

一种新的面向 H. 264 低比特流的码率控制方法

贺建新 曹岳辉

(中南大学信息科学与工程学院,长沙 410083)

摘要 JVT-G012 算法存在两个不足:一是当存在快速运动或场景变换时,线性 MAD(平均绝对偏差)预测法的效果变得很差;二是 H. 264 存在更多的运动补偿模式,更多的比特用于非纹理数据的编码,而 G012 算法简单地采用已编码帧的非纹理数据来预测当前帧的非纹理数据,这种方法已不再有效。本文给出了一种基于宏块的差图像直方图法,并根据图像复杂度在编码单元之间合理分配比较率。同时提出了一种时空结合的 MAD 预测模型和一种更加灵活的非纹理数据的预测方法。实验结果表明,该方法能更加接近目标码率,并获得更高的视觉质量和 PSNR。

关键词 H. 264/AVC 码率控制 差图像直方图 时空域 MAD 预测模型

中图法分类号: TP391 文献标识码: A 文章编号: 1006-8961(2009)06-1119-05

A Novel Rate Control Scheme for Low Bit Rate Stream of H. 264

HE Jian-xin, CAO Yue-hui

(School of Information Science and Engineering, Central South University, Changsha 410083)

Abstract There are two limitations in JVT-G012 scheme: firstly, the linear MAD prediction model performs poorly due to high motion or scene changes. Secondly, there exist much more complex motion compensation strategies in H. 264, and a higher percentage of bits is required to encode non-texture data. Thus the method that simply predicts non-texture bits through the previous coded frames is no longer efficient. First of all, a macroblock-based histogram of difference is introduced, and bits are allocated rationally among coding units according to image complexities. Furthermore, a novel rate control scheme named Spatio-Temporal MAD prediction model is presented, and a more flexible non-texture prediction scheme is derived. As shown in experiments, the proposed algorithm achieves a bit rate closer to the target, and provides an improved visual quality and the higher PSNR.

Keywords H. 264/AVC, rate control, histogram of difference, spatio-temporal MAD prediction model

1 引言

码率控制是视频编码器的一个必不可少的组成部分,并因此在以往的视频编码标准中被广泛而深入地研究,如 MPEG-2 中所采用的 TM5^[1], H. 263 中所采用的 TMN8^[2], MPEG-4 中所采用的 VM8^[3-4], 和 H. 264^[5] 所采用的 JVT-G012^[6] 算法。

相对于以往的视频编码标准,基于 H. 264^[5] 的码率控制算法存在以下两种难点。一是由于 H. 264

将量化参数 QP 同时用于率失真优化 RDO 和码率控制,导致了所谓的蛋鸡悖论,并使码率控制时所用的 MAD 只能根据已编码帧预测得来。二是 H. 264 采用了一些更加复杂的算法和特点(如多预测参考帧、1/4 像素精度运动补偿、多种模式的帧内/帧间预测编码等),运动补偿后的残差帧的能量变得很小,相对于以前更多的比特用于运动向量、编码模式等非纹理数据,并由此导致非纹理数据在各帧之间波动很大,更加难以精确预测,这种问题在低码率编码时更加突出。

收稿日期:2007-07-16;改回日期:2007-11-13

第一作者简介:贺建新(1982 ~),男,2008 年于中南大学获计算机应用技术专业硕士学位。主要研究方向为视频编码与传输技术。

E-mail:he_jianxin@126.com

H. 264 所采用的 JVT-G012 码率控制算法^[6]通过采用二次率失真模型^[7]、假想参考解码器 HRD (hypothetical reference decoder)、流体流量阻塞模型 (fluid flow traffic model)、线性 MAD 预测模型等技术,实现了率失真优化与码率控制的结合,并达到了较好的实用效果。然而该算法仍存在以下不足之处。一是当视频序列存在快速运动或场景变换时,各帧之间的 MAD 波动很大,G012 所采用的线性 MAD 预测模型所得到的 MAD 和实际相差很大。在这些不规则帧的数据参与预测模型的参数更新后,更会引起该模型的失效。二是 G012 算法简单地采用以前编码帧的非纹理数据比特数来预测当前帧的非纹理数据,而由于 H. 264 采用更加复杂的编码模式,使得实际编码所用的非纹理数据比特数和预测到的比特数相差太大,尤其在低码率等情况下,这种预测误差会给码率控制带来极坏的影响。

