

基于人眼立体视觉特性的多视视频码率控制算法

卢山 孙军 王嘉

(上海交通大学电子工程系图像通信与信息处理研究所, 上海 200240)

(上海交通大学上海市数字媒体处理与传输重点实验室, 上海 200240)

摘要 提出了一种基于人眼立体视觉特性的多视视频码率控制算法。引入了立体视觉质量的客观评价标准,并基于此推导出一种帧级目标码率分配方法,同时根据相邻宏块量化参数来调整当前宏块量化参数,以消除立体视频的块效应。仿真结果显示,采用本文算法的多视视频编码器,在取得更稳定的比特率的同时,有效地提高了立体视觉质量。

关键词 立体视觉特性 码率控制 多视视频

中图法分类号: TP301.6 文献标识码: A 文章编号: 1006-8961(2009)11-2396-05

Multiview Video Rate Control Scheme Based on Human Stereo Visual Feature

LU Shan, SUN Jun, WANG Jia

(Institute of Image Communication and Information Processing, Department of Electronic Engineering,
Shanghai Jiaotong University, Shanghai 200240)

(Shanghai Key Laboratory of Digital Media Processing and Transmissions, Shanghai Jiaotong University, Shanghai 200240)

Abstract This paper proposed a multiview video rate control scheme based on human stereo visual feature. First, the new quantitative measure for stereo video quality is introduced. Then, the bit allocation scheme is derived on the basis of this measure. Finally, the macroblock quantization parameters are modified according to the quantization parameter of the neighboring macroblock on the purpose of eliminating block effect in stereo video. The experiment results demonstrate that the proposed scheme can improve the stereo video quality and obtain quite stable rate.

Keywords human stereo visual feature, rate control, multiview video

1 引言

多视视频技术被认为是数字视频技术的一个重要发展趋势。在我国即将举行的2010年世博会上,将采用立体电视进行世博会转播。然而,多视视频有多路视频信号,视频码率将成倍增加,增大了网络带宽的压力,同时多视视频编码中,既进行运动预测,又进行视差预测,因此具有更多的编码图像类

型,使原来针对单视视频编码的码率控制算法,如MPEG2的TM5算法^[1],H.263的TMN8算法^[2]等,都不能直接应用于多视视频编码上。

目前,国内外学者针对多视视频码率控制已有一定研究。Natio等人提出统一缓冲区模型^[3],用TM5算法进行码率控制,并用人眼视觉特性给主视辅视分配比特;Lim等人提出基于二次模型的码率控制算法^[4],根据视差预测和运动预测的结构关系,将图像分成多种类型分别建模,计算目标比特和

基金项目:国家高技术研究发展计划(863)项目(2008AA01A318,2008AA01A319)

收稿日期:2009-06-19;改回日期:2009-08-31

第一作者简介:卢山(1984~),男,上海交通大学电子工程系图像通信与信息处理研究所信号与信息处理专业硕士研究生。主要研究方向为立体视频编解码。E-mail:lu_shan@jtu.edu.cn

量化参数;Zhu 等人提出一种基于人眼视觉特性的码率控制算法^[5],根据帧率、码率和图像类型决定图像分辨率。这些算法对人眼视觉特性的利用,都基于 Julesz 的一个基本假设^[6]:立体视频的人眼视觉质量取决于立体视频中质量较高的视频。然而,该假设只在模糊降质时成立^[8]。

本文针对立体视觉的这一问题进行探讨,并对 TM5 算法进行改进,使之适用于多视视频编码。

2 人眼视觉特性

2.1 一种立体视频质量的客观标准

基于人眼视觉特性,即低质量图像对整体质量影响较小,Ozbek 等人对主、辅视峰值信噪比 (PSNR) 赋以不同的权重,提出了一种立体视频质量的客观标准:^[7]

$$Q_n = (1 - \alpha) \cdot \text{PSNR}_n^{\text{Main}} + \alpha \cdot \text{PSNR}_n^{\text{Left}} \quad (1)$$

其中, $\text{PSNR}_n^{\text{Main}}$ 和 $\text{PSNR}_n^{\text{Left}}$ 分别表示主、辅视 PSNR 的第 n 视, α 是给主、辅视 PSNR 权重,取 α 为 1/3。对一个序列的质量,可以用 Q_n 的平均值来表示, N 是主视或辅视的帧数:

$$Q = \frac{1}{N} \sum_{n=0}^{N-1} Q_n \quad (2)$$

2.2 双目视觉合成中单目输入的不同权重

针对立体视频的人眼视觉质量取决于立体视频中质量较高视频的假设,Meegan 等人对不同情况降质对整体质量的影响做了大量研究^[8],发现有两种类型降质:第一种是模糊,即通过空域低通滤波;第二种是块效应,主要由视频压缩算法产生。

