文章编号 1674-2915(2009)06-0531-07

基于平均预测残差的 H.264 宏块层码率控制算法

陈 静,赵 建,周 游
(中国科学院 长春光学精密机械与物理研究所,吉林 长春 130033)

摘要:针对 H.264 码率控制模型的"蛋鸡悖论",提出一种用平均预测残差来表示当前宏块编码复杂度的算法。首先根据各个宏块不同的编码复杂度进行准确的比特分配,然后根据改进的二次 R-D 预测模型来计算宏块的量化参数,并进行率失真优化,最后进行码率控制模型系数的更新。将算法在 JM8.6 平台上进行实验,结果显示,在相同码率下,本文算法的平均控制精度比 JVT-H017 提高 0.165 kb/s,平均 PSNR 比 JVT-H017 提高 0.13 dB,表明本算法能够更准确控制码率并获得更高质量的主观视觉。

关 键 词:码率控制;视频编码;预测残差
中图分类号:TN919.81 **文献标识码**:A

H.264 macro-block layer rate control algorithm based on average prediction residual

CHEN Jing, ZHAO Jian, ZHOU You

(*Changchun Institute of Optics, Fine Mechanics and Physics, Chinese Academy of Sciences, Changchun* 130033, *China*)

Abstract: In order to solve the problem of "Egg-Chichen Absurdity" in the rate control model H.264, this paper presents an algorithm which can express the coding complexity of current macro-block using the average prediction residual. Firstly, a more precise bit allocation scheme is implemented according to the coding complexity of each macro-block, and then the quantization parameters of macro-block are calculated using the improved quadratic Rate-Distortion(R-D) prediction model and R-D optimization is achieved. Finally, the coefficients of the rate control model have been updated. This algorithm is implemented in JM8.6, and the experimental results show that the proposed method can improve 0.165 kb/s than JVT-H017 in the control accuracy, and improve 0.13 dB in PSNR on the average at the same rate. It can be concluded that the proposed method can control the rate more accurately and can obtain a higher quality subjective vision.

Key words: rate controlling; video coding; prediction residual

收稿日期:2009-08-21;**修订日期**:2009-11-10
基金项目:中国科学院知识创新工程领域前沿资助项目(No.060Y32T060)

1 引 言

H.264 标准中每个视频帧都是以宏块为单位进行编码的,每个宏块产生的编码比特数会根据视频内容的变化而变化,最终导致输出比特率不恒定。为了在固定比特率的信道末端解码出高质量的视频图像,码率控制已经成为编码器中不可或缺的一部分,并广泛地用于 MPEG-4 和H.263 等标准中[1]。通常情况下,编码器通过改变宏块的量化参数对变化的码率进行调节,但是 H.264 的码率控制比其它标准要复杂许多,因为 JM 中码率控制和模式选择都要涉及量化参数。对当前帧进行率失真优化(Rate-distortion Optimization)时,需要利用当前宏块的 MAD 值(Mean Absolutely Difference)计算出每个宏块的量化参数;而当前宏块的 MAD 值又只有在率失真优化完成后才能得到,因此导致了著名的"蛋鸡悖论"(Egg-Chicken Absurdity)[2,3]。

JVT-G012 标准中提出利用时间和空间相关性来解决蛋鸡悖论,用线性 MAD 模型预测编码复杂度,用传统的二次 R-D 预测模型来计算量化参数[4],然后再对当前宏块进行率失真优化。JVT-G012 之后的码率控制标准 JVT-H017 基本沿用了这一方案,但是,预测得出的 MAD 不能完全地反映当前宏块的编码复杂度,由其计算出的量化参数也不够精确,这可能导致实际比特数与目标比特数不相符。本文考虑直接从预测残差的角度来实现码率控制,并提出平均预测残差 APR(Average Prediction Residual)的概念,将 APR 值应用在 H.264 中的目标比特分配,R-D 预测模型以及系数更新中,从根本上摆脱"蛋鸡悖论"的束缚。由于 APR 可以在率失真优化前得到,又能比较准确地反映编码复杂度,因此获得的码率控制精度也更加精确。

