

基于标准化量化步长 θ 的数据率模型

杨天武 彭强 诸昌铃

(西南交通大学信息科学与技术学院, 成都 610031)

摘要 为了更精确地进行视频实时流控,通过分析视频编码的数据率特征,提出了一个新颖的基于标准化量化步长 θ 的数据率模型。在对离线实验获得的 θ 数据率曲线进行平滑处理后,即可创建一个数据率查找表,然后再基于查找表设计出简单高效的实时 θ 模型流控算法。因为 θ 模型估计精确,且无任何自适应参数,所以 θ 模型流控受视频场景切换影响很小。实验结果显示, θ 模型流控的性能明显优于TMN8流控。与同样是基于精确数据率模型的 ρ 域流控相比,虽性能相当,但是 θ 模型流控不仅计算复杂度更低、更灵活,且更易于在硬件中实现。

关键词 视频编码 数据率模型 量化步长 流控

中图分类号: TN919.8 **文献标识码:** A **文章编号:** 1006-8961(2007)05-0841-06

Rate Model Based on Standard Quantization Step Size

YANG Tian-wu, PENG Qiang, ZHU Chang-qian

(School of Information Science & Technology, Southwest Jiaotong University, Chengdu 610031)

Abstract A novel rate model of standard quantization step size (θ) for real-time rate control is proposed based on the analysis of rate characteristic in video coding. By smoothing the θ rate curve obtained offline, a look-up table of coding rate is created first. A simple and effective rate control algorithm for real-time video is developed based on the look-up table. Because the θ rate model is an accurate model without any adaptive parameter, the θ model rate control will not suffer from the impact of scene change. Our experiments show that the performance of θ model rate control is much better than the TMN8 rate control. Comparing with the ρ domain rate control which also adopts an accurate coding rate model, the θ model rate control has lower computational complexity and offers significant advantages in hardware implementation.

Keywords video coding, coding rate model, quantization step size, rate control

1 引言

各种不同的视频业务,例如可视电话和视频监控等,由于受到网络带宽和编解码器缓存大小的限制,因此需要控制编码比特数。视频流控旨在带宽和缓存受限的条件下最大限度地提高视频编码质量。

流控算法的好坏直接影响视频质量。一些流控算法虽可通过剩余比特数、缓存占用率等参数来控制编码比特数,但经常无法精确控制,由此就引起了

跳帧和带宽闲置的问题。为了改善流控性能,现已相继提出了一些基于实验数据率模型的流控算法^[1-4]。这些模型以不同函数形式表示了数据率随量化步长 Q 和失真度而变化的关系,简称 Q 模型。一些 Q 模型^[1,2]由于计算很复杂,需要通过多次编码来选择最佳参数,因此不适合实时流控。另一些 Q 模型^[3,4]本身虽并不十分精确,但利用视频间的特征相似性来动态地调整模型参数,就可在很大程度上弥补模型的不足,它们不仅能在更精确地控制编码比特数的同时,大幅度提升流控性能,并且算法简单,其中一些流控算法已被国际视频标准采纳,例

基金项目:国家自然科学基金项目(60672099)

收稿日期:2005-10-14;改回日期:2006-04-10

第一作者简介:杨天武(1971~),男。1993年获甘肃农业大学农学学士学位,现为西南交通大学交通信息工程及控制专业博士研究生。主要研究方向为视频图像压缩编码及传输控制。E-mail:yskyw@tom.com

如 MPEG-2 的 TM5、MPEG-4 的 VM8 和 H. 263 + 的 TMN8, 但由于这些算法模型本身不够精确, 当运动剧烈或场景切换时, 因模型参数变化剧烈而难以动态适应, 从而导致流控性能下降。文献[5~7]提出了 ρ 域数据率模型, 该模型虽然精确性很高, ρ 域流控性能也很理想, 但是由于 ρ 域流控要求在实际编码前确定所有宏块的编码模式, 从而使其在 H. 264/AVC 等具有帧内块间预测的编码器中的应用受到某些限制。本文提出了一种标准化量化步长 θ 的数据率模型, 由于该模型是无任何动态适应参数的固定模型, 且具有很高的精确性和稳定性, 因此受场景切换影响很小。利用该模型, 通过建立 $1/\theta$ 到数据率 R 的快速查找表, 即可设计出简单高效的流控算法。 θ 模型流控的计算复杂度仅为宏块级, 其不仅算法流程简单高效, 且非常易于硬件实现。

