

视频压缩标准中的可分级编码技术^{*}

王相海^{1,2} 谭茹¹

(辽宁师范大学计算机与信息技术学院 大连 116029)¹

(南京大学计算机软件新技术国家重点实验室 南京 210093)²

摘要 视频的渐进传输、视频数据库浏览等多分辨率环境下的多媒体应用产生了视频可分级编码。本文首先对视频可分级编码的含义和种类进行了说明,然后对重要的视频压缩国际标准进行了概述,在此基础上,对视频压缩国际标准中各类视频可分级编码的编码思想和研究进展进行了分析和讨论,最后对视频可分级编码技术的发展进行了展望。

关键词 视频,可分级编码,国际标准

The Scalable Coding Algorithms in Video Data Compression Standards

WANG Xiang-Hai^{1,2} TAN Ru¹

(College of Computer and Information Technology, Liaoning Normal University, Dalian 116029)¹

(State Key Laboratory for Novel Software Technology, Nanjing University, Nanjing 210093)²

Abstract The multimedia applications in a multiresolution environment, such as progressive transmission of video, video database browsing, etc., arise video scalable coding algorithm. This paper first explains the concepts of all kinds of video scalable coding, and makes a summary for significant video data compression international standards. In addition, the coding algorithm and research progress on all kinds of video scalable coding in international standards are analyzed and discussed. Finally, the future development of video coding is also analyzed and conjectured in the paper.

Keywords Video, Scalable coding, International standard

1 引言

在许多实际应用中,由于用户要求不同、终端能力不同、异构网络的不同支路所能提供的 QoS 不同,或网络传输条件的变化(比如噪声、拥塞等)等,需要提供不同质量的图像和视频信号,比如视频会议、视频点播、多媒体数据库浏览等等。解决此类问题的最好方法是用单个编码器产生分层次的压缩码流。对不同层次的码流解码,可以获得不同的图像质量。例如低档次终端只对码流的一部分进行解码,获得低分辨率的图像;而处理能力高的终端,对整个码流进行解码,获得高分辨率的图像。这种编码机制通常称为可分级编码。一个视频编码的比特流具有可分级性(Scalability),是指利用该比特流的一部分(或子集)可以产生对该视频图像一个有用表示的能力,即解码器能够根据分辨率的情况对码流的一部分进行解码。视频的可分级编码通常包括以下情况^[1,2]:空间分辨率可分级编码、数率可分级(或称 SNR 可分级)编码和时间分辨率可分级编码。视频的空间可分级编码是指以不同空间分辨率对图像进行解码;视频的数率可分级指对一个压缩的码流,用户可根据带宽或系统能力的具体情况从码流的开始端来截取相应的一段码流进行解码。在对码流的截取过程中,如果可在一定范围内可任意选择数率,则此时的数率可分级称为连续数率可分级。视频的时间分辨率可分级是指在解码端以不同的帧率进行解码。

视频的可分级编码机制在早期的一些视频国际标准中就

已经体现,比如 MPEG-2^[3],到了近期的 MPEG-4^[4]这一编码思想被明确提了出来,并且随着流媒体技术的应用,视频的可分级编码技术越来越受到重视。本文首先对目前常见的视频压缩国际标准进行了概述,然后对标准中的可分级编码技术的思想和方法进行了分析和讨论,最后对视频编码技术的发展方向进行了展望。

2 常见的视频压缩国际标准概述

ITU-T 与 ISO/IEC 是制定视频编码标准的两大组织^[5]。ITU-T 的视频编码标准主要包括 H. 261、H. 263 和 H. 264 等,主要应用于实时视频通信领域,如电视会议等;ISO/IEC 的主要视频标准包括 MPEG-1、MPEG-2、MPEG-4 等,主要应用于视频存储、广播电视、因特网或无线网上的流媒体等。下面分别对这些标准进行简单概述。

(1) H. 261

H. 261 是 1990 年 12 月完成并予以通过的在 ISDN 上开展双向声像业务的国际标准,主要用于可视电话和视频会议等,即“采用 $p \times 64\text{kb/s}$ 的声像业务的图像编解码”。该标准是最早的运动图像压缩标准,它只对 CIF 和 QCIF 两种图像格式进行处理,其编码过程融合了基于 DCT 的编码方法和带运动估计的 DPCM 预测编码算法。每帧图像被分成图像层、块组层、宏快层和块层来处理。但在标准中并没有提出可分级的编码机制。

(2) MPEG-1

^{*}国家自然科学基金(60372071)、辽宁省自然科学基金(20032105)、辽宁省高等学校优秀人才支持计划(RC-04-11)和大连市科技基金资助。王相海 博士、教授,CCF 高级会员,主要研究领域为 CG、CAGD 和多媒体信息处理;谭茹 硕士生,研究方向为计算机图形学、多媒体信息处理。

MPEG-1 标准是 ISO 的活动图像专家组 MPEG 于 1993 年 8 月公布,用于传输 1.5Mbps 数据传输率的数字存储媒体运动图像及其伴音的编码。MPEG-1 中定义了三种图像类型:I、P、B 图像。I 图像及帧内图像,采用帧内编码,不参考其它图像;P 图像即预测图像,采用帧间编码,参考前一幅 I 或 P 图像,用作运动补偿;B 图像即双向预测,参考前后两个方向的图像。MPEG-1 的编码结构与 H. 261 类似,也采用分层结构,不同的是 MPEG-1 多出片层(或 Slice 层),用于防止误码在一帧内扩展。MPEG-1 的视频编码器与 H. 261 并无太大区别,但需考虑双向预测编码和解码。图像 MPEG-1 中也没有考虑可分级编码机制。

(3) MPEG-2

MPEG-2 是 MPEG 于 1994 年推出的编码率从 4Mbps 到 15Mbps 的压缩标准,该标准是针对标准数字电视和高清晰度电视在各种应用下的压缩方案和系统层的详细规定,因而特别适应于广播级的数字电视的编码和传送,被认定为 SDTV 和 HDTV 的编码标准。

MPEG-2 的最显著特点是通用性,它是 MPEG-1 的超集,基本算法相同,但增加了 MPEG-1 所没有的许多功能,例如增加了对隔行扫描的编码,提供了可分级(scalable)的功能。MPEG-2 视频的编码按处理方法的复杂程度分成“类”(profile)。所谓类是指 MPEG-2 的不同处理方法,每一类都包括压缩处理方法的一个集合,不同的类意味着使用不同集合的压缩编码工具,类越高,编码越精细。MPEG-2 视频编码共分成五类:第一类是简单类,仅使用 I、P 图像。第二类是主类,使用 I、P、B 图像,比简单类增加了双向预测方法,在相同码率的情况下,能给出比简单类更好的图像。第一类和第二类都没有可分级性。第三类是质量可分级类,它使用 I、P、B 图像,按信噪比分级。第四类是空间可分级类,它也使用 I、P、B 图像,按空间可分级。第五类是高类,同样使用 I、P、B 图像,按空间或信噪比均可分级。