根据人眼视觉特性,双视视觉会给信息量更大的视以更大权重。当发生模糊降质时,模糊的视的信息量会减少;当使用视频压缩算法时,在滤除空域高频信息的同时,会在宏块边缘增加空域高频信息,这是人眼视觉系统中一种重要信息,换言之,块效应图像会比原图像产生更多信息。因此,当辅视发生模糊降质时,不会影响立体视觉质量,而当发生块效应时,就会严重影响立体视觉质量。

在 MPEG-2 标准的压缩过程中,离散余弦变换 (DCT) 变换和低量化阶量化会减少空域高频信息,而高量化阶量化会产生块效应。量化阶大小通过码率控制调节,在低码率情况,高量化阶量化不可避免,但码率控制算法可以调节相邻宏块量化阶来避免块效应。

3 本文算法

3.1 SGOP 结构

单视视频码率控制算法和 GOP (group of pictures) 结构有密切关系,定义 SGOP 为主、辅视 GOP 组合作为多视视频编码的基本结构,其结构采用如图 1 所示的参考关系。其中,主视帧只做运动估计,辅视帧做视差估计和运动估计,共分为 6 种帧类型:主视有 I 帧、P 帧、B 帧,辅视有 Is 帧、Ps 帧、Bs 帧。图 1 中,箭头方向表示当前图像对相应图像参考。

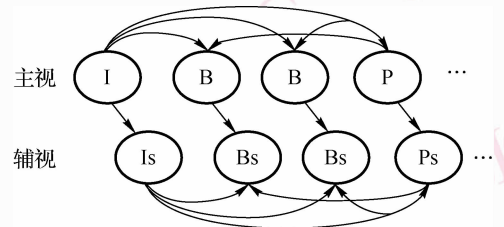


图 1 SGOP 结构
Fig. 1 Structure of SGOP

3.2 比特分配

在 TM5 的算法中,仅考虑 3 种帧类型,而本文多视视频的编码结构有 6 种帧类型。如果直接将 Is 帧、Ps 帧和 Bs 帧视作 I 帧、P 帧、B 帧,会造成比特分配的不合理。同时,常数 K_I, K_P, K_B 也会影响比特分配,而在质量评价标准中主视和辅视有不同的权重,所以这几个常数,包括辅视帧类型对应的 K_{Is}, K_{Ps}, K_{Bs} 都需要重新确定。

下式给出了对 SGOP 的比特分配方法:

$$T_I = \frac{R \times \frac{X_I}{K_I}}{\frac{X_I}{K_I} + N_P \cdot \frac{X_P}{K_P} + N_B \cdot \frac{X_B}{K_B} + N_{Is} \cdot \frac{X_{Is}}{K_{Is}} + N_{Ps} \cdot \frac{X_{Ps}}{K_{Ps}} + N_{Bs} \cdot \frac{X_{Bs}}{K_{Bs}}}$$

$$T_P = \frac{R \times \frac{X_P}{K_P}}{\left(N_P \cdot \frac{X_P}{K_P} + N_B \cdot \frac{X_B}{K_B} + N_{Ps} \cdot \frac{X_{Ps}}{K_{Ps}} + N_{Bs} \cdot \frac{X_{Bs}}{K_{Bs}} \right)}$$

$$T_B = \frac{R \times \frac{X_B}{K_B}}{\left(N_P \cdot \frac{X_P}{K_P} + N_B \cdot \frac{X_B}{K_B} + N_{Ps} \cdot \frac{X_{Ps}}{K_{Ps}} + N_{Bs} \cdot \frac{X_{Bs}}{K_{Bs}} \right)}$$

