

# 基于可伸缩视频编码 Hierarchical-B 结构的 恒定质量控制

谈永敏 杨小康 宋利

(上海交通大学电子工程系图像通信与信息处理研究所, 上海 200240)

(上海交通大学上海市数字媒体处理与传输重点实验室, 上海 200240)

**摘要** 针对 H. 264 视频编码标准可伸缩扩展部分采用 Hierarchical-B 结构实现了时域上的可分级特性. 同时与传统的 IPPP 或者 IBP 编码结构相比, 在编码增益上也有较大的提高. 但其特殊的预测结构会造成图像质量的抖动, 在有快速运动和场景切换时此现象尤甚. 提出了一种简单的基于 PSNR 和 MAD 的恒定质量控制算法, 通过在帧级选择合适的量化参数来限制图像质量的波动范围. 该算法在 JVT 规定的标准基础上进行了仿真, 实验结果表明, PSNR 的方差大幅减小, 并基本接近目标 PSNR.

**关键词** 恒定质量 Hierarchical-B 结构 可伸缩视频编码

中图分类号: TN919.81 文献标识码: A 文章编号: 1006-8961(2007)10-1827-05

## Constant Quality Control for Hierarchical-B Structure in Scalable Video Coding

TAN Yong-min, YANG Xiao-kang, SONG Li

(Institute of Image Communication and Information Processing, Department of Electrical Engineering,

Shanghai Jiaotong University, Shanghai 200240)

(Shanghai Key Laboratory of Digital Media Processing and Transmissions, Shanghai Jiaotong University, Shanghai 200240)

**Abstract** The scalable extension of H. 264 video coding standard (SVC) has achieved temporal scalability using Hierarchical-B structure. Compared to traditional coding structures like IPPP and IBBP, it has a significant coding gain as well. However, due to its hierarchical prediction dependency, quality fluctuation among consecutive frames may sometimes occur. Such situation usually gets worse when high motion or scene changes exist. In this paper, we propose a simple but effective constant quality algorithm based on PSNR and MAD measures. Restriction to the possible fluctuation is realized by adapting the quantization parameter (QP) at frame level. We simulate our method based on JVT official specification. Experimental results show that our algorithm has greatly reduced the PSNR variance and obtained the target PSNR simultaneously.

**Keywords** constant quality, Hierarchical-B structure, scalable video coding

## 1 引言

多媒体信号的传输与存储通常需要对编码压缩参数进行准确地选择. 就视频信号而言, 根据所要

求的压缩效率和编码质量, 需要运用不同的编码技术, 并合理地设置相应的编码器参数. 压缩比率和质量均和多个参数有关, 如 GOP (group of pictures) 结构, 图像时间和空间分辨率, 宏块的编码模式, 量化参数等. 总体而言, 对视频流的压缩比率和编码

基金项目: 国家自然科学基金项目(60625103); 国家高技术研究发展计划 863 项目(2006AA01Z124)

收稿日期: 2007-07-05; 改回日期: 2007-07-23

第一作者简介: 谈永敏(1983~), 男. 现为上海交通大学通信与信息系统专业硕士研究生. 主要研究方向为视频编解码. E-mail: angelwings@sjtu.edu.cn

质量呈反比例关系,且在大多数情况下该关系是非线性的。实际上,直接影响这两个量的就是每一帧的量化参数的选择。在编码过程中,无论以比特率还是 PSNR (peak signal to noise ratio) 为目标,都可以在局部范围内进行优化,进而使整体效果接近目标值。依具体应用场合的不同,信源的编码可以只考虑比特率或者质量,即可分为恒定比特率(CBR)编码与恒定质量编码。

在恒定比特率编码中,编码器主要的约束是达到尽可能精确的目标比特率。这一方法对于 CBR 信道下的视频传输至关重要,可以避免由于比特率的抖动而可能造成的信息传输丢失和延迟,也是大多数 CBR 假设下的码率控制算法的前提。

恒定质量编码的主要目的是以相对稳定的 PSNR 级别来对视频内容编码,而不考虑编码所需的比特数。虽然这一偏向由于可能导致网络带宽利用率的上下波动而不太适合真实的视频传输场景,但同样也有其可用之处。在远程遥控机械,视频监控中,数据存贮业务如数码电影的转录中,质量往往是最关键的因素,而传输带宽通常是不受限的。

