

H. 264编码下基于 DCT系数量化值的 PSNR估计方法

沙 亮, 王贵锦, 赵安邦, 林行刚, 唐 昆

(清华大学 电子工程系, 北京 100084)

摘 要: 针对广泛用于测量视频信号质量的指标峰值信噪比 (PSNR) 因需源信号而受限的问题, 提出一种无需参考源的 PSNR 估计方法。该方法从视频码流中提取离散余弦变换 (DCT) 系数的量化值进行统计, 得到其概率密度函数, 并以此依据 Laplace 模型估计 PSNR。该方法针对最新的视频压缩标准 H. 264 的帧结构、变换和量化结合方式做出了优化, 使得不同的编码速率和帧类型的码流均可适应。使用不同码率和不同运动方式的 H. 264 码流进行测试, 比较使用新方法得到的 PSNR 与实际 PSNR 间的差异。结果显示 CIF 格式的中低码率情况下二者的均值在 1 dB 以内。

关键词: 离散余弦变换 (DCT) 系数; Laplace 分布; H. 264; 峰值信噪比 (PSNR); 量化

中图分类号: TN 919.81

文献标识码: A

文章编号: 1000-0054(2008) 04-0510-04

H. 264 coding PSNR estimation method based on quantized DCT coefficients

SHA Liang, WANG Guijin, ZHAO Anbang, LIN Xinggang,
TANG Kun

(Department of Electrical Engineering, Tsinghua University,
Beijing 100084, China)

Abstract A peak signal-to-noise ratio (PSNR) estimation algorithm that does not use a reference source was developed to address restrictions caused by source signals in calculating the PSNR, which is a widely used index for video quality. The algorithm enables PSNR estimates based on the Laplacian probability density function of the quantized discrete cosine transform (DCT) coefficients extracted from a video bit stream. The algorithm is optimized for the video coding standard H. 264 based on its frame structure and its combination of transform and quantization, which accommodates different bit rates and frame types. Tests on various rates and motion videos give average difference of 1 dB between the estimated PSNR and the real PSNR for moderate bit rates.

Key words discrete cosine transform (DCT) coefficients; Laplacian distribution; H. 264; peak signal-to-noise ratio (PSNR); quantization

像间的误差, 是视频编码质量 (特别是中低码率的视频编码质量) 的主要客观测度之一; 但其计算却需要原始视频参考, 实际监测条件下, 往往原始视频是不可获得的, 这使得 PSNR 不能直接计算, 应用场合受到限制。因此, 提供能够在接收端可以独立计算, 无需源视频做参考的 PSNR 估计方法^[1]是很有实际意义的。无参考方法分为两类: 使用已解码视频的方法和直接使用码流信息的方法。前者对块作用、边缘能量等时空域效应假设出噪声模型, 可用于基于人类视觉系统 (HVS) 的主观视频质量监测。Wang 等通过分析频域块信号的能量提出了块效应测量算法^[2]。Wu 等提出考虑亮度掩闭效应的块边缘失配值 (BIM) 的途径^[3]。但由于计算量偏大, 受到限制。后者利用码流中提取的编码参数来估计图像质量。该方法特别适合广播框架和网络视频服务的质量监测, 应用面广, 计算量适中, 发展较快。

Turaga 等较早地提出了一种无需原始视频的 PSNR 估计方法^[4]。他们利用离散余弦变换 (DCT) 系数的统计特性来估计原始图像的编码质量。尽管该方法对固定量化参数下 I 帧的质量做出了很好的估计, 但对于 P 帧却不适用。

Knee 提出了一种基于 MPEG2 码流信息的 NR 质量测量方法^[5-6]。该方法使用 DCT 系数数据、系数的比特数、量化器参数对视频统计, 从而很好地估计出 PSNR。然而该方法需要进行预备性的实验确定校准参数, 并且对每一个 I 帧都要更新参数来估计其后 P 帧和 B 帧的 PSNR。Asuro 等对 MPEG2 码流下的估计方法做了进一步优化, 提供 HDTV 形式的高码率测试结果, 但对中低码率的视频适用性仍

收稿日期: 2007-01-30

基金项目: 国家自然科学基金资助项目 (60472028);

