

H.264/AVC 码率控制优化算法

袁 武¹⁾ 林守勋²⁾ 牛振东¹⁾ 罗海勇²⁾ 张勇东²⁾

¹⁾(北京理工大学计算机科学技术学院 北京 100081)

²⁾(中国科学院计算技术研究所 北京 100080)

摘 要 提出了一种新颖的编码特性预测机制,较为充分地利用了视频信源的时空相关性,改进了率失真建模的有效性;利用 Lagrangian 优化理论推导出两种率失真优化的位分配方案,并实现了相应的码率控制算法,即线性模型算法和二次模型算法.大量实验数据表明:线性模型算法和二次模型算法的编码效率基本上相同,而前者的码率控制能力稍优于后者;和 H.264/AVC 参考软件中所采用的 JVT G012 码率控制算法相比,两种新算法在获得更高编码效率的同时,能够更加准确地控制输出码率.

关键词 码率控制;位优化分配;H.264/AVC

中图法分类号 TP391

Efficient Rate Control Schemes for H.264/AVC

YUAN Wu¹⁾ LIN Shou-Xun²⁾ NIU Zhen-Dong¹⁾ LUO Hai-Yong²⁾ ZHANG Yong-Dong²⁾

¹⁾(School of Computer Science, Beijing Institute of Technology, Beijing 100081)

²⁾(Institute of Computing Technology, Chinese Academy of Sciences, Beijing 100080)

Abstract In this work, a novel coding characteristics prediction scheme is presented to improve R-D modeling, by exploiting spatio-temporal correlations. Two different approaches to the problem of optimum bit allocation at a macroblock-by-macroblock basis are achieved, one of which is developed on a modified MPEG-4 Q2 rate model and the other on a linear rate model. Extensive experiments show that the linear scheme is a bit more accurate than the quadratic one while they achieve similar coding performance. It's also shown that both the two schemes significantly exceed JVT G012, the current standardized RC scheme.

Keywords rate control; optimum bit allocation; H.264/AVC

1 引 言

码率控制技术是影响视频编码器编码效率的关键因素之一,其主要任务就是有效地控制视频编码器,使其输出码流的大小满足传输信道实际带宽的限制,并且,尽可能获得最优的解码图像.鉴

于码率控制技术的重要性,它一直是视频编码技术研究领域中的热点问题,目前比较有名的算法主要有 MPEG-2 TM-5^[1]、MPEG-4 VM8^[2]以及 H.263 TMN-8^[3]等.

一般而言,我们可以把码率控制算法的主要功能分解为两个步骤,第一步就是所谓的“位分配(bit allocation)”,即把有限的带宽分配到图像组(GOP)、

收稿日期:2006-07-20;最终修改稿收到日期:2007-11-27. 本课题得到国家“八六三”高技术研究发展计划项目基金(2007AA12Z239)、国家自然科学基金(40501057)、霍英东基金会优秀青年教师资助项目(91101)及教育部 2006 年新世纪优秀人才计划等项目资助. 袁 武,男,1974 年生,博士,主要研究方向为视频编码、网格计算、多媒体信息资源管理. E-mail: yuanwu@bit.edu.cn. 林守勋,男,1948 年生,博士,研究员,主要研究领域为视频编码、视频内容分析和检索、统计机器翻译. 牛振东,男,1968 年生,博士,教授,主要研究领域为多媒体信息资源管理、数字图书馆、数字化教育、复杂软件体系结构等. 罗海勇,男,1967 年生,博士研究生,主要研究方向为视频处理、普适计算、嵌入式系统. 张勇东,男,1973 年生,博士,副研究员,主要研究方向为图像和视频处理技术.

帧以及宏块等图像单位上;第二步就是计算量化参数(QP),以使实际输出码率和目标码率基本一致.码率控制优化算法的设计目标就是要实现位优化分配以及输出码率的精确控制,其关键在于建立精确的率失真数学模型^[4].

作为最新一代的视频编码技术国际标准,H.264/AVC以编码结构复杂和计算复杂度增加的代价,换取了比其他标准较高的编码效率^[5].在H.264/AVC编码器的参考软件实现中,采用了Lagrangian编码器控制技术对编码预测(帧内/间预测)进行优化计算,它需要量化参数作为控制参数^[6].这种要求对码率控制算法的设计施加了较大的影响,这是因为量化参数的计算依赖于率失真模型,而残余图像的统计信息是率失真建模的重要参数,其只能在编码预测之后才能获得.这就形成了在H.264/AVC技术研究中比较有名的因果矛盾问题.

目前,适用于H.264/AVC的码率控制算法主要有JVT-F086^[7]和JVT-G012^[8].在JVT-F086的控制过程中,可能需要进行迭代的编码处理,它首先利用预测技术选择量化参数进行编码处理,然后根据编码效果决定是否需要调整量化参数再次进行编码处理,其计算复杂度较高.而且,JVT-F086采用了比较简单的TM-5码率模型,对输出码率的控制存在较大的误差.

在JVT-G012中则采用了MPEG-4 Q2码率模型,可以根据信源特征及时地调整模型参数,预测误差较小.为了获取残余图像的编码统计信息,JVT-G012根据历史信息,利用线性MAD模型预测当前未编码图像块的编码复杂度.和JVT-F086相比,JVT-G012仅需要进行一次编码处理,计算复杂度较低;而且,其实验数据也进一步表明,JVT G012的编码效率优于JVT-F086.但是,JVT-G012尚存在一些不足:(1)未充分利用编码特性的时空相关性,率失真模型的预测误差较大;(2)在宏块之间,简单地按照编码复杂度的比例分配带宽,没有实现优化的位分配,编码性能还存在较大提升空间.

针对JVT-G012的这些不足,本文提出了一系列的改进措施.首先利用时空相关性改进率失真建模的准确性;然后,利用Lagrangian优化技术,针对二次码率模型和线性码率模型,分别推导出优化的位分配方案.实验数据充分表明,两种新算法的编码效率基本相同,而线性模型算法的码率控制能力稍强;和JVT-G012相比,两种新算法都具有更高的编码效率,对输出码率的控制更加精确.本文的部分成果已

经在国际期刊《IEEE Transaction on CSVT》^[9-10]上发表.

本文第2节介绍编码特性预测机制;在第3节建立编码特性的数学模型,包括头信息码率模型、编码复杂度模型、残余图像码率模型和图像失真模型;在第4节,根据Lagrangian优化技术,从理论上导出位优化分配算法;在第5节,简要地说明码率控制算法的实现过程;第6节对实验及实验结果进行分析;文章最后对全文进行总结.

2 编码特性预测机制

在VM-8以及JVT-G012等码率控制算法中,一般都采用相邻图像块率失真相关性较高的统计假设,并根据该假设,利用回归分析技术建立率失真模型,然后利用所得到的模型预测未编码图像块的率失真特性.该类算法的效能主要取决于率失真模型的有效“建立”及有效“使用”,也就是是否建立了比较精确的率失真模型,是否能够根据图像块的实际特性选择合适的模型进行预测.

