

# 一种基于 H.264 的视频转码方案

刘兆广, 彭玉华, 杨 阳

(山东大学信息科学与工程学院, 山东济南 250100)

**摘 要:** 本文提出了一种基于 H.264 降低图像尺寸转码中的宏块类型快速选择方案. 文中以原始图像中对应四个宏块的非零系数总个数为准则选择宏块类型, 首次提出以指数模型来描述阈值和重新量化参数的关系 ( $Thr-Q_r$  模型), 并将其线性化为一元线性回归模型, 进一步使用最小二乘法来估计模型中的参数. 帧内模式选择中,  $Thr-Q_r$  模型是用来区分  $I4 \times 4$  和  $I16 \times 16$  宏块; 帧间模式选择中,  $Thr-Q_r$  模型区分出当前宏块所处区域的运动性质, 进而简化候选宏块类型个数. 实验结果表明, 与全搜索法相比, 在保证压缩性能损失不大的情况下, 本文提出的帧内模式选择方法的速度能提高 4 到 5 倍, 帧间模式选择方法的速度能提高约 35 倍.

**关键词:** 视频转码; H.264; 宏块类型选择; 线性回归

**中图分类号:** TP391.41 **文献标识码:** A **文章编号:** 0372-2112 (2008) 05-0848-04

## A Video Transcoder Scheme Based on H.264

LIU Zhao-guang, PENG Yu-hua, YANG Yang

(School of Information Science and Engineering, Shandong University, Jinan, Shandong 250100, China)

**Abstract:** A fast macroblock mode selection method in downscaling transcoder based on H.264 is proposed in this paper, which the total non-zero coefficients of responding four macroblocks in pre-coded frame is utilized as the criterion to select macroblock mode. We propose an exponent function model between threshold and re-quantization parameter ( $Thr-Q_r$  model). This model is converted into a linear regression model, and least square method is adopted to estimate parameters in the model. In the intra mode selection part,  $Thr-Q_r$  model is used to classify  $I4 \times 4$  or  $I16 \times 16$ ; In the inter mode selection part,  $Thr-Q_r$  model is used to classify the motion activity in downscaled frame, and different candidate macroblock modes are used according to classified motion activity areas. We compared the PSNR performance and computational complexity of proposed method with the full search algorithm. For the intra mode selection part, we can speed up about 4~5 times, and for the inter mode selection part, we can speed up about 35 times.

**Key words:** video transcoder; H.264; macroblock decision; linear regression

## 1 引言

在 H.264 帧内预测中, 宏块分为  $I4 \times 4$  和  $I16 \times 16$  两种.  $I4 \times 4$  类型中宏块被划分成 16 个  $4 \times 4$  子块, 每个子块有各自的预测模式; 而宏块在  $I16 \times 16$  类型中会使用一个预测模式. 在帧间预测中, 宏块可以划分为  $P16 \times 16$ 、 $P16 \times 8$ 、 $P8 \times 16$ 、 $P8 \times 8$  等类型. 如何选择宏块类型是基于 H.264 的视频编码的研究热点之一<sup>[1,2]</sup>.

视频转码技术中通常分为图像尺寸转码<sup>[3,4]</sup>、帧率转码<sup>[5]</sup>、重新量化<sup>[6~8]</sup>三个方面. 在视频转码中也需要重新选择宏块类型. 在文献<sup>[3]</sup>中, 宏块类型的选择是基于常见的 R-D 曲线最小化代价准则. 首先为每种宏块类型计算各自的初始运动矢量, 并利用初始运动矢量来选择具有最小代价的宏块类型. 文献<sup>[4]</sup>则使用重构图像的整数余弦变换系数的能量作为选择宏块类型的依据. 在重新量化转码中<sup>[6~9]</sup>, 不同的量化步长, 宏块类型的使用也有所不同. 比如在帧内预测中, 当量化步长较

小时, 会使用较多的  $I4 \times 4$  类型; 而随着量化步长增加, 会使用越来越多的  $I16 \times 16$ . 在文献<sup>[6]</sup>中, 当前宏块的比特流长度被作为衡量标准来选择在重新量化后的宏块类型. 另外, 如果选择开环的视频转码系统, 重新量化误差会逐帧积累, 形成漂移效应. 文献<sup>[7]</sup>引入了一个与宏块类型相关的矩阵来对重新量化误差进行补偿. 文献<sup>[9]</sup>专门分析了 H.264 重新量化转码中存在的问题.

