

可伸缩视频编码的自适应 QP 级联算法

万卫星, 李厚强

(中国科学技术大学电子工程与信息科学系, 合肥 230027)

摘要: 针对可伸缩视频编码标准中采用的分级 B 帧预测结构, 提出一种图像级的内容自适应量化参数(QP)级联算法。在该算法中, 任意一帧图像的最终 QP 值由分级预测的结构及该图像内所有宏块的运动预测模式决定。实验结果证明, 与目前 SVC 标准采用的技术相比, 该算法最多可以提升 0.363 dB 的编码性能。

关键词: 可伸缩视频编码; 分级预测; 内容自适应量化参数

Adaptive QP Cascading Algorithm for Scalable Video Coding

WAN Wei-xing, LI Hou-qiang

(Dept. of Electronic Engineering and Information Science, University of Science and Technology of China, Hefei 230027)

【Abstract】 Aiming at hierarchical B prediction structure utilized in Scalable Video Coding(SVC) standard, this paper proposes a picture-level Content-Adaptive Quantization Parameter Cascading(CAC) algorithm. The Quantization Parameter(QP) value for each picture is decided based on the hierarchical prediction structure and prediction modes of the macroblocks in the picture. Experimental result demonstrates that compared with the technique used in SVC standard, the algorithm increases rate distortion performance by 0.363 dB.

【Key words】 Scalable Video Coding(SVC); hierarchical prediction; content-adaptive Quantization Parameter(QP)

1 概述

随着计算机网络的不断发展, 视频编码技术的研究方向不再局限于视频数据的压缩特性。网络条件的异构性、终端设备的多样性以及用户需求的复杂性使不同的应用场景对视频的帧率、分辨率和质量有不同的要求, 这些都对传统的视频编码标准提出了挑战。为了解决这一系列问题, 联合视频组(JVT)将可伸缩视频编码(Scalable Video Coding, SVC)^[1]纳入 H.264/AVC 标准的扩展中, 并正式将其作为国际标准。SVC 一方面追求提高视频的压缩特性, 另一方面又注重提供视频的可伸缩性服务。为了实现码流的可伸缩性, SVC 编码器产生一段包含一个基本层和多个增强层的码流, 其中基本层与 H.264/AVC 标准兼容, 而增强层可以在 3 个方向分别对基本层进行增强: 时间分辨率(帧率), 空间分辨率和峰值信噪比 PSNR^[2-3]。

时间分辨率的可伸缩性可以通过分级 B 帧预测(hierarchical B prediction)的编码框架来实现^[4]。在这种编码框架下, 图像的重要性与其所处的时间级相关, 因此, 通过对不同时间级的图像设置不同的量化参数(Quantization Parameter, QP)可以达到较高的编码效率。然而, 在当前 SVC 标准中, 并没有考虑具体视频序列和图像内容的特性, 不同时间级的图像之间 QP 差值是固定的, 因此, 还有很大的改进空间。本文提出一种图像级内容自适应 QP 级联(Content-Adaptive QP Cascading, CAC)算法, 不仅利用了一个图像组(Group Of Pictures, GOP)中高时间级图像参考低时间级图像的预测结构, 还考虑了图像由于做运动预测而引起的能量变化。其编码性能比 SVC 现有的方法有一定的提高, 并且算法复杂度增加较少。

2 分级 B 帧预测编码框架

SVC 在时间上的可伸缩性主要通过分级 B 帧预测编码框

架来实现。分级 B 帧预测结构可以实现同一个视频在不同时间分辨率或帧率下的表示。图 1 给出了一个由 8 帧图像组成 GOP 的分级预测结构, 整个视频序列的第 1 帧是瞬时解码刷新(Instantaneous Decoding Refresh, IDR)帧。关键帧是指每一个 GOP 的终止帧, 其之前所有被编码的帧在显示顺序上都先于该帧, 如图 1 中的第 0、第 8 和第 16 帧, 这些关键帧等距离分布在整个序列里。当前关键帧和所有处于当前关键帧与前一个关键帧之间的帧共同构成一个 GOP。在分级 B 帧预测结构中, GOP 中的图像可以分为两部分: (1)关键帧, 采用帧内编码(I 帧)或者采用前一个关键帧作为预测的帧间编码(P 帧); (2)非关键帧, 采用低时间级或同时间级上距该帧最近的 2 帧作为参考进行双向预测编码(B 帧)。

