

文章编号:1000-582X(2011)04-113-05

实时视频编码中的低复杂度码率控制算法设计

刘晓明,李毓蕙,仲元红

(重庆大学 通信工程学院,重庆 400044)

摘要:针对 H.264 原有码率控制提案中的不足,并考虑到 H.264 实时编码对编码效率的要求,设计了低复杂度的宏块层码率控制方案。对于图像复杂度的预测,充分挖掘图像的时空相关性,结合运动矢量对 MAD 进行预测,将 G012 的 MAD 预测算法中线性回归的算法复杂度降低为加权平均的算法复杂度。对于率失真模型,采用运算复杂度低的线性拟合代替 G012 算法中的二次拟合,并结合 SSIM 讨论了这种简化的可能性。针对运动量大的图像编码后 PSNR 的波动问题,对 GOP 层初始 QP 确定和 P 帧比特分配进行了改进,并充分利用已编码单元信息对当前量化参数进行调整,有效的控制码率。实验证明,改进后的算法对平稳输出序列的 PSNR 波动和降低运算复杂度有比较好的效果。

关键词:R-D 模型;MAD 模型;预测;计算复杂度;比特分配

中图分类号:TN919.8

文献标志码:A

Low complexity rate control scheme for real-time video coding

LIU Xiao-ming, LI Yu-hui, ZHONG Yuan-hong

(College of Communication Engineering, Chongqing University, Chongqing 400044 P. R. China)

Abstract: Considering the deficiency of H.264 rate control scheme and the coding efficiency requirements for real-time encoding, a low complexity rate control algorithm in MB-layer is proposed. MAD is predicted by combining motion vector and the temporal and spatial correlation to reduce the complexity of G012-MAD prediction algorithm from linear regression to the complexity of the weighted average. For the rate-distortion model, the linear model with low computational complexity instead of the quadratic fitting algorithm in G012 is used, while the possibility of is discussed combining with SSIM. For the fluctuation of the PSNR of the coded image with large movement, the determination of QP in GOP layer and the bit allocation of P frame are improved, and the current quantization parameter is adjusted by making full use of the information of coded MB to control the rate effectively. The experimental results show that the algorithm is more effective in resisting the fluctuation of PSNR for smoothly output sequence, and can decrease computational complexity.

Key words: R-D model; MAD model; prediction; computational complexity; bits allocation

H.264 是目前最高效和应用最多的编码标准^[1-3], H.264 编码过程中,引入了基于拉格朗日优

化的编码控制,将量化参数同时用于码率控制和率失真优化,而导致了蛋鸡悖论,即:对宏块执行率失

收稿日期:2010-09-26

基金项目:国家自然科学基金重点资助项目(51035008)

作者简介:刘晓明(1963-),男,重庆大学教授,主要从事电子自动化、通信与测控中的三化、信号与图像处理、汽车电子等方向研究,(Tel)13983099959;(E-mail) lxm_cqu@126.com。

真优化(rate distortion optimization, RDO),要先通过宏块的平均绝对误差(mean absolute differences, MAD)值决定宏块的量化参数(quantization parameter, QP)值,而宏块的 MAD 值只有在 RDO 执行后才能获得。因此,经典的 TM5^[3], TMN8^[4]等码率控制算法在 H. 264 中并不适用。

目前,适用于 H. 264 的码率控制算法^[5-6]主要有 JVT 的几个会议提案。JVT-F086^[7]继续了经典算法 TM5,采用二次模式选择的办法避免蛋鸡悖论,不过这也增加了编码复杂度,难以应用于实时传输,同时,它的率失真模型过于简单,难以得到精确的码率控制。JVT-G012^[8]是应用比较广泛的 H. 264 码率控制算法,在这些提案中,解决蛋鸡悖论的方法是利用图像的时空相关性,即用线性 MAD 模型来估计帧的复杂度,用二次 R-Q 模型来计算量化参数。JVT-O016^[9]对 G012 提案的改进主要在于充分挖掘视频序列的时空相关性,对编码质量的提高作出了贡献。JVT-W042^[10]是最新的码率控制提案,将码率控制分为几种模式,以适应不同结构的图像组(group of pictures, GOP)系列。

