

一种分层判决结构的 H.264/AVC 快速帧间模式选择方法

何书前^{1,2},倪江群¹,石 春²

(1. 中山大学信息科学与技术学院, 广东广州 510275; 2. 海南师范大学信息科学技术学院, 海南海口 571158)

摘 要: H.264/AVC 视频编码标准采用了基于率失真最优化(RDO)技术的多种编码模式选择的预测编码方法,取得了很高的率失真效率,同时也引入了非常高的计算复杂度.为了降低 RDO 全搜索方法的预测复杂度,本文提出了一种分层判决结构的快速帧间模式选择方法.根据 H.264/AVC 帧间预测编码的结构特点,将宏块的各种模式分为 4 个层次.研究运动估计预测原理,利用预测编码的时间复杂度与空间复杂度之间的耦合关系,得到每个层次的自适应判决阈值,使得快速模式选择方法应用于每一个层次,在保证率失真性能的前提下,及时退出多余的全搜索率失真代价函数计算.实验结果表明,该算法平均减少了 45.79% 的编码时间,同时,保持了视觉质量和输出码率不变.对比结果也显示,与 Zeng 方法相比,得到了 0.05dB 的 PSNR 增益;相比 Choi 方法,本文方法节省了约 20% 的计算时间;本文方法优于这两种快速模式选择方法.

关键词: H.264/AVC; 模式选择; 率失真最优化; 帧间预测

中图分类号: TN919.8 **文献标识码:** A **文章编号:** 0372-2112 (2013) 11-2199-08

电子学报 URL: <http://www.ejournal.org.cn>

DOI: 10.3969/j.issn.0372-2112.2013.11.014

A Novel Fast Inter-Frame Mode Decision Scheme for H.264/AVC Based on a Hierarchical Decision Process

HE Shu-qian^{1,2}, NI Jiang-qun¹, SHI Chun²

(1. School of Information Science and Technology, Sun Yat-sen University, Guangzhou, Guangdong 510275, China;

2. School of Information Science and Technology, Hainan Normal University, Haikou, Hainan 571158, China)

Abstract: The rate distortion optimization (RDO) which uses mode-selection based prediction coding can achieve the high coding efficiency in H.264/AVC video coding standard. However, exhaustive checking of all the prediction modes increases computational complexity significantly. In this paper, we propose a hierarchical decision structure for fast inter-frame mode selection to reduce the complexity of the exhaustive RDO mode decision process. Different modes of macroblock are first classified into 4 layers in terms of the characteristics of the structure of inter-frame prediction coding in H.264/AVC. Using the coupled relationship between the spatial and temporal complexity of the motion-estimation prediction coding, the adaptive threshold for each layer can be calculated. The obtained threshold can then be applied to each layer to terminate the mode selection process while maintaining almost the same R-D performance in H.264. Experimental results show a 45.79% reduction on computational time and the same RD performance as the baseline coder by using the proposed fast selection method. The results also demonstrate that the proposed algorithm is superior to two recent fast mode decision algorithms, with the peak signal-to-noise ratio (PSNR) being improved by 0.05dB on average when compared to Zeng's algorithm and the entire encoding time being reduced by about 20% on average when compared to Choi's mode decision algorithm.

Key words: H.264/AVC; mode decision; rate distortion optimization (RDO); inter-frame prediction

1 引言

最新的视频编码标准 H.264/AVC 实现了高效的编

码效率,扩展了视频服务的网络应用,如移动网络视频服务、流式媒体服务和高清交互媒体服务等.相比以往的视频编码标准,H.264/AVC 采用多种预测编码技术,

如多参考帧、可变块大小、1/4 精度运动估计等;这些技术实现了高编码效率,但引入严重的计算复杂度^[1,2].特别是使用率失真最优化(RDO)技术穷举所有的编码模式,得到最佳的编码模式以获得最优率失真性能的编码效率,造成极大的计算复杂度,占了编码器 90% 的计算资源.

随着 H.264/AVC 的广泛应用,高复杂度编码器在移动终端中的实时应用受到限制;为了降低编码器的复杂度并保持编码效果,近年来,很多有效的帧间快速模式选择方法被提出^[3-12].其中,一些算法建议优化 RDO 函数,搜索所有的模式,获得较好的编码效率;另外一些方法则利用时空相关性,预测模式搜索子集,避免检验所有的模式,大大降低了编码复杂度;本文的工作属于后一种方式.在编码器中,视频空间复杂度和时间运动程度的高低决定该类内容更适合于某种编码模式,如空间复杂度较低,时间域运动状态较低的宏块,更适合较大块编码模式;而空间复杂度高,时间运动程度较剧烈,则更适合更小分块的编码模式.视频序列中存在高时空相关性,从已编码宏块信息可准确预测当前宏块的时空复杂度,从而确定最佳模式;具体分为以下几种情况:(1)空间特性预测方法:文献[6]中利用当前宏块的水平和垂直方向的边缘矢量信息作为检测空间复杂度的测度,获得当前宏块的空间复杂度信息,文献[8]则利用方差作为空间复杂度信息的测度;文献[7]提出零块模式的概念,用于确定提前中止阈值.(2)时间特性预测方法:文献[3,9]利用已编码块的运动矢量信息,构建了运动度量模型,对编码模式进行归类,大大简化了 RDO 模式选择方法.文献[5]则采用从底向上的思路,通过运动估计得到的 4×4 块运动矢量信息,确定当前宏块最佳编码模式.(3)时空相关特性:文献[11]对 P 帧的时空相关性进行随机分析,得到空间信息估计模型,确定第一阶段的阈值;通过分析编码信息的组成部分,得到了时间相关性模型,结合当前宏块的时空相关性,得到第二阶段的阈值.文献[9,10]也采用类似的分阶段时空相关性预测思路;其中,文献[9]在残差与运动矢量均为差零条件下选择 skip 模式为最佳模式,并对帧间最佳模式的时空复杂度与帧内模式空间复杂度比较判断是否退出帧内模式处理;文献[10]则利用空间相关性预测 skip 模式和帧内模式,再利用时间相关性度量预测帧间运动估计编码模式.综上所述,现有的研究均指出了时间特性与空间特性对预测结果的决定关系.空间特性用于零块提前退出方法取得了非常好的效果,但是,缺少了运动信息,预测精度有限;时间特性方法则是利用时空已编码块的运动矢量信息,预测当前宏块的最佳编码模式,未考虑残差信息的影响,容易导致预测误差传播;两者均考虑,将时空特性

分开建模,但未涉及到时间与空间特性之间的关系,从而影响了预测结果的准确性.