MPEG-2 的编码码流分为类个层次,并用句法规定了一个层次性结构,从上到下分别是:图像序列层、图像组、图像、宏块条、宏块和块。

(4) H. 263

H. 263 是 1996 年 ITU-T 为低于 64kb/s 的窄带通信信道制定的视频编码标准,但实际上该标准可以用在很宽的码流范围。该标准是在 H. 261 基础上发展起来的,但与 H. 261 相比,H. 263 在提高编码性能和纠错能力等方面得以大大增强,主要体现在以下几方面:H. 263 的运动补偿采用了半像素精度;支持 CIF、QCIF、SQCIF、4CIF 或 16CIF 的彩色 4:2:0 亚取样图像;包含了四个可协商的选项以改善性能;采用无限制的运动向量和基于语法的算术编码;数据流层次结构的某些部分是可选的,从而使得编解码可以配制成更低的数据流或更好的纠错能力。1998 年 ITU-TL 在 H. 263 之后又推出了 H. 263 的版本 2,记为 H. 263+^[6],它提供了 12 个新的可协商模式和其它特征,进一步提高了压缩编码性能。比如其中一个重要的特性是增加了时间可分级、信噪比可分级和空间可分级的编码机制,增强了视频信息在易误码、易丢包的异构网络环境下的传输能力。此外,在 H. 263+ 基础上推出的 H. 263++^[6],为了增强码流在恶劣信道上的抗误码性能和提高编码效率,又在 H. 263+ 的基础上增加了 3 个选项。

(5) MPEG-4

MPEG 于 1999 年 2 月公布了 MPEG-4 标准的第一版

本,于 2000 年年初公布了第二版本,并正式成为国际标准。MPEG-4 与以前标准的显著差别在于它是采用基于对象的编码理念,即在编码时将景物分成若干在时间和空间上相互联系的视频音频对象,将这些对象分别编码后再经过复用传输到接收端,对不同的对象进行解码后再组合成所需要的视频和音频,从而方便对不同的对象用不同的编码方法和表示方法,同时也可以方便地对各种对象进行操作和编辑。

MPEG-4 的结构与 MPEG-1 和 MPEG-2 不同,它构造一个工具箱的形式,通过定义框架(Profile)、级(Level)、算法(Algorithm)和工具(Tool)等四层结构来分层次描述具体应用中的编码方案。MPEG-4 标准具有高效压缩、基于内容交互以及基于内容的空间、时间等可分级的特点,可应用于因特网视音频广播、无线通信、交互式视频游戏和多媒体邮件等领域。

(6) H. 264

H. 264 是由 ISO/IEC 与 ITU-T 组成的联合视频组 JVT 制定的新一代视频压缩编码标准^[7]。它不仅具有优异的压缩性能,而且对网络传输具有更好的支持功能^[8];它引入了面向 IP 包的编码机制,有利于网络中的分组传输,支持网络中视频的流媒体传输。此外,H. 264 具有较强的抗误码特性,可适应无线信道中的视频传输。该标准支持不同网络资源下的分级编码传输。H. 264 标准分为三个档次^[7]:基本档次、主要档次和扩展档次。基本档次版本简单,利用 I 帧和 P 帧支持帧内和帧间编码,支持利用基于上下文的自适应的变长编码进行的熵编码 CAVLC,其应用面广,可用于可视电话、会议电视、无线通信等实时视频通信等;主要档次采用了多项提高图像质量和增加压缩比的技术措施,比如支持隔行视频,采用了 B 帧的帧间编码和加权预测的帧内编码,支持利用基于上下文的自适应的算术编码 CABAC,可用于 SDTV、HDTV 和 DVD 等;扩展档次支持码流之间有效地切换、改进误码性能,可用于各种网络的视频流传输,但该档次不支持隔行视频和 CABAC。

与 H. 261、H. 263 一样,H. 264 也采用基于 DCT 变换编码和 DPCM 的差分编码的混合结构。但在该混合结构框架下,H. 264 引入了新的编码方式,大大提高了编码效率,同时具有适应多种信道的能力。

3 视频压缩标准中可分级编码技术分析

3.1 MPEG-2 中的可分级编码分析

MPEG-2 中的质量可分级类、空间可分级类和高类的视频编码器都是采用分层机制通过 2~3 个流的组合来表示一个视频^[9,10]。MPEG-2 的这种灵活的支持多种不同分辨率的编码能力,为解决 HDTV 和 SDTV 之间的互操作性提供了解决方案,使这两个产品可以互相兼容。

3.1.1 质量可分级类

MPEG2 中视频信号的质量可分级编码方案的实现是在 DCT 域进行的。编码器中图像经过 DCT 变换后,先进行一次粗量化处理,经可变长编码和熵编码后,得到一个基层的码流。同时,粗量化处理得到的中间结果经逆量化后,与原 DCT 频域图像相减,得到的差分图像再进行一次细量化。量化结果经可变长编码和熵编码后,得到一个增强层的码流。基层表示编码图像的基本数据,但代表的图像质量低。增强层则用来改进信噪比,这就意味着必要时可以忽略比特流中的增强部分,而只解码比特流中的基础部分,仍可得到有用

的图像序列,只不过质量差一些。图 1 是 MPEG2 实现视频

信号质量可分级的一个编、解码方案。

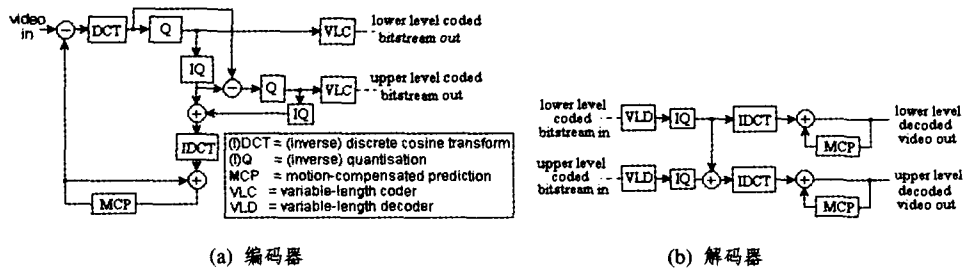


图 1 视频信号的质量可分级编码

3.1.2 空间可分级类

MPEG2 空间可分级编码的一个特点在于使用来自低层的解码图像作为高层图像的一个预测。图 2 给出了一个实现空间可分级的编解码方案。方案中总体包括两个大的“环形”编码器,分别对应基层的低分辨率图像和增强层的高分辨率图像的编码,基层编码器所产生的码流可以与非可分级编码形式相同的方式进行解码;而来自基层的局部解码图像通过“向上转换”后也可以作为预测图像对上层图像进行预测,这种预测作为对上层运动补偿预测器的补充,为增强层提高预