在 G012 中,对 MAD 预测模型和二次 R-Q 模型的参数更新都是采用线性回归的办法,这在一定程度上影响了编码的实时性。在帧层码率控制中,没有考虑到 P 帧的实际编码复杂度,在运动量大或场景切换的情况下,某一 P 帧编码完后缓冲区充盈度较高时,极易使得后续帧的峰值信噪比(peak signal noise ratio, PSNR)值连续大幅下降,甚至引起跳帧。针对这些不足,提出了一系列的改进措施,以下将分别对 GOP 层,帧层,宏块层的改进算法进行介绍。

1 GOP 层初始 QP 值确定

在 G012 的 GOP 层码率控制中, I 帧及后续第一个 P 帧都采用同一量化参数进行量化,如式(1)计算

$$QP_s = \frac{\text{Sum}_{\text{PQP}}}{N_p} - 1 - \frac{8T_r(n_{i-1}, N_{\text{GOP}})}{T_r(n_i, 0)} - \frac{N_{\text{GOP}}}{15}, \quad (1)$$

式中, Sum_{PQP} 是前一 GOP 的 P 帧量化参数和, N_p 是前一 GOP 中的 P 帧总数。在运动量比较小的视频序列中, I 帧及后续第一个 P 帧采用同一量化参数进行量化并无不妥,但对于运动量比较大的视频序列,比如本设计的实际应用环境,如果第一个 P 帧也采用 I 帧的量化参数,极有可能造成缓冲区充盈度过高甚至溢出,使得后续 P 帧的质量持续下降, PSNR 波动较大。所以对第一个 P 帧采用主动增大量化参数的办法,并考虑到视频的一致性,第一

个 P 帧的量化参数设为 $QP_s + 2$ 。

2 刷新目标缓冲级别的 P 帧比特分配

在 G012 帧层码率控制中,对各帧进行目标比特分配,其算法思想是将当前 GOP 的第一个 P 帧编码完后缓冲区的实际占有度平均分配给当前 GOP 中所有未编码的 P 帧,没有考虑到 P 帧的实际编码复杂度,在运动量大或场景切换的情况下,当某一 P 帧编码完缓冲区占有度较高时,这样的目标缓冲级别设置会使得后续连续数帧图像的 PSNR 值连续大幅度下降,甚至可能会引起跳帧。改进的基本思想是在每帧编码完后,目标缓冲级别随着缓冲区实际充盈度的变化更新,将当前帧编码完后缓冲区充盈度对目标比特分配的影响平均分摊给当前 GOP 中所有未编码的 P 帧。

在 G012 中,每编码完 1 个 P 帧,目标缓冲级别下降 1 个定值

$$\frac{Tbl(n_{i,2}) - B_s/8}{N_p - 1}. \quad (2)$$

将这一项修改为

$$\frac{B_c(n_{i,j}) - B_s/8}{N_p(j)}, \quad (3)$$

$B_c(n_{i,j})$ 是第 i 个 GOP 的第 j 帧编码完后缓冲区的实际充盈度, $N_p(j)$ 是第 j 帧编码完后剩余未编码的 P 帧数目。实验表明,这样的修改对各类视频序列的 PSNR 值波动都有一定的平滑作用,对运动剧烈的序列更为明显。

3 低复杂度的宏块层码率控制策略

3.1 结合运动矢量的时空域 MAD 预测法

MAD 的线性预测模型在每次编码完 1 个基本单元后都需要对模型参数进行刷新^[11]。刷新的方法是线性回归,模型参数的精确性由一定数目的输入参考点来保证,一般由坐标平面上至少 20 个数据点拟合出 1 条直线,1 次拟合完成后,去除其中与直线误差过大的奇异点,进行第二次直线拟合。可以看出,这种处理过程的运算代价非常大,在实时编码中,这种方式将给其他模块的效率要求带来很大的压力。