绝大多数已有文献中的恒定质量控制算法都是基于传统的 IPPP 和 IBBP 结构。例如采用简单的基于 PSNR 滑动窗口的算法来直接选择合适的帧级量化参数<sup>[1]</sup>。或者利用每帧图像的平均绝对误差(MAD)作为帧复杂度的度量准则来建立目标比特,失真和量化参数之间的模型<sup>[2,3]</sup>。

目前,ISO 和 ITU-T 组织联合开发的 H. 264 标准的可伸缩扩展部分(SVC)已处于标准化过程的后期<sup>[4]</sup>。和之前的视频标准相比,SVC 具有一些新的特性。而针对 SVC 的恒定质量控制,目前几乎还是空白。因此,本文提出了一种简单的基于 PSNR 和 MAD 的恒定质量控制算法,通过在帧级选择合适的量化参数(QP)来限制图象质量的波动范围。

## 2 SVC 的 Hierarchical-B 结构

SVC 中的时域可分级特性可以通过 Hierarchical-B 的概念来高效实现,如图 1 所示。

图 1 下部的序号代表编码顺序,符号  $T_x$  代表相应的第  $x$  时域分层。基本层  $T_0$  采用全 I 帧或者 IPPP 的结构,而除  $T_0$  层以外的其他增强层通常编为 B 帧。两个连续  $T_0$  层之间的所有增强层图像帧加上后一个  $T_0$  图像帧组成了一组图像(GOP)。实现时域可分级

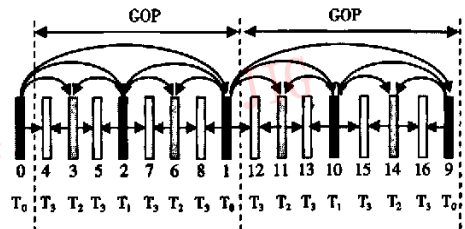


图 1 实现时域可分级的 Hierarchical-B 预测结构

Fig. 1 Hierarchical-B prediction structure for enabling temporal scalability

的主要约束是对属于某个时域增强层  $k$  的图像帧而言,其参考帧队列中只能包含时域层数小于  $k$  的参考帧。依此规则组织而成的参考依赖关系使得解码器能够通过简单地丢弃所有时域级别大于目标时域层  $T_{target}$  的图像帧来达到所需要的时间分辨率,而不会因为部分帧的丢失而影响解码过程。

虽然 Hierarchical-B 可以灵活地实现时间可分级特性,但这样类似金字塔式的结构很容易造成编码质量较大幅度的波动。其原因主要有以下几点:

(1) 沿时间顺序的各帧不属于同一时域分级,且通常高低时域级别交错出现。

(2) 较低时域级别的图像帧会直接和间接地被所有较高时域级别的图像帧参考,如果该帧编码质量不理想,会影响当前 GOP 内后续的所有待编码帧。

由(1),(2)可知, $T_0$  层编码对整个图像序列质量的一致性起到了至关重要的作用。但另一方面,如果假设实际低时延的应用场合, $T_0$  层采用 IPPP 结构,GOP 大小为 16,则相邻  $T_0$  帧的距离相隔 16 帧。如果不施加额外的恒定质量控制算法而简单地采用固定 QP 的方法来编码,如此远的参考距离会造成预测的不准确。如果视频序列是高速运动的,或者含有场景切换时,经过 16 帧之后的视频内容可能已经发生了很大的变化,可以预见,此时在这样的 GOP 中编码质量将会发生巨大的抖动。图 2 中,对 Stefan 序列进行了测试,其中,QP 值恒定。从图 2 中可以清晰地观察到以上描述的现象,尤其在序列的后半段。

