

压缩视频感觉质量的计算

宋建新, 毕厚杰

(南京邮电学院 166 信箱, 南京 210003)

摘 要: 论文提出了基于视觉特性、视频内容和模糊学的压缩视频感觉质量计算方法. 方法的核心是利用视觉掩蔽特性. 论文分析了与压缩视频质量有关的视觉特性及视频图像内容特性, 提出了视觉掩蔽计算结构及用模糊学方法进行视觉阈值提升的计算方法, 并给出了感觉峰值信噪比的压缩视频质量的计算方法, 因而该方法使压缩视频质量的度量更符合人的视觉特性.

关键词: 视觉特性; 图像特性; 压缩视频质量

中图分类号: TN919

文献标识码: A

文章编号: 0372-2112 (2000) 07-0079-05

Perceptual Quality Metric for Compressed Video

SONG Jian-xin, BI Hou-jie

(173 # 38 GuangDong Rd. Nanjing 210003, China)

Abstract: This paper proposes the method to measure perceptual quality of compressed video which is based on human visual features, video contents and Fuzziness. The key of the method is to compute the visibility threshold. The paper analyses the visual features and the video features which are related to quality metric for compressed video. We design the architecture for measuring perceptual video quality. The elevation factors for visibility threshold are calculated based on fuzziness. Perceptual PSNR is proposed for measure of compressed video.

Key words: visual features; video features; compressed video quality

1 引言

实际的视频压缩编码均采用失真编码. 如何度量这种失真, 以便将失真恰当地分配, 是保证压缩视频质量, 提高编码效率的重要一环, 尤其对以质量为目标视频编码应用场合更是如此, 如 VBR (Variable Bit Rate) 视频编码等.

在以人眼为最终信宿的视频应用中, 视频质量应与人的视觉相匹配, 评价视频质量应以人的视觉为依据. 由于视频质量的客观度量方法^[1] (如基于均方误差的评价方法) 独立于视频的内容, 度量结果与人眼的视觉感受往往不一致. 主观评价由评价者直接从感觉上去度量图像的失真, 通过打分评价, 其结果必然符合人的感觉. 然而这种方法不能直接用于视频压缩编码过程中的质量评价与控制. 因此要寻找一种既符合人眼感觉、又可定量计算的压缩视频质量的度量方法, 其关键就是要寻找导致主观评价与客观评价不一致的根源所在, 以便“对症下药”.

由于人的视觉能力有限, 在一些条件下可以容忍一些失真, 有些失真人眼根本看不出来, 所以要研究哪些损伤人眼能觉察, 哪些损伤人眼看不到, 只有这样才能真正区分压缩视频的质量. 依据人眼感觉所确定的压缩视频的质量称为视频感觉质量. 压缩视频的感觉质量依赖于视频编码方案、编码控制

方案及视频内容.

MPEG 算法的图像损伤由量化引起. 从视觉的角度, 量化误差是一种视觉激励, 它以未量化的原始图像为背景. 视频质量的度量就是研究这种视觉激励的大小. 本文提出一种基于图像子块内容的、结果符合人眼感觉的压缩视频质量的计算方法.

视频质量客观度量与主观度量不同的关键是视觉掩蔽. 由于视觉掩蔽效应, 在一定条件下的一定量的编码损伤被掩蔽掉了, 从质量评价的角度, 看不见的编码损伤就不应计及到量化损伤中, 所以研究视频感觉质量计算的基础是视觉掩蔽^[2, 3]. 本文的贡献在于提出了一种基于人眼视觉特性和模糊数学的、可计算的视频感觉质量的度量方法, 其意义在于: (1) 解决了视频感觉质量的计算问题, 为视频编码的质量控制奠定了基础^[9]. (2) 将模糊数学的思想引入到压缩视频感觉质量的度量中.

2 视觉特性与视频特性

视觉是图像作用于人眼的结果, 视频是一组相关图像按一定速率播放的序列. 所以研究压缩视频感觉质量的基础是视觉特性和视频特性.

收稿日期: 1999-03-29; 修订日期: 2000-02-15

基金项目: 原邮电部中青年教师基金资助课题

2.1 视频感觉质量的特性

(1) 静态性与动态性

空间域上量化损伤给人的感觉称为静态损伤,如一视频序列冻结在某帧上,看到的该帧损伤就是静态损伤,而视频连续播放时量化损伤给人的感觉称为动态损伤.人眼对静态损伤比对动态损伤更为敏感.

