

符合人眼视觉特性的视频质量评价模型

王楠楠 李桂苓

(天津大学电信学院, 天津 300072)

摘要 视频技术的发展为其质量评价提出了新的课题,但由于评价图象质量的关键在于所用视觉模型是否符合人的感知特性,因此评价图象质量必须考虑以视觉锐度、对比度敏感度、多通道结构和掩盖特性为基础的人眼视觉特性(HVS)。为了使人们对基于人眼视觉特性的视频质量评价模型研究现状有所了解,介绍了几种目前比较成功的基于HVS的视频质量评价模型,并分析和总结了它们的性能,最后展望了评价模型的发展。

关键词 视频图象 质量评价 HVS 视觉模型

中图法分类号: TP918.8 文献标识码: A 文章编号: 1006-8961(2001)06-0523-05

Video Quality Evaluation Models Based on Human Visual Properties

WANG Nan-nan, LI Gui-ling

(Electronic Information Engineering college of Tianjin University, Tianjin 300072)

Abstract Picture quality evaluation methods could be utilized to assess image algorithms' performance. The evaluation of image quality includes subjective evaluation and objective test. The traditional subjective evaluation depends on observer's experiences and motivations, and therefore it is complex and its results always change. With the development of the video compression technology, it is very urgent to research new video quality evaluation method. The key point of the video quality evaluation is that its model should match the perceptive characteristics of human. Based on the visual acuity, contrast sensitivity function, multi-channel structure and masking effect, this paper summarizes main properties of Human Visual System (HVS) taken into account in the assessment of the video quality. It also presents several comparatively satisfactory video quality models based on HVS, including spatio-temporal model, RAI model, three-dimension weighted signal-to-noise model and three-layered bottom-up noise weighting model. Their performances are also analyzed respectively. The perspective of video quality evaluation is given finally.

Keywords Video image, Quality evaluation, HVS, Visual model

0 引言

如今基于 MPEG-2^[1]标准的数字视频压缩技术正迅速发展,其目标是获取高压缩比和高质量的图象,但高压缩比会带来图象质量损伤,如何评定这种损伤成为人们普遍关心的问题,与模拟视频系统相比,评价数字重建图象质量更为复杂。近年来,图象质量评价普遍采用主观评价方法,国际上也制订了相应标准 ITU-R. BT. 500^[2],但由于主观评价需许多评

估者参与和需按复杂程序和步骤进行,因而耗时长,结果稳定性也较差,难于生产测试和故障排除等,因此研究一种在一定程度上能替代主观评价,并能实现定量测定图象质量的客观方法则具有重要意义。

目前对图象的客观测量通常采用信噪比(SNR)或均方误差(MSE)这样的指标,但 SNR 和 MSE 与主观评价的相关性较差,Eskicioglu 等通过对 14 种图象质量评价方法进行研究后指出,引入人眼视觉特性(HVS)可提高测试结果与主观评价的一致性^[3],因此本文首先描述了人眼视觉基本特性,

然后概述了基于这些特性的视频质量评价模型。

1 视觉基本特性

1.1 视觉锐度

视觉锐度表征视觉系统分辨细节的能力,并能分辨的相邻两点视角的倒数 θ 表示,其值与该相邻两点在视网膜上的成像位置有关^[4],即由于光敏细胞在视网膜的黄斑区分布较密,故中央凹(黄斑区中心)分辨力最高,随着偏离中央凹距离 x 的增加,视觉锐度以 $1/x$ 形式单调、快速下降,且在视网膜外围区域,由于多个光敏细胞连接在一条视神经上,故视觉锐度下降最快。