针对 JVT-G012 码率控制算法存在的不足,提出了一种新的码率控制方法:首先为了较好地比较各基本单元的编码复杂度,检测出视频序列中的快速运动和场景变换,给出了基于宏块的差图像直方图法。这种方法能根据各基本单元的编码复杂度,合理地进行帧级和宏块级的比特分配。再根据基于宏块的差图像直方图法,提出了一种时空结合的预测 MAD 的新方法和一种改进的非纹理数据预测方法。

2 时域相关性检测

残差帧的 MAD 是对编码复杂度的一个很好的度量标准,然而要想达到较好的效果必须得到精确的 MAD。JVT-G012 算法对于基本单元的 MAD 是通过简单的线性预测方法得到的,这种方法在一些非平稳视频序列的编码中性能不佳。当发生快速运动特别是场景变换时,当前帧和前一帧相关性极小,这将导致预测到的 MAD 和实际的 MAD 有很大的误差。以下介绍几种两帧之间相关性检测的典型算法。

直方图差法^[8] (DOH) 中帧 f_n 和帧 f_m 的直方图差定义为

$$doh(f_n, f_m) = \sum_{i=0}^{q-1} |h_n(i) - h_m(i)| \quad (1)$$

式中, q 为图像的灰度级数, h_n 和 h_m 分别为帧 f_n 和 f_m 的直方图。灰度直方图能很好地表现出图像内容的复杂度,因此 DOH 法是衡量图像帧之间相关性的一个很理想的方法,对于镜头的伸缩、平移和场景

变换等全局运动有很好的检测效果。但 DOH 对于图像的局部运动,不管是快速运动还是平缓运动都不敏感。

差图像直方图法^[8] (HOD)。令残差图像 $f_n - f_m$ 的直方图为 $dhod(i)$, $i \in [-q+1, q-1]$, HOD 定义为

$$dhod(f_n, f_m) = \frac{\sum_{i \notin [-\alpha, \alpha]} dhod(i)}{\sum_{i=-q+1}^{q-1} dhod(i)} \quad (2)$$

式中, α 是一个由实验得出的阈值,通常取 32, f_n 和 f_m 分别取当前未编码帧和前一帧的重构帧。差图像直方图 HOD 法相对于 DOH 法,对于图像的局部运动更加有效,能较好地反映出图像运动的剧烈程度。

HOD 法能很好体现帧图像的相对复杂度,可以根据 HOD 来分配帧的比特率。但考虑到宏块层的码率控制时也需要衡量宏块的图像复杂度,在这里采用基于宏块的差图像直方图法 (BHOD),以减少计算复杂度。差图像帧的 BHOD 定义为

$$bhod(f_n, f_m) = \sum_{i=0}^{N_{mb}-1} hod_i(f_n, f_m) \quad (3)$$

式中, f_n 和 f_m 一般为当前帧和前一帧的重构帧, N_{mb} 为一帧中的宏块个数, $hod_i(f_n, f_m)$ 则为第 i 个宏块的 HOD。

3 新的 H. 264 码率控制算法

3.1 帧级比特率分配

图像帧的位分配由缓冲区的填充度和剩余比特数决定,同时为了使各帧之间的图像 PSNR 趋于一致,进行位分配的时候也要考虑图像复杂度。

$$R_i = \beta \times R'_i + (1 - \beta) \times R'' \quad (4)$$

式中, R_i 为分配给第 i 帧的目标比特数, β 是一个常数,可以设为 0.5, R'_i 为由 HRD (假想参考解码器模型) 计算得到的帧比特数, R'' 为由剩余比特数和图像相对复杂度计算得来的帧比特数,计算为

$$R''_i = \alpha \times \frac{R_{\text{remain}}}{N_{\text{remain}}} \quad (5)$$

$$\alpha = \min \left\{ \max \left\{ \frac{bhod_i}{bhod_{i-1}}, k_1 \right\}, k_2 \right\} \quad (6)$$

式中, R_{remain} 为该视频序列中的剩余比特数, N_{remain} 为该视频序列中的剩余帧数。 k_1 和 k_2 为常数,由实验得出,在本文中分别取为 0.5 和 2。 $bhod_i$ 为第 i 帧的未编码图像和前一帧的重构图像之间的

