

H. 264 标准中的宏块编码模式的时空相关性预测算法^{*})

周敬利 向东 余胜生 陈加忠

(华中科技大学计算机科学与技术学院 武汉 430074)

摘要 与以往的各种视频编码标准相比, H. 264 视频编码标准在编码效率上获得了相当大的提高。但这是以提高编码的计算复杂度为代价的, 其中运动估计和模式选择的计算复杂度最大。文中提出一种宏块编码模式的快速预测算法, 该算法首先检查根据宏块编码模式的时空相关性预测得到的模式, 再与相应的阈值进行比较, 判断是否结束模式选择过程, 而不用检查所有的模式。试验结果表明: 在保持相近率失真性能的同时, 该算法可以有效地降低编码的计算复杂度。

关键词 视频编码, H. 264, 模式, 时空相关性, 阈值

A Predicting Algorithm of Macroblock Coding Mode Based on Spatio-Temporal Correlation for H. 264

ZHOU Jing-Li XIANG Dong YU Sheng-Sheng CHEN Jia-Zhong

(School of Computer Science and Technology, HUST, Wuhan 430074)

Abstract H. 264 has achieved a significant improvement in rate-distortion efficiency relative to existing standards. However, this comes at a cost in considerably increased complexity at the encoder mainly due to motion estimation and mode decision. In this paper, a fast predicting algorithm of macroblock coding mode was proposed. This method first searches the modes predicted by the spatio-temporal correlation of macroblock coding mode, then decides whether to terminate the mode decision procedure according to a corresponding threshold instead of searching all modes. Experimental results show that this method can efficiently reduce the computational complexity while keeping similar RD performance.

Keywords Video coding, H. 264, Mode, Spatio-temporal correlation, Threshold

1 引言

H. 264 视频编码标准是目前最新的国际视频编码标准^[1], 其主要目标是提高视频压缩的性能, 提供网络友好的视频流表示, 以及适应各种不同应用的需求。H. 264 标准和以前的标准均采用了基于块的混合编码结构, 但它采用了可变块尺寸的运动补偿、1/4 像素精度运动估计、多参考帧、4×4 整数变换和 CABAC (Context-based Adaptive Binary Arithmetic Coding) 等新技术。与以往的各种视频编码标准相比, H. 264 标准在编码效率上获得了最大 50% 的提高^[2]。

基于固定大小尺寸块的运动估计和补偿在 MPEG-1、2 和 H. 261/H. 263 等标准中获得了广泛的运用, 但这种技术并不能适应视频场景特性的变化, 如在运动剧烈区域和纹理区域希望用小尺寸的块进行编码, 而在静止背景和变化缓慢区域则采用大尺寸的块进行编码。为了更好地适应视频场景的变化, H. 264 标准采用了可变块尺寸的运动补偿技术, 最多允许 7 种不同尺寸大小的运动补偿块。在相同的视频质量条件下, 该技术能够获得 4%~20% 的比特率节省, 但是其计算复杂度是随着允许的模式数目而线性增加的, 每增加一种模式, 复杂度增加 2.5% 左右^[2], 特别是采用多参考帧时, 其计算复杂度更是随着参考帧数成倍增加。

在 H. 264 标准中, 为了得到宏块的编码模式, 模式选择过程对所有的模式进行全搜索, 因此其计算复杂度非常大。文中提出一种宏块编码模式的时空相关性预测算法(简称模式预测算法), 该算法首先检查根据宏块编码模式的时空相关性预测得到的模式, 再与相应的阈值进行比较, 判断是否提前终止, 而不用搜索所有的模式, 因此大大降低了模式选择的计

算复杂度。

2 模式选择过程

在 H. 264 标准中, 运动估计和模式选择采用了率失真优化技术。在对宏块进行编码时, 首先对帧间预测 (INTER) 模式的所有的块做运动估计, 得到最佳运动矢量和参考帧, 接着计算所有 INTER 模式和帧内预测 (INTRA) 模式的率失真代价, 选取代价最小的作为编码模式。在运动估计过程中, 相应块的最佳运动矢量和参考帧通过最小化代价 J_{motion} 选择^[3]:

$$J_{motion} = SAD(s, c) + \lambda_{motion} R(MVD, REP) \quad (1)$$

式中, λ_{motion} 是用于运动估计的拉格朗日乘法算子, MVD 表示当前运动矢量与运动预测矢量的差, RFP 表示参考帧, R (MVD, RFP) 表示编码运动矢量差和参考帧所需的比特数, SAD (Sum of Absolute Difference) 在原始块信号 s 和运动估计块信号 c 之间进行。