所以提出了一种结合运动矢量的时空域 MAD 预测法,使用相对简单的加权模型来进行 MAD 预测,改进了运算复杂度。并且考虑到 G012 对 MAD 模型进行参数更新时,其数据是在最近的编码块中得到,这样只利用到了空间的相关性,而在实际的编码中,没有空间相关性或者空间相关性较弱但时间相关性较强的情况经常发生, MAD 预测法中的加权

模型是基于时空域的。

参考 H.264 标准运动矢量的预测方法,假设 E 为当前宏块, A, B, C 分别为 E 的左, 上, 右上方的 3 个对应块, 则当前宏块运动矢量的预测值 MV 取 A, B, C 3 个宏块的 MV 的中值。

借用这种思想, 同时考虑 2 帧之间的时间相关性, 利用当前宏块 cur 和前 1 帧相应位置的宏块 pre 的邻域, 如图 1 所示, 提出如下 MAD 时空加权预测模型

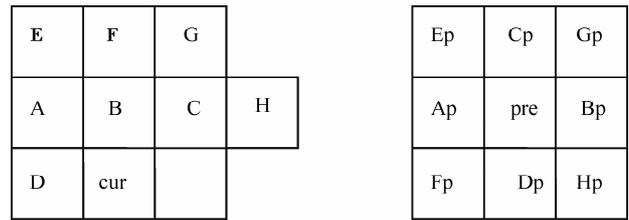


图 1 当前宏块及前 1 帧对应位置宏块的邻域

$$MAD_{cur} = \begin{cases} \frac{a(MAD_B + MAD_D) + b(MAD_A + MAD_C) + MAD_{pre}}{3}, & MV_{pre} \leq \text{Thres}; \\ \frac{aMAD_{A+B+C+D} + bMAD_{E+F+G+H} + aMAD_{pre} + bMAD_{Ap+Bp+\dots+Gp+Hp}}{2}, & MV_{pre} > \text{Thres}. \end{cases} \quad (4)$$

式中, $MAD_{A+B+C+D}$ 表示 A, B, C, D 4 个宏块的 MAD 平均值, a, b 是加权系数, $a+b=1$, a, b 的比例可根据实际图像进行调整。

Thres 是表征运动剧烈程度的运动矢量阈值, 作为 MAD 计算范围的评判标准, 用于对 MAD 进行分级预测。如果 $MV_{x_{pre}}, MV_{y_{pre}}$ 都小于 16, 即水平垂直的运动矢量均不超过 1 个宏块, 则在当前帧中取如图 1 所示 A, B, C, D 4 个宏块 MAD 对当前宏块 MAD 进行预测, 按公式(4)中 $MV_{pre} \leq \text{Thres}$ 计算, 如果 $MV_{x_{pre}}, MV_{y_{pre}}$ 都大于 16, 即水平垂直的运动量都超过 1 个宏块时, 经过对大量序列的实验, 同时考虑 MAD 预测的准确性, 计算的复杂度以及硬件实现的方便, 按公式(4)中 $MV_{pre} > \text{Thres}$ 进行计算。

在这个预测模型中, 模型参数为常量, 运算复杂度很低, 并且充分考虑了 MAD 的空间以及时间相关性, 并结合运动矢量大小对加权预测进行分级处理, 使 MAD 的计算范围更加合理。

3.2 R-D 曲线的线性拟合

一般而言, 视频信源在经过编码预测及 DCT 变换之后, 得到的 DCT 系数的相关性很低, 可以近似认为不相关, 而且符合拉普拉斯概率分布, 可以得到 DCT 量化系数的信息熵 $H(\alpha)$, 其中 $\alpha = MAD/Q$ 。 $H(\alpha)$ 具有非常复杂的数学表达式, 是 1 个关于 α 单调递增的函数。