2 编码复杂度的概念与平均预测残差 APR 的定义

首先简单介绍一下编码复杂度的概念。在大多数 R-D 模型中,编码复杂度是反映视频内容本质的参数,可以用不同的统计方法来表示编码复杂度。在 MPEG-4 的二次 R-D 预测模型中编码复杂度由 MAD 表示,一个宏块的 MAD 值定义如下:

$$\mathrm{MAD} = \frac{1}{I \times J} \sum_{|i| \leqslant \frac{I}{2}} \sum_{|j| \leqslant \frac{J}{2}} | f(i,j) - g(i,j) |$$

$$I = J = 16, \tag{1}$$

式(1)中,$f(i,j)$是原始图像的像素值,$g(i,j)$是重建图像的像素值。H.264 中进行宏块的编码模式判别时,必须先给出一个量化参数,而量化参数则要根据编码模式判别后的 MAD 来计算,这样就导致了蛋鸡悖论。JVT-G012 中,为解决这一矛盾,待编码宏块的 MAD 值可以通过上一帧中相同位置宏块的 MAD 值预测得到,如图 1 所示。

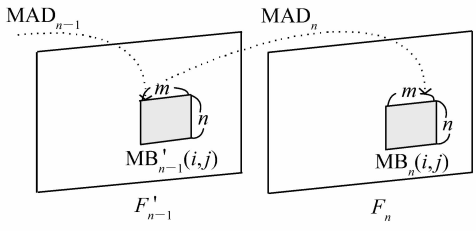

图 1 MAD 的预测示意图

Fig. 1 Sketch of prediction for MAD

线性预测 MAD 公式如下:

$$\mathrm{MAD}_n = \rho \mathrm{MAD}_{n-1} + \gamma, \tag{2}$$

式(2)中,MAD_n 代表当前宏块 MAD 的预测值,MAD_{n-1}代表上一帧相同位置的宏块的 MAD 值,ρ 和 γ 代表两个模型参数,在一个宏块编码完成之后即被更新。由于 MAD 是预测得到的值,并不能够准确地反映当前宏块的编码复杂度,所以没有解决"蛋鸡"悖论的关键。为了克服这一难题,本文提出平均预测残差的概念。

原始宏块经过运动估计后,得到参考宏块,二者相减所得到的残差信息就可直接反映出图像的运动剧烈情况和内容的复杂程度。通常较大的残差说明图像运动剧烈或者内容变化较大;较小的残差意味着图像运动平缓或者在内容上基本没有变化,所以预测残差可以表示宏块的编码复杂度。一个宏块的平均预测残差值定义如下:

$$K_{\mathrm{APR}} = \frac{1}{m \times n} \sum_{x=0}^{m-1} \sum_{y=0}^{n-1} | f(x,y) - f(x - d_x, y - d_y) |, \quad m = n = 16, \tag{3}$$

式(3)中,$f(x,y)$表示原始宏块内坐标为(x,y)的像素点的像素值,$f(x - d_x, y - d_y)$表示参考宏块内对应点的像素值。由于 APR 值可以直接在率失真优化之前得到,又能准确地反映图像的编码复杂度,因此确定的量化参数更加准确。

3　基于宏块层的码率控制实现

本节利用上节提出的平均预测残差的概念来实现宏块层的码率控制算法。此算法包括以下 3 个步骤:(1)根据编码宏块的 APR 值,对每个宏块的目标比特进行更加合理的分配,以减少因比特分配不合理而造成图像质量的下降;(2)根据本文所提出的 R-D 预测模型来计算量化参数;(3)采用滑动窗口机制,筛选有效合理的宏块数据,对二次 R-D 模型参数进行更新。该算法流程图如图 2 所示。