回归分析技术是数学统计建模中更新模型参数常用的办法,其有效性依赖于样本数据的数量和质量,因此,利用回归技术建立率失真模型,其有效性依赖于样本数据选择方法.视频信源具有复杂的率失真统计特性:不同的图像块可能分属于率失真特性各异的不同视频对象;即使属于相同视频对象,在不同位置也可能表现出不同的运动特征,率失真特性的差异也可能较大.鉴于此,按照率失真相关性选择样本数据,尽可能地提高回归分析的有效性,是率失真精确建模的重要原则.

一般地,视频编码器按照自上而下、从左到右的光栅扫描顺序依次对每个宏块进行编码处理.在利用线性回归更新率失真模型参数时,JVT-G012、VM-8等码率控制算法按照编码处理时间的先后顺序选择样本数据,所选择的样本数据在图像帧上一般呈线状排列,如图1(a)所示.很显然,该种简单的样本数据选择方式没有很好地利用时空相关性,所获得的率失真模型存在比较大的预测误差.

经过深入分析,我们发现JVT-G012、VM-8等算法选择样本数据的方式主要存在3个问题:(1)只对空域相关性进行了利用,缺乏对时域相关性的利用;基本上只在同一帧图像内选择样本数据,没有利用前一帧或几帧的有用信息.而在视频信源中,经常可能出现时域相关性高而空域相关性较低

的情况。(2) 利用空间相关性的方法存在很大的缺陷:如图 1(a)所示,样本数据所对应的宏块在图像帧内呈线状排列,空域相关性较高的可能性较低;以上两点,导致了 JVT G012 算法率失真模型的预测误差较大。(3) 在算法中只利用了空域相关性,而空域相关性与位置关系紧密相关,一般而言,空间距离较远的图像块的相关性较低。因此,我们很难精确地预测和样本数据距离较远宏块的率失真信息,而在 JVT G012 中使用线性 MAD 模型预测时,没有考虑模型的适用性,对所有未编码图像块用相同模型参数进行预测,其预测误差是比较大的。

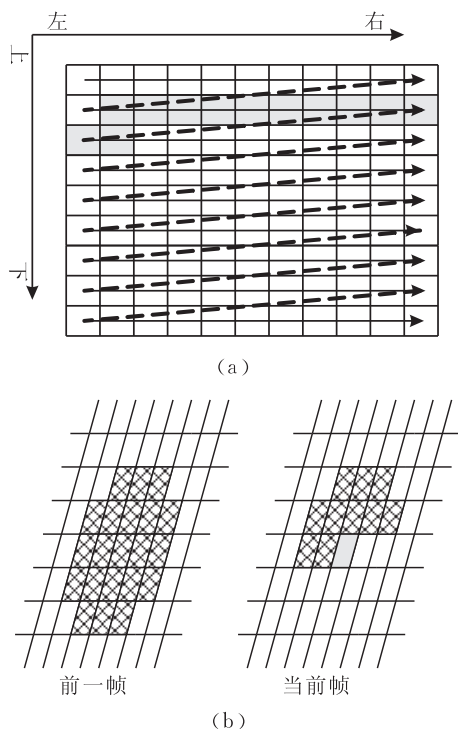


图 1 样本数据选择示意图

针对这些问题,我们提出了编码特性预测机制:(1) 联合使用时间相关性和空间相关性,从当前帧和同类型编码的前几帧中选择样本数据。(2) 按照时空距离远近选择样本数据。选择样本数据的理想方式,应该是首先按照时空相关性划分图像空间,然后根据时空相关性大小选择样本数据。由于时空相关性与视频信源的信号特征密切相关,但是编码特性具有非常复杂的统计特性,要精确地根据相关性划分图像空间非常困难。考虑到时空相关性与时间和空间距离密切相关,距离越近,相关性大的可能性越高,反之亦然。这个特点启发我们可以按照空间距离小于某个搜索阈值在当前帧和前几帧已编码图像数据中选择样本数据,即 $|\Delta x| + |\Delta y| \leq \text{SearchDiameter}$,其中 $|\Delta x|$, $|\Delta y|$ 分别表示样本宏块与当前宏块之间

的横向位置偏移及纵向位置偏移, SearchDiameter 表示搜索阈值。图 1(b)显示了预设阈值设为 2(本文实验中所采用的搜索阈值)时,样本数据在当前帧和前一帧的分布情况。(3) 在利用所获得的率失真模型进行预测时,不滥用模型,根据预测对象所处的时空位置选择合适的模型参数进行计算。通过上述改进,我们可以有效地改进率失真建模的有效性。在后续章节中,我们将针对不同的编码特性验证编码特性预测机制的有效性。

3 率失真建模

鉴于 H.264/AVC 采用了 Lagrangian 编码器控制技术、头信息所需占用的码率增加等特点,我们把率失真建模问题分解为 4 个子问题,包括头信息码率预测、编码复杂度预测、码率-量化步长模型和失真量化步长模型。

3.1 头信息码率预测

在视频编码输出码流中,除了残余图像 DCT 系数的压缩数据信息,还包括宏块模式 mode(或者宏块模式 mode、运动矢量 motion)和量化参数 QP 等头信息的压缩数据信息。当传输信道带宽比较低时,头信息所用比特数占总码流的比例比较大,此特点在 H.264/AVC 中尤为明显,这是因为它采用了多项复杂的编码技术,例如,可以使用多至 16 个的运动矢量(motion vector)预测一个宏块的运动特征,需要较多的比特用于头信息编码。和其他标准相比,H.264/AVC 在降低总码率的情况下,用于头信息编码的比特数不减反增,根据文献[11]中的实验统计,其在总码率中所占比例可达 30%~50%。由此可见,头信息码率预测的准确性对码率控制性能具有重要的影响作用。

文献[12-13]阐明了对头信息码率进行数学建模非常困难。在常见的码率控制算法中,普遍采用历史平均值估计头信息码率,它是一种比较粗糙的预测方法,已经不能满足 H.264/AVC 标准的需要。较精确地预测头信息码率,是在 H.264/AVC 标准中实现较为精确的码率控制所必需解决的一个重要问题。

文献[14]认为,在时域和空域相邻的图像宏块,其运动矢量之间存在相关性。对实验数据作进一步的分析,我们可以发现在时域和空域相邻的图像块间极有可能选择相似的编码模式。因此,在时域和空域相邻的图像块之间,头信息码率可能存在较强的

相关性. 下面, 我们设计了一个实验来验证时域相关性. 我们按照搜索半径所划定的时空范围内在前一帧搜索与目标宏块头信息码率相差最小的宏块. 在实验中, 搜索半径分别设为 1 和 0. 搜索半径设为 0 时, 在前一帧搜索和目标宏块位置相同的宏块; 搜索半径设为 1 时, 在前一帧搜索和目标宏块位置空间距离小于等于 1 的宏块. 实验结果如图 2 所示, 可看出, 头信息码率存在较强的时域相关性. 因此, 采用前一帧的历史信息预测头信息码率是可行的. 为简化搜索计算, 我们直接使用前一帧相同位置宏块的头信息码率预测当前宏块的头信息码率, 如下式所示:

$$H_{i,j}^k = H_{i,j}^{k-1} \tag{1}$$

在上式中, 上标 k 表示 P 帧序号, 下标 i, j 表示该宏块位于第 i 行第 j 列. 表示方法以下相同.

