

基于人眼感知特性的亮度系数压缩方法

喻 莉 郭 珊 徐士麟 周 刚 李 荣

(华中科技大学电信系, 武汉光电国家实验室(筹), 武汉 430074)

摘 要 在传统视频编码系统中, 尽管失真大多由均方误差 (MSE) 度量, 但是基于 MSE 的失真往往难以衡量不同视频流的主观差异, 因此, 人眼视觉系统对视频流的感知特性有必要被编码器利用。为了进一步提高编码效率, 针对人眼对不同亮度的信号敏感程度不同的特性, 提出了一种基于人眼感知特性的亮度系数压缩算法, 该算法通过前向量化将人眼不能察觉的冗余信息丢掉来提高编码器压缩效率, 并保证了人眼对损失的信息不可见。实验结果表明, 采用该算法的 AVS 参考编码器, 其输出码率的下降幅度达到 8% ~ 40%, 而解码图像的主观观测质量却同未采用该算法的编码器相当。

关键词 视觉感知编码 人眼视觉系统 视觉冗余 韦伯定律

中图法分类号: TN919.81 文献标识码: A 文章编号: 1006-8961(2009)03-0452-06

A Perceptual Coding Method Based on the Compression of Luma Coefficients

YU Li, GUO Shan, XU Shi-lin, ZHOU Gang, LI Rong

(Electronics & Information Engineering Department, Huazhong University of Science and Technology,
Wuhan National Laboratory for Optoelectronics, Wuhan 430074)

Abstract In the conventional video coding system, the distortion is always measured by Mean-square-error (MSE). However, the MSE-based distortion fails to measure the subjective difference of videos. As a result, the property and interest of human visual system (HVS) should be considered by video coder. This paper proposes a new perceptual coding method aiming at compressing the luma coefficients in terms of the eye's sensitivity. By using the pre-quantization strategy, it discards the imperceptible information and improves the compression performance of video codec without contaminating the video subjective quality. The experiment result proves that this perceptual coding method can efficiently reduce the output bit rate of AVS reference software by 8% ~ 40% while guaranteeing the quality of decoded video.

Keywords perceptual coding, human visual system, visual redundancy, Weber's law

1 引言

迄今为止, 已出现了许多优秀的视频压缩标准, 这些标准在保证视频解码质量的同时, 极大提高了视频数据的压缩效率, 以方便码流在有限的信道中

传输。比如国际上广泛使用的 MPEG-4、H. 264^[1] 以及由我国自行设计开发的数字音频视频编解码技术标准 AVS (advanced audio/video coding standard)^[2]。其中, 如何去掉时间冗余、空间冗余和编码冗余是现今大多数编码器所关心的主要问题。其中时间冗余一般可以通过前向预测和后向预测、插值等方法去

基金项目: 国家自然科学基金项目 (NSFC 60502023); 新世纪优秀人才支持计划项目 (NCET070339); 湖北省自然科学基金项目 (2005ABA254)。

收稿日期: 2007-05-21; **改回日期:** 2007-08-31

第一作者简介: 喻 莉 (1970 ~), 女, 教授, 博士生导师。现为国家“数字音视频编解码技术标准”专家组成员, 中国图象图形学协会会员, 湖北省通信专家。一直从事计算机网络、移动通信、多媒体信息处理、数据压缩编码等方面的科研和开发工作, 已发表论文 30 余篇, 申请发明专利 7 项。E-mail: lyuyxt@sina.com

除;空间相关性可通过空间变换(如离散余弦变换(DCT)、小波变换)等方式消除;编码冗余则是由熵编码器消除。然而,人眼才是所有视频的最终接受体,而根据人眼视觉系统的特性,图像或视频数据中还存在较多的视觉心理冗余成分。因此,如何有效地消除视觉冗余就成为了视频编码领域的一大挑战。基于视觉特性的视频编码器可以概述为:在保证输出视频具有一定的主观质量的前提下,用最少的比特来编码视频信号。

人眼视觉系统的生理神经结构非常复杂,现在人们对其的理解和研究还不是很深入,但大量的实验结果表明,人类的视觉系统并不能感知图像中出现的所有变化细节。由于大多数视频编码器都会产生失真,因此可将能被视觉系统感知的最小失真(或不能被感知的最大失真)定义为 JND (just noticeable distortion)^[3]。JND 提供了一个门限,而在此门限以下的误差则被视为不可见的。