国家教育部博士点基金项目 (20040003015)

作者简介: 沙亮 (1982-), 男 (汉), 辽宁, 博士研究生。

通讯联系人: 林行刚, 教授, E-mail: xglin@tsinghua.edu.cn。

峰值信噪比 (PSNR) 测量原始图像和已处理图

没有实验数据支持^[7]。

H.264是迅速成熟的视频编码标准,正在成为主流。本文提出在H.264框架下估计PSNR的实现方法。该方法基于DCT系数量化值的统计特性并利用若干码流中参数来估计DCT量化噪声。通过分析视频码流信息的统计特性而不必完全解码即可获取计算压缩视频的PSNR,可适用于各种类型的帧和预测、变换模式,不需要预备性实验校准参数。

1 无参考PSNR估计的基本思路与概率模型

1.1 无参考PSNR估计的基本思路

图1显示了无参考PSNR估计的总体思路。

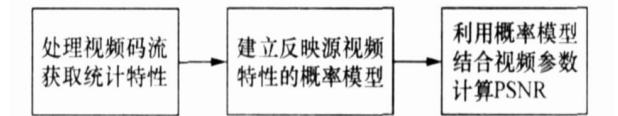


图1 无参考PSNR估计框图

在已编码视频中,源视频的信息已经丢失,但通过对视频码流的统计,依然可以获取源视频的数字特征。进一步地,对源视频数据的分布进行合理的假设,确定适合的概率模型,带入前面得到的数字特征,以形成对源视频的完整估计。在此基础上,即可结合编码视频中可以获取的数据计算PSNR。尽管在实际系统中,由码流结构、变换和量化方式、帧类型等带来实现上的诸多区别,但主线仍可依据图1进行。在图1的三个步骤中,建立概率模型是关键的一步,直接影响了统计参数的选择和PSNR计算的准确性。

1.2 DCT系数分布的概率模型

原始视频DCT交流系数的分布可假设为泛化Gauss分布^[8]。归一化Gauss分布的概率密度函数定义为

$$p(x)=\frac{e^{-\left(\frac{1}{c}\right)^{\frac{1}{V}}}}{2U\Gamma\left(\frac{1}{c}\right)}\exp\left[-\left(\frac{|x|}{U}\right)^{\frac{1}{V}}\right],\quad c>0. \tag{1}$$

其中 $\Gamma(\cdot)$ 为标准Gamma函数。参数 c 和 U 与分布的形状和标准差有关。参数 c 定义了峰度 $V^{[9]}$:

$$V=\frac{\Gamma\left(\frac{5}{c}\right)\Gamma\left(\frac{1}{c}\right)}{\Gamma\left(\frac{3}{c}\right)^2}. \tag{2}$$

文[10]认为DCT变换交流系数分布可大大简化,近似认为是Laplace分布,即泛化Gauss分布的一种。Laplace分布的概率密度函数即为式(1)中当

$c=1$ 且 $U=\frac{e}{2}$ 的情况。此时概率密度函数仅由标准差 e 表征,即

$$p(x)=\frac{1}{2e}\exp\left[-\frac{2|x|}{e}\right]. \tag{3}$$

分析DCT系数量化值的特性的预备性实验^[7]表明:由于峰度不能用于估计分布,因此基于Laplace分布的PSNR估计模型比泛化Gauss模型更合适。又因Laplace分布由标准差 e 唯一确定,所以我们可以量化值的统计中估计未量化的信源DCT系数分布。从而在本估计方案里,DCT系数的分布由Laplace模型确定并作为各种帧类型和预测模式下的基础。

2 H.264 PSNR估计实现方法

2.1 基于DCT变换的视频PSNR估计原理

以H.264为代表的视频编码降质的主要原因是DCT系数的量化。PSNR表征了源视频和解码视频间编码误差的大小。其定义如下:

$$P_{\text{SNR}}=20\lg\left[\frac{S_{p-p}}{X_{\text{MSE}}}\right], \tag{4}$$
$$X_{\text{MSE}}=\frac{1}{N}\sum_{i=1}^N(x_i-x_i')^2.$$

