

帧间编码模式选择及其择优早期终止的 H. 264/AVC 快速算法

章国宝 李亮

(东南大学自动化学院, 南京 210096)

摘要 通过研究 H. 264/AVC 帧间编码中最佳模式分布, 提出了 3 种模式选择早期终止策略, 分别对 SKIP, $P16 \times 16$, $P16 \times 8$ 和 $P8 \times 16$ 这三类模式成为最佳编码模式进行判断, 适当地结束编码模式选择过程, 避免了对后面不可能模式率失真代价的计算, 从而减少了帧间编码的计算复杂度。实验结果表明, 该算法可以减少 H. 264/AVC 编码器 50% ~ 70% 的编码时间, 同时保持与全搜索算法基本一致的编码性能, 峰值信噪比 PSNR 平均下降 0.17 dB 左右, 编码比特率平均增加 2.42%。

关键词 视频编码 H. 264/AVC 帧间编码 快速编码模式选择 编码器优化

中图法分类号: TP391 **文献标识码:** A **文章编号:** 1006-8961(2009)01-0059-06

A Fast Coding Algorithm for H. 264/AVC Based on Best Intermode Decision and Early-terminated Strategies

ZHANG Guo-bao, LI liang

(School of Automation, Southeast University, Jiangsu Nanjing 210096)

Abstract According to the distribution of the best inter coding modes, we give three early-terminated strategies for deciding whether one of the SKIP, $P16 \times 16$, $P16 \times 8$ or $P8 \times 16$ can be selected as the best inter coding mode. If the conditions can be satisfied, the computation of RD cost of the rest mode can be saved. Experimental results show that the proposed algorithm can save about 50% to 70% time compared with a full-searching algorithm in H. 264/AVC, meanwhile the decreasing of PSNR is about 0.17dB and the increases of bit-rate is about 2.42%.

Keywords video coding, H. 264/AVC, inter prediction, fast mode decision, encoder optimization

1 引言

H. 264/AVC 是由 ISO/ICE 和 ITU-T 组成的联合视频编码小组 JVT 共同推出的最新视频编码国际标准^[1]。由于它采用了一系列先进编码技术, 如帧内预测、可变块大小运动估计/补偿、多参考帧、1/4精度运动矢量、 4×4 整数 DCT 变换、环路去块效应滤波以及 CABAC(基于上下文的自适应二进制算术编码)等, 极大地提高了视频编码效率, 与以前的视

频编码标准 MPEG-2 相比, 在相同重建图像质量条件下, H. 264/AVC 所需要的比特率仅为 MPEG-2 的三分之一左右^[2]。然而这些编码效率的取得是以增加编码器的计算复杂度为代价的。H. 264/AVC 为了适应不同特征的视频序列采用了复杂的帧内和帧间预测技术, 为宏块提供了许多可选编码模式, 并采用基于率失真优化的方法来选择宏块的最佳编码模式。对于帧间编码来讲, 由于率失真代价函数的计算涉及许多耗时操作, 使得帧间编码模式选择部分消耗了整个编码过程一半以上的编码时间。因此,

收稿日期: 2007-11-29; 改回日期: 2008-09-05

第一作者简介: 章国宝(1965 ~), 男, 副教授, 院长助理, 主任。1999 年于东南大学获模式识别与智能系统专业工学博士学位。研究方向为图像处理、模式识别、智能系统。承担 863 等课题 20 多项, 公开发表论文 30 余篇。E-mail: guobaozh@seu.edu.cn

针对这一部分进行算法优化可以显著降低 H. 264/AVC 编码器的计算复杂度,对 H. 264/AVC 视频编码标准的实际应用具有重要意义。

目前针对帧间编码模式选择问题已经提出了很多快速算法,这些算法的基本思想就是根据视频序列的特征来减少当前宏块的可选编码模式,避免对其他编码模式计算率失真代价函数,从而降低帧间编码计算量。文献[3]中提出的算法分别根据宏块空间上的一致性和时间上的相似性,来选择大块预测模式编码,跳过对小块编码模式的检查。文献[4]利用几种编码模式率失真代价值的大小和相互关系来决定宏块尺寸的划分,避免尝试计算所有模式的率失真代价。文献[5]提出了一种 SKIP 模式早期终止策略,对于满足条件的宏块直接用 SKIP 模式编码,避免计算任何率失真代价。本文算法利用最佳编码模式分布不均匀的特点,分别针对 SKIP, $P16 \times 16$, $P16 \times 8$ 和 $P8 \times 16$ 这 3 类模式提出了 3 种模式选择早期终止条件,通过适当地结束编码选择过程来避免对后面多余模式的率失真代价计算,在几乎不损失编码性能的前提下极大地减小了编码器的计算复杂度。