$$\begin{aligned}
 T_{I_s} &= \frac{R \times \frac{X_{I_s}}{K_{I_s}}}{\left(N_P \cdot \frac{X_P}{K_P} + N_B \cdot \frac{X_B}{K_B} + N_{I_s} \cdot \frac{X_{I_s}}{K_{I_s}} + N_{P_s} \cdot \frac{X_{P_s}}{K_{P_s}} + N_{B_s} \cdot \frac{X_{B_s}}{K_{B_s}} \right)} \\
 T_{P_s} &= \frac{R \times \frac{X_{P_s}}{K_{P_s}}}{\left(N_P \cdot \frac{X_P}{K_P} + N_B \cdot \frac{X_B}{K_B} + N_{P_s} \cdot \frac{X_{P_s}}{K_{P_s}} + N_{B_s} \cdot \frac{X_{B_s}}{K_{B_s}} \right)} \\
 T_{B_s} &= \frac{R \times \frac{X_{B_s}}{K_{B_s}}}{\left(N_P \cdot \frac{X_P}{K_P} + N_B \cdot \frac{X_B}{K_B} + N_{P_s} \cdot \frac{X_{P_s}}{K_{P_s}} + N_{B_s} \cdot \frac{X_{B_s}}{K_{B_s}} \right)}
 \end{aligned} \tag{3}$$

其中, $T_I, T_P, T_B, T_{I_s}, T_{P_s}, T_{B_s}$ 分别是各帧目标比特数, R 是 SGOP 的剩余比特数, $X_I, X_P, X_B, X_{I_s}, X_{P_s}, X_{B_s}$ 分别是各帧复杂度。这些值的更新方法和 TM5 算法相似。但 $K_I, K_P, K_B, K_{I_s}, K_{P_s}, K_{B_s}$ 需要重新确定, 其推导的过程如下。

比特分配问题可以用下式表示:

$$\begin{aligned}
 \text{Maximize} \quad & Q = \frac{1}{S} \sum_{s=0}^{S-1} Q_s \\
 \text{subject to} \quad & \sum_{s=0}^{S-1} (R_{\text{Main},s} + R_{\text{Sub},s}) < R_{\text{Total}}
 \end{aligned} \tag{4}$$

其中, $R_{\text{Main},s}$ 和 $R_{\text{Sub},s}$ 分别是主、辅视第 s 视比特数, R_{Total} 是当前序列给定比特数, Q 是各个帧质量 Q_s 的均值。

根据文献[7]的质量评价标准, 比特分配问题最终归结为率失真代价方程的最小化问题。

$$\begin{aligned}
 C_s &= \text{MSE}_s + \lambda \cdot (R_{\text{Main},s} + R_{\text{Sub},s}) \\
 C &= \sum_{s=0}^{S-1} C_s
 \end{aligned} \tag{5}$$

问题的最优解在下式(6)成立时取得:

$$w \frac{\partial \text{MSE}_k}{\partial q_k} + \lambda \frac{\partial R_k}{\partial q_k} = 0 \tag{6}$$

其中, k 表示帧类型, w 表示 PSNR 计算时的权重, 按下式(7)计算。

$$w = \begin{cases} 1 - \alpha & \text{主视} \\ \alpha & \text{辅视} \end{cases} \tag{7}$$

在一定的 q_k 值范围内, MSE_k 与 q_k 间的关系是近似分段线性的, 如图 2 所示。

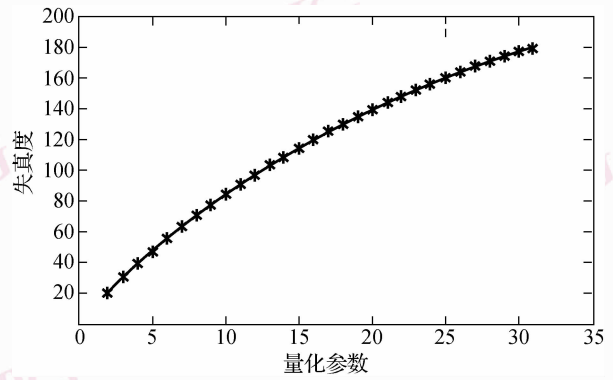


图 2 失真度和量化参数模型
Fig. 2 Distortion-quantization parameter model

$$\frac{\partial \text{MSE}}{\partial q} = \begin{cases} \frac{\text{MSE}_{10} - \text{MSE}_2}{q_{10} - q_2} & q \in [2, 10] \\ \frac{\text{MSE}_{20} - \text{MSE}_{11}}{q_{20} - q_{11}} & q \in [11, 20] \\ \frac{\text{MSE}_{31} - \text{MSE}_{21}}{q_{31} - q_{21}} & q \in [21, 31] \end{cases} \tag{8}$$

对于每帧的 $\frac{\partial \text{MSE}_k}{\partial q_k}$, 假定宏块的量化阶数是稳定的, 于是可以用前一帧的 $\frac{\partial \text{MSE}_k}{\partial q_k}$ 近似得到。