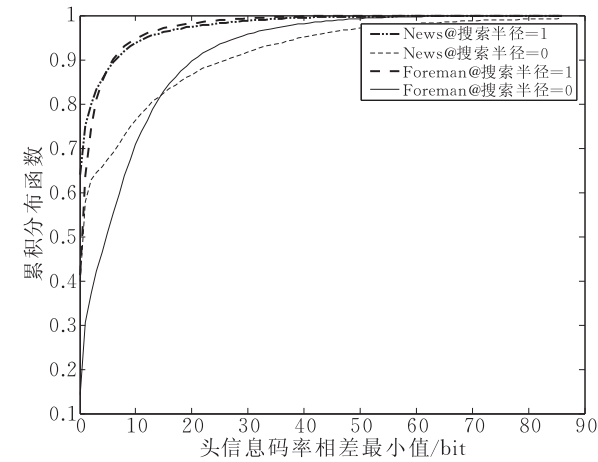


图 2 头信息码率空域相关性

为了验证新模型, 我们进行了大量实验. 在图 3 中, 显示了典型的部分实验结果. 可以看到, 新的方法明显地优于历史平均值预测方法.

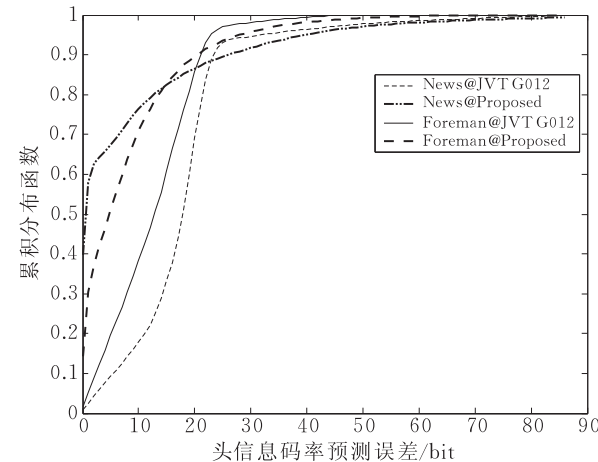


图 3 头信息码率预测误差

3.2 编码复杂度预测

目前, 在常用的率失真模型中, 普遍引入了编码复杂度, 它反映了信源的统计特性, 是位优化分配的重要参考量^[15]. 一般而言, 从概率统计角度, 我们可以使用残差图像的均方差 (σ 或 MSD) 或者标准误差 (MAD) 表示图像编码复杂度, 在 TMN-8 中, 采用 σ 表示图像编码复杂度, 在 TM-5 中用 MAD 表示图像编码复杂度.

在传统标准中, 视频编码器在帧内预测 (或者帧间预测) 之后, 根据所产生的图像预测残差的统计信息调整率失真模型, 然后计算量化参数; 但是 Lagrangian 编码器控制技术要求在对编码预测进行优化计算之前, 码率控制算法已经计算出量化参数. 可以采用逐步求精的办法, 通过多次编码处理获取编码复杂度信息, JVT F086 采用了该类思路.

考虑到视频图像内容的变化具有较强的时空连续性, 我们可以利用编码复杂度的历史信息及其变化趋势来预测图像编码复杂度. 一般而言, 在视频图像帧、块之间的差异主要是由视频对象相对镜头的移动而造成的. 这些移动可以分解成各方向上的平移, 可以用线性关系近似地表示编码复杂度的变化关系. 在 JVT G012 中就采用了线性 MAD 模型根据历史信息预测当前未编码图像块的编码复杂度, 公式如下所示.

$$MAD_{i,j}^k = \rho_{i,j}^k MAD_{i,j}^{k-1} + \gamma_{i,j}^k \tag{2}$$

在回归分析更新模型参数时, 采用编码特性预测机制改进样本数据的选择方式. 大量的实验表明, 编码特性预测机制对线性 MAD 模型的预测性能有所提高, 如图 4 所示, 其中编码复杂度预测误差 =

$$\left| \frac{MAD_{actual}}{MAD_{predicted}} - 1.0 \right|.$$

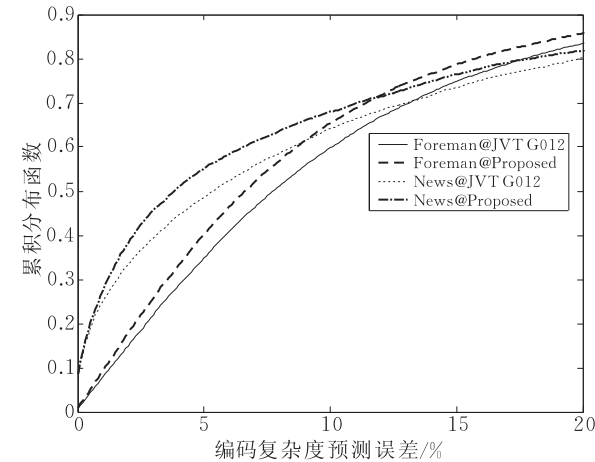


图 4 编码复杂度预测误差比较

3.3 残余图像码率模型

一般而言,视频信源在经过编码预测以及 DCT 变换之后,所得到的 DCT 系数的相关性很低,我们可以认为它们“近似”不相关,而且符合 Laplacian 概率分布函数.根据文献[16],可以按照以下公式计算 DCT 量化系数的信息熵.

$$H(\alpha) = -\left(1 - \exp\left(-\frac{2}{\alpha}\right)\right) \log_2\left(1 - \exp\left(-\frac{2}{\alpha}\right)\right) - 2 \sinh\left(\frac{2}{\alpha}\right) \left[\frac{\log_2\left(\sinh\left(\frac{2}{\alpha}\right)\right)}{\exp\left(\frac{1}{\alpha}\right) - 1} - \frac{\exp\left(\frac{1}{\alpha}\right)}{\alpha \ln 2 \left(\exp\left(\frac{1}{\alpha}\right) - 1\right)^2} \right] \quad (3)$$

其中 $\alpha = \frac{MAD}{Q}$, Q 表示量化步长, MAD 表示残差图像的绝对值平均误差.

$H(\alpha)$ 具有非常复杂的数学表达式,为了了解它的函数特性,我们绘制出它的抽样曲线,如图 5 所示.可看出, $H(\alpha)$ 是一个关于 α 单调递增的函数.在图 5 中,还绘制出线性拟和曲线以及二次拟和曲线.可看出,二次曲线拟和的精确度优于线性曲线拟和.

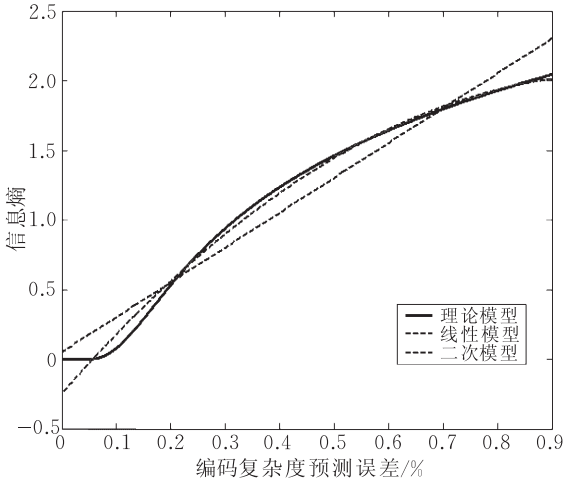


图 5 残余图像的码率信息熵曲线及其拟和

如果采用二次曲线拟和,我们可以获得著名的 MPEG-4 Q2 码率模型:

$$R = X_1 \frac{MAD}{Q} + X_2 \left(\frac{MAD}{Q}\right)^2 \quad (4)$$

可以利用线性回归技术获得模型参数 X_1 和 X_2 .在回归分析更新模型参数时,采用编码特性预测机制改进样本数据的选择方式.