在本文, 我们提出了一种快速的同时支持降低图像尺寸和重新量化的宏块类型选择方法. 该方法中适用的图像缩放因子为 2, 输入的比特流需要完全解码, 在重新编码阶段充分利用了解码得到的信息, 从而提高了编码速度.

## 2 $Thr-Q_r$ 模型

降低尺寸后图像中的一个宏块(当前宏块)对应着原始图像中的四个宏块. 本文将统计这四个宏块的非零系数总个数 ( $num\_nz$ ).  $num\_nz$  值越大, 意味着残差数

据越多. 本文提出的  $Th\_Q_r$  模型(重新量化参数  $Q_r$ ) 分别应用到帧内模式选择中( $Th\_FQ_r$  模型)和帧间模式选择中( $Th\_PQ_r$  和  $Th\_SQ_r$  模型).  $Th\_FQ_r$  模型是用来区分当前宏块的类型;  $Th\_PQ_r$  模型和  $Th\_SQ_r$  模型是用来区分当前宏块所在区域的运动性质.

### 2.1 $Th\_I-Q_r$ 模型

在帧内模式选择中, 首先需要为当前宏块选择类型, 然后根据不同的宏块类型选择预测模式. 通过大量的统计结果, 我们发现当前宏块对应的  $num\_nz$  值较小时, 它选择为  $116 \times 16$  的可能性较大; 否则, 选择  $14 \times 4$  的概率更高. 因此本文使用  $num\_nz$  作为划分  $116 \times 16/14 \times 4$  的依据, 如果当前宏块对应的  $num\_nz$  值小于阈值( $Th\_I$ ), 就选择  $116 \times 16$ , 否则选择  $14 \times 4$ . 图 1 以序列 'Garden' 为例, 给出了  $Th\_I$  与  $Q_r$  的关系. 实验中为每个  $Q_r$  选择一个固定阈值  $Th\_I$ , 如果当前宏块对应的  $num\_nz$  小于这个固定阈值, 就使用  $116 \times 16$ , 否则使用  $14 \times 4$ . 图中左半部分表示这种固定阈值方法与全搜索法的比较, 两者压缩性能几乎完全相同; 右半部分表示  $Q_r$  和固定阈值的关系.

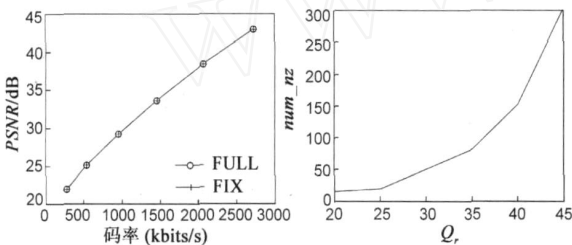


图 1  $Q_r$  与  $Th\_I$  的关系

根据实验结果, 在保证压缩性能不受损失的前提下,  $Q_r$  与  $Th\_I$  的关系很接近于指数曲线, 本文称之为  $Th\_FQ_r$  模型, 见等式(1), 其中  $a$  和  $b$  是模型的参数.

$$Th\_I = ae^{bQ_r} \tag{1}$$

对等式(1)进行线性化处理, 得到一元线性回归模型<sup>[10]</sup>, 然后利用最小二乘法估计模型中的参数  $a$  和  $b$ . 对等式两边同时取对数, 可以得到下式:

$$\ln Th\_I = \ln a + bQ_r \tag{2}$$

将  $y = \ln Th\_I$  和  $c = \ln a$  代入, 则有:

$$y = c + bQ_r \tag{3}$$

根据最小二乘法估计, 参数  $c$  和  $b$  可以由下式得出:

$$b = \frac{\sum_{i=1}^n Q_{ri} y_i - \frac{\sum_{i=1}^n Q_{ri}}{n} \frac{\sum_{i=1}^n y_i}{n}}{\sum_{i=1}^n Q_{ri}^2 - \left(\frac{\sum_{i=1}^n Q_{ri}}{n}\right)^2} \tag{4}$$

$$c = \frac{\sum_{i=1}^n y_i - b \frac{\sum_{i=1}^n Q_{ri}}{n}}{n} \tag{5}$$

其中,  $n$  是帧数目;  $Q_{ri}$  是第  $i$  帧的量化参数;  $y_i$  对应第  $i$

帧的  $num\_nz$  阈值, 由  $Th\_I_i$  计算得出.

在转码前无法预知适合当前视频序列的  $Th\_FQ_r$  关系. 因此, 需要对大量的视频进行测试, 得到一个初始的  $Th\_FQ_r$  关系, 并在转码过程中更新. 下面给出计算和更新参数  $a$  和  $b$  的伪代码.

```

a: 输入初始  $Th\_FQ_r$  集合.
b: 使用当前  $Th\_FQ_r$  集合, 根据等式和估计参数  $a$  和  $b$ .
c: 使用参数  $a$  和  $b$  根据等式计算  $Th\_I$ .
d: 计算阈值  $Th\_low$  和  $Th\_high$ :
     $Th\_low = 0.9 \times Th\_I$ ;  $Th\_high = 1.1 \times Th\_I$ .
e: for (当前帧所有宏块) {
    为当前宏块统计  $num\_nz$  值;
    if ( $num\_nz < Th\_low$ ) 选择  $116 \times 16$ ;
    else if ( $num\_nz > Th\_high$ ) 选择  $14 \times 4$ ;
    else {
        选择  $14 \times 4$  和  $116 \times 16$  中最佳者;
        if (最佳类型是  $14 \times 4$ )  $num\_4 \times 4 ++$ ;
        else  $num\_16 \times 16 ++$ ;
    }
}
if ( $num\_16 \times 16 > num\_4 \times 4$ ) 将  $Th\_high$  和  $Q_r$  添加到  $Th\_FQ_r$  集合;
else 将  $Th\_low$  和  $Q_r$  添加到  $Th\_FQ_r$  集合;
f: go to b, 转码下一帧.
    
```

从上述伪代码可以看出, 实际使用的阈值分别为  $Th\_low$  和  $Th\_high$ . 如果  $num\_nz$  值处在这两者之间, 则会在  $116 \times 16$  和  $14 \times 4$  中选择较好的类型, 最后根据选择的结果来更新  $Th\_FQ_r$  集合.

若利用  $Th\_FQ_r$  模型选择了  $116 \times 16$ , 本文将对全部 4 种预测模式进行测试, 如果选择了  $14 \times 4$ , 整个宏块将被分成 16 个  $4 \times 4$  子块, 每个子块最多有 9 种候选模式. 由于每个降低尺寸后图像中的  $4 \times 4$  子块对应着原始图像中 4 个  $4 \times 4$  子块. 如果这 4 个  $4 \times 4$  子块位于  $116 \times 16$  宏块内, 本文将只测试模式 0、1、2, 否则, 本文将测试这 4 个  $4 \times 4$  子块的预测模式, 从而至少为每个  $4 \times 4$  子块省略 5 种候选模式.

### 2.2 $Th\_PQ_r$ 模型和 $Th\_SQ_r$ 模型

在 H.264 中, P 帧允许使用的宏块类型包括  $14 \times 4$ 、 $116 \times 16$ 、SKIP、 $16 \times 16$ 、 $16 \times 8$ 、 $8 \times 16$ 、 $8 \times 8$  等. 将宏块类型分为两个集合  $set1 = \{116 \times 16, SKIP, 16 \times 16, 16 \times 8, 8 \times 16\}$ ,  $set2 = \{14 \times 4, 8 \times 8\}$ . 显然  $set1$  中对应的宏块类型主要位于运动较平缓的区域,  $set2$  位于运动较剧烈的区域.