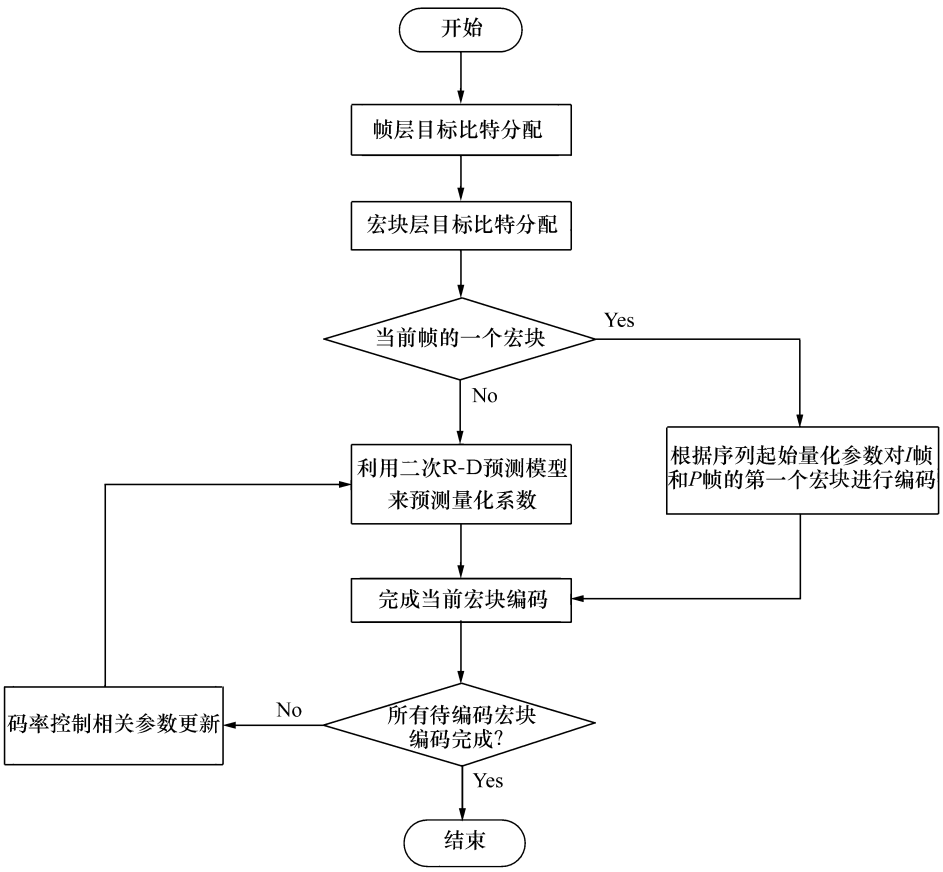

图 2　码率控制算法流程图

Fig. 2　Flow chart of proposed rate control algorithm

3.1　宏块层比特分配

在给宏块进行比特分配之前,首先要对包含该宏块的帧进行比特分配。JVT-H017 标准中,图像组 GOP(Group of Pictures)层的目标比特采用固定值;帧层的目标比特由网络带宽、缓存大小、缓存占用量以及剩余比特共同决定;宏块层的目标比特根据平均绝对差值 MAD 进行分配。假设给定 GOP 结构是第 1 帧为 I 帧,后续帧为 P 帧,且第一个 I 帧量化参数为 24,第一个 P 帧量化参数与之相同,后续 P 帧的目标比特定义为:

$$T_r = \frac{U_r}{N_P} \times (1 - \rho) + T_{r-1} \times \rho, \tag{4}$$

式(4)中,U_r 表示编码第 j 帧时 GOP 内剩下的可分配比特,N_P 表示 GOP 内未编码的 P 帧个数,T_{r-1}表示上一帧的实际比特数,ρ 为权重常数,一般设为 0.5。

对于每一帧的宏块而言,第一个宏块的量化参数由前一帧的各个宏块的平均量化参数确定。对于后续宏块而言,则先根据式(3)计算当前编码宏块 APR 的值,再根据下式来计算当前宏块的目标比特率:

$$R_i = T_r \times \frac{K_{\mathrm{APR}}^2(i)}{\sum_{i=0}^{N_{mb}} K_{\mathrm{APR}}^2(i)}, \tag{5}$$