1.2 对比度敏感度

视觉分辨力还与相对对比度有关,相对对比度用视觉系统的对比度敏感度函数 CSF 表示(或称人眼视觉系统空间调制转移函数 MTF ^[5])。 CSF 是频率的函数,表示为

$$CSF(\mu) = \frac{\text{输出对比度}}{\text{输入对比度}} \quad (1)$$

此处 μ 为空间频率,单位是 cpd(周/度)。实验中常以探测阈值来得到对比度的大小,探测阈值对应激励刚好可见或不可见,并以观察者检测到的概率为 50% 时的激励强度来定义。 CSF 为探测阈值 $C_T(\mu)$ 的倒数

$$CSF(\mu) = \frac{1}{C_T(\mu)} \quad (2)$$

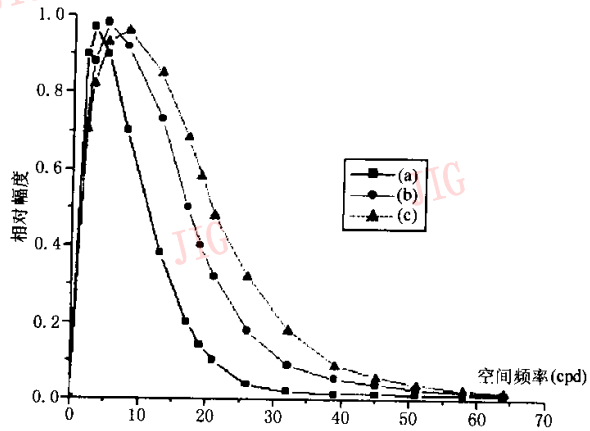
不同实验所得的 CSF 虽略有不同,但基本与 Ngan 等给出的函数形式一致^[6]

$$CSF(\mu) = (0.31 + 0.69)e^{-0.29\mu} \quad (3)$$

在图象编码与评价中使用最广的 CSF 函数已分别由 Ngan、Mannos 和 Sakrison、Nill 给出^[6~8](见图 1)。由图表明, HVS 具有带通滤波器性质,其高、低频端敏感度下降。

1.3 多通道结构

生理学、心理学的实验研究指出,视觉皮层细胞的响应在频域呈带通特性,并给出了敏感度峰值位置和视觉细胞响应带宽。据此推知,人的大脑具有将独立的视觉机制聚合起来的能力,而各视觉机制对频率域之某一部分敏感^[9],故可用一滤波器组来模拟人的视觉特性。而该滤波器组则是各视觉机制的近似,它们将可见数据分解成一组各个方向、空间频率和时间频率带宽均有限的信号,这种有限带宽信



(a) Ngan 的 CSF , 峰值在 3cpd 处; (b) Nill 的 CSF , 峰值在 5cpd 处; (c) Mannos 和 Sakrison 的 CSF , 峰值在 8cpd 处

图 1 CSF 特性曲线

号即称为通道。

Daugman 和 Hess 等通过研究认为,可以按对数形式将空间频率域分为 4~8 个频带,并且存在同样数目的方向带,而时间频率则由 2 个或 3 个通道组成^[10,11]。由于时间频率的第 3 个通道只在极低空间频率下存在,因此关于是否存在这个通道一直有很多争论,近来的研究倾向于只有两个时间视觉机制^[12],即瞬时视觉机制和持续视觉机制,其中瞬时视觉机制对快速运动模式敏感;而持续视觉机制对静止或慢速运动的图象敏感。

在多数情况下,可将通道看成是彼此独立的,但一些研究者认为,通道之间也存在着相互作用^[13]。另外,空间对比度函数还存在着空间-时间的可分性问题,例如,运动物体的时间感知特性依赖于空间特性,这种时间特性在低频端是带通,而在高频端是低通。如今有两种观点来解释这种空间-时间感知的相互关系,其第 1 种称为敏感度标度假设,它将这种关系解释为时间滤波器的峰值敏感度与空间频率的依赖关系,按照这种假设,在时间域和空间域中,可使用分离的滤波器组;另一种称为共变假设,它将这种关系理解为在滤波器组的时间特性中空间-时间的共变,而且这种假设产生的滤波器组比第一种更为复杂。

1.4 掩盖效应

CSF 解释了对单一激励的感知,即几个激励同时存在时,激励间会产生相互作用,一个激励的存在将导致另一激励探测阈值的改变,这种现象称为掩盖效应^[14,15],由于通常只考虑两个激励包含在同一通道时的情形,因此掩盖效应可看作是一个非线性

变换器,如图 2 所示.