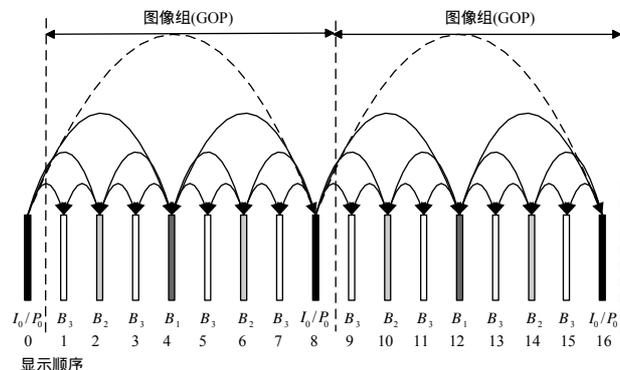


图 1 具有 4 个时间级的分级预测结构

基金项目: 国家自然科学基金资助项目(60672161); 国家“863”计划基金资助项目(2006AA01Z317)

作者简介: 万卫星(1984 -), 男, 硕士研究生, 主研方向: 视频编码, 视频传输; 李厚强, 教授、博士生导师

收稿日期: 2009-08-10 **E-mail:** libra@mail.ustc.edu.cn

需要注意的是, 分级 B 帧预测结构并不要求每个 GOP 中含有 2 的指数次幂帧图像。为了实现视频在时间上的可伸缩性, B 帧只能采用较低时间级或同时间级上的图像作为参考。从图 1 中可以看出, 图像的重要性与它所处的时间级相关。时间级越低的帧被其他帧直接或间接参考的次数越多, 其编码质量对整个序列的影响也越大, 所以, 应该为它提供一个相对较好的质量; 而那些位于最高时间级上的帧由于未被其他任何帧参考, 因此允许有相对较差的质量。由以上分析可知, 通过为不同时间级的图像分配不同的编码质量, 可以进一步提高视频的编码效率。而视频质量是由量化参数 QP 决定的, 因此, 如何合理地分配每一帧的 QP 直接影响视频的编码效率。

3 内容自适应 QP 级联算法

在 CAC 算法中, 假设 GOP 内共有 T 个时间级, 位于最高时间级 $T-1$ 上的帧的 QP 由系统根据用户目标码率设置, 记为 QP_{T-1} ; 为时间级低于 $T-1$ 的帧定义一个伸缩因子, 用来决定该图像应采用的 QP 值与 QP_{T-1} 的差值。该伸缩因子主要由分级预测结构以及运动预测带来的能量变化所决定。

3.1 伸缩因子

在分级 B 帧预测结构中, 对于单向预测, 运动预测可以表示为在预测方向上使用一个高通滤波器 $\{1, -1\}$; 对于双向预测, 同样可以表示为使用一个高通滤波器 $\{-1/2, 1, -1/2\}$ ^[5]。由于非归一化滤波器的使用使图像的能量在做完运动预测之后发生了变化, 因此提出对每帧图像使用一个伸缩因子来控制其 QP, 以平衡不同时间级上图像的能量。这些伸缩因子主要由运动预测中使用的滤波器的能量以及图像内所有宏块 (macroblock) 的运动预测模式共同决定。在每帧图像做完运动预测之后, 与之对应的平均能量变化由下式计算:

$$E_{t,m} = \frac{1}{N} \sum_{i=1}^N \alpha_i \quad (1)$$

其中, $t=0,1,\dots,T-1$ 为时间级; m 表示在同一时间级 t 上的图像索引; N 是一帧内宏块的总数目; α_i 是加权因子, 表示第 i 个宏块做运动预测时带来的能量变化。 α_i 与宏块 i 的预测模式有关。如果该宏块采用帧间单向预测 (表示为高通滤波器 $\{1, -1\}$), α_i 取值为 $\sqrt{2}$; 如果该宏块采用帧间双向预测 (表示为高通滤波器 $\{-1/2, 1, -1/2\}$), α_i 取值为 $\sqrt{3}/2$; 而对于采用帧内预测的宏块, 对应的 α_i 为 1。

GOP 内每帧图像的平均能量变化 $E_{t,m}$ 确定之后, 位于时间级 t 上的第 m 帧图像的伸缩因子可用下式计算:

$$SF_{t,m} = \overline{SF_{t+1}} \div E_{t,m} \quad (2)$$

其中, $t=0,1,\dots,T-2$ 为时间级; m 表示在同一时间级 t 上的图像索引; $\overline{SF_{t+1}}$ 是 GOP 内第 $t+1$ 时间级上所有图像的 $SF_{t+1,j}$ (j 为索引值) 的平均值; 最高时间级图像对应的 $SF_{T-1,j}$ 初始化为 1.0。