2 基于标准化量化步长 θ 的数据率模型

2.1 θ 模型的理论分析

众所周知, 视频编码的变换系数通常服从拉普拉斯分布 (Laplacian distribution)^[8], 其概率密度函数表示为

$$p(x) = ke^{-(\lambda|x|)^{\alpha}} \quad (1)$$

由于绝大多数情况下, α 近似为 1, 因此概率密度函数可简化为

$$p(x) = \frac{\lambda}{2} e^{-\lambda|x|} \quad (2)$$

定义失真度为 $D(x) = |x - \hat{x}|$, 其中 x 和 \hat{x} 分别为原始和重建的变换系数。如果允许失真度为 D , 那么根据香农 (Shannon) 信息论, 表示系数所需的最小数据率为

$$R(D) = \text{lb}\left(\frac{1}{\lambda D}\right) \quad (3)$$

考虑典型的均一量化方式, 即宏块内所有变换系数的量化步长相同。假定量化步长为 Q ; w 为确定量化系数重建值位置的常数, H. 263 编码器的 w 值为 0.25; d 为零量化区域 (dead zone) 大小, H. 263 编码器 P 宏块的 d 为 1.25, 那么根据定义, 失真度可表达为

$$\begin{aligned} D(Q) &= 2 \int_0^{dQ} xp(x) dx + 2 \sum_{i=0}^{\infty} \int_{(i+d)Q}^{(i+d+1)Q} |x - (i+d+w)Q| p(x) dx \\ &= \int_0^{dQ} \lambda x e^{-\lambda x} dx + \sum_{i=0}^{\infty} \int_{(i+d)Q}^{(i+d+1)Q} \lambda |x - (i+d+w)Q| e^{-\lambda x} dx \end{aligned}$$

$$= \frac{1}{\lambda} + e^{-\lambda dQ} \left(Q \left(\frac{(w-1)e^{-\lambda Q}}{1-e^{-\lambda Q}} - d \right) + \frac{2}{\lambda} \frac{e^{-\lambda Q} - 1}{1-e^{-\lambda Q}} \right) \quad (4)$$

将式(4)带入式(3), 得到的数据率表达式为

$$R(Q) = \text{lb} \left(\left(\left(1 + e^{-\lambda dQ} \left(\lambda Q \left(\frac{(w-1)e^{-\lambda Q}}{1-e^{-\lambda Q}} - d \right) + \frac{2}{\lambda} \frac{e^{-\lambda Q} - 1}{1-e^{-\lambda Q}} \right) \right)^{-1} \right) \right) \quad (5)$$

令 $\theta = \lambda Q$, 数据率可表示为 θ 的函数, 即

$$R(\theta) = \text{lb} \left(\left(\left(1 + e^{-\theta d} \left(\theta \left(\frac{(w-1)e^{-\theta}}{1-e^{-\theta}} - d \right) + \frac{2}{\lambda} \frac{e^{-\theta} - 1}{1-e^{-\theta}} \right) \right)^{-1} \right) \right) \quad (6)$$

值得注意的是, 式(6)中 w 和 d 是与编码器相关的常数, 只有 θ 为变量, 即数据率 R 为 θ 的函数。视频变换编码基本上都采用正交变换, 例如 DCT (discrete cosine transform) 等。因此视频残差帧的像素值方差等于变换系数的方差。大多数情况下, 运动估计的块匹配原则为最小平均绝对差值 s , 即

$$s = 1/\lambda \Rightarrow \theta = \lambda Q = Q/s \quad (7)$$

如果用以下最小均方误差 σ^2 作为块匹配原则

$$\sigma^2 = 2/\lambda^2 \Rightarrow \theta = \lambda Q = \sqrt{2/\sigma^2} Q = \sqrt{2} Q/\sigma \quad (8)$$

这里将 θ 称作为标准化量化步长。若将各宏块的像素值除以该宏块的 s , 则各宏块的数学期望标准化为 1, 若再用标准化步长 θ 来量化标准化后的各宏块, 则该量化值就等于原变换系数用量化步长 Q 进行量化的量化值。这就消除了因各宏块的数学期望变化而造成的对数据率估计精度的影响, 因此可能在视频剧烈运动和镜头切换时精确估计数据率。