测精度提供了可能,增强层通过一种附加的“权函数 W”对来自上层和下层的预测作出选择。

与前面介绍的质量可分级一样,空间可分级编码器所形成的低层码流能够以非可分级情况相同的方式进行解码。而为了了解上层码流,低层码流首先被解码并将被解码的低分辨率图像通过“向上转换”后提供给上层解码器。上层解码器根据上层码流中传输过来的权函数 W 的取值在其自身的运动补偿预测和来自低层的运动预测之间作出选择,具体过程参见图 2(b)。

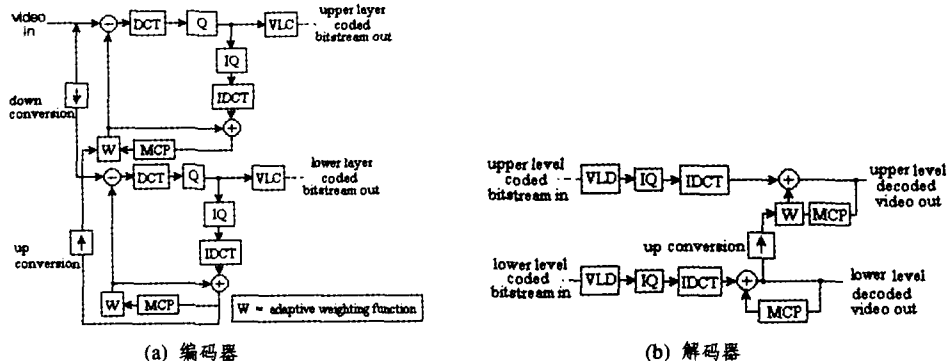


图 2 视频信号的空间可分级编码

3.1.3 高类

与前面介绍的质量可分级类和空间可分级类所使用的色度信号格式为 4 : 2 : 0 相比,MPEG2 中的高类增加了 4 : 2 : 2 的色度信号格式,此外它按质量或空间均可分级。

3.1.4 频率可分级(数据分割)

在 MPEG2 中,频率可分级也称为数据分割^[11],它是指在视频帧的多层表示中,每一层包含不同的频率分量,即基本层包含低频分量,而其它层包含逐渐增加的较高的频率分量,这样基本层提供一个模糊的图像。随着增强层的不断加入,图像会逐渐清晰。在基于块的混合编码中,实现这个思想的一种方法是:在基本层中包括模式信息、运动信息和每个宏块的前几个 DCT 系数,而在增强层中则包括其余的 DCT 系数。当然该种可分级编码也可通过子带或小波变换等整帧变换来实现。

3.2 H. 263+ 中的可分级编码分析

H. 263+ 支持时间、质量和空间可分级编码^[6]。时间可分级提供了一种通过增加显示帧率来提高图像可视质量的编码机制。H. 263+ 中的时间可分级编码机制是通过双向预测帧(即 B 帧)来实现的。由于 B 帧图像由双向预测得到,分别用前向 MV、后向 MV、前后向 MV 平均进行运动补偿,得到 3 个预测帧差,然后取其最小者作为 B 帧的预测误差进行编码。由于其它帧从不被 B 帧预测,所以 B 帧的舍弃不会影响

后面各帧的解码质量。这样,可以通过控制 B 帧的取舍来调整帧率,获得解码的时间可分级特性(参见图 3)。

值得注意的是,与 P 帧相比,使用 B 帧可以改进编码效果,但同时会增加编码时的复杂度和内存需求,并会带来一些额外的延迟。

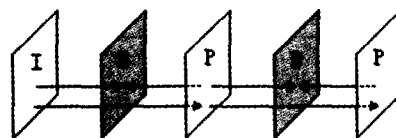


图 3 时间可分级实现示意图

H. 263+ 的数率可分级编码机制与 MPEG2 中的数率编码机制基本相同,都是通过使用更细的量化器来编码增强层的帧差图像。这些附加的信息增加了所有解码图像的 SNR,所以质量可分级也被称为 SNR 可分级。图 4 给出了一个质量可分级编码的结构图。

H. 263+ 的空间可分级通过创建一个多分辨的码流来适应客户端不同的显示需求。图 5 给出了一个空间可分级编码结构图,从图中可以看出,该结构图与图 4 的数率可分级编码结构图极其相似,所不同的是在空间可分级编码中一个空间增强层所要恢复的是重构参考层图像的上采样版本和原图像高分辨率版本之间的帧差信息。比如,如果参考层有一个

QCIF 分辨率, 并且增强层有一个 CIF 分辨率, 那么参考层图像必须被适当地按比例调整, 以便使增强层的图像能被它预测。对于一个单一的增强层, 标准中允许在水平方向、垂直方向或水平和垂直方向以 2 的系数来增加分辨率, 同时允许有多个增强层。

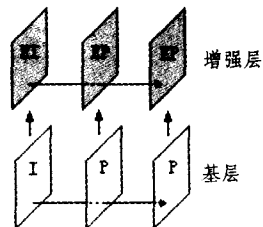


图 4 速率可分级实现示意图

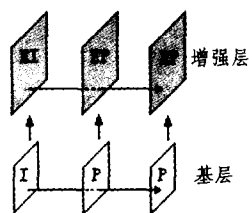


图 5 空间可分级实现示意图

在 H. 263+ 中, 无论是质量可分级还是空间可分级, 在增强层中都存在着两种类型的图像 EI 和 EP。其中 EI 图像表示该图像采用的预测方式是来自下层的参考图像, 即只采用向上的预测方式。而 EP 图像则表示该图像既能够采用来自前面增强层中的图像进行预测, 也能采用来自下层参考层中的图像进行预测。这样, 对于 EP 图像, 可以通过计算向上和向前预测的平均值来提供一种双向预测选项。值得注意的是, 对于 EI 和 EP 图像, 如果采用的是来自参考层图像的向上预测, 则隐含着不需要确定运动矢量。而对于 EP 的向前预测情况, 运动矢量的计算则是必须的。

3.3 MPEG-4 中的可分级编码分析

3.3.1 MPEG-4 中的一般可分级编码技术

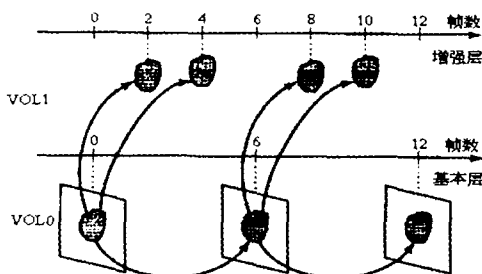


图 6 类型 1 的进行 P-VOP 预测的时间可分级结构

MPEG-4 中提供了基于对象的可分级编码机制^[12], 图 6~图 8 分别给出了 MPEG-4 中具有两类增强形式的时间可分级编码方案^[11]; 其中第一类是指视频对象层 VOL0 具有一个物体和一个背景的整体帧, 而 VOL1 是表示 VOL0 中的特殊对象, 通过以低帧率编码 VOL0, 而以高帧率来编码 VOL1, 来实现基于对象的可分级。其中图 6 中 VOL1 的对象是通过利用 VOL0 中的帧来进行预测的, 而图 7 则是通过利用 VOL0 中的帧通过双向预测来获得 VOL1 中增强层中的对象。第二类是指视频对象 VO0 是仅包含背景且没有可分级性层的整个帧的序列, 而 VO1 是一个具有两个可分级性

