

基于残差能量的 H.263 与 H.264 间的编码转换

张 健, 沈庭芝, 姜 微

(北京理工大学信息科学技术学院电子工程系, 北京 100081)

摘 要: 论述当前的视频转换编码技术, 提出一种低复杂度的从 H.263 到 H.264 的编码转换结构。该结构基于视频编码中运动估计的残差能量最小的原理, 使用自适应能量分类的方法对转换编码进行优化。实验结果显示, 该方法相对于 H.264 全模式搜索能够节省 50% 以上的时间。与其他快速算法相比, 具有较高的编码转换性能。

关键词: 编码转换; 帧间预测; 模式选择; 离散余弦变换

Transcoding from H.263 to H.264 Based on Residual Energy

ZHANG Jian, SHEN Ting-zhi, JIANG Wei

(Department of Electronic Engineering, School of Information Science and Technology, Beijing Institute of Technology, Beijing 100081)

【Abstract】 This paper discusses the current transcoding, proposes a low complexity transcoding structure from H.263 to H.264. The method is based on the principle of the minimum residual energy of motion estimation. Adaptive energy method is layered to reduce the time of inter-frame prediction modes selection on re-motion estimation. Experimental results show that compared with full modes search method, it can save more than 50% of the time. With other fast transcoding algorithms, the method has lower complexity, higher video quality, and a shorter time transcoding.

【Key words】 transcoding; interprediction; mode decision; Discrete Cosine Transform(DCT)

1 概述

随着 Internet 和多媒体通信业务的发展, 网络视频的应用越来越广泛。在通用的多媒体访问中, 介入网络的形式十分多样化, 包括 LAN、ADSL、无线网络、ISDN 和拨号网络。这些不同的网络具有不同的信道特征, 包括带宽、位误码率和报文丢失率。同时, 不同的用户终端, 如手持计算机、PDA、机顶盒等具有不同的计算和显示能力。为了适应在不同的网络条件下向不同的终端用户传输数据, 视频参数必须能够根据用户环境进行动态的调整。视频转码就是完成此类任务的关键性技术之一^[1]。

针对最新的 H.264/AVC 视频编码标准, 国际视频联合组织发布的一种更高压缩率以及更高视频质量的视频编码标准, 越来越多地应用在这些设备中, 并且逐步代替数字视频系统中的 H.263 与 MPEG2 视频编码标准。比较 H.263、MPEG2、MPEG4, H.264/AVC 能够完成更高的编码效率, 这是因为它具有的帧内编码和帧间编码模式, 即使这些模式会产生更高的编码复杂度。因此, 如何减小 H.264 的复杂度成为一个研究的热点^[2]。

随着 H.264 标准的提出, 关于 H.264 的异构转换编码技术变得越来越困难。在 H.264 中引入了新的 4×4 块的整数离散余弦变换(1CT)。此外, 转换编码还必须在像素域内进行非线性的去块效应滤波。被滤波后的图像被送回来以重构原图像, 避免在转换编码和连续解码之间的“漂移”现象出现。就复杂度而言, 在频率域内采用 DCT-to-1CT 的方法并不尽如人意。另外一点困难是, H.264 比起使用 DCT 变换的其他标准来说在帧内和帧间宏块预测上都有更高的参数范围。

本文提出一种低复杂度的宏块模式选择方法。该方法基于残差能量的评估进行帧间预测, 用来组成低复杂度的异构转换编码器的一部分。

2 H.264 视频编码中的帧间预测

在 H.264 标准中, 帧间预测的宏块模式选择是最花费时间的处理。这是由于可变块运动估计、四分之一像素运动补偿以及帧内预测这些方法的使用而引起的。

自从 H.261 以来, 基于块的方法使用在每一个主要的编码标准中。与早期的标准最大的不同包括: H.264 支持更多的块尺寸(最低到 4×4), 更好的子像素运动向量(亮度部分四分之一像素)。H.264 支持运动补偿块大小从 16×16 到 4×4。每一个宏块的亮度部分可能被分割成 4 种方式: 16×16, 16×8, 8×16 或者 8×8。每一个被分割的区域叫作宏块的子块。如果采用了 8×8 的模式, 每一个 8×8 的宏块子块又可以被分割成更小的 4 种方式: 8×8, 8×4, 4×8 或者 4×4。在每一个宏块中, 这些子块的分割产生了非常多的组合方式。众所周知, 这种分割宏块成各种子块进行运动补偿的方法被称为树型结构运动补偿。

