

一种可伸缩性视频编码的帧速率有效控制方法

赵志杰^{1),2)} 林茂六¹⁾ 朱斌¹⁾ 林亚¹⁾ 黄亮²⁾ 高鑫²⁾

¹⁾(哈尔滨工业大学电子信息技术研究院, 哈尔滨 150001) ²⁾(哈尔滨商业大学计算机与信息工程学院, 哈尔滨 150028)

摘要 可伸缩性视频编码(SVC)技术是图像和视频处理中一个新的研究领域,近年来取得了很多成果。在介绍 SVC 的有关概念和国内外研究现状的基础上,提出了一种有效的帧速率控制方法。该方法对视频序列建立了图像组(GOP)结构,并给出了对其采用分层运动补偿时间滤波器(MCTF)实现方案以记录帧间的相关性信息。实验结果表明,分层运动补偿时间滤波器结构能提供较高的处理速度和良好的性能,并且能灵活地实现 SVC 的时间伸缩性。

关键词 视频压缩 可伸缩编码 运动补偿时间滤波 小波变换

中图分类号: TN919.81 **文献标识码:** A **文章编号:** 1006-8961(2006)08-1156-05

An Effective Frame Rate Control Method for Scalable Video Coding

ZHAO Zhi-jie^{1),2)}, LIN Mao-liu¹⁾, ZHU Bin¹⁾, LIN Ya¹⁾, HUANG Liang²⁾, GAO Xin²⁾

¹⁾(School of Electronics and Information Technology, Harbin Institute of Technology, Harbin 150001)

²⁾(School of Computer and Information Engineering, Harbin University of Commerce, Harbin 150028)

Abstract Scalable video coding(SVC) technology is a new domain in the processing of image and video, and there are many results of SVC at present. This paper introduces the main concept of SVC and brings forward an effective method to control the velocity of the frame. The result of experiments indicates the Motion Compensated Temporal Filter(MCTF) framework can perform higher velocity of processing and better performance, and also neatly achieve the temporal scalability of SVC.

Keywords video compression, scalable video coding, motion compensated temporal filter, wavelet transform

1 引言

现在广泛使用的传统视频信号编解码标准如 MPEG-1, 2, 4 及 H. 26X 虽然对于不同的用户需求/网络传输条件提供了多种码率选择,但当用户需求/网络传输条件改变时,高码率数据流不能被直接编辑从而生成低码率数据流,码率的转换必须使用特定的技术和设备。也就是说,视频信号一旦经过编码,生成的数据流就不再适用于不同的用户需求/网络传输条件了。SVC(scalable video coding)就是针对上述问题提出的一种全新的编码技术。这种编码

技术所产生的数据流可以根据不同的用户需求/网络传输条件,使用简单高效的提取模块对其进行编辑,并使用统一的解码方案,从而为用户提供完全个性化服务。

为了解决视频编码的可伸缩性问题,近年来,很多学者做了多方面努力,取得了丰富的成果。对具有预测的编码框架,Ohm^[1]提出了增加对网络和设备变化适应能力的编码方案,具有一定的伸缩性。在这些具有预测的编码方法中,比较有代表性的是针对 MPEG-4 标准,实现空间可伸缩性和颗粒度可伸缩性(FGS)^[2]。但是,这些可伸缩性方法与传统的视频信号编解码标准相比,编码效率较低。采用

基金项目: 中国航天科技创新基金(200416);黑龙江省自然科学基金(F200501)

收稿日期: 2005-09-12; **改回日期:** 2006-03-06

第一作者简介: 赵志杰(1963 ~),男,副教授。现为哈尔滨工业大学信号与信息处理专业博士研究生。目前主要研究方向为视频图像处理等。E-mail: zhijiezz@yahoo.com.cn

该类编码方法的图像信息不适合在低带宽环境下传输。

小波变换视频编码方案可以从不同角度在一定的码率范围内提供连续的空间、时间、PSNR 和复杂度可伸缩性,并能保持较高的编码效率。在采用帧内小波变换编码方法去除时间方面信息冗余研究方向上,Ohm^[3]首次提出不采用闭环结构,而是采用基于非时间递归结构的运动补偿时间滤波(MCTF)。Choi 和 Woods^[4]在后来的研究中对 MCTF 进行了改进。

单层 MCTF 只能提供有限的时间伸缩性,即帧速率只能减半。为了扩展这种有限的时间伸缩性,提出了分层 MCTF 结构,给出了一种基于运动补偿时间滤波(MCTF)来实现 SVC 帧速率控制的有效方法。该方法对视频序列建立了图像组(GOP)结构,并给出了对其采用分层 MCTF 记录帧间相关性信息的实现方案。通过保留或舍弃这些相关性信息,可以灵活地实现 SVC 的帧速率控制。

