

面向对象的 H. 264 码率控制算法

薛 卫¹ 都思丹² 叶锡君¹

(南京农业大学计算机系 南京 210002)¹ (南京大学电子科学与工程系 南京 210093)²

摘 要 提出了 H. 264 码率控制改进算法。首先,基于帧间运动估计残差(mean absolutely difference, MAD)的空间和时间相关性提出了“十字”MAD 预测模型;为了平衡缓冲的占有率,在分配帧比特时将前一帧的比特使用情况作为当前帧比特分配时的惩罚因子;最后,结合视频文字对象提取技术实现了基于文字对象的宏块比特分配以及量化步长的计算。实验结果表明,改进算法能在准确控制码率的情况下,提高输出序列的峰值信噪比。

关键词 H. 264, 码率控制, 文字对象

中图法分类号 TN919.81 **文献标识码** A

Object-based Rate Control Scheme for H. 264 Video Coding

XUE Wei¹ DU Si-dan² YE Xi-jun¹

(Department of Science and Technology, Nanjing Agricultural University, Nanjing 210002, China)¹

(Department of Electronics Science and Engineering, Nanjing University, Nanjing 210093, China)²

Abstract A new rate control scheme in H. 264 coding standard was proposed. The scheme includes + shape MAD prediction model based on space and time relativity of MAD. We used former frame bit usage as penalty factor to modify the bit-allocation rule to balance buffer share. Furthermore, combining video text extraction techniques, the method realized macro bit allocation and quantization step based on text area. Experimental results show that the new rate-control scheme can control the bits very well and the average PSNR is improved.

Keywords H. 264, Rate control, Text object

1 引言

码率控制贯穿整个 H. 264 视频编码过程,好的码率控制算法能够在保证编码缓冲区不产生溢出的同时,尽可能地提高视频编码的重建质量。与 MPEG 和其它 H 系列标准一样, H. 264 码率控制也是通过调节量化步长(Quantization parameters, QP)使编码码率尽可能与目标码率一致。H. 264 中,由于率失真优化技术的引入产生了著名的鸡蛋悖论问题,因此现有的码率控制策略,如 MPEG2 TM5, H. 263 TMN8, MPEG4 VM8 均无法适用于 H. 264 中。

现有的 H. 264 码率算法^[1-10]主要是针对帧比特分配、MAD,以及率失真模型的研究。比特分配基本上都是基于网络带宽、缓存占用量、缓存大小或剩余比特等因子的模型。基于 H. 264 率失真优化的复杂性,帧内预测或帧间运动估计残差可通过已编码帧获取, JVT(Joint Video Team)^[6]提出用线性回归,两步编码法^[7],前一帧的 QP 被用来产生残差,然后用估计残差求出 QP,但这两种方法计算比较复杂。而率失真模型相对比较成熟,常见的有基于宏块的率失真、二次率失真模型。以上算法虽从一定程度上解决了比特分配、残差预测问题,但在实际应用中仍存在改进的空间。

MPEG-4 编码的基本单元是视频对象(Video Object,

VO),对不同 VO 的运动信息、形状信息和纹理信息分配不同的码字。由于对象分割算法的复杂性,基于对象的视频编码更适合于电脑生成的 3D 动画,普通视频难以适用。H. 264 标准未定义对象编码,且从目前技术角度考虑,如果在 H. 264 中实现不同对象的编码也是很难的。鉴于视频文字的特殊性,相对于其它对象文字更易分割^[11,12],而且与图像其它部分相比具有较低的受关注度。实际使用中,可将基于文字对象的宏块比特分配策略作为 H. 264 码率控制的可选项,在帧内宏块比特分配中使用,赋予该区域较少的比特。

基于以上分析,本文提出 H. 264 码率控制的多个关键算法:十字 MAD 预测模型、帧码率分配、基于文字对象的宏块码率分配等。

2 H. 264 码率控制算法

当码率控制的基本单元为一帧时,码率控制只包含图像组(Group Of Pictures, GOP)层和帧层;否则,在完成帧层码率控制后,将进行基本单元层的码率控制。而后者更具有实际意义,能比较精确有效地控制码率,因此,改进算法的码率控制涉及到帧层及宏块层,基本单元层内的码率控制策略结合视频对象分割技术,根据预先设定的对象重要性,分配不同对象区域的比特流。基本单元层码率控制主要包括如下 3 个