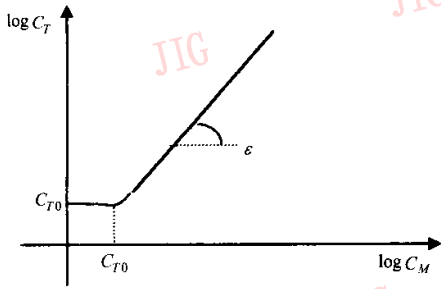


图 2 掩盖效应模型

图 2 中, C_{T_0} 是由 CSF 给出的对于目标的对比度探测阈值(即在没有背景时的对比度探测阈值); C_T 是存在掩盖效应时目标的对比度探测阈值; C_M 是背景的对比度. 由图 2 可见:

(1) C_M 较低时, 探测阈值为一常数, 即

$$C_T = C_{T_0};$$

(2) 当 C_M 接近 C_{T_0} 时, 探测阈值略微下降, 呈现一勺状;

(3) 随着 C_M 的增加, C_T 以对比度掩盖器(即背景)幂的形式增加, 其用对数-对数的图形表示则为线性关系, 斜率为 ϵ .

实际探测阈值 C_T 可由下式计算

$$C_T = \begin{cases} C_{T_0} & \text{若 } C_M < C_{T_0} \\ C_{T_0} \left(\frac{C_M}{C_{T_0}} \right)^\epsilon & \text{其他} \end{cases} \quad (4)$$

上面这种激励可见度与激励所在处的平均背景亮度有关的掩盖效应, 则称为对比度掩盖效应. 此外, 还有一种纹理掩盖效应, 它是指由于背景亮度在空间或时间上的非均匀性所导致的可见度阈值的改变. 据研究, 大多数图象都包含复杂而不均匀的亮度背景, 且在有较大亮度变化的区域和具有丰富活动性的区域, 其可见度阈值增大, 也即激励的可见度减小.

2 典型的客观评价模型

2.1 三维滤波模型

Van den Branden 等在人眼视觉特性的基础上, 建立了一种用于评价视频编码质量的视觉时空模型, 并提出了 MPQM (Moving Picture Quality Metric) 质量度量^[16,17]. 该模型的主要特点是用一个三维滤波器组来模拟 HVS 的多通道特性(如图 3 所示). 滤波器模拟的过程是先把原始图象序列粗略地分割成均匀区域、轮廓和纹理, 同时, 使用滤波器组

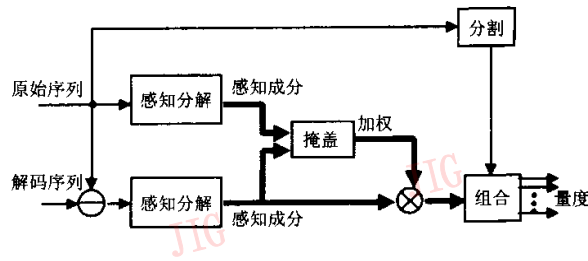


图 3 三维滤波模型示意图

按照 5 种空间频率、4 种取向、2 种时间频率将原始图象序列和编码误差序列分解为感知成分, 且每一通道中都要考虑对比度敏感度和掩盖效应; 然后, 将感知通道的输出进行掩盖, 这一步骤可预测到皮层 V_1 区域(即初级视觉皮层)细胞的响应; 把所得数据收集起来用以解释高层感知, 此过程称为组合; 最后提出了 MPQM 度量法. 考虑到注意力集中和观看距离的影响, 这种度量以三维序列块为基础进行计算, 其从本质上说, 就是用 Minkowski 求和法把通道输出的幅值结合在一起, 按照失真的程度, 予以相应的指数加权. 一个给定序列块的实际失真 E 可按下式计算

$$E = \left(\frac{1}{N} \sum_{c=1}^N \left(\frac{1}{N_x N_y N_t} \sum_{t=1}^{N_t} \sum_{x=1}^{N_x} \sum_{y=1}^{N_y} |e[x; y; t; c]| \right)^\beta \right)^{\frac{1}{\beta}} \quad (5)$$

