

H.264 中基于先验预测的帧间编码模式选择算法研究

千宗良 齐丽娜 朱秀昌

(南京邮电大学图像处理与图像通信重点实验室 南京 210003)

摘要 帧间编码的模式选择算法是 H.264 视频编码器中计算量较大的部分。该文首先简要介绍了 H.264 帧间多模式编码原理, 然后对帧间编码的最佳模式进行统计、分析, 提出了一种基于先验预测的帧间编码模式选择快速算法。仿真实验结果表明该文提出的快速模式选择策略可以在保证视频编码质量的同时, 很大程度上降低编码器的计算复杂度。

关键词 H.264, 运动估计, 模式选择, 率失真优化

中图分类号: TN919.8

文献标识码: A

文章编号: 1009-5896(2006)10-1883-05

Interframe Video Coding Mode Decision Algorithm Based on Prior Detection in H.264

Gan Zong-liang Qi Li-na Zhu Xiu-chang

(Image Processing & Image Comm. Key Lab, Nanjing Univ. of Posts and Telecomm., Nanjing 210003, China)

Abstract Interframe video coding mode decision algorithm is one of the most complex part in H.264 encoder. In this paper, the theory of variable block-size motion compensation in H.264 is given and the best mode statistical law of inter-frame video coding is analyzed at first. Then a novel fast mode decision algorithm based on prior detection is presented. The results of simulation test show that the proposed fast mode decision strategy can highly improve the efficiency of real time H.264 encoder while maintaining video coding quality.

Key words H.264, Motion estimation, Mode decision, Rate distortion optimization

1 引言

H.264^[1]是ISO/IEC(国际标准化组织 / 国际电子技术委员会)的 MPEG(联合运动图像专家组)和 ITU-T(国际电信联盟)联合成立的JVT(联合视频专家组)制定的一个新的国际视频压缩标准, 同时也是MPEG-4 的第 10 部分(Advance Video Coding)。

H.264 与以往的视频编码标准相比具有更好的压缩性能和面向网络传输的“友好性”。H.264 具有优秀的压缩性能是多种新技术所产生效果的积累所致^[2,3]。这些新技术包括: 多模式帧内预测、可变尺寸块的运动补偿、多参考帧的运动补偿、4×4 整数变换、高效的熵编码、环路滤波等。其中H.264 的运动估计和模式选择是编码器中最为复杂的编码模块之一。本文就H.264 中帧间模式选择算法进行了较为深入的分析, 提出一种基于先验预测的帧间编码模式选择快速算法。实验结果表明本文提出的快速算法可以在保证视频编码质量的同时, 大幅度降低编码器的复杂度和计算量。

2 H.264 标准中帧间多模式编码原理

H.264 标准的帧间预测可以采用多模式运动估计、1/4 像素精度运动补偿和可供选择的多参考帧, 对当前编码

帧进行精确预测, 提高压缩效率。在帧间预测编码时, 每个亮度宏块可划分成形状不等的区域, 作为运动描述区域。其划分方法如图 1 所示: 当宏块级运动估计时有 16×16, 16×8, 8×16 等 3 种模式; 当亚宏块级运动估计时, 可以进一步划分成 8×8, 8×4, 4×8 和 4×4 等 4 种模式。

由于H.264 标准中针对每个宏块可以提供多种编码模式, 所以在编码端就需要对编码的宏块选取最佳的编码模式, 以达到最佳压缩效果。在JM参考模型^[4]中提供了两种最佳模式选择算法: 基于失真的判决和基于率失真优化^[5](Rate Distortion Optimization, RDO)的判决。

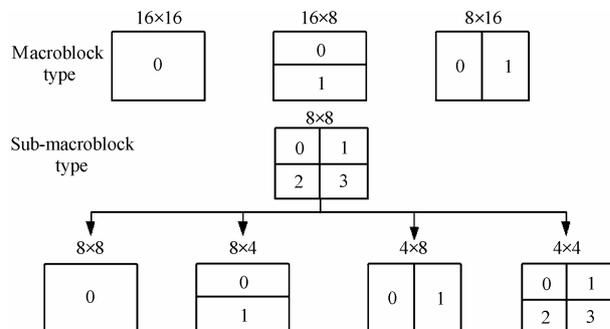


图 1 H.264 运动补偿的宏块划分模式
Fig.1 Segmentations of the macroblock for motion compensation in H.264