## 3 恒定质量控制算法的实现

### 3.1 质量控制的一般方法

要取得恒定的视频质量,最简单的方法就是采

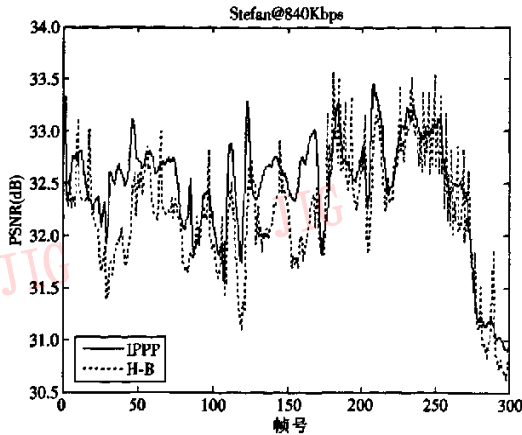


图 2 IPPP 结构与 Hierarchical-B 结构的比较  
Fig.2 PSNR comparison of IPPP structure and Hierarchical-B structure(Stefan sequence)

用固定 QP 编码。但正如上节所描述,视频序列的内容缺乏一致性,同一序列中不同场景段的特性也不尽相同。因此,在编码过程中始终使用同一个 QP 并不一定会随之带来较稳定的 PSNR 输出。然而,对于序列中的某一段连续场景,通常各帧之间具有较为相似的运动程度和纹理复杂度,此时采用恒定的 QP 值可以近似地达到恒定的质量输出。

因此,PSNR 值可以作为一种有效的指示。若观察到当前帧的 PSNR 与过去某一窗口长度的平均 PSNR 相比有明显的变化,视频内容很有可能已经发生了变化,如一个新场景的开始。此时,就需要通过一些步骤来调整 QP,以保证新视频内容的质量仍然维持在目标 PSNR 附近。

完成这一调整的基本方法是用不同的 QP 值重复对 PSNR 远离目标值的同一图像帧进行编码,直到获得最接近的效果。显然,这种方法能够使得 PSNR 的波动范围最小,但是其计算复杂度很高,尤其在场景切换的边界可能需要很多次循环编码。

另一个可行的方法就是基于对过去 PSNR 历史值的观察,如果当前 PSNR 值与平均值的差超过了某一个预先设定的阈值,就需要相应地调整 QP。该方法的性能虽然并非最优,但质量控制算法额外带来的计算复杂度却可以几乎忽略不计。

### 3.2 初始 QP 的选择

初始 QP 的选择应尽可能使 PSNR 值接近目标 PSNR,这样对之后所有  $T_0$  层图像帧以及每个 GOP 内各时间增强层的 QP 调整更为有利。

对于零均值的 *i. i. d.* 信源,失真 ( $D$ ) 和量化步长  $Q_s$  的关系可以近似为<sup>[5]</sup>

$$D(Q_s) = \frac{Q_s^2}{\epsilon} \quad (1)$$

其中,  $Q_s$  是量化步长,  $\epsilon$  是和信源有关的参数。

在 SVC<sup>[4]</sup> 中,  $Q_s$  和 QP 的关系为

$$Q_s = 2^{QP/6} \times v(QP \bmod 6) \quad (2)$$

其中,  $v(x)$  为标准中指定的离散函数。

根据式(1),式(2),若使用均方误差 (MSE) 来表示失真,PSNR 值可由下式推导出:

$$\begin{aligned} PSNR &= 10 \lg \frac{255^2}{MSE} = 10 \lg \frac{255^2}{\frac{Q_s^2}{\epsilon}} \\ &= 10 \lg \frac{\epsilon \times 255^2}{(C_1 \times C_2^{QP})^2} = a \times QP + b \quad (3) \end{aligned}$$

其中,  $C_1, C_2, a, b$  均为模型中的常数。从图 3 中可以观察到这一近似的线性关系。

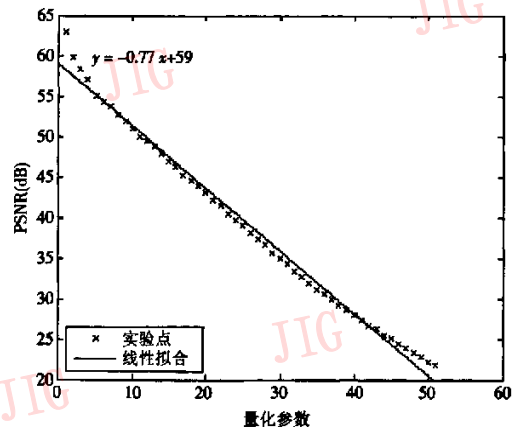


图 3 PSNR 与 QP 的近似线性关系 (Stefan 序列)

