

一种面向 SVC 的码率控制算法

徐 龙¹⁾ 高 文^{1),2)} 季向阳¹⁾ 赵德斌³⁾

¹⁾(中国科学院计算技术研究所 北京 100190)

²⁾(北京大学数字媒体研究所 北京 100871)

³⁾(哈尔滨工业大学计算机科学与技术系 哈尔滨 150001)

摘 要 H. 264/AVC 可伸缩性扩展视频编码系统(JSVM)提供了一种可伸缩视频编码(SVC)的解决方案,然而它本身并没有提供一种有效的码率控制算法.文中基于 JSVM 的分层预测结构,提出了一种全新的码率控制算法.首先在码率分配方面,考虑到分层 B 帧预测(或运动补偿时间域滤波(MCTF))结构,给出一种分层的码率分配方案;然后,针对不同类型和不同时间分解层各自的统计特性,分别为它们设计了不同的率失真(R-D)模型.实验结果表明,文中算法能够有效地控制码率,使得目标码率跟实际产生码率之间的偏差最大不超过 2%;同时文中算法较大地提高了解码图像的质量,使得峰值信噪比(PSNR)在低码率端可提高 1dB;另外,JSVM 是通过不断调整量化参数(QP)使得实际产生的码率逐次逼近目标码率,较之这种尝试型的码率控制算法,文中基于模型的码率控制是在一次编码中产生最终的目标码率,从而大大降低了计算复杂度.

关键词 可伸缩编码;码率控制;率失真优化;率失真模型

中图法分类号 TP391

A Rate Control Algorithm for Scalable Video Coding

XU Long¹⁾ GAO Wen^{1),2)} JI Xiang-Yang¹⁾ ZHAO De-Bin³⁾

¹⁾(Institute of Computing Technology, Chinese Academy of Sciences, Beijing 100190)

²⁾(Institute of Digital Media, Peking University, Beijing 100871)

³⁾(Department of Computer Science and Technology, Harbin Institute of Technology, Harbin 150001)

Abstract The joint scalable video model (JSVM) as a scalable extension of H. 264/AVC provides a high efficient solution for scalable video coding (SVC). However, there is not efficient rate control scheme is specified. This paper presents a new rate control scheme for the JSVM based on the special structure of hierarchical B decomposition (or motion compensation temporal filtering (MCTF)). In a sense of bit allocation, a novel hierarchical bit allocation strategy is proposed. Meanwhile, the different rate-distortion (R-D) model is provided for each kind picture with different picture type or temporal decomposition level. Experiment results show that the mismatch of target bit rate and real bit rate generated for coding doesn't exceed 2%; meanwhile, the PSNR improvement can be up to 1dB in low bit rate; furthermore, comparing to the original rate control method which achieves the target bit rate through regulating quantization parameter (QP) step by step, the computation complexity of the proposed rate control algorithm base on experimental model is reduced drastically.

Keywords scalable video coding; rate control; rate distortion optimization; rate distortion model

收稿日期:2006-09-11;最终修改稿收到日期:2008-06-06. 本课题得到北京市自然科学基金(4041003)、国家自然科学基金(60333020)、国家科技攻关计划(2002BA906A35)和国家“八六三”高技术研究发展计划项目基金(2004AA119010)资助. 徐 龙,男,1977 年生,博士,主要研究方向为可伸缩视频编码、3D 小波视频编码与图像处理. E-mail: lxu@jdl.ac.cn. 高 文,男,1956 年生,博士,教授,博士生导师,主要研究领域为多媒体数据压缩、图像处理、计算机视觉、多模式接口、人工智能与虚拟现实等. 季向阳,男,1976 年生,博士,主要研究方向为视频编码、图像处理、3D 小波视频编码. 赵德斌,男,博士,教授,博士生导师,主要研究领域为视频压缩、图像处理与模式识别.

1 引言

随着计算机、通信和各种娱乐媒体的交互、融合,如今各种数字媒体应用通过通信网络连为一体.在网络环境里(比如互联网),各种终端系统可能具有不同的再现能力和应用需求.所以为了某种特定的应用而一次压缩的码流并不是令人满意和有效的,对一些特定的用户或设备而言,甚至是没有意义的.解决这个问题一个有效的方法就是可伸缩编码.在可伸缩编码里,为低端应用而产生的低码率码流作为子集嵌入在为高端应用而产生的高码率码流中.这样,对于一次性压缩的高码率码流,通过选择性地传输和解码其中的部分码流,就能够得到适用于多种应用和多种设备的解码结果.可伸缩功能主要包括信噪比、空间分辨率、时间分辨率和复杂度可伸缩性.

目前,基于 H. 264 可伸缩扩展的视频编码系统^[1]被采纳作为可伸缩视频编码的标准,称为 Joint Scalable Video Model (JSVM). JSVM 在时间域上采用基于运动补偿时域滤波(MCTF)的开环预测或分层 B 帧(Hierarchical B)的闭环预测^[2-3].以一组图像(GOP)为单位,经过分层 B 帧预测(或 MCTF^[4-6]),产生一组不同频率的子带图像(subband),包括低通(low-pass)子带和若干高通(high-pass)子带,子带的编码采用 H. 264/AVC 的方法.通过依存选择性地传输或解码部分子带(频率从底到高)实现时间域分辨率的可伸缩性;为了取得空间域可伸缩性,对原始视频序列经过多次下采样得到多个分辨率的视频序列,这些序列经过过去相关一起压缩到一个码流文件,实际应用环境中选择性地传输或解码相应分辨率的序列可以获得多个空间分辨率的解码图像;信噪比(SNR)可伸缩性通过传统的 FGS 技术实现.

JSVM 提供了可伸缩编码的有效解决方案,通过一次编码产生的高码率码流,可以根据用户需求和应用环境,得到各自时域(帧率)、空域和信噪比的解码.但是这种可分级毕竟是有限的,不可能得到任意连续(无缝的)分辨率的解码,只有码率控制可以产生任意精细程度的码流,并且,获得可伸缩码流本身需要码率控制.除此之外,码率控制涉及的一系列优化策略能够有效地提高编码的性能.总之,作为任何编码标准的应用,码率控制是不可缺少的,同样对于 JSVM 码率控制的研究价值是肯定的.传统的码率

控制模型有 MPEG2 的 TM5^[7]、H. 263 的 TMN8^[8]和 TMN12^[9]以及 MPEG4 的 VM8^[10]等.对于 H. 264/AVC,一种基于 VM8 改进的码率控制算法^[11]和一种基于率失真优化和假设参考模型(HRD)的码率控制算法^[12],被采纳并广泛应用.