根据 TM5 算法中的一阶率失真模型, R_k, q_k 有以下关系。

$$\frac{\partial R_k}{\partial q_k} = -X_k \cdot q_k^{-2} \tag{9}$$

$$\text{式中, } X_k = \frac{1}{n} \cdot \sum_{i=1}^n q_i^k R_i^k \tag{10}$$

联立式(6)、式(8)、式(9), 可以得到

$$q_k = \sqrt{\frac{\lambda X_k}{w_k \frac{\partial \text{MSE}_k}{\partial q_k}}} \tag{11}$$

不同类型帧的量化参数满足

$$\frac{q_I}{K_I} = \frac{q_P}{K_P} = \frac{q_B}{K_B} = \frac{q_{I_s}}{K_I K_{I_s}} = \frac{q_{P_s}}{K_P K_{P_s}} = \frac{q_{B_s}}{K_B K_{B_s}} = c \tag{12}$$

令 $K_I = 1$, 便可利用式(11)、式(12), 得到 $K_P, K_B, K_{I_s}, K_{P_s}, K_{B_s}$ 的值, 并在每个 SGOP 更新。

3.3 宏块级量化阶调整

根据文献[8], 辅视模糊不会造成立体视觉降质, 而块效应会引起明显降质。根据文献[7]提出的质量评价标准, 辅视会因为分配较少的比特增加

发生块效应的机率。编码过程中,造成块效应的最重要因素是相邻宏块量化阶数差异过大。所以,在得到各宏块的量化参数后,将其调整到和前一宏块量化阶数不超过 2。调整方法如下:

$$q_k = \begin{cases} q_{k-1} - 2 & q_k < q_{k-1} - 2 \\ q_{k-1} + 2 & q_k > q_{k-1} + 2 \\ q_k & \text{其他} \end{cases} \quad (13)$$

4 实验结果

在基于 MPEG-2 的 MVP 的编码器上,对本文算法的性能进行了测试。采用 ballroom 和 toy 序列,分辨率为 640 × 480,帧率为 25 fps,GOP 为 12 帧,共编码 125 帧,比较对象为文献[3]算法。为了合理评估本文算法,采用多个评价标准:码率偏离程度、解码图像客观质量和主观质量。第一个标准用来评价码率稳定程度,用下式计算,

$$bit_{deviation} = \left| \frac{R_{actual} - R_{target}}{R_{target}} \right| \times 100\% \quad (14)$$

其中, R_{actual} , R_{target} 表示实际码率和目标码率。第二个标准是文献[7]提出的标准。

4.1 码率控制

表 1 给出了文献[3]算法和本文算法在码率控制方面的性能差异。可以看出,对于不同的测试序列,本文算法在实际码率偏离程度上都优于文献[3],可以将偏离程度控制在 0.3% 左右。

表 1 不同方法下码率比较

Tab. 1 Bit rate comparison of two methods for ballroom

序列	目标码率 (Kbps)	文献[3]算法		本文算法	
		实际码率	偏离程度 (%)	实际码率	偏离程度 (%)
ballroom	512	507.75	0.83	510.19	0.35
	256	254.10	0.74	255.13	0.34
toy	512	508.00	0.78	510.88	0.22
	256	254.44	0.61	255.38	0.24

4.2 客观质量

表 2 给出了文献[3]算法和本文算法在质量控制方面的性能差异。可以看出,对不同测试序列,本文算法都优于文献[3],可以将 Q 值提高 0.4 dB 到 2.04 dB。在某些情况,给辅视分配较少比特数却可以获得比文献[3]辅视更好的 PSNR。

表 2 不同方法下 Q 值比较

Tab. 2 Q comparison of different methods for ballroom

	目标码率 (ballroom)		目标码率 (toy)		
	512	256	512	256	
文献[3]方法	主视 (PSNR)	37.27	34.95	35.83	31.29
	辅视 (PSNR)	36.80	34.43	35.31	31.06
	Q 值	37.11	34.77	35.65	31.21
本文方法	主视 (PSNR)	38.67	35.71	36.81	33.28
	辅视 (PSNR)	36.48	34.09	36.18	33.19
	Q 值	37.94	35.17	36.60	33.25

同时,我们研究了 α 等于 1/2, 2/5, 1/3 时主、辅视解码图像 PSNR 的关系。从图 3 可以看出,对不同测试序列,在不同码率情况下,尽管随着 α 的减小,辅视分配比特数随之减小,但由于参考的主视 PSNR 的提高,辅视质量 PSNR 稳定。