层 VOL0 和 VOL1 的特殊的对象序列, VOL1 表示具有与 VOL0 相同的视频对象但被以高于 VOL0 的帧率编码的视频对象层, 其中 VOL1 被认为是基本层, 而 VOL1 则被认为是增强层。

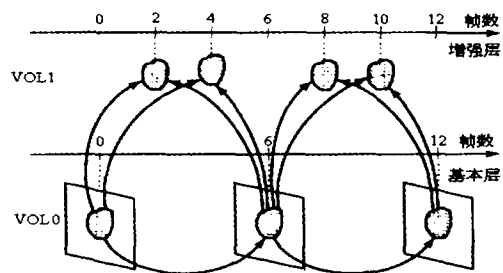


图 7 类型 1 的进行 B-VOP 预测的时间可分级结构

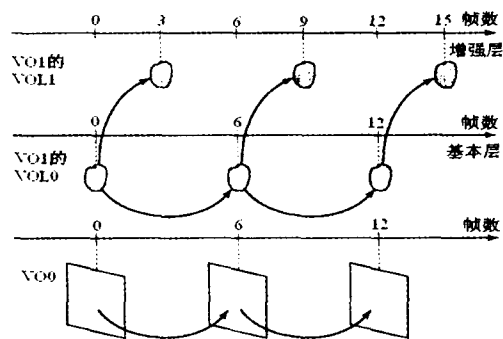


图 8 类型 2 的时间可分级增强结构

3.3.2 MPEG-4 中的 FGS 编码技术

(1) FGS 的基本思想

上述 MPEG-4 中的一般可分级编码技术的一个共同特性是要想使得解码精度得以提高, 增强层的码流必须完全解码, 否则不会增强解码的质量。MPEG 组织注意到了网络传输对视频编码的新的要求, 开始征集精细的可分级视频编码方案。该方案要求将视频编码成一个可以单独解码的基层码流和一个可以在任何地点截断的增强层码流, 其中基层码流适应最低的网络带宽, 而增强层码流用来适应网络带宽变化的动态范围。世界各地的学者提出了许多解决方案, 最终 Weiping Li 的 FGS 方案^[13] 被选中。

FGS 视频编码的基本思想是将一个视频序列编码为一个基层码流和一个增强层码流, 基层采用非可分级编码方式进行编码, 其码率作为解码端可分级解码的下界; 增强层则采用了基于 DCT 系数的位平面编码机制对原始图像和重构图像的帧差图像进行编码。MPEG-4 修订版中 FGS 编码器和解码器框架参见图 9 和图 10。

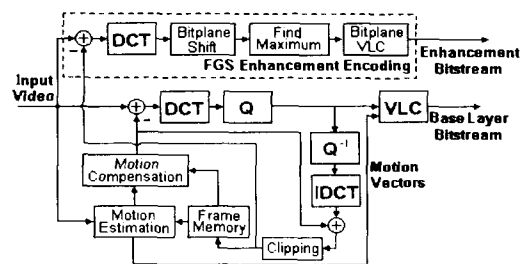


图 9 FGS 编码器结构

尽管 FGS 编码机制也是将视频序列分为两层, 但与传统

的可分级编码不同,FGS增强层可以根据对每一帧的解码精度要求,任意截取其码流,从而实现连续可分级的解码机制。在MPEG-4中的FGS编、解码器框架中,其基层采用的是MPEG-4ASP编解码方案,而增强层则是对原始DCT系数与基层粗量化系数之差用Bit-Planned的编码方式,由于采用了位平面编码技术,使得每一个系数的较高位,也就是该系数的较重要部分优先得到编码。这样,在码流截断时可以保证当前所获得的解码信息是重要信息,从而提供了精细可分级的特性。

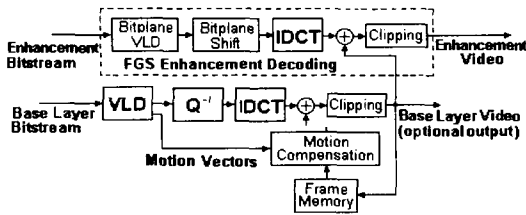


图10 FGS解码器结构

(2) FGS基础上的时间可分级编码机制

在前面FGS编码结构中,由于对时间增强帧的时间预测是在基层进行的,这样每个时间增强帧的质量不会影响到其它帧的质量,从而可以在上述FGS编码结构中结合时间可分级编码机制,这种混合的可分级编码方案被称作FGST^[14]。采用该可分级编码方案,不仅保持了FGS的精细可分级特性,而且支持帧率的变化。图11、12给出了FGST示意图。

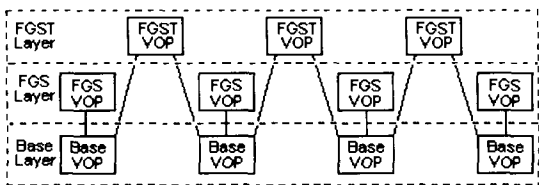


图11 FGST与FGS相分离的编码结构

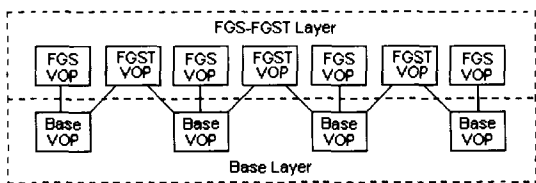


图12 FGST与FGS相合并的编码结构

(3) FGS的改进

前面介绍的FGS编码结构由于预测总是在基于最低质量的基层上进行,对增强层的预测一般不是很准确,形成的DCT参差较大,编码效率不是很高,有时甚至比传统的SNR可分级方案的编码质量还要糟糕^[15]。近年来,人们在FGS编码结构的基础上对细粒度的视频可分级编码技术进行了较深入的研究,提出了许多有效的改进算法。

为了改进FGS的编码效率,文[16,17]提出了一种渐进的精细可分级编码方案(PFGS, Progressive Fine Granular Scalable),该方法在保持了FGS所具有的细粒度可分级特性、对信道的自适应特性和对错误的可恢复能力的同时,提高了较FGS方案1dB左右的编码效率^[18]。文[19,20]提出了所谓基于宏块的渐进细粒度可分级编码,即在上述PFGS的运动预测和重建中,在宏块一级选择所使用的参考信息,使每个增强层宏块在运动补偿和重建中使用的参考宏块可灵活地