码率控制不但涉及量化参数的选择而且涉及最优模式决策和最优码率分配,同时还应该考虑解码器的溢出.首先,率失真模型表示码率和量化参数之间的关系,给定目标码率,根据率失真模型计算编码所用的量化参数从而使得实际编码所用的比特数尽可能地接近目标码率.例如在 TM5 中,采用了一个简单的基于缓冲区变化的线性率失真模型;而 H. 264 采用了更为精确的二次率失真模型;也有方法是根据量化参数和量化后零系数所占比例的关系建立率失真模型,称为 ρ 域模型.其次,码率控制通常跟率失真优化紧密地联系在一起,二者互为一个矛盾统一体.另外,在广泛运用的帧间预测编码里,运动估计产生的运动向量和残差信息之间需要一个折中,通常是用拉格朗日乘数法来实现此率失真优化的.率失真优化通过选择最优编码模式和最优码率分配取得最优的编码性能.

在当前的 JSVM 中,码率控制从给定的初始量化参数出发,通过二分搜索算法得到最终的量化参数.这种算法产生的多次循环增加了时间复杂度,并且算法本身没有考虑到编码图像(或宏块)的属性.本文针对 JSVM 提出了一种全新的基于图像内容的自适应码率控制算法,码率的分配和率失真模型的构造都充分考虑到了 JSVM 特殊的分层结构.

本文第 2 节展示部分测试序列的率失真性能曲线以及在此统计结果上建立的率失真模型;第 3 节详细描述分层的码率分配算法;实验结果和分析在第 4 节;第 5 节对全文进行了总结.

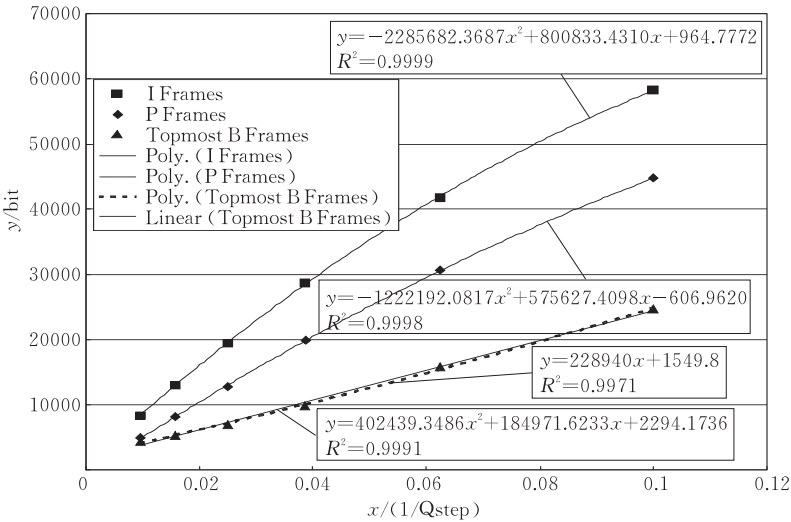
2 分层率失真模型

码率控制所要达到的一个目标是:(1)决定适当的量化参数使得所编码的比特流满足给定的目标码率;(2)解码图像的失真最小.对于(1),由于编码中码率的产生由量化参数决定,而实际应用中往往给定码率,这就需要用给定的码率求得相应的量化参数,这个问题的解决取决于率失真模型,通常率失真模型是一个码率和量化参数的方程.

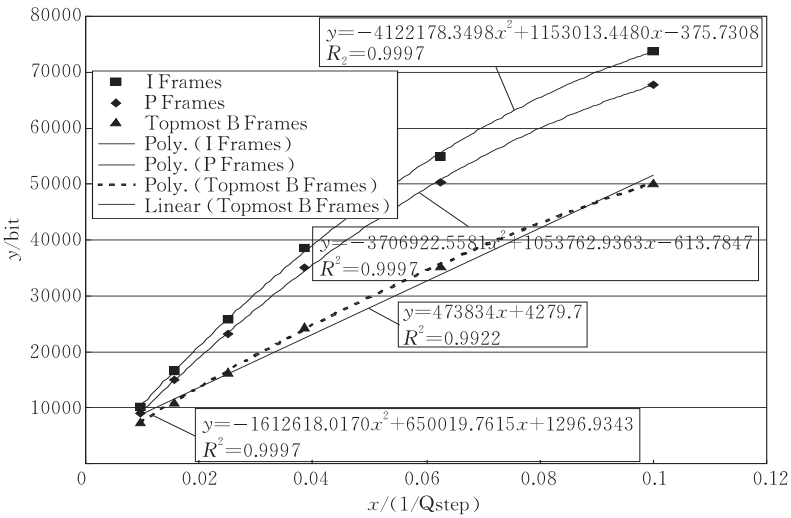
在 JSVM 分层 B 帧的预测结构(或 MCTF)中,产生了一系列特征(主要是频率)不同的帧,包括低

通子带(I、P 帧)和不同层上的高通子带(B 帧). 我们对一些序列作了细致的统计分析,得到分别针对于不同类型帧的率失真性能曲线(如图 1 所示). 实验的目的是给不同类型的帧指定不同的率失真模型,从结果可以看出,B 帧的率失真性能的确跟 I 帧和 P 帧相差很大,I 帧和 P 帧亦有较大的差别. 所以我们提出的解决方案是:(1)给定 I 帧和 P 帧不同的率失真模型,模型参数在相邻若干个 GOP 的各自

帧类型范围内更新;(2)给定不同层次 B 帧不同的率失真模型,模型参数除了在相邻 GOP 间的相应层次的帧间更新外,还在本层内帧间更新;(3)最底层 B 帧的量化参数由次底层插值得到,建议不使用码率控制;(4)关于 I、P 和 B 帧采用的率失真模型的类型为:I 帧和 P 帧分别采用二次率失真模型,而 B 帧采用线性的率失真模型.



(a) foreman (QCIF)



(b) football (QCIF)

图 1 率失真性能曲线及其线性、二次曲线拟合

图 1 中“I Frames”代表 I 帧的量化步长与编码的比特数之间的关系,“P Frames”代表 P 帧的量化步长与编码比特数之间的关系,二者的率失真曲线比较接近,而与“B Frames”(代表 B 帧的量化步长与编码比特数之间的关系)相差较远. 图 1 中“Poly I Frames”、“Poly P Frames”和“Poly B Frames”分别代表 I、P 和 B 帧率失真性能的二次拟合曲线,可以

看出三者率失真性能的拟合都非常符合二次方程,相关性在 0.99 以上,所以在码率控制中使用二次率失真模型能够很好地控制码率,使得实际产生的码率与目标码率尽可能匹配. 然而,二次率失真模型的更新比较复杂,并且相对线性模型需要更多的样本点,故而对于 B 帧,我们选择了线性率失真模型. 就 B 帧而言,我们做了一系列测试,统计得出 B 帧使用

线性模型,相关性在 0.99 以上,而 I、P 帧使用线性模型相关性只在 0.97 以上,所以在我们提出的码率控制中,所有 B 帧使用线性率失真模型.图 1 中只给出了最高层 B 帧(“foreman”和“football”)的率失真曲线以及它的线性拟合,可以看出其最高层率失真性能曲线非常符合线性模型,相关性达到 0.99.

