷歌学术](#)
- [25] 五. 卡亚尔加德, J. 马滕斯  
**模糊和噪声降解的图像的感知表征: 实验**  
美国光学学会杂志, 13 (1996年), 第1166-1177页  
[CrossRef](#) [在Scopus中查看记录](#) [谷歌学术](#)
- [26] 美国. 卡鲁纳塞克拉, N. 金斯伯里  
**基于人类视觉敏感性的屏蔽图像伪影的失真度量**  
IEEE图像处理交易, 4 (6) (1995年), 第713-724页  
[在Scopus中查看记录](#) [谷歌学术](#)
- [27] M. 卡内克, P. 勒卡莱, D. 巴尔巴  
**一种基于结构信息感知的图像质量评估方法**  
IEEE国际图像处理会议记录, 第2卷 (2003年), 第2页。185-188  
[在Scopus中查看记录](#) [谷歌学术](#)
- [28] D. 钱德勒, S. 赫马米  
**VSNR: 一种基于小波的自然图像视觉信噪比**  
IEEE 图像处理事务, 16 (9) (2007), 第 2284-2298 页  
[在Scopus中查看记录](#) [谷歌学术](#)
- [29] E. 拉尔森, D. 钱德勒  
**最明显的失真: 全参考图像质量评估和策略的作用**  
电子成像杂志, 19 (1) (2010)  
pp.011006-1-011006-21  
[谷歌学术](#)
- [30] 美国. 奇克鲁尔, V. 桑达拉姆, M. 赖斯莱因, L. 卡拉姆  
**客观视频质量评估方法: 分类、回顾和绩效比较**  
IEEE 广播交易, 30 (2005), 第17-26  
[在Scopus中查看记录](#) [谷歌学术](#)
- [31] W. 林, C. 郭  
**感知视觉质量指标: 一项调查**  
视觉传播和图像表示杂志, 22 (2011年), p.297-312  
[文章](#)  [下载PDF](#) [在Scopus中查看记录](#) [谷歌学术](#)

- [32] 美国。谢赫, A. 博维克  
**图像信息和视觉质量**  
IEEE图像处理交易, 15 (2006年), 第430-444页  
[CrossRef](#) [在Scopus中查看记录](#) [谷歌学术](#)
- [33] 美国。温克勒, P. 莫汉达斯  
**视频质量测量的演变: 从PSNR到混合测量**  
IEEE广播交易, 54 (3) (2008年), 第660-668页  
[CrossRef](#) [在Scopus中查看记录](#) [谷歌学术](#)
- [34] M.平森, S. 狼  
**一种客观测量视频质量的标准化新方法**  
IEEE 广播交易, 50 (2004), p.312-322  
[在Scopus中查看记录](#) [谷歌学术](#)
- [35] Z. 王, A. 博维克, H. 谢赫, E. 西蒙切利  
**图像质量评估: 从误差可见性到结构相似性**  
IEEE图像处理交易, 13 (2004年), 第600-612页  
[CrossRef](#) [在Scopus中查看记录](#) [谷歌学术](#)
- [36] Z. 王, E. 西蒙切利  
**复小波域中的平移不敏感图像相似性**  
IEEE国际声学、语音和信号处理会议 (2005年), p.573-576  
[在Scopus中查看记录](#) [谷歌学术](#)
- [37] Z. 王, E. 西蒙切利, A. 博维克  
**图像质量评估的多尺度结构相似性**  
IEEE Asilomar 信号、系统和计算机会议记录 (2003 年), p.1398-1402  
[CrossRef](#) [在Scopus中查看记录](#) [谷歌学术](#)
- [38] Z. 王, L. 卢, A. 博维克  
**基于结构失真测量的视频质量评估**  
信号处理: 图像通信, 19 (2) (2004), p.121-132  
[文章](#)  [下载PDF](#) [CrossRef](#) [在Scopus中查看记录](#) [谷歌学术](#)
- [39] P.Vu, C. Vu, D. 钱德勒  
**视频质量评估的时空最明显失真模型**  
IEEE国际图像处理会议记录 (2011年), p.2505-2508  
[CrossRef](#) [在Scopus中查看记录](#) [谷歌学术](#)
- [40] R.曼蒂克, K. 金, A. 伦佩尔, W. 海德里希  
**HDR-VDP-2: 用于所有亮度条件下可见性和质量预测的校准视觉指标**  
ACM图形交易 (SIGGRAPH'11程序), 30 (4) (2011)  
(第1条。40)  
[谷歌学术](#)