3.2 分级 QP 设置

得到伸缩因子 $SF_{t,m}$ 之后, 其对应的 QP 值的计算如下:

$$QP_{t,m} = \text{Round}(QP_{T-1} + 61bSF_{t,m}) \quad (3)$$

其中, $t=0,1,\dots,T-2$ 为时间级; m 表示在同一时间级 t 上的图像索引; QP_{T-1} 即系统输入的最高时间级图像对应的 QP 值; $QP_{t,m}$ 的取值范围是 0~51。需要注意的是, 所有位于最高时间级的图像具有相同的 QP 值。

因为宏块的运动预测模式在没有做完运动预测和模式选择之前是未知的, 所以本文算法先对目标序列分 GOP 进行预

处理。在预处理过程中, 根据当前 SVC 标准采用的 QP 设置技术对 GOP 内每帧进行运动预测和模式选择, 由式(1)~式(3)计算每帧对应的最佳 QP。然后利用这些最佳 QP 重新对 GOP 内图像进行正式的编码。

4 实验结果与分析

模拟实验在最新的 SVC 参考软件 JSVM_9_15(JSVM)中实现了本文的 CAC 算法。其中, GOP 长度设为 16, 系统为最高时间级图像设置的 QP 值为 28, 32, 36 和 40 这 4 种情况。实验的测试序列如下: container, foreman, hall_monitor, irene, mobile, news, paris, silent, tempete。所有序列都是 30 Hz 的 QCIF, 每个序列只有第 1 帧采用帧内编码 (即 I 帧)。本文对以下 2 种技术做了比较: (1)CAC 算法; (2)SVC 参考软件 JSVM_9_15: 当前 SVC 标准采用的技术, 层与层之间的 QP 差值是固定不变的, 没有考虑视频序列和图像组之间的差异性。并采用 Bjontegaard 方法^[6]计算峰值信噪比增益和码率减少率。由表 1 可以看出, 与当前 SVC 标准技术相比, CAC 平均有 0.2 dB 的编码性能提高或 3% 的码率节省, 对于效果最好的序列有将近 0.4 dB 的编码性能提高。

表 1 CAC 相对 JSVM 的性能改进

| 测试序列 | 峰值信噪比增益/dB | 码率减少率/(%) |
|--------------|------------|-----------|
| container | 0.110 | 2.02 |
| foreman | 0.202 | 3.55 |
| hall_monitor | 0.168 | 2.85 |
| irene | 0.087 | 1.57 |
| mobile | 0.172 | 3.45 |
| news | 0.256 | 3.97 |
| paris | 0.363 | 5.67 |
| silent | 0.207 | 3.46 |
| tempete | 0.106 | 2.15 |
| 均值 | 0.186 | 3.19 |

根据测试序列中运动信息的多少, 可以把所有测试序列分为 3 类: 低运动或静止背景序列, 中等运动序列以及高运动序列。对于低运动或静止背景序列, 如 paris, CAC 相比 JSVM 编码性能提高了 0.3 dB~0.4 dB, 如图 2 所示; 对于中等运动序列, 如 news, CAC 的编码性能提高了 0.2 dB~0.3 dB, 如图 3 所示; 对于高运动序列, 如 foreman, CAC 的性能提升比前 2 类序列稍低, 但也有 0.1 dB~0.2 dB, 如图 4 所示。