因此,本文方法通过 RDO 预测编码分析,结合率失真模型,映射为时空相关性模型,既考虑了空间特性,又考虑了编码模式之间的时间特性,利用空间与时间特性的耦合关系,推导出可提前退出的编码条件,提高了预测编码模式的准确性,从而提高编码效率.

2 H.264 帧间模式选择原理

H.264/AVC 采用的 RDO 模式选择,提供 7 种帧间运动补偿块大小.这些模式分为两级结构,宏块首先分割为 4 种模式: $16 \times 16, 16 \times 8, 8 \times 16, 8 \times 8$;其次 8×8 模式进一步分割为 $8 \times 8, 4 \times 8, 8 \times 4, 4 \times 4$ 更小的块模式.因此,加上其他类型的模式,帧间编码宏块共有 11 种候选模式.RDO 处理是找出最小的率失真代价模式,以达到最高的率失真性能.RDO 编码框架中,当前宏块需在已编码帧中选择最佳的参考帧,通过帧间或帧内预测,得到残差信号与运动信号,最后经过变换、量化、熵编码输出总的码率 R ^[1,2].

在预测编码过程中,为了简化编码器,运动估计与模式选择分别采用了不同的率失真代价函数:

(1)模式选择:

$$J(o, p, m | Qp) = \text{Min} \{ D(o, p, m | Qp) + \lambda \cdot R(o, p, m | Qp) \} \quad (1)$$

其中, J 代表模式选择率失真代价函数. D 采用 $SSD = \sum_{i,j} |Diff(i, j)|^2$, 代表原始帧宏块 o 与重构帧预测宏块 p 的平方残差和, $Diff(i, j) = o(i, j) - p(i, j)$, Qp 为量化参数, m 为编码模式, R 表示宏块编码信息, λ 为量化参数相关的经验值,即: $\lambda = 0.85 \times 2^{(Qp - 12)/3}$.

(2)运动估计:

$$J_{motion}(m, \sqrt{\lambda}) = \text{Min} \{ SAD(o, p(m)) + \sqrt{\lambda} \cdot R(mvd) \} \quad (2)$$

$SAD = \sum_{i,j} |Diff(i, j)|$ 为残差绝对值和; $mvd = mv - mvp$, $mv = (mv_x, mv_y)$ 是运动矢量, $mvp = (mvp_x, mvp_y)$ 为预测的运动矢量, $R(mvd)$ 表示编码 mvd 信息.

3 基于耦合关系的快速模式选择方法

RDO 代价模型是综合了残差信息、失真和运动信息的联合优化模型.近年来,在 RDO 率失真快速模式选择方法中,有许多时间复杂度和空间复杂度度量方法,并取得了较好的预测效果,但未考虑两者之间的关系,影响了编码的率失真性能.

3.1 率失真模型分析

文献[16]分析了混合编码框架下运动估计率失真模型.假定视频编码器输出总码率 R 由 i 个不同类型的

码率组成:

$$R = R_1 + R_2 + \cdots + R_i \quad (3)$$

同时,假设失真率函数: $D(R_1, R_2, \cdots, R_i)$ 为严格凹且连续可导的代价函数,符合 $\frac{\partial D}{\partial R_i} \leq 0, \forall i$, 即单独增加某个码率将降低总的失真,反之亦然. 最佳比特分配问题是在目标码率条件下获得最小失真,即:

$$dD = \frac{\partial D}{\partial R_1} dR_1 + \frac{\partial D}{\partial R_2} dR_2 + \cdots + \frac{\partial D}{\partial R_N} dR_N = 0 \quad (4)$$

当 R 为固定目标码率时:

$$dR_1 + dR_2 + \cdots + dR_i = 0 \quad (5)$$

式(6)为式(4)、(5)方程组的解,可达到最佳的率失真性能:

$$\frac{\partial D}{\partial R_1} = \frac{\partial D}{\partial R_2} = \cdots = \frac{\partial D}{\partial R_N} \quad (6)$$

当前宏块的编码信息可分为两个部分:残差信息 $R(e)$ 与头信息 $R(h)$; 其中,头信息包含运动信息. 假定

头信息由非负函数 $Mt(x_1, x_2 \cdots x_n)$ 表示, 其中, x_i 为影响头信息函数的参数, x_i 与运动信息存在线性关系; 如给定 R , 残差失真随着头信息的增加而降低; 即式(4)、(5)简化为 $R = R(e) + R(h)$, 满足式(6)中的最佳率失真的条件. 从式(5)可以得到, $dR(e) = -dR(h)$, 得到:

$$\frac{dR(e)}{dMt} = -\frac{dR(h)}{dMt} \quad (7)$$

从式(7)中可知, 相对于运动信息, 头信息的增加将导致残差信息的减少, 称为两者之间的耦合关系.