其中, $e[x; y; t; c]$ 是在位置 (x, y) 、当前块的时间 t 、通道 c 处经掩盖的误差信号; N_x, N_y 和 N_t 分别表示块的水平、垂直和时间维数; N 是通道数; β 是 Minkowski 求和法的指数.

这种 MPQM 质量度量方法较全面地模拟了人眼的视觉特性, 能够在相当大的比特率范围内对视频质量做出较好的评价. 但由于频率分解而增加了算法的运算量, 计算的时间较长. 因此若想做到对视频质量进行实时评价, 还需对此模型做进一步的简化.

2.2 预测模型

RAI-Research Centre 开发了一种研究视频压缩系统主、客观相关性的模型^[18], 它可通过客观测量来评价主观图象质量. 考虑到人眼对不同空间频率激励的敏感性以及空间掩盖效应, 首先用 3×3 的 Sobel 滤波器将图象划分为平坦区和边缘, 随后用包含平坦区、边缘和整帧的 M 组训练序列, 依据 DSCQS (Double Stimulus Continuous Quality Scale) 方法完成主观评价, 得到主观评价真值 DEG ; 其次, 从 M 组中抽出它的 N 个子序列, 并按下式计算评价的预测值 deg

$$deg = C_0 + C_1 m_1 + \dots + C_n m_n \quad n \leq m \quad (6)$$

其中, C_i 为系数; m_i 为组成子集 N 的客观测量值; 最后用最小均方差的方法计算出模型的系数, 进而得到整个模型。这种模型已应用于那些用来进行主观测试的序列, 且将测试结果与在 MPEG 内完成的主观评价结果进行了比较, 结果比较令人满意, 但这种模型的缺点是模型本身只适用于用来计算它的那些序列, 且在一个压缩系统下得到的模型, 不能用于评价另外一个系统, 因此可移植性较差。

2.3 三维加权信噪比模型

Okomoto 等提出了一种客观感知方法——三维信噪比 (3D-SNR)^[19] 方法。这种方法是在考虑了视觉时空频率响应和视觉掩盖效应的基础上由普通的 SNR 形成的。时空频率特性 $W(\mu, \nu, f)$ 是由空间频率特性 $W_s(\mu, \nu)$ 和时间频率特性 $W_T(F)$ 相乘而得。

$$W_s(\mu, \nu) = 2.46(0.1 + 0.25f) \exp(-0.25f) \quad (7)$$

$$W_T(F) = 0.134 \times [1 + (F/0.5)] / [1 + (F/7.8)^2]^{1.2} \quad (8)$$

式中, $f = (\mu^2 + \nu^2)^{1/2}$, μ 是水平空间频率, ν 是垂直空间频率, 单位均是 cpd(周/度), F 是时间频率, 单位是 Hz。由于 W_s 与 W_T 都近似符合 HVS, 因此三维加权信噪比模型比起其他模型更易理解, 算法也比较简单, 依据三维信噪比的模型进行计算机模拟实验的结果表明, 客观测试指标与原有的物理信噪比相比, 其与主观评价结果的相关性有显著提高^[20]。

2.4 3层加权模型

这是由 Hamada 等人提出的一种评价视频质量的方法^[21], 其基本思想是, 鉴于人眼不是一次看清整个图象, 而只是看到其中注视点附近的一个点域, 因此人眼可依赖于这个点域附近, 一些混在纹理结构中噪声的程度和特点, 来识别这个点域的纹理结构和质量, 若沿着目标移动注视点, 就可感受到整个图象的质量。在此基础上, 他们提出了由下至上的 3 层噪声加权模型(见图 4 所示)。

