

JVT 草案中的核心技术综述*

郑翔^{1,2+}, 叶志远¹, 周秉锋¹

¹(北京大学 计算机科学技术研究所,北京 100871)

²(武汉市广播电视局 总工办,湖北 武汉 430015)

An Overview of the Core Technology of JVT Draft's

ZHENG Xiang^{1,2+}, YE Zhi-Yuan¹, ZHOU Bing-Feng¹

¹(Institute of Computer Science and Technology, Peking University, Beijing 100871, China)

²(Technology Department, Wuhan Broadcast & Television Bureau, Wuhan 430015, China)

+Corresponding author: Phn: +86-10-82529988 ext 5235, E-mail: zhengxiang@founder.com, <http://www.founder.com>

Received 2003-01-17; Accepted 2003-10-14

Zheng X, Ye ZY, Zhou BF. An overview of the core technology of JVT draft's. *Journal of Software*, 2004, 15(1):58~68.

<http://www.jos.org.cn/1000-9825/15/58.htm>

Abstract: As one stream standard which will be produced by the ITU-T & ISO in the future, Draft JVT (joint video team) is a new video-audio coding standard. Its goal is to produce a significant compression performance relative to all the existing video coding standards in bit saving, procession quality and compression efficiency by some measures which include slice technology, high resolution variable pixels, different block sizes and shapes, intra/inter prediction and coding from multiple reference pictures, etc. On the other hand, JVT would be designed to offer a transparent translation over different networks and increase network friendliness by information encapsulation and precedence control technologies. Although the basic coding framework of the standard is similar to that of currently popular video standards, JVT includes many new coding features which are presented in the paper. First, a broad view on video-audio coding standards is given. Second, the structure of JVT is presented. Then, both the detailed technologies which improve on coding efficiency of the VCL (video coding layer) and the key terms which are used in NAL (network abstraction layer) are introduced, which is the core in the paper. Finally, the current problems and the challenges of JVT's future research are discussed and analyzed.

Key words: JVT (joint video team) draft; VCL (video coding layer); NAL (network abstraction layer); codec; bit rate; network friendliness

摘要: 作为一种由ITU-T国际视频组织和ISO的MPEG小组联合推出的未来视频处理标准,JVT(joint video team)的主要功能是改进当前压缩标准,通过采用切图技术、高分辨率的亚像素技术以及不同尺寸的块和形状与

* Supported by the National PostDoctor Foundation of China under Grant No.2003033077 (中国博士后研究基金)

作者简介: 郑翔(1968—),男,湖北枣阳人,博士,主要研究领域为流媒体技术,计算机网络;叶志远(1963—),男,高级工程师,主要研究领域为流媒体,信息安全技术;周秉锋(1963—),男,博士,研究员,博士生导师,主要研究领域为多媒体信息处理,网络计算,计算机图形学,图像处理。

使用多个参考图进行 Intra/Inter 预测编码等技术来降低码率,提高压缩质量和压缩效率.同时在网络层通过对信息的封装与优先级控制等技术来实现数据在不同网络之间的透明传输与网络友好.重点介绍了 JVT 草案中的核心技术,包括 JVT 体系结构、层次模型以及与 VCL(video coding layer)和 NAL(network abstraction layer)两层相关具有代表性的关键技术,分析和探讨了当前 JVT 存在的问题和进一步的研究方向.

关键词: JCL 草案;视频编码层;网络抽象层;编解码器;比特率;网络友好

中图法分类号: TP391 文献标识码: A

目前主要有两个制定视频编码标准的国际组织:ITU-T 和 ISO/IEC.ITU-T 的标准称为建议,以字母排序,视