### 3.3 Linear relationship between PSNR and QP (Stefan sequence)

取实验数据中较接近拟合线性曲线的中间段,令 QP 为 10 和 40, 编码两次, 分别得到  $PSNR_1$  和  $PSNR_2$ 。这样就可以计算出线性曲线的斜率  $k$  和截距  $b$ 。再将  $PSNR_{target}$  代入, 即可得到初始 QP:

$$QP_{init} = \text{round} \left[ \frac{PSNR_{target} - b}{k} \right] \quad (4)$$

其中,  $\text{round}[\cdot]$  表示四舍五入运算。

### 3.3 时域 $T_0$ 层图像帧的恒定质量控制

保持  $T_0$  层 IPPP 预测结构的图像质量的稳定性对整个序列恒定质量控制的效果十分关键。QP 的选择主要基于过去若干个已编码  $T_0$  层图像帧的平

均 PSNR 值。若该平均值与目标 PSNR 值的差异在阈值  $\Delta$  之内,当前帧将使用和上一帧相同的 QP 值进行编码:

$$\left| \frac{1}{N} \sum_{i=1}^N \text{PSNR}_{j,i} - T \right| < \Delta \Rightarrow \text{QP}_j = \text{QP}_{j-1} \quad (5)$$

其中,  $j$  为下一待编码帧的序号,  $T$  为目标 PSNR 值,  $N$  为滑动窗口的长度, 决定计算平均值时使用已编码帧的个数。而在差异大于阈值  $\Delta$  时, 根据该差值的正负, 编码当前帧的 QP 将相应的增加或减小:

$$\left| \frac{1}{N} \sum_{i=1}^N \text{PSNR}_{j,i} - T \right| > \Delta \Rightarrow \text{QP}_j = \text{QP}_{j-1} + \delta \quad (6)$$

$\delta$  可由下式计算出:

$$\delta = \pm \min \left\{ \left\lceil \gamma \left| \frac{1}{N} \sum_{i=1}^N \text{PSNR}_{j,i} - T \right| \right\rceil, K \right\} \quad (7)$$

其中, 运算符  $\lceil \cdot \rceil$  表示截尾取整。  $\delta$  和  $\left( \frac{1}{N} \sum_{i=1}^N \text{PSNR}_{j,i} - T \right)$  具有相同的正负性。QP 调整的最大步长由参数  $K$  控制, 以防止算法过分剧烈地自适应而造成局部质量效果的大幅变化。参数  $\gamma$  可以看作是一个与视频内容有关的尺度因子。

对于式(5)中滑动平均窗口的长度  $N$  而言, 如果  $N$  较小, 算法会对任何原因所造成的 PSNR 值的偏差而迅速做出反应和调整。另一方面, 如果  $N$  较大, 当新场景开始时, 由于需要逐帧顺序替换过去被记忆的 PSNR 值, 算法自适应的反应会较慢。考虑到 Hierarchical-B 结构的  $T_0$  层编码虽看似为传统 IPPP 预测, 但相邻帧的预测距离为一个 GOP 的长度, 经过实验测试, 将  $N$  的值设为 3。

对于阈值  $\Delta$  而言, 显而易见的是当  $\Delta$  值较小时, 算法的调整会非常敏感。因为 QP 只能在整数范围变化,  $\text{QP} \pm 1$  通常导致 PSNR 变化 0.5dB 左右, 较小的  $\Delta$  值只会造成无意义的调整和 PSNR 的宽幅振荡。相反, 如果  $\Delta$  值较大, 算法的反应会比较迟缓。经过实验测试, 将  $\Delta$  值设为 0.5。另外,  $K$  和  $\gamma$  分别取经验值 2 和 1.8。

以上方法仍存在一定的缺陷。由于 PSNR 值在完整重建一帧之后才能得到, 故具有一定的滞后效应, 尤其在场景切换时和高速运动序列中可能会造成 QP 调整不及时而导致后续图像帧连续发生本不应有的质量波动。