在高质量参考帧和低质量的参考帧间选择,从而给PFGS方案带来更大的灵活性,并且能够更好地在消除误差传递和提高编码效率间寻求平衡。文[21]在上述基于宏块的PFGS方案的基础上,给出了一种同时支持时域和精细SNR可分级的PFGST编码方案。与前面所介绍的FGST方案相类似,PFGST方案中基层仅包含以基层帧率进行编码的I帧和P帧,所有的时域可分级帧采用B帧形成增强层码流,其中I帧和P帧的编码与基于宏块的PFGS相同,B帧的方法与FGST类似,经过双向预测后,其帧差使用位平面技术形成时域可分级的增强层码流。与FGST不同的是,在B帧的编码过程中,每个宏块都可以选择SNR可分级的低质量或高质量的参考宏块来进行运动补偿,因而可以获得更好的编码性能。

前面所介绍的PFGS编码方案中,总是假设基本层能够得到充分的解码,这样对于带宽小于基本层码流的码率的用户而言,PFGS就不再具有网络带宽的自适应能力了。为此,文[22]中提出了一种精细的空域可分级编码方案(FGSS)。FGSS将PFGS和传统的空域可分级编码相结合。与PFGS相同,所有基层图像使用重建的基层图像来做运动预测和补偿,而增强层图像则使用前一帧重建的某个增强层图像来做运动预测和补偿。即FGSS使用两个参考图像,做两次运动预测和两次运动补偿。与PFGS相比,FGSS的基本层码流的码率更低。而与传统的空域可分级编码不同,低分辨率的视频和高分辨率的视频质量可以随码率的增加连续地加以改进。

3.4 H.264中的可分级编码分析

尽管在H.264的第一版中并不支持可分级特性,但它作为被标准支持的重要工具被列入到标准的工作计划中^[23]。2005年1月,MPEG和ITU-T的VCEG(Video Coding Experts Group)同意共同确定视频可分级编码方案作为对H.264/MPEG-4AVC标准的一个修订,并将HHI(Heinrich Hertz Institute)的图像通讯组所开发的可分级编码方案作为第一个工作草案(WD-1)^[24]。

3.4.1 H.264/MPEG-4AVC可分级扩充的基本结构

H.264/AVC可分级扩充的基本设计思想是把H.264/AVC的混合视频编码方法^[25]扩展为采用提升框架的运动补偿时间滤波器MCTF(Motion-Compensated Temporal Filtering)。图13为2分支MCTF情况,其中“P”和“U”分别代表“预测”和“修改”步骤, S_k 对应时间间隔为 k 取样的视频信号。多相分解器将 S_k 分解为 S_{2k} 和 S_{2k+1} , S_{2k+1} 采用运动补偿进行预测,获得高通信号 $H_k = S_{2k+1} - P(S_{2k})$, S_{2k+1} 和 $P(S_{2k})$ 之间的差再一次进行运动补偿,获得低通信号 $L_k = S_{2k} + U(S_{2k+1} - P(S_{2k}))$ 。对 $P(\cdot)$ 和 $U(\cdot)$ 的讨论,可参见文[26]。值得注意的是,当修改步骤被删除时,上述结构与传统混合视频编码的开环结构是相同的。在这种提升的MCTF结构中,“预测”和“修改”是分开进行的。由于“提升”是互逆的,这样任何运动补偿技术可以结合到滤波器组的“预测”和“修改”步骤中。进一步,MCTF的这种开环结构为结合空间、时间和SNR可分级编码提供了可能^[27]。

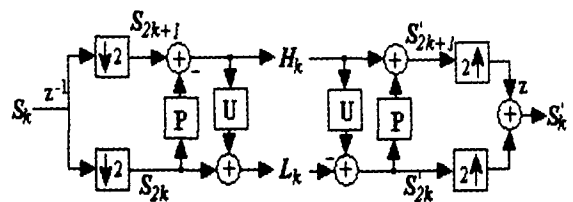


图13 2分支MCTF

H. 264/MPEG-4AVC的可分级扩充具有以下几个关键特征:分级预测结构;具有可转接的层内预测结构的分层编码机制;与 H. 264/MPEG-4AVC 相兼容的基层编码;采用渐进提炼片的细粒度质量可分级特性;H. 264/MPEG-4AVC 的NAL(Network Abstraction Layer)的使用和扩充。图 14 给出了 H. 264/MPEG-4AVC 的具有三个空间层的可分级扩充的基本结构图。图中每一层都包含一个带有特殊运动参数的独立的分级运动补偿预测结构,该结构提供了一个视频序列的时间可分级表示,同时能适应于空间和质量的混合可分级特性。不同层间的冗余通过采用包括预测参数和纹理数据的不同层内的预测予以去除。每一层输入图像的基层编码均采用与 H. 264/MPEG-4AVC 相似的变换编码,对应的 NAL 单元包含运动信息和纹理数据,这些基层表示决定了一个时空分辨率能够被解码的最小比特率。

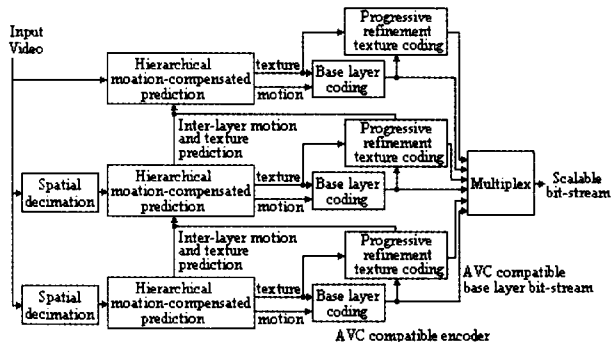


图 14 H. 264/MPEG-4AVC 可分级扩充的基本编码结构

3.4.2 时间可分级

实际上,一般的时间可分级编码机制在 H. 264/ MPEG-4AVC 标准中已经得以支持,不同的是在 H. 264/ MPEG-4AVC 可分级扩充结构中用 MCTF 代替了简单的分层图像编码。图 15 给出了一个图像序列的分层预测结构的例子,其中关键帧(第 8 帧)和该帧与其前面的关键帧(第 0 帧)之间的各帧组成了一个 GOP(Group of Pictures)。关键帧可以采用帧内编码,也可采用利用前面关键帧作为参考帧进行运动补偿预测的帧间编码,一般它们是作为最低时间分辨率被解码的。此外,关键帧也可以考虑作为编码器与解码器的重同步点来使用。一个 GOP 的其它图像采用的是分层预测的方案,比如在例子中第 4 帧利用其前后的关键帧(第 0 帧和第 8 帧)作参考被预测,由于仅仅依靠关键帧,因而它与关键帧一起可以作为下一个高一级的时间分辨率解码图像;而下一个高一级的时间分辨率图像(第 2 帧和第 6 帧)则仅仅使用前一个低分辨率图像(第 0、4、8 帧)作为参考帧进行预测,以此类推。显然,这种图像序列的分层预测结构内在提供了时间可分级特性。