如果采用二次曲线拟合, 就得到了 G012 提案中的二次模型, 二次 R-D 模型的模型参数 X_1 和 X_2 利用线性回归技术来获得。显然, 二次模型的曲线拟合比一次模型的曲线拟合会带来更大的计算开销。在对失真度的要求并不苛刻的情况下, 并且码率控制并非编码传输的核心算法, 因此可以最大限度地降低运算复杂度。

基于以上分析, 对 R-D 关系采用线性曲线拟合, 并结合 JVT-O016 中的头信息预测, 得到以下预测模型

$$R = X_{21} \frac{MAD}{Q_{step}} + X_{22} + H. \quad (5)$$

虽然线性曲线拟合的精确度总体上不如二次曲线, 但如果 QP 的取值保持在不太大的范围内, 线性曲线可以拟合到与理论模型具有较高的近似度。

峰值信噪比数学表达简单直观, 反映了整幅图像的失真程度, 一直以来都被作为评价图像质量的主要标准。但图像最终由人接受, 质量评价应该考虑 HVS, 即人类视觉系统的各种特性。结构相似度是 Wang 等学者最先提出的一种图像质量评价方法^[12-13]。这种基于整体图像结构信息变化而非像素误差的结构相似度 (structural similarity, SSIM) 质量评价法, 比均方误差 MSE 更加符合人眼主观感觉, 而且计算也比较简单。

通过对一系列标准测试序列 news, tempete, mobile, stefan, highway, waterfall 等采用不同 QP 值进行编码, 并对解码后图像与原图像进行 SSIM 评价, 对不同序列得到的结果取均值, 得到 SSIM-QP 趋势线如图 2 所示。

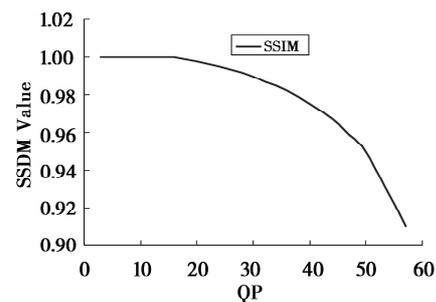


图 2 SSIM-QP 趋势线

由于对图像失真度的要求并不苛刻,通过对实时采集的视频图像进行编码实验,当 QP 值小于 20 时,输出码率将会超过传输系统的信源码率指标,而量化参数大于 45 的情况也极少发生。在这一量化参数范围内,线性拟合的准确度是比较高的,对 R-D 曲线采用线性拟合以降低运算复杂度完全可行。

3.3 量化参数调整

在 G012 中,量化参数是由 R-D 模型直接确定的,没有充分利用已编码单元的信息,考虑到这一点,引入 1 个 QP 调整因子 η ,根据已编码宏块的实际编码比特数和目标比特数的比值情况来确定当前宏块的 QP 调整因子。设前 1 编码宏块的实际编码比特数为 bits_actual ,目标比特数为 bits_predict ,则 $\eta = \text{bits_actual}/\text{bits_predict}$ 。当前宏块量化参数值按下式调整

$$\text{QP} = \begin{cases} \text{QP}, 0 < \eta < 0.8; \\ \text{QP} + 1, 0.8 \leq \eta \leq 1.5; \\ \text{QP} + 2, \eta > 1.5. \end{cases} \quad (6)$$

这样,提高了码率控制对图像空间相关性的利用,使得输出码率更为接近目标码率。

4 实验结果与分析

将算法在适用于 DSP 芯片 DM642 的实时编码器 T264^[14]中实现,并采用一系列标准测试序列进行仿真,与原 T264 中使用的算法进行比较。图像格式均为 CIF (352 × 288),GOP 结构均为 IPPPP ……IPPPP…,I 帧间隔为 30 帧。这里将对运动量较小的 news 序列和运动量较大的 stefan 序列的实验结果进行分析,对于 news 序列,目标码率分别设为 150 kbit/s,250 kbit/s,对于 stefan 序列,目标