宏块编码模式通过最小化率失真代价 J_{mode} 选择^[3]:

$$J_{mode} = SSD(s, c, MODE|QP) + \lambda_{mode} R(s, c|MODE|QP) \quad (2)$$

式中, λ_{mode} 是用于模式选择的拉格朗日乘法算子, QP 为量化参数, MODE 是允许的宏块模式, s 是原始的宏块信号, c 是根据不同模式重建的宏块信号, $R(s, c, MODE|QP)$ 表示对宏块编码所需的比特数, 包括宏块头, 运动矢量和变换系数信息, SSD (Sum of Square Difference) 在原始宏块信号和重建的宏块信号之间进行计算。

在 H. 264 标准中, 对于 P 帧而言, $MODE \in \{INTRA-4 \times 4, INTRA-16 \times 16, SKIP, INTER-16 \times 16, INTER-8 \times 16, INTER-16 \times 8, INTER-8 \times 8\}$, 当采用 INTER-8×8 模式时, 每

^{*} 本课题得到国家“八六三”高技术研究发展计划(2002AA11901010)和湖北省自然科学基金项目(2002ABB027)资助。周敬利 教授, 博士生导师, 研究方向为多媒全通信技术及其应用, 高性能网络存储技术; 向东 博士研究生, 研究方向为视频压缩与通讯; 余胜生 教授, 博士生导师, 研究方向为计算机网络与存储, 多媒体处理与通信; 陈加忠, 博士, 研究方向为小波理论和信号分析, 图像与初频编码技术。

个8×8大小的子宏块又可以采用 INTER-8×8, INTER-8×4, INTER-4×8 和 INTER-4×4 这四种宏块子模式。在这些模式中, INTER 模式的计算复杂度远大于 INTRA 模式和 SKIP 模式, 这是因为必须对每一种 INTER 模式进行运动估计; INTRA 模式的计算复杂度处于 INTER 模式和 SKIP 模式之间; SKIP 模式无需作运动估计, 也无需对预测残差进行变换、量化、反量化和反变换, 因此其计算复杂度最小。

3 模式预测算法

3.1 宏块编码模式的时空相关性

在视频编码中, 对于一帧图像中静止背景区域和变化缓慢的区域, 一般采用大尺寸的模式进行编码。在这些区域中, 当前宏块和它的时空相邻宏块可能位于同一运动物体上或相同的背景中, 因此它们的编码模式很可能相同或者相近。例如当图1中宏块a(宏块d的左宏块), 宏块b(宏块d的上宏块), 宏块c(宏块d的前一帧相同位置宏块)的编码模式是SKIP(记为模式0)或INTER-16×16(记为模式1)时, 宏块d的编码模式很有可能是0或1; 若不是, 也趋向于相近的INTER-16×8或INTER-8×16, 我们称这种关系为宏块编码模式的时空相关性。

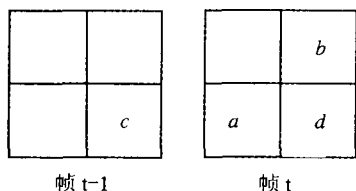


图1 宏块最佳模式预测示意图

为了研究这种相关性以及选择那些相邻宏块作为预测宏块, 我们对5个CIF(Common Intermediate Format)格式的视频流的编码结果进行了统计, 统计结果表明:

- 1) 在这些相邻的宏块中, 宏块a、宏块b和宏块c的编码模式与当前宏块d的编码模式的相关性最大。采用更多的相邻宏块固然能够提高预测精度, 但会使符合相关性条件的宏块数目大大减少, 因此我们用以上3个宏块作为预测宏块。
- 2) 当宏块a、宏块b和宏块c的编码模式均为0时, 宏块d的编码模式为0的概率远大于其它模式, 其次模式1。
- 3) 当宏块a、宏块b和宏块c的编码模式是0或1时(a、b、c的编码模式不同时为0), 宏块d的编码模式为0和1的概率和远大于其它模式。

根据统计结果2和3, 并考虑到模式0的计算复杂度最小, 因此在模式预测算法中, 首先检查模式0, 再检查模式1。每检查完一种模式时, 该模式的率失真代价与下面一节将要讨论的阈值进行比较, 以判断是否终止模式预测算法。