2 H. 264/AVC 帧间预测编码

H. 264/AVC 中的 P 帧编码一共有 7 种编码模式,分别是:SKIP、 $P16 \times 16$ 、 $P16 \times 8$ 、 $P8 \times 16$ 、 $P8 \times 8$ 、I16MB 和 I4MB。为了对每个宏块选择最佳的编码模式,H. 264/AVC 采用基于率失真优化的方法计算所有编码模式的率失真代价,从中选择代价最小的模式作为宏块的最佳编码模式,具体过程如下:

(1) 计算运动估计和模式选择的 Lagrange 系数 λ_{MOTION} 和 λ_{MODE} 。

(2) 对以上各种块尺寸做运动估计,根据运动估计率失真代价公式求出各种块尺寸的最佳运动矢量。其中对于 $P8 \times 8$ 模式,要对 16×16 块中的每个 8×8 块,从 8×8 、 8×4 、 4×8 、 4×4 中选择最佳划分。

(3) 对各种模式进行编码得到宏块的重建误差和编码比特率,据此计算各自的率失真代价。

(4) 选择率失真代价最小的模式作为宏块的最佳帧间编码模式。

从上面的过程可以看出,要完成对一个宏块的帧间编码需要进行运动补偿预测、DCT 变换量化、

反量化、反 DCT 变换、熵编码等复杂操作。由于 H. 264/AVC 中引入了可变块尺寸的运动估计、多参考帧、1/4 像素精度运动矢量等精确预测技术,更是加剧了运动估计的计算复杂度。有关统计表明 H. 264/AVC 编码器的计算复杂度是 MPEG-2 编码器的 5~10 倍,其中运动估计和编码模式率失真代价的计算占了 H. 264/AVC 编码器计算量的绝大部分。本文就是针对这一问题提出快速帧间编码模式选择算法,从而降低编码器的计算复杂度。

3 帧间编码模式快速选择算法

3.1 帧间编码模式分布

由于视频序列具有很强的时间和空间相关性,例如视频中存在大量的静止背景、全局运动以及空间一致性区域,最佳编码模式并不是在这 7 种模式中均匀分布,往往 SKIP 模式和较大块尺寸的模式占了绝大多数比例。另外由于帧间预测的精确度比帧内预测高得多,P-Slice 中的宏块几乎很少采用 I4MB 和 I16MB 这两种帧内模式编码。图 1 是对标准测试序列进行编码(编码器为 Baseline profile,编码帧结构为 IPPPP,参考帧数目为 5,运动估计搜索范围为 16,量化参数为 28,使用率失真优化)得到的 P-Slice 中宏块的最佳模式分布。

从图 1 中可以看出,对于不同运动剧烈程度和纹理复杂度的视频序列,SKIP 模式和大尺寸宏块模式在最佳编码模式中占有绝大部分比例,而帧内预测模式 I4MB 和 I16MB 在 P-Slice 的宏块中几乎不使用。利用这种模式分布的差异,如果能够较早地判断出概率大的编码模式为最佳模式,将能够避免大量的其他模式的率失真代价计算。特别对于运动比较缓慢或者纹理简单的视频序列(如 News),它们的最佳模式绝大部分是 SKIP,如果能预先确定 SKIP 为最佳编码模式,将节省很多计算量。本文给出的快速帧间编码模式选择算法就是根据这样的模式分布规律,用 3 个阶段的早期终止策略来结束模式选择过程,尽可能避免对不可能编码模式的率失真代价计算。

3.2 SKIP 模式早期终止策略

P-Slice 中的 SKIP 宏块也称为拷贝宏块,它不需要对预测运动矢量误差和预测残差系数编码,因为它的运动矢量预测误差为 0,宏块预测残差经过 DCT 变换和量化后也全部为 0,SKIP 宏块的重建值