为了满足率失真优化的位分配算法的需求,本文对它作了进一步的修改.众所周知,为了保证图像质量的连续性,量化参数在相邻帧相邻图像块之间变化不能太大.不妨假设量化参数只能在 $\{QP_{\min},$

$QP_{\max}\}$ 范围内取值,其中 QP_{ave} 是取值范围的中间值, Q_{\min} , Q_{\max} 和 Q_{ave} 是与之对应的量化步长.利用泰勒级数把 MPEG-4 Q2 码率模型在 Q_{ave} 展开,并舍去三阶以上高阶项,可以获得具有如下形式的残差图像码率模型:

$$\tilde{R}_{i,j}^K(Q) = A_{i,j}^K (Q_{i,j}^K)^2 + B_{i,j}^K Q_{i,j}^K + C_{i,j}^K \quad (5)$$

其中, $A_{i,j}^K = \frac{R''(Q_{\text{ave}})}{2}$, $B_{i,j}^K = R'(Q_{\text{ave}}) - R''(Q_{\text{ave}}) \times Q_{\text{ave}}$ 和 $C_{i,j}^K = R(Q_{\text{ave}}) - A_{i,j}^K \times Q_{\text{ave}}^2 - B_{i,j}^K \times Q_{\text{ave}}$.

如果 $Q_{i,j}^K$ 和 Q_{ave} 相差比较大,上式和 MPEG-4 Q2 存在较大的舍予误差,不过它对算法的有效性影响不大,这主要是因为 MPEG-4 Q2 和上式都是近似计算,而且, $Q_{i,j}^K$ 只能在 $\{QP_{\min}, QP_{\max}\}$ 区间内取值,这种限制进一步降低了模型误差.在后续章节,我们将看到实验结果也充分地验证了上式的有效性.

如果采用线性曲线拟和,我们可以得到

$$\tilde{R}_{i,j}^K = a_{i,j}^K \alpha_{i,j}^K + b_{i,j}^K = a_{i,j}^K \frac{MAD_{i,j}^K}{Q_{i,j}^K} + b_{i,j}^K \quad (6)$$

可以利用线性回归技术获得模型参数 $a_{i,j}^K$ 和 $b_{i,j}^K$.在回归分析更新模型参数时,采用编码特性预测机制改进样本数据的选择方式.

可看出,由于二次模型存在极点,二次模型不满足单调性,而且从式(4)到式(5)需要额外的泰勒公式展开;而线性模型比较简单,满足单调性.

结合头信息码率预测模型,我们可以得到以下两种码率模型:

$$\tilde{R}_{i,j}^K(Q) = A_{i,j}^K (Q_{i,j}^K)^2 + B_{i,j}^K Q_{i,j}^K + C_{i,j}^K + H_{i,j}^{K-1} \quad (7)$$

和

$$R_{i,j}^K(Q) = a_{i,j}^K \frac{MAD_{i,j}^K}{Q_{i,j}^N} + b_{i,j}^K + H_{i,j}^{K-1} \quad (8)$$

3.4 图像失真模型

原始图像与重建图像之间的图像损失,主要是由于对残余图像 DCT 系数量化引起的.一般地,可以采用原始图像与重建图像之间像素差值的方差 (MSD) 和峰值声噪比 ($PSNR$) 来评估图像失真.

在经典算法中,采用了方差作为图像失真的评估手段,假设图像失真符合平均分布概率,可以导出图像失真关于量化步长之间的函数关系为 $D = \frac{Q^2}{12}$.

该数学模型表明,图像失真仅仅与量化参数有关.然而,事实上,图像失真的大小由量化参数和 DCT 系数共同决定.

由于视频图像变化的时空连续性,图像失真在相邻图像帧之间的分布也具有一定的时空连续性.

鉴于此,并且考虑到失真与量化步长之间的正比关系,下面,我们提出一种启发式的失真模型.

$$D_{i,j}^K = \chi_{i,j}^K Q_{i,j}^K \quad (9)$$

$$\text{其中 } \chi_{i,j}^K = \frac{D_{i,j}^{K-1}}{Q_{i,j}^{K-1}}.$$

为了验证该公式的有效性,我们进行了大量的实验,其中部分的典型实验结果显示在图 6 中. 实验结果充分表明,利用新模型对图像失真进行预测具有一定的可行性.

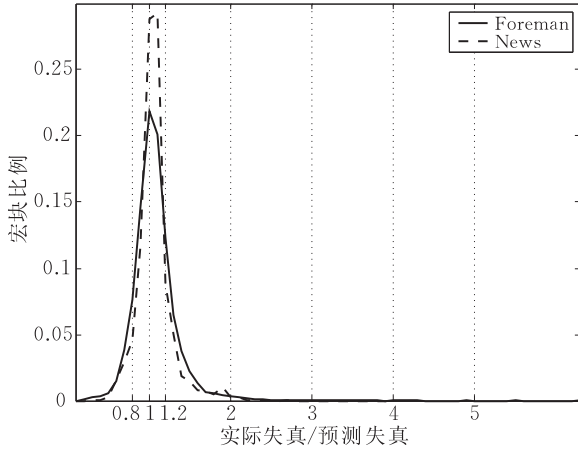


图 6 失真模型验证

4 率失真优化的位分配算法

根据上述率失真模型,我们利用 Lagrangian 优化技术推导率失真优化的位分配算法. 实质上,位优化分配算法就是要解决以下形式的数学问题:

$$\begin{aligned} \min_{Q_{1,1}^K, \dots, Q_{M,N}^K} & \sum_{i=1, j=1}^{M,N} D_{i,j}^K \\ \text{s. t.} & \sum_{i=1, j=1}^{M,N} (R_{i,j}^K) = T_{\text{frame}} \end{aligned} \quad (10)$$

对于二次码率模型和线性图像失真模型,位分配优化问题的求解也就是下式的求解:

$$\begin{aligned} \min_{Q_{1,1}^K, \dots, Q_{M,N}^K} & \sum_{i=1, j=1}^{M,N} \chi_{i,j}^K Q_{i,j}^K \\ \text{s. t.} & \sum_{i=1, j=1}^{M,N} (A_{i,j}^K (Q_{i,j}^K)^2 + B_{i,j}^K Q_{i,j}^K + C_{i,j}^K + H_{i,j}^{K-1}) = T_{\text{frame}} \end{aligned} \quad (11)$$

对于线性码率模型和二次图像失真模型,位分配优化问题的求解也就是下式的求解:

$$\min_{Q_{1,1}^K, \dots, Q_{M,N}^K} \sum_{i=1, j=1}^{M,N} \chi_{i,j}^K Q_{i,j}^K$$

$$\text{s. t.} \quad \sum_{i=1, j=1}^{M,N} \left(a_{i,j}^K \frac{MAD_{i,j}^K}{Q_{i,j}^K} + b_{i,j}^K + H_{i,j}^{K-1} \right) = T_{\text{frame}} \quad (12)$$