3.2 头信息编码分析

在 H.264/AVC 编码器中, 将编码输出残差之外的信息归类为头信息, 由模式, 运动矢量, 参考帧, 量化参数等信息组成, 如图 1 所示, 在这些信息中, 主要组成部分为模式信息 $R(\text{Mode})$ 和运动矢量信息 $R(\text{Inter})$, 其他的(如量化参数等)信息 $R(Q)$ 占的比重较小且稳定, 可用常数 θ 来表示; 因此, 头信息的定义如下:

$$R(h) = R(\text{Mode}) + R(\text{Inter}) + \theta \quad (8)$$

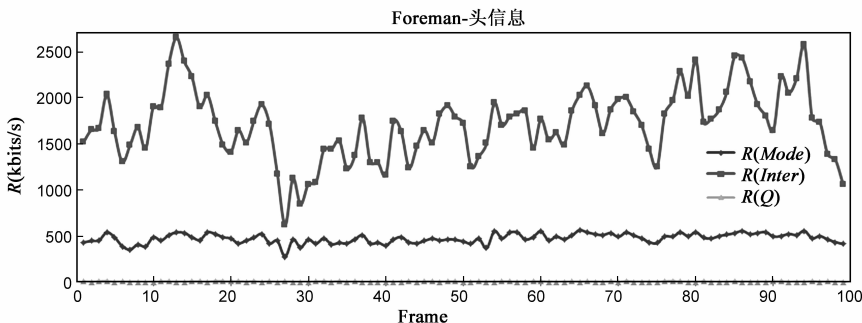


图1 Foreman视频序列头信息组成部份

定义 N_B 为每一帧所有宏块最佳模式的子宏块数

之和, 即 $N_B = \sum_{i=1}^{M_B} N_i$, M_B 为每帧的宏块数, N_i 为第 i 个宏块的分块数, 按照帧间每个宏块子块计算, 定义如下: skip 为 0, 16×16 为 1, 8×16 和 16×8 为 2, 8×8 为 4, 帧内模式为 4, 以此类推. 编码模式分块的数量 N_i 越大, 则 $R(\text{Mode})$ 和 $R(\text{Inter})$ 产生的信息量越大, 同时, 同一种模式下, 运动矢量差值 mvd 越大, 则 $R(\text{Inter})$ 的信息量越大. 本文对 QCIF 格式的 news, foreman 和 football 三种不同运动剧烈程度的视频序列进行测试, 选择 16 至 40 的偶量化参数, 随机选出第 1, 10, 20 帧的数据, 给出了运动矢量、模式与编码块数之间的关系, 横坐标为当前量化参数条件下每帧编码子宏块数的总和 N_B . 由于篇幅有限, 如图 2 中只列出了 Foreman 序列关系图, 图中显示帧的模式和运动矢量信息与 N_B 之间的呈线性关系, 即:

$$R(\text{Mode}) = \alpha \cdot N_B \quad (9)$$

$$R(\text{Inter}) = \beta(mvd) \cdot N_B \quad (10)$$

其中, α 为常量, $\beta(mvd)$ 为该模式平均 mvd 信息量. 将

式(9)、(10)代入式(8)可得头信息模型:

$$R(h) = \alpha' \cdot N_B + \theta \quad (11)$$

其中 $\alpha' = \alpha \cdot (1 + \beta(mvd))$. 假定 $\beta(mvd)$ 稳定, 用已编码相同类型的平均值预测, 式(11)可进一步简化为头信息与编码块数量之间的线性关系.

3.3 时空耦合关系

根据文献[15]的失真模型, 残差信息与运动信息在各种模式下的关系如图 3 和图 4 所示, 横坐标为宏块分块数 N_i , 纵坐标为头信息、残差信息和失真归一化值, 即当前值与平均值之间的比值. 该实验测试在量化参数为 28 时对多种视频序列进行编码, 随机选取 football (QCIF) 某个宏块的编码结果, 显示了残差信息与头信息存在耦合关系, 分块模式越细, 编码块数越多, 运动信息则越大; 反之亦然, 失真与头信息的关系类似.

在实际的编码器中, 假定残差服从拉氏分布, 求解率失真优化问题^[15]. 残差信息由残差的统计特性和量化参数决定 $R(e | \sigma, Qp)$, σ 为残差的方差; 同样, 失真函数可表示为 $D(\sigma, Qp)$, 残差 RDO 代价为: $D(\sigma, Qp) + \lambda R(e | \sigma, Qp)$, 通过 λ 参数将率与失真量纲统一;

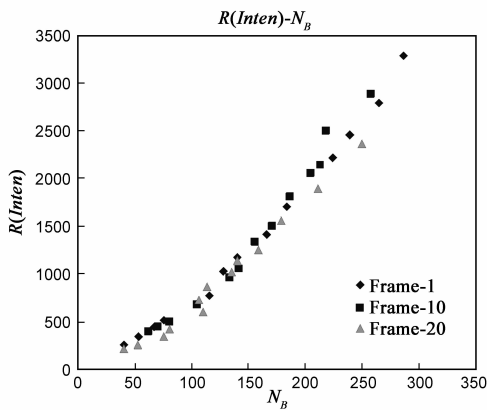
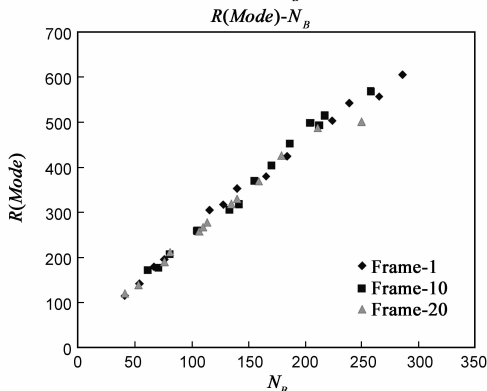
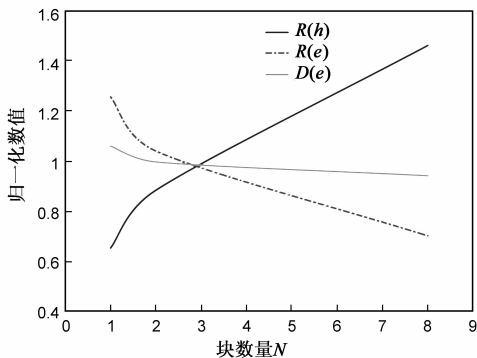
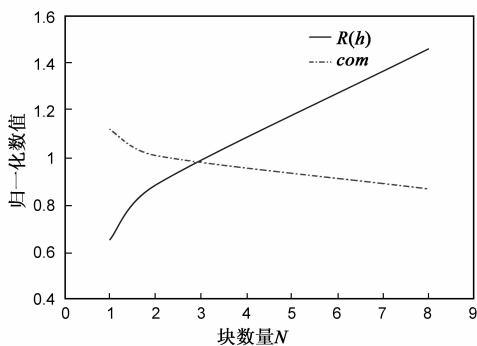
(a) $R(\text{Inter})-N_b$ 关系图(b) $R(\text{Mode})-N_b$ 关系图