这种模型在每一层都需使用一个考虑了视觉感知的特定加权函数, 即:

① 首先在噪声层, 针对局部域 (j, m) 中的噪声特性, 用 h_i 进行噪声加权, 然后计算这些经过平均加权噪声 $WMSE_{NL}$

$$WMSE_{NL}(j, m) = \frac{1}{I} \sum_{i=1}^I h_i \cdot P(j, m, i) \quad (9)$$

② 在纹理层, 针对局部域 (j, m) 中的纹理 $C(j, m)$, 用 t_c 进行加权, 再计算每一帧 $WMSE_{TL}$ 的平均

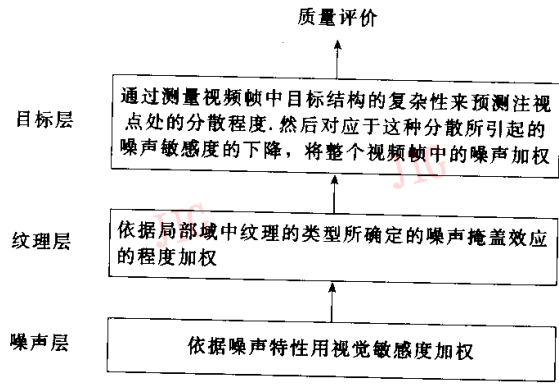


图4 由下到上的 3 层噪声加权模型
加权值 $WMSE_{TL}$

$$WMSE_{TL}(j) = \frac{1}{M} \sum_{m=1}^M t_c(j, m) \cdot WMSE_{NL}(j, m) \quad (10)$$

式中, $t_c(j, m)$ 是纹理层 $C(j, m)$ 的噪声加权函数。

③ 在目标层, 对应于注视点的分散程度, 用 $G(j)$ 对 j 帧进行加权, 随后计算 1 到 J 帧上 $WMSE_{TL}$ 的平均加权值 $WMSE_{OL}$;

$$WMSE_{OL} = \frac{1}{J} \sum_{j=1}^J q_G(j) \cdot WMSE_{TL}(j) \quad (11)$$

式中, $q_G(j)$ 为噪声加权函数, $j=1, 2, \dots, J$

最后将 $WMSE_{OL}$ 转化为加权信噪比 $WSNR$, 再按照 ITU-R Rec. 500-7 定义的双重刺激连续质量标度方法(DSCQS)用下式进行计算

$$D(\%) = \int(WSNR) \quad (12)$$

式中, D 表示 DSCQS。

这种方法的特点在于可适用于不同类型的视频图象, 并支持模拟/数字的复合/分量格式, 可做成硬件, 因而易于实现实时评价。

3 展望

在视频质量评价中应引入 HVS 这一点已被众多研究者所证实, 尽管研究 HVS 已有相当长的时间, 且已采用了很多相互补充的方法, 但人们对 HVS 的认识还存在许多未知领域, 由于几乎不可能建立起比较精确、统一的数学或物理模型, 因此这在一定程度上影响了视频质量客观评价的准确性。另外, 广播系统对视频质量进行实时评价的要求越来越高, 这意味着评价模型必须简洁, 而且便于硬件实现, 而现有的视频评价模型常常是有效性和复杂性的折中, 由此可见, 应用不断更新的数学工具, 对生

理学和心理学进行更加深入的研究,是完善视频质量评价的必要前提。

视频质量客观评价模型虽然多种多样,但迄今为止尚无客观评价的国际标准,而数字视频压缩技术在地面广播、卫星广播、节目制作、有缆和无线业务网等方面的广泛应用则急需制定一套统一的国际标准。为此,1997年,由致力于视频质量评价方法和视频质量评价标准的各个领域和组织的专家组成了VQEG(Video Quality Expert Group,视频质量专家组)^[22]。目前,VQEG正在世界范围内广泛征集各种提案,并对各种客观模型的性能进行综合分析。可以预见,随着众多研究者的不懈努力,制定视频质量客观评价标准的步伐将会越来越快。