3.2 阈值的选取

选取适当的阈值, 对于模式预测算法的性能是至关重要的。我们可以用已编码图像中不同模式率失真代价的累加和的平均值作为不同模式的阈值。这种方法虽然比较简单, 但它不能很好地适应视频场景的变化, 因为编码模式即使相同, 其率失真代价有可能相差很大。

从上一节统计结果1可知, 宏块a, 宏块b和宏块c的编码模式与宏块d的编码模式的相关性最大, 因此使用它们编码模式的率失真代价作为阈值函数的参数, 能够更好地反映视频场景的变化。几种可用的阈值函数可以表示为:

$$T_k = a_k \times \min(J_a, J_b, J_c) + b_k \quad (3)$$

$$T_k = a_k \times \text{median}(J_a, J_b, J_c) + b_k \quad (4)$$

$$T_k = a_k \times \text{avg}(J_a, J_b, J_c) + b_k \quad (5)$$

$$T_k = a_k \times \max(J_a, J_b, J_c) + b_k \quad (6)$$

式中, a_k, b_k 可以是常数, 如 $a_k=1, b_k=0$, 也可以根据视频场景特性自适应变化, J_a, J_b, J_c 分别为宏块a、宏块b和宏块c的编码模式的率失真代价。我们可以选取最小值式(3), 中值式(4), 平均值式(5)或者最大值式(6)作为提前结束的准则。但一个有效的阈值在降低模式选择的计算复杂度的同时, 应尽可能保持小的率失真性能损失。下面我们从以上两个方面, 以宏块a、宏块b和宏块c的编码模式均为0为例来说明阈值的选取。

当计算得到模式0或模式1的率失真代价时, 我们与(3)~(6)中的某个阈值进行比较以判断是否终止模式预测算法。但是, 有可能在提前终止时得到的率失真代价大于该宏块的最小率失真代价, 从而导致率失真性能的损失。如图2所示, 图中左边为采用阈值式(3)~(6)以及阈值(5)且偏移值为100和200时, 错误提前终止的宏块数, 右边为正确提前终止的宏块数。一般来说, 错误提前终止的宏块数越多, 其率失真损失就越大, 但是不同的阈值对率失真性能的影响也不尽相同。因此我们用错误提前终止的宏块数的加权和来评价对率失真性能的影响:

$$N = \sum_i \alpha_i N_i \quad (7)$$

其中, $0 \leq \alpha_i \leq 1$, 当选取最小值作为阈值时, 我们认为其对率失真性能影响很小, 此时 $\alpha=0.25$, 当采用中值和平均值作为阈值时, $\alpha=0.5$, 当采用最大值时 $\alpha=1$, N_i 为采用不同的阈值时错误提前终止的宏块数。

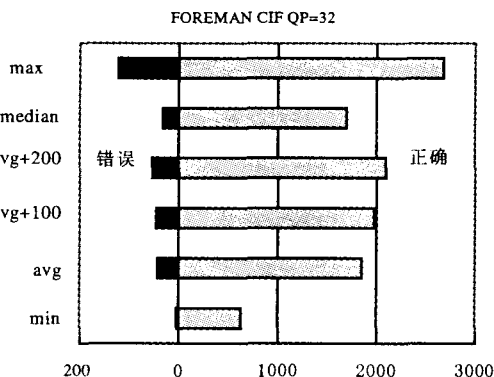


图2 正确提前终止和错误提前终止的宏块数目对比

在模式预测算法中, 一般说来, 提前终止的宏块数越多, 模式选择的计算复杂度降低就越大。在H.264标准中, 每一种帧间模式的计算复杂度基本相同^[2], 模式0的计算复杂度相对于帧间模式非常小, 因此可以假设其计算复杂度为0。我们用必须搜索的宏块模式的节省率来衡量计算复杂度的降低, 节省率越大, 计算复杂度降低也越大, 节省率可以近似表示为(这里不考虑帧内模式):

$$S = (M_0 + M_1 \times 6/7) / M \quad (8)$$

其中 M_0 和 M_1 分别为在模式0和模式1时提前终止的宏块数, M 为满足宏块a、宏块b和宏块c的编码模式均为0的宏块总数。

采用阈值式(3)~(6)作为模式预测算法中模式0和模式1的判断准则时, 它们可以采用以下几种组合方式:

A. 模式0时阈值使用最小值; 模式1时, 如果模式1的率失真代价不大于模式0的率失真代价, 则模式1的阈值使用平均值; 否则, 模式0的阈值使用最大值。

B. 模式0时阈值使用平均值; 模式1时, 如果模式1的率失真代价不大于模式0的率失真代价, 则模式1的阈值使用平均值; 否则, 模式0的阈值使用最大值。

C. 模式0时阈值使用中值;模式1时,如果模式1的率失真代价不大于模式0的率失真代价,则模式1的阈值使用中值;否则,模式0的阈值使用最大值。

D. 模式0时阈值使用最大值;模式1时,如果模式1的率失真代价不大于模式0的率失真代价,则模式1的阈值使用最大值。

表1 使用不同阈值组合时的试验结果

组合		A	B	C	D	A	B	C	D	A	B	C	D
		$b_k=0$				$b_k=100$				$b_k=200$			
Foreman	S(%)	68.8	73.8	73.2	78.5	71.9	76.5	76.1	81.2	74.9	79.2	78.9	83.8
	N	26	36	35	134	27.5	39	37.5	138	30.5	45	41	145
Container	S(%)	70	74.6	74.2	79.4	74.6	79.2	79.3	84	78.6	82.9	83.4	87.7
	N	64	72	71.5	157	66.75	73.5	72.5	166	73.75	80	80.5	176

我们对 Foreman 和 Container 序列采用以上四种组合时的 S 和 N 值进行了计算,表1给出了计算结果。我们的目标是找到最佳的组合,使 S 最大化的同时, N 最小化。但是这两个函数是相互矛盾的,即增大 S 的方法必将使 N 增大。如表1所示:在偏移值为0的情况下,组合 B、组合 C 和组合 D 的 N 和 S 值均大于组合 A;在 Foreman 序列中,组合 B、组合 C 和组合 D 的 S 值比组合 A 分别增加了 5%、4.4% 和 9.7%,而 N 值分别增加了 10、9 和 109;在 Container 序列中,组合 B、组合 C 和组合 D 的 S 值比组合 A 分别增加了 4.6%、4.4% 和 9.4%,而 N 值分别增加了 8、7.5 和 94。因此,组合 B 和组合 C 在获得较大的计算复杂度降低的同时,而能保持较小的率失真性能损失。作为计算复杂度和率失真性能的折衷,我们选择组合 B 和组合 C 作为最优的阈值组合。

此外,我们从图2发现,当阈值函数采用一定的偏移值 b_k 时,正确提前退出的宏块数将会大大增加,而错误提前退出的宏块数的增加将会很小,从而使 S 增加的同时, N 的增加较小。表1列出了采用偏移值 100 和 200 时的 N 和 S 值,从表中可以看出,采用偏移值 100 的性能明显优于偏移值 200,特别是在 Container 序列中,组合 B 和组合 C 的 S 值的增幅约为偏移值 200 的 55%,但是 N 仅仅增加了 1.5 和 1。

4 试验结果及分析

我们采用组合 B 和偏移值 100,在 H.264 参考软件 JM7.3 的平台上执行了模式预测算法,测试的序列有 CIF 格式的 Container、Paris、Foreman、News、Akiyo、Flowergarden 和 Tempete 序列,编码帧率为 30Hz,量化参数为 24、28、32 和 36,每 15 帧插入一个 I 帧,采用 2 个参考帧,运动估计的搜索范围为 32,仅对 P 帧进行了试验,H.264 标准中的模式选择采用全搜索算法。图3是模式预测算法和全搜索的率失真曲线比较,表2是模式预测算法相对于全搜索的平均 PSNR (Peak Signal-to-Noise Ratio)增益、平均比特节省和计算复杂度降低。

表2 模式预测算法与全搜索的性能比较

测试序列	平均 Y 信号 PSNR 增益 L (dB)	平均比特节省 (%)	计算复杂度降低 (%)
Container	-0.024	-0.62	42~74
Foreman	-0.05	-1.19	22~44
Paris	-0.046	-0.77	45~57
Flowergarden	-0.015	-0.24	31~41
Tempete	-0.02	-0.47	14~37
News	-0.06	-1.05	56~67
Akiyo	-0.048	-0.93	72~83
平均值	-0.038	-0.753	

