

中图分类号: TN919.8 文献标志码: A 文章编号: 1006-8961(2011)03-0305-05

论文索引信息: 顾梅花, 余宁梅, 寇立康, 朱磊. H.264 快速模式选择算法中的提前终止策略 [J]. 中国图象图形学报, 2011, 16(3): 305-309

H.264 快速模式选择算法中的提前终止策略

顾梅花^{1),2)}, 余宁梅¹⁾, 寇立康¹⁾, 朱磊²⁾

¹⁾ (西安理工大学自动化与信息工程学院, 西安 710048) ²⁾ (西安工程大学电子信息学院, 西安 710048)

摘要: 针对 H.264 中模式选择计算复杂度高的问题, 提出一种快速模式选择算法。根据帧间编码最佳模式的统计规律, 找出宏块的时间空间特性与最佳模式的关系, 在帧间模式选择的各个阶段, 检测宏块的时间空间特性, 一旦满足相应的残差一致性条件或空间一致性条件, 则提前终止模式选择。实验结果表明, 该方法在输出比特率平均升高 2.11% 的情况下, 编码时间能平均减少 23.11%, 峰值信噪比平均提高 0.058 dB。

关键词: H.264; 模式选择; 残差一致; 空间一致; 早停止

Fast mode decision algorithm for H.264 based on early termination method

Gu Meihua^{1),2)}, Yu Ningmei¹⁾, Kou Likang¹⁾, Zhu Lei²⁾

¹⁾ (School of Automation and Information Engineering, Xi'an University of Technology, Xi'an 710048 China)

²⁾ (School of Electronic Information, Xi'an Polytechnic University, Xi'an 710048 China)

Abstract: To simplify the computational complexity in mode selection process for H.264, a fast mode decision algorithm is proposed in this paper. According to the statistical rule of the best mode chosen by inter frame, the relationship of the temporal and spatial characteristics with the best mode is found. At each stage of inter mode selection, the temporal and spatial characteristics of macroblock (MB) are detected, once the residual homogeneous or spatial homogeneous conditions are satisfied, the mode selection process is terminated. Experimental results show that this method can make the encoding speed accelerate about 23.11% and peak signal to noise ratio (PSNR) improve about 0.058 dB, with only 2.11% increase about bit rate.

Keywords: H.264; mode selection; residual homogeneous; spatial homogeneous; early termination

0 引言

视频编码标准 H.264/AVC 由联合视频专家组 (JVT) 于 2003 年 5 月完成^[1]。为获取高效的编码效率, H.264 采用率失真优化 (RDO) 技术来检测所有可能的帧间和帧内模式寻找最佳的编码结果^[2]。穷尽的模式选择过程通常大约占 JM (joint model) 参考软件总计算负荷的 90%^[3]。因此, 在保持编码

质量和总比特率的同时降低计算复杂度是设计 H.264 快速模式选择算法的主要目标。许多快速而有效的模式选择方法在近年被提出^[4-8]。主要有两方面的思路, 一是 RDO 函数的优化, 另一个思路是跳过概率较小的模式判断从而减少计算复杂度。其中文献[6]提出一种快速模式判定算法, 包含两个早停止阶段。第 1 阶段, 检测 16×16 尺寸块的模式来尽早停止。如果满足早停止条件, SKIP (跳过) 模式被设置; 否则, 检测余下的帧间模式来寻找最佳帧

收稿日期: 2010-05-31; 修回日期: 2010-08-27

基金项目: 陕西省教育厅专项科研项目 (2009JK474, 2010JK558); 陕西省重点学科建设专项资金 (107080903)。

第一作者简介: 顾梅花 (1980—), 女, 讲师。西安理工大学微电子学与固体电子学博士研究生, 主要研究方向为视频编解码算法, VLSI 设计。E-mail: gu_meihua@163.com。

间模式。然后执行第 2 阶段的早停止检测来尽早跳过所有帧内模式。这种算法对于存在大量静止背景或简单运动的视频序列非常有效,但并不适合含有中高速运动且纹理复杂的视频序列。本文提出一种适用于上述各种序列的快速模式选择算法,在帧间模式选择的各个阶段,利用宏块(MB)的时间空间特性,多次使用各种早停止策略,加速了模式选择过程。该方法与文献[6]的算法相结合,能够进一步提升编码器的编码效率。实验结果表明,与 JM 参考软件中的穷尽模式选择算法相比,该算法在保持视频编码质量和比特率基本不受影响的同时能够显著加快编码器的编码速度。