图2 Foreman视频序列关系图

图3 $R(h)-R(e)-D(e)$ 关系图图4 $R(h)-com$ 关系图

运动信号包括了运动矢量、模式和其他辅助信号,这些信号的度量用信息的方式表示,即时间复杂度;残差信号包括残差码率和残差失真,用 $com(\sigma, Qp) = D(\sigma, Qp) + \lambda R(e | \sigma, Qp)$ 来表示空间残差复杂度.图4显示空间复杂度和时间复杂度相对于分块数量存在时空耦合关系.

本文按照分块数进行分类,讨论时空耦合关系在RDO模式选择中的影响.参照JM模型^[13]中按照分块数从小到大的编码顺序,即 $N_i < N_{i+1}$,当前模式的代价为: $J_i = [D_i(\sigma_i, Qp) + \lambda R_i(e | \sigma_i, Qp)] + \lambda R_i(h)$,下一层模式的代价为: $J_{i+1} = [D_{i+1}(\sigma_{i+1}, Qp) + \lambda R_{i+1}(e | \sigma_{i+1}, Qp)] + \lambda R_{i+1}(h)$;根据耦合关系,在计算 $i+1$ 模式代价之前,无法判断结果.随着 N 的不断增加, $R(h)$ 不断增加,而 $R(e)$ 不断减少,达到 $R(e) = 0$ 时, N 的增加将不能换取 $R(e)$ 减少,耦合关系达到极限条件.利用该极限作为进一步处理的条件,分为两种情况 $R(e) = 0$ 和 $R(e) \neq 0$; (1) 当 $R(e) \neq 0$ 时,无法确定 i 与 $i+1$ 模式代价的大小,证明请参见附录A; (2) 当 $R(e) = 0$ 时,且预测值相同 $\sigma_{i+1} = \sigma_i$,则 $D_{i+1}(\sigma_{i+1}, Qp) = D_i(\sigma_i, Qp)$;由式(11)可知 $R_{i+1}(h) > R_i(h)$,代入(1)可知 i 模式为最优模式;当 $R(e) = 0$,预测值不同 $\sigma_{i+1} \neq \sigma_i$,如 $J_{i+1} - J_i < 0$,得到 $D_i - \lambda [R_{i+1}(h) - R_i(h)] > D_{i+1}$.利用 i 模式的分块数和运动矢量信息可得到 $i+1$ 模式的头信息的预测值 $R'_{i+1}(h)$,利用同一种最佳模式的平均失真作为当前模式的失真 D'_{i+1} ,得到阈值 $T_1 = D'_{i+1} + \lambda [R'_{i+1}(h) - R_i(h)]$.如 $D_i > T_1$,则 $J_{i+1} - J_i < 0$;否则, $J_{i+1} - J_i > 0$.

3.4 快速模式选择算法

本文将块大小同为 16×16 的 skip 和 16×16 模式归为一类,用帧内模式表示相对于 8×8 更小的分块模式;将所有模式分为4个层次: L_1 慢运动(skip或 16×16), L_2 中速运动(16×8 或 8×16), L_3 快速运动(8×8 [$8 \times 8, 8 \times 4, 4 \times 8, 4 \times 4$]), L_4 高速运动或高纹理区域(Intra4 $\times 4$, Intra16 $\times 16$).参照JM模型^[13]中的编码顺序,逐层编码,并定义每一层的提前中止阈值.帧间编码模式提前中止包括以下3种:

3.4.1 16×16 模式与 Skip 模式中中止阈值

在 L_1 层中,先对 16×16 宏块进行处理,得到运动估计和模式编码,并判断是否满足中止条件,以进一步处理其他模式;中止条件分为以下几种:

(1) 当 $R(e) = 0, mvd = 0$ 时,skip模式的率失真代价:

$$J_{skip} = \sum_{i,j} |o(i,j) - p(i,j)|^2 + \lambda R_{skip}(h) \quad (12)$$

16×16 模式的最小率失真代价:

$$J_{16 \times 16} = \sum_{i,j} |o(i,j) - p(i,j)|^2 + \lambda R_{16 \times 16}(h) \quad (13)$$

由于 $R(h_{16 \times 16}) > R_{skip}(h)$, 因此, $J_{16 \times 16} > J_{skip}$, 最佳模式为 Skip 模式。