(2) 内容相关性

一定量的编码损伤是否可见,既取决于损伤的大小,也取决于图像的内容.因为同样幅度的损伤位于不同内容的图像背景下,给人的感觉不同.

(3) 编码方法相关性

不同的编码方法,信息丢失的类型不同,损伤的形式不同,给人的感觉也不同.

2.2 与视频图像质量度量有关的视觉特性

视觉处理的复杂性目前还没有为人类所真正理解和掌握,但是人们已发现了一些视觉现象,本文研究的是如何将些现象用到压缩视频的质量度量中.涉及压缩视频感觉质量的视觉特性有多通道结构^[8]、视觉阈值^[6]及掩蔽^[5].

(1) 多通道结构与视觉阈值

人的视觉系统是一个多通道结构,它把输入的图像分解成不同感觉分量.每个感觉通道都有其自己的阈值,称之为视觉阈值.给定通道中激励,如果其值低于通道所对应的视觉阈值,人眼就感觉不到该激励.本文研究图像编码的损伤度量问题,也就是把编码损伤视作一种激励,由于这种激励通常很小,视觉系统并不是总能感觉到它们的,只有激励超过视觉阈值,人才能感觉到损伤的存在,所以研究压缩视频的感觉质量,视觉阈值是一个很重要的值.

(2) 掩蔽效应

当存在多个激励时,它们之间就会互相干扰,导致视觉阈值发生变化,这就是掩蔽现象.对编码的视频图像,原图像是掩蔽者,编码损伤是目标.实际的视频图像千变万化,编码后的损伤分量的背景也就各式各样,所以损伤分量的可见性也是变化的.

视觉阈值的存在使得低于阈值的损伤不被觉察到,掩蔽的存在则使视觉阈值提升,从而使不可见的损伤增大.这是研究基于 HVS(Human Visual System)的压缩视频质量度量的一个重要的出发点.换句话说,如果设法使损伤出现在人眼看不见的地方,也就改善了图像的质量.

掩蔽有各种形式^[2~4].人眼对损伤的敏感度在非常亮或非常暗的区域下降,这称为对比掩蔽;人眼对损伤的敏感度在图像空间域变化大的区域比变化小的区域低,这称为纹理掩蔽;内容随时间变化大的图像块,人眼对其损伤的敏感度低,这是运动掩蔽;场景切换后的瞬间 1~3 帧,人眼对损伤的可见性下降,这是切换掩蔽.

可见掩蔽的大小与图像的内容有关.

2.3 视频图像特性度量

视频图像的特性可分别在空间域和时间域描述.视频图像块空间域的特性可用亮度、活动性来表示.视频图像块时间域的特性可用运动性及场景切换来表示.

(1) 图像子块亮度

图像子块的亮度 b 定义为图像子块的灰度平均值,即

$$b = \frac{1}{N^2} \sum_{m=0}^{N-1} \sum_{n=0}^{N-1} f(m, n) \quad (1)$$

式中, $f(m, n)$ 是子块中各象素的灰度值.

(2) 图像子块复杂度

用空间变化率来描述图像子块的复杂度,其定义如下:对 8×8 图像子块,设行方向上灰度变化率为 f 、列方向上灰度变化率为 g 、图像子块的空间变化率为 sp ,则有

$$\begin{aligned} f &= \sqrt{\frac{1}{56} \sum_{j=0}^7 \sum_{k=1}^7 [f(j, k) - f(j, k-1)]^2} \\ g &= \sqrt{\frac{1}{56} \sum_{k=0}^7 \sum_{j=1}^7 [f(j, k) - f(j-1, k)]^2} \\ sp &= \sqrt{(f)^2 + (g)^2} \end{aligned} \quad (2)$$

(3) 图像子块运动性

图像子块的运动性指相邻两帧图像同一位置上的图像子块内容的变化.即

$$m = \frac{1}{8 \times 8} \sum_i \sum_j |f_k(i, j) - f_{k-1}(i, j)| \quad (3)$$

(4) 场景切换

场景切换应根据视频序列中各帧图像之间所包括的运动信息来确定,这里采用简单的“帧间距离”的概念来度量.所谓帧间距离是指两帧图像间对应位置亮度的差异,可用帧差直方图来度量.设有两帧图像, f_n 和 f_m , 它们之间差的直方图记为 $hg[i], i = [q-1, q-1]$, q 为图像象素的灰度级.