参 考 文 献

- ISO/IEC 13818, Generic coding of moving pictures and associated audio, 1994.
- CCIR Rec. 500-5. Method for the subjective assessment of the quality of television pictures. Recommendations of the ITU, Telecommunications Standardization Sector, 1990.
- Eskicioglu A M, Fisher P S. Image quality measure and their performance. IEEE Trans. Comm., 1995,43(12):713~724.
- 俞斯乐,郭福云,李桂苓等.电视原理.北京:国防工业出版社,1994.
- Eude T, Cherifi H. On quality metrics for low bitrate coding. SPIE Visual Communication and Image Processing, 1997,3016:70~81.
- Ngan K N, Rao K S, Singh H. Cosine transform coding incorporating human visual system model. SPIE fiber'86, Cambridge, MA, U. S. A., 1986:165~171.
- Mannos J L, Sakrison D J. The effect of a visual fidelity criterion on the encoding of images. IEEE Trans. Inform Theory., 1974,20:525~536.
- Nill N B. A visual model weighted cosine transform for image compression and quality assessment. IEEE Trans. Comm., 1985,33(3):551~557.
- Daugman J G. Two-dimensional spectral analysis of the cortical receptive field profiles. Vision Research, 1980,20:847~856.
- Hess R F, Snowden R J. Temporal properties of human visual filters; Number, shapes and spatial covariation. Vision Research, 1992,32(1):47~59.
- Burbek C A, Kelly D H. Spatiotemporal characteristics of visual mechanisms:Excitatory-inhibitory Model. JOSA, 1980,70(9):1121~1126.
- Hammett S T, Smith A T. Two temporal channels or three? A re-evaluation. Vision Research, 1992,32(2):285~291.
- Teo P C, Heeger D J. Perceptual image distortion. In:Proceedings of the International Conference on Image Processing, Austin, TX, U. S. A., 1994:982~986.
- Legge G E, Foley J M. Contrast masking in human vision. JOSA, 1980,70(12):1458~1471.
- Comes S, Macq B. Human visual quality criterion. SPIE Visual Communication and Image Processing, 1990,1360:2~7.
- Lindh P, Van den Branden C J. Efficient spatio-temporal decomposition for perceptual processing of video sequences. In:Proceedings of the International Conference of Image Processing, Lausanne, Switzerland, Sponsored by the IEEE Signal Processing Society, 1996:331~334.
- Van den Branden C J. A working spatio-temporal model of the human visual system for image restoration and quality assessment applications. In:Proceedings of the International Conference on Acoustics, Speech, and Signal Processing, Atlanta, Georgia, USA, Sponsored by the IEEE Signal Processing Society 1996:2293~2296.
- Ardito M, Visca M. Correlation between objective and subjective measurements for video compressed system. SMPTE, 1996,105(12):768~773.
- Okomoto J. A study on subjective and objective evaluation methods for coded moving picture quality. International Picture Coding Symposium, Melbourne, Australia, 1996:519~523.
- 张强.主、客观统一的MPEG-2系统运动图象质量评价技术的研究[硕士学位论文].天津:天津大学,1999.
- Hamada T, Miyaji S, Matsumoto S. Picture quality assessment system by three-layered bottom-up noise weighting considering human visual perception. SMPTE, 1999,108(1):20~26.
- Corriveau P, Webster A. VQEG evaluation of objective methods of video quality assessment. SMPTE, 1999,108(9):645~648.

王楠楠 1972年生,1998年获哈尔滨工业大学物理电子学与光电子学专业硕士学位,现为天津大学信号与信息处理专业博士研究生。研究方向为多媒体技术、图象质量评价。

李桂苓 1938年生,1962年天津大学电子工程系毕业,现任天津大学电子信息工程学院教授、博士生导师。研究方向为声音与图象信息处理与传输、数字视频与多媒体技术。