$BHOD$, \overline{bhod}_{i-1} 为前 $i-1$ 帧的平均 BHOD, 可计算为

$$\overline{bhod}_{i-1} = \frac{1}{i-1} \sum_{j=1}^{i-1} bhod_j \quad (7)$$

3.2 宏块级码率控制

Ning^[9] 提出了一种新的宏块级图像复杂度预测方法, 令 σ_p 为前一帧相应位置宏块的标准差, σ_p^l 为前一帧相应位置左边宏块的标准差, σ_p^u 为前一帧相应位置上面宏块的标准差, 可以求得当前帧宏块的预测标准差 σ 为

$$\sigma = \frac{8 \times \sigma_p + \sigma_p^l + \sigma_p^u}{10} \quad (8)$$

Ning 的方法完全依靠前一帧的信息来预测当前帧宏块的复杂度, 当存在快速运动或场景变换时这种方法会出现较大的偏差。在实时低码率视频编码中, 运动信息占了帧间编码比特数的很大一部分。当前宏块的编码复杂度可以由局部运动程度近似表示, 而 BHOD 是一个有效的方法。本文采用 $bhod_{i,j}$ 表示第 i 帧第 j 个宏块的 BHOD, $\overline{bhod}_{spatial}(i,j)$ 为第 i 帧中当前待编码宏块的前 j 个宏块的平均 BHOD。

$$\overline{bhod}_{spatial}(i,j) = \frac{1}{j} \sum_{k=1}^j bhod_{i,k} \quad (9)$$

定义第 i 帧的 BHOD 相对率 $bhod_{ratio}(i)$, 第 i 帧第 j 个宏块的 BHOD 空间相对率 $bhod_{spatial_ratio}(i,j)$ 为

$$bhod_{ratio}(i) = \frac{bhod_i}{\overline{bhod}_{i-1}} \quad (10)$$

$$bhod_{spatial_ratio}(i,j) = \frac{bhod_{i,j}}{\overline{bhod}_{spatial}(i,j-1)} \quad (11)$$

注意, 这里的 $bhod_{ratio}(i)$ 反映的是当前未编码的第 i 帧相对于它之前的视频帧的时间相关性, 如果当前未编码帧相对于在它之前的图像序列的运动较为剧烈或存在场景变换, $bhod_{ratio}(i)$ 的值会突然变大。而 $bhod_{spatial_ratio}(i,j)$ 反映的是当前未编码的第 i 帧中的第 j 个宏块, 相对于同一帧中的前面 $j-1$ 个宏块的编码复杂度的相对率, 表示的是空间相关性。为了使视频序列存在快速运动或场景变换时, 能较好估计出当前宏块的 MAD, 借鉴文献[10]的 MAD 预测方法, 可以结合当前宏块的左边宏块和上面宏块的实际 MAD_{left} 、 MAD_{top} 来预测。

$$MAD = \frac{\omega_{pre} \times MAD_{pre} + \omega_{left} \times MAD_{left} + \omega_{top} \times MAD_{top}}{\omega_{pre} + \omega_{left} + \omega_{top}} \quad (12)$$

式中, MAD_{pre} 为根据 JVT-G012 算法的线性预测模型

得到的 MAD 值, 系数 w_{pre} 、 w_{left} 、 w_{top} 分别为 MAD_{pre} 、 MAD_{left} 、 MAD_{top} 的权重。由 $bhod_{ratio}(i)$ 的取值大小情况, 可以得知当前帧相较于过去已编码帧的相对运动程度和相对复杂度。如果 $bhod_{ratio}(i)$ 的值远远偏离均值 1 的话, 则当前帧的相对运动程度和相对复杂度都会发生较大变化, 此时可以相应减少时间相关性的权值 w_{pre} , 或者增加空间相关性的权值 w_{left} 和 w_{top} 。

在本文的计算中, 取 $w_{pre} = 3 - |bhod_{ratio}(i) - 1|$,

$$w_{left} = w_{top} = \frac{3 \times bhod_{curr}}{bhod_{curr} + bhod_{left} + bhod_{top}} \circ bhod_{curr},$$