到稿日期:2009-03-19 返修日期:2009-06-12 本文受国家自然科学基金(60472026),南京农业大学青年科技创新基金(KJ09027)资助。

薛 卫(1979-),男,博士,讲师,主要研究方向为音视频处理技术、信息安全、嵌入式系统, E-mail: xwskyzyq@sina.com; 都思丹 女,教授,博士生导师,主要研究方向为信号处理等。

部分。

2.1 MAD 预测——十字预测模型

JVT-H014^[6]中采用线性模型(2)预测,每处理完一个帧、基本单元或宏块要对模型的参数进行更新操作。 $f(i, j)$, $g(i, j)$ 分别为当前帧、前一帧块的灰度矩阵, i, j 为块的高和宽。

$$MAD = \frac{1}{I \times J} \sum_{|x| < \frac{I}{2}} \sum_{|y| < \frac{J}{2}} |f(i, j) - g(i-x, j-y)| \quad (1)$$

$$MAD_{cb} = x_1 \times MAD_{pb} + x_2 \quad (2)$$

JM 推荐的线性预测模型要求对预测参数 x_1 和 x_2 实时刷新,预测模型的精确性由一定数目的输入参考点来保证,一般是由坐标平面上的至少 10 个点拟合出一条直线,一次拟合完成后,还要去除其中与直线误差过大的点,进行第二次直线拟合,可以预见到这样的处理过程运算开销较大。

基于 MAD 的空间和时间相关性,本文提出一种新的 MAD 预测方法——十字预测模型。假设当前宏块是第 n 帧的第 m 个块, (i, j) 是其坐标, $MAD_n(i, j)$ 是其估计残差。

$MAD_{n-1}(i-1, j)$, $MAD_{n-1}(i+1, j)$, $MAD_{n-1}(i, j-1)$, $MAD_{n-1}(i, j+1)$, $MAD_{n-1}(i, j)$ 分别为 $n-1$ 帧的与 (i, j) 相对应的上下左右即同一位置块的 MAD 值。定义权重 a , 预测模型为:

$$MAD_n(i, j) = (1-4a)MAD_{n-1}(i, j) + a[MAD_{n-1}(i-1, j) + MAD_{n-1}(i+1, j) + MAD_{n-1}(i, j-1) + MAD_{n-1}(i, j+1)] \quad (3)$$

2.2 帧层率控制

HRD(Hypothetical Reference Decoder)^[6]是一个假想的解码器,它是 JVT 标准的重要组成部分之一,它描述了在视频传输中客户端缓冲区的状况,并要求在对编码码流进行解码的过程中缓冲区既不上溢又不下溢,这一要求一般要通过码率控制来实现。HRD 从直观上看它由编码图像缓冲区(Coded Picture Buffer, CPB)和瞬时解码器以及解码图像缓冲区(Decoded Picture Buffer, DPB) 组成。

定义 $u(i, j)$ 为第 i 个 GOP 第 j 帧时的码率, Fr 为帧率, $b(i, j)$ 为第 i 个 GOP 第 j 帧实际用去比特, $t_r(i, j)$ 表示第 i 个 GOP 中第 j 帧从缓冲中移出的时间, $be(t)$ 表示 t 时间内的数据量,下界和上界为 $V(i, j)$, $U(i, j)$ 初始值为:

$$V(i, 1) = T_r(i, 0) + \frac{u(i, 0)}{Fr} \quad (4)$$

$$U(i, 1) = (T_r(i, 0) + be(t_r(1, 1))) \times \bar{\omega}$$

$T_r(i, 0)$ 是第 $i-1$ 个 GOP 缓冲残留比特数, $T_r(1, 0) = 0$, 调节因子 $\bar{\omega} = 0.9$ 。 $V(i, j)$, $U(i, j)$ 更新公式如下:

$$V(i, j) = V(i, j-1) + \frac{u(i, j-1)}{Fr} - b(i, j-1) \quad (5)$$

$$U(i, j) = U(i, j-1) + (\frac{u(i, j-1)}{Fr} - b(i, j-1)) \times \bar{\omega}$$

帧比特分配步骤:

1) 为 P 帧分配比特数。

编完 GOP 中第 $j-1$ 帧后,其它 P 帧理论上可获得平均比特数为:

$$\bar{T}_r(i, j) = \frac{T_r(i, j)}{N_{gop} - (j-1)} \quad (6)$$

N_{gop} 表示一个 GOP 的帧数。表示考虑到相邻帧间的影响,引入带权重的前一帧预测比特数与实际比特数之差, γ 一

般取较小值,帧目标比特数可通过下式求得:

$$T_f(i, j) = (1-\gamma) \cdot \bar{T}_f(i, j) + \gamma \cdot [T_f(i, j-1) - b(i, j-1)] \quad (7)$$