基于传统 H. 264/AVC 码率控制的率失真模型,我们对于 I 帧和 P 帧的二次率失真模型为

$$R = c_1 \times \frac{\sigma^2}{QP} + c_2 \times \frac{\sigma^2}{QP^2} \tag{1}$$

而对于各个时域分解层上的 B 帧,实验统计数据结果表明可以用简单的线性曲线拟合,其线性模型为

$$R = \frac{c \times \sigma^2}{QP} \tag{2}$$

其中 R 为分配给当前帧(或宏块)除去编码头信息(header bits)的比特数, σ^2 为当前帧(宏块)的 MAD (Mean Absolute Distortion), QP 为当前帧(宏块)的量化参数(Quantization Parameter, QP), c, c_1 和 c_2 是常数.时域分解的 B 帧在各自层次上使用各自的率失真模型,模型参数在各自层次的帧上更新,其中二次模型参数的更新使用线性回归方法^[13-14],一次模型的更新简单地使用若干历史数据和当前数据的加权平均.

3 分层的码率分配策略

在 JSVM 的时域分解中,有两种不同的分解方式:(1) MCTF. 它是一个倒立的金字塔结构,低通子带处于金字塔的最底层,是原始若干帧低通信息的汇聚,这里的 I、P 帧都属于时域分解的低通子带;(2) Hierarchical B. 它是一个金字塔结构,最底层为最高频 B 帧,从低到高,频率依次降低,其对整个编码性能的影响也越来越大,金字塔的顶端为帧内编码的 I 帧或帧间前向预测编码的 P 帧.如图 2(a)所示 3 级 MCTF,8 帧原始图像经过 3 级 MCTF 分解得到 1 帧低通子带(I 帧或 P 帧)和分布在 3 级上的高通子带;图 2(b)所示为 4 层 Hierarchical B 分解,原始 16 帧经过 4 级 Hierarchical B 分解得到一个 I (或 P)以及分布在 4 层上的 B 帧.

在我们以前的工作^[15]中,基于 MCTF 的结构,每个子带的码率分配由当前 GOP 所剩的比特数和缓冲区变化来决定;同时,各个子带间的码率分配由一个权重参数来调整,这个权重参数由以下两个因素决定:一方面,由于 MCTF 所用的非归一化的滤波器,滤波器的归一化被放置到量化里来实现;另一

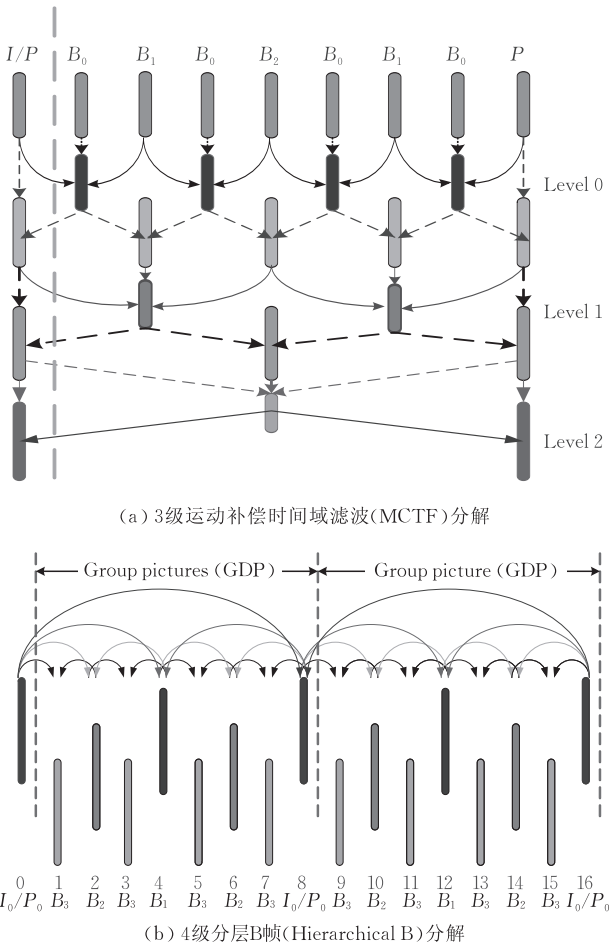


图 2 JSVM 时域分解结构

方面,不同频率的帧对于整体编码性能的影响力不同,低频帧在编码中的影响力最大.码率在这些子带间按照各自的复杂度信息(宏块系数哈达玛变换的平方累加和)和重要性信息(权重参数)来分配.依图中表示,最底层的帧为 MCTF 分解的低通子带,含有原始视频序列的低频信息,它对于编码性能的影响是至关重要的,接下来是次底层的高通子带,依存从下到上,各个层上的高通子带频率依次升高,对整体编码性能的影响依存递减.基于这种考虑,指定各个子带一个权重因子(scaling factor),参考文献[1],本文给出了 MCTF 分解子带从高层到底层(频率从低到高)对应的权重因子为: $\{1.000f, 1.216f, 1.478f, 1.794f\}$.除了 scaling factor,我们还考虑了滤波器的归一化问题,JSVM 的 MCTF 编码结构中所涉及的滤波器系数和相应的滤波器能量为: Haar 低通:系数为 $\{1/2, 1/2\}$,能量为 $\sqrt{1/2}$; Haar 高通:系数为 $\{1, -1\}$,能量为 $\sqrt{2}$; 样条 5/3 低通:系数为 $\{-1/8, 1/4, 3/4, 1/4, -1/8\}$,能量为 $\sqrt{23/32}$; 样条 5/3 高通:系数为 $\{-1/2, 1, -1/2\}$,能量为 $\sqrt{3/2}$.