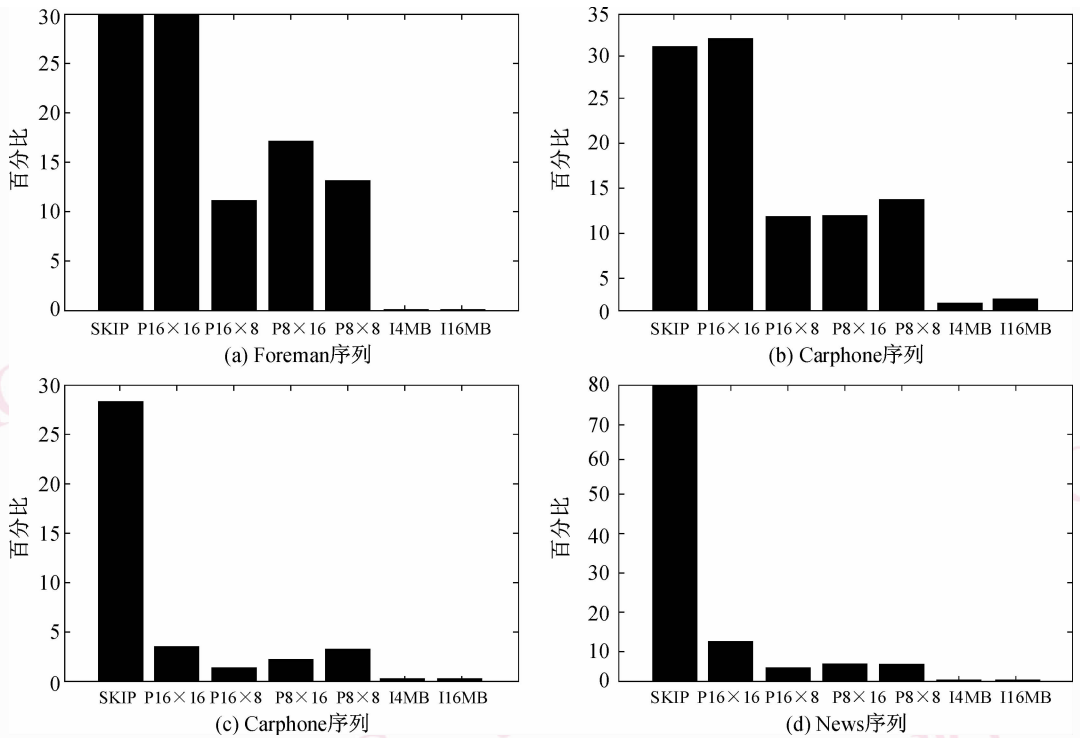


图 1 4 种标准序列的最佳编码模式分布

Fig. 1 The best mode code of four standard sequence distribution

就是预测运动矢量指向的前一参考帧中的宏块值。SKIP 宏块通常出现在静止背景区域或运动一致区域,它一般具有很高的预测精度并且不需要对运动矢量编码,由率失真代价的计算公式可推断最佳模式为 SKIP 的编码代价 RDcost 要比最佳模式为其他模式的 RDcost 要低,图 2 和图 3 是对 Carphone 序列用全搜索的率失真优化方法得到的最佳模式的 RDcost 统计情况:

从图 2 中可以看出最佳模式为 SKIP 的平均 RDcost 远小于其他模式的 RDcost,因此可以用 SKIP

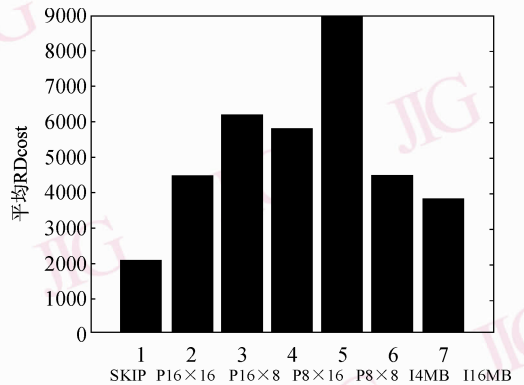


图 3 各种最佳模式的平均 RDcost 值

Fig. 3 The average RDcost of a variety of the best modes

模式的率失真代价为准则初步筛选出典型的 SKIP 模式,具体决策方法为:

(1) 计算 SKIP 模式的率失真代价 RDcost_{SKIP}。

(2) 如果 RDcost_{SKIP} < TH_{SKIP}, 最佳编码模式选为 SKIP 并结束模式选择过程; 否则进入下一阶段决策。

这里的 TH_{SKIP} 是根据实验得到的阈值,它一般取很小的值,目的是避免将适合其他模式编码的宏块错误地选为用 SKIP 编码,虽然会将真正的

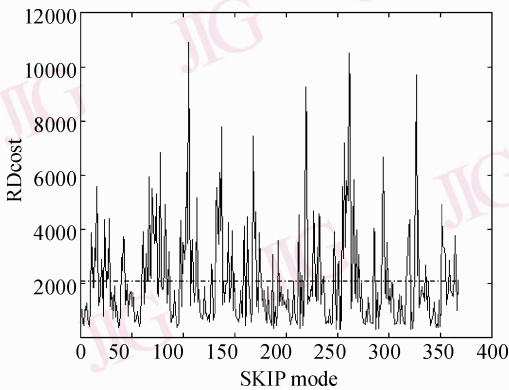


图 2 SKIP 为最佳模式的 RDcost 值

Fig. 2 RDcost when SKIP is the best mode

SKIP 模式排除在外,但本文给出的算法仍然可以在下一阶段的决策中被选择。由于 SKIP 模式率失真代价的计算不需要进行运动估计,DCT 变换及熵编码等耗时操作,它的计算复杂度比其他模式低得多,如果能在这个阶段终止模式选择将节省很多的编码时间,特别是对运动缓慢或纹理简单的视频序列效果更加明显。

3.3 P16 × 16 模式早期终止策略

从上节中的最佳编码模式分布图可以看出,最佳模式为 P16 × 16 仍然占有很大的比例,如果能在终止模式选择过程,也能避免大量无用模式的率失真代价计算。为了保证选择的最佳模式的准确性,首先分析最佳模式为大块模式(SKIP, P16 × 16)的宏块特征,图 4 是 Paris 序列某一帧中宏块的编码模式划分图(采用基于 RDO 的全搜索方法得到的结果),可以看出以下两种类型的宏块适合用大块尺寸模式编码:

(1) 位于静止或运动较小区域的宏块,不管宏块的纹理是简单还是复杂。

(2) 位于整体运动区域的具有简单纹理的宏块。

P16 × 16 模式早期终止策略的思想就是判断当前宏块是否满足上面的两种特征之一,如果满足就结束模式选择过程,否则继续进入下一阶段的判断。

首先用 16 × 16 宏块的运动矢量来定义宏块的运动大小 $Motion_{16}$, $Motion_{16}$ 的值取 0 或 1, 分别表示轻微运动和剧烈运动,具体定义如下:

$$Motion_{16} = \begin{cases} 0 & |mv_x| \leq 2, |mv_y| \leq 2 \\ 1 & \text{其他} \end{cases} \quad (1)$$

16 × 16 宏块的纹理复杂度可以用宏块 DCT 变换后的 AC 系数能量之和 E_{AC} 来表示,即

$$E_{AC} = \sum_{u=0, v=0}^{15, 15} (F_{uv}^2) - F_{00}^2 \quad (2)$$

其中, F_{uv} 是宏块 DCT 变换后的系数, F_{00} 是系数中的直流分量。为了计算简单采用式(3)来求 AC 系数能量和:

$$E_{AC} = \sum_{x=0}^{N-1} \sum_{y=0}^{N-1} f(x, y)^2 - \frac{1}{N \times N} \left(\sum_{n=0}^{N-1} \sum_{m=0}^{N-1} f(x, y) \right)^2 \quad (3)$$

这样宏块的纹理复杂度就可以定义为

$$Text = \begin{cases} 0 & E_{AC} \leq TH_{AC} \\ 1 & \text{其他} \end{cases} \quad (4)$$

式中, TH_{AC} 为复杂度度量阈值,根据实验测定取值为 50 000。同时考虑到 P16 × 16 作为最佳模式一般

具有比较好的预测效果,具体表现就是它具有比其他帧间预测模式更小的 RDcost,因此可以选择一个较小的阈值 $TH_{P16 \times 16}$, 作为终止模式选择的另一个条件。