$bhod_{left}$ 和 $bhod_{top}$ 分别为当前未编码宏块、当前未编码宏块的左边和上面宏块的 BHOD。当前宏块没有可利用的左边宏块时取 $w_{left} = 0$, 当前宏块没有可利用的上面宏块时取 $w_{top} = 0$ 。最后取

$$MAD = \min \{ \max \{ MAD, \alpha_1 MAD_{pre} \}, \alpha_2 MAD_{pre} \} \quad (13)$$

式中, 参数 α_1 、 α_2 为预测模型的系数值, 在本文中分别初始化为 0.9 和 1.1, 并随着以后的编码过程不断迭代修正。

MAD^2 相对于 MAD 更能反映出一帧图像的复杂度, 所以本文在计算宏块的目标比特率时采用了 MAD^2 来表示宏块的相对复杂度。宏块的位分配也和帧的位分配类似, 由缓冲区的填充度和剩余比特数决定, 同时也要考虑宏块的相对复杂度。令 R'' 为由剩余比特数和宏块的相对复杂度计算得来的宏块比特数, 得

$$R''_i = \frac{R_r}{N_{r,mb}} \times \frac{MAD_i^2}{\sum_{k=1}^{i-1} MAD_k^2} \quad (14)$$

式中, R_r 为该帧中剩余的未编码比特数, $N_{r,mb}$ 为该帧中的未编码宏块数, MAD_i 为第 i 个宏块的 MAD。

3.3 非纹理数据预测

在 H.264 编码标准中, 由编码模式和运动向量等所组成的非纹理数据会随着视频序列内容的改变而发生急剧变化, 其数据量会随着待编码图像的运动程度和内容复杂度成比例地增加。因此和预测 MAD 类似, 在这里引用 BHOD 法来预测宏块的非纹理数据。今 $R_{overhead}(i,j)$ 为第 i 帧第 j 宏块的非纹理数据比特数, $R_{texture}(i,j)$ 为第 i 帧第 j 宏块的纹理数据比特数, 则第 i 帧第 j 宏块的目标比特数 $R_{total}(i,j)$ 为

$$R_{total}(i,j) = R_{overhead}(i,j) + R_{texture}(i,j) \quad (15)$$

首先计算出前 $i-1$ 帧的实际非纹理数据的平均比特数 $\bar{R}_{frame}(i-1)$, 当前帧前 $j-1$ 个宏块的实际

非纹理数据的平均比特数 $\bar{R}_{mb}(j-1)$

$$\bar{R}_{frame}(i-1) = \frac{1}{(i-1) \times N_{mb}} \sum_{k=1}^{i-1} \sum_{l=1}^{N_{mb}} R_{overhead}(k,l) \quad (16)$$

$$\bar{R}_{mb}(j-1) = \frac{1}{j-1} \sum_{k=1}^{j-1} R_{overhead}(i,k) \quad (17)$$

再由 $bhod_{ratio}(i)$ 和 $bhod_{spatial_ratio}(i,j)$ 的值来预测

$$R_{overhead}(i,j) = \beta \times bhod_{ratio}(i) \times \bar{R}_{frame}(i-1) + (1-\beta) \times bhod_{spatial_ratio}(i,j) \times \bar{R}_{mb}(j-1) \quad (18)$$

式中, β 为表示权重的常系数, 由实验得出。可以随着比值 $\frac{j}{N_{mb}}$ 的增大, 相应小幅度地减少 β 的取值大小, 如 $\beta = 0.5 \times \left(1 - \frac{j-1}{N_{mb}-1}\right) + 0.5$ 。计算出当前宏

块的 MAD 和目标比特数 $R_{total}(i,j)$ 后, 再根据预测到的非纹理数据比特数 $R_{overhead}(i,j)$ 和二次率失真模型^[7] 计算出 $Q_{step}(i,j)$ 的值

$$R_{texture}(i,j) = R_{total}(i,j) - R_{overhead}(i,j) = \alpha_1 \frac{MAD}{Q_{step}(i,j)} + \alpha_2 \frac{MAD}{Q_{step}^2(i,j)} \quad (19)$$