hg 偏离 0 的值越多,表明两帧间的变换越大,所以两帧间的帧间距离为

$$d(f_n, f_m) = \sum_{i \in [q-1, q-1]} \frac{hg[i]}{\sum_{i=q-1}^{q-1} hg[i]} \quad (4)$$

式中,分母是总的象素数,分子表示差值超过某一阈值 $|i|$ 的象素数,所以 $d(f_n, f_m)$ 反映了两帧图像偏离超过阈值的程度.

3 基于图像内容的视觉阈值像计算

视觉掩蔽体现在视觉阈值的上升上.对基于 MPEG 算法产生的编码图像的图像感觉质量度量的核心就是根据图像特性来确定每个 DCT 系数的视觉阈值. MPEG 编码中已经使用了这种概念,它体现在量化权重矩阵上.本文以此为基础来进一步研究特定图像子块 DCT 系数的视觉阈值.

3.1 独立于图像内容基本视觉阈值

对一个 8×8 图像子块作 DCT 变换得到一个 8×8 的 DCT 系数块 C_{uv} , 其中, u 和 v 为空间频率分量.将 8×8 DCT 系数近似地看成一个多通道结构,每个系数一个通道.设每个通道的视觉阈值为 t_{uv} ,从而形成一个 8×8 的视觉阈值矩阵 $[t_{uv}]$.

MPEG 中的量化过程为

$$\bar{c}_{uv} = \left\lfloor \frac{c_{uv}}{\frac{w_{uv}}{w_{00}} Mquant} \right\rfloor \quad (5)$$

式中, c_{uv} 和 \bar{c}_{uv} 分别是量化前后的 DCT 系数值, $Mquant$ 是量化尺度因子, w_{uv} 是量化的视觉权重矩阵.在 MPEG 中给出了缺

省量化矩阵,它们是在大量图像进行视觉心理实验的基础上,经过统计分析得出的,来源于图像,又独立于具体图像的.本文就以它为基本依据来确定基本视觉阈值 t_{uv}].

c_{uv} 与 \bar{c}_{uv} 之差称为量化误差 e_{uv} ,由式(5)可知,当 $M_{quant} = 1$ 时,最大可能的量化误差是 $w_{uv}/(w_{00} \times 2)$,当实际的量化误差小于这一值时,人眼就感觉不到误差的存在,所以这里取 $M_{quant} = 1$ 时量化误差最大可能值为基本视觉阈值,即

$$t_{uv} = e_{uv}|_{\max} = w_{uv}/(w_{00} \times 2) \quad (6)$$

基本视觉阈值是独立图像内容的,而实际的视觉阈值随着图像块特性的变化而变化,为正确度量图像的感觉质量,必须在基本视觉阈值基础上,再基于图像的内容来确定最终的视觉阈值.

3.2 基于视频内容的视觉阈值

特定视频中图像的视觉阈值是对独立于图像的视觉阈值的增强,用公式表示为

$$M(u, v) = (1 +) T(u, v) \quad (7)$$

式中, $M(u, v)$ 是这个特定图像块的视觉阈值矩阵, $T(u, v)$ 是基本视觉阈值矩阵,它与图像子块的特性无关.式中的标量为视觉阈值提升因子,这是本文要求的量.

3.2.1 视觉阈值提升因子的模糊计算 尽管人们发现了视觉的掩蔽现象,也建立了一些掩蔽的数学模型^[7],但是由于掩蔽现象太复杂,而且它还和视觉心理有关,所以要建立一个精确的数学模型还有很长的路要走,目前还没有一个被普遍接受的模型,在这种情况下,本文提出用模糊逻辑的思想来描述掩蔽现象可能是比较适合的,因为它把人的经验、知识变成计算机可接受的规则,从而可有效地进行掩蔽计算.

(1) 规则库

用模糊逻辑来推理判决一个具体的 DCT 系数块各系数的视觉阈值的关键是建立判决规则库.根据上面的研究可建立如下的规则库:

规则 1:if sp is BG_1 then increase;

规则 2:if b is D or b is B then increase, and will increase more when b is D than when b is B ;

规则 3:if m is BG_2 then increase;

规则 4:if d is BG_3 then increase.