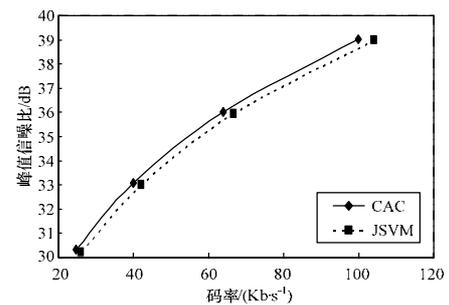


图 2 paris 序列的编码性能曲线

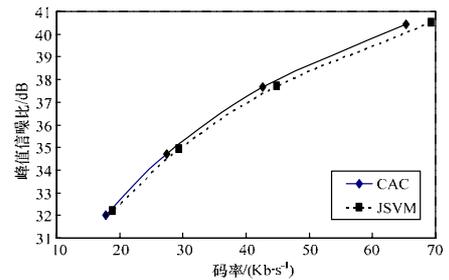


图 3 news 序列的编码性能曲线

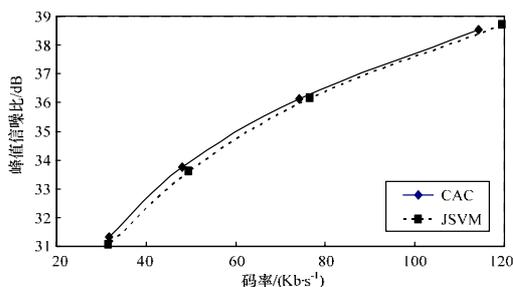


图4 foreman 序列的编码性能曲线

由此可以看出,本算法对所有序列都有较大的编码性能提升,尤其是对于运动特征较小的序列,因为在这类序列中,视频图像内采用帧内编码的宏块数目比运动剧烈的序列少。

5 结束语

本文针对当前 SVC 标准技术中采用的分级 B 帧预测编码框架,提出一种基于视频内容的自适应 QP 级联算法。该方法同时考虑了一个 GOP 中各图像之间的参考关系及运动预测给图像带来的能量变化,从而为 GOP 中每帧图像选择最佳 QP 值。通过与 SVC 参考软件的比较可看出,本算法相比标准技术最高可有 0.4 dB 的峰值信噪比提升或 6% 的码率节省。

编辑 张帆

(上接第 212 页)

参考文献

- [1] 赵知劲, 郑仕链, 尚俊娜, 等. 基于量子遗传算法的认知无线电决策引擎研究[J]. 物理学报, 2007, 56(11): 6760-6766.
- [2] 赵知劲, 郑仕链, 尚俊娜. 认知无线电技术[M]. 北京: 科学出版社, 2008.
- [3] Peng Chunyi, Zheng Haitao, Zhao B Y. Utilization and Fairness in Spectrum Assignment for Opportunistic Spectrum Access[J]. ACM

Mobile Networks and Applications, 2006, 11(4): 555-576.

- [4] Rahimi-Vahed A, Mirzaei A H. Solving a Bi-criteria Permutation Flow-shop Problem Using Shuffled Frog-leaping Algorithm[J]. Soft Computing, 2008, 12(5): 435-452.
- [5] Eusuff M, Lansey K, Pasha F. Shuffled Frog-leaping Algorithm: A Memetic Meta-heuristic for Discrete Optimization[J]. Engineering Optimization, 2006, 38(2): 129-154.

编辑 陈晖

(上接第 214 页)

表 1 人眼垂直积分投影波形特征参数值

| 垂直积分投影波形 | 波形特征参数 | | | 眼睛状态判定 |
|-------------|--------|------|------|--------|
| | 面积比 | 高宽比 | 平坦度 | |
| 图 5(b)的左眼波形 | 0.58 | 3.02 | 0.33 | 张开 |
| 图 5(c)的右眼波形 | 0.59 | 2.87 | 0.40 | 张开 |
| 图 6(b)的左眼波形 | 0.71 | 5.56 | 0.58 | 闭合 |
| 图 6(c)的右眼波形 | 0.74 | 6.62 | 0.64 | 闭合 |

表 2 不同识别方法的性能比较

| 方法 | 性能 | 是否适用人眼倾斜 | 预存信息 |
|------|----|----------|----------------|
| 本文方法 | 优 | 是 | 不需要预存 |
| 模板匹配 | 劣 | 是 | 需要预存多个模板, 信息量大 |
| 积分投影 | 优 | 否 | 不需要预存 |

5 结束语

本文提出一种识别人眼疲劳的方法,在定位眼睛后用活动轮廓检测眼睛,结合积分投影与活动轮廓的方法对人眼疲劳进行识别。该方法具有简单、快速、有效和鲁棒性等优点,

是疲劳驾驶识别系统的关键技术之一,但在活动轮廓误差较大时,其识别效果较差,这也是下一步的研究方向。

参考文献

- [1] 张荣国, 刘小君, 王蓉, 等. 自适应梯度矢量流轮廓提取方法研究[J]. 模式识别与人工智能, 2008, 21(6): 799-805.
- [2] 王蓉, 张荣国, 王艳艳, 等. 一种改进的 GVF 模型[J]. 计算机应用, 2008, 28(S1): 226-229.
- [3] Kass M, Witkin A, Terzopoulos D. Snakes: Active Contour Models[C]//Proc. of the 1st International Conference on Computer Vision. [S. l.]: IEEE Computer Society Press, 1987.
- [4] 黎云汉, 朱善安. 一种改进的 Snake 眼睛边缘检测算法[J]. 光电工程, 2007, 34(10): 78-83.
- [5] 曹菊英. 基于积分投影的司机图像眼睛定位方法[J]. 信息技术, 2007, 31(7): 20-22.

编辑 顾姣健