基于失真的判决准则在 JM 参考模型中的误差衡量函数为绝对误差和(Sum of Absolute Differences, SAD), 选取编码模式中绝对误差和最小的编码模式为最佳帧间编码模式, 如式(1)所示:

$$D_{SAD}(X_i, M) = \sum_{s \in X_i} |Y(s) - \hat{Y}_M(s)| \quad (1)$$

式(1)中 $Y(S)$ 表示原始点 S 的像素值, $\hat{Y}_M(S)$ 表示在编码模式 M 下, 重建图像点 S 的像素值。

基于率失真优化的判决的开销函数如式(2)所示:

$$J(M, \lambda_{MOTION}) = SAD(S, c(M)) + \lambda_{MOTION} \times R(M - P) \quad (2)$$

其中 $M = (m_x, m_y)^T$ 为模式 m 的运动矢量, $P = (p_x, p_y)^T$ 为模式 m 的运动矢量的预测值, λ_{MOTION} 为 Lagrange 乘数因子, SAD 函数同式(1), $R(M - P)$ 为模式 M 所需的 bit 数。此时在所有可供选择的编码模式中选取开销函数最小的编码模式作为最终编码模式。

基于率失真优化的模式判决在编码性能上比基于失真的模式判决更为出色。本文主要研究的是基于率失真优化的模式选择算法, 同样适用于失真模式判决。

JM 参考模型的率失真优化的模式选择算法步骤如下:

步骤 1 计算宏块级帧间编码模式的开销函数, 选择最佳模式。其中可选模式有 16×16, 8×16, 16×8;

步骤 2 计算亚宏块级帧间编码模式的开销函数, 选择最佳模式。其中可选模式有 8×8, 8×4, 4×8, 4×4;

步骤 3 计算帧内编码模式的开销函数, 选择最佳模

式, 其中可选模式有 Intra 16×16, Intra 4×4;

步骤 4 根据率失真准则, 在可供选择编码模式中选择率失真最小的编码模式为最终编码模式。

3 H.264 中帧间编码最佳模式统计规律分析

在 JM 参考模型中, 帧间最佳模式的选择算法非常复杂、耗时, 因此在实际应用中非常有必要对这个编码单元优化。本文从帧间编码最佳模式的统计规律出发, 获得足够的先验知识, 为模式选择快速算法提供依据。

3.1 宏观上的帧间编码最佳模式统计规律

本文根据 JVT 提供的测试代码 JM86 对不同视频序列宏块所采用的最佳模式进行统计, 统计结果如表 1 所示。可以发现无论是运动较为缓慢的视频序列如: Missa, Claire, 还是运动较为复杂的序列如: Foreman, Football, 这些序列的帧间编码最佳模式的统计规律都满足宏块级帧间预测模式最多, 其次是亚宏块级帧间预测模式, 帧内预测模式相对较少。

视频图像内容可以粗略的分成 3 种: 平坦背景区域、复杂背景区域和运动区域。通常情况下, 平坦背景区域在视频内容中占比较大的成分, 一般用 16×16 的帧间预测模式可以获得较好的效果; 而在运动区域情况较为复杂的情况下, 则需要使用更多的编码模式, 从中选择出最佳的模式进行编码。根据对不同视频序列的最佳编码模式的统计, 本文总结出在不同的视频内容下可选用的编码模式, 如表 2 所示。

表 1 QP=26 时, P 帧的宏块最佳模式统计

Tab.1 Best mode statistics data of macroblock in P frame(QP=26)

序列	SKIP 模式		16×16, 16×8, 8×16 模式		8×8, 8×4, 4×8 4×4 模式		帧内预测模式	
	宏块数	百分比(%)	宏块数	百分比(%)	宏块数	百分比(%)	宏块数	百分比(%)
Missa	5478	55.33	4312	43.56	110	1.11	0	0
Claire	22506	71.94	7571	24.20	1153	3.69	54	0.17
Foreman	8628	27.93	12968	41.98	8945	28.97	347	1.12
Football	4057	12.97	14704	47.00	6294	20.12	6229	19.91

表 2 不同视频内容下的可用选择模式

Tab.2 Optional code mode in different video content

视频内容	可用的选择模式
平坦背景区域	16×16 帧间预测模式, 16×16 帧内预测模式
复杂背景区域	4×4 帧内预测模式, 16×16 帧内预测模式, 16×16 帧间预测模式, 16×8 帧间预测模式, 8×16 帧间预测模式
运动区域	4×4 帧内预测模式, 16×16 帧内预测模式, 16×16 帧间预测模式, 16×8 帧间预测模式, 8×16 帧间预测模式, 8×8 帧间预测模式及各种子模式