利用 Lagrangian 优化技术,可以把上述约束优化问题转化为非约束问题,分别表示为

$$\begin{aligned} \min_{Q_{1,1}^K, \dots, Q_{M,N}^K} & \sum_{i=1, j=1}^{M,N} \chi_{i,j}^K Q_{i,j}^K + \\ \lambda & \left(\sum_{i=1, j=1}^{M,N} (A_{i,j}^K (Q_{i,j}^K)^2 + B_{i,j}^K Q_{i,j}^K + C_{i,j}^K + H_{i,j}^{K-1}) - T_{\text{frame}} \right) \end{aligned} \quad (13)$$

$$\begin{aligned} \min_{Q_{1,1}^K, \dots, Q_{M,N}^K} & \sum_{i=1, j=1}^{M,N} \chi_{i,j}^K Q_{i,j}^K + \\ \lambda & \left(\sum_{i=1, j=1}^{M,N} \left(a_{i,j}^K \frac{MAD_{i,j}^K}{Q_{i,j}^K} + b_{i,j}^K + H_{i,j}^{K-1} \right) - T_{\text{frame}} \right) \end{aligned} \quad (14)$$

经过数学运算之后,我们可以分别得到下列两个公式:

$$\begin{aligned} Q_{i,j}^{K*} &= -\frac{B_{i,j}^K}{2A_{i,j}^K} - \frac{\chi_{i,j}^K}{A_{i,j}^K} \cdot \\ & \sqrt{\frac{T_{\text{frame}} - \sum_{m=i, n=j}^{M,N} H_{m,n}^{K-1} + \sum_{m=i, n=j}^{M,N} \frac{(B_{m,n}^K)^2}{4A_{m,n}^K} - \sum_{m=i, n=j}^{M,N} C_{m,n}^K}{\sum_{m=i, n=j}^{M,N} \frac{(\chi_{m,n}^K)^2}{A_{m,n}^K}}} \end{aligned} \quad (15)$$

$$Q_{i,j}^{K*} = \frac{\sum_{m=1, n=1}^{M,N} \sqrt{a_{m,n}^K MAD_{m,n}^K \chi_{m,n}^K}}{T_{\text{frame}} - \sum_{m=1, n=1}^{M,N} (b_{m,n}^K + H_{m,n}^{K-1})} \sqrt{\frac{a_{i,j}^K MAD_{i,j}^K}{\chi_{i,j}^K}} \quad (16)$$

对于式(15),需要对以下两种情况进行特殊处理:

(1) 如果最终计算得到的式(5)衰减为一次函数,需要进行升次操作,把其调整为二次函数.

(2) 如果 $T_{\text{frame}} - \sum_{m=i, n=j}^{M,N} H_{m,n}^{K-1} + \sum_{m=i, n=j}^{M,N} \frac{(B_{m,n}^K)^2}{4A_{m,n}^K} - \sum_{m=i, n=j}^{M,N} C_{m,n}^K < 0$, 则式(15)无实数解,说明目标码率太大或者过小.

5 码率控制算法

本文主要关注宏块级码率控制技术,在 GOP 及帧级上沿用 JVT-G012 码率控制算法的位分配策略. 对 GOP 及帧级位分配策略感兴趣的读者,可参阅文献[8]. 对于 I 帧和 GOP 中的第一个 P 帧,我

们采用固定的量化参数进行编码处理;对于其它 P 帧,则根据目标码率和视频信源的实际编码特性为每个宏块逐个地计算最优的量化参数.

二次算法和线性算法的处理流程基本相似,限于篇幅,下面主要介绍基于线性码率模型的码率控制算法,对二次算法感兴趣的读者可参阅文献[9-10]:

1. 初始化
令列标变量 $i=0$,行标变量 $j=0$.
2. 计算 $MB_{i,j}^K$ 的量化参数 $Q_{i,j}^{K*}$.
 - 2.1. 令变量 $ii=i$,变量 $jj=j$,变量 $ST=0$ 以及变量 $SU=0$;然后执行下列循环操作.
 - 2.2. 利用线性回归技术计算线性 MAD 模型参数 $\rho_{ii,jj}^K$ 和 $\gamma_{ii,jj}^K$,然后计算 $MAD_{ii,jj}^K = \rho_{ii,jj}^K MAD_{ii,jj}^{K-1} + \gamma_{ii,jj}^K$,如果 $MAD_{ii,jj}^K$ 小于零,把 $MAD_{ii,jj}^K$ 置为 0.
 - 2.3. 利用线性回归技术计算码率模型参数($a_{ii,jj}^K$ 和 $b_{ii,jj}^K$).
 - 2.4. 计算失真模型参数 $\chi_{ii,jj}^K$.
 - 2.5. 计算 $ST=ST+b_{ii,jj}^K + H_{ii,jj}^{K-1}$.
 - 2.6. 计算 $SU=SU+\sqrt{a_{ii,jj}^K MAD_{ii,jj}^K \chi_{ii,jj}^K}$.
 - 2.7. 循环条件:如果当前宏块有右邻宏块,计算 $ii=ii+1$;否则计算 $ii=0$ 和 $jj=jj+1$.如果已经处理完剩下的所有宏块,跳转到步 3,否则跳转到步 2.2.
3. 如果 $T_{\text{frame}}-ST$ 小于零,令 $Q_{i,j}^{K*}$ = 最近处理宏块的量化参数,然后跳转到步 6.
4. 否则计算 $Q_{i,j}^K = \frac{SU}{T_{\text{frame}}-ST} \sqrt{\frac{A_{i,j}^K MAD_{i,j}^K}{\chi_{i,j}^K}}$,然后根据 $Q_{i,j}^K$ 计算 QP .
5. 为了获得平滑的解码图像质量,调整 QP ,使 QP 和前一个宏块量化参数之间相差不超过 1.
6. 宏块编码.
7. 事后处理:
记录编码信息,包括头信息码率 $H_{i,j}^K$ 、失真模型系数 $\chi_{i,j}^K = \frac{D_{i,j}^K}{(actualQ_{i,j}^K)^2}$,码率模型回归分析样本数据 $\frac{MAD_{i,j}^K}{actualQ_{i,j}^K}$, $MAD_{i,j}^K$ 以及残差码率 $R_{i,j}^K$ (注意, $actualQ_{i,j}^K$ 对应于最后所采用的量化参数 QP).
如果 $MB_{i,j}^K$ 有右邻宏块,计算 $i=i+1$;否则计算 $i=0$ 和 $j=j+1$.如果当前帧所有宏块已经处理完毕,则转到下一帧的编码处理;否则转到步 2.

6 实验结果与分析

为了评估新算法的编码性能,我们在 JVT 参考软件 JM10.1^① 的基础上实现了二次模型算法和线性模型算法,然后对它们和 JVT G012 算法进行了大量的对比实验,测试范围涵盖了目前常用的视频序列和带宽条件.

限于文章篇幅,这里只给出了“Foreman”,“Paris”,“News”和“Silent”等 4 个典型视频序列的

部分实验数据^②,其中前两者为 CIF(4:2:0)格式,帧率为 30 帧/s;后两者为 QCIF(4:2:0)格式,“News”的帧率为 15 帧/s,“Silent”的帧率为 10 帧/s.对 I 帧和第一个 P 帧使用固定的初始量化参数进行编码处理,初始量化参数分别设置为 24,26,28,⋯,40.我们首先使用各个初始量化参数对序列进行量化参数固定的编码处理,所得到的输出码率作为后续实验的目标码率.在表 1 中给出了其他的主要实验参数设置.