2.2 平滑数据率曲线生成

式(6)给出了当视频源服从拉普拉斯分布时的数据率的理论下线, 但由于受到实际压缩编码方法的影响, 因此编码性能可能达不到理论下线。但是各种实际的编码器的数据率曲线仍然保持与理论曲线相似的特征, 例如对 H. 263 编码器, 当式(6)中的对数变为 1.7 时, 则数据率的理论曲线就接近实验曲线。由于不同编码器的编码特征不同, 数据率曲线也有所差异, 因此很难用简单公式表达, 这正是存在大量经典数据率的实验公式, 且各不相同的原因。但是经大量实验证明, 对于特定编码器, 由于根据 θ 确定的数据率曲线相当稳定, 因此本文采用离线实验方式来生成特定编码器数据率曲线。

数据率 R 是 θ 的函数, 即也是 $1/\theta$ 的函数。因为容易确定 $1/\theta$ 到数据率 R 的一一对应查找表, 所以实际应用中采用 R 随 $1/\theta$ 而变化的数据率变化曲线。图 1 显示了“Carphone”序列在不同编码器中的数据率曲线。

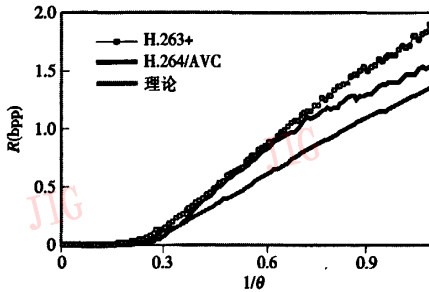


图 1 “Carphone”序列在不同编码器中的数据率曲线
Fig. 1 Rate curves in different Encoders for “Carphone”

实际编码中, 由于 $1/\theta$ 大于 6.0 的情况极少, 且 $1/\theta$ 大于 6.0 时, 数据率非常相近, 因此限定 $1/\theta$ 的变动范围为 $[0, 6.0]$ 。本文将该区间合理地分成宽为 0.01 的 600 个小区间, 并认为小区内数据率相同。对大量视频序列采用不同参数进行编码, 同时计算每个宏块的 $1/\theta$ 值, 并记录产生的比特率, 最后求每个 $1/\theta$ 区间内的所有样本的数据率均值。数据率曲线应为单调增函数。当 $1/\theta$ 较大时, 小区内样本通常为极少几个, 特别当 $1/\theta$ 大于 1 时, 甚至很多小区间没有样本, 由于这会引起数据率曲线波动, 因此需要平滑数据率曲线。当 $1/\theta$ 等于 γ 时, 平滑后的数据率 \bar{R}_γ 可用区间 $[\gamma - N \times 0.01, \gamma + N \times 0.01]$ 的实验数据得到:

$$\bar{R}_\gamma = \bar{R} - \bar{i} \frac{\sum_{i=-N}^N ((0.01i + \gamma) K_{\gamma+0.01i} (R_{\gamma+0.01i} - \bar{R}))}{\sum_{i=-N}^N ((0.01i + \gamma)^2 \times K_{\gamma+0.01i})} \quad (9)$$

其中,

$$\bar{i} = \frac{\sum_{i=-N}^N (0.01i + \gamma) K_{\gamma+0.01i}}{\sum_{i=-N}^N K_{\gamma+0.01i}}$$

$$\bar{R} = \frac{\sum_{i=-N}^N R_{\gamma+0.01i} K_{\gamma+0.01i}}{\sum_{i=-N}^N K_{\gamma+0.01i}}$$

其中, N 为整数, 指用作平滑 \bar{R}_γ 的数据的变化范围,

N 点应随 $1/\theta$ 的增大而增大。 \bar{R}_γ 为 $1/\theta$ 等于 γ 时的实验数据率均值, K_γ 为 $1/\theta$ 位于小区间 $[\gamma - 0.005, \gamma + 0.005]$ 时的样本数。

3 基于 θ 数据率模型的实时流控

流控技术可以在帧级和宏块级进行, 其中帧级流控对所有视频编码系统都是必不可少的。对于视频点播系统等对时延要求不严格的应用而言, 可采用基于率失真优化的帧级流控来获得理想的视频质量。对实时交互式视频, 例如可视电话和网络会议等, 通常帧级流控是简单地将缓存保持在一个较低水平上, 同时为每帧分配基本相等的比特数, 但是要求严格的缓存管理和极小的系统时延, 且需要进行宏块级流控, 而精确的宏块级流控则很困难, 因为其要求有精确的数据率模型。实验证实, 由于宏块的编码数据率可通过其 $1/\theta$ 精确估计, 因此利用数据率查找表可以实现非常精确的宏块级流控。具体算法步骤如下:

(1) 初始化 在编码第 1 帧前, 初始化数据率查找表。在编码每帧首宏块前, 设置量化步长 Q 等于前一帧首宏块的量化步长。令量化步长调整量 ε_Q 等于 0, 则估计编码 DCT 系数的可用比特数 B_A (下角 A 代表 available), 就等于编码当前帧的可用比特数减去编码运动向量及宏块模式等的比特数。

(2) 确定基准量化步长 利用数据率查找表来估计所有剩余宏块用 Q 量化时的编码 DCT 系数比特数 B_Q 。令新的 $Q = B_Q/B_A$, 如果 Q 变化不足 ± 1 , 则令 Q 变化等于 ± 1 。重复步骤 (2), 直到找到一个 Q , 使得 B_A 介于 B_{Q-1} 和 B_Q 之间。

(3) 确定当前宏块量化步长 如果当前帧还有未编码宏块, 则令当前宏块量化步长为 $Q + \varepsilon_Q$, 并确定当前宏块的量化参数, 否则转步骤 (1)。

(4) 更新数据 在每编码完 30 个宏块, 再计算已编码宏块的实际编码比特数 B_C (下角 C 代表 coded) 和模型估计比特数 B_E (下角 E 代表 estimate), 令 $\varepsilon_Q = \lfloor 0.1Q(B_C - B_E) / \min(B_C, B_E) + 0.5 \rfloor$ 。更新 B_A 、 B_{Q-1} 和 B_Q 。当 B_A 不介于 B_{Q-1} 和 B_Q 之间时, 则转到步骤 (2), 否则转到步骤 (3)。

上述实时流控的主要思想是尽量用相同量化步长来量化所有宏块, 以降低因局部量化步长过大而导致的块效应。虽然 θ 模型是无模型参数的固定模型, 但是受视频编码特征影响很小, ε_Q 绝大多数情

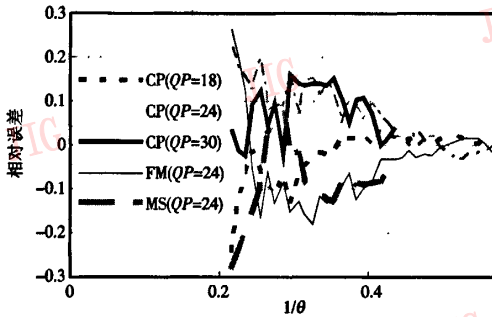
况常下为 0, 只在极低码率时 ε_0 才有很小变化, 即只需要对模型估计误差做轻微调整。

4 模拟实验和讨论

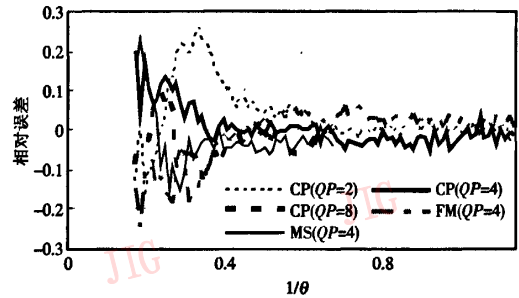
4.1 θ 数据率模型的精确度

为了验证本文算法的性能, 采用 H. 264/AVC 测试模型 JM7.6 和 H. 263+ 测试模型 TMN8 进行了实验。实验时, 选取了多个典型的视频测试序列, 在各种编码条件下进行编码测试, 并存储每个宏块的 $1/\theta$ 和数据率值, 最后计算与每个 $1/\theta$ 对应的数据率均值, 并且生成平滑后的数据率曲线。假定给定 $1/\theta$ 值, 平滑后的数据率为 \bar{R} , 一次编码测试中的数据率均值为 R_i (下角 t 代表 test)。该次测试的数据率相对误差定义为 $\bar{R}/R_i - 1.0$ 。为了确定不同编码

条件下数据率的稳定性, 还考察了多个测试的数据率相对误差, 部分实验结果如图 2 所示。图中 CP、FM 和 MS 分别表示“Carphone”、“Foreman”和“Miss”视频序列, 各测试的量化参数 (quantization parameter, QP) 用于量化所有宏块。无论是对 H. 264/AVC 还是对 H. 263+, 当 $1/\theta$ 大于 0.2 时, 则所有测试的相对误差小于 30%, 当 $1/\theta$ 大于 0.4 时, 相对误差基本上小于 5%。要注意在 $1/\theta$ 小于 0.2 时, 绝大多数宏块不会编码。当 $1/\theta$ 大于 0.4 时, 会产生较多比特, 通常情况下流控性能取决于对这些宏块数据率的估计精度, 但是由于在极低码率下, 有大量宏块的 $1/\theta$ 值为 0.2~0.4, 数据率的整体估计精度基本上由这些宏块的数据率估计误差决定, 且它们的数据率估计误差相对较大, 因此实际流控中用量化步长调整量 ε_0 来调整固定数据模型的不足。