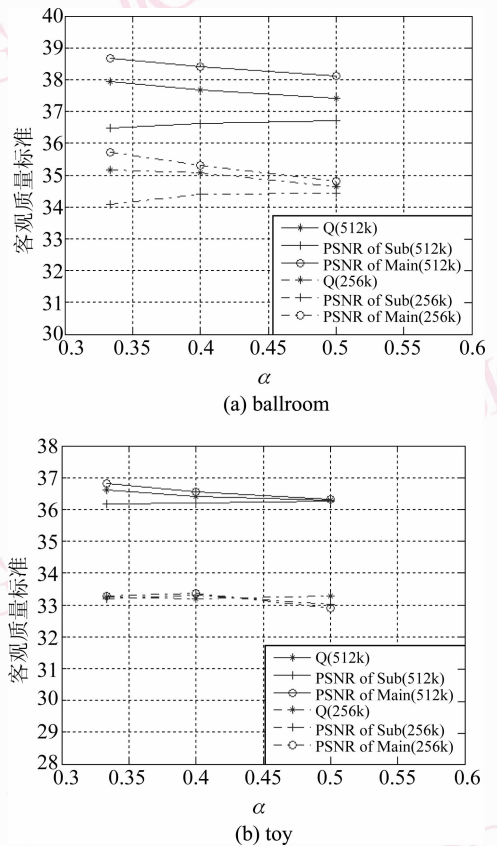
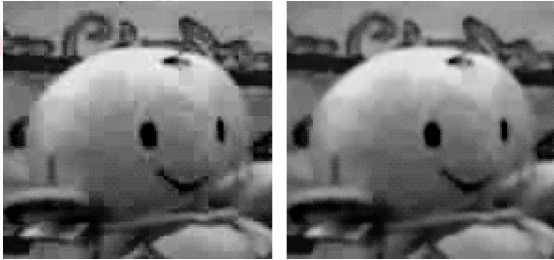


图 3 不同码率下主、辅视 PSNR 及 Q 值的关系
Fig. 3 Relationship between Q values and PSNR at difference bit-rate

4.3 主观质量

从视频序列中随机抽取一帧进行比较,可以看出,与文献[3]相比,本文算法明显减小了块效应。



(a) 文献[3]算法

(b) 本文算法

图 4 文献[3]算法与本文算法主观质量比较(toy)

Fig. 4 Subjective quality comparison of two methods for toy

5 结 论

通过对立体视觉特性的分析,针对在辅视的降质没有块效应降质的前提下,立体视觉质量由主视决定的特点,引入了一种对 PSNR 加权的客观质量评价标准,并推导对比特分配时各帧关系。同时,考虑到自适应量化可能引入块效应,提出了宏块级量化的调整方法,能有效防止辅视产生块效应。实验结果表明,本文算法在有效控制码率和提高立体视觉质量上有明显效果。

参考文献 (References)

- 1 ISO/IEC JTC/SC29/WG11, MPEG Test Model 5 [S].
- 2 ITU-T/SG15, Video Codec Test Model, TMN8 [S].
- 3 Natio S, Matsumoto H. Advanced rate control technologies for 3D-HDTV digital coding based on MPEG-2 multi-view profile [A]. In: Proceedings of IEEE International Conference on Image Processing [C], Kobe, Japan, 1999, 1: 281-285.
- 4 Lim J, Kim J, Ngan KN, Sohn K. Advanced rate control technologies for 3D-HDTV [J]. IEEE Transactions on Consumer Electronics, 2003, 49(4): 1498-1507.
- 5 Zhu Z J, Liang F, Jiang G Y, et al. Bit allocation and rate control for stereoscopic video coding [A]. In: Proceedings of 3DTV Conference [C], Kos, Greece, 2007: 1-4.
- 6 Julesz B. Foundations of Cyclopean Perception [M]. Chicago, IL: University of Chicago Press, 1971.
- 7 Ozbek N, Tekalp A M. Unequal inter-view rate allocation using scalable stereo video coding and objective stereo video quality measure [A]. In: IEEE International Conference on Multimedia and Expo [C], Hannover, Germany, 2008: 1113-1116.
- 8 Meegan D V, Stelmach L B, Tam W J. Unequal weighting of monocular inputs in binocular combination: Implications for the Compression of Stereoscopic Imagery [J]. Journal of Experimental Psychology, 2001, 7(2): 143-153.