使用 MAD 对于 I 帧和 I 宏块, 使用原始像素的绝对值来代替残差像素的绝对值) 和 MAD 比率作为帧复杂度的度量, 并以此来检测序列中视频性质变化剧烈的时刻。

$$\text{MAD}_j = \frac{\sum_{\text{pix} \in F} |\text{pix}_{\text{orig}} - \text{pix}_{\text{ref}}|}{w \times h} \quad (8)$$

其中,  $F$  为当前帧所有像素点的集合,  $\text{pix}_{\text{orig}}$  为原始像素值,  $\text{pix}_{\text{ref}}$  为运动补偿后参考帧中的相应像素值,  $w, h$  为图像的宽度和高度。

$$\text{MAD}_{\text{ratio}}(j) = \frac{\text{MAD}_j}{\frac{1}{j-1} \left( \sum_{i=1}^{j-1} \text{MAD}_i \right)} \quad (9)$$

$\text{MAD}_{\text{ratio}}(j)$  为当前帧  $j$  的预测 MAD 与先前已编码的各 P 帧的实际 MAD 的平均值之比。  $\text{MAD}_j$  用线性模型预测<sup>[3]</sup>:

$$\text{MAD}_j = \alpha \times \text{MAD}_{j-1} + \beta \quad (10)$$

$\alpha, \beta$  的初始值为 1, 0。之后用线性回归的方法逐帧更新。QP 进一步调整的策略为

$$\begin{cases} \text{QP}_j = \text{QP}_j - 1 & \text{when } \text{MAD}_{\text{ratio}} < 0.6 \\ \text{QP}_j = \text{QP}_j + 1 & \text{when } \text{MAD}_{\text{ratio}} > 1.5 \end{cases} \quad (11)$$

### 3.4 时域增强层图像帧的恒定质量控制

由 Hierarchical-B 结构可知, 一个 GOP 内各时域增强层的参考关系是确定的, 编码时的顺序也是固定的。如果只考虑最基本的单参考帧情况, 则直接影响具体某一帧编码效果的就是该帧的前后两个参考帧, 如图 5 所示。

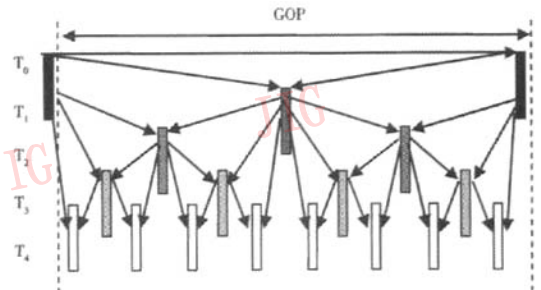


图 5 GOP 内各时间层的预测依赖关系

Fig. 5 Prediction dependencies for temporal layers

由图 5 可见, 前一个 GOP 的  $T_0$  图像帧是当前 GOP 内所有待编码帧的预测基础。对于 GOP 内具体的某一待编码帧  $F_j$ , 其左右两个参考帧可能属于不同的时域增强层, 其中在编码顺序上更靠近  $F_j$  的那一帧在参考中应占主导地位, 根据实验测试, 这一因素也可以用 MAD 来度量。随着时域分级的增加 (即预测距离缩短), 每一帧的 MAD 值也逐渐变小。可采取的主要策略就是通过前后两个参考帧的已编码信息来决定当前帧的量化参数  $\text{QP}_{j-1}$ 。

$$\begin{cases} QP_{j-1} = QP_L & (MAD_L < MAD_R) \\ QP_{j-1} = QP_R & \text{其他} \end{cases}$$

式中,下标 L,R 分别对应当前帧的左参考帧和右参考帧,如图 5 所示。随后,用与式(5),式(6),式(7)类似的方法调整 QP,得到  $QP_j$ ,显然此时  $N$  取值为 2。

另外,当式(5)起作用的情况下,还需要额外考虑了时域级别(即预测距离)的因素。以  $T_2$  为界,在编码  $T_1$  和  $T_2$  层时,因为预测距离相对较远,通过实验测试,发现对于某些 MAD 值很大的参考帧,即使其本身的 PSNR 在阈值范围内,但若不调整 QP,参考该帧的待编码帧的 PSNR 会产生较大的偏离。