随着  $num\_nz$  值递增, 当前宏块使用  $set1$  的概率递减. 另外, 随着量化参数的递增, 具有相同  $num\_nz$  值的宏块使用  $set1$  的概率也在递增. 该性质与 2.1 中所述的  $Th\_FQ_r$  极为类似. 因此, 引入阈值  $Th\_P$ , 若当前宏块对应的  $num\_nz$  值小于  $Th\_P$ , 则在选择宏块类型时, 直

接跳过  $set2$  中的宏块类型. 这里同样使用指数模型 ( $Th.P-Q_r$  模型) 来描述  $Q_r$  与  $Th.P$  的关系:

$$Th.P = ae^{bQ_r} \quad (6)$$

$P16 \times 16$  类型对应着图像中运动缓慢的区域, 这里引入一个阈值  $Th.S$ , 当  $num.nz$  值小于阈值  $Th.S$  时, 当前宏块直接选择为  $P16 \times 16$ , 从而节省计算时间. 这里也使用指数模型 ( $Th.S-Q_r$  模型) 来描述  $Th.S$  与  $Q_r$  之间的关系:

$$Th.S = ae^{bQ_r} \quad (7)$$

$Th.P-Q_r$  和  $Th.S-Q_r$  模型中的参数  $a$  和  $b$  的计算和更新方法与  $Th.T-Q_r$  类似.

本文提出的整个帧间宏块类型选择的过程见图 2.

宏块类型选择中根据提出的指数模型跳过了部分候选类型的测试, 进而节省了运算时间, 宏块类型选择的准则是计算下式的最小值:

$$COST = SAD + MV_{bits} \quad (8)$$

其中:  $SAD$  是预测值和参考值间的差绝对值的和 ( $SAD$ : Sum of Absolute Difference). 预测值根据初始运动矢量计算得到. 是拉格朗日算子.  $MV_{bits}$  是运动矢量所消耗的比特长度. 所有候选宏块类型都以等式 (8) 为准则来比较, 并选择其中最小值作为最终类型. 初始运动矢量根据原始图像中的运动矢量计算得到, 具体方法见文献 [3].

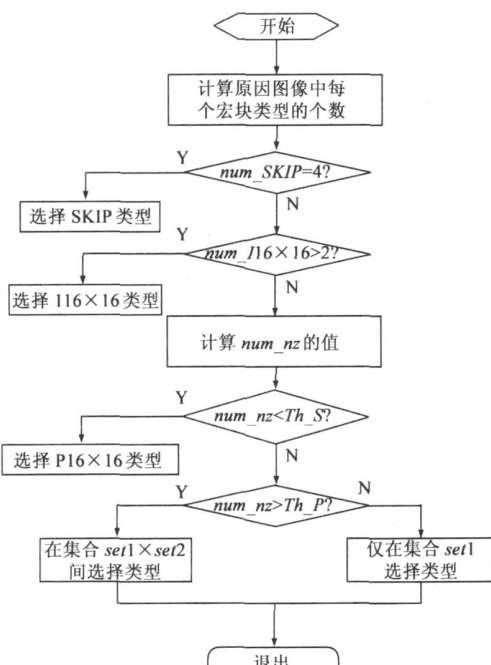


图 2 帧间模式选择流程

### 3 实验结果

实验中原始图像尺寸为 CIF (352 × 288), 帧率为 30 帧/秒, 输入量化参数为 20. 原始视频由 H.264 参考软件 JM12.1 编码. 符号“FULL”表示使用全搜索法重新进

行宏块类型选择得到的结果. “FAST”代表本文提出的方法.

#### 3.1 帧内模式选择实验

所有帧都使用 I 帧进行编码. 图 3 和表 1 分别给出了压缩性能和消耗时间的比较, 从图中可以看出, 两种方法的压缩性能差别很小.