每一个子块或者子块的子块都要求有一个单独的运动向量。每一个运动向量也必须被编码和传输, 此外, 子块的选择方法也必须被编码传输。如果选择一个很大的子块尺寸(如 16×16、16×8、8×16), 则只需要很少的比特来标记运动向量的选择和子块的划分类型。但是, 在高细节的区域, 运动补偿的残差能量可能会显著地增加。如果选择一个小尺寸的分割(8×4、4×4 等), 则在运动补偿之后能够得到更低的残差能量, 但是它需要很多的比特来编码运动向量以及子块的选择方法。因此, 子块尺寸的使用在压缩性能上有一个很重要的影响。总之, 大尺寸子块适合于图像的均匀区域而小尺寸的子块适用于高细节的区域。

基金项目: 国家自然科学基金资助项目(60772066)

作者简介: 张 健(1981 -), 男, 博士研究生, 主研方向: 多媒体视频编解码; 沈庭芝, 教授、博士生导师; 姜 微, 博士研究生

收稿日期: 2010-02-20 **E-mail:** zpower007@163.com

3 提出的方法

3.1 残差能量特性

帧间模式的性能是由残差的能量大小来决定的。在视频的运动估计与运动补偿中,需要计算当前块和参考帧中一系列相邻块的差值的能量,通常,均方差可以用来衡量残差块中的能量。一个 $N \times N$ 块的 MSE 计算如下:

$$MSE = \frac{1}{N^2} \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} (C_{ij} - R_{ij})^2$$

其中, C_{ij} 是当前块的像素; R_{ij} 是参考区域的像素。

在运动估计中,当计算出最佳模式后,需要对残差进行 DCT 变换以及量化,量化后的 DCT 系数经过编码作为输出码流。而在转换编码中,对接收码流进行反解码后得到的就是经过量化后的 DCT 系数。

对残差进行 DCT 变换的公式如下:

$$F_{xy} = \frac{C(x)C(y)}{4} \sum_{i=0}^7 \sum_{j=0}^7 f_{ij} \cos\left(\frac{(2i+1)x\pi}{16}\right) \cos\left(\frac{(2j+1)y\pi}{16}\right)$$

其中, $f_{ij} = C_{ij} - R_{ij}$ 。

图像的能量在 DCT 域内有

$$F_{xy}^2 = \left(\frac{C(x)C(y)}{4} \sum_{i=0}^7 \sum_{j=0}^7 f_{ij} \cos\left(\frac{(2i+1)x\pi}{16}\right) \cos\left(\frac{(2j+1)y\pi}{16}\right) \right)^2$$

$$F_{xy}^2 = T_8 f_{ij} T_8^T f_{ij} T_8^T = T_8 f_{ij}^2 T_8^T = \frac{C(x)C(y)}{4} \sum_{i=0}^7 \sum_{j=0}^7 f_{ij}^2 \cos\left(\frac{(2i+1)x\pi}{16}\right) \cos\left(\frac{(2j+1)y\pi}{16}\right)$$

根据图像的能量守恒定律以及 DCT 变换的正交性可知,使用的宏块残差能量的测量方法是:在 H.263 中,运动补偿预测后再进行量化的 DCT 系数的绝对平方和。

3.2 模式统计特性

前文分析了实际传输的宏块残差是能量最小的判断结果,那么,从 H.263 到 H.264 的转换编码中,传统的方法需要在 H.264 中重新进行运动估计并找出所有模式中最好的结果。而实际中,如果 H.264 中残差的能量已经很小了,而在 H.264 中重新做运动估计时找到的最好的模式仍然是 H.263 中的 16×16 模式或者 8×8 模式。而这实际上形成了一个重复的工作,增加了计算的复杂度。

在 H.264 中,还引入了一个新的模式判断方法——比特失真最优化。实际上,在增加宏块分割的复杂度的同时,还会增加运动向量的编码,因此,在 H.264 中,并不是宏块分割得越细编码效果越好。考虑到运动向量编码的关系,采用 RDO 的方法平衡两者之间的关系。引入大量的分析和统计,如图 1 所示,用来观察实际的编码中各个模式所占的比重。

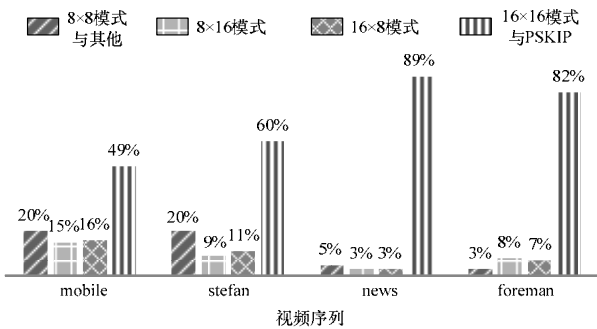


图 1 4 种视频序列中各个模式所占的百分比

从图 1 中可以看出,在 H.264 中,即使运动很复杂的 mobile 与 stefan 视频序列, 16×16 模式以及 PSKIP(也是 16×16) 模式占据了 50% 以上。随着运动程度的减小, 16×16 模式所