3.2 相邻宏块间的最佳模式统计规律

一幅图像中的内容往往是渐进变化的, 相邻的宏块的特征也基本相同^[6], 所以相邻的宏块所选择的最佳编码模式相同的可能性较大。H.264 编码器中, 同一Slice中宏块的编码顺序通常是光栅扫描(raster scan)顺序编码的, 即从左到右, 从上到下。图 2 所示相邻宏块示意图中, 设当前需要编码的宏块为宏块D, 正上方已编码宏块A的编码模式为Bestmode(A), 右上方已编码宏块B的编码模式为Bestmode(B), 左边已编码宏块C的编码模式为Bestmode(C)。定义当前编码宏块D的最佳编码模式的预测值为Premode(D), 定义如式(3)所示。

$$Premode(D) = \begin{cases} Bestmode(A), & Bestmode(A)=Bestmode(B) \\ & =Bestmode(C) \\ Bestmode(B), & Bestmode(A)=Bestmode(B) \\ Bestmode(C), & others \end{cases} \quad (3)$$

统计宏块 D 的最佳编码模式的预测值 Premode(D)和根据 JM86 仿真实验得出的最佳编码模式 Bestmode(D)在采用不同视频序列的仿真实验中, 预测命中率数据统计如表 3 所示。可以看到对于运动较缓慢的序列如 Claire, Missa, 预测最佳编码模式的命中率一般都在 85% 以上, 而对于运动剧烈的序列如 Foreman, Football, 预测最佳编码模式的命中率也在 55% 以上。

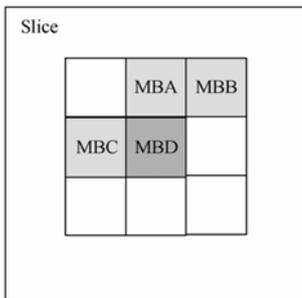


图 2 相邻宏块示意图

Fig.2 Neighbouring macroblocks for the current macroblock

4 H.264 中基于先验检测的帧间模式选择快速算法

上文对 H.264 编码器中的宏块最佳编码模式进行分析, 给出了最佳编码模式的统计分布规律。基于先验检测的帧间模式选择算法是根据这些分布规律, 结合宏块

的实际特性, 采用门限控制的办法, 依次对统计概率上可能是最佳编码的模式逐一筛选。

基于先验检测的帧间模式选择算法具体步骤如下:

步骤 1 以当前宏块 D 的先验预测模式 Premode(D)对该宏块进行预测编码。如果所得到的先验预测模式的编码开销小于门限 TH 就直接得到最优的编码模式选择结果, 跳至步骤 5; 如果先验预测模式的编码开销大于门限 TH 则跳至步骤 2;

步骤 2 依次进行 16×16 帧间预测模式、16×8 帧间预测模式、8×16 帧间预测模式的预测编码。如果某种编码模式的开销小于门限 TH 就直接得到最优编码模式, 跳至步骤 5; 如果这三种编码模式的开销都大于门限 TH 则跳至步骤 3;

步骤 3 依次进行 8×8 帧间预测模式、4×8 帧间预测模式、8×4 帧间预测模式、4×4 帧间预测模式的预测编码。如果某种编码模式的开销小于门限 TH 就直接得到最优编码模式; 如果这 4 种编码模式的开销都大于门限则跳到步骤 4;

步骤 4 依次进行 4×4 帧内预测模式、16×16 帧内预测模式的预测编码。根据先前所有可供选择编码模式的开销, 选择最优的编码模式。如果最佳编码模式是 16×16 帧间预测模式, 跳至步骤 5; 如果最佳编码模式不是 16×16 帧间预测模式, 模式选择算法结束, 得到当前编码宏块的最佳编码模式;

步骤 5 如果编码的最佳模式是 16×16 帧间预测模式, 则判断是否可用 P 帧中的 SKIP 模式编码。

基于先验检测的帧间模式快速算法中, 门限的确定是影响本文算法的关键因素, 如果门限过大会造成编码性能急剧下降, 如果门限过小会造成编码器计算时间过长, 起不到优化的效果。考虑到衡量视频编码器优化效果的主要 3 个指标是编码图像质量、编码速度和压缩效率, 本文的门限 TH 确定由下面的方法得到。

设在 H.264 编码器没有使用优化算法时, 对某一视频序列所需的编码时间为 T₁, 输出码流为 Bits₁, 编码信噪比为 PSNR₁。对编码器做了本文模式选择算法优化之后, 编码器对同一视频序列的编码时间为 T₂, 输出码流为 Bits₂, 编码信噪比为 PSNR₂。算法有效性的指标函数如式(4)所示:

表 3 P 帧中预测最佳模式的正确率统计(QP=26)

Tab.3 Accuracy rate statistics data of predicting best code mode in P frame (QP=26)

序列	采样 7 种帧间预测模式			采用 3 种帧间预测模式		
	宏块总个数	预测准确总个数	百分比(%)	宏块总个数	预测准确总个数	百分比(%)
Missa	9900	8596	86.83	9900	9510	96.06
Claire	39600	33704	85.11	39600	36814	92.96
Foreman	31680	19107	60.31	31680	22621	71.40
Football	31680	18022	56.89	31680	21769	68.72

$$E = \frac{(\text{PSNR}_1/\text{PSNR}_2)^m}{(\text{Bits}_2/\text{Bits}_1)^n (T_2/T_1)^q} \quad (4)$$

其中 m, n, q 均大于零, 本文取 $m=n=q=1$ 。

设当前编码宏块的预测编码模式为 $\text{Premode}(X)$, J_p 为相邻宏块中编码模式为 $\text{Premode}(X)$ 的宏块开销函数的均值, 其中开销函数由式(2)得到。门限 TH 定义如式(5)所示, 因此门限 TH 不是一个固定的常数, 而是根据预测信息变化而变化的。

$$\text{TH} = \alpha J_p, \quad \alpha > 0 \quad (5)$$

式(5)中 α 是一个乘数因子, 在本文算法中是一个固定常数, 直接影响算法性能。给定一个 α 就有对应一个门限 $\text{TH}(\alpha)$, 同时也可以得到算法的性能指标函数 $E(\alpha)$ 。 $E(\alpha)$ 是随着 α 的变化而变化的函数。当 $E(\alpha)$ 取得最大值 $E_{\max}(\alpha^*)$ 时, 算法的最佳 α_{opt} 等于 α^* , 此时有最佳门限 TH_{opt} 等于 $\text{TH}(\alpha_{\text{opt}})$ 。经过大量的实验数据统计, 本文取最佳 $\alpha_{\text{opt}} = 1.15$, 算法的性能基本上达到最优, 同时适合大多数的视频序列。

5 实验结果与分析

在 H.264 测试代码 JM86 基础上加入基于先验检测帧间模式选择算法后, 分别从图像质量、编码速度和压缩效率 3 个方面和原始程序相比较, 检验算法的有效性。本文采用 H.264 的 Baseline Profile 编码, 选用 1 个参考帧, 选取 4 个有代表性的序列 Missa, Claire, Foreman, Football, 这些序列各取 80 帧, 第 1 帧为 I 帧(帧内编码), 其余 79 帧全部为 P 帧(帧间编码)。实验平台是 Pentium-4 处理器, 主频 2.4G, 内存 256M DDR。

表 4 本文算法编码模式计算次数和 JM 算法的比较

Tab.4 Comparison of operation times between proposed algorithm and JM algorithm

序列	量化参数	JM86	JM86 加上本文算法	减少模式计算次数
		编码模式的计算次数	编码模式的计算次数	
Missa	26	70389	13430	56959
	33	70389	11054	59335
Claire	26	281556	57828	223728
	33	281556	48735	232821
Foreman	26	281556	64568	216988
	33	281556	49640	231916
Football	26	281556	93315	188241
	33	281556	89816	191740

表 5 使用全运动搜索时, 本文算法和 JM 算法性能比较(QP=26)

Tab.5 Performance comparison between proposed algorithm and JM algorithm using full motion search algorithm (QP=26)

序列	JM86				JM86 加上本文算法			
	总编码时间(s)	运动估计时间(s)	PSNR_Y (dB)	输出码率 (kbit/s)	总编码时间(s)	运动估计时间(s)	PSNR_Y (dB)	输出码率 (kbit/s)
Missa	41.389	11.043	41.19	35.96	14.684	6.795	41.19	36.64
Claire	167.292	44.819	42.72	152.14	57.347	30.350	42.70	155.44
Foreman	190.181	40.409	37.77	434.92	68.384	27.672	37.79	445.90
Football	194.05	42.485	38.48	1333.19	79.527	30.004	38.49	1387.27