(2) 当 $R(e) = 0, mvd \neq 0$ 时, 可得到, $J_{16 \times 16} - J_{skip} < \frac{\Delta\sqrt{\lambda}}{N}(N\sqrt{\lambda} + \frac{\Delta\sqrt{\lambda}}{N} - 2SAD_{skip})$, 该不等式的证明参见附录 B. 其中: $\frac{\Delta\sqrt{\lambda}}{N} > 0$, 因此, $J_{16 \times 16} - J_{skip}$ 的正负值与 $(N\sqrt{\lambda} + \frac{\Delta\sqrt{\lambda}}{N} - 2SAD_{skip})$ 一致, 即 $(N\sqrt{\lambda} + \frac{\Delta\sqrt{\lambda}}{N} - 2SAD_{skip})$ 决定 $J_{16 \times 16} - J_{skip}$ 的大小. 当 $T_0 = (N\sqrt{\lambda} + \frac{\Delta\sqrt{\lambda}}{N})/2 < SAD_{skip}$, 则 $J_{16 \times 16} - J_{skip} < 0$, 选出 L_1 层最佳模式 $Mode_{L1}$ 并进入下一层编码处理; 当 $T_0 = (N\sqrt{\lambda} + \frac{\Delta\sqrt{\lambda}}{N})/2 > SAD_{skip}$, 则 $J_{16 \times 16} - J_{skip} > 0$, 选择最佳模式并退出。

(3) 当 $R(e) \neq 0$ 时, 直接进入下一层编码处理。

3.4.2 帧间运动估计模式逐层中止阈值

在 L_2 层中, 对 16×8 和 8×16 模式进行 RDO 处理, 得到本层的最佳编码模式, 并与 L_1 的最小代价 J_{L1} 对比, 确定 L_2 层的最佳编码模式 $Mode_{L2}$ 、失真 D_{L2} 和代价 J_{L2} . 根据时空耦合关系, 确定以下 3 种条件下的处理结果:

(1) 当 $R(e) = 0, mw_{L2} = mw_{L1} = 0$ 时, 得到 $D_{L1} = D_{L2}$, $J_1 = D_{L1} + \lambda R_{L1}(h)$, 以及 $J_2 = D_{L2} + \lambda R_{L2}(h)$; 由 $R_{L2}(h) > R_{L1}(h)$ 可得 $J_{L1} - J_{L2} < 0$, 最佳模式为 $Mode_{L1}$. L_2 与 L_3 层之间最佳模式的选择处理一样。

(2) 当 $R(e) = 0, mw_{L1} \neq mw_{L2}$ 时, 该条件同样适用于 L_2 与 L_3 层之间最佳模式的选择, 因此, 用当前层 i 与下一层 $i+1$ 层表示层间模式选择的关系, 预测阈值 T_1 . 如果 $D_i > T_1$, 则提前中止; 否则, 继续下一层次的检查, 直到帧间模式搜索完毕。

(3) 当 $R_e \neq 0$ 时, 直接进入下一层编码处理。

3.4.3 帧间与帧内模式中中止阈值

平均内容复杂度^[14]表示为: $\overline{Com}_{Intra} = \frac{1}{M \times N} \sum_{i,j}^{N,M} (|o_{i,j} - o_{i+1,j}| + |o_{i,j} - o_{i,j+1}|)$, M 和 N 代表宏块的水平和垂直的像素尺寸. 内容复杂度与码率存在线性关系 $R_{Intra} = \gamma \overline{Com}_{MB}$, 因此, 可得帧内平均内容复杂度 $\overline{Com}_{MB} = R_{Intra}/\gamma$. 同理, 帧间运动估计最佳编码模式, 其编码信息为 R , 平均内容复杂度为: $\overline{Com}_{Intra} = T_2 = \frac{\lambda \cdot R}{M \times N}$, 对比帧间与帧内内容复杂度的大小, 如 $\overline{Com}_{Intra} < T_2$, 则进行帧内编码处理. 否则, 提前中止模式选择编码。

3.5 算法流程

本文建议的模式选择逐层中止算法的步骤如下:

(1) 计算 16×16 模式的代价函数, 得到 16×16 的

代价; 判断残差信息是否为零, 如为非零进入下一层模式计算(4) 否则, 计算阈值 T_0 , 进入(2);

(2) 当前 16×16 的 mvd 如为 0, 最佳模式为 skip 模式, 并中止退出; 否则, 进入(3);

(3) 如 $SAD_{skip} < T_0$, $Mode_{L1}$ 为选择最佳模式并中止退出; 否则, 进入(4)。

(4) 计算 16×8 和 8×16 的代价, 比较两种模式与上一层模式的代价, 选出该层最小的代价 J_{L2} 。

(5) 如 $R_{L2}(e) = 0, mw_{L2} = mw_{L1} = 0$, $Mode_{L1}$ 为最佳模式并中止退出; 否则, 进入(6)。

(6) 计算阈值 T_1 , 如 $D_{L1} < T_1$, $Mode_{L2}$ 为最佳模式并中止退出; 否则, 进入(7)。

(7) 计算 8×8 的代价, 如 $R_{L3}(e) = 0, mw_{L3} = mw_{L2} = 0$, $Mode_{L2}$ 为最佳模式并中止退出; 否则, 进入(8)。

(8) 更新阈值 T_1 , 如果 $D_{L2} < T_1$, $Mode_{L2}$ 为最佳模式并中止退出; 否则, 进入(9)。

(9) 计算 \overline{Com}_{Intra} , 得到 T_2 , 如 $\overline{Com}_{Intra} > T_2$, $Mode_{L3}$ 为最佳模式并中止退出; 否则, 进入帧内编码模式。

(10) 帧内编码, 选择最佳的编码模式。

4 实验结果分析

4.1 实验环境

本文算法在 JVT 参考软件 JM10.2 平台^[13]上实现, 对不同的运动激烈程度的 QCIF 和 CIF 格式的视频序列进行了测试, 视频序列的分类为: 低运动或低空间复杂度视频序列(Akiyo), 中等空间复杂度或运动程度视频序列(Foreman, News)和高空间复杂度和快速运动视频序列(Coastguard, Flower). 运行仿真实验的计算机的配置为: Intel 2.0 GHz 双核 CPU, 2GB DDR III 内存, 操作系统为 Windows XP SP2. 参考软件的测试条件如下: 编码结构采用 IPP 结构, 参考帧数选择为 1, RDO 最优化编码, 运动估计搜索范围为 16, 量化参数分别为: 24, 28, 32, 36, 40.