通过研究发现,在实际图像中,存在大量的人类视觉系统不能察觉的描述细节变化的数据,即使经过变换、量化和熵编码后,仍然存在一些为人眼所不能察觉的冗余信息。尤其是在背景亮度较高的区域,人眼对于失真的敏感程度大大降低,这就为进一步压缩视频中的冗余亮度信息提供了可能。另外,在以 YUV 域为基础的视频编解码系统中,由于亮度系数(Y分量)携带了大部分的信息数据,因此能有效压缩视频中的亮度信息,也就使编码系统中压缩率大幅提高。

总而言之,基于视觉感知的视频压缩算法的核心思想是区分能否被人眼感知的信号分量,以尽量去除人眼不能察觉的或者对人眼观测贡献较小的冗余信息,并在保证人眼主观观测效果一致的情况下,提高系统对视频的压缩率和降低传输码率。这种基于人眼视觉特性的编码方式我们称之为视觉感知图像编码(PIC)^[4]方法。其实早先的一些编码方法^[5]也考虑到了人眼视觉特性,例如将 DCT 系数的低频部分采取细量化,而对高频部分则采取粗量化^[6],即利用人眼对高频信息不敏感的特性,使得 DCT 变换和量化能更有效地降低输出码率。其他一些已知的基于人眼视觉系统(HVS)的压缩算法也主要用于小波变换域或 DCT 变换域^[7]。

本文提出的方法是针对视频信号中的亮度信息部分,首先区分高亮度区域和普通亮度区域,并对高亮度区域的亮度系数进行预压缩。此方法直接在像

素域内处理视频信号,通过基于 JND 的前向量化的方法预压缩残差系数,再进行变换量化。同时,由于本方法不参与模式选择和 RDO (rate distortion optimization),因此其算法实现的复杂度并不高。实验结果表明,在解码视频主观观测质量基本未改变的情况下,比特率下降了 8% ~ 40%,从而提高了系统的压缩效率。

2 基于韦伯原则的 JND 建模

通过对人眼视觉系统的生理神经结构的理解和研究发现,由于人类的视觉系统并不能感知图像中出现的所有变化细节,而在实际图像中却存在大量人类视觉系统不能察觉的用于描述细节变化的数据,因此,图像数据中存在着大量视觉冗余。为此,本文引入了阈值化操作,即只有当像素的像素值变化大于某个阈值的时候,人眼才能感知到变化的发生。早在 19 世纪初,Weber 就给出了这一现象的数学描述,即

$$\Delta S = K \cdot S \quad (1)$$

其中, K 是常数, S 是参考信号的幅值, ΔS 是人能区分当前信号与参考信号差异的门限值。换句话说,人区别变化的能力同参考信号的幅值是成反比的,参考信号的幅值越大,人的敏感程度越低,能区别变化的能力就越弱。

19 世纪 60 年代,Fechner 将 Weber 法则应用于人的感官实践中,用外界的刺激强度当作参考信号,发现人的感觉强度(intensity of a sensation)同外界刺激的强度的对数成正比,于是公式(1)可修改为

$$I = K \cdot \log \frac{S}{T} \quad (2)$$

其中, I 是人的感知敏感度度量, K 是常数, T 是一门限值,即人能感知的信号的最小强度,而 S 则是大于 T 的外界刺激强度。人的感觉强度同外界刺激强度的关系式即被称为 Weber-Fechner 法则^[8],该法则适应于人的各个感官,如视觉、听觉、触觉等。

因此,可以定义一个信号误差的可视门限值,这样当某个信号的误差大于该门限时,即能被人眼所感知。应该指出的是,可视门限并不是一个固定的值,应随着信号时间、空间、频率上的变化而变化。显然,在有损的视频编码器中,如果失真小于 ΔS ,那么此类失真可以允许存在;反之,大于 ΔS 的失真应

该尽量避免。因此,结合式(1),在 PIC 方法中,有

$$J = \Delta S = K \cdot S \tag{3}$$

式中,符号 J 代表 JND 的值。由式(3)可知, J 同信号强度 S 是成正比的,即信号强度越大, J 的值也越大,人眼能接受的信号误差也越大,而相应的信号中的视觉冗余信息也越多,信号可供压缩的幅度也越大。