表 4 给出了本文的快速模式选择算法和 JM86 的模式编码计算次数的比较, 可以看出本文算法和原始算法相比, 模式编码计算次数有了很大程度上的减小, 这取决于两个方面: 第一, 由于有了先验知识的预测, 帧间预测模式编码次数减少; 第二, 在帧间预测模式足够准确的情况下, 不再做复杂的帧内模式编码。

表 5 和表 6 分别是本文算法和 JM 算法的计算机仿真实验性能比较, 其中表 5 使用全运动搜索, 表 6 使用快速运动搜索^[7]。可以看出采用本文算法的编码器的 PSNR 只有很少的下降(大约 0.02dB), 输出的码率有小范围的变化, 最大的浮动不超过 5%, 但是与此同时可以节约大量总编码时间和运动估计时间。图 3 给出了本文算法和 JM 原始算法的主观图像性能对比。总体来说本文算法相当有效, 与 JM 原始算法相比编码的性能下降较少。



图 3 本文算法与 JM 编码主观视频对比

(QP=26, Claire 序列, 第 10 帧)

(a) 原始图像 (b) JM 编码重建图像 (c) 本文算法编码重建图像
Fig.3 Visual performance comparison between proposed algorithm and JM algorithm (QP=26, Claire video sequence, the 10th frame)
(a) Original picture (b) Reconstruction picture with JM (c) Reconstruction picture with proposed algorithm

表 6 使用 JM 模型提供的快速运动搜索法与本文算法的实验性能比较(QP=26)

Tab.6 Performance comparison between proposed algorithm and JM algorithm using fast motion search algorithm (QP=26)

序列	JM86				JM86 加上本文算法			
	总编码 时间(s)	运动估计 时间(s)	PSNR_Y (dB)	输出码率 (kbit/s)	总编码 时间(s)	运动估计 时间(s)	PSNR_Y (dB)	输出码率 (kbit/s)
Missa	30.144	1.211	41.17	35.71	5.672	0.582	41.16	35.67
Claire	118.717	4.453	42.66	152.28	22.38	2.452	42.63	155.84
Foreman	150.691	7.127	37.75	433.36	29.032	2.902	37.74	447.73
Football	154.577	9.491	38.48	1342.21	35.273	4.329	38.48	1395.18

6 结束语

本文首先对 H.264 视频标准帧间编码的最优编码模式分布规律进行研究, 得出其统计分布规律, 然后根据所得的分布规律, 提出基于先验预测帧间编码的模式选择快速算法。仿真实验结果表明: 本文算法应用在 H.264 编码器中, 可以在基本保证编码质量的同时, 很大程度上降低编码器的运算复杂度, 同时预测模式的获取和门限 TH 的计算也相对简单, 没有额外增加编码器的计算量。

参 考 文 献

- [1] Draft ITU-T Recommendation and Final Draft International Standard of Joint Video Specification [S](ITU-T Rec. H.264 | ISO/IEC 14496-10 AVC), 7th Meeting: Pattaya, Thailand, 7-14 March, 2003.
- [2] Wiegand T, Sullivan G J, Bjøntegaard G, Luthra A. Overview of the H.264/AVC video coding standard [J]. *IEEE Trans. on Circuits And Systems Video Technology*, 2003, 13: 560-576.
- [3] Richardson E G. H.264 and MPEG-4 Video Compression: Video Coding for Next Generation Multimedia [M]. John Wiley & Sons, 12 August, 2003.
- [4] H.264/AVC Reference Software JM86 [EB/OL]http://iphome.hhi.de/suehring/tml/download/.
- [5] Sullivan G J, Wiegand T. Rate-distortion optimization for video compression [J]. *IEEE Signal Processing Magazine*, 1998, 15(6): 74-90.
- [6] Ndjiki-Nya, Makai P, Blattermann B. Improved H.264/AVC coding using texture analysis and synthesis [C]. *Image Processing*, 2003. Proceedings. 2003 International Conference on, 14-17 Sept., 2003, vol.3: 849-852.
- [7] Tourapis H Y C, Tourapis A M. Fast motion estimation within the H.264 codec [C]. *Multimedia and Expo*, 2003. ICME '03. Proceedings. 2003 International Conference on, 6-9 July, 2003, 3: 517-520.

干宗良: 男, 1979 年生, 博士生, 主要研究方向为图像通信、视频压缩编码技术。

齐丽娜: 女, 1979 年生, 博士生, 主要研究方向为超宽带无线通信系统。

朱秀昌: 男, 1947 年生, 教授, 博士生导师, 中国通信学会、电子学会高级会员, “江苏省图像处理与图像通信重点实验室”主任。主要研究方向为图像通信、视频处理。