占据的百分比也逐渐增加。由此,在转换编码时,很多的宏块不需要重新再做运动估计,而直接使用 H.263 中的 16×16 模式或者 8×8 模式以及运动向量。所以,如何判断 H.263 中哪些宏块可以不需要重新编码是下文讨论的问题。

首先,分析 H.263 中宏块的能量统计特性。从图 2 中可以看出,在 H.263 中,大部分的宏块残差能量都很低或者等于零,即使在有复杂的运动背景的视频系列中,很多宏块的残差能量依然很低。

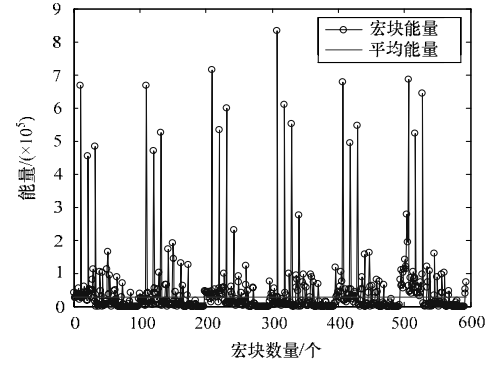


图 2 mobile 序列中宏块的能量分布

因此,可以设定一个阈值,用来判断哪些宏块不需要重新进行运动估计。但是,如果事先设定阈值,会造成该算法的使用范围受到限制,因此,根据整幅图像的残差能量特性来自适应地调整阈值。

3.3 算法

通过上面的观察和分析,本文采用自适应能量分层的方法来节省转换编码的时间。如果 H.263 中宏块的残差能量很小,可以认为之前模式的性能足够好,可以关闭运动估计功能,直接使用 H.263 中的运动向量与宏块模式。另一方面,如果 H.263 中宏块残差的能量非常大,那么需要在 H.264 编码器中对该宏块重新进行运动估计。此外,当重新进行运动估计时,使用全模式搜索法来处理。

3.3.1 阈值设定

阈值设定步骤如下:(1)计算当前块所在该帧图像的所有 8×8 块的能量。(2)计算当前帧图像的宏块的平均能量 E 。(3)设置能量分层阈值。设定参数 A ,分层阈值为 AE (一般情况 A 设为 1)。

3.3.2 早期中断

每个宏块分为 4 个 8×8 的子块,计算 4 个子块的残差能量和是否大于阈值,如果大于阈值,则需要重新进行运动估计。当计算到某个子块的和,若时残差能量已经大于阈值,则中断程序,不需要再累加计算,直接判断该宏块需要重新进行运动估计。

3.3.3 分层处理

为了提高效率,整个处理分为 2 层。第 1 层不需要重新进行运动估计。宏块的模式和运动向量直接使用 H.263 中的信息,减小了重新编码的时间。第 2 层在 H.264 标准下重新进行运动估计。为了提高编码效率,采用 RDO 的方法对 9 种帧间模式以及帧内模式依次进行计算,找到最优的模式和运动向量进行编码。

4 仿真结果

将提出的低复杂度宏块模式选择编码应用到 H.264 参考软件 JM12.2 中。H.263 视频首先被解码,在算法中需要的宏块信息被提出并存储起来。提出的信息包括宏块的残差 DCT

系数、宏块的模式(16×16 或者 PSKIP)以及宏块的运动向量。低复杂度的宏块选择并入在 H.264 的编码端。H.264 编码器读入解码后的 H.263 视频数据,宏块的模式通过提出的方法做出选择。

4.1 本文算法与 H.264 全模式搜索法的比较

首先把提出的算法与在 JM12.2 中对所有的模式进行搜索(FMS)后得到的转换编码的性能进行比较。对比了 3 个视频序列:mobile, carphone, stefan。视频图像的格式为 QCIF (176×144),在 H.264 编码端量化参数为 28,视频图像的帧数为 300 帧。表 1 显示了在上面条件下测得的 2 种方法的结果。