如果计算得到的纹理数据 $R_{texture}(i,j)$ 比非纹理数据 $R_{overhead}(i,j)$ 还小, 则简单地由前一帧宏块的平均 QP 值加 2 作为当前宏块的 QP 值, 反之则根据 H. 264 的标准算法得出 QP。

$$QP = \begin{cases} \bar{QP}_{pre} + 2 & R_{texture}(i,j) < R_{overhead}(i,j) \\ \text{round}(6 \log_2 Q_{step}(i,j)) + 4 & \text{其他} \end{cases} \quad (20)$$

4 实验结果

本文在文献[10]提出的加权 MAD 预测方法的基础上, 给出了一种时空结合的 MAD 预测模型。通

过对比分析和实验可以验证, 提出的 MAD 预测算法, 不论是在预测精度还是在适应范围上都有了较大的改进。

在 JM86 上实现了该算法, 并和 JVT-G012 算法进行了比较。考虑到实时低码率下的应用, 两种算法都采用宏块作为基本单元, 编码帧数为 60 帧。通过对不同类型的 H. 264 标准测试序列的编码结果进行统计, 使实验具有一定的普遍性和代表性。测试序列的纹理和运动复杂度各不相同, akiyo, news, salesman, carphone, foreman 的图像纹理细节和运动程度总体上是依次从小到大的, akiyo 的图像纹理和运动都较为平缓, 而 carphone 和 foreman 都存在较为快速的运动。除了 foreman 为 CIF 格式外, 其余的都为 QCIF 格式, 其初始 QP 分别为 24, 24, 24, 30, 34, 目标码率分别为 32 kbps, 48 kbps, 48 kbps, 64 kbps, 192 kbps。具体测试条件列于表 1。

表 1 实验条件

Tab. 1 Experiment conditions

| 哈达玛变换 | 使用 |
|------------|-------|
| 率失真优化(RDO) | 使用 |
| MV 搜索窗大小 | 16 |
| 熵编码 | CAVLC |
| 参考帧数 | 5 |
| GOP 结构 | IPPP |

为了更加清楚地比较这两种码率控制算法的性能, 定义码率预测误差率 $E_{rate_predict}$ 为

$$E_{rate_predict} = \frac{|R_{acture} - R_{predict}|}{R_{predict}} \times 100\% \quad (21)$$

式中, R_{acture} 为编码所实际产生的码率, $R_{predict}$ 为目标码率。从表 2 的实验结果可以得知, 本文算法明显优于 JVT-G012 算法。本文算法各帧的编码比特率的波动明显减少, 不仅使实际输出码率更加靠近目标码率, 同时峰值信噪比 PSNR 值也提高了 0.19 db 到 0.38 db 之间。

表 2 JM 86 算法与本文算法性能比较

Tab. 2 Performance comparison for the proposed algorithm with JM 86

| 序列名 | JM 86 | | | 本文算法 | | | 码率误差减少(%) | PSNR 增加 |
|----------|--------|---------|-------|--------|---------|-------|-----------|---------|
| | 实际码率 | 码率误差(%) | PSNR | 实际码率 | 码率误差(%) | PSNR | | |
| akiyo | 32.47 | 1.47 | 39.16 | 32.20 | 0.63 | 39.35 | 0.84 | 0.19 |
| news | 48.58 | 1.21 | 35.65 | 48.12 | 0.25 | 35.93 | 0.96 | 0.28 |
| salesman | 49.02 | 2.13 | 33.79 | 48.44 | 0.92 | 34.08 | 1.21 | 0.29 |
| carphone | 65.55 | 2.42 | 35.16 | 64.48 | 0.75 | 35.54 | 1.67 | 0.38 |
| foreman | 193.43 | 0.74 | 33.78 | 192.42 | 0.22 | 34.13 | 0.52 | 0.35 |

图 1 和图 2 分别给出了测试序列为 carphone 和 foreman 时,两种算法下各测试序列每帧图像的 PSNR。从图中可以看出,由于本文算法能更好地根据每帧图像的复杂度合理地在各帧之间分配码率,同时各个宏块之间进一步调整比特率,从而使整个序列的 PSNR 波动较小,平均 PSNR 也得到了大幅度的提高。由理论分析和实验可以得出,本文算法在时空复杂度方面也有不错的表现,计算简单,不引入额外的编码延迟,完全适用于基于 H. 264 低比特流的实时视频编码。