1 H. 264 帧间模式选择

H. 264/AVC 中的 P 帧和 B 帧编码,共有 10 种候选模式:SKIP/DIRECT, P 16 × 16, P 16 × 8, P 8 × 16, P 8 × 8 (8 × 8, 8 × 4, 4 × 8, 4 × 4), INTRA (I4MB, I16MB),既支持帧间预测模式又支持帧内预测模式。H. 264 为帧间编码的运动估计提供了 7 种块尺寸,这些帧间模式构成了一个 MB 的两级结构,如图 1 所示。第 1 级 L1 包括的模式有 P 16 × 16, P 16 × 8, P 8 × 16, P 8 × 8, 而第 2 级 L2 中, P 8 × 8 模式下的每个 8 × 8 块又可以有 4 种子分割方式,分别为 8 × 8, 8 × 4, 4 × 8 和 4 × 4。

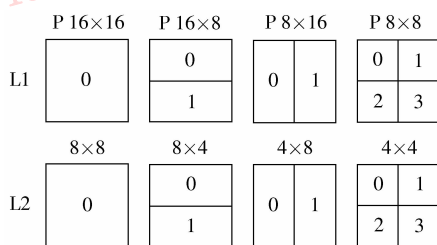


图 1 MB 的两级结构

Fig. 1 Two level structure of MB

为得到较好的率失真性能, H. 264 采用 RDO 技术进行模式选择。在所有可能的帧间帧内模式中选择,将最小率失真(RD)代价对应的模式选为最佳模式。RD 代价函数定义为

$$J(s, c, M | Q, \lambda_M) = S(s, c, M | Q) + \lambda_M \cdot R(s, c, M | Q) \quad (1)$$

式中, s 和 c 分别代表源视频信号和重建视频信号, M 表示一个 MB 模式, 属于 { SKIP/DIRECT, P 16 ×

16, P 16 × 8, P 8 × 16, P 8 × 8, INTRA } , Q 是量化参数 ($0 \leq Q \leq 51$) , λ_M 是拉格朗日乘数, S 是原始信号 s 和重建信号 c 的方差和, $R(s, c, M | Q)$ 是所选模式、MB 头信息以及残差信息编码所需要的全部比特数。

2 帧间编码最佳模式统计及分析

一般来说, 选择一个较大块尺寸通常会导致较大的运动补偿残差, 特别是在那些包含不同运动的区域。相反, 选择一个较小块尺寸通常会引起较小残差。对于连续运动的区域, 包括物体的整体运动, 移动背景的平滑运动, 以及静止背景为零运动, 用较大块尺寸编码就已足够, 不会引起较大的残差。对于包含复杂运动以及不同物体运动的区域, 就应当表示更细节的运动, 使用较小块尺寸。选择了 4 个序列, 包括慢速简单运动 (News, Miss), 中高速运动 (Foreman), 以及高速复杂运动 (Football), 用穷尽模式选择算法得到最佳模式的平均概率分布, 如表 1 所示。

表 1 $Q = 28$ 时, P 帧的宏块最佳模式分布 (%)

Tab. 1 Best mode distribution for frame P when $Q = 28$

序列	SKIP	P 16 × 16	P 16 × 8	P 8 × 16	P 8 × 8	INTRA
News	74.68	6.25	2.98	3.91	11.87	0.31
Miss	72.39	17.03	3.67	3.17	1.30	2.44
Foreman	28.49	25.78	10.66	12.82	21.04	0.21
Football	12.37	15.35	7.99	7.82	30.41	26.15
平均	46.98	16.10	6.33	6.93	16.11	7.28

实验结果表明, 由于视频序列具有很强的时间和空间相关性, 最佳编码模式并不是在各种模式中均匀分布的。据观察, 许多自然视频序列会包含大量背景或一致性运动的区域, 这些区域的 MB 基本会选择帧间大块尺寸模式进行编码 (SKIP, P 16 × 16)。利用穷尽模式选择法对序列 News 和 Football 进行实验, 得到各个 MB 的最佳模式。图 2(a)(b) 分别显示了 News 第 10 帧和 Football 第 15 帧中 MB 的最佳块尺寸模式分布。图 2(a) 中含有大量的背景和空间一致性区域, 其中包括静止背景和播音员的衣服, 这些区域的 MB 全部采用 16 × 16 大块帧间模式编码。尽管播音员的衣服边缘并不平滑, 但因为大多数时候没有运动或只有少量运动, 所以依然采用 16 × 16 大块模式。播音员后面的舞者相对较小,