表 1 实验参数设置

参数	值	参数	值
运动估计精度	1/4 pel	参考帧	5
Hadamard 变换	ON	熵编码模式	CABAC
率失真优化	ON	GOP 编码结构	IPPP
运动估计搜索范围	±32	IntraPeriod	0
Restrict Search Range	2	编码帧数	100

码率控制算法的性能主要从两个方面进行评估,包括图像质量损失的程度以及控制输出码率的能力.

码率控制算法的首要任务是控制输出码流的大小,以满足带宽条件的限制.码率预测误差(Rate Prediction Error, RPE)是评价码率控制能力的常用方式,它的计算如下所示. $RPE = |(R_a - T_p)/T_p| \times 100\%$,其中 R_a 为图像帧的实际输出码率, T_p 为它的目标码率.如果码率预测误差越小,说明码率控制算法对输出码率的控制能力越强,反之亦然.

图 7 显示了各个码率控制算法在不同带宽条件下码率控制能力的曲线图,其中,码率预测平均

误差是 $(\sum_{i=2}^{99} RPE_i)/98$.可明显地看出,新算法的码率预测平均误差比较低,并且在多数情况下,JVT G012 的预测平均误差比新算法高出很多,这说明新算法的码率控制能力强于 JVT G012;新算法码率预测平均误差的变化比较平缓,JVT G012 的预测平均误差变化比较大,特别在低码率的情况下,曲线变化陡峭,这说明新算法能够更好地适应不同的带宽条件.图 8 显示了初始量化参数设为 32 时,“Foreman”和“News”的码率预测误差变化情况.这些实验数据充分说明新算法对码率控制的能力优于 JVT-G012.在对实验数据的进一步分析过中,我们也可以发现,

① JM10.1. http://iphome.hhi.de/suehring/tml/download/old_jm/jm10.1.zip.
② 更多的实验数据可参见 JVT-O016tests.doc,该文档位于 JVT 组织的 ftp 网站“http://ftp3.itu.ch/av-arch/jvt-site/2005_04_Busan”上.

线性模型算法的码率控制能力好于二次模型算法,部分原因在于后者建立码率模型时存在模型误差——舍弃 MPEG-4 Q2 码率模型泰勒展开式高阶项而造成的;再加之,二次模型存在极点,在计算量化参数时不能保证码率模型的单调性,这也降低了码率控制计算的有效性.

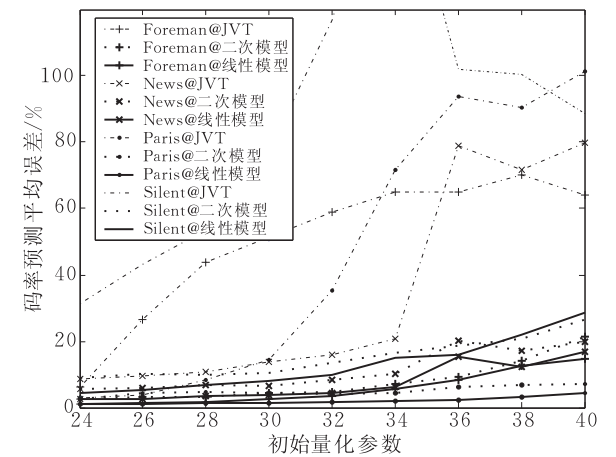
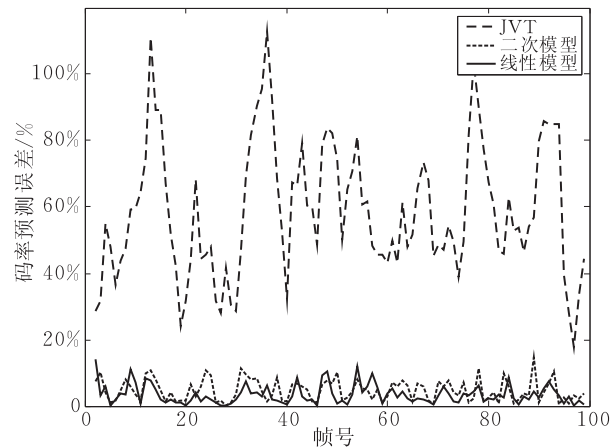
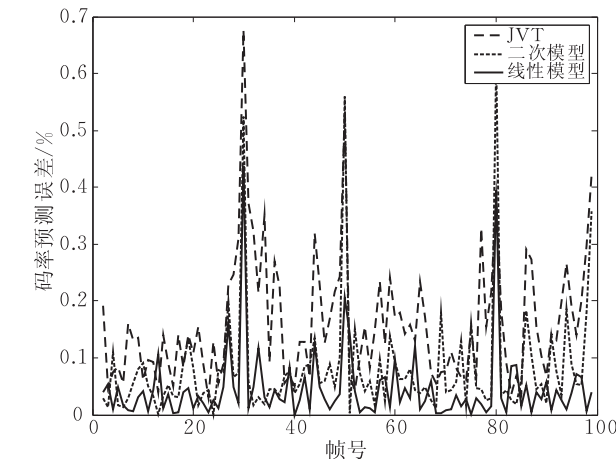