$b(i, j-1)$ 为第 i 个 GOP 的第 $j-1$ 帧实际比特数。

2) 利用 HRD 限制,帧最终比特数为:

$$T_f(i, j) = \max\{V(i, j), T_f(i, j)\} \\ T_f(i, j) = \min\{U(i, j), T_f(i, j)\} \quad (8)$$

2.3 基于文字对象的宏块比特分配策略

分配之前必须通过字幕检测算法^[11]提取出文字区域,然后进行比特分配。假设分配给帧内文字区域宏块比特为 b_{ext} ,文字区域宏块个数为 N_{ext} ,其它区域宏块比特为 b_{roi} ,文字区域宏块个数为 N_{roi} 。

$$b_{ext}N_{ext} + b_{roi}N_{roi} = T_f(i, j) \quad (9)$$

其中, $b_{ext} = \theta \cdot b_{roi}$, θ 为分配比例。通过式(9)可以得到帧中文字区域与其它区域所获得的总比特数。

1) 分配宏块比特数

文字区域平均分配比特,而其它区域宏块以块 MAD 作为权重单独分配。

文字区域:

$$r_m = b_{ext}$$

其它区域:

$$r_m = \frac{MAD_m}{\sum_{k=1}^N MAD_k} \cdot b_{roi} \cdot N_{roi}$$

2) 计算宏块量化参数

文字区域:

$$\frac{r_m \cdot N_{ext}}{MAD} = \frac{a}{Q} + \frac{b}{Q^2}$$

MAD 为文字区域各宏块 MAD 之和,区域内使用固定 QP 可以减少运算时间。

其它区域:

$$\frac{r_m}{MAD_m} = \frac{a}{Q_m} + \frac{b}{Q_m^2}$$

3) 量化步长调整

为了限制相邻块波动太大,对 Q_m 进行调整。

如果 $Q_m < Q_{m-1}$:

$$Q_m = \max(Q_{m-1} - 2, Q_m)$$

否则:

$$Q_m = \min(Q_{m-1} + 2, Q_m)$$

3 实验

3.1 本文 MAD 预测方法与线性预测模型比较

在 H. 264 参考软件 JM8.6^[13]上实现本文提出的 MAD 预测方法,并与 JM8.6 采用的预测算法作比较。实验采用 4:2:0 格式标准序列,帧率 15 帧/s,编码器设置:使用 RDO, CAVLC 编码,运动搜索范围 32,参考帧 1。

在两种算法支持下,实验统计出 100 帧内各序列的帧平均位数和 PSNR。如图 1 所示,采用十字预测模型后,PSNR 和帧平均位数无明显波动,总体趋势与线性预测模型一样。表 1 给出了 7 个序列的实验结果,与线性预测模型相比,绝大部分序列的 PSNR 得到一定的提高,而且每帧比特数更加接近设定的码率。

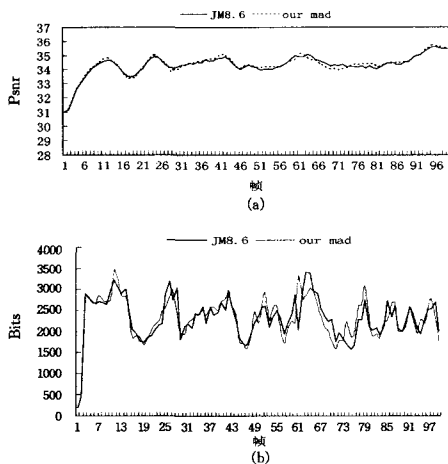


图1 (a)为CIF格式 silent 序列基于MAD线性预测模型、十字预测模型 H.264 编码器产生的 PSNR, (b)为实际比特率

表1 不同序列在线性预测模型与本文 MAD 方法下比特率、PSNR 数据列表

码主	序列	JM8.6 帧比特率	JM8.6 PSNR	本文 MAD 帧比特率	本文 MAD PSNR
48	foreman	46.91	34.41	46.94	34.49
24	carphone	22.75	32.79	22.78	32.76
24	claire	23.38	40.41	23.38	40.45
64	coastguard	30.94	29.20	30.98	29.26
36	silent	34.94	34.33	34.90	34.34
48	hall	47.00	39.35	46.90	39.36
24	hall	22.67	35.98	22.69	35.99