综上, P16 × 16 模式早期终止具体过程如下:

(1) 对 16 × 16 块做运动估计求出最佳运动矢量。

(2) 根据式(1)求出 $Motion_{16}$, 如果为 0 则执行步骤 5; 否则进入下一步。

(3) 根据式(4)求出 $Text$, 如果为 0 则执行步骤 5; 否则进入下一步。

(4) 计算 P16 × 16 的率失真代价, 如果小于 $TH_{P16 \times 16}$ 则执行步骤 5; 否则执行步骤 6。

(5) 从 SKIP 和 P16 × 16 中选择率失真代价最小的作为最佳编码模式并结束模式选择过程。

(6) 进入下一阶段的模式选择过程。

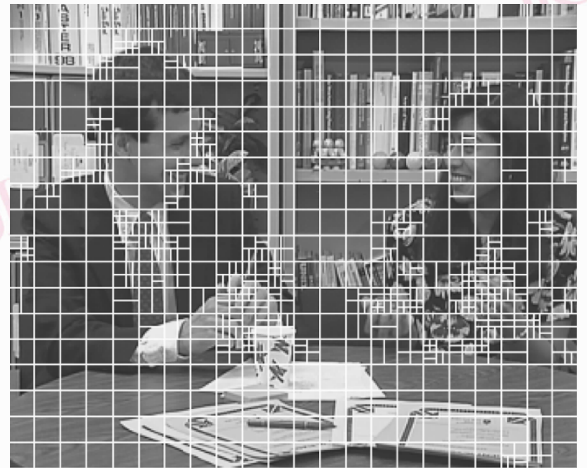


图 4 Paris 序列某帧图像的宏块编码模式

Fig. 4 Macroblock coding model of Paris sequence picture frame

3.4 P16 × 8 和 P8 × 16 模式早期终止策略

H. 264/AVC 为了提高宏块的预测精度采用了树型的宏块划分方式, P8 × 8 模式宏块的每个 8 × 8 子块又可以进一步划分为 8 × 8, 8 × 4, 4 × 8, 4 × 4 等 4 种形式, 以保证对具有复杂运动的视频序列进行精确的运动估计补偿, 从而提高编码效率。但这也导致了 P8 × 8 模式的率失真代价计算非常复杂, 消耗很多的编码时间。为此提出了 P16 × 8 和 P8 × 16 模式早期终止策略, 尽量避免计算 P8 × 8 的率失真代价。首先给出宏块采用 P8 × 8 模式编码要满足的条件即(1)宏块具有较大的运动;(2)宏块位于运动

物体的边界或宏块内的物体具体不同的运动方向。如果满足以上条件则表明宏块还需要进一步划分,否则结束模式选择过程,最佳帧间编码模式只在 SKIP, P16 × 16, P16 × 8, P8 × 16 中选择。

宏块的运动大小 $Motion_{P8 \times 8}$ 根据 16×8 和 8×16 两种尺寸块的运动矢量来确定, $Motion_{P8 \times 8}$ 等于 1 表示宏块具有较大的运动,具体定义如下式:

$$Motion_{P8 \times 8} = \begin{cases} 1 & (|mv_{P16 \times 8x}| > 3 \& \& |mv_{P16 \times 8y}| > 3) \parallel \\ & (|mv_{P8 \times 16x}| > 3 \& \& |mv_{P8 \times 16y}| > 3) \\ 0 & \text{其他} \end{cases} \quad (5)$$

式中, $mv_{P16 \times 8x}$ 和 $mv_{P16 \times 8y}$ 分别表示左右两个 16×8 块中运动矢量的 x 和 y 分量的最大值, $mv_{P8 \times 16x}$ 和 $mv_{P8 \times 16y}$ 分别表示上下两个 8×16 块中运动矢量的 x 和 y 分量的最大值。

如果宏块位于运动对象的边界或宏块内的运动对象具有不同的运动方向,则宏块更适合用更小的块尺寸做运动估计,并且各个小块都有不同的运动矢量,如果采用一致的运动矢量就会导致各个小块运动补偿预测误差有很大差异。下面分别将 P16 × 8 和 P8 × 16 模式的预测误差分为图 5 和图 6 所示的 4 组。