根据滤波器能量,两个相邻层子带(对应滤波前和滤波后)之间量化参数(实际涉及到计算都是指量化步长,量化参数只是作为量化步长的索引)的关系由以下公式表示:

$$\begin{cases} QP_{H,n} = QP_{L,n} \times (1/\alpha) \times \beta \\ QP_{L,(n-1)} = QP_{L,n} \times (1/\alpha) \end{cases} \quad (3)$$

这里 n 为 MCTF 的级数, $QP_{L,n} (t=L, H)$ 为 n 级 t 类型子带的量化参数, 常数 $\alpha = \sqrt{23/32}$, $\beta = 3/2$ 分别表示低通和高通滤波器的能量. 从而相邻层高通子带与低通子带间能量的比例关系为 β/α , 相邻两个低通层间能量的比例关系为 $1/\alpha$. 式(3)这种层间量化参数的关系类似地在 JSVM 固定量化参数的编码中已经出现^[1-2]. 可以看出, 上层子带的量化参数可以由底层子带(最底层为 MCTF 分解得到的低通子带)的量化参数依照式(3)导出; 同时, 两层高通子带量化参数之间的关系也可以根据式(3)导出:

$$\begin{aligned} QP_{H,n-1} &= QP_{L,n-1} \times (1/\alpha) \times \beta \\ &= (QP_{L,n} \times (1/\alpha)) \times \beta \times (1/\alpha) \\ &= QP_{H,n} \times 1/\alpha \end{aligned} \quad (4)$$

假设码率与量化参数之间为简单的反比例关系, 则第 i 个高通子带的目标码率可以表示为

$$\begin{cases} T_i^h(j) = \frac{\alpha^l \times \sigma_i^h(j) \times B_i(j)}{\sum_{m=i}^N \alpha^l \times \sigma_m^h(j)} \\ i = 0, 1, \dots, N-2, N-1; l = 0, 1, \dots, d-1 \end{cases} \quad (5)$$

对于如图 2(a) 中唯一的低通子带, 式(5)相应地修改为

$$\begin{cases} T_0^{l/p}(j) = \frac{\sigma_0^{l/p}(j) \times B_0(j)}{\sigma_0^{l/p}(j) + \sum_{m=1}^N \beta \times \alpha^l \times \sigma_m^b(j)} \\ l = 0, 1, \dots, d-1 \end{cases} \quad (6)$$

这里 $B_i(j)$ 是编码第 i 个子带之前, 当前(第 j) GOP 剩余的比特数; $\sigma_i^t(j) (t=I, p, b)$ 是当前 GOP 第 i 子带(类型为 t)的复杂度(Sum Absolute Distortion, SAD); l 为 MCTF 分解的层次, 假设分解层次为 d , 如图 2, 则低频帧处于分解的顶层(记为 d), 其它高通子带分别位于的 $d-1$ 到 0 层, GOP 的大小为 N .

最后, 综合考虑归一化和权重因子, 则相应的式(5)、式(6)修改为

$$\begin{cases} T_i^h(j) = \frac{S_l \times (23/32)^l \times \sigma_i^h(j) \times B_i(j)}{\sum_{m=i}^N S_l \times (23/32)^l \times \sigma_m^b(j)} \\ S_l = \{1, 1.2, 1.4, 1.7\} \end{cases} \quad (7)$$

和

$$\begin{cases} T_0^{l/p}(j) = \frac{S_d \times \sigma_0^{l/p}(j) \times B_0(j)}{\sigma_0^{l/p}(j) + \sum_{m=1}^N S_l \times (3/2) \times (23/32)^l \times \sigma_m^b(j)} \\ S_l = \{1, 1.2, 1.4, 1.7\} \end{cases} \quad (8)$$

式中参数 l, d 和 i 如式(5)、式(6)所示, S_l 表示第 l 层的权重因子.

对于 Hierarchical B 结构, 码率分配类似于 MCTF 码率分配的推理过程, 只是涉及到的滤波器不同, 从而式(3)~式(8)中涉及到的比例因子需要修改. 简单地来说, Hierarchical B 结构中不包含 MCTF 分解中提升格式小波变换的更新(update)^[5-6]过程, 从而式(3)~式(8)中代表低通滤波器(样条 5/3)能量的常数因子 α 始终为 1.

4 实验结果及讨论

为了评价本文的码率控制算法, 我们分别在一系列序列(包括“football”, “foreman”, “mobile”和“harbour”等)上进行了测试. 测试结果表明, 该算法可以有效地控制码率; 并且相对于原始搜索量化参数的码率控制(min-max)算法, 该算法提高了编码的性能, 特别是在低码率端, 图像的 PSNR 提高了近 1dB; 同时, 由于该算法中量化参数的决策是在一次过程中取得, 所以时间复杂度降低了.

根据文献[16]的性能分析和文献[17-18]的实验结果, 对于低频帧使用较小 QP, 而高频帧使用较大 QP 的策略较之所有帧使用相同 QP(后面称为固定 QP)的方法能够极大提高编码的性能, 这种量化称为分层或层叠(cascade)量化. 其中, 在文献[17]中, 我们针对单层 H. 264/AVC 的 Hierarchical B 编码, 基于与本文相同的思想提出了使用伸缩因子(scaling-factor)进行最优码率分配的码率控制算法; 而且, 给出了所有帧使用相同 QP、传统的码率控制、分层的多级 QP 和我们提出的码率控制的性能对比结果. 如图 3 和图 4 分别展示了“football”和“bus”(176×144(QCIF), 30Hz)的 H. 264/AVC 单层编码结果(JSVM 基本层编码与 H. 264/AVC 完全兼容), 可以看出, 使有多级 QP 和我们提出的使用 scaling-factor 的码率控制算法, 可以大大提高编码的性能, 图中“Constant QP”表示所有帧使用相同 QP 的编码, “Existing RC”表示使用传统码率控制, “Hierar QP”表示使用多级 QP. 与所有帧使用相同 QP 编码比较, 多级 QP 和我们的码率控制算法可以提高 PSNR 2dB 多; 而且, 如果使用传统的码

率控制^[12],可以看出它并不适合新的分层 B 帧编码结构,PSNR 较固定 QP 编码下降超过 1dB. 在文献[18]中,针对空间分辨率层,我们提出了一种新的率失真模型,以更好地利用层间编码信息.

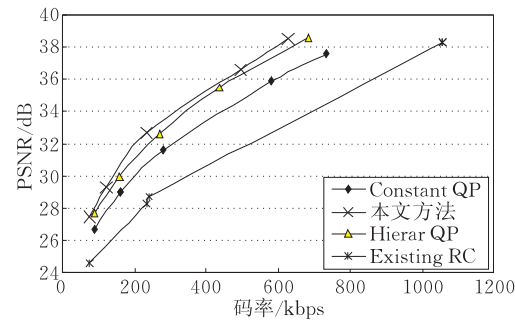


图 3 “football”PSNR 对比图

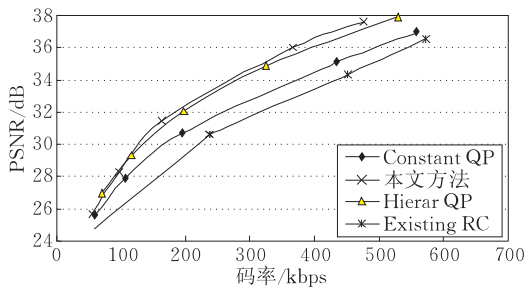
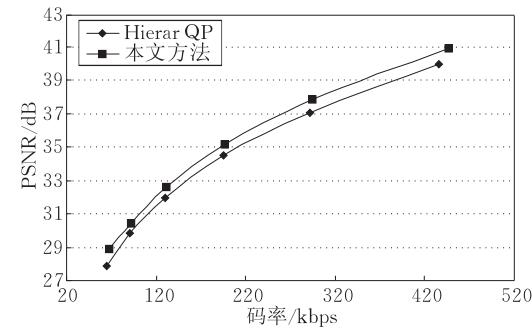
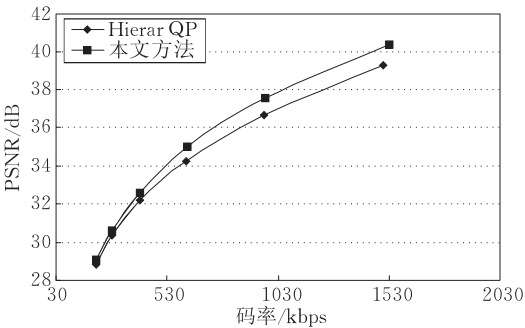