图 7 GOP 级码率预测平均误差对比



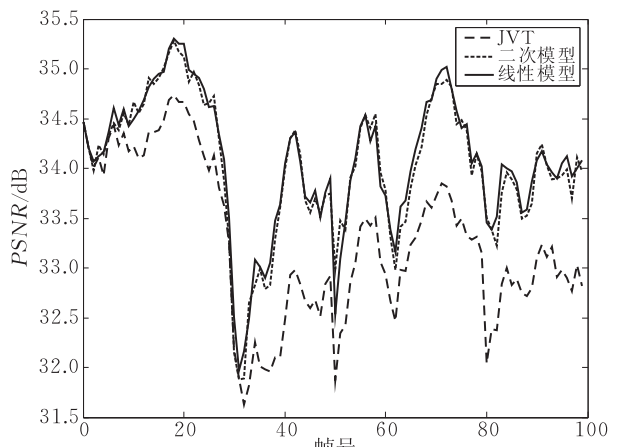
(a) Foreman, CIF, 30Hz, 初始量化参数=32



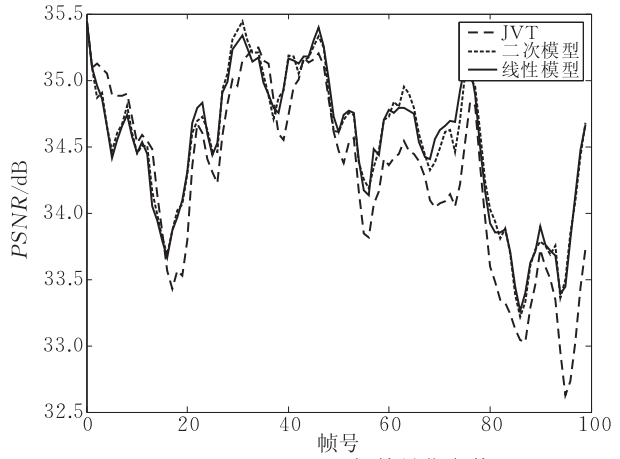
(b) News, QCIF, 3.33 Hz, 初始化参数=32

图 8 帧级码率预测平均误差比较

另一个评价码率控制算法性能的重要因素是图像失真. 图像编码损失的主要原因是 DCT 系数量化产生的, 只有有效地分配量化参数, 才能获得质量优良的视频图像. *PSNR* 是常用的评估图像失真的统计方式. 我们在图 10 中绘出各个测试序列的码率-*PSNR* 曲线图, 相关具体数据参见表 2. 可看出, 新算法的曲线基本上都位于 JVT-G012 的上方. 图 9 显示了初始量化参数设为 32 时, “Foreman”和“News”的 *PSNR* 变化情况. 从表 2 中可以看到, 相对于 JVT G012, 二次模型算法的 *PSNR* 增益最高可达 0.97dB, 而线性模型算法的 *PSNR* 增益最高可达 1.08dB; “Foreman”的二次模型算法平均增益是 0.18dB, 它的线性模型平均增益是 0.18dB; “Paris”的二次模型算法平均增益是 0.36dB, 它的线性模型平均增益是 0.19dB; “News”的二次模型算法平均增益是 0.69dB, 它的线性模型平均增益是 0.76dB; “Silent”的二次模型算法平均增益是 0.19dB, 它的线性模型平均增益是 0.15dB. 这些实验数据充分说明新算法的编码效率优于 JVT-G012.



(a) News, QCIF, 3.33 Hz, 初始量化参数=32



(b) Foreman, CIF, 30Hz, 初始量化参数=32

图 9 帧级 PSNR 变化比较

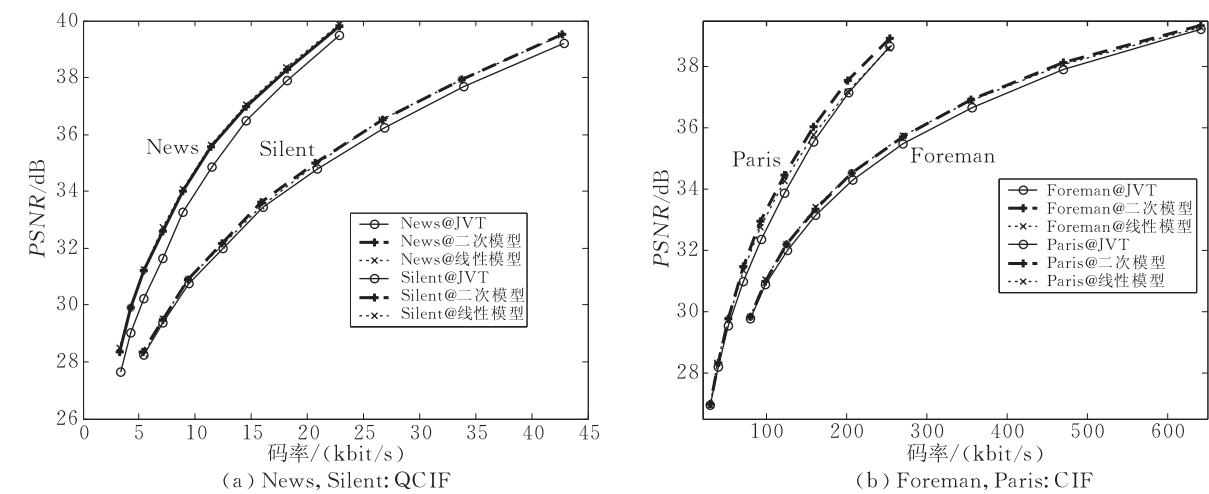


图 10 编码效率比较

表 2 实验数据统计表

初始 量化 参数	Foreman								News							
	目标 码率	JVT		二次模型		线性模型		目标 码率	JVT		二次模型		线性模型		目标 码率	目标 码率
		PSNR	码率	PSNR	码率	PSNR	码率		PSNR	码率	PSNR	码率	PSNR	码率		
24	641.42	39.21	641.76	39.35	641.71	39.3	641.56	22.84	39.5	22.87	39.79	22.87	39.88	22.85		
26	469.54	37.92	470.78	38.13	469.73	38.08	469.67	18.15	37.91	18.2	38.29	18.18	38.37	18.16		
28	354.63	36.67	355.91	36.93	354.86	36.9	354.77	14.51	36.5	14.55	36.99	14.53	37.06	14.52		
30	269.20	35.49	270.4	35.71	269.38	35.74	269.31	11.43	34.86	11.46	35.55	11.45	35.65	11.44		
32	206.20	34.3	207.07	34.52	206.39	34.53	206.32	8.88	33.27	8.92	34.02	8.92	34.07	8.89		
34	160.82	33.15	161.45	33.34	160.92	33.4	160.92	7.06	31.64	7.1	32.6	7.08	32.72	7.07		
36	124.85	32	125.38	32.18	125.04	32.19	124.94	5.42	30.24	5.45	31.21	5.44	31.26	5.43		
38	98.51	30.88	98.9	30.98	98.66	31.06	98.67	4.26	29.01	4.29	29.92	4.27	29.92	4.27		
40	78.88	29.78	79.19	29.84	79.01	29.82	78.99	3.30	27.66	3.33	28.36	3.31	28.5	3.31		

初始 量化 参数	Paris								Silent							
	目标 码率	JVT		二次模型		线性模型		目标 码率	JVT		二次模型		线性模型		目标 码率	目标 码率
		PSNR	码率	PSNR	码率	PSNR	码率		PSNR	码率	PSNR	码率	PSNR	码率		
24	252.59	38.66	253.15	38.91	252.98	38.6	252.68	42.89	39.21	42.89	39.51	42.69	39.47	42.65		
26	201.14	37.15	201.66	37.55	201.47	37.19	201.24	33.99	37.68	33.99	37.94	33.81	37.94	33.77		
28	158.21	35.55	158.96	36.05	158.46	35.77	158.28	26.86	36.24	26.86	36.52	26.72	36.51	26.66		
30	121.58	33.87	122.44	34.47	121.77	34.26	121.65	20.93	34.78	20.93	34.99	20.74	34.93	20.72		
32	92.16	32.37	93.02	32.96	92.27	32.77	92.23	16.05	33.45	16.05	33.62	15.91	33.49	15.89		
34	70.07	30.98	70.79	31.47	70.17	31.35	70.12	12.44	32	12.44	32.18	12.36	32.1	12.36		
36	51.74	29.53	52.38	29.77	51.8	29.75	51.79	9.40	30.75	9.4	30.88	9.34	30.88	9.33		
38	38.67	28.2	39.01	28.32	38.8	28.31	38.72	7.12	29.38	7.12	29.5	7.09	29.52	7.09		
40	29.23	26.94	29.55	26.98	29.28	26.95	29.3	5.38	28.26	5.38	28.34	5.36	28.28	5.37		