3 基于视觉感知的编码模型

3.1 基于 JND 的前向量化

由第 1 节、第 2 节的叙述可以容易地推断出以下结论:当一个像素的像素值的变化幅度在当前像素的 JND 值以内时,则人眼不能发觉其变化。换句话说,若当前像素的真实像素值为 V_{curr} ,则当 $V \in [V_{curr} - V_{JND}, V_{curr} + V_{JND}]$ 时,人眼是不能分辨 V_{curr} 同 V 的差别的,这也就为压缩视频序列中的视觉冗余信息提供了可能。另一方面,由于 J 的值同信号本身的强度呈线性关系,即信号强度越大, J 的值就越大,人眼就越不能察觉出信号的变化,因此对于各种亮度级别的信号最好采用不同尺度的压缩方法,才能尽可能地去掉信息中的视觉冗余部分,以提高编码器的压缩率。

设当前像素的亮度分量的真实值为 Y ,其对应

的 JND 值为 Y_{JND} ,经过式(4)压缩后变为 Y_C ,在解码端,若要由 Y_C 得到 Y 的重构值 Y_R ,则只需进行下式的逆变换即可(如式(5)所示):

$$Y_C = \left\lfloor \frac{(Y + Y_{JND})}{2Y_{JND}} \right\rfloor \tag{4}$$

$$Y_R = Y_C \cdot (2Y_{JND}) \tag{5}$$

此时重构的像素值与真实值之间的误差为

$$\xi = Y - Y_R = Y - \left\lfloor \frac{(Y + Y_{JND})}{2Y_{JND}} \right\rfloor \cdot (2Y_{JND}) \tag{6}$$

由于 Y 和 Y_{JND} 均为正整数,因此由式(6)可以很明显的看出, $\xi \in [-Y_{JND}, Y_{JND}]$ 之间。由于小于 Y_{JND} 的误差人眼不能察觉,因此此时的压缩对于人眼来说是无损压缩,而且式(4)将大大降低编码器的输出比特率,从而可实现编码器压缩率的提升。

由式(4)和式(5)描述的过程可知,在笔者设计的 PIC 编码器中称为前向量化。如图 1 所示,对于基于运动估计和运动补偿的编码器来说,当一个宏块的编码模式由率失真优化(RDO)确定后,每个残差系数先将依据式(4)进行前向量化,然后被前向量化后的残差系数再进行标准规定的 DCT 变换和量化。那么在重构端,经过反量化和反 DCT 变换的残差系数将依据式(5)进行反前向量化。

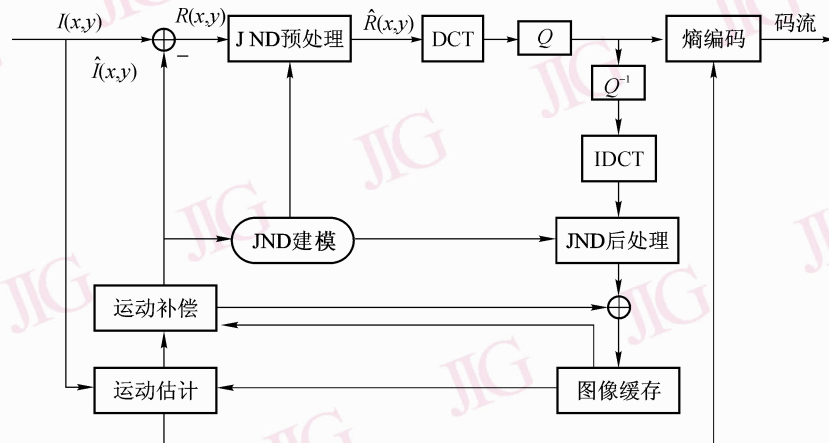


图 1 视觉感知编码器(PIC)结构

Fig. 1 Perceptual coding codec framework

3.2 具体实现

由前面的论述可知,基于 JND 的前向量化虽然在客观上会降低信噪比,但对于人眼来说却是无损的,这样就可取得用尽量少的比特来表征图像或视

频而不影响重构图像的主观质量的效果。根据式(3),理论上不同的亮度值对应不同的 JND 值 Y_{JND} ,但是考虑到算法的复杂度,特将 Y_{JND} 进行了“分段化”处理,即将在 $[0, 256]$ 中取值的亮度系数分为几