4.2 结果分析与比较

本文方法从三个方面比较编码性能: 图像质量、输出码率、编码时间. 均利用 JM 模型中的全搜索 RDO 编码方法作为参考, 最终得到平均亮度偏差 Δ PSNR(dB), 码率偏差 Δ BR(%) 和总编码时间节省 Δ TS(%).

表 1 显示了本文方法、Choi^[9]和 Zeng^[10]提出的快速模式选择算法相对于 JM 穷举方法的比较结果. 表中列出了 24, 28 和 32 时结果, 其他参数类似, 相对于 JM 全搜索方法的性能, Choi^[9]方法平均 PSNR 值降低了 0.03dB, 平均码率降低了 0.6%, 平均编码时间减少了 38.01%, 编码性能主要决定于 skip 模式和帧内模式的比例; Zeng^[10]方法利用时空相关性, 提高了预测精度, 在保证图像质量的前提下, 大大降低了编码时间, 平均

表 1 本文算法、Choi 等^[9]和 Zeng 等^[10]算法的比较

视频序列	AM	Choi's algorithm			Zeng's algorithm			Proposed algorithm		
	Qp	Δ PSNR(dB)	Δ BR(%)	Δ TS(%)	Δ PSNR(dB)	Δ BR(%)	Δ TS(%)	Δ PSNR(dB)	Δ BR(%)	Δ TS(%)
Akiyo(QCIF)	24	-0.01	-1.81	-59.45	-0.07	-0.34	-71.75	-0.01	-0.64	-63.15
	28	-0.05	-2.29	-57.59	-0.07	0.12	-72.24	0.00	-2.40	-61.18
	32	-0.01	-0.97	-57.25	-0.06	0.32	-73.41	0.04	-0.79	-60.38
Foreman(QCIF)	24	0.00	-0.12	-34.66	-0.07	0.30	-55.12	-0.01	0.12	-54.35
	28	0.03	0.04	-33.87	-0.09	0.27	-56.05	-0.02	0.40	-50.43
	32	-0.05	-0.46	-35.68	-0.10	0.36	-57.37	0.02	0.50	-46.69
News(QCIF)	24	-0.04	-0.50	-58.09	-0.05	0.61	-65.49	-0.05	-0.17	-64.00
	28	-0.07	-0.75	-56.59	-0.06	0.22	-66.64	-0.08	-0.44	-60.87
	32	-0.04	-0.35	-55.86	-0.05	0.01	-66.92	-0.06	-0.39	-58.55
Foreman(CIF)	24	-0.03	-0.38	-42.42	-0.07	0.41	-54.57	-0.03	-0.33	-55.04
	28	-0.03	-0.43	-42.26	-0.09	0.57	-56.98	-0.03	-0.39	-52.73
	32	-0.08	-0.57	-44.10	-0.13	0.42	-60.31	-0.01	0.02	-51.68
Coastguard(CIF)	24	0.00	-0.14	-24.47	-0.02	0.30	-50.28	-0.01	-0.04	-54.65
	28	-0.01	-0.17	-22.03	-0.02	0.34	-50.62	-0.02	-0.22	-37.11
	32	-0.03	-0.70	-26.62	-0.03	0.25	-52.52	-0.04	-0.44	-36.96
Flower(CIF)	24	-0.01	-0.04	-44.59	-0.02	0.08	-56.82	0.01	-0.05	-64.24
	28	-0.01	0.07	-33.11	-0.02	0.07	-56.95	0.00	0.01	-58.30
	32	-0.02	-0.29	-32.27	-0.02	0.16	-57.01	0.00	-0.55	-46.36