式(5)中,T_r 表示分配给当前帧的目标比特,$K_{\mathrm{APR}}(i)$表示第 i 个宏块的 APR 值,N_{mb}表示一帧图像内所有宏块的个数。JVT-H017 算法采用预测得到的 MAD 来计算当前宏块目标比特数,而本文算法直接采用宏块的 APR 值来计算。相比之下,本文算法更具有准确性。

3.2 量化参数的计算

本文沿用了 JVT-G012 算法中的二次 R-D 预测模型结构,把该模型中的 MAD 值用 K_{APR}值替代。新的 R-D 预测模型定义如下:

$$\frac{R_i - H_i}{K_{\mathrm{APR}}} = a_1 \cdot Q_i^{-1} + a_2 \cdot Q_i^{-2}, \tag{6}$$

式(6)中,R_i表示第 i 个宏块的目标码率,H_i表示宏块头、运动矢量和语法元素占用的比特数,Q_i表示当前宏块的量化步长,而 a_1、a_2为模型参数,在编码完一个宏块后即被更新。由量化步长计算出当前宏块的量化参数 Q_{pi}之后,为保持连续的视觉平滑感,需要调整 Q_{pi},调整公式如下:

$$Q_{pi} = \min\{Q_{p(i-1)} + 2, \max\{Q_{p(i-1)} - 2, Q_{pi}\}\}, \tag{7}$$

式(7)中,Q_{pi}为当前宏块的 QP 值,$Q_{p(i-1)}$是前一个宏块的 QP 值,最终 Q_{pi}值还要被下式限制:

$$Q_{pi} = \min\{51, \max\{Q_{pi}, 1\}\}, \tag{8}$$

当量化参数 Q_{pi}确定以后,H.264 的编码器通过最小化 Lagrangian 代价函数实现对当前宏块编码模式的确定[5],完成率失真优化。

3.3 二次 R-D 模型参数的更新

Lee 等人在文献[6]中提出将滑动窗口机制应用于 R-D 模型参数的更新,它可以适应平滑场景切换的影响。如果复杂度变换较大,则采用含有最近数据点的较小窗口。通过这样的机制,编码器可以灵活地选取那些比较典型的数据用于 R-D 模型系数的更新。滑动窗口的大小可按照如下方式进行选择:

$$w = \begin{cases} \dfrac{K_i^{\mathrm{APR}}}{K_{i-1}^{\mathrm{APR}}} \times W_{\max}, K_{i-1}^{\mathrm{APR}} > K_i^{\mathrm{APR}} \\ \dfrac{K_{i-1}^{\mathrm{APR}}}{K_i^{\mathrm{APR}}} \times W_{\max}, K_i^{\mathrm{APR}} > K_{i-1}^{\mathrm{APR}} \end{cases}, \tag{9}$$

式(9)中,$W_{\max}$表示最大滑动窗口长度,一般设为 20。取得了滑动窗口大小以后,在窗口的数据集中去除偏差较大的数据点来减小离群数据点对模型参数更新的影响。选取的标准是比较模型预测得到的比特数和实际被使用的比特数之差$\mathrm{err}(i)$是否大于某一基准值 std,根据最小均方估计理论,std 为滑动窗口内所有宏块的目标比特数与实际比特数的标准偏差,二者定义如下:

$$\mathrm{err}(i) = \left(\frac{a_1}{Q_i} + \frac{a_2}{Q_i^2}\right) \times K_{\mathrm{APR}} - R_i \quad i = 1,2\cdots\cdots w$$

$$\mathrm{std} = \sqrt{\frac{1}{w}\sum_{i=0}^{w}\left[\left(\frac{a_1}{Q_i} + \frac{a_2}{Q_i^2}\right) \times K_{\mathrm{APR}} - R_i\right]} \quad i = 1,2\cdots\cdots w \tag{10}$$

其中 Q_i和 R_i 是已编码帧的量化级数和实际比特数,w 是滑动窗口大小。若 $\mathrm{err}(i) > \mathrm{std}$ 则说明第 i 个宏块的实际比特数与目标比特数偏差较大。由此可知这个宏块的数据精确度不够,不能用于 R-D 预测模型参数的更新。

