

基于卡尔曼滤波的 H. 264 头信息量估计 及其在帧级码率控制中的应用

陈 杰 张文军 杨小康 谈永敏

(上海交通大学图像通信与信息处理研究所, 上海 200240)

摘 要 由于 H. 264 算法中宏块划分的灵活性导致了头信息的数据量变化剧烈而难以预测, 同时由于变换系数的数据量大大的减小, 又使头信息量在码流中所占的比重远高于在先前各个标准的码流中的比重, 因此准确估计头信息量能有效提高 H. 264 码率控制算法的性能。为了准确估计头信息, 提出了一种应用卡尔曼滤波器估计帧级总头信息比特数的方法, 并将该方法应用于帧级的码率控制。实验结果表明, 该算法能准确估计头信息, 并能在一定程度上提高相同码率下的总体图像质量。

关键词 H. 264 码率控制 头信息 卡尔曼滤波

中图分类号: TN919. 81 **文献标识码**: A **文章编号**: 1006-8961(2006)11-1640-05

Estimation of the Number of Header Bits Using Kalman Filter and Its Application in Frame Layer Rate Control for H. 264

CHEN Jie, ZHANG Wen-jun, YANG Xiao-kang, TAN Yong-min

(Institute of Image Communication and Information Processing, Shanghai Jiao Tong University, Shanghai 200240)

Abstract H. 264/MPEG-4 AVC features flexible macroblock mode decision, which makes the number of data bits produced by encoding header information in H. 264 video stream vary irregularly and hard to be predicted. As H. 264 encoder reduces the data bits for transform coefficients effectively, the number of header bits takes a larger portion in the overall bits than that of any former video compression standard. Thus precise estimation of header bits can effectively improve the performance of the H. 264 rate control algorithm. In this paper, we apply the Kalman Filtering theory to improve the precision of estimating the number of the header bits at the frame layer, and incorporate the estimation model to frame layer rate control.

Keywords H. 264, rate control, header information, kalman filtering

1 引 言

1.1 码率控制及其对视频通信的意义

图像通信中码率控制的目的, 就是通过调节编码参数来控制单位时间内的编码视频流的数据量, 以使产生的比特流符合各种应用需求。

在许多视频应用中, 包括已有相当基础的数字

电视广播和正在兴起的手机电视, 都对码率有较严格的要求, 其中信源编码可生成码率相对恒定和可控的原始码流, 其有利于在经过其后的多道传输步骤(复用-信道编码-调制-发送-接受-解调-信道解码-解复用等)后, 保证最终实时信源解码的播放效果。

1.2 码率控制原理

MPEG-2 TM5 模型^[1]与率失真优化(rate-distortion optimization, RDO)方法是码率控制的基

基金项目: 国家自然科学基金重点项目(6033203)

收稿日期: 2006-06-15; 改回日期: 2006-08-10

第一作者简介: 陈杰(1981~), 男, 2004 年获上海交通大学学士学位, 现为上海交通大学在读硕士研究生。研究方向为 H. 264 视频传输系统与码率控制。E-mail: jie.chen@sjtu.edu.cn

础, TM5 模型中的码率控制策略可通过预分配的目标比特数来控制每帧图像的编码比特数, 以保证编码和解码缓冲区不发生上溢及下溢; 而率失真优化方法则更进一步, 可在限制编码码率的前提下求得最小的编码失真度, 并可通过调节量化参数 (quantization parameter, QP), 来达到控制码率的目的。文献[2]给出了基于率失真优化的码率控制的一般方法, 即拉格朗日最优化方法。

文献[3]基于率失真优化原理, 给出低缓存延迟的码率控制方案, 是 H.264 码率控制的重要基础。

1.3 H.264 码流头信息特点与码率控制策略

H.264 视频压缩码流中, 除了亮度分量 Y 和色差分量 Cb 、 Cr 在运动补偿后的残差变换系数外, 还包括: 帧头信息 (header, 一般固定)、宏块模式 (macroblock mode)、运动信息 (motion information)、编码块样式 (coded block pattern, CBP)、量化参数变化值 (delta quantization) 和填充比特 (stuffing bits)。为了描述方便, 本文中, (广义) 头信息指码流中除亮度、色差系数外的所有信息, 包括相对固定的帧头信息 (狭义头信息)。