3.2 基于对象的宏块比特率分配策略

实验在 JM8.6 程序中实现本文帧比特分配、QP 模型,并与标准 JM8.6 模型作比较,如表 2 所列。由于标准序列中无字幕,因此通过修改 4:2:0 格式标准序列获得嵌有文字的视频,帧率 15 帧/s。编码器设置:使用 RDO, CAVLC 编码,运动搜索范围 32,参考帧 1。

为了测试效果,设置文字区域与其它区域宏块比特数比值为 0.5。结果表明,基于对象的宏块比特率分配策略的 PSNR 平均值大于 JM8.6 分配模型的,码流也较平稳。

表2 不同序列在 JM8.6 模型与本文码率控制方法下比特率、PSNR 数据列表

码率	序列	基于文字对象 比特分配帧 比特率	基于文字对象 PSNR	基于 JM8.6 模型 帧比特率	JM8.6 模 型 PSNR
24	carphone	24.13	31.64	23.54	31.59
48	carphone	48.97	34.75	47.62	34.55
24	foreman	24.08	30.81	23.50	30.70
48	foreman	49.06	33.97	47.58	33.80
24	akiyo	22.48	40.97	23.60	40.89
48	akiyo	46.07	43.70	47.65	44.31

结束语 本文主要研究了 H.264 中的关键技术: MAD 预测、帧比特分配及宏块比特分配等。利用宏块在空间和时

间上的相关性,提出了复杂度相对较低的十字加权残差预测模型,该模型在大部分测试序列中获得了较好的效果。缓冲中的占有率对视频传输影响较大,因此本文在分析 JVT 帧比特分配的同时,提出一个新的分配方法,将前一帧的比特使用情况作为当前帧分配时的奖惩,同时,利用基于文字对象的编码方法,适当减少字幕区域比特数,使得输出画面峰值信噪比得到提高。本文只是在 H.264 中尝试了基于文字对象的码率控制技术,而对其它对象的码率控制仍需进一步的研究。

参考文献

- [1] Li Z G, Pan F, Lim K P, et al. Adaptive Frame Layer Rate Control for H.264[C]//IEEE ICME. 2003;230-240
- [2] Xu Jianfeng, He Yun. A novel rate control for H.264[J]. IEEE Transactions On Image Processing, 2004, 23(5):809-812
- [3] Sullivan G J, Wiegand T, Stockhammer T. Using the draft H.26L video coding standard for mobile applications[C]//IEEE International Conference on Image Processing 2001. Thessaloniki, Greece, Oct. 2001
- [4] Kamaci N, Altunbasak Y, Mersereau R M. Frame Bit Allocation for the H.264/AVC Video Coder Via Cauchy-Density-Based Rate and Distortion Models[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2005, 15(8):994-1006
- [5] Altunbasak Y, Kamaci N. An analysis of the DCT coefficient distribution with the H.264 video coder[J]. Proc. IEEE Int. Conf. Acoustics, Speech, Signal Processing, 2004, 3(3):177-180
- [6] Li Z, Gao W, et al. Adaptive Rate Control with HRD Consideration[S]. ISO/IEC JTC1/SC29/WG11 and ITU-T SG16 Q.6 Document JVT-H014, May 2003
- [7] Peng Yin, Boyce J. A New Rate Control Scheme for H.264 Video Coding[C]//Proc. IEEE Int. Conf. Image Process. Oct. 2004:449-452
- [8] Marpe D, Schwarz H, Wiegand T. Context-Based Adaptive Binary Arithmetic Coding in the H.264/AVC Video Compression Standard[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2003, 13(7):620-636
- [9] Hong Shina, Leeb Y-L, Park H W. Rate control using linear rate-r model for H.264[J]. Signal Processing: Image Communication, 2004, 16:341-352
- [10] 韩峥, 唐昆, 崔慧娟. 基于 H.264 的码率控制算法[J]. 清华大学学报:自然科学版, 2008, 48(1):59-61
- [11] Jain A K, Yu B. Automatic text location in images and video frames[C]//Proc. ICPR. 1998:1497-1499
- [12] Lyu M R, Song Jiqiang, Cai Min. A Comprehensive Method for Multilingual Video Text Detection, Localization, and Extraction [J]. IEEE Transactions Circuits and Systems for Video Technology, 2005, 15(2):243-255
- [13] H.264 reference model software version JM8.6[EB/OL]. <http://bs.hhi.de/~suehring/tml/>, March 2003