SE_{11}	SE_{12}
SE_{13}	SE_{14}

SE_{21}	SE_{22}
SE_{23}	SE_{24}

图 5 16×8 块的预测误差 图 6 8×16 块的预测误差

Fig. 5 16×8 forecast error Fig. 6 8×16 forecast error

定义它们各自的预测误差如式(6)所示:

$$SE_{ij} = \sum_{x=0}^7 \sum_{y=0}^7 |Org_{ij}(x,y) - Pred_{ij}(x,y)| \quad (6)$$

$i = 1, 2; j = 1, 2, 3, 4$

式中, $Org_{ij}(x,y)$ 是原始宏块在位置 (i,j) 处的像素值, $Pred_{ij}(x,y)$ 是预测宏块在位置 (i,j) 处的像素值, SE_{ij} 为子块预测误差绝对值的累计和,它的差异性可以反映各个子块运动的不同,定义 SSE_i 来度量各 SE_{ij} 之间的差异,

$$SSE_i = \sum_{j=1}^4 \left(SE_{ij} - \frac{1}{4} \sum_{k=1}^4 SE_{ik} \right)^2 \quad i = 1, 2 \quad (7)$$

这样宏块是否处在运动物体边界或内部是否有不同运动方向的对象,就可以根据 SSE_i 的大小来判断,具体判别方法如下:

$$Motion_C = \begin{cases} 1 & SSE_1 > TH_{Motion} \parallel SSE_2 > TH_{Motion} \\ 0 & \text{其他} \end{cases} \quad (8)$$

式中, TH_{Motion} 是根据实验得到的阈值,一般取 2 000, $Motion_C$ 等于 1 表示宏块内部有复杂的运动,需要进行小块尺寸的运动估计。综上所述, P16 × 8 和 P8 × 16 模式早期终止过程如下:

(1) 分别对 16×8 和 8×16 块做运动估计;

(2) 计算 $Motion_{P8 \times 8}$ 的值,如果为 0 则跳到第 4 步,否则执行下一步;

(3) 计算 $Motion_C$ 的值,如果为 1 则跳到第 5 步,否则执行下一步;

(4) 在 SKIP, P16 × 16, P16 × 8, P8 × 16 中选择率失真代价最小的模式作为最佳帧间编码模式并结束模式选择过程。

(5) 继续做 P8 × 8 模式的运动估计并计算其率失真代价,从 SKIP, P16 × 16, P16 × 8, P8 × 16, P8 × 8 中选择率失真代价最小的模式作为最佳帧间编码模式。

4 算法测试与分析

为了测试快速算法的性能,将算法在 H.264/AVC 参考软件 JM8.6 上实现^[6],并对各种格式的标准测试序列进行编码实验^[7],编码器的配置如下:帧结构为 IPPPP,编码帧数为 100/150,量化参数有 (28,32,36,40),参考帧数目为 5,运动估计搜索范围为 ±16,使用 Hadamad 变换和率失真优化,熵编码采用 CAVLC(基于上下文的自适应可变长编码),未使用快速运动估计。实验中对 QCIF 格式的标准测试序列 Foreman, Silent, News, Claire 和 CIF 格式的标准测试序列 Foreman, Stefan 进行了编码,结果如表 1 所示。其中, $\Delta PSNR$ 表示采用本文算法引起的 PSNR 的下降; bitrate 比特率, $\Delta Bitrate$ 表示采用本文算法引起的比特率增加在 JM8.6 比特率中所占的比例; TS 表示快速算法节省的编码时间。从表 1 中可以看出:对于 QCIF 格式中运动比较平缓的视频序列(如 Claire),本文快速算法平均可以减少 70.29% 编码时间,同时仅带来 0.14 dB 的 PSNR 下降