- [41] A.张, D.公牛  
**感知视频压缩质量评价方法**  
IEEE国际图像处理会议记录 (2013年), p.39-43  
[在Scopus中查看记录](#) [谷歌学术](#)
- [42] A.张, D. Bull, 一个基于感知的视频质量评估混合模型, IEEE视频技术电路和系统交易, 提交出版。  
[谷歌学术](#)
- [43] A.奥尔特加, K.拉姆钱德兰  
**图像和视频压缩的速率失真方法**  
IEEE信号处理杂志, 15 (6) (1998年), p.23-50  
[在Scopus中查看记录](#) [谷歌学术](#)
- [44] G.沙利文, T.维冈德  
**视频压缩速率失真优化**  
IEEE信号处理杂志, 15 (6) (1998年), p.74-90  
[在Scopus中查看记录](#) [谷歌学术](#)
- [45] Y.王, J.奥斯特曼, Y.张Q  
**视频处理和通信**  
普伦蒂斯大厅 (2002)  
[谷歌学术](#)
- [46] T.维冈德, H.施瓦茨, A.乔赫, F.科森蒂尼, G.沙利文  
**速率约束编码器控制及视频编码标准对比**  
IEEE 视频技术电路和系统交易, 13 (7) (2003), p.688-703  
[在Scopus中查看记录](#) [谷歌学术](#)
- [47] E.H.杨, X.于  
**H.264帧间编码速率失真优化: 通用框架**  
IEEE 图像处理事务, 16 (7) (2007), 第 1774-1784 页  
[在Scopus中查看记录](#) [谷歌学术](#)
- [48] H.埃弗雷特  
**广义拉格朗日乘数法, 解决资源最优分配问题**  
运营研究, 11 (1963年), p.399-417  
[CrossRef](#) [谷歌学术](#)
- [49] Y.肖汉姆, A.格尔绍  
**任意一组量化器的有效位分配**  
IEEE 声学、语音和信号处理交易, 36 (1988年), p.1445-1453  
[在Scopus中查看记录](#) [谷歌学术](#)
- [50] D.-K.权, M.-Y.沈, C.郭  
**具有增强速率和失真模型的H.264视频速率控制**

- [51] G.沙利文, T. 维冈德, K. Lim, 联合模型参考编码方法和解码隐藏方法, JVT-I049, 圣地亚哥, 2003年。

[谷歌学术](#)

- [52] K. 林景芳, G. 沙利文, T. Wiegand, ISO/IEC MPEG和ITU-T VCEG文档JVT-O079的联合视频团队 (JVT), 联合模型参考编码方法和解码隐藏方法的文本描述, 2005年。

[谷歌学术](#)

- [53] S.S.桑斯-罗德勒格斯, T. 席尔  
一种具有分层GOP结构的HEVC速率控制算法

IEEE国际学术语音和信号处理会议记录 (2012年) 第1710-1722页



下载

- [54] B.李, H. 李, L. 李, J. 张, HEVC的R-lambda模型速率控制, JCTVC-K0103, 在: 第11届JCTVC会议, 中国, 2012年10月。

[谷歌学术](#)

- [55] A.萨德卡  
压缩视频通信  
威利 (2002)

[谷歌学术](#)

- [56] D.Agrafiotis, D. 公牛, C. 卡纳加拉贾, N.卡姆农瓦塔纳  
使用H.264编码多个优先兴趣区域

IEEE图像处理国际会议记录 (2006年), p.53-56

[在Scopus中查看记录](#) [谷歌学术](#)

★ 有关图10.14和10.15的彩色高质量版本, 请参阅电子版或网站。

<sup>1</sup> 可以说, 由ANSI标准化的VQM[34]比其他公司更广泛地接受, 尽管它不是性能最好的。

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。

FEEDBACK 