由于早前的视频编码标准的宏块较大, 无法做到完美的运动补偿, 从而造成运动补偿后的各像素位置上的残差较大, 但由于头信息与运动补偿残差变换系数相比, 其数据量小且相对固定, 因此传统码率控制策略一般只针对运动补偿后的残差阵列, 而头信息数据量则被当做一个常数处理。H.264 算法中由于宏块划分的灵活性决定了其头信息的数据量变数大而难以预测, 且在码流中所占的比重远高于在先前各个标准的码流中的比重, 因此必须改进传统上的码率控制算法, 这也是本研究工作的重点。

H.264 联合编码模型 (joint model, JM) 采用 Li 等在文献[4]中提出的码率控制方法: 用已编码帧中宏块的平均绝对残差 (mean absolute difference, MAD) 值来线性预测当前帧的 MAD 值, 以便得出预测量化步长供 RDO 使用。文献[5]对文献[4]进行了改进, 其不使用已编码帧预测, 而是考虑到了码流头信息与 QP 的关系, 对当前 P 帧, 采用帧间 16×16 的模式进行一遍运动估计预编码, 并将其结果作为 RDO 的基础, 来得到 QP, 再分配码率, 从而提高了准确性。而文献[6]又通过对最终未采用帧间 16×16 的宏块进行更准确的 RDO 来进一步改进了上述算法。

文献[5,6]提出的方法虽然考虑到了头信息 (包括运动信息) 在码率中所占的重要比例, 但没有进行预测, 仅根据平均头信息比特数与 $P16 \times 16$ 预测模式得到的纹理残差总和的近似线性关系, 来粗略确定待编码帧的头信息数据量, 而且在帧级码率分配中, 仍将头信息与残差系数合并作为率失真优化的参数, 故在码率控制精确度上仍有提升的必要和余地。

1.4 H.264 码流头信息量估计

本文对 H.264 码流头信息量的估计进行了研究, 其出发点是基于视频序列各帧中的运动信息量与运动的剧烈程度关系密切, 即决定于视频的内容, 而不能单纯依靠率失真优化得到的量化参数值控制。在传统的通过 QP 控制码率的基础上, 对各帧的广义头信息量 (除纹理残差信息外的剩余比特数, 包括运动信息) 根据时间变化规律进行预测, 可以提高码率控制精确度。

基本方法是利用已编码帧的头信息比特数来预测当前帧的头信息比特数, 其中还应用最优估计方法 (卡尔曼滤波器) 修正预测值。

同时, 尝试将取得的成果应用到对文献[6]提出的算法的改进上, 并基于 H.264 算法的标准参考模型 JM (版本 9.3) 进行了对比分析。

2 最优估计方法与卡尔曼滤波的应用

最优估计起源于时间序列分析, 是基于时间序列模型对时间序列进行预报和控制, 包括信号和状态估计, 也称最优滤波。卡尔曼滤波方法是一种时域方法, 基于状态空间模型和射影理论, 可用来解决状态估计问题, 因为是递推算法, 所以便于在计算机上实现, 且可处理多变量、时变、非平稳时间序列滤波问题^[7,8]。卡尔曼滤波方法适用的系统包括状态和观测模型, 其可以分别用下列线性离散方程描述:

$$\begin{aligned} \mathbf{x}(k+1) &= \Phi(k+1, k)\mathbf{x}(k) + \mathbf{B}(k+1, k)\mathbf{u}(k) + \mathbf{\Gamma}(k+1, k)\mathbf{w}(k) \\ \mathbf{z}(k+1) &= \mathbf{H}(k+1)\mathbf{x}(k+1) + \mathbf{v}(k+1) \end{aligned} \quad (1)$$

式中, $\mathbf{x}(k+1)$ 为 n 维随机序列; $\mathbf{u}(k)$ 为 r 维控制输入序列; $\mathbf{w}(k)$ 为 p 维系统噪声序列; $\mathbf{z}(k+1)$ 为 m 维观测序列; $\mathbf{v}(k+1)$ 为 m 维观测噪声序列, 且系统噪声 $\mathbf{w}(k)$ 和观测噪声 $\mathbf{v}(k+1)$ 为零均值白噪声序列, 即对所有 k, j , 有如下公式^[8]:

$$\begin{cases} E[\mathbf{w}(k)] = 0 \\ E[\mathbf{v}(k)] = 0 \\ E[\mathbf{w}(k)\mathbf{w}^T(j)] = Q(k)\delta_{k,j} \\ E[\mathbf{v}(k)\mathbf{v}^T(j)] = R(k)\delta_{k,j} \\ E[\mathbf{w}(k)\mathbf{v}^T(j)] = 0 \end{cases} \quad (2)$$