个子区间,对于每个子区间中的所有亮度系数, Y_{JND} 值是不变的。大量的测试数据和实验结果表明,如果将图像或序列的像素亮度分量划分为3个不同的亮度级别分别建立 JND 模型,则对于绝大多数测试序列都能取得较好的压缩效果,其对应的 Y_{JND} 取值即分别为:0,1,2(根据实验的经验数据得到)。当然, Y_{JND} 的取值可以随着以后研究的不断深入而改进,以求达到压缩率和主观观测效果的最佳平衡。

为了便于描述算法的具体实现步骤,本文先定义 θ_1 和 θ_2 ($\theta_1 > \theta_2$) 两个门限,若当前像素的亮度分量大于 θ_1 时,则认为当前像素属于高亮度区域;若当前像素的亮度分量小于 θ_1 且大于 θ_2 时,则认为当前像素属于普通亮度区域;若当前像素的亮度分量小于 θ_2 时,则认为当前像素属于低亮度区域。其中 θ_1 和 θ_2 的取值主要依靠实验测试数据得出的经验值确定。

值得注意的是:由于在图像的解码过程中,待解码像素的信息并不能预先得到,因此无法在解码前就进行如上叙述的 JND 建模,致使解码器将无法通过式(5)描述的方法正确恢复待解码像素。所以,应考虑利用待解码像素的参考像素来估计 JND 值,而且参考块本身就是当前块在前(后)向帧中的最佳匹配块,事实证明,用参考块中的亮度系数来大致估计当前块的亮度水平的准确率还是很高的。

运动估计的目标之一就是找寻当前块在参考帧中的最佳匹配块,由于可采取适当的方法和准则来保证参考块选取的精度,因此可以认为当前编码像素的 JND 等于其参考像素的 JND。另外,AVS-M 和 H.264 中均采用 4×4 的整数 DCT 变换,为适应整个编码器的结构,要尽可能小地改动整个编码流程,可以 4×4 块为基本单位逆向找出其对应的参考块,以方便其后续的处理过程。

该算法的主要实现步骤如下:

(1)通过参考像素进行 JND 建模

设与当前编码像素对应的参考像素的亮度分量为 Y_{pred} ,则 Y_{JND} 可由下式计算:

$$Y_{\text{JND}} = \begin{cases} 2 & Y_{\text{pred}} > \theta_1 \\ 1 & \theta_1 > Y_{\text{pred}} > \theta_2 \\ 0 & Y_{\text{pred}} < \theta_2 \end{cases} \quad (7)$$

(2)PIC 设计

记当前像素的亮度残差系数为 R_{curr}^Y ,其经过前向量化后变为 \hat{R}_{curr}^Y 。量化原则是利用式(4)对高亮

度区域的 R_{curr}^Y 进行粗量化和对普通亮度区域的 R_{curr}^Y 进行细量化,而对低亮度区域的 R_{curr}^Y 则不进行任何处理。

根据式(7)和式(4),压缩过程可以描述为

$$\hat{R}_{\text{curr}}^Y = \begin{cases} (R_{\text{curr}}^Y + 2) \gg 2 & Y_{\text{pred}} > \theta_1 \\ (R_{\text{curr}}^Y + 1) \gg 1 & \theta_1 > Y_{\text{pred}} > \theta_2 \\ R_{\text{curr}}^Y & Y_{\text{pred}} < \theta_2 \end{cases} \quad (8)$$

此后,对 \hat{R}_{curr}^Y 进行后续的 DCT 变换、量化、熵编码等过程,编码器结构框图如图 1 所示。其中 $I(x,y)$ 和 $\hat{I}(x,y)$ 分别为原始的和重建的当前像素值, $R(x,y)$ 和 $\hat{R}(x,y)$ 分别为未经过和经过前向量化的残差系数。式(7)、式(4)和式(5)描述的处理过程分别为“JND 建模”、“JND 预处理”和“JND 后处理”3 个模块。