图 4 “bus”PSNR 对比图

图 5~图 8 给出“football”,“foreman”,“mobile”和“harbour”在两级空间分辨率层(176×144 (QCIF)15Hz, 352×288 (CIF)15Hz)上从低码率到

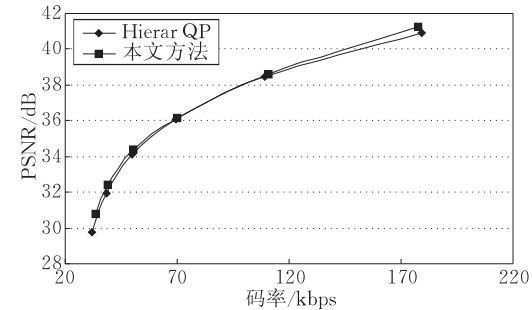


(a) football (176×144 15Hz)

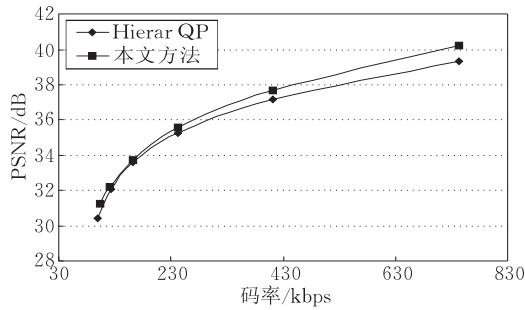


(b) football (352×288 15Hz)

图 5 多级 QP 和码率控制结果 PSNR 对比(football)



(a) foreman (176×144 15Hz)



(b) foreman (352×288 15Hz)

图 6 多级 QP 和码率控制结果 PSNR 对比(foreman)

高码率的 PSNR 对比. 其中“football”在各个码率点上(QCIF 在 60kbps 和 450kbps 之间,CIF 在 200kbps 和 1500kbps 之间)的 PSNR 较多级 QP 的结果提高了 0.3~1.1dB;“foreman”在基本层(QCIF,15Hz)低码率端(小于 38kbps)的 PSNR 提高了 1.0dB;“mobile”和“harbour”在基本层(QCIF 15Hz)低码率端也分别有大于 0.75 dB 和 1.0dB 的增益;“football”,“foreman”和“mobile”在增强层(CIF,15Hz)最高码率点上的 PSNR 分别有 1.09dB, 0.96dB 和 0.84dB 的增益. 总体来看,本文所提出的分层结构的码率分配方案做到了码率的优化分配,从而使得编码的性能大大提高;同时,本文所提出的码率控制算法(RC)较好地工作于从低码率到高码率的各个码点上. 表 1~表 4 分别详细给出了以上各个测试序列的图像 PSNR 和码率以及与使用多级 QP 编码的图像的 PSNR 和码率的对比结果;同时,码率控制的误差($\Delta BITS_Error$)和 PSNR 的增益($\Delta PSNR_Y$)也在表 1~表 4 中给出. 可以看出总体码率控制的误差保持在正负 2%误差范围内,对于基本层(QCIF, 15Hz) PSNR 的增益在 0.3~1.1dB 之间,对于增强层,PSNR 的增益在 0~1.1dB 之间. 图 5~图 9 和表 1~表 4 中,为了与码率控制结果对比,都给出了多级 QP 量化的编码结果,因为在 JSVM,多级 QP 相对固定 QP 方法的编码增益达到了 1~2dB.

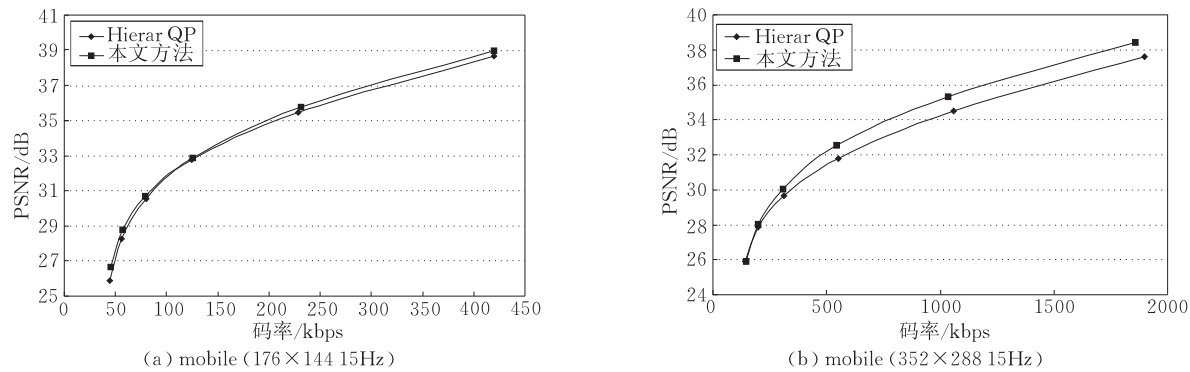


图 7 多级 QP 和码率控制结果 PSNR 对比(mobile)

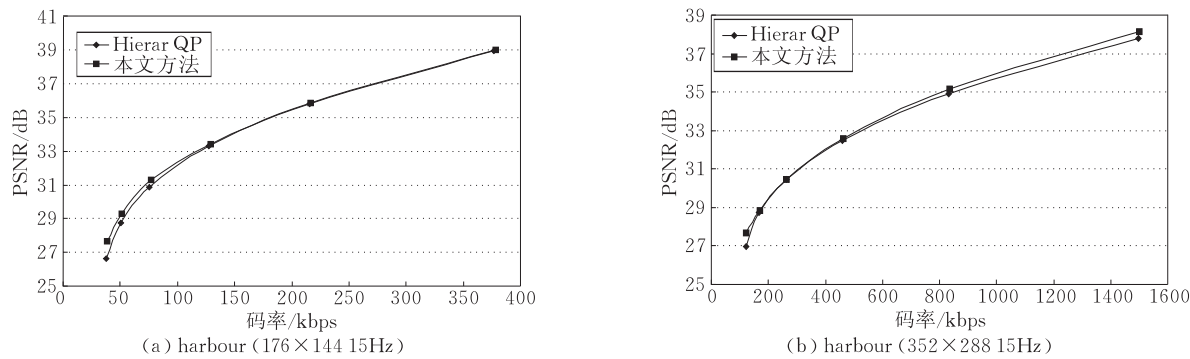


图 8 多级 QP 和码率控制结果 PSNR 对比(harbour)