这些规则建立在各自的模糊集上.

(2) 模糊集

图像子块亮度模糊集

人眼对暗区、亮区及中等亮区的损伤的敏感性不同,我们只对属于暗和亮的块感兴趣,所以本文定义了两个模糊子集对应图像子块的亮度 b : D (Dark) 和 B (Bright),它们的隶属函数分别为 μ_D 和 μ_B .

图像子块复杂度模糊集

根据图像子块的空间变换率,将子块分为两类:平滑块和纹理块.由于我们只对纹理块感兴趣,所以将平滑块称为非纹理块.这样,图像子块的复杂度只用一个模糊集 BG_1 就可表示了,其隶属函数为 μ_A .

图像子块运动性模糊集

同样,图像子块的运动性也用一模糊集 BG_2 就可表示

了,其隶属函数记为 μ_M .

场景切换模糊集

前面用“帧间距离” d 来度量两帧间的变化,并规定一个阈值判别是否发生场景切换,实际上,由于视频内容的千变万化,很难用一个固定的阈值来判别是否发生场景切换,这里定义一个模糊集 BG_3 来表示两帧间的不同,其论域为两帧间的距离 d ,隶属函数记为 μ_S .

隶属函数 μ_B 、 μ_A 、 μ_M 和 μ_S 的形式如式(8),形状如图 1(a) 左边所示.隶属函数 μ_D 形式如式(9),形状如图 1(b) 左边所示.不同的隶属函数,参数 L 、 R 不同, x 代表不同的图像特性.

$$\mu_1(x) = \begin{cases} (x - L)/(R - L), & L \leq x \leq R \\ 1, & x > R \\ 0, & \text{otherwise} \end{cases} \quad (8)$$

$$\mu_2(x) = \begin{cases} (x - R)/(L - R), & L \leq x \leq R \\ 1, & x \leq L \\ 0, & \text{otherwise} \end{cases} \quad (9)$$

不同的图像特性对视觉阈值提升因子的影响不同,为了进行模糊推理,还要为不同属性的提升因子定义不同的模糊集,各个属性所对应的提升因子的隶属函数形式统一表示为式(10)所示,如图 1 右边所示.

$$\mu(i) = \begin{cases} i/i_{\max}, & i \leq i_{\max} \\ 1, & i > i_{\max} \\ 0, & \text{otherwise} \end{cases} \quad (10)$$

式中, i 分别是四个属性之一, i 是各个属性所对应的提升因子,属性不同,提升因子不同.

3.2.2 模糊推理 有了规则库和模糊集就可通过模糊推理获得各个图像块属性所对应的阈值提升因子,推理过程如图 1 所示.最终阈值的综合提升因子是四个特性确定的各个提升因子的和.

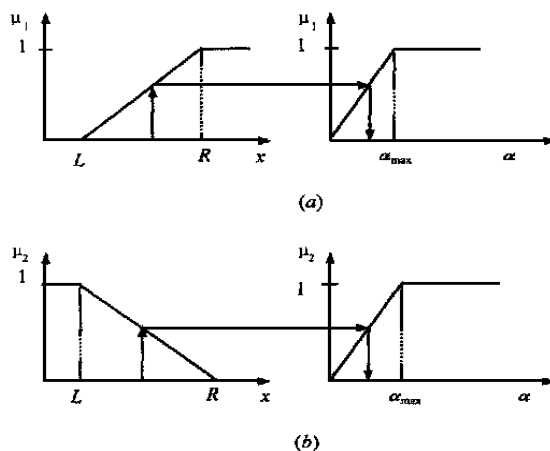


图 1 模糊推理过程.(a)用于图像子块复杂度、运动性、亮度特性中暗区和场景切换的模糊推理;(b)用于图像子块亮度特性中亮区的模糊推理

4 视频感觉质量的计算

4.1 计算模型

综上所述,图像子块的量化损伤由两部分构成:视觉可见部分与视觉不可见部分,一个损伤属于何者取决于视觉阈值.根据图像子块特性及基本视觉阈值求得实际的视觉阈值后,将小于阈值的损伤掩蔽了,剩余的损伤就是影响图像感觉质量的损伤.基于这样的基本思想,得出如图2所示的计算模型.