选取具有代表性的宏块数据后,本文采用线性衰落方法来计算式(6)中两个模型参数,其式如下:

$$a_2 = \frac{n \times \sum_{i=1}^{n} R_i - \left(\sum_{i=1}^{n} Q_i^{-1}\right) \times \left(\sum_{i=1}^{n} R_i \times Q_i\right)}{n \times \sum_{i=1}^{n} Q_i^{-2} - \left(\sum_{i=1}^{n} Q_i^{-1}\right)^2}, \tag{11}$$

$$a_1 = \frac{\sum_{i=1}^{n}\left[R_i \times Q_i - a_2 \times Q_i^{-1}\right]}{n},$$

式(11)中,n 表示被选取的宏块个数。

4 实验结果与讨论

在 JVT 的校验模型 JM8.6 上实现本文提出

的码率控制策略,并和 JVT-H017 方案进行比较。本实验使用 QCIF 格式的测试序列 akiyo 和 news 以及 CIF 格式的测试序列 Foreman,帧率为 15 frame/s,给定第一个 *I* 帧的 *QP* 为 24。实验环境如表 1 所示。

表 1　实验环境

Table 1　Experiment conditions

MV 预测	1/4 pixel
Hadamard 变换	ON
RD 优化	ON
搜索范围	16
参考帧	1
熵编码方式	CABAC
GOP 结构	IPPP

表 2 是序列 Foreman 在不同码率下的实际比特率和平均峰值信噪比情况。从表格中可以明显看出,本文算法的实际比特率更靠近给定目标值。经计算得,平均控制精度比 JVT-H017 高 0.165 kb/s左右,而且本文的方法还可以获得更好的峰值信噪比值,序列 Foreman 的峰值信噪比比原算法平均提高约 0.1 dB。

表 2　两种方法的码率控制效果对比

Table 2　Performance comparison of the two rate control methods

目标比特率(kb/s)	实际比特率(kb/s)		PSNR(dB)	
	JVT-H017	本文算法	JVT-H017	本文算法
75.97	76.16	76.12	43.75	43.84
43.64	43.75	43.70	41.12	41.22
26.25	26.32	26.25	35.71	35.82
16.48	16.58	16.49	28.47	28.52

在视频压缩领域中,经常利用峰值信噪比(PSNR)来衡量重建图像的质量[7],PSNR 计算公式如下:

$$\mathrm{PSNR} = -10\lg\frac{\dfrac{1}{M\times N}\sum_{i=1}^{M}\sum_{j=1}^{N}[f(i,j)-g(i,j)]^2}{255\times 255}, \tag{12}$$

式(12)中,M 和 N 分别是图像的宽度和高度,$f(i,j)$是原图像的像素值,$g(i,j)$是重建图像的像素值。图 3(a)是 akiyo 序列在 69 kb/s 下的 PSNR 波动图,(b)是 news 序列在 45 kb/s 下的 PSNR 波动图。从图中可以明显看出,本文算法编码的PSNR明显高于H017。结果表明,在

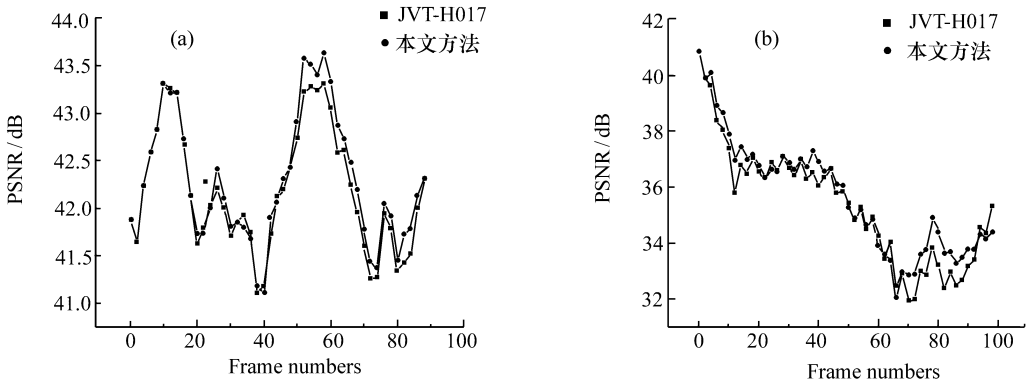

图 3　本文算法和 JVT-H017 算法的编码性能比较曲线

Fig. 3　Comparison of coding performance curve for JVT-H017 and proposed algorithm

akiyo 序列中,本文算法的平均 PSNR 值比 H017 算法高 0.128 dB;在 news 序列中,本文算法的平均 PSNR 值比 H017 算法高 0.15 dB,即本文提出的算法能够获得较高的图像质量。