(a) H. 264/AVC 编码器



(b) H. 263+ 编码器

图 2 不同编码条件下的数据率相对误差

Fig. 2 Relative error of rate with different coding parameters

4.2 不同数据率模型的流控性能比较

为了比较不同模型的流控的性能, 本文在 H. 263+^[9] 测试模型 TMN8 中实现了本文提出的 θ 模型流控及其文献[7]中的 ρ 域流控, 并与 TMN8 中的流控算法进行了比较。所有测试均采用 IPPP 编码, 即仅第 1 帧为 I 帧, 量化参数取 31; 其余帧为 P 帧, 量化步长由各种流控算法控制, 同时设置快速搜索, 且所有测试的视频序列均为 QICF 格式。

4.2.1 不同数据率模型的估计精度比较

假定编码第 i 帧 DCT 系数的实际编码比特数为 B_c , 模型估计比特数为 B_e , 则数据率模型的帧估计误差定义为

$$E_i = \begin{cases} B_e/B_c - 1 & B_e > B_c \\ B_c/B_e - 1 & B_c \geq B_e \end{cases} \quad (10)$$

若视频序列共有 M 帧, 则帧平均估计误差定

义为

$$E = \frac{1}{M} \sum_{i=1}^M E_i \quad (11)$$

在数据率为 384kbps、帧频为 30fps 时, 编码不同的视频测试序列, 并将平均帧估计误差及其最大帧估计误差分别列于表 1 和表 2。 ρ 域模型非常精确, 平均帧估计误差在 3% 以内, 最大帧估计误差小于 10%。 θ 模型估计精度略低于 ρ 域模型, 但也相当精确, 平均帧估计误差在 6% 以内, 最大帧估计误差小于 30%。同时注意到 θ 模型是固定模型, 由于其无任何模型参数, 因此无需动态适应。TMN8 模型估计精度较低, 其对“Carphone”序列的平均帧估计误差达到 68%, 对“Foreman”序列的最大帧估计误差为 520%。虽然 θ 模型的估计精度略低于 ρ 域模型, 但是通过剩余比特数等参数进行动态流控后,

表 1 不同数据率模型的帧平均估计误差

Tab.1 Average relative error of rate estimation with different rate models

视频序列	ρ 域模型	θ 模型	TMN8 模型
Carphone	0.02	0.03	0.68
Foreman	0.023	0.059	0.5
Miss	0.028	0.056	0.16

表 2 不同数据率模型的帧最大估计误差

Tab.2 Maximum relative error of rate estimation with different rate models

视频序列	ρ 域模型	θ 模型	TMN8 模型
Carphone	0.076	0.138	3.20
Foreman	0.099	0.298	5.20
Miss	0.038	0.162	0.69

能完全消除模型的估计误差对流控性能的影响。

4.2.2 实时流控性能比较

由于实时流控要求很小的端到端时延,因此要求精确地控制编码比特。假定 B_A 为当前帧分配的编码比特数,而实际的编码比特数为 B_C ,则定义流控相对误差为 $(B_C - B_A)/B_C$,其反映了流控对编码比特的控制能力,是判断流控性能好坏的重要依据。图 3 比较了各种流控在“Foreman”序列中,编码为 48kbps 和 10fps 时的流控相对误差。在第 99 编码帧周围,因镜头转动导致编码复杂度增加,由于此时用最大量化参数仍不能满足编码比特的控制要求,因此所有流控都出现了较大的流控相对误差。除此以外, ρ 域和 θ 模型的流控相对误差均在 2% 以内,大多数情况在 1% 以内。TMN8 虽然采用了大量动

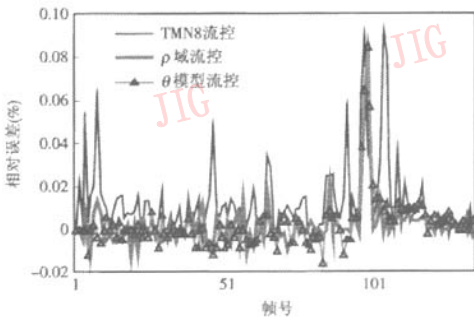


图 3 不同流控在“Foreman”序列中的流控相对误差比较
Fig.3 Comparison of relative rate control error in different rate control algorithm for “Forman”