2 基于宏块的运动补偿时间滤波器

2.1 当前基于宏块的运动估计(ME)视频编码

在视频序列中,相邻的帧间有很多相关的信息。这意味着这些帧对之间能提取出大量冗余信息。通过运动估计得出帧之间的相关性可以有效地去除这种帧之间的冗余。当前多数采用基于宏块的运动估计,就是一帧图像按照一定的大小分成若干块(块的大小一般 8×8 像素或 16×16 像素等),通过帧之间比较宏块之间的相关性得出宏块到宏块的运动向量。例如,目前普遍采用的 IPB 帧模式,就是通过记载参考帧的数据,参考帧到预测帧的运动向量以及重构误差来达到重构预测帧的目的。

2.2 基于平均绝对误差标准的宏块匹配

运动估计需要实现对运动矢量的方向和距离的检测和跟踪,例如确定宏块的运动趋势或某些特征点。当进行运动估计时,一般预测帧被划分为许多宏块。由于亮度运动信息与色度运动信息的特征极其相近,故色度块和亮度块用相同的运动向量,因此,一般只需要对亮度进行运动估计。通过对每个亮度宏块进行 2 维空间搜索来确定运动矢量,该搜索实现在相邻参考帧中寻找最佳的匹配宏块。进行 2 维空间搜索时如图 1 所示,首先确定预测帧当前宏块的位置,然后在参考帧中的相应位置按照一定的搜索方法依次遍历搜索范围内所有的宏块,并计

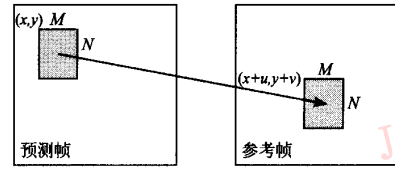


图 1 运动估计中的两帧

Fig. 1 Two frames in the motion estimate

算和预测帧的当前宏块的误差,确定最小误差宏块,从预测帧的当前宏块到参考帧的最小误差宏块的位移即为所要寻找的运动向量。

运动估计可以使用不同的误差衡量标准,如:绝对误差和(SAD)、分级倍差和(HSDD)、平均绝对误差(MAE)和均方误差(MSE)。在这些方法中,通常使用 SAD 和(MSE)。心理视觉学研究发现 MAE 并不是人眼对图像感知到的误差,它虽然不是最优方法,但它的计算成本相对较低^[1]。在此,采用了 MAE 误差衡量标准。

2.3 基于宏块的全面搜索方法

根据宏块移动方式的不同,ME 可以采用以下几种方法:

- (1) 全面搜索法;
- (2) 局部阶梯-下降搜索法;
- (3) 菱形搜索;
- (4) 六边形搜索;
- (5) 十字形搜索;
- (6) 分层大小可变块匹配(HVSBM)。

在以上方法中,ME 经常使用的是全面搜索法和 HVSBM 方法。全面搜索法给出最佳匹配结果,但计算成本很高;而 HVSBM 方法计算速度很快,但精确度略低些。在效率和精确度之间折衷考虑,在此选用了全面搜索方法。

在图 2 中,宏块的大小为 $N \times M$ (典型大小为 8×8 或 16×16 像素,这种选择是在精确度和计算成

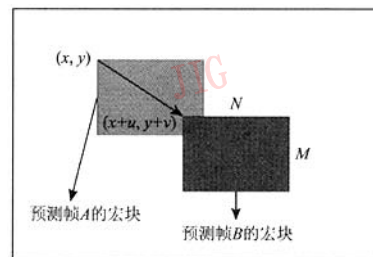


图 2 运动估计块匹配

Fig. 2 Block-matching in the motion estimate

本中间的很好折衷)。在参考帧中,搜索范围为 P , 则计算 MAE 的公式如下:

$$MAE(u, v) = \frac{1}{MN} \sum_{k=0}^{M-1} \sum_{l=0}^{N-1} |P(x+k, y+l) - R(x+k+u, y+l+v)| \quad (1)$$

$u, v \in [-P, P]$

$MAE(u, v)$ 为最小的 u 和 v , 即是要找的运动向量 (u, v) 。

3 运动补偿时间滤波器 (MCTF)

在进行运动估计时,通常情况下,仅由粘贴这些宏块并不能很好地重建参考帧,因为旋转和块变换信息并不包含在运动矢量中。为解决该问题,必须引入另一帧来记录预测帧和运动补偿的帧之间的差异。而实际上传输的是预测帧和运动补偿的帧的均值,而不是参考帧。这种操作称为运动补偿时间滤波(MCTF),产生的均值帧称为低通帧,差异帧称为高通帧。MCTF 结构如图 3 所示。