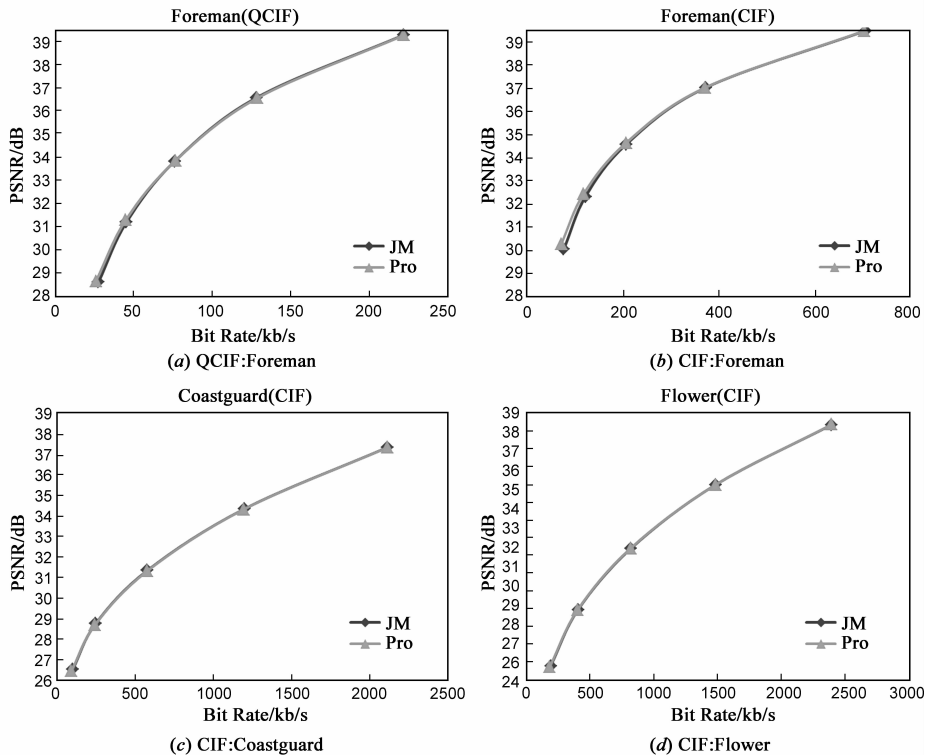


图5 JM和本文方法的率失真曲线

PSNR 值降低了 0.06, 平均码率增加了 0.17%, 平均编码时间减少了 52.76%, 节省的编码时间比例较稳定. 本文方法采用了逐层中止结构, 因此, 带来了更高的编码增益, 平均编码时间降低了 45.79%, 平均 PSNR 值降低了 0.01dB, 平均码率降低了 0.8%. 从以上数据可见, 不论是编码时间, 还是率失真性能上, 该方法整体性能优于 Choi^[9]方法, 与 JM 穷举搜索方法基本一致, 编码时间最大节省 64.24%. 对于运动程度较平稳的视频序列, 如 Akiyo 序列, 由于 skip 模式占的比重较大, 因此, 本文方法的增益较小. 而对于运动剧烈程度较高的序列, 本文方法带来的编码时间节省较大, 达到了 64.24%; 本文方法的率失真性能明显优于 Zeng^[10]方法, 但是, 需要增加检验更多的编码模式, 达到非耦合关系状态, 平均编码时间较 Zeng^[10]方法多了 6.97%, 而较 Choi^[9]方法减少了 7.78%. 此外, 图 5 列出了本文方法与 JM 方法的 R-D 性能比较, 本文方法的曲线几乎与 JM 方法曲线重叠, 率失真性能基本不变.

以上的实验结果显示, 本文方法在保持率失真效果的同时有效的降低了计算复杂度, 主要归结为本方法有效利用已编码模式信息, 并结合时空耦合关系, 由时空内容复杂度和量化参数确定中止阈值, 提前确定最佳编码模式, 节省了计算复杂度.

5 结论

在本文中, 结合时空复杂度的耦合关系, 利用率失真最优化理论, 提出了一种分层提前中止模式选择方法. 仿真实验结果表明, 该方法的整体性能明显优于现有方法, 相对于 JM 穷举方法, 平均降低了差不多 45.79% 计算复杂度, 同时保持了 PSNR 值和输出码率基本不变, PSNR 值降低 0.01dB, 码率平均降低了 0.8%. 算法还利用了头信息的编码规律, 结合残差的方差以及量化参数等决定因素, 根据当前编码模式的已编码信息, 自适应的调整中止阈值, 保证了 R-D 特性与穷举方法基本一致.

附录:

A. 当 $R(e) \neq 0$ 时, 无法确定层间代价的大小

由于 $N_i < N_{i+1}$, 残差方差的关系为: $\sigma_{i+1} \leq \sigma_i$; 可得: $R_{i+1}(e) < R_i(e)$, $D_{i+1}(e) < D_i(e)$ 和 $R_i(h) < R_{i+1}(h)$. 代入(1)可得 J_i 和 J_{i+1} , 差值: $J_i - J_{i+1} = \Delta D + \lambda \Delta R(e) - \lambda \Delta R(h)$, 其中, $\Delta D = D_i(e) - D_{i+1}(e)$, $\Delta R(e) = R_i(e) - R_{i+1}(e)$, $\Delta R(h) = R_i(h) - R_{i+1}(h)$; 因为 $\Delta D > 0$, $\Delta R(e) > 0$, $\Delta R(h) < 0$, 无法判断 J_i 和 J_{i+1} 的大小.

B. L_1 层的 $R(e) = 0$, $mvd \neq 0$ 时的推导

因为模式 16×16 运动估计的最小代价为: $SAD_{16 \times 16}(s, p(mv)) + \sqrt{\lambda} \cdot R_{16 \times 16}(mvd) = \text{Min}\{SAD(s, p(mv)) +$

$\sqrt{\lambda} \cdot R(mvd)\}$, 其中 $mvd \neq 0$, $R_{16 \times 16}(mvd) > R_{skip}(0)$, 因此:

$$SAD_{16 \times 16}(s, p(mv)) < SAD_{skip}(s, p_{skip}(mv)) \quad (B1)$$

由式 (B1) 可得: $\sum_{i,j} |s(i, j) - p(i, j)| < \sum_{i,j} |s(i, j) - p_{skip}(i, j)|$, 可推出:

$$SSD_{16 \times 16}(s, p(mv)) = \sum_{i,j} |s(i, j) - p(i, j)|^2 < SSD_{skip}(s, p_{skip}(mv)) = \sum_{i,j} |s(i, j) - p_{skip}(i, j)|^2$$
 从而, 模式选择的代价函数为:

$$J_{16 \times 16} = \sum_{i,j} |o(i, j) - p(i, j)|^2 + \lambda(R_{16 \times 16}(h))$$

$$J_{skip} = \sum_{i,j} |o(i, j) - p'(i, j)|^2 + \lambda(R_{skip}(h))$$
 由式(2)和 (B1)得到:

$$SAD_{16 \times 16} < SAD_{skip} - \sqrt{\lambda} \Delta = \sum_{i,j} |x_{i,j}| - \sqrt{\lambda} \Delta \quad (B2)$$

其中, $R_{16 \times 16}(mvd) - R_{skip}(0) = \Delta$, $o(i, j) - p'(i, j) = x_{i,j}$.