表 1 快速算法的编码性能

Tab. 1 Coding performance of fast algorithm

视频序列格式	视频序列	QP	Δ PSNR (dB)	Δ Bitrate (%)	TS (%)
QCIF	Foreman	28	-0.27	+1.43	-56.45
		32	-0.17	+0.95	-50.44
		36	-0.12	+1.32	-47.62
		40	-0.09	+2.41	-46.97
QCIF	Silent	28	-0.29	+2.76	-62.18
		32	-0.24	+1.92	-57.67
		36	-0.16	+2.03	-51.34
		40	-0.04	+5.76	-49.93
QCIF	News	28	-0.29	+2.94	-72.08
		32	-0.20	+2.09	-63.86
		36	-0.18	+1.85	-59.73
		40	-0.13	+1.74	-55.45
CIF	Claire	28	-0.16	+2.42	-71.11
		32	-0.15	+2.63	-73.90
		36	-0.03	+2.76	-64.00
		40	-0.22	+3.65	-72.13
CIF	Foreman	28	-0.31	+2.73	-61.85
		32	-0.14	+2.93	-54.86
		36	-0.09	+3.17	-55.44
		40	-0.02	+4.90	-51.82
CIF	Stefan	28	-0.19	+3.03	-59.46
		32	-0.19	+3.98	-57.12
		36	-0.17	+3.84	-56.16
		40	-0.13	+3.23	-52.52

和 2.87% 的比特率的增加;对运动剧烈的视频序列(如 Foreman),平均可以减少 50.37% 编码时间,同时 PSNR 的下降 0.16 dB,比特率增加 1.53%。图 7 给出了快速算法与全搜索算法编码性能的率失真曲线。可以看出,快速算法在基本不损失编码性能的前提下,能够降低编码器 50% ~ 70% 的编码时间。

5 结论

根据 H. 264/AVC 帧间编码中最佳模式分布的特征,提出了一种基于早期终止策略的帧间编码模式选择快速算法。实验结果表明,该算法对于各种

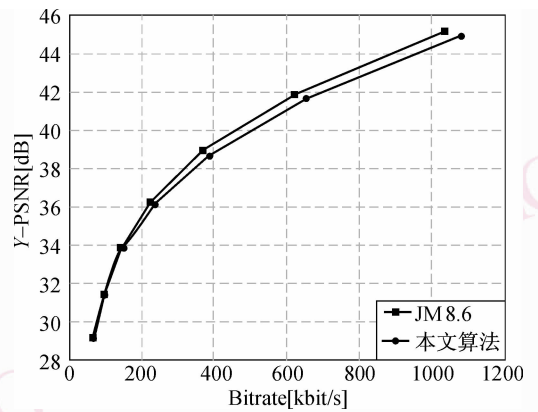


图 7 QCIF 序列 Foreman 的率失真曲线

Fig. 7 Foreman's rate-distortion curve of QCIF sequence

类型的视频序列在不同的比特率限制下,均可以有效降低编码器的复杂度,并保证与原来算法基本相同的编码性能,对 H. 264/AVC 视频编码标准的实际应用具有重要意义。

参考文献 (References)

- 1 ITU-T Rec. H. 2641 ISO/IEC 14496-10 AVC. Draft ITU-T Recommendation and Final Draft International Standard Joint Video Specification[S].
- 2 Wiegand T, Schwarz H, Joch A, et al. Rate constrained coder control and comparison of video coding standards[J]. IEEE Transactions on Circuits and System for Video Technology, 2003, 13(7):688-703.
- 3 Wu D, Pan F, Lim K P, 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.
- 4 Xu Li-dong, Lin Xing-gang. Fast mode decision for inter frames in H. 264/AVC[A]. In: Proceedings of IEEE International Symposium on Communications and Information Technology (ISCIT 2005) [C], Beijing, China, 2005, 1:433-436.
- 5 Lee Jeyun, Jeon Byeungwoo. Fast mode decision for H. 264[A]. In: Proceedings of the 2004 IEEE International Conference on Multimedia and Expo[C], Taipei, Taiwan, 2004; 1131-1134.
- 6 Joint Video Team (JVT) of ISO/ICE MPEG and ITU-T VCEG. JVT Reference Software version JM8.6 [CP/OL]. <http://iphome.hhi.de/suehring/tml/>, 2005-09-26/2006-03-01.
- 7 Choi I, Lee J, Choiand W, et al. Performance evaluation of fast mode Decision Algorithm for H. 264 [EB/OL]. http://wftp3.itu.int/av-arch/jvt-site/2004_10_Palma/JVT-MO13.doc, 2004-10-28/2006-03-01.