其中 x_i 和 x_i' 是源信号和对应解码信号在对应位置 i 的样本值。 S_{p-p} 是信号峰值幅度,在8b采样的条件下等于255。

实际上,在H.264编码中,下面3个特点决定了以式(4)进行PSNR估计是困难的:

- 1) 在已编码视频中,就只能有DCT变换系数的量化值,而没有源视频的像素值。
- 2) 在编码中质量的下降是且仅是对DCT变换系数的量化过程引起的。
- 3) 在编码过程中,很少对像素直接变换,通常都先进行预测,而变换是对预测的残差形式进行的。

于是,就需要PSNR的DCT域等价形式。据Parseval定理,因为DCT是标准正交变换,像素域MSE等于DCT域的均方量化误差 X_{MSQE} ,所以可以在DCT域测量量化误差,其公式如下:

$$P_{\text{SNR}}=20\lg\left[\frac{S_{p-p}}{X_{\text{MSQE}}}\right], \tag{5}$$
$$X_{\text{MSQE}}=\frac{1}{N}\sum_{i=1}^N(X_i-X_i')^2.$$

其中 X_i 和 X_i' 为源信号和对应解码信号的DCT系数。当信源DCT系数的分布给定,量化误差值就可

由下式得到:

$$\begin{cases} X_{\text{MSQE}} = \frac{1}{N} \sum_{v=0}^{V-1} \sum_{u=0}^{U-1} \sum_{P_Q=0}^{N_{P_Q}-1} \left(n(u,v,P_Q) \times \sum_i \int_{T_i}^U f_i(X) dX \right), \\ f_i(X) = p_{(u,v,P_Q)}(X)(X - q_i)^2, \\ N = \sum_{v=0}^{V-1} \sum_{u=0}^{U-1} \sum_{P_Q=0}^{N_{P_Q}-1} n(u,v,P_Q). \end{cases} \quad (6)$$

其中: $p_{(u,v,P_Q)}(X)$ 表示对每个量化参数 P_Q DCT 系数 X 在位置 (u,v) 的概率密度函数

2.2 H. 264 编码 PSNR 估计步骤与特殊问题

在 H. 264 编码框架下, DCT 变换是以 4×4 整数 DCT 变换为主的, 因此式 (6) 对各位置 (u,v) 的统计都是以 U/V 均为 4 进行。量化参数 P_Q 总数 N_{P_Q} 为 52, 每个 P_Q 对应的量化区间为 $[T_i, U_i]$, 该区间内的量化值表示为 q_i , $n(u,v,P_Q)$ 为各量化参数 P_Q 在位置 (u,v) 的量化值的总样本数。

一般来说, 对各个量化参数 P_Q 都进行估计是必要的, 包括全部 P_Q 的整体分布才体现完整视频的特性。在带有码率控制的编码条件下, 使用相同 P_Q 的部分图像与完整视频的特性有很大区别。也就是说, 标准差 σ 依赖于 P_Q , 即对各量化参数 P_Q 都进行计算才能正确计算 X_{MSQE} 。

需要强调的是, 由于 H. 264 采用多种变换模式和集成变换及量化的特殊性, 使其 DCT 系数的提取需要特别处理。图 2 表示出 H. 264 下的 DCT 系数提取过程。

H. 264 特殊的 DCT 变换和量化结合决定对 H. 264 码流提取 DCT 系数量化值必须在逆量化与 IDCT 之间完成, 然而逆量化又与 IDCT 的一部分结合。因而有必要在此之前单独还原量化值。而针对 I 帧由于 DC 系数进一步采用了 Hadamard 变换, 必须在解码过程中对 DC 和 AC 系数分别提取。提取后的 DCT 系数才能运用在后面的统计过程中。

在 H. 264 编码方案下, 有 3 种类型帧, 分别使用不同的预测模式: I 帧使用自身信息, P 和 B 帧使用参考帧的运动补偿信息。

I 帧由于采用帧内预测的方式即

$$Y[0][0] = \text{dct-depred} + \text{dct-diff}, \quad (7)$$

DC 系数残差 dct-diff 又进行了 4×4 Hadamard 变换, 进一步去除空间冗余, 因此在量化时也大多成为残差形式, 分布集中, 因而可假设为 Laplace 分布。