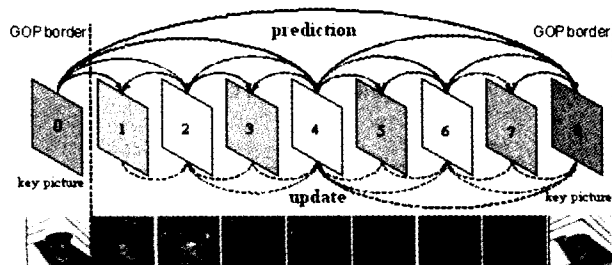


图 15 分层预测结构实例

上述图像序列的分层编码机制可以被扩展到 MCTF 中,即除了采用上面介绍的运动预测操作外,引进运动补偿的“修改”操作(参见图 15 的虚线操作)。在编码器端,MCTF 分解过程从最高时间分辨率开始,图像组被分解为 A 和 B 两部分,B 中的图像用 A 中的图像进行预测,并被运动补偿帧差图像所代替。而所获得的 B 中图像的帧差图像再一次被 A 中图像进行运动补偿,并将获得的运动补偿帧差图像加到 A 中图像中,于是 A 中的图像被一个沿着运动轨迹的低通滤波器所获得的低通版本图像所取代。在所获得的低通图像集合上重复上述过程,直到产生一个单独的低通图像作为关键帧图像。图 16 给出了一个分解过程的例子。

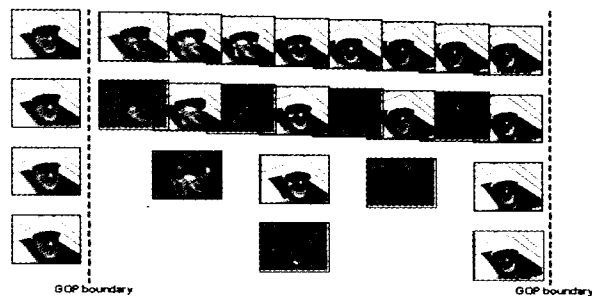


图 16 重复的 MCTF 时间分解过程

事实上,分层预测编码和 MCTF 并不只限于二进制时间可分级的情况,图 17 给出了将图像组以不同的形式划分为 A、B 两部分并进行 MCTF 分解的例子,其中假设 A、B 集合中分别为包含 N_A 个和 N_B 个图像的图像组 $\{A_k\}$ 和 $\{B_k\}$, $\{H_k\}$ 和 $\{L_k\}$ 分别为所获得的高通和低通图像。

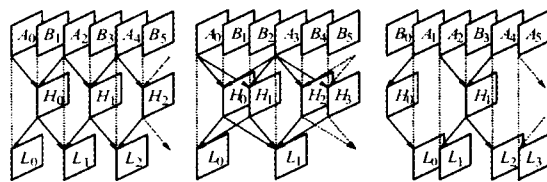


图 17 图像组按时间分解为高通和低通图像

3.4.3 空间可分级

文[28]的实验结果发现,在基于金字塔的视频层次编码结构中,允许编码器通过一种选择预测机制自由地选择空间分辨率层间的依赖关系,会使得空间可分级编码更加有效。为此,文[26]采用一种“可选择的层间预测技术”实现了一种水平和垂直分辨率为 2 的倍数的多分辨率视频空间可分级(如 QCIF, CIF 和 4CIF 等)编码方案。图 18 给出了这种层间预测方法的示意图,一个基层图像的编码信息可以同时用作增强层图像的预测信息。这种可选择的层间预测包括以下三种方式:一是采用上采样低分辨率信号的宏块预测;二是采用上采样低分辨率运动矢量的运动矢量预测;三是采用上采样低分辨率帧差信号的帧差图像预测方法。

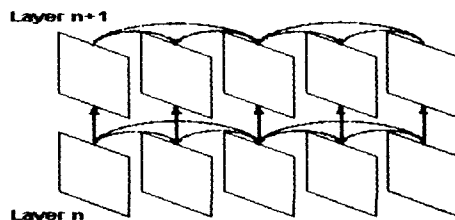


图 18 层间预测示意图

实际上,上面所述的层间预测技术也可用于基层与当前层具有相同空间分辨率的情况,只不过此时可以简单地省略上采样操作。

3.4.4 SNR 可分级

对于 SNR 基层,与混合视频编码方案一样,其高通图像包含帧内或帧差宏块。对帧差宏块,编码采用与 H. 264/AVC 中相同的变换和量化过程。而对帧内宏块,其编码也同样采用了与 H. 264/AVC 相同的编码模式。对于低通图像,则采用与 H. 264/AVC 帧内图像相似的独立的编码方式,或与 H. 264/AVC 帧间图像相似的帧间编码方式。对于 SNR 的增强层,SNR 基层和原始子带图像的量化误差采用与基层相同的方案进行重新量化,所不同的是这里采用了更加精细的量化步长。进一步,该增强层和基层合起来,可以考虑作为下一个增强层的基层,这样可以再一次应用前面所介绍的方法。这种实现 SNR 的简单方法,之所以在增强层被删除情况下仍然具有很好的编码效果,原因是方案中所采用的开环编码结构以及前面所介绍的时间分解结构。

3.4.5 细粒度 SNR 可分级

文[29]扩展了文[28,30]中 SNR 可分级方案,使其具有细粒度可分级的特性。方案中基层包通常对应一个最小的可以接受的重构质量,并且这个基本的重构质量可以通过在任意点截取增强层的 NAL 单元(包)来改进解码质量。每一个增强层包包含一个对应量化步长二等分的提炼标识,对低通和高通图像的提炼符号直接在变换域进行编码,这样在解码端对子带图像的每一个变换块只需进行一次逆变换。