且包含了不同方向的运动,适合于用小块尺寸编码。图 2(b) 包含有大量的剧烈运动和丰富的纹理。其中包括各个运动员高速运动的身体,衣服和头部的细节纹理,这些区域的 MB 全部采用小于 8×8 的块尺寸或者 INTRA 模式(标注 I 的 MB)。虽然存在少许镜头的平移,但因为背景的纹理比较简单,而且属于整体平移,所以背景区域的 MB 依然采用 16×16 的大块尺寸模式。另外,运动员衣服较为平坦的部分,以及身体的边界部分在空间上较为均匀,而且存在部分整体运动,这些区域的 MB 则选择 16×8 或 8×16 的块尺寸模式。



图 2 穷尽模式选择后的最佳模式分布

Fig. 2 Best mode distribution after full mode selection

通过上面的分析,得出 MB 的时间空间特性以及对应模式如表 2 所示。

表 2 不同 MB 特性对应的编码模式

Tab. 2 Different characteristics of MB and its modes

类型	MB 的时间空间特性	对应模式
1	静止背景、运动较小区域,不管其纹理是简单还是复杂;或整体运动、具有简单纹理	SKIP, P 16×16
2	空间基本均匀且少量连续整体运动	P 16×8 , P 8×16
3	位于运动物体边界且物体具有不同的运动方向	P 8×8
4	快速运动且纹理丰富,或场景变换	INTRA

3 帧间编码中的早停止策略

3.1 残差一致特性检测

帧间编码过程中,最佳帧间模式下得到的预测残差的能量一定较小。反之,如果某一种模式下的预测残差能量较小,则该模式很有可能是最佳帧间模式。利用这一思想,本文利用 MB 的时间特性,通过检测运动估计后的预测残差是否具有残差一致特性^[7]判断当前 MB 能否采用当前尺寸模式。如果

MB 具有残差一致特性,则说明该 MB 不管是简单纹理还是复杂纹理,但静止或运动平滑。判断 MB 是否具有残差一致特性的主要思想如下,如果 $N \times N$ 块中的 4 个预测残差子块 $R_1 \sim R_4$ (如图 3 所示)满足

$$\sum_{i=1}^{N/2} \sum_{j=1}^{N/2} |r_k(i, j) - M_k| < \zeta \quad k = 1, 2, 3, 4 \quad (2)$$

式中, ζ 表示一个待定阈值, $r_k(i, j)$ 是 $N \times N$ 预测后的第 k 个 $N/2 \times N/2$ 子块中第 (i, j) 个像素的预测残差,则认为该 $N \times N$ 块的残差一致。

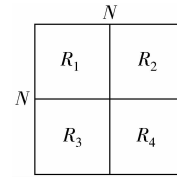


图 3 残差块及其子块, $N = 16$ 或 8

Fig. 3 Residual block and the sub-blocks, $N = 16$ or 8

定义

$$r_k(i, j) = \hat{f}(i + v_x, j + v_y) - f(i, j) \quad (3)$$

式中, $\hat{f}(i + v_x, j + v_y)$ 和 $f(i, j)$ 分别表示第 k 个子块中第 (i, j) 个预测像素值,原像素值。

第 k 个子块的平均预测残差定义为

$$M_k = \frac{4}{N^2} \sum_{i=1}^{N/2} \sum_{j=1}^{N/2} r_k(i, j) \quad (4)$$

对于式(2)中的待定系数 ζ , 较大的 ζ 值会产生较多的残差一致块,从而可以加速宏块模式选择,但也会降低编码性能,对于较小的 ζ 值,虽然可以保证编码性能,但不能减少编码时间。所以 ζ 的值需要根据实际应用进行折中。另一方面,因为残差块来自于运动估计后的预测误差,较大的量化步长 Q_s 会使预测残差块在变换、量化后更容易变成全零块,从而可以加速编码速度并减少编码比特数,但同时会降低图像质量,反之,较小的 Q_s 能保证编码图像质量,但不能提高编码速度。所以 ζ 值与量化步长 Q_s 密切相关。为简单起见,认为 ζ 与 Q_s 满足线性关系。选择 $Q_s, 2Q_s, 4Q_s, 8Q_s, 12Q_s$ 以及 $16Q_s$ 分别作为阈值对 16×16 和 8×8 块进行实验,并比较相应的编码质量和速度,结果表明,较为合理的 ζ 值设定为

$$\zeta = \begin{cases} 8Q_s & N = 16 \\ 2Q_s & N = 8 \end{cases} \quad (5)$$

3.2 空间一致特性检测

如果 MB 空间具有整体相似性,则其运动会具

有一致性而且不易分裂成小块。本文对 MB 空间一致性的检测主要利用 MB 的边缘信息^[8],这些编码信息来自于由 Sobel 算子产生的边缘图。对于一个亮度像素 $p_{i,j}$,定义相关的边缘矢量 $\mathbf{D}_{i,j} = \{dx_{i,j}, dy_{i,j}\}$,其中 $dx_{i,j}, dy_{i,j}$ 分别表示垂直与水平方向的变化程度。