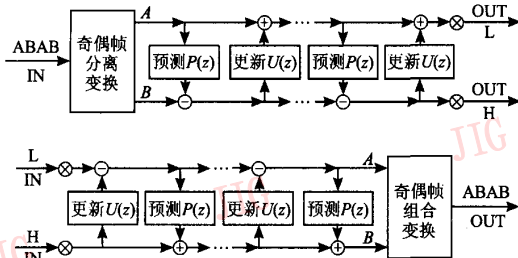


图 3 采用迭代方法实现 MCTF 的编解码功能框图

Fig. 3 Implement the function of coding and decoding using iterative method

MCTF 在消除帧间的冗余时,采用基于 ME 方法。并建立在运动补偿的概念上,采用小波变换手段。小波滤波采用迭代结构来完成,如图 3 结构中,预测和更新算法包含了运动补偿(MC)和逆运动补偿(IMC)。

在迭代滤波结构中,采用 Harr 小波滤波方法完成 MCTF 操作。首先将输入视频序列分为两大类:预测帧 A (奇数帧)和参考帧 B (偶数帧),相邻的 A 和 B 构成一对,然后使用以下公式实现每对帧对的预测运算^[4]:

$$H(m, n) = \frac{1}{\sqrt{2}}(A(m, n) - B(m+u, n+v)) \quad (2)$$

其中, H 为高通帧(帧对间的差异), (m, n) 为帧 A 中当前像素的位置, (u, v) 为相应的运动矢量。

高通帧记录了当前像素和相应的运动补偿像素间的差异。

获得高通帧后,采用下式对帧对实施更新运算,以获得低通帧。

$$L(m, n) = \frac{1}{\sqrt{2}}(B(m, n) + A(m-u, n-v)) \quad (3)$$

其中, L 为低通帧(帧对间的平均), (m, n) 为帧 B 中当前像素的位置, (u, v) 为相应的运动矢量。

式(2)、(3)是基于这样的假设:帧 B 中的像素仅和帧 A 中的一个像素存在联系。但实际上,很多情况下,帧 B 中并非所有的像素都能满足这一假设条件。像素间的联系存在着以下几种情况,针对不同情况进行如下处理:

(1) 非参考像素 参考帧中不用于参考的像素。此类像素仅定义在参考帧。

如果非参考像素在预测帧中没有相对应的像素,它的更新算法为

$$L(m, n) = \sqrt{2}B(m, n) \quad (4)$$

(2) 单一关联像素 参考帧中的一个像素仅作为预测帧中一个像素的参考。

(3) 多关联像素 参考帧中的一个像素作为预测帧中多个像素的参考。

多关联像素的更新算法为

$$L(m, n) = \frac{1}{S(m, n)} \sum_{i=1}^{S(m, n)} H(m-u_i, n-v_i) + \sqrt{2}B(m, n) \quad (5)$$

其中, S 为多关联像素个数(对帧 B 中像素而言),元素值等于帧 A 中与当前参考帧的 (m, n) 位置的像素有联系的像素的数目(取非参考帧像素为 0,单一关联像素为 1,多关联像素大于 1)。

(4) 帧内像素 预测帧中的像素在参考帧没有可接受的参考像素。

如果帧内像素没有参考像素,则预测算法为

$$H(m, n) = \frac{1}{\sqrt{2}} A(m, n) \quad (6)$$

4 分层 MCTF 和帧速率控制

由于低通帧记录了低频成分,高通帧保留了高频成分,并且人眼对低频成分要比高频成分敏感得多,所以只要保留前者而舍弃后者来减少帧的数目,

即可以控制帧速率。

但是单层 MCTF 只能提供有限的时间伸缩性,帧速率只能被减半。为了扩展这种有限的时间伸缩性,提出了分层 MCTF 结构。对一个视频序列,进行多层 MCTF 来更有效地消除冗余。对前一层得到的低通帧再应用 MCTF 来产生新的低通帧和高通帧序列,直到只剩下一个低通帧和高通帧为止。每个图像组的帧数都被提前定义,这里定义为 16。分层 MCTF 结构如图 4 所示。