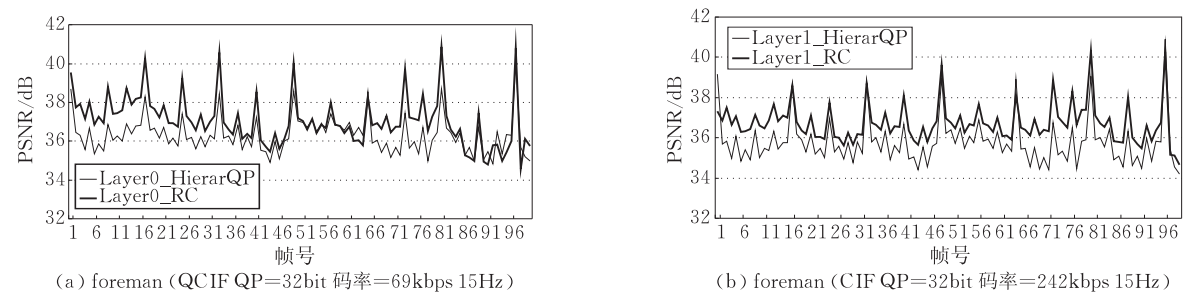


图 9 多级 QP 和码率控制单帧图像的 PSNR 对比

表 1 “football” 多级 QP 和码率控制结果对比

固定 QP 的方法				本文方法				$\Delta PSNR_y$ /dB	$\Delta BITS$ Error
码率	Y	U	V	码率	Y	U	V		
435.8400	39.9776	43.2618	44.4631	446.9967	40.8796	43.7484	45.0040	0.9020	0.026
291.4745	37.0344	40.9381	42.6291	294.9219	37.8552	41.6172	43.2944	0.8208	0.012
195.7992	34.4698	39.0138	41.0352	196.6982	35.1303	39.8065	41.7985	0.6605	0.005
129.8266	31.9160	37.4470	39.8182	130.7967	32.6426	38.1627	40.4865	0.7266	0.007
90.4556	29.8129	36.1262	38.7643	91.8373	30.4225	36.9022	39.5049	0.6096	0.015
65.4254	27.9089	34.8166	37.6010	67.0496	28.8588	36.0051	38.7952	0.9499	0.025

(a) FOOTBALL(BL) QCIF 15Hz

固定 QP 的方法				本文方法				$\Delta PSNR_y$ /dB	$\Delta BITS$ Error
码率	Y	U	V	码率	Y	U	V		
1504.1710	39.2840	43.4958	44.6463	1531.0360	40.3792	44.0526	45.2599	1.0952	0.018
967.3877	36.6979	41.6218	43.0915	969.9562	37.5899	42.1937	43.6932	0.8920	0.003
617.2327	34.2149	39.9831	41.7505	620.2177	34.9948	40.6282	42.3457	0.7799	0.005
406.8951	32.2008	38.7358	40.7266	406.8429	32.5639	39.1259	40.9964	0.3631	0.000
282.3251	30.3349	37.6961	39.7693	282.3224	30.6006	37.7789	39.8871	0.2657	0.000
209.5056	28.8003	36.7053	39.0242	211.4201	29.0838	37.0097	39.3686	0.2835	0.009

(b) FOOTBALL(EL) CIF 15Hz

表 2 “foreman” 多级 QP 和码率控制结果对比
(a) FOREMAN(BL) QCIF 15Hz

固定 QP 的方法				本文方法				$\Delta PSNR_Y$ /dB	$\Delta BITS Error$
码率	Y	U	V	码率	Y	U	V		
179.4290	40.8809	44.5038	45.3993	177.7142	41.2554	44.3040	45.3094	0.3745	−0.010
109.1164	38.4442	42.7266	43.5938	110.9023	38.6072	42.6475	43.5474	0.1630	0.016
69.4692	36.1096	40.9989	41.8184	69.9633	36.1777	41.1232	41.9813	0.0681	0.007
50.0830	34.1054	39.8165	40.5413	50.6494	34.3577	40.2937	40.9839	0.2523	0.011
38.7319	31.9184	38.7112	39.1933	39.2910	32.4312	40.1043	40.7231	0.5208	0.014
32.1352	29.7734	37.6559	37.8835	33.5926	30.7861	39.6252	40.2051	1.0127	0.045

(b) FOREMAN(EL) CIF 15Hz

固定 QP 的方法				本文方法				$\Delta PSNR_Y$ /dB	$\Delta BITS Error$
码率	Y	U	V	码率	Y	U	V		
743.3007	39.3526	43.9486	45.5160	744.7697	40.3175	43.8648	45.6939	0.9649	0.001
411.8044	37.2076	42.4796	44.1270	415.8894	37.9885	42.4788	44.3142	0.7089	0.009
242.4091	35.2865	41.0023	42.6004	242.6463	35.5666	41.0335	42.6929	0.2801	0.001
161.4538	33.5769	39.9270	41.3453	161.5183	33.7635	39.8104	41.2319	0.1866	0.000
122.1650	32.0669	38.9163	40.0420	122.1065	32.2274	39.1321	40.3685	0.1605	0.000
100.3283	30.4314	37.9203	38.7142	103.8132	31.2700	38.8217	40.0131	0.8386	0.035

表 3 “mobile” 多级 QP 和码率控制结果对比
(a) MOBILE(BL) QCIF 15Hz

固定 QP 的方法				本文方法				$\Delta PSNR_Y$ /dB	$\Delta BITS Error$
码率	Y	U	V	码率	Y	U	V		
419.3909	38.6913	40.2244	39.907	419.5986	39.0365	40.7915	40.3437	0.3452	0.000
228.7068	35.4655	37.4848	37.0782	231.0977	35.81	37.9788	37.3698	0.3445	0.010
124.6494	32.7649	35.2973	34.6777	125.6698	32.9132	35.7340	34.9287	0.1483	0.008
79.5972	30.5634	33.7425	33.0993	79.4830	30.7057	34.0260	33.2425	0.1423	−0.001
56.1012	28.2703	32.2183	31.5536	56.6450	28.7894	33.0541	32.2630	0.5191	0.010
44.1203	25.8811	30.7784	30.1328	44.9603	26.6389	32.3416	31.5389	0.7578	0.019