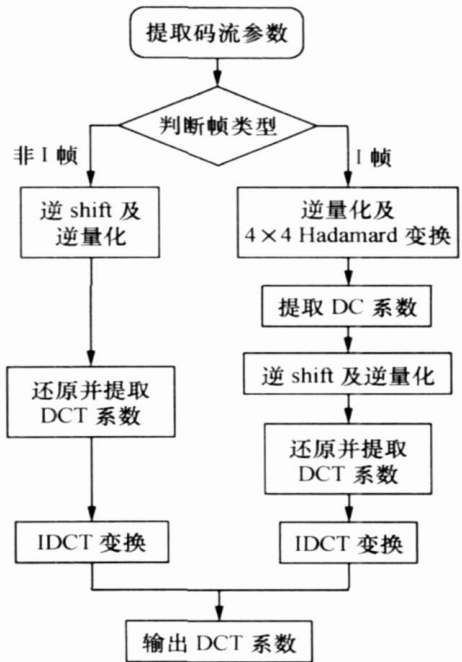


图 2 DCT 系数提取流程图

P 帧和 B 帧作为预测编码帧, 解码图像可以表示为 $\hat{S}_{k+1} = w(S_k - \hat{S}_k)$, S_k 为第 k 帧输入图像, \hat{S}_k 为对应预测图像, 二者表示 DCT 量化和其逆过程。因此编码误差 e_k 可表示为

$$\begin{aligned} e_k &= S_k - [\hat{S}_{k+1} + w(S_k - \hat{S}_k)] = \\ &= (S_k - \hat{S}_k) - w(S_k - \hat{S}_k). \end{aligned} \quad (8)$$

式 (8) 说明对 P 帧和 B 帧, 计算预测误差信号的 X_{MSQE} 等于解码信号的 X_{MSQE} 。由于在 P 帧和 B 帧中, 存在着非编码块, 不传输 DCT 系数, 因而在码流中就没有信息可用于估算 X_{MSQE} 。因而 X_{MSQE} 只能使用编码块计算:

$$X_{\text{MSQE}} = aX_{\text{MSQE-Intra}} + bX_{\text{MSQE-NonIntra}}, \quad (9)$$

a 和 b 分别表示帧内编码块和非帧内编码块在全部编码块中的比例。

H. 264 标准中并无关于量化区间的强制规定, 在码流中也不提供相应的信息, 从而量化区间也要作为 PSNR 估计的一部分。如果假设量化值 q_i 是量化区间 $[T_i, U_i]$ 的中点, Q_{step} 为量化步长, 即

$$\begin{cases} T_i = q_i - 0.5Q_{\text{step}}, \\ U_i = q_i + 0.5Q_{\text{step}}. \end{cases} \quad (10)$$

基于上述分析可通过下列步骤估计 PSNR

1) 提取阶段

- 从码流中提取 CT 系数和量化参数 P_Q
- 以 4×4 块为单位获得辅助量 $\sum X_i$ 与 $\sum X_i^2$

- 将辅助量填入各 16×16 宏块的 P_Q
- 2) 建模阶段。
- 对各 P_Q 和 4×4 块计算该帧 ϵ 值。
- 应用式 (7)求得源视频 DCT 系数 $p_{(u,v,P_Q)}(X)$ 。
- 3) 计算阶段。
- 应用式 (6)求该帧 X_{MSQE}
- 应用式 (5)求该帧 PSNR
- 结合各帧 X_{MSQE} 并应用式 (9) (5)得到整个码流平均 X_{MSQE} 和 PSNR

3 实验结果

本文选择 H. 264 JM10.0产生编码率分别为 2 Mb/s、1.2 Mb/s和 400 kb/s含相对集中和中等运动的“Stephan”和含复杂运动的“Mobile”CIF格式测试文件。表 1为测试的统计结果。