在没有控制量 $\mathbf{u}(k)$ 的情况下, 最优滤波估计 $\hat{\mathbf{x}}(k) = \mathbf{x}(k|k)$ 由下列递推关系式得到:

$$\begin{cases} \mathbf{x}(k|k) = \Phi(k, k-1)\mathbf{x}(k-1|k-1) + \\ \quad \mathbf{K}(k)[\mathbf{z}(k) - \mathbf{H}(k)\Phi(k, k-1) \times \\ \quad \mathbf{x}(k-1|k-1)] \\ \mathbf{K}(k) = \mathbf{P}(k|k-1)\mathbf{H}^T(k)[\mathbf{H}(k) \times \\ \quad \mathbf{P}(k|k-1)\mathbf{H}^T(k) + \mathbf{R}(k)]^{-1} \\ \mathbf{P}(k|k-1) = \Phi(k, k-1)\mathbf{P}(k-1|k-1) \times \\ \quad \Phi^T(k, k-1) + \mathbf{I}(k, k-1) \times \\ \quad \mathbf{Q}(k-1)\mathbf{I}^T(k, k-1) \\ \mathbf{P}(k|k) = [\mathbf{I} - \mathbf{K}(k)\mathbf{H}(k)]\mathbf{P}(k|k-1) \\ k = 0, 1, \dots \end{cases} \quad (3)$$

卡尔曼滤波原来是在自动控制中用来处理多变量、时变、非平稳时间序列滤波问题而发展起来的方法, 由于其递推算法的特征决定了它不需要保留已知的观测序列 $\mathbf{z}(k)$, 而只需要很少的用于存储中间变量的内存空间, 就能准确地估计实际状态, 同时计算量也相对较小, 因此能满足对复杂度限制较多的视频编码——特别是实时编码的码率控制需要。同时码率控制中也需要不断更新模型参数, 而这种相似性则有利于卡尔曼滤波器与码率控制的整合。文献[9, 10]对其在视频编码——包括运动估计和码率控制中的应用进行了初步尝试, 并验证了将其转到本文的视频码率控制的理论可行性。

3 头信息比特数估计模型

3.1 对文献[6]算法头信息比特估计模型的分析

在文献[6]的码率控制算法中, 头信息比特数独立于变换系数比特数用于进行码率估计, 在进行了帧间 16×16 预测之后, 即可使用如下的简单的宏块头信息比特数估计模型:

$$H_i = C \times com_i \quad (4)$$

其中

$$com_i = \begin{cases} H_{\text{vd}}/C, & \sigma_i^2 \leq \sigma_{\text{vd}}^2 \\ [\log(\sigma_i^2)]^2, & \text{else} \end{cases} \quad (5)$$

上式中 H_i 是当前帧中第 i 个宏块的头信息比特数, σ_i 是用帧间 16×16 模式预编码之后运动补偿残差的标准差, com_i 为当前宏块的复杂度, C 是表征 H_i 与 com_i 线性关系的常数。记前一帧中头信息比特数 H_i 小于预定义常数 ($H_i = 11$) 的宏块的 H_i 和 σ_i^2 的平均值分别为 H_{vd} 和 σ_{vd}^2 。当宏块 σ_i^2 小于或等于 σ_{vd}^2 时, 则认为该宏块的头信息量很小, 而直接用 H_{vd} 作为当前宏块 H_i 的估计值。否则认为宏块的头信息比特数与 $com_i = [\log(\sigma_i^2)]^2$ 成线性。

此外, 在完成每个宏块的编码后, 自适应地更新 C , 以保持码率控制模型健壮稳定。

式(6)在运动剧烈的情况下 H_i 与 $[\log(\sigma_i^2)]^2$ 成线性关系的假设是基于经验的, 缺乏说服力。

通过使用 Matlab 对已编码的几个序列进行分析可以发现, 无论从帧级还是从宏块级来看, C 的变化相当大, 与其常数的假设是相当矛盾的。

因此本文针对头信息比特估计提出以下码率控制改进方案。

3.2 帧级头信息比特数直接估计模型

本文估计头信息比特数时, 完全放弃原来的头信息比特数估计模型(式(4), 式(5)), 而直接用前一帧的头信息比特数经过卡尔曼滤波来预测当前帧的头信息比特数。同时在宏块层根据每帧中已编码宏块的头信息比特数来自适应更新全帧的头信息比特数的预测值。