$$dx_{i,j} = p_{i-1,j+1} + 2 \times p_{i,j+1} + p_{i+1,j+1} - p_{i-1,j-1} - 2 \times p_{i,j-1} - p_{i+1,j-1} \quad (6)$$

$$dy_{i,j} = p_{i+1,j-1} + 2 \times p_{i+1,j} + p_{i+1,j+1} - p_{i-1,j-1} - 2 \times p_{i-1,j} - p_{i-1,j+1} \quad (7)$$

边缘矢量幅度为

$$A(\mathbf{D}_{i,j}) = |dx_{i,j}| + |dy_{i,j}| \quad (8)$$

利用边缘矢量幅值来判定一个 $N \times N$ 块是否具有空间一致性,如果块中所有像素位置的边缘矢量幅度之和小于一个阈值 T ,则认为该块具有空间一致性。阈值 T 是一个预设的参数。令 r 和 c 为块 $B_{r,c}$ 的行列索引,定义

$$H_{r,c} = \begin{cases} 1 & \sum_{i,j \in N \times N} A(\mathbf{D}_{i,j}) < T \\ 0 & \sum_{i,j \in N \times N} A(\mathbf{D}_{i,j}) \geq T \end{cases} \quad (9)$$

$H_{r,c} = 1$ 表明 $N \times N$ 块 $B_{r,c}$ 具有空间一致性, $H_{r,c} = 0$ 说明该块纹理较为复杂。大量实验结果表明,对于 16×16 块,阈值 T 设置为 20 000 能够取得较好的编码性能。对于 8×8 块,合理的阈值 T 设置为 5 000^[9]。

4 算法的详细过程

将上述提前终止策略应用于 H. 264 帧间编码模式选择过程中,并与文献[6]中的 SKIP 模式预判以及选择性帧内模式预判结合。具体步骤为

1) $P 16 \times 16$ 模式下运动估计,判断 SKIP 条件是否满足,若满足则提前结束模式选择;

2) 判断是否满足残差一致条件或空间一致条件,若满足,进入步骤 6);

3) $P 16 \times 8$ 模式下运动估计,判断是否满足残差一致条件或空间一致条件,若满足,进入步骤 6);

4) $P 8 \times 16$ 模式下运动估计,判断是否满足残差一致条件或空间一致条件,若满足,进入步骤 6);

5) 亚宏块级 $P 8 \times 8$ 下的运动估计及模式判定,每个子块的 8×8 运动估计后,判断是否满足残差一致条件或空间一致条件,若满足,认为当前子块的最

佳模式为 8×8 ,并退出当前子块的模式判定过程。4 个子块的模式选择完成后,进入步骤 6);

6) 计算当前及之前的所有帧间模式下的 RD 代价,选择最小代价对应的模式为最佳帧间模式;

7) 选择性帧内模式选择预判,如果满足相关条件,则跳过帧内模式选择过程,将最佳帧间模式作为当前 MB 的最佳模式,否则,继续执行帧内模式选择过程,并计算各个帧内模式下的 RD 代价,与最小帧间 RD 代价进行比较,选择最小 RD 代价对应的模式为最佳模式。

5 实验结果与分析

本文算法在 H. 264 参考软件 JM16. 1 上实现,仿真实验的硬件平台为 Intel Core 处理器,主频 2.0 GHz,内存 1 G。测试条件设置为:每个序列各取 100 帧;编码结构为 IPPP;使用 RDO;QP 设置为 24, 28, 32, 36;使用 CABAC 熵编码;参考帧数量为 1。实验中主要采用 3 个指标评价算法性能^[10],分别为与原始参考软件中的穷尽模式选择算法相比重建图像质量降低的分贝数 $\Delta\text{PSNR}/\text{dB}$;总比速率升高的百分比 $\Delta B\%$;平均编码时间降低的百分比 $\Delta T\%$;"+"表示增加;"-"表示减少。