码率分别设为 150,250,350 kbit/s。测试结果统计表如表 1 所示。

表 1 算法与 T264 码率控制的效果比较

测试序列(码率/kbits ⁻¹)	平均 psnr/dB		码率偏差 /kbit		编码速度/(帧·s ⁻¹)	
	T264	本算法	T264	本算法	T264	本算法
news(150)	35.99	36.24	-2	-0.5	36.45	36.03
stefan(150)	23.69	24.13	-1	+0.3	33.30	32.86
news(250)	38.68	39.02	-5	-2	37.64	37.15
stefan(250)	27.60	28.70	-8	+4	32.73	32.09
stefan(350)	29.49	29.81	-6	-2	31.30	30.74

由表 1 可见,由于按实际缓冲区充盈度来刷新目标缓冲级别,本算法有时会超出预定的码率,但总的来说,码率偏差较 T264 算法更小。采用了宏块级的码率控制以后,本算法较 T264 的帧层码率控制在平均 psnr 值上得到了平均 0.51dB 的提升,并且该算法是低复杂度的,在编码速度上对原编码器影响很小,不超过 2%。

图 3 和图 4 是 news 序列在 150 kbit/s,250 kbit/s 目标码率下, stefan 序列在 250 kbit/s,350 kbit/s 目标码率下的 PSNR 图。由图可见,算法对序列的 PSNR 值波动都有一定的平滑作用,特别是对于运动量较大的 stefan 序列,在插入 I 帧后, T264 算法的 PSNR 值大幅下降,而主动增大第一个 P 帧量化参数的策略以及按实际缓冲区充盈度刷新目标缓冲级别的策略补偿了这一点,使得插入 I 帧后的 PSNR 下降变得比较平缓。由图中也可以看到,平均 PSNR 值也有一定的提高。