为了使 SNR 增强层的 NAL 单元可在任意点被截取,修改变换系数的编码顺序。为了传输变换系数的提炼表示,方案中采用了 H. 264 中基于上下文的自适应的算术编码 CABAC,只是附加了六个额外的上下文。方案中变换系数提炼层的编码过程被分成三次扫描:第一次扫描针对具有如下两个特性的所有变换系数的提炼层进行编码:一是在基层和所有的从属增强层表示中已经被编码的系数层等于零,这些变换系数也被称为非重要变换系数;二是变换系数被定位在至少包含一个变换系数的“变换块”中,这些“变换块”中不等于零的变换系数层在基层或任何从属增强层表示中已经被传输。这些变换系数块也被称为重要的变换系数块。第二次扫描针对具有如下两个特性的所有变换系数的提炼层进行编码,即不等于零的变换系数层已经在基层或任何前面的增强层表示中已经被编码。这样的变换系数也被称为重要变换系数。最后,在第三次扫描中,所有剩余的提炼层被编码。对应的变换系数具有如下性质:一是在基层和所有从属增强层表示中已经被编码的变换系数层等于零,这些变换系数也被称为非重要变换系数;二是变换系数被定位在不包含任何变换系数的“变换块”中,这些“变换块”中不等于零的变换系数层在基层或任何从属增强层表示中已经被传输。这些变换系数块也被称为非重要的变换系数块。在每一次扫描中,对应的变换系数将以一定的顺序被传输。具体传输的伪代码可参见文[29]。

3.4.6 灵活的联合可分级

利用前面介绍的时间、空间和细粒度 SNR 可分级方案可以实现 H. 264 灵活的联合可分级编码机制^[27,29],其总体方案见图 19。其中,为了提供时间可分级并且获得一个视频序列的有效的可分级码流,图像间的视频依赖关系通过一个开环子带方法来实现,相关的时间分解-合成滤波器组结构可以方

便地实现 Haar 滤波器和 5/3 滤波器运动补偿提升表示间的基于块的自适应选择,此外还具有多参考帧的能力。进一步,选择一种基于块的帧内模式(Intra Mode),以便能够对使用运动补偿不能合理预测的块进行有效的表示。对于每一个空间层,图像组经 MCTF 分解,通过采用 H. 264 中高效的运动模型,连同 Haar 小波和 5/3 样条小波的自适应选择使得 MCTF 中的预测和提升步骤与 H. 264 中 B 分片(双向预测分片)运动补偿预测相似。对于空间基层,采用一种不带修改步骤的 MCTF,以便支持 H. 264 适应的基层表示。

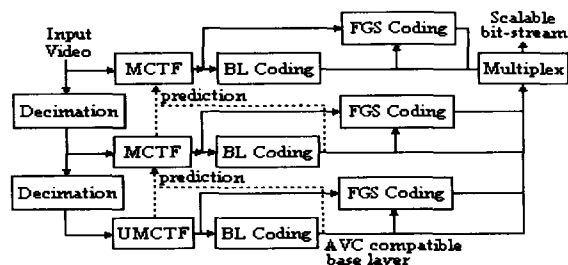


图 19 灵活的联合可分级编码总体结构图

MCTF 内在地提供了时间可分级特性。为了提供空间可分级,方案中使用一个金字塔结构。尽管 MCTF 是独立地应用在每一个空间分层,在方案中结合一个很大程度的帧间预测,表示时间高通信号的帧内宏块和帧差宏块可以使用对应的先前空间层的插值重构信号来预测,而每一个 MCTF 层的运动描述也可以用于随后增强层运动描述的预测。

进一步,一个时间子带表示的开环结构也为有效地结合 SNR 可分级提供了可能。为了获得图像组的 SNR 可分级,在方案中结合前面介绍的细粒度 SNR 可分级。对每一个空间层,子带图像的基层编码(BL Coding)是通过使用包括基于块的变换和上下文自适应二进制算术编码(CABAC)在内的 H. 264 的帧内和帧差编码工具对低通和高通图像 L 和 H 进行编码来实现的。最低空间层的基层表示与 H. 264 的比特流是一致的。子带图像的增强层则采用细粒度可分级编码方式,可以在任意点上截取对应的 NAL。对于一个任意时间-空间分辨率,必须传送一个表示空间基层的最小比特流,它可以根据实际应用中可接受的最低视频质量的重构表示来进行调整。而对于高于这个最低比例率的任意比特率,都可以以一个合适的方式截取对应空间-时间分辨率和所有较低分辨率层的细粒度可分级的 NAL 单元来实现。

结束语与展望 多分辨率环境下的多媒体应用导致了视频可分级编码技术的提出和发展,近年来人们在这一领域进行了积极的研究,出现了许多有效的视频可分级编码方案。但由于该研究领域还处于不断发展阶段,在许多地方还有待进一步完善和研究。我们认为,未来在该领域的研究过程中,以下几个问题值得重视:

(1)可分级编码的效率还有待进一步提高,与目前可分级编码框架相适应的一些配套技术还有待进行深入研究。目前所提出的视频可分级编码技术,总体而言编码效率还不是很,究其原因方案中一些关键技术,诸如运动预测和补偿、对关键帧和帧差图像的编码技术等,大多是采用传统非可分级编码方案中的技术,适应性还不是很。特别对细粒度可分级编码的需求,导致误差漂移等影响编码效率的现象发生。

(2)基于小波的视频可分级编码技术将会得到进一步的

发展^[31]。近年来,基于小波的图像可分级编码技术得到了很好的发展,目前尽管基于小波的视频可分级编码技术已经表现出了非常强的发展势头,但其技术的发展还远不如图像可分级编码那样成熟,其理论体系尚需进一步完善。

(3)适应于无线信道环境下视频可分级编码技术还有待进一步研究。无线信道所具有的时变、带宽有限、误码率高等特性使得在无线网络上实时传输高质量的视频成为一个挑战性的工作。但随着无线信道嵌入式系统的应用,对这一技术的需求越来越迫切。在无线信道通信中,可分级编码除了要适应网络状态的动态变化外,有限的移动终端电池能量和信道的高误码率所造成的影响也是需要考虑的一个因素。