(b) MOBILE(EL) CIF 15Hz

固定 QP 的方法				本文方法				$\Delta PSNR_Y$ /dB	$\Delta BITS Error$
码率	Y	U	V	码率	Y	U	V		
1900.3530	37.5987	39.6892	39.6241	1858.2790	38.4437	40.1970	40.0883	0.8450	−0.022
1057.5760	34.4895	37.5660	37.3202	1033.2660	35.3362	37.7929	37.5216	0.8467	−0.023
550.4276	31.8012	35.9059	35.4749	544.2943	32.5659	35.9572	35.5392	0.7647	−0.011
311.6772	29.6854	34.6673	34.1165	307.1561	30.0454	34.3884	33.8439	0.3600	−0.015
198.8160	27.8423	33.2860	32.7687	196.8952	28.0058	33.1804	32.6374	0.1635	−0.010
142.3945	25.8788	31.9902	31.4661	144.5059	25.9241	32.4015	31.7194	0.0453	0.015

表 4 “harbour” 多级 QP 和码率控制结果对比
(a) HARBOUR(BL) QCIF 15Hz

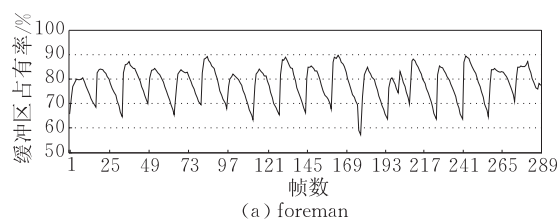
固定 QP 的方法				本文方法				$\Delta PSNR_Y$ /dB	$\Delta BITS Error$
码率	Y	U	V	码率	Y	U	V		
377.1228	38.9736	44.9030	46.5748	378.2781	39.0295	45.3112	46.9725	0.0559	0.003
215.6954	35.8099	43.1356	44.8971	216.5652	35.8649	43.2576	45.0385	0.0550	0.004
127.6775	33.3178	41.3927	43.3965	128.8461	33.4209	41.8052	43.9067	0.1031	0.009
75.6720	30.8712	40.2703	42.4406	76.8430	31.2859	40.4910	42.6246	0.4147	0.015
50.3785	28.7376	39.3906	41.3970	51.1937	29.2865	39.8674	41.9061	0.5489	0.016
37.8439	26.6311	38.4179	40.1485	38.7327	27.6765	39.3548	41.3490	1.0454	0.023

(b) HARBOUR(EL) CIF 15Hz

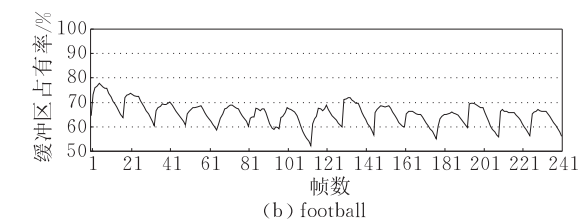
固定 QP 的方法				本文方法				$\Delta PSNR_Y$ /dB	$\Delta BITS Error$
码率	Y	U	V	码率	Y	U	V		
1498.0110	37.8054	44.6855	46.1063	1501.915	38.1254	44.4927	45.9115	0.3200	0.003
831.1183	34.9216	43.3233	44.9152	834.7092	35.1831	42.7532	44.2671	0.2615	0.004
458.1087	32.4796	41.9872	43.5392	461.3926	32.5798	41.7656	43.3946	0.1002	0.007
262.0883	30.4646	40.9378	42.5697	264.3012	30.4456	40.6954	42.3332	−0.0190	0.008
168.4701	28.7352	39.7960	41.4275	169.8968	28.8228	39.7794	41.4149	0.0876	0.008
121.7081	26.9631	38.7822	39.9450	122.9892	27.6967	39.3448	41.0302	0.7336	0.011

进一步分析结果可知,本文提出的码率控制算法对于运动较剧烈的序列(例如“football”),PSNR 有更大的提高.如图 5 所示,“football”在两层空间分辨率层上从低码率到高码率,PSNR 都有不同程度的提高,其中基本层上码率从 60~450kbps 之间,图像的 PSNR 提高了 0.66~0.94dB,增强层码率从 200~1500kbps 之间,图像的 PSNR 提高了 0.26~1.09dB.相反地,“harbour”和“mobile”运动相对缓慢,中高码率 PSNR 提高幅度不大.以上对于“football”,“mobile”和“harbour”实验结果的分析可以说明本文所提出的码率控制算法在率失真优化、码率的优化分配(包括运动矢量和编码系数的码率分配)方面对于整体编码性能的影响起到了关键的作用,正因为如此,对于运动平缓的序列,每帧的码率变化不大,这样使用码率分配方案起到的作用不如前者大,所以使用本文的码率控制算法对于运动平滑的序列的 PSNR 提高不如运动剧烈的序列的 PSNR 高.

图 9 是“foreman”单帧图像 PSNR 对比图,两条曲线分别是用固定 QP 和本文所提出的码率控制算法得到的各帧图像的 PSNR 对比.从图 9 可以看出,不论是用多级 QP 编码或以码率控制编码,单帧



(a) foreman



(b) football

图 10 缓冲区的变化曲线(增强空间分辨率层)

总之,使用了分层的率失真模型,使得编码产生的码率能够准确地匹配目标码率;使用了分层的码率分配方案,做到了有限码率的优化分配,使得编码的整体性能大大提高;另外,分层的 B 帧(或 MCTF)预测结构产生不同重要性的帧,使得码率控制在 GOP 内更加容易,实际上,码率控制中为了取得 GOP 级的码率匹配,对于次重要性帧的码率调整较大,从而校正了由于重要性较高帧编码产生码率的不匹配.

5 结 论

本文针对 JSVM 提出了一种有效的码率控制算法,其中在提高图像的 PSNR 和码率的精确匹配方面,本算法都表现出了较好的性能.本文主要在码率分配和率失真模型设计两大关键技术提出了我们自己的方法.首先,基于 Hierarchical B(或 MCTF)

图像的 PSNR 在各个低通(I/P)帧都对应一个波峰,这是由于低通帧对于一个 GOP 中的所有高通帧的质量都有影响,从而高质量的低通帧可以保证整体的编码性能,已有实验结果分析说明了这种 PSNR 波动并不影响解码图像的视觉效果.

图 10 给出了测试序列“foreman”和“football”在两层空间分辨率(QCIF 15Hz, CIF 15Hz),总体码率分别为 242kbps(基本层码率 69kbps,增强层码率 173kbps)和 617kbps(基本层码率 195kbps,增强层码率 422kbps)时的编码器缓冲区变化曲线.可以看出,编码缓冲区处于一个较平稳的周期变化的状态.变化周期为一个 GOP 时间的长度.并且由于每个 GOP 的低频帧(I/P 帧)消耗最多的码率,从而曲线变化上有一个缓冲区填充的尖峰,然后缓冲区填充度逐渐降低,直到另一个 GOP 的低频帧开始进入.另外,“football”的缓冲区较“foreman”变化大,这是由于相对“foreman”而言,“football”有更剧烈的运动,表现出各个编码帧复杂度较大幅度的变化,以及相应地编码使用比特数较大的差别,从而缓冲区数据不同的流入和流出速率使得其填充度有较大的变化.