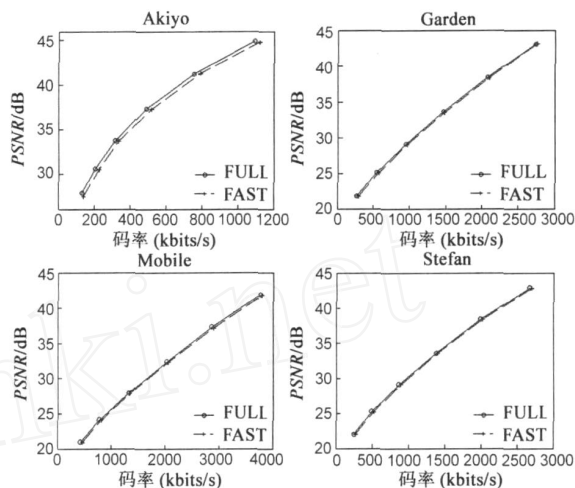


图 3 视频序列率失真比较

根据表 1 的结果, 本文提出的方法总编码时间是全搜索法的 70% 左右; 如果仅考虑选择宏块类型部分的耗时, 本文方法的耗时仅为全搜索法的 20% ~ 25%.

表 1 “FAST”和“FULL”耗时比较

序列	总编码时间(秒)		选择宏块类型时间(秒)	
	FAST	FULL	FAST	FULL
Akiyo	3.218	4.690	0.478	2.344
Garden	3.557	4.898	0.469	1.989
Mobile	5.105	6.630	0.650	2.520
Stefan	4.210	5.833	0.578	2.377

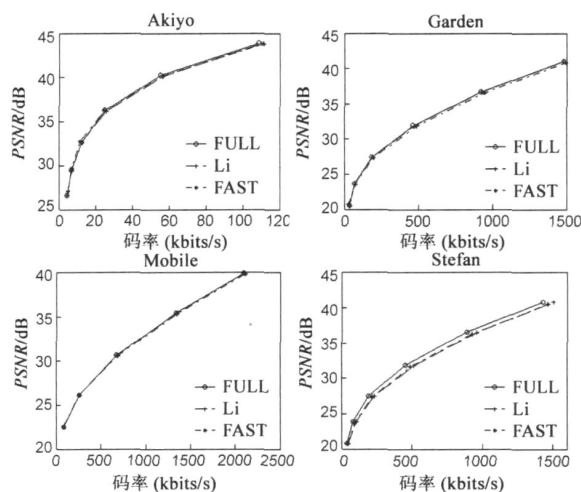


图 4 视频序列率失真比较

#### 3.2 帧间模式选择实验

只有第一帧使用 I 帧, 其余全部使用 P 帧编码. 全

搜索法中运动预测的搜索步长为 16. 本文还实现了文献<sup>[3]</sup>提出的方法(Li). 根据图 4 的实验结果, 本文的方法和 Li 的方法压缩性能非常相近, 两者都与全搜索法差别不大.

表 2 给出了三种方法的消耗时间对比. 相对于全搜索法, Li 的方法可以将总编码时间提高约 5.5 倍, 本文方法可以提高约 15 倍; 只考虑选择宏块类型时间, Li 的方法平均能将速度提高 6 到 7 倍, 而本文方法可以提高 35 倍左右, 从而极大的节省了选择宏块类型的时间.

表 2 “FAST”和“FULL”耗时比较

序列	重新编码时间(秒)			选择宏块类型时间(秒)		
	FAST	Li	FULL	FAST	Li	FULL
Akiyo	2.19	6.48	37.05	0.65	4.54	35.29
Garden	4.09	9.58	58.37	1.86	7.13	56.18
Mbible	5.64	11.78	83.48	2.56	8.45	80.33
Stefan	5.13	20.48	75.45	2.22	17.51	72.86

#### 4 结论及后续工作

在基于 H.264 的降低图像尺寸的转码系统中, 需要为降低尺寸后图像中的宏块重新选择类型. 本文提出了一种快速的宏块类型选择方法, 其中包含帧内和帧间模式选择两部分. 在帧内模式选择部分, 本文提出的方法利用原始图像中四个宏块的非零系数总个数为准则来选择宏块类型. 在帧间模式选择部分, 利用  $num_{nz}$  为准则划分当前宏块所在区域的运动性质, 根据不同的运动性质, 跳过部分候选宏块类型, 从而节省了计算时间. 本文使用指数模型来描述  $num_{nz}$  与重新量化参数的关系. 最后的实验结果表明, 与全搜索法相比, 在压缩性能相差不大的情况下, 本文提出的方法可以大幅度的提高宏块类型选择的速度. 很多研究人员已经提出了基于 H.264 的快速帧内/帧间模式选择的方法<sup>[1], [2]</sup>, 在下一步的研究中, 可以探讨如何将这选择方法应用到转码系统中.