码效率和码率控制能力上,两种新算法均明显优于 JVT-G012;而且,两种新算法具有相近的编码效率,而基于线性模型的算法控制码流大小的能力更强。

7 结束语

码率控制算法的好坏是视频编码器性能高低的关键因素之一。本文针对 JVT G012 码率控制算法的不足之处,先提出了一个新颖的编码特性预测机制,改进了编码特性相关性的利用方式,有效地减少了率失真模型的预测误差;根据编码特性预测机制提出了新的头信息码率预测方法和图像失真模型;在 MPEG-4 Q2 码率模型和线性码率模型的基础上,利用 Lagrangian 优化技术分别推导出两种高效的位率失真优化分配算法。实验数据充分表明,在编

参 考 文 献

[1] MPEG-2, Test Model 5, Doc. ISO/IEC JTC1/SC29 WG11/93-225b. Test Model Editing Committee, Apr. 1993

[2] Video Group. Text of ISO/IEC 14496-2 MPEG-4 Video VM-Version 8.0. ISO/IEC JTC1/SC29/WG11 Coding of Moving Pictures and Associated Audio MPEG 97/W1796, Stockholm, Sweden, July 1997

[3] Ribas-Corbera J, Lei S. Rate control in DCT video coding

for low-delay communications. *IEEE Transactions on Circuits and System Video Technology*, 1999, 9(1): 172-185

- [4] He Z, Mitra S K. Optimum bit allocation and accurate rate control for video coding via ρ domain source modeling. *IEEE Transactions on Circuits and Systems for Video Technology*, 2002, 12(10): 840-849
- [5] Wiegand T, Sullivan G J. Draft ITU-T recommendation H.264 and final draft international standard of joint video specification (ITU-T Recommendation H.264 | ISO/IEC 14496-10 AVC). Joint Video Team of ISO/IEC JTC1/SC29/WG11 and ITU-T SG16/Q. 6 Doc. JVT-G050, Pattaya, Thailand, Mar. 2003
- [6] Sullivan G J, Wiegand T. Rate-distortion optimization for video compression. *IEEE Signal Processing Magazine*, 1998, 15(6): 74-90
- [7] Ma S W, Gao W, Lu Y, Lu H Q. Proposed draft description of rate control on JVT standard. JVT-F086, 6th Meeting. Awaji, Japan, 5~13 December, 2002
- [8] Li Z G, Pan F, Lim K P, Feng G N, Lin X, Rahardaj S. Adaptive basic unit layer rate control for JVT. JVT-G012, 7th Meeting. Pattaya II, Thailand, 2~14 March, 2003
- [9] Yuan W, Lin S X, Zhang Y D, Yuan W, Luo H Y. Optimum bit allocation and rate control for H.264/AVC. *IEEE Transactions on Circuits and System for Video Technology*, 2006, 16(6): 705-715
- [10] Yuan W, Lin S X, Zhang Y D, Yuan W, Luo H Y. Opti-

mum bit allocation and rate control for H.264/AVC. JVT-O016, 15th Meeting. Busan, KR, 16~22 April, 2005

- [11] Zhang J, Yi X, Ling N, Shang W. Context adaptive lagrange multiplier (CALM) for motion estimation in JM. JVT-S028, 19th Meeting. Geneva, Switzerland, 31 March ~ 7 April, 2006
- [12] Ramchandran K, Ortega A, Vetterli M. Bit allocation for dependent quantization with applications to multiresolution and MPEG video coders. *IEEE Transactions on Image Processing*, 1994, 3(5): 533-545
- [13] Ortega A, Ramchandran K. Forward-adaptive quantization with optimal overhead cost for image and video coding with applications to MPEG video coders//*Proceedings of the IS&T/SPIE Dig. Video Comp. : Alg. & Tech.*. San Jose, CA, 1995: 129-138
- [14] Chalidabhongse J, Jay Kuo C-C. Fast motion vector estimation using multiresolution-spatio-temporal correlations. *IEEE Transactions on Circuits and Systems for Video Technology*, 1997, 7(3): 477-488
- [15] Lee H J, Chiang T H, Zhang Y Q. Scalable rate control for MPEG-4 video. *IEEE Transactions on Circuits and Systems for Video Technology*, 2000, 10(6): 878-894
- [16] Moscheni F, Dufaux F, Nicolas H. Entropy criterion for optimal bit allocation between motion and prediction error information//*Proceedings of the SPIE Visual Commun. and Image Proc.*, 1993: 1548-1557



YUAN Wu, born in 1974, Ph. D. . His research interests include video coding, grid computing.

LIN Shou-Xun, born in 1948, Ph. D. , professor. His research interests include video coding, video content analysis and retrieval, statistic machine translation.

Background

Rate control (RC) is one of the key problems in regard of coding performance, and has drawn significant research attention. For rate control, its main task is to regulate DCT coefficients quantization to adapt to the actual channel bandwidth, and meanwhile achieve high-quality reconstructed pictures at the decoder. Several RC schemes have been reported in the literature, including TMN-8, TM-5 and VM-8 etc.

Generally speaking, a typical RC scheme can be decomposed into two steps, which are bit allocation and subsequent

NIU Zheng-Dong, born in 1968, Ph. D. , professor. His research interests include multimedia information management, digital library, digital education and software architecture.

LUO Hai-Yong, born in 1967, Ph. D. candidate. His research interests include video processing, pervasive computing and embedded systems.

ZHANG Yong-Dong, born in 1973, Ph. D. , associate professor. His research interests include image processing and video processing.

achievement of the bits target. The insight behind optimal bit allocation and accurate achievement is to precisely approximate rate-distortion (R-D) behaviors of video content via mathematical modeling. Because video coding is a highly non-linear process, it is a difficult task to precisely approximate R-D behaviors by using a close-form formula, and an empirical approach is most often used indeed. Based on R-D models, Lagrangian optimization or linear programming methods are employed to achieve optimum bit allocation.

As a new generation of video coding standards, H.264/AVC greatly outperforms prior coding standards in coding performance by exploiting lots of complicated coding methods. As a side-effect, H.264/AVC gets highly complicated. The introduction of Lagrangian coder control method into a H.264/AVC-complained coder greatly challenges the design of RC schemes. The Lagrangian coder control method demands quantization parameter as a prerequisite to the RDO for inter/intra prediction, and consequently couples inter/intra prediction and RC much tightly. Since accurate R-D modeling on the residual can only be conducted after inter/intra prediction, the peculiarity of Lagrangian coder control method leads to the chicken and egg dilemma. Thus, the design of RC scheme for H.264/AVC is quite different from prior standards, and conventional RC scheme cannot be directly employed in a H.264/AVC-complained coder.

Several works related to RC have contributed to H.264/

AVC, such as JVT-F086, JVT-G012 and JVT-O016. In JVT-F086, a multiple-pass scheme is presented to circumvent the dilemma, while in JVT-G012 and JVT-O016, a linear MAD model is used indeed. In JVT-G012, the conventional MPEG-4 Q2 model is employed to calculate the quantization parameter. Compared with JVT-G012, JVT-O016 improves the accuracy of MPEG-4 Q2 and MAD models by using an effective coding characteristics prediction scheme. Moreover, JVT-O016 achieves optimum bit allocation at a macroblock-by-macroblock basis in a TMN-8-alike way, and therefore significantly improves the coding performance.

In this paper, an approach similar to JVT-O016 is presented with significant difference in using a linear rate model instead of a quadratic rate model and a quadratic distortion model instead of a linear distortion model. Based on the R-D models, we develop a close-form formula to the problem of optimum bit allocation.