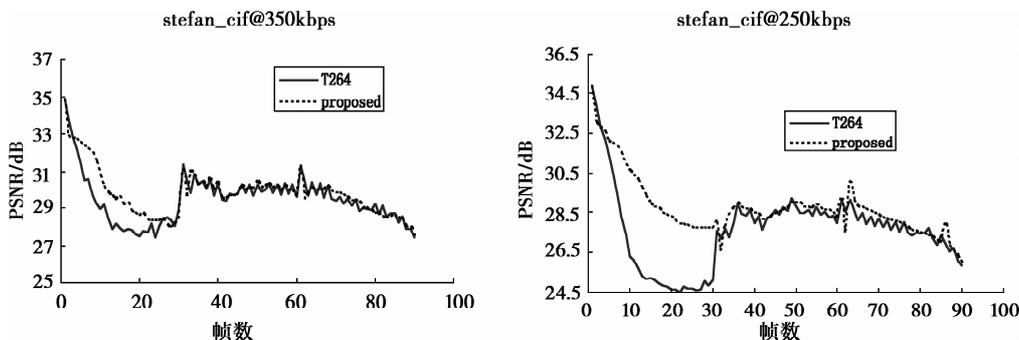


图 3 news 序列的 PSNR 对比图

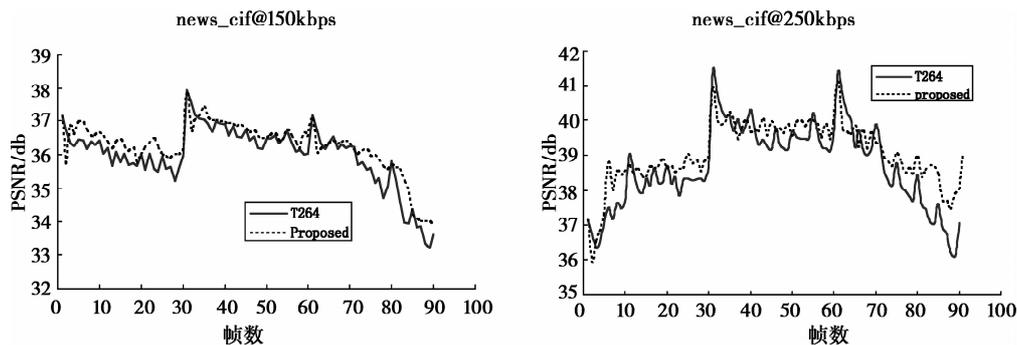


图4 stefan序列的PSNR对比图

5 结 语

在码率控制的宏块级提出了一种结合运动矢量的时空域MAD预测法,并对其中的率失真模型进行了简化,在提高了码率控制精度的同时,减少了算法的运算复杂度。对帧层的目标缓冲级别进行了改进,平滑了大运动量序列编码后的峰值信噪比波动。实验证明,低复杂度码率控制算法有效可行。

参考文献:

- [1] ITU-T Recommendation H. 264 & ISO/IEC 14496-10 (MPEG-4) AVC [S]. Geneva, Switzerland: Advanced Video Coding for Generic Audiovisual Services, 2003.
- [2] WIEGAND T, SULLIVAN G J, BJONTEGAARD G, 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.
- [3] Test Model 5 [CP]. ISO/IEC JTC1/SC29/WG2/No 400, MPEG93/457, Apr 2003.
- [4] Video Codec Test Model [CP]. ITU-T/SG15. TMN8. Portland, OR, June, 1997.
- [5] CHEN Z, NGAN K N. Recent advances in rate control for video coding [J]. Signal Processing: Image Communication, 2007, 22(1): 19-38.
- [6] WU W, KIM H K. Rate control initialization scheme for H. 264/AVC [C] // Proceedings of ICEIC 2008, June 24-27, 2008, Tashkent, Uzbekistan. [S. l.]: IEEE, 2008: 534-537.
- [7] MA S, GAO W. Proposed draft description of rate control on JVT standard, JVT-F086 [S]. Awaji: Joint Video Team (JVT) of ISO/IEC MPEG & ITU-T VCEG 6th Meeting, 2002: 1-5.
- [8] LI Z G, PAN F, LIM K P, et al. Adaptive basic unit layer rate control for JVT, JVT-G012 [S]. Pattaya, Thailand: Joint Video Team (JVT) of ISO/IEC MPEG & ITU-T VCEG 7th meeting, 2003: 1-6.
- [9] YUAN W, LIN S X, ZHANG Y D. Optimum bit allocation and rate control for H. 264/AVC, JVT-O016 [S]. Busan, KR: Joint Video Team (JVT) of ISO/IEC MPEG & ITU-T VCEG 15th meeting, 2005.
- [10] LEONTARIS A, TOURAPIS A M. Rate control reorganization in the joint model (JM) reference software, JVT-W042 [S]. San Jose, California, USA: Joint Video Team (JVT) of ISO/IEC MPEG & ITU-T VCEG 23rd meeting, 2007.
- [11] LIU Y, LI Z G, SOH Y C, et al. Adaptive MAD prediction and refined R-Q model for H. 264/AVC rate control [C/OL] // 2006 IEEE International Conference on Acoustics, Speech and Signal Processing, Toulouse, May 14-19, 2006 [2006-07-24]. http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=1660490&tag=1.
- [12] WANG Z, BOVIK A C, SHEIKH H R, et al. Image quality assessment: from error visibility to structural similarity [J]. IEEE Transactions on Image Processing, 2004, 13(4): 600-612.
- [13] WANG Z, LU L, BOVIK A C. Video quality assessment using structural distortion measurement [C] // Proceedings of the IEEE International Conference on Image Processing, 2002, Rochester, Rochester: Signal Processing Society, 2002: 65-68.
- [14] T264 reference software [CP/OL]. http://sourceforge.net/project/showfiles.php?group_id=119399.

(编辑 侯 湘)