#### 参考文献:

[1] Tierr Ying Kuo, Chen-Hung Chan. Fast variable block size motion estimation for H.264 using likelihood and correlation of

motion field[J]. IEEE Trans on Circuit and system for video technology, 2006, 16(10): 1185 - 1195.

[2] 李世平, 蒋刚毅, 郁梅. 快速帧内预测模式选择新方法[J]. 电子学报, 2006, 34(1): 141 - 146.

LI Shi-ping, JIANG Gang-yi, YU Mei. New fast mode selection for intra prediction[J]. Acta Electronical Sinica, 2006, 34(1): 141 - 146. (in Chinese)

[3] Chi-Hung Li, Chung-Neng Wang, Tihao Chiang. A fast downsizing video transcoder based on H.264/AVC standard[A]. PCM. 2004[C]. Tokyo, Japan, November 2004, LNCS 3333: 215 - 223.

[4] Huifen Shen, Xiaoyan Sun, Feng Wu, Houqiang Li, Shipeng Li. A fast downsizing video transcoder for H.264/AVC with rate-distortion optimal mode decision[A]. IEEE International conference on Multimedia and Expo (ICME '2006)[C], Toronto, Ontario, Canada, 2006. 2017 - 2020.

[5] Kai-Tat Fung, Yui-Lam Chan, Warr-Chi Siu. New architecture for dynamic frame-skipping transcoder[J]. IEEE Trans Image processing, 2002, 11(8): 886 - 900.

[6] Damien Lefol, Dave Bull. Mode refinement algorithm for H.264 inter frame requantization[A]. IEEE International conference on Imaging Process (ICIP '2006)[C]. Atlanta, Ga, USA, 2006. 845 - 848.

[7] Jan De Cock, Stijn Notebaert, Peter Lambert, Davy De Schrijver, Rik Van de Walle. Requantization transcoding in pixel and frequency domain for intra 16 × 16 in H.264/AVC[A]. ACIVS 2006[C]. Antwerp, Belgium, September 2006, LNCS 4179: 533 - 544.

[8] 修晓宇, 卓力, 沈兰荪. 一种基于 PID 控制器的 H.264 比特率转码方案[J]. 电子学报, 2006, 34(6): 1062 - 1065.

Xiaoyu Xiu, Li Zhuo, Lansun Shen. An H.264 bit rate transcoding scheme based on PID controller[J]. Acta Electronical Sinica, 2006, 34(6): 1062 - 1065. (in Chinese)

[9] Damien Lefo, Dave Bull, Nishan Canagarajah. Performance evaluation of transcoding algorithms for H.264[J]. IEEE Trans on Consumer electronics, 2006, 52(1): 215 - 222.

[10] 王松桂, 陈敏, 陈立萍. 线性统计模型: 线性回归与方差分析[M]. 北京: 高等教育出版社, 2004.

#### 作者简介:



刘兆广 男, 1977 年生于山东菏泽, 分别于 1998 年和 2001 年在西安理工大学获得工学学士学位和工学硕士学位. 现为山东大学信息科学与工程学院博士研究生. 主要研究方向为视频编解码、视频转码等.

E-mail: liuzhg1031@yahoo.com.cn



彭玉华 女, 1966 年生于山东郓城, 教授、博士生导师, 中国电子学会会士、中国电子学会高级会员、电子测量与仪器学会微波毫米波测试专业委员会副主任委员、中国通信学会会员. 现为山东大学信号处理与计算机视觉研究所所长, 研究兴趣包括多尺度分析、小波分析与数字信号处理、低信噪比信号分析、视频编解码技术、数字图像处理及在生物医学工程、生物特征

识别领域的应用等.