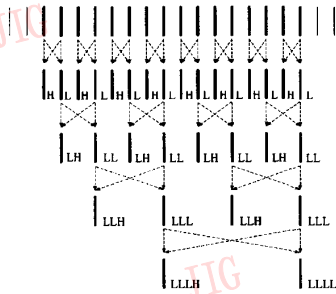


图 4 分层 MCTF 结构

Fig.4 Delamination structure of MCTF

在对视频序列应用分层 MCTF 之后,最后得到 1 个低通帧 t-LLLH 和一系列的高通帧,包括 1 个 t-LLLH,2 个 t-LLH,4 个 t-LH 和 8 个 t-H。如果要求帧速率减半,可舍弃 8 个 t-H,用剩余帧来重构视频序列。重构操作是 MCTF 的逆过程(IMCTF)。重构的帧是采用做第 1 层 MCTF 后得到的 8 个低通帧,它们保留了原始视频序列的大部分重要的低频信息,所以可看作是原始帧的近似。根据这种原理,如果要求帧速率减少到初始的 1/4,则可以舍弃 8 个 t-H 和 4 个 t-LH,而仅保留剩余的帧。使用这种分层的 MCTF 结构,帧速率可以得到有效的控制。

5 实验结果和讨论

为检验本文提出的帧速率控制方法,使用 QCIF 格式的标准测试序列 foreman 进行测试。其帧速率为 30Hz。表 1 给出了采用不同的搜索范围重建的视频序列的 PSNR。相应的 PSNR 曲线如图 5 所示。

PSNV、PSNU、PSNY 分别代表 V、U、Y 3 个分量的 PSNR 值,实线搜索范围为 16,点划线搜索范围为 8。

由表 1 可见,应用分层 MCTF 方法实现视频序

表 1 不同搜索范围的 PSNR

Tab.1 PSNR in different search ranges

| | 15fps | 7.5fps | 3.75fps | 1.875fps |
|--------------------|--------|--------|---------|----------|
| 搜索范围 = 8 PSNY(dB) | 43.488 | 40.727 | 38.709 | 37.106 |
| PSNU(dB) | 51.569 | 48.908 | 47.291 | 46.203 |
| PSNV(dB) | 52.536 | 49.746 | 48.035 | 46.866 |
| 搜索范围 = 16 PSNY(dB) | 44.337 | 42.527 | 40.882 | 39.437 |
| PSNU(dB) | 52.636 | 50.630 | 48.884 | 47.742 |
| PSNV(dB) | 53.445 | 51.464 | 49.437 | 48.235 |

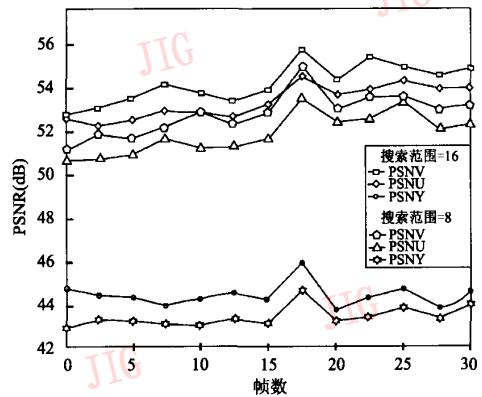


图 5 不同搜索范围的 PSNR 的比较

Fig.5 Comparing PSNR in different search ranges

列压缩,产生的比特流可被有效处理(超过6400fps)来达到不同的帧速率,并且重建视频序列的 PSNR 在 40 ~ 50dB 附近,这表明重建的序列相当成功。如果扩大搜索范围,可以提高编解码器的性能。而运动估计模型的性能越好,实现的结果越理想。同时,帧速率的调整级对 PSNR 有相似的影响,因为由舍弃高通成分导致的 ME/MCTF 的误差在每一层都在被累加。

6 结论

在本文中,我们提出了分层 MCTF 结构,给出了针对可伸缩性视频编码采用分层 MCTF 结构实现时间伸缩性的研究方法。实验结果证明,该方案是一个有效的选择,处理速度快并且编码性能好。同时,运动估计的效率将直接影响整个系统的性能,在以后的工作中有待于进一步改进。

参考文献 (References)

- 1 Ohm Jens-Rainer, Li W, *et al.* Summary of Discussions on Advanced Scalable Video Coding [R]. JTC1/SC29/ WG11 MPEG01/M7016, Geneva, Switzerland; ISO/IEC, 2001.
- 2 Radha Hayder, van der Schaar Mihaela, Chen Ying wei. The MPEG-4 fine-grained scalable video coding method for multimedia streaming over IP [J]. IEEE Transactions on Multimedia, 2001, 3(1): 53 ~ 68.
- 3 Ohm Jens-Rainer. Three-dimensional subband coding with motion compensation [J]. IEEE Transactions on Image Processing, 1994, 3(5): 559 ~ 571.
- 4 Choi Seung-Jong, Woods J W. Motion compensated 3-D subband coding of video [J]. IEEE Transactions on Image Processing, 1999, 8(2): 155 ~ 167.
- 5 Mehrsersht Nagita, Taubman David. Adaptively weighted update steps in motion compensated lifting based on scalable video compression [A]. In: Proceeding of the IEEE International Conference on Image Processing [C], Sydney, Australia, 2003: 771 ~ 774.