态适应策略,但由于模型不够精确,经常出现较大的流控相对误差,从而导致数据在缓存中大量积累,可能会引起跳帧。同时会导致帧内量化参数变化巨大,使得视频质量下降。

表 3 给出了不同流控算法的编码图像的峰值信噪比 (peak signal noise ratio, PSNR) 性能。 θ 模型和 ρ 域流控的性能几乎相同。这说明虽然 θ 模型较 ρ 域模型的帧估计误差略大,但仍完全在流控算法的动态调整范围以内,不影响流控性能。 ρ 域流控需要随时计算未编码宏块的变换系数分布表,其计算复杂度为像素级,但是由于 θ 模型利用运动估算可得到最小平均绝对差值 s ,因此其不涉及任何像素级运算,计算复杂度仅为宏块级。 θ 模型无须编码前获得任何变换系数,就能很好地融合 H.264/AVC 等编码器的帧内预测特性。相对而言, θ 模型流控最灵活,且最易在各种编码器和硬件中实现。

表 3 不同流控算法的 PSNR 性能比较

Tab.3 Comparison of PSNR performance with different rate control algorithm

测试序列	目标比特率 (kbps)	帧频	PSNR (dB)		
			TMN8 流控	ρ 域流控	θ 模型流控
Container	32	10	33.39	33.56	33.76
	64	30	33.79(1)	34.06	34.12
	128	30	36.95(5)	36.98	37.23
Foreman	32	10	28.42(6)	28.67(3)	28.67(3)
	64	30	28.94(6)	29.11(3)	29.10(3)
	128	30	31.66	31.69	31.70
Carphone	32	10	30.71	30.77	30.74
	64	30	30.57	30.66	30.67
	128	30	33.33	33.32	33.33

注:括号内的数字为跳帧数,未标明表示未跳帧

5 结 论

θ 数据率模型由于消除了因宏块平均绝对差值变化而引起的数据率估计误差,因此比 Q 模型有更高的估计精度,虽然视频数据源的分布特征变化在一定程度上影响了 θ 数据率模型的估计精度,但是由于相对估计误差都不大于 30%,且绝大多数情况小于 10%,因此经过动态调整,流控算法能够完全精确控制编码比特率,同时不影响流控性能。 θ 模型流控方法简单,计算复杂度仅为宏块级,但是却获

得了比 TMN8 更好的流控性能, 已与 ρ 域流控的流控性能相当。

参考文献 (References)

- 1 Ding W, Liu B. Rate control of MPEG video coding and recording by rate-quantization modeling[J]. IEEE Transactions on Circuit System for Video Technology, 1996, 6(1): 12 ~ 19.
- 2 Wang Li-min. Rate control for MPEG video coding [J]. Signal Processing: Image Communication, 2000, 15(3): 493 ~ 511.
- 3 Ribas-Corbera J, Lei S. Rate control in DCT video coding for low-delay communications[J]. IEEE Transactions on Circuit System for Video Technology, 1999, 9(1): 172 ~ 185.
- 4 Chiang T, Zhang Y Q. A new rate control scheme using quadratic rate distortion model[J]. IEEE Transactions on Circuit System for Video Technology, 1997, 7(1): 246 ~ 250.
- 5 Kim Tae-yong, Roh Byeong-hee, Kim Jae-kyoon. An accurate bit-rate control for real-time MPEG video encoder[J]. Signal Processing: Image Communication, 2000, 15(3): 479 ~ 492.
- 6 He Zhi-hai, Mitra Sanjit K. A unified rate-distortion analysis framework for transform coding [J]. IEEE Transactions on Circuit System for Video Technology, 2001, 11(12): 1221 ~ 1236.
- 7 Kim Yong Kwan, He Zhihai, Mitra S K. A novel linear source model and a unified rate control algorithm for H.263/MPEG-2/MPEG-4 [A]. In: Proceedings of International Conference on Acoustics, Speech, and Signal Processing (ICASSP2001) [C], Salt Lake City, UT, USA, 2001: 1777 ~ 1780.
- 8 Lam Edmund Y, Goodman d Joseph W. A mathematical analysis of the DCT coefficient distributions for images [J]. IEEE Transactions on Image Processing, 2000, 9(10): 1661 ~ 1666.
- 9 ITU-T. Video Coding for Low Bit Rate Communications [S]. ITU-T Recommendation H.263, Version 2. January 1998.