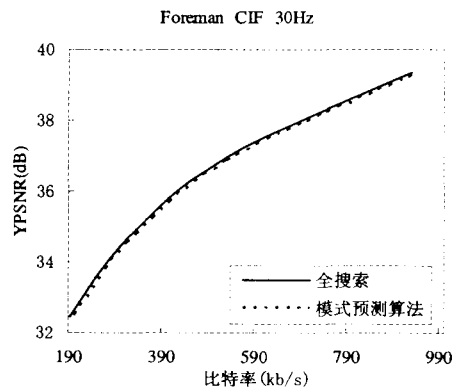


图3 模式预测算法和全搜索的率失真曲线比较

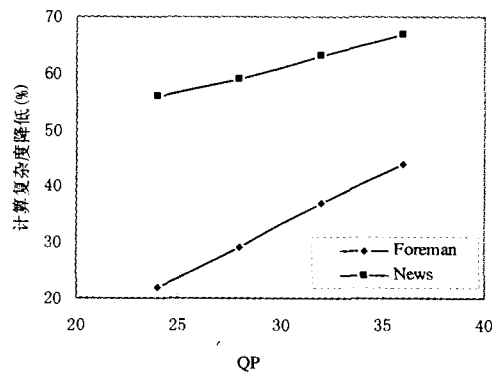


图4 计算复杂的降低随QP变化示意图

从图3和表2可以看出:模式预测算法与全搜索相比,率失真曲线非常接近,PSNR 损失非常小。对于变化平缓或有较多静止背景区域的视频序列,在这些序列中满足时空相关性条件的宏块数目比摄像机运动、复杂物体运动或复杂纹理的序列多,因此模式选择计算复杂度的降低也大,如 Container 序列的最大计算复杂度降低为 74%,Tempete 序列的最大计算复杂度降低为 37%。并且随着 QP 的增大,计算复杂度降低也相应增加,如图4所示,这是因为随着 QP 的增大,在率失真代价中运动矢量所占的比重增大,模式选择趋向于选择运动矢量少的模式,即编码模式为 0 或 1 的概率增加,从而满足时空相关性的宏块数目增加。

此外,模式预测算法的性能远好于通过减少模式数来降低编码计算复杂度的方法,如只采用模式 0,模式 1 和 INTER-8×8(只采用 INTER-8×8 宏块子模式)这 3 种模式时,Akiyo 序列的最大计算复杂度降低约为 18%,但是其 PSNR 损失为采用模式预测算法的两倍,达到了 0.12dB;Foreman

(下转第 127 页)

• 利用统计语言模型,将语言知识、用户背景和查询上下文融入到信息检索中,建立更为复杂的短期和长期用户模型,以使得检索结果最大限度地符合用户实际信息需求。

• 将语言模型应用到不同的信息检索领域,如信息过滤(Information filtering)、多语言检索(Cross-language informa-

tion retrieval)、Web 信息检索等等。

目前,我们正进行的研究工作有:基于语言模型的文档相似性研究;基于语言模型的用户模型的研究;将语言模型应用到个性化推荐中。

表 2 不同信息检索模型的比较

	特征项权重	结构化查询	语义扩展	匹配方法	相关性反馈	其他
布尔模型	-0 或 1	-显式 -AND,OR,NOT...	-	-精确匹配 -BIR	-	结构化查询
向量空间模型	-实验获得 -标准化	-隐含 -AND	-字典 -聚集分析	-模糊匹配 -基于距离	-权重更新 -Rocchio 方法	特征项权重
传统概率模型	-相关性文档估计 -获得困难	-隐含 -AND	-	-模糊匹配 -基于概率	-重新估计权重	相关性反馈
语言模型	-直接估计 -容易获得 -不需要标准化 -体现语言特点	-隐含 -AND,OR,NOT	-查询语言模型	-模糊匹配 -基于概率分布或 Bayes 决策	-查询语言模型	综合框架