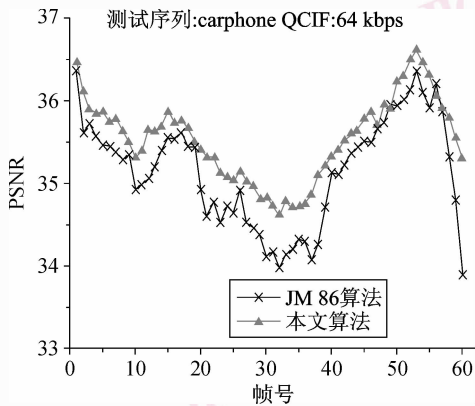


图 1 测试序列为“carphone”时两种算法的 PSNR

Fig. 1 Comparison of PSNR for sequence “carphone” between two algorithms

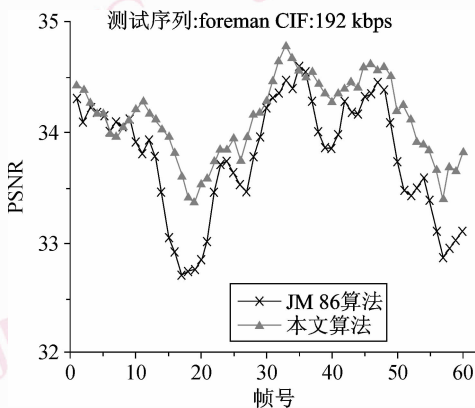


图 2 测试序列为“foreman”时两种算法的 PSNR

Fig. 2 Comparison of PSNR for sequence “foreman” between two algorithms

5 结论

首先分析了 H. 264 相对于以前编码标准给码率控制带来的不利影响,并指出了 JVT-G012 算法存在的不足。给出了一种基于宏块的差图像直方图法,根据编码复杂度在编码单元之间合理分配比较率。并通过对快速运动和场景变换的检测,给出了一种时空结合的 MAD 预测模型。针对低码率下非纹理数据的预测问题,给出了一种更加灵活的非纹理数据的预测方案。实验结果表明,本文算法优于 JVT-G012 算法,不但使实际编码所产生的码率更加精确地接近于目标码率,而且各帧的 PSNR 波动较小,主观视觉质量得到了较大程度的改善,平均 PSNR 也得到提高,具有较强的实用性和鲁棒性。

参考文献 (References)

- 1 ISO/IEC JTC1/SC29 WG11/93-400, MPEG-2 Test Model 5 [S].
- 2 Corbera J R, Lei S. Rate control in DCT video coding for low-delay communication [J]. IEEE Transactions on Circuits and Systems for Video Technology, 1999, 9(1): 172-185.
- 3 Lee H, Chiang T, Zhang Ya-qin. Scalable rate control for MPEG-4 video [J]. IEEE Transactions on Circuits and Systems for Video Technology, 2000, 10(9): 878-894.
- 4 Vetro A, Sun H, Wang Y. MPEG-4 rate control for multiple video objects [J]. IEEE Transactions on Circuits and Systems for Video Technology, 1999, 9(1): 186-199.
- 5 Thomas W, Gary J S, Gisle B, et al. Overview of the H. 264/AVC video coding standard [J]. IEEE Transactions on Circuits and Systems for Video Technology, 2003, 13(7): 560-576.
- 6 Li Zheng-guo, Pan Feng, Lim K P, et al. Adaptive basic unit layer rate control for JVT [A]. In: Proceeding of the 7th JVT Meeting [C], Pattaya, Thailand, 2003: 2-14.
- 7 Chiang T, Zhang Ya-qin. A new rate control scheme using quadratic rate distortion model [J]. IEEE Transactions on Circuits and Systems for Video Technology, 1997, 7(1): 246-250.
- 8 Lee J, Dickinson B W. Temporally adaptive motion interpolation exploiting temporal masking in visual perception [J]. IEEE Transaction on Image Processing, 1994, 3(5): 513-526.
- 9 Wang Ning, He Yun. A new bit rate control strategy for H. 264 [A]. In: Proceedings of Pacific-Rim Conference on Multimedia [C], Singapore, 2003: 1370-1374.
- 10 Chiang T H, Wang S J. A novel macroblock-layer rate-distortion model for rate control scheme of H. 264/AVC [A]. In: Proceedings of Computer Vision, Graphic, and Image Processing [C], Taiwan, 2006: 908-918.