参考文献

- 1 王相海. 基于小波的图像和视频可分级编码研究:[南京大学博士后研究报告]. 2001. 6
- 2 王相海. 基于小波的视频空间可分级编码研究. 计算机科学, 2002, 29(7): 115~117
- 3 ISO/IEC 13818-2. Generic coding of moving pictures and associated audio information. MPEG(Moving Pictures Expert Group). International Organization for Standardisation 1994. (MPEG2 Video)
- 4 Fukunaga S, et al. MPEG-4 Video Verification Model Version 16.0. ISO/IEC document JTC1/ SC29/WG11 N3312, March 2000/ Noordwijkerhout
- 5 郭晓强, 门爱东. 视频编码标准的发展: 从 H. 261 到 H. 264. <http://www.ict.ac.cn>
- 6 Cote G, Erol B, Gallant M, et al. H. 263+: Video coding at low bit rates. IEEE Trans on Circuits and systems for video Technology, 1998, 8(7): 849~866
- 7 毕厚杰. 新一代视频压缩编码标准—H. 264/AVC. 北京: 人民邮电出版社, 2005
- 8 张鲲, 吴正茂, 卢官明. H. 264 标准详解. 中国多媒体视讯. <http://searchmobilecomputing.techtarget.com.cn/tips/252/1913752.shtml>
- 9 张福炎. 数字媒体的原理与应用. 南京大学计算机科学与技术系讲义, 2002. 8
- 10 Tudor P N. MPEG-2 video compression. http://www.bbc.co.uk/rd/pubs/papers/paper_14/paper_14.shtml
- 11 Wang Y, Ostermann J, Zhang Ya-Qin. 视频处理与通信. 侯正信, 杨喜, 王文全译. 电子工业出版社, 2003
- 12 王相海. MPEG-4 自然视频编码工具和算法. 计算机科学, 2001, 1: 1~7
- 13 Li W. Overview of fine granularity scalability in MPEG-4 video standard. IEEE Transactions on Circuits and Systems for Video Technology, 2001, 11(3): 301~317
- 14 van der Schaar M, Radha H, Chen Y. An all FGS solution for hybrid temporal-SNR scalability. ISO/IEC JTC1/ SC29/ WG11, MPEG99/M5552, Dec. 1999
- 15 Macnicol J, Frater M, Arnold J. Results on fine granularity scalability. Melbourne, Australia, ISO/IEC JTC1/SC29/ WG11, MPEG99/m5122, Oct. 1999
- 16 Li S, Wu F, Zhang Y Q. Study of a new approach to improve FGS video coding efficiency. ISO/IEC JTC1/ SC29/ WG11, MPEG99/M5583, Dec. 1999
- 17 Wu F, Li S, Zhang Y Q. A framework for efficient progressive fine granularity scalable video coding. IEEE Trans on Circuits and Systems for Video Technology, 2001, 11(3): 332~344
- 18 Li S, Wu F, Zhang Y Q. Experimental results with progressive fine granularity scalable (PFGS) coding. ISO/IEC JTC1/SC29/ WG11, MPEG99/ M5742, 2000
- 19 Sun X, Wu F, Li S, et al. Macroblock-based progressive fine granularity scalable video coding. In: IEEE International Conference on Multimedia and Expo(ICME), Tokyo, August 2001
- 20 孙晓艳, 高文, 吴枫, 等. 基于宏块的渐进、精细可伸缩的视频编码. 软件学报, 2002, 13(11): 2134~2141
- 21 孙晓艳, 高文, 吴枫, 等. 基于宏块的具有时域和 SNR 精细可伸缩的视频编码. 计算机学报, 2003, 26(3): 345~352
- 22 Wang Q, Wu F, Li S, et al. Fine-granularity spatially scalable video coding. In: IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Salt Lake City, 2001, 3: 1801~1804
- 23 Joint Video Team(JVT) of ISO/IEC MPEG. ITU-T VCEG: Final draft international standard of joint video specification (ITU-T Rec. H. 264/ISO/IEC 14 496-10 AVC). JVT G050, 2003
- 24 Reichel J, Schwarz H, Wien M. Scalable Video Coding - Working Draft 1, Joint Video Team(JVT), Doc JVT-N020, Hong Kong, CN, Jan. 2005
- 25 ITU-T Recommendation H. 264 & ISO/IEC 14496-10 AVC. Advanced video coding for generic audiovisual services, (version 1: 2003, version 2: 2004) version 3: 2005
- 26 Schwarz H, Marpe D, Schierl T, et al. MCTF and scalability extension of H. 264/AVC. In: Proc. Picture Coding Symposium (2004), San Francisco, CA, USA, 2004
- 27 Schwarz H, Marpe D, Schierl T, et al. Combined scalability support for the scalable extension of H. 264/AVC. In: Proc. ICME 2005, Amsterdam, The Netherlands, July 2005
- 28 Schwarz H, Marpe D, Wiegand T. Further Experiment for an MCTF extension of H. 264. ITU-T VCEG-W06, July 2004
- 29 Schwarz H, Hinz T, Marpe D, et al. Further Progress on Scalable Extension of H. 264. ITU-T VCEG, Doc VCEG-X08, October 2004
- 30 Schwarz H, Marpe D, Wiegand T. Subband Extension of H. 264/AVC. ITU-T VCEG, Doc VCEG-V04, March 2004
- 31 王相海. 图像可分级编码研究进展. 中国图象图形学报, 2006, 11(8): 1051~1061

(上接第 6 页)

- 18 Bhattacharya S, Dey S, Brglez F. Fast true delay estimation during high level synthesis. Computer-Aided Design of Integrated Circuits and Systems, 1996, 15(8): 1088~1105
- 19 Gao Youxin, Wong D F. Wire-sizing optimization with inductance consideration using transmission-line model. Computer-Aided Design of Integrated Circuits and Systems, 1999, 18(12): 1759~1767
- 20 Lakshminarayana G, Jha N K. High-Level Synthesis of Power-Optimized and Area-Optimized Circuits from Hierarchical Data-Flow Intensive Behaviors. IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems, 1999, 18(3): 265~281
- 21 Smith J, Micheli D G. Polynomial Circuit Models for Component Matching in High level synthesis. IEEE Transactions on VLSI, 2001, 9(6): 783~799
- 22 Smith J, Micheli D G. Polynomial Methods for Component Matching and Verification. In: Proc. of ICCAD, 1998. 678~685
- 23 Smith J, Micheli D G. Polynomial Methods for Allocating Complex Components. In: Proc. of DATE, 1999. 217~222
- 24 Peymandoust A, Micheli D G. Using Symbolic Algebra in Algorithmic Level DSP Synthesis. In: Proc. of DAC, 2001. 277~282
- 25 Peymandoust A, Micheli D G. Symbolic Algebra and Timing Driven Data-flow synthesis. In: Proc. of ICCAD, 2001. 300~305
- 26 Peymandoust A, Micheli D G. Application of Complex Algebra in High level Data-flow Synthesis. IEEE Transactions on CAD of IC and Systems, 2003. 22(9): 1154~1165
- 27 Hosangadi A, Fallah F, Kastner R. Energy Efficient Hardware Synthesis of Polynomial Expressions. In: Proc. of VLSI, 2005. 653~658
- 28 Hosangadi A, Fallah F, Kastner R. Factoring and Eliminating Common Subexpressions in Polynomial Expressions. In: Proc. of ICCAD, 2004. 169~174
- 29 冯刚, 马光胜, 杜振军. 动态串扰优化的开关盒布线. 半导体学报, 2005, 26(2): 399~405
- 30 冯刚, 马光胜, 杜振军. 信号相关的串扰优化详细布线. 计算机辅助设计与图形学学报, 2005, 17(5): 1074~1078
- 31 Shen V R L. A PN-based approach to the high-level synthesis of digital systems. INTEGRATION, the VLSI journal, 2006, 39: 182~204
- 32 Grun P, Eles p, Kuchcinsk K, Peng Zebo. Automatic Parallelization of a Petri Net-Based Design Representation for High-Level Synthesis. In: Proceedings of EUROMICRO-22. 185~192