表 1 “Stephan”与“Mobile”测试结果					
测试文件	编码率 Mb s ⁻¹	实际平均 PSNR dB	估计平均 PSNR dB	平均值 误差 dB	误差 标准差
Stephan	0.4	30.90	31.31	0.41	0.654
	1.2	35.66	35.67	0.01	0.272
	2.0	38.70	39.32	0.62	0.052
Mobile	0.4	26.78	28.19	1.41	0.791
	1.2	31.72	32.38	0.66	0.677
	2.0	34.14	34.95	0.81	0.082

分析表 1可见,以 Mobile测试文件为例,随着编码率由 2 Mb/s、1.2 Mb/s减小到 400 kb/s,估计的误差从 0.81 dB、0.66 dB到最大的 1.41 dB,前两者码率较高,误差在 1 dB以内,后者码率较低,则误差高于 1 dB;从各帧统计的误差标准差来看,则由 0.082、0.677扩大到 400 kb/s时的 0.791,估计的误差波动性也在增强。我们注意到,在视频码率降低时,其 PSNR也在相应地降低,如 Mobile在码率从 2 Mb/s到 400 kb/s时其实际 PSNR降低了 7.36 dB,这意味着在编码量化过程中大量信息损失,而通过已编码视频去估计降质前的信息模型就会出现较大的误差,其波动性也会扩大。另一方面,对比运动集中的 Stephan和运动复杂的 Mobile测试,可见在运动集中的情况码率误差都较小,而两者在 400 kb/s估计的平均值误差则相差 1 dB左右,这除了量化损失信息程度不同外,高速运动使得统计估计源视频的概率模型不够准确也是一大原因。

上述分析得出两点结论: 1) 高码率的视频比低码率的视频估计误差较小,误差波动也较小; 2) 中等和相对集中运动的视频比复杂运动的视频估计误差较小。

4 结 论

本文给出基于 H. 264码流提取信息的数字视频 PSNR估计方法。该算法基于 DCT系数量化值的统计特性和假设 DCT系数的分布符合 Laplace分布。

通过不同序列在不同码率条件下的实验验证了估计的性能。实验结果显示,除了视频内容运动复杂的“Mobile”400 kb/s的甚低视频质量下(事实上此时的视频质量主观上已无法接受,这里仅用于测试),估计 PSNR与实际 PSNR之间平均误差在 1 dB以内。本文的方法适用于各种码率和帧类型。

尽管本文提出的方法在监测系统中测试性能可观,但在高压缩低速率的网络需求上,本文提供的方法在精度上还有待提高。

参考文献 (References)

[1] Rec. J 143. User requirements for Objective Perceptual Video Quality Measurements in Digital Cable Television [S]. ITU-T, May 2000.

[2] Wang Z, Bovil A. Blind measurement of blocking artifacts in images [J]. *Int Conf Image Process*, 2000, **III**: 981–984.

[3] Wu H, Yuen M. A generalized block-edge impairment metric for video coding [J]. *IEEE Signal Process Lett*, 1997, **4**(1): 317–320.

[4] Turaga D, Chen Y, Caviedes J. No reference PSNR estimation for compressed pictures [J]. *Signal Process Image Commun*, 2004, **19**: 173–184.

[5] Knee M. A single-ended picture quality measure for MPEG-2 [C]// *Proc Int Broad. Convention (IBC 2000)*. Amsterdam, 2000: 95–100.

[6] Knee M, Diggins M, World Intellectual Property Bureau. Improvements in Data Compression [P]. *Int Patent Appl WO 00/22 834*, 2000.

[7] Atsuro I, Masaaki K, Naohiro H, et al. A method of estimating coding PSNR using quantized DCT coefficients [J]. *IEEE Trans, Circuit and System for Video Technology*, 2006, **2**: 251–259.

[8] Pham T, Figueiredo R. Maximum likelihood estimation of a class of non-Gaussian densities with application to lp deconvolutions [J]. *IEEE Trans Acoust, Speech, Signal Process*, 1999, **37**(1): 73–82.

[9] Lam E, Goodman J. A mathematical analysis of the DCT coefficient distributions for images [J]. *IEEE Trans Image Process*, 2000, **4**: 1661–1666.

[10] Edue T, Grisel R, Cherifim H, et al. On the distribution of the DCT coefficients [C]//*Proc IEEE ICASSP*. Adelaide, Australia, 2004: V365–V368.