最后,在 GOP 内依次编码的过程中,再次使用 PSNR 滑动平均窗口来监测 PSNR 的波动情况,即使经过先前的算法,QP 值没有变化,但如果此时的 PSNR 差异超过了阈值,仍需要对 QP 做出调整,以此来保证质量控制的总体效果。

### 4 实验结果和性能分析

根据 JVT-V201 文档所规定的标准对提出的算法进行了仿真。用整体 PSNR 的方差以及平均 PSNR 与目标 PSNR 的偏离程度作为衡量标准。表 1 给出了针对几个序列的实验结果,其中涵盖了场景切换,高速运动和低速运动等情况。图 6 则详细给出了逐帧 PSNR 比较的结果。从实验结果证实了算法的有效性。

表 1 改进算法与 Hierarchical-B 结构的性能比较

Tab.1 Performance comparison of the proposed scheme and Hierarchical-B structure

测试序列	PSNR (dB)		PSNR 的方差		比特率 (Kbps)
	目标值	实验值	H-B	本文算法	
Stefan	32.2	32.16	0.34	0.08	840
	29.5	29.47	0.35	0.10	454
Foreman	34.1	34.02	1.91	0.73	241
	31.7	31.59	2.14	0.77	141
Tempete	35.1	35.07	0.03	0.02	1098
	31.9	31.88	0.04	0.02	545
Football	33.75	33.6	2.81	0.92	934
	31.2	31.12	2.93	0.88	581

本文恒定质量控制算法的主要计算负担在于每帧编码之前的 MAD 估计和编码之后有关 PSNR 的

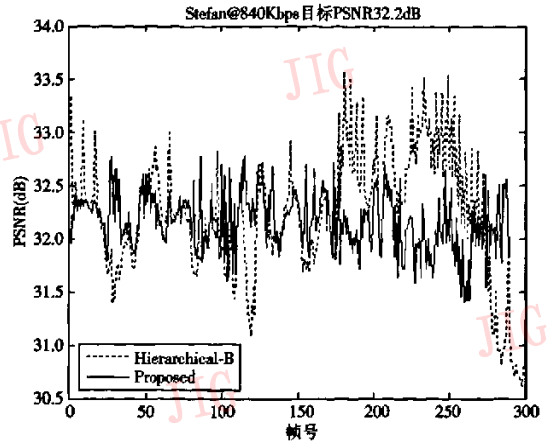


图 6 PSNR 逐帧的比较 (Stefan 序列)

Fig. 6 PSNR comparison frame by frame (Stefan)

统计,它所带来的额外的计算复杂度几乎可以忽略不计。

### 5 结论

本文针对 SVC 中实现时间可伸缩特性的 Hierarchical-B 结构提出了一种恒定质量控制算法,充分考虑了其特有的预测结构,利用已编码帧的平均 PSNR 并结合 MAD 值动态调整帧量化参数 QP。实验结果表明,该算法有效地降低了编码过程中的 PSNR 波动,同时和预先设定的目标 PSNR 值也十分接近。将来的工作可以考虑继续在宏块级建立 QP 调整模型,进一步提高算法的性能。

### 参考文献 (References)

- 1 Fabio D V, Juan Carlos De Martin. PSNR control for GOP-level constant quality in H. 264 video coding [A]. In: Proceedings of the 5<sup>th</sup> IEEE International Symposium on Signal Processing and Information Technology [C], Athens, Greece, 2005; 612 ~ 617.
- 2 Xie B, Zeng W. Sequence based rate control for constant quality video [A]. In: Proceedings of International Conference on Image Processing [C], Rochester NY, USA, 2002; 77 ~ 80.
- 3 Jiang M, Yi X, Ling N. Frame Layer bit allocation scheme for constant quality video [A]. In: Proceedings of International Conference on Multimedia Expo [C], Taipei, China, 2004; 1055 - 1058.
- 4 Heiko Schwarz, Detlev Marpe. Overview of the scalable extension of the H. 264/MPEG-4 AVC video coding standard [R]. Joint Video Team, DOC. JVT-W132, San Jose, CA, USA, April, 2007.
- 5 Jayant N, Noll P. Digital Coding of Waveforms [M]. Englewood Cliffs, NJ: Prentice Hall, 1994.