参 考 文 献

- 1 Ponte J M,Croft W B. A Language Modeling Approach to Information Retrieval. In:Proc. of the 21st Intl. Conf. on Research and Development in Information Retrieval(SIGIR98)
- 2 Berger A,Lafferty J. Information Retrieval as Statistical Translation. In:Proc. of the 22st Intl. Conf. on Research and Development in Information Retrieval(SIGIR99)
- 3 Lafferty L, Zha C. Probabilistic Relevance Models Based on Document and Query Generation. Language Modeling for Information Retrieval. Kluwer Academic Publishers
- 6 Leek M D,Schwartz R. A Hidden Markov Model Information Retrieval System. In:Proc. of the 22st Intl. Conf. on Research and Development in Information Retrieval(SIGIR99)
- 7 Zhai C,Lafferty J. Model-Based Feedback in The Language Modeling Approach to Information Retrieval. www. google. com
- 8 Lafferty J,Zhai C. Document Language Models, Query Models, and Risk Minimization for Information Retrieval. In:Proc. of SIGIR'2001
- 9 Zaragoza H, Hiemstra D, Tipping M. Bayesian Extension to The Language Model for Ad Hoc Information Retrieval. In:Proc. of the 26st Intl. Conf. on Research and Development in Information Retrieval(SIGIR03)
- 10 Jin R, Hauptmann A G, Zhai C. Title language model for information retrieval. SIGIR,2002. 42~48
- 11 Song F, Croft W. A general language model for information retrieval. In: Proc. of Eighth Intl. Conf. on Information and Knowledge Management (CIKM99)
- 12 Lavrenko V, Croft W. Relevance-Based Language Models: Estimation and Analysis. In: Proc. the ARDA Workshop on Language Modeling and Information Retrieval, Pittsburgh, PA, May 31-June 1, 2001
- 13 Jin R, Si L, Hauptmann G, Callan P. Language model for IR using collection information. SIGIR,2002. 419~420
- 14 Bahl L R, Brown P F. A Tree-Based Statistical Language Model

- for Natural Language Speech Recognition. IEEE Transactions on Acoustics, and Signal Processing, 1989
- 15 Rosenfeld R. Adaptive Statistical Language Modeling; A Maximum Entropy Approach; [CMU Technical Report CMU-CS-94-138].
- 16 Zhai C,Lafferty J. A Risk Minimization Framework for Information Retrieval. citeseer. nj. nec. com
- 17 David J C. A Hierarchical Dirichlet Language Model. Natural Language Engineering, 1995,1(3):1~19
- 18 Chen S F,Goodman J. An Empirical Study of Smoothing Techniques for Language Modeling; [Technical Report Tr-10-98]. Harvard University
- 19 Zhai C,Lafferty J. A Study of Smoothing Methods for Language Models Applied to Ad Hoc Information Retrieval. In:Proc. of SIGIR'2001
- 20 Zhai C,Lafferty J. Two-Stage Language Models for Information Retrieval. In: Proc. of the 25th Intl. Conf. on Research and Development in Information Retrieval (SIGIR02)
- 21 Hiemstra D,de Jong F M G. Disambiguation Strategies for Cross-Language Information Retrieval. In:Proc. of the third European Conf. on Research and Advanced Technology for Digital Library (ECDL), 274~293
- 22 Hiemstra D. Term-Specific Smoothing for the Language Modeling Approach to Information Retrieval. The Importance of a Query Term. In:25th Annual Intl. ACM SIGIR Conf. on Research and Development in Information Retrieval, Tampere, Finland, Aug. 2002
- 23 Hiemstra D. Using Language Models for Information Retrieval;[PH. D thesis]. University of Twente, 2001
- 24 Fuhr N. Probabilistic Models in Information Retrieval. The Computer Journal,1992,35(3):243~255
- 25 Allan J, Croft B, et al. Challenges in Information Retrieval and Language Modeling. Report of a Workshop held at the Center for Intelligent Information Retrieval, University of Massachusetts Amherst, Sep. 2002

(上接第 120 页)

序列的最大计算复杂度降低约为 19%,但是其 PSNR 损失竟达到了 0.34dB。

结论 文中提出了一种用于 H. 264 标准中的宏块编码模式的时空相关性预测算法,该算法的有效性取决于满足时空相关性条件的宏块数和判断阈值。对于变化平缓或有较多静止背景区域的视频流,模式选择的计算复杂度大大降低,并且随着 QP 的增加,计算复杂度的降低也相应增加。试验结果表明:宏块编码模式的时空相关性预测算法,既能降低编码的计算复杂度,又能保持相近的率失真性能。

参 考 文 献

- 1 ISO/IEC 14496-10;2003. Coding of audiovisual objects-part 10: advanced video coding,2003
- 2 Ostermann J, Bormans J, List P, et al. Video coding with H. 264/AVC; tools, performance and complexity. IEEE Circuits and Systems Magazine, 2004, 4(1): 7~28
- 3 Wiegand T, Schwarz H, Joch A, et al. Rate-constrained coder control and comparison of video coding standards. IEEE Transactions on Circuits and Systems for Video Technology, 2003, 13(7): 688 ~703