将式 (B2) 转换为一维方式:

$$SAD_{16 \times 16} < \sum_{i=1}^N |x_i| - \sqrt{\lambda} \Delta = \sum_{i=1}^N [|x_i| - \frac{\sqrt{\lambda} \Delta}{N}], N = 256.$$

$$\text{因为: } 0 < \sum_{i=1}^N [|x_i| - \frac{\sqrt{\lambda} \Delta}{N}] \leq \sum_{i=1}^N \left| x_i - \frac{\sqrt{\lambda} \Delta}{N} \right|,$$

$$\text{所以 } SAD_{16 \times 16} = \sum_{i=1}^N |x'_i| < \sum_{i=1}^N \left| x_i - \frac{\lambda \Delta}{N} \right|$$

其中, $o(i, j) - p(i, j) = x'_{i,j}$, x'_i 为 $x'_{i,j}$ 的一维形式.

$$\text{则: } SSD_{16 \times 16} = \sum_{i=1}^N |x'_i|^2 < \sum_{i=1}^N \left| x_i - \frac{\sqrt{\lambda} \Delta}{N} \right|^2$$

从而:

$$SSD_{16 \times 16} - SSD_{skip} = \sum_{i=1}^N |x'_i|^2 - \sum_{i=1}^N |x_i|^2 < \frac{\Delta \sqrt{\lambda}}{N} (\frac{\Delta \sqrt{\lambda}}{N} - 2SAD_{skip}) \quad (B3)$$

因此,

$$J_{16 \times 16} - J_{skip} = SSD_{16 \times 16} - SSD_{skip} + \lambda(R_{16 \times 16}(h) - R_{skip}(h))$$

其中, SSD_{skip} 已知, 得:

$$SSD_{16 \times 16} - SSD_{skip} < \max(SSD_{16 \times 16}) - SSD_{skip} = \max(SSD_{16 \times 16} - SSD_{skip})$$

由式 (B3) 可得:

$$J_{16 \times 16} - J_{skip} < \frac{\Delta \sqrt{\lambda}}{N} (N \sqrt{\lambda} + \frac{\Delta \sqrt{\lambda}}{N} - 2SAD_{skip})$$

参考文献

- [1] Recommendation ITU-T H.264|ISO/IEC 14496-10:2009. Advanced Video Coding for Generic Audio-visual Services[S].
- [2] T Wiegand, G J Sullivan, et al. Overview of the H.264/AVC video coding standard[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2003, 13(7): 560 - 576.
- [3] T Zhao, et al. Fast mode decision based on mode adaptation

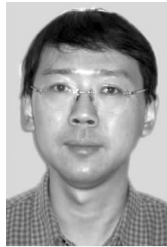
- [J]. IEEE Transactions on Circuits and Systems for Video Technology, 2010, 20(5): 697 – 705.
- [4] S H Ri, Y Vatis, et al. Fast inter-mode decision in an H. 264/AVC encoder using mode and lagrangian cost correlation[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2009, 19(2): 302 – 306.
- [5] Z Liu, L Shen, et al. An efficient intermode decision algorithm based on motion homogeneity for H. 264/AVC [J]. IEEE Transactions on Circuits and Systems for Video Technology, 2009, 19(1): 128 – 132.
- [6] D Wu, F Pan, et al. Fast intermode decision in H. 264/AVC video coding [J]. IEEE Transactions on Circuits and Systems for Video Technology, 2005, 15(7): 953 – 958.
- [7] Z Xie, Y Liu, et al. A general method for detecting all-zero blocks prior to DCT and quantization [J]. IEEE Transactions on Circuits and Systems for Video Technology, 2007, 17(2): 237 – 241.
- [8] A C W Yu, G R Martin, et al. Fast inter-mode selection in the H. 264/AVC standard using a hierarchical decision process [J]. IEEE Transactions on Circuits and Systems for Video Technology, 2008, 18(2): 186 – 195.
- [9] I Choi, et al. Fast coding mode selection with rate-distortion optimization for MPEG-4 part-10 AVC/H. 264 [J]. IEEE Transactions on Circuits and Systems for Video Technology, 2006, 16(12): 1557 – 1561.
- [10] H Zeng, C Cai, et al. Fast mode decision for H. 264/AVC based on macroblock motion activity [J]. IEEE Transactions on Circuits and Systems for Video Technology, 2009, 19(4): 491 – 499.
- [11] K I Lee, et al. Efficient Inter Mode Prediction Based on Model Selection and Rate Feedback for H. 264/AVC [J]. IEEE Transactions on Circuits and Systems for Video Technology, 2011, 21(6): 708 – 716.
- [12] 宋彬, 常义林, 李春林. H. 264 帧间预测模式的快速选择算法 [J]. 电子学报, 2007, 35(4): 697 – 700.
S Bin, C Yi-lin, L Chun-lin. Novel fast selecting inter prediction mode algorithm based H. 264 [J]. Acta Electronica Sinica, 2007, 35(4): 697 – 700. (in Chinese)
- [13] JM joint model [CP/OL]. <http://iphome.hhi.de/suehring/ttml/download/2007-08-07>.
- [14] B Yan, K R Sun. Joint complexity estimation of I-frame and P-frame for H264AVC rate control [J]. IEEE Transactions on Circuits and Systems for Video Technology, 2012, 22(5): 790 – 798.
- [15] X Li, et al. Laplace Distribution based lagrangian rate distortion optimization for hybrid video coding [J]. IEEE Transaction on Circuits and Systems for Video Technology, 2009, 19(2): 193 – 205.
- [16] Girod B. Rate-constrained motion estimation [A]. SPIE VCIP [C]. Chicago: SPIE, 1994. 1026 – 1034.

作者简介



何书前 男, 1978 年生于海南琼海. 中山大学信息科学与技术学院博士. 研究方向为视频编码压缩与传输.

E-mail: heshuqian05@126.com



倪江群 男, 1963 年生于江苏无锡. 中山大学教授, 博士生导师, 研究方向为多媒体信息处理和通信、信息隐藏和数字取证.