本文把预测的每帧头信息比特数作为状态序列元素 $\mathbf{x}(k)$, 其中 k 为帧序号; 而以实际已编码的前一帧的头信息比特数作为当前状态的观测值 $\mathbf{z}(k)$ 。假设系统为线性定常时不变的, 又因为输入输出都是 1 维的序列, 且没有控制量 $\mathbf{u}(k)$, 所以系统模型(式(1))可简化为

$$\mathbf{x}(k+1) = \Phi(k)\mathbf{x}(k) + \mathbf{w}(k) \quad (6)$$

$$\mathbf{z}(k+1) = \mathbf{H}(k+1)\mathbf{x}(k+1) + \mathbf{v}(k+1)$$

其中, $\Phi = 1, H = 1$ 为常量; \mathbf{w} 和 \mathbf{v} 的均方差 $Q(k)$ 与 $R(k)$ 也设为常量 Q 和 R , 取 $Q = 0.1, R = 22$ 。

因为本模型主要用于低延时的码率控制, 故采用 IPPP... 序列。卡尔曼滤波器的输入参数为已编码帧的头信息比特数, 而输出则作为下一帧的头信息比特数的预测值, 采用式(4)的模型方程。

同时当第 i 块宏块编码完成后, 再自适应更新总头信息比特数估计值为

$$H_{\text{pred}} = \sum_{j=1}^i h_j + H_{\text{pred}} \times \frac{N-i}{N} \quad (7)$$

其中, h_j 是已编码的第 j 个宏块的真实头信息比特数; N 为一帧的总宏块数, 对 CIF 序列, $N = (352 \times 288) / (16 \times 16) = 396$ 。

4 实验结果与分析

为了检验所提出的基于卡尔曼滤波的头信息量估计方法的准确性及其对整个码率控制算法的影响, 对 5 个 CIF 标准序列进行了测试, 已涵盖了多种运动复杂度。实验中对每个标准序列的所有帧进行了编码。具体的测试序列信息与编码参数分别如表 1、2 所示。

表 1 测试序列信息

Tab. 1 Information of video sequences for the test

测试序列	格式	帧率 (fps)	I 帧量化参数值	总编码帧数	帧类型
paris	CIF	30	20	1065	IPPP
tempete	CIF	30	28	260	IPPP
highway	CIF	30	28	2000	IPPP
foreman	CIF	30	36	300	IPPP
mobile	CIF	30	36	300	IPPP

表 2 测试编码选项

Tab. 2 Encoder parameters for the test

运动矢量分辨率	1/4 pixels	受限搜索范围	2
Hadarnard	使用	参考帧	5
RDO	使用	熵编码模式	CABAC
搜索范围	16	条带模式	未用

表 3 给出了应用该模型的每帧头信息比特数估计值与实际值的比较, 同时列出不使用卡尔曼滤波器直接用前一帧的头信息比特数作为当前帧头信息比特数估计值的误差, 以及采用式 (4), 式 (5) 的原模型, 同时用已编码各帧 C 的平均值 \bar{C} 以作为估计下一帧头信息比特数的初始 C 值, 而将得到的下一帧总头信息比特数估计值的误差作为参考。

可见, 采用卡尔曼滤波器估计的误差百分数比直接预测减少三分之一以上, 与原模型相比, 准确度更是大大提高。

表 4 比较了将本文所提出的头信息量估计方法应用于文献 [6] 算法后对 PSNR 的影响。由表 4 可见, 与文献 [6] 算法 (2-step QP refinement) 相比, 其在各种码率及各种测试序列下的编码压缩亮度分

量的信噪比均有稳定提高。特别对一些高码率或高运动复杂度的序列, 原算法效果不佳, 平均信噪比较 JM 的标准码率控制策略的结果差了不少, 而采用卡尔曼滤波对码率头信息进行预测后, 则缩小了信噪比上的差距, 这是因为在这种情况下, H.264 编码器较少采用预编码的帧间 16×16 模式作为最终的宏块预测模式的缘故, 由于本文的算法对头信息数据量的预测独立于预编码提供的信息, 故能有效减小头信息部分的预测误差。

表 3 头信息比特数预测误差比较

Tab. 3 Comparison of header bits prediction error

预测模型	$C_{avg} \sum O_i^{[6]}$		直接前一帧替代		卡尔曼滤波	
	$C_{avg} \sum O_i - H_j$	$ H_j - H_{j+1} $	$ H_j - H_{j+1} $	$ KF(H_j) - H_{j+1} $		
误差式	平均值	平均相对误差 (%)	平均值	平均相对误差 (%)	平均值	平均相对误差 (%)
foreman	4 572.3	93.74	1 114.68	20.11	859.93	16.00
highway	1 033.81	29.43	871.24	24.74	547.1	15.19
mobile	1 861.21	18.78	1 530.97	17.76	1 122.96	12.80
paris	669.22	18.55	1 047.96	33.72	683.12	21.48
tempete	10 769.80	123.89	1 750.29	19.06	1 202.56	13.22