解码的过程即为编码的逆过程,即在解码端,将接收到的码流进行反熵编码、反量化、反 DCT 变换等视频标准中所定义的过程,即首先得到初始的残差系数;然后将其中的亮度残差部分通过反前向量化来得到最后的重构残差系数。设恢复后的亮度残差系数为 $\tilde{R}_{\text{curr}}^Y$,其解码过程可以表述为

$$\tilde{R}_{\text{curr}}^Y = \begin{cases} \hat{R}_{\text{curr}}^Y \ll 2 & Y_{\text{pred}} > \theta_1 \\ \hat{R}_{\text{curr}}^Y \ll 1 & \theta_1 \geq Y_{\text{pred}} \geq \theta_2 \\ \hat{R}_{\text{curr}}^Y & Y_{\text{pred}} < \theta_2 \end{cases} \quad (9)$$

4 实验结果

本文算法已在 AVS-M 标准参考软件 wm2.5a 中实现。本实验的测试对象是“bus”、“coastguard”、“foreman”、“paris”、“tempeste”、“football”等 6 个 CIF 标准视频序列(图像大小为 352×288),每个序列编码 50 帧,除第 1 帧为 I 帧外,其余均为 P 帧,即序列结构为 IPPPPPP…。在帧率为 30fps, RDO 打开的状态下,量化参数(QP)值逐个取 15~30 时分别统计测试结果,并同未使用本文算法的 AVS-M 参考软件进行比较的统计结果表明,在门限值 θ_1 和 θ_2 (见 3.2 节)分别取 140 和 110 时,测试结果较好。此时的实验结果如图 2 所示,其中横轴表示 QP 的取值,纵轴表示使用本文算法的 PIC 编码器编码的比特率较未使用本文算法的 AVS-M 参考软件中的编码器的编码比特率的下降幅度,不同符号的曲线

则表示不同序列的测试结果。图 2 直观地反映出 PIC 编码器编码的比特率同 AVS-M 标准编码器的比特率相比有大幅度的下降,达到 8% ~ 41%,同时 PIC 编码器还保证了解码端同样的主观视觉效果,并有效平衡了主观解码质量和编码比特率的问题,图 3 则给出了 3 个典型的主观观测效果图。

视频序列,比特率均存在大幅度的下降,其中 foreman 序列在 QP 值等于 18 时的下降幅度高达 35% 以上(如图 3(a)、图 3(b)所示),coastguard 序列的下降幅度基本也在 35% 以上(如图 3(c)、(d)所示),tempeste 序列在 QP 值等于 30 时,码率下降了 13.6% (如图 3(e)、(f)所示),而主观观测效果却与 AVS 标准解码器一致。

为了更为公平地评判 PIC 方法和 AVS-M 标准编码器解码图像的主观视觉效果,本文引用国际电信联盟 (ITU) 推荐的电视图像主观质量评分方法,将图像质量分为 5 个等级 (见表 1),在同样的观测条件下找 10 人,观测相同的 QP 值下的 AVS-M 解码器和 PIC 解码器的解码码流,并按评分标准为每个观测序列的主观质量打分。图 4 给出了评分结果,其中横轴表示图 2 中的 6 个测试序列,纵轴表示每个序列的得分情况。结果显示,绝大部分人认可本文算法解码的主观视觉效果,因此本文算法的使用并不影响视频序列的主观视觉效果。

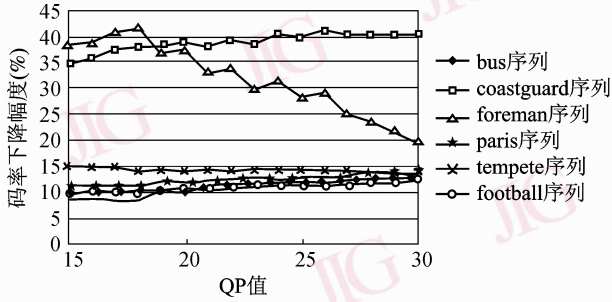


图 2 6 个 CIF 标准序列的测试结果
Fig. 2 Six CIF sequences' test results

从图 2 还能清楚地看到,采用本文算法编码的



(a) QP 值为 18, foreman 序列
第 25 帧 AVS-M 解码图
(编码比特数为 73 680 bits)