表 1 全模式搜索方法与提出的方法数据对比

视频序列	方法	信噪比/dB	比特率/(Kb·s ⁻¹)	时间/s
mobile	FMS 方法	34.33	1 121.80	83.378
	本文方法	34.34	1 122.22	41.351
carphone	FMS 方法	37.35	173.46	55.679
	本文方法	37.37	174.59	24.356
stefan	FMS 方法	34.79	871.93	78.553
	本文方法	34.78	872.35	34.736

表 2 显示了对应的节省时间、改变的信噪比值以及改变的比特率的结果。对比 H.264 中的全部模式的搜索方法,提出的方法很好地保持了原来的图像质量,并且比特率的增加非常得小。在低比特率的情况下,视频的比特率增加较多;在高比特率的信道中,视频的比特率仅仅只增加 0.02%左右。而提出的方法比起 H.264 中的全模式搜索,提高了 50%以上的转换编码时间,这大大减少了转换编码的时间。

表 2 全模式搜索方法与提出的方法的性能对比

序列	时间/(%)	信噪比/dB	比特率/(%)
mobile	50.4	+0.01	0.03
carphone	56.3	+0.02	0.60
stefan	55.8	-0.01	0.05

图 3 显示了 2 种方法的 RD 性能图,序列模式为 178×144,帧数为 300 帧。从图中可以看出,相比于性能最好的 H.264 中的全模式搜索法,提出的方法的 RD 曲线基本上能与其保持一致。

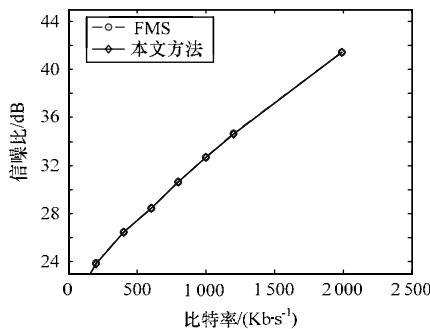


图 3 本文方法与 FMS 的 RD 曲线

4.2 本文算法与不同的快速宏块模式选择算法的比较

把目前提到的一些快速算法综合在一起做比较。

从表 3 可以看出,对 mobile 序列,提出的算法平均损失的图像质量和增加的比特率都是最小的,但是却能节省最大的转换编码时间(超过 50%)。

表 3 本文算法与各种不同快速算法之间的比较

算法	时间/(%)	信噪比/dB	比特率/(%)
文献[3]的算法	-10.96	-0.01	0.01
文献[4]的算法	-22.83	-0.02	0.07
文献[5]的算法	-46.93	-0.01	0.10
文献[6]的算法	-87.60	-0.18	3.86
本文算法	-50.40	+0.01	0.03

5 结束语

本文提出了一种简单高效的 H.264 帧间预测宏块模式选择的算法,实现了从 H.263 到 H.264 的快速转换。该算法具有低复杂度的特点,它只要求对残差的能量进行统计,从而简化模式选择的判断。同时,提出的转换编码算法采用了自适应能量阈值的方法,对每一帧图像的阈值都基于该图像本身,而不受其他参考帧的影响。

实验结果显示,提出的算法保持了一个非常高的图像质量。同时,提出的算法只需对 H.264 的参考软件做较小的改动就能实现新的算法。最后,比较了一些其他的帧间预测快速算法。结果显示,提出的算法能够完成最好的视频转换编码的结果。

参考文献

- [1] Fernandez-Escribano G, Kalva H, Cuenca P, et al. A Fast MB Mode Decision Algorithm for MPEG-2 to H.264 P-frame Transcoding[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2008, 18(2): 172-185.
- [2] Zhang Jian, Shen Tingzhi, Jiang Wei, et al. A Fast Algorithm for Intra Downsizing in H.264/AVC Transcoding[C]//Proc. of International Colloquium on Computing, Communication, Control, and Management. Guangzhou, China: [s. n.], 2008: 496-499.
- [3] Lim K P. Fast Inter Mode Selection[Z]. Joint Video Team of MPEG and VCEG, 2003.
- [4] Jeon B, Lee J. Fast Mode Decision for H.264[Z]. Joint Video Team of MPEG and VCEG, 2003.
- [5] Kim Y H, Yoo J W, Lee S W. Adaptive Mode Decision for H.264 Encoder[J]. Electronic Letters, 2004, 40(19): 1172-1173.
- [6] Lu Xiaolan, Tourapis A M, Yin Peng, et al. Fast Mode Decision and Motion Estimation for H.264 with a Focus on MPEG-2/H.264 Transcoding[C]//Proc. of IEEE International Symposium on Circuits and Systems. Kobe, Japan: [s. n.], 2005: 1246-1249.

编辑 顾逸斐