图 4(a)和图 4(b)分别为 news 序列第 59 帧和第 60 帧在两种方法下的解码图像。其中,图 4(a)为 JVT-H017 算法下的解码图像,图 4(b)为本文算法下的解码图像。将图 4(a)和图 4(b)进行比较可知,本文算法使解码图像质量得到了改善,具有更高的视觉效果。

(a) JVT-H017算法下的图像质量

(a) Image quality by JVT-H017

(b) 本文算法下的图像质量

(b) Image quality by proposed method

图 4 两种方法下的解码图像质量比较

Fig. 4 Comparison of decoding image quality for two methods

5 结 论

本文引入了 APR 概念,并将其应用在H.264 中的目标比特分配、R-D 预测模型以及系数更新中,从根本上摆脱“蛋鸡悖论”的束缚。实验表明,本文方法能够在准确控制码率的同时提高序列的 PSNR。在相同码率下本文算法的平均 PSNR 比 JVT-H017 提高 0.13 dB,平均控制精度比 JVT-H017 提高 0.165 kb/s,说明本算法在获得更好的视频质量的同时,码率控制的精度也随之提高。另外,由于视频编码的计算复杂度主要集中在运动估计和帧内预测上,所以本算法不会影响编码器的实时性。

参考文献:

[1] LI Z G, PAN F, LIM K P. Adaptive basic unit layer rate control for JVT. JVT-G012[A]. Joint Video Team of ISO/IEC and ITU 7th Meeting[C]. Pattaya, Thailand, 2003, 1-33.

[2] 周全,刘文予,崔春晖. 一种低码率下的新型宏块级码率控制算法[J]. 中国图象图形学报, 2006, 11(9): 1217-1222.

ZHOU Q,LIU W Y,CUI CH H. A novel macroblock layer rate control algorithm in low bitrate[J]. *J. Image Graphics*, 2006,11(9):1217-1222. (in Chinese)

[3] 李蔷,余松煜. 一种高效的 H.264 码率控制方法[J]. 上海交通大学学报,2004,38(11):1837-1840.
LI Q,YU S Y. A novel rate control method for H.264[J]. *J. Shanghai Jiaotong University*,2004,38(11):1837-1840. (in Chinese)

[4] 段大高,李文正,莫倩,等. 一种改进的 H.264 码率控制算法[J]. 微电子学与计算机,2008,25(6):90-93.
DUAN D G,LI W ZH,MO Q,*et al*.. An improved H.264 rate control algorithm[J]. *Microelectronics Computer*,2008,25(6):90-93. (in Chinese)

[5] MA S W,WU F,LI Z G. Proposed draft of adaptive rate control[A]. Joint Video Team(JVT) of ISO/IEC MPEG & ITU-T VCEG(ISO/IEC JTC1/SC29/WG11 and ITU-T SG16 Q.6) 8th Meeting[C]. Geneva,Switzerland,May 20-26,2003, 1-25.

[6] LEE H J,CHIANG T H,ZHANG Y Q. Scalable rate control for MPEG-4 video[J]. *IEEE Trans. Circuits Systems Video Technol.*,2000,(10):878-894.

[7] RIBAS J,LEI S. Rate control in DCT video coding for low – delay communications[J]. *IEEE Trans. Circuits Syst. Video Technol.*,1999,9:172-185.

作者简介:陈　静(1986—),女,山东日照人,硕士研究生,主要从事实时图像处理与视频压缩等方面的研究。
E-mail:jingjingchen211@ yahoo. cn