特殊的分解结构,考虑了各个时域分解层上 B 帧(或高通子带(相当于 MCTF))对整体编码性能的影响和作为参考帧对后续编码 B 帧的影响,分别为每个帧指定了一个权重因子.在此基础上,根据每个帧的权重因子和它的复杂度得到各自的码率配额;其次,针对不同类型帧的统计特性,分别为它们指定了不同的率失真模型.实验结果证明本算法不但能够有效地控制码率,而且极大地提高了编码性能.本文涉及码率控制的其他方面,如缓冲区管理,沿用了已有 TMN8 的方法,另外,目前未考虑码率在多个空间分辨率层间的分配问题,所有这些都需要进一步考虑.

参 考 文 献

- [1] Reichel J, Schwarz H, Wien M. Joint scalable video model JSVM-6. Joint Video Team; Doc. JVT-S202, Geneva, Switzerland, 2006
- [2] Schwarz H, Marpe D, Thomas W. Overview of the scalable

- video coding extension of the H. 264/AVC standard. IEEE Transactions on CSVT, 2007, 17(9): 1103-1120
- [3] Schwarz H, Marpe D, Wiegand T. Comparison of MCTF and closed-loop hierarchical B pictures; ITU-T VCEG 16th Meeting, 2005
- [4] Taubman D. Successive refinement of video; Fundamental issues, past efforts and new directions; VCIP 2003//Proceedings of the SPIE. Lugano, Switzerland, 2003, 5120: 649-663
- [5] Flierl M, Girod B. Video coding with motion-compensated lifted wavelet transforms//Proceedings of the PCS. Saint-Malo, France, 2003: 59-62
- [6] Flierl M. Video coding with lifted wavelet transforms and frame-adaptive motion compensation//Proceedings of the VLBV. Madrid, Spain, 2003: 243-251
- [7] ISO-IEC/JTC1/SC29/WG11, MPEG2 Test model 5 Draft, April 1993
- [8] Ribas-Corbera J, Lei S. Rate control in DCT video coding for low-delay communications. IEEE Transactions on CSVT, 1999, 9(1): 172-185
- [9] Appendix II of H. 263 Version 3, Test model 12, Tenth Meeting, Portland, OR, USA, Aug. 22-24, 2000
- [10] Clziang T, Zhang Y Q. A new rate control scheme using quadratic rate distortion model. IEEE Transactions on CSVT, 1997, 7(1): 287-311
- [11] Ma Si-Wei, Gao Wen, Lu Yan. Proposed draft description of rate control. Joint Video Team (JVT) of ISO/IEC MPEG & ITU-T VCEG, Doc. JVT-F086-L, 6th Meeting, Awaji, Island, JP, Dec. 5-13, 2002
- [12] Ma Si-Wei et al. Proposed draft adaptive rate control. Joint Video Team (JVT) of ISO/IEC MPEG & ITU-T VCEG, Doc. JVT-H017r3, 8th Meeting, Geneva, May 20-26, 2003
- [13] Shin Ji-Hong, Lee Yung-Lyul, Park Hyunwook. Rate control using linear rate-model for H. 264. Signal Processing, Image Communication, 2004, 19(14): 341-352
- [14] Kim H M. Adaptive rate control using nonlinear regression. IEEE Transactions on CSVT, 2003, 13(5): 432-439
- [15] Xu Long, Ma Si-Wei, Zhao De-bin, Gao Wen. Rate control for scalable video model; VCIP2005//Proceedings of the SPIE. Beijing, China, 2005, 5960: 525-534
- [16] Wien M, Schwarz H, Oelbaum T. Performance analysis of SVC. IEEE Transactions on CSVT, 2007, 17(9): 1194-1203
- [17] Xu Long, Gao Wen, Ji Xiang-Yang, Zhao De-Bin. Rate control for hierarchical B-picture coding with scaling-factors//Proceedings of the ISCAS'2007. New Orleans, USA, 2007: 49-52
- [18] Xu Long, Gao Wen, Ji Xiang-Yang, Zhao De-Bin, Ma Si-Wei. Rate control for spatial scalable coding in SVC//Proceedings of the PCS2007. Lisboa, Portugal, 2007



XU Long, born in 1977, Ph. D.. His research interests include video coding, wavelet and image processing.

GAO Wen, born in 1956, Ph. D., professor. His cur-

rent research interests are multimedia data compression, image processing, computer vision, multimodal interface, artificial intelligence and virtual reality.

Ji Xiang-Yang, born in 1976, Ph. D.. His research interests include video coding, image processing and 3D wavelet video coding.

ZHAO De-Bin, Ph. D., professor. His current research interests are video compression, image processing and pattern recognition.

Background

This paper is concentrated on the rate control for joint scalable video model (JSVM). In the beginning of standardization of scalable video coding (SVC), there is not efficient rate control scheme proposed for JSVM, where the bit rate controlling is realized by multi-pass encoding. Certainly, the target bit rate can be obtained by truncated the bit stream compressed with high picture quality or high bit rate, which is an important character of SVC. However, this method is hard to achieve the optimized coding performance, so a one-pass rate control scheme is proposed for JSVM in the authors' previous work. For rate control of hybrid video coding H. 264/AVC, the laboratory that the authors is currently dedicated to making important contributions in the past several years. A two-pass rate control scheme proposed individually by Siwei Ma from the Lab. is adopted by JVT to be a reference solution of rate control of H. 264/AVC. And also, a one-pass rate control scheme popularly employed in the current reference codec of H. 264/AVC is proposed by Zhengguo Li, Singapore and Siwei Ma.

In the beginning of standardization of SVC, the authors have provided the rate control scheme for JSVM, which is

published in VCIP2005. The algorithm of VCIP2005 is based on the traditional rate distortion (R-D) model, and with a layered bit allocation strategy which can brings much more coding performance. The algorithm in this paper makes a bigger improvement than the previous one, due to a set of scaling-factors employed for bit allocation. In the new proposed bit allocation, a set of weight parameters derived from the importance of each subband/picture for the whole coding performance is used to help bit allocation; moreover, based on the experiments, the authors propose a leveled R-D model for each hierarchical level of subbands. For low-pass subbands, the quadratic R-D model is employed, and for high-pass subband, the linear R-D model is used. From the experiments provided by the authors, it can be seen that the leveled R-D model can reflect the relationship between R and D more accurately.

This research is supported by National Natural Science Foundation of China under grant Nos. 4041003, 60333020, National Key Scientific and Technological Project under grant No. 2002BA906A35, National Basic Research Program of China (973 Program) under grant No. 2004AA119010.