表 4 改进后算法与原算法及 JM 标准码率控制的性能比较

Tab. 4 Results of our Rate Control scheme with Kalman Filter, Rate Control with 2-step QP Refinement and standard JM scheme

测试序列	码率控制方案	PSNR-Y (dB)	增益 (dB)	30fps 时的码率 (bps)	码率变化率 (%)
	JM9.3 RC	34.94		1 076 276	
tempete	RC-QP refine	34.92	-0.02	1 021 135	-5.12
	RC with KF	35.05	0.11	1 045 641	-2.85
paris	JM9.3 RC	42.02		968 614	
	RC-QP refine	42.32	0.3	959 129	-0.98
	RC with KF	42.41	0.39	960 664	-0.82
mobile	JM9.3 RC	27.71		338 392	
	RC-QP refine	27.56	-0.15	335 594	-0.83
	RC with KF	27.72	0.01	335 380	-0.89
highway	JM9.3 RC	38.05		220 313	
	RC-QP refine	38.07	0.02	219 099	-0.55
	RC with KF	38.15	0.1	219 010	-0.59
foreman	JM9.3 RC	31.06		118 681	
	RC-QP refine	31.31	0.25	116 926	-1.48
	RC with KF	31.4	0.34	117 042	-1.38

5 结 论

卡尔曼滤波器虽能在几乎不增加复杂度的条件下,大大提高对各帧头信息数据量的预测准确度,但限于已存在的码率控制系统本身的结构设计,算法对相同码率下总体图像质量的提高相对有限。如何利用预编码时的帧间 16×16 模式预测的结果,结合卡尔曼滤波等最优估计方法来进一步提高码率控制的效果,将是一个有前途的研究方向。

参考文献 (References)

- 1 Motion Picture Expert Group. MPEG-2 Test Model 5 [S]. Doc ISO/IEC JTCl/SC29/WG11/N0400. 1993.
- 2 Wiegand T, Schwarz H, Joch A, *et al.* Rate-constrained coder control and comparison of video coding standard [J]. IEEE Transactions on Circuits and Systems for Video Technology, 2003, 13(7): 688 ~ 703.
- 3 Ribas-Corbera J, Lei S. Rate control in DCT video coding for low-delay communications [J]. IEEE Transactions on Circuits and Systems for Video Technology, 1999, 9(1): 172 ~ 185.
- 4 Li Z G, Pan F, Lim K P, *et al.* Adaptive rate control for H. 264 [A]. In: Proceedings of International Conference on Image Processing(ICIP)[C], Singapore, 2004.
- 5 Li P, Yang X K, Lin W S, *et al.* Buffer-constrained R-D model-based rate control for H. 264/AVC [A]. In: Proceedings of IEEE International Conference on Acoustics Speech, and Signal Processing (ICASSP)[C], Philadelphia, Penn. USA, 2005, 2: 3531 ~ 3534.
- 6 Yang X K, Tan Y M, Ling N. Rate control for H. 264 with two-Step quantization parameter determination but single-pass encoding [J]. In: EURASIP Journal on Applied Signal Processing [EB/OL], <http://www.hindawi.com/GetArticle.aspx?doi=10.1155/ASP/2006/63409>, 2006.
- 7 Deng Zi-li. Optimal estimation theory with applications—Modeling filtering and information fusion estimation [M]. Harbin, China: Harbin Industrial University Press, 2005: 490. [邓自立. 最优估计理论及其应用——建模、滤波、信息融合估计 [M]. 哈尔滨: 哈尔滨工业大学出版社, 2005: 490.]
- 8 Shi Zhong-ke. Computational methods of optimal estimation [M]. Beijing: Science Press, 2001. [史忠科. 最优估计的计算方法 [M]. 北京: 科学出版社, 2001.]
- 9 Chan D Y, Lin S J, Lin C W. Improved rate control via rate-quantization modeling with Kalman Filter [A]. In: Proceedings of IEEE Asia-Pacific Conference on Circuit and System [C], Tainan, Taiwan, China, 2004: 1145 ~ 1148.
- 10 Kuo C M, Chung S C, Shih P Y. Kalman filtering based rate-constrained motion estimation for very low bit rate video coding [J]. IEEE Transactions on Circuits and Systems for Video Technology, 2006, 16(1): 3 ~ 18.