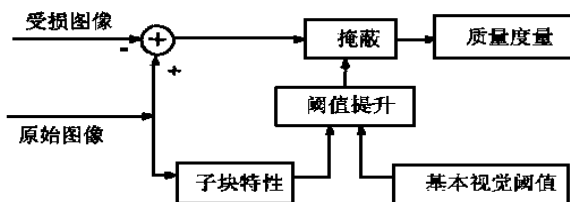


图2 压缩视频感觉质量计算模型

4.2 感觉峰值信噪比

对视觉可见部分的量化损伤可以用感觉峰值信噪比(PPSNR)来表示其视觉效果,其定义如下:

$$PPSNR = 20 \log_{10} \frac{255}{\sqrt{E[|C(u, v) - C(\hat{u}, v)| - M(u, v)]^2 \cdot (u, v)}} \quad (11)$$

其中

$$(u, v) = \begin{cases} 1, & \text{if } |C(u, v) - C(\hat{u}, v)| > M(u, v) \\ 0, & \text{if } |C(u, v) - C(\hat{u}, v)| \leq M(u, v) \end{cases} \quad (12)$$

式中, $C(u, v)$ 表示在 8×8 DCT 系数块中的系数, $C(\hat{u}, v)$ 为去量化后的 DCT 系数. 上式的含义为只用量化误差幅度大于视觉阈值的量化误差值来计算压缩视频的质量.

5 实验结果

图像质量度量实验用一个 50 帧、352 × 288 的序列图像进行, 其中前 25 帧为 FlowerGarden 序列, 后 25 帧为 Mobile&Calendar 序列. 将这一序列分别用 TM5 MPEG2 和作者提出的均匀质量编码算法^[9]进行编解码后, 再进行质量的度量. 实验中确定的参数如表 1 所示.

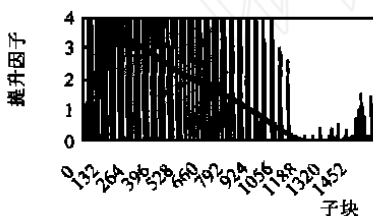


图3 第2帧亮度提升因子

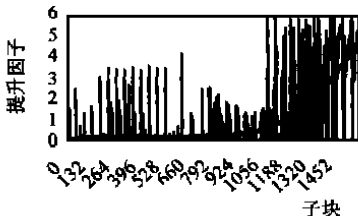


图4 第30帧亮度提升因子

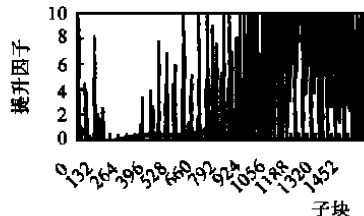


图5 第2帧复杂度提升因子

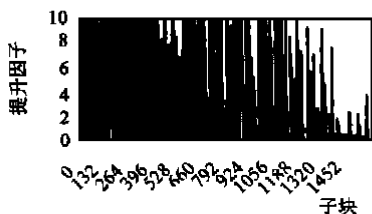


图6 第30帧复杂度提升因子

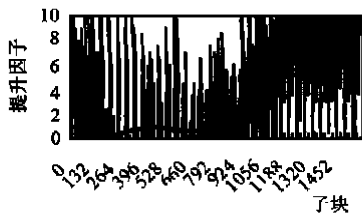


图7 第2帧运动提升因子

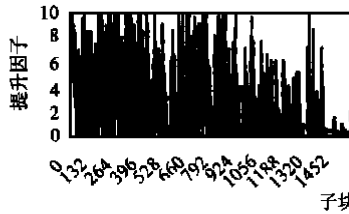


图8 第30帧运动提升因子

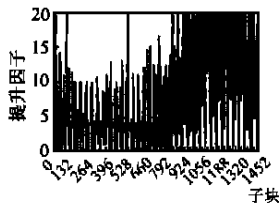


图9 第2帧综合提升因子

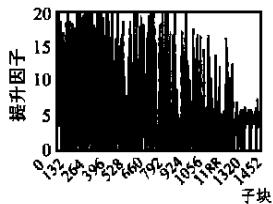


图10 第30帧综合提升因子

表1 模糊集实验参数

图像特性	输入隶属函数参数(L, R)	输出隶属函数参数(L, R)
亮度暗区	(30, 80)	(0, 6)
亮度亮区	(190, 230)	(0, 4)
子块活动性	(15, 30)	(0, 10)
子块运动性	(10, 30)	(0, 7)
场景切换	(0.4, 0.7)	(0, 5)