(b) QP 值为 18, foreman 序列
第 25 帧 PIC 解码图
(编码比特数为 39 888 bits)



(c) QP 值为 25, coastguard 序列
第 33 帧 AVS-M 解码图
(编码比特数为 88 056 bits)



(d) QP 值为 25, coastguard 序列
第 33 帧 PIC 解码图
(编码比特数为 51 136 bits)



(e) QP 值为 30, tempeste 序列
第 16 帧 AVS-M 解码图
(编码比特数为 54 504 bits)



(f) QP 值为 30, tempeste 序列
第 16 帧 PIC 解码图
(编码比特数为 45 624 bits)

图 3 解码端图像主观视觉效果对比图

Fig. 3 Contrast decoded pictures subjectively

表1 ITU-R 质量和受损程度表

Tab.1 ITU-R quality and impairment scales

| 质量评价 | 受损程度 |
|-------|---------|
| 5 非常好 | 无察觉 |
| 4 好 | 刚察觉但不讨厌 |
| 3 一般 | 有点讨厌 |
| 2 差 | 讨厌 |
| 1 非常差 | 非常讨厌 |

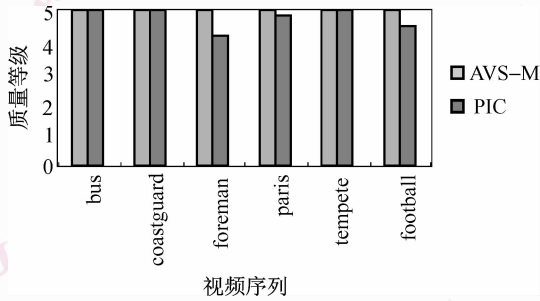


图4 视频序列主观质量评分对比图

Fig.4 Subject assessment for video sequences

5 结论

本文针对人眼对高亮度区域不敏感的特性,提出了一种基于人眼视觉感知的亮度系数压缩算法,用该算法将人眼不能感知的高亮度区域的信息进行适当的压缩,并保证失真范围在人眼可观测的门限值以下,由于有效去除了部分视觉冗余信息,因而不仅降低了视频编码端的码率,也提高了整个编解码器的压缩率。由于人眼对此部分信息的丢失不可见,

因此可保证解码端视频图像的重构质量。实验结果证明,本文算法同标准 AVS-M 编解码器的压缩结果相比,foreman 序列和 coastguard 序列的码率下降了 30% ~ 40%,但主观视觉几乎没有差异。本文算法用于视频流在有限带宽信道中的传输,可提高视频数据的传输速率。

参考文献 (References)

- 1 Wiegand T, Sullivan G J, Bjontegaard Gisle, *et al.* Overview of the H. 264/AVC video coding standard [J]. IEEE Transactions on Circuits and System, Video Technology, 2003, **13**(7):560-576.
- 2 AVS. Information Technology-Advanced Coding of Audio and Video-Part 7: Mobile Video [S]. AVS-M, Mar, 2005. [AVS. 信息技术——先进音视频编码第 7 部分:移动视频(送审稿) [S]. AVS-M, 2005 年 3 月].
- 3 Chou C H, Li Y C. A perceptually tuned subband image coder based on the measure of just-noticeable-distortion profile [J]. IEEE Transactions on Circuits System Video Technology, 1995, **5**(6): 467-476.
- 4 Höntsch Ingo, Karam Lina J. Adaptive image coding with perceptual distortion control [J]. IEEE Transactions on Image Processing, 2002, **11**(3): 213-222.
- 5 Tong H Y, Venetsanopoulos A N. A perceptual model for JPEG applications based on block classification, texture masking, and luminance masking [A]. In: Proceedings of IEEE International Conference on Image Processing (ICIP98) [C], Chicago, Illinois, USA, 1998: 428-431.
- 6 Ahumada Jr A J, Peterson H A. Luminance-model-based DCT quantization for color image compression [J]. Proceedings of the SPIE, 1992, **1666**:365-374.
- 7 Höntsch Ingo, Karam Lina J. Locally adaptive perceptual image coding [J]. IEEE Transactions on Image Processing, 2000, **9**(9): 1472-1483.
- 8 Boring E G. Sensation and Perception in the History of Experimental Psychology [M]. New York: Appleton-Century, 1942.