表 3 和表 4 是采用 QCIF 序列和 CIF 序列分别测试本文算法以及文献[6]算法的实验结果,其中 A 代表文献[6]算法,B 代表本文算法。由表 3 中数据可知,本文算法测试 QCIF 序列在输出比特率平均升高 1.77% 的情况下,峰值信噪比平均提高了 0.033 dB,编码时间减少了 25.77%,极大地提升了编码器的整体性能,总体上优于文献[6]算法。由表 4 中数据可知,文献[6]对于所有 CIF 序列几乎没有什么效果,而本文算法在码率平均增加 2.57% 的情况下,能够使编码时间减少 19.88%,PSNR 提高 0.087 dB。值得注意的是,本文算法在降低计算复杂度的同时,还能使重建图像的质量在一定程度上得到改善,其中 Miss 序列的效果最好,其次依次是 Football, Paris, Tempete, Foreman, Container, Silent。由表 3 和表 4 综合来看,除了个别快速运动且纹理复杂的序列(例如 Mobile),本文算法适用于表 2 中除了类型 4 以外的所有其他各种特性的序列。

表3 QCIF序列的实验结果

Tab.3 Experiment results of QCIF sequences

QCIF 序列	PSNR/dB		$\Delta B/\%$		$\Delta T/\%$	
	A	B	A	B	A	B
News	-0.023	-0.037	0.77	1.07	-18.67	-25.54
Foreman	-0.018	0.051	0.14	2.23	-5.53	-28.74
Silent	0.007	0.009	0.40	2.63	-14.76	-23.00
Container	-0.012	0.037	0.96	1.15	-18.91	-23.91
Football	-0.006	0.107	0.09	1.64	-0.78	-27.64
平均	-0.011	0.033	0.54	1.77	-11.73	-25.77

表4 CIF序列的实验结果

Tab.4 Experiment results of CIF sequences

CIF 序列	PSNR/dB		$\Delta B/\%$		$\Delta T/\%$	
	A	B	A	B	A	B
Miss	-0.012	0.211	0.18	2.50	-1.74	-33.9
Paris	-0.011	0.089	-0.16	2.45	4.75	-25.87
Mobile	-0.002	-0.01	0.05	1.79	-0.56	-0.15
Tempete	-0.004	0.058	-0.06	3.57	-0.37	-19.58
平均	-0.0073	0.087	0.0025	2.57	0.52	-19.88

6 结论

根据帧间编码最佳模式的统计规律,以及宏块的时间空间特性与最佳模式的关系,提出一种基于宏块时间空间特性的早期终止策略的快速模式选择算法。实验结果表明,该算法在只有少许码率增加的情况下,对于多种不同类型的视频序列都能显著降低编码器的计算复杂度,并保持良好的编码质量。尤其对于存在静止背景,纹理简单且运动较小,或存在一致运动的视频序列具有更好的效果。

参考文献 (References)

- [1] Joint Video Team. Draft ITU-T Recommendation and Final Draft International Standard of Joint Video Specification[S]. Geneva: ITU-T Recommendation H.264 and ISO/IEC 14496-10 AVC, 2003.
- [2] Sullivan G J, Topiwala P, Luthra A. The H.264/AVC advanced video coding standard: overview and introduction to the fidelity range extensions [C]//Proceedings of SPIE Conference on Applications of Digital Image Processing. Washington, USA: SPIE, 2004: 454-475.
- [3] Tu Y K, Yang J F, Sun M T. Statistical rate-distortion estimation for H.264/AVC coders [C]//Proceedings of IEEE International Symposium on Circuits and Systems. New York, USA: IEEE Press, 2006: 365-368.
- [4] Moon J M, Moon Y H, Kim J H. A computation reduction method for RDO mode decision based on an approximation of the distortion [C]// Proceedings of IEEE International Conference on Image Processing. New York, USA: IEEE Press, 2006: 2481-2484.
- [5] Kannangara C S, Richardson Iain E G, Bystrom M, et al. Low-complexity skip prediction for H.264 through Lagrangian cost estimation [J]. IEEE Transactions on Circuits and Systems for Video Technology, 2006, 2(16): 202-208.
- [6] Choi I, Lee J, Jeon B. 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, 12(16): 1557-1561.
- [7] Wang H M, Lin J K, Yang J F. A successive termination and elimination method for fast H.264/AVC SATD-based inter mode decision [J]. IET Signal Processing, 2009, 3(3): 165-176.
- [8] Pan F, Lin X, Susanto R, et al. Fast Mode Decision for Intra Prediction [EB/OL]. (2003-03-07) [2009-11-27]. http://ftp3.itu.int/av-arch/jvt-site/2003_03_Pattaya/JVT-G013.doc.
- [9] 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.
- [10] Bjontegaard G. Calculation of average PSNR differences between RD-curves [EB/OL]. (2001-04-02) [2009-11-27]. http://ftp3.itu.int/av-arch/video-site/0104_Aus/VCEG-M33.doc.