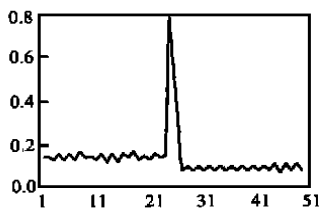


图11 场景切换因子

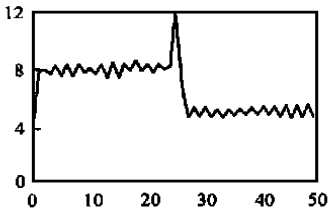


图12 50帧序列图像的提升因子

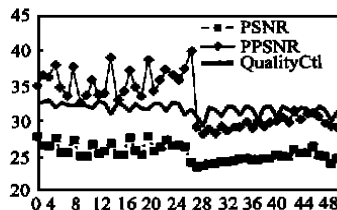


图13 50帧序列图像的图像质量

下面给出部分实验结果和变化的趋势.限于篇幅,这里以第 2 帧(FlowerGarden)和第 30 帧(Mobile)为例给出了部分实验结果.实验中,取阈值为 40,图 11 是相邻帧间的距离度量值,图中横坐标 0 表示第 1 帧相对于第 0 帧,依此类推,可见在视频序列的第 25 帧发生场景切换.图 12 是每帧图像的平均提升因子.图 13 是计算度量结果,图中 QualityCtl 是作者用本文的度量方法控制编码质量所得的结果^[9].

6 结束语

本文提出了编码图像感觉质量的计算度量方法.不同的编码方法.为了控制其编码质量,应采用与编码方法相适应的感觉质量度量方法.基于 DCT 块的视频图像压缩以 8×8 子块为量化单元,所以本文所提出的质量度量方法也是以子块为基础的.本文提出的图像质量度量的核心是视觉阈值及其提升,掩蔽发生在图像的局部,所以图像中的不同位置有不同程度的视觉阈值提升.实验表明这种度量方法是符合人眼视觉特性的.

参考文献

- [1] A. M. Eskicioglu, P. Fisher, Image quality measures and their performance, IEEE Trans. Commun. 1995, 43(12):2959~2965
- [2] B. Zhu, A. H. Tewfik, O. N. Greek, Image coding with mixed representations and visual masking, ICASSP '95:2327~2330
- [3] B. Macq, M. Mattavelli, O. V. Calster, Image visual quality restoration by cancellation of the unmasked noise, ICASSP '94, v53~v56
- [4] W. J. Tam, L. Stelmach, L. Wang, et. al. Visual masking at video scene cuts, SPIE V2411:111~118
- [5] P. C. Teo, D. J. Heeger, Perceptual image distortion, ICIP-94:982~986
- [6] T. D. Tran, R. Safranek, A locally adaptive perceptual masking threshold model for image coding, ICASSP '96:1882~1885
- [7] C. J. Van, B. Lambrecht, A working spatio-temporal model of the human visual system for image restoration and quality assessment applications, ICASSP '96:2291~2294
- [8] S. J. Westen, R. L. Lagendijk, J. Biemond, Perceptual image quality based on a multiple channel HVS model, ICASSP '95:2351~2354
- [9] 宋建新. 基于 ATM 的变比特率视频预编码与传输关键技术与理论的研究, 博士论文, 上海交通大学, 1998. 6



宋建新 1959 年生于江苏, 分别于 1982 年和 1990 年在南京邮电学院无线电工程系获学士、硕士学位, 1998 年在上海交通大学电子工程系获博士学位. 现为南京邮电学院副教授. 出版著作两部, 发表论文 10 余篇. 目前主要研究方向: IP 网和 ATM 网上图像与多媒体通信. E-mail: songjx@126.com



毕厚杰 1932 年生. 1953 年毕业于上海交通大学电子工程系. 现为上海交通大学、南京邮电学院和南京大学通信学院教授, 博士生导师. 发表图像通信工程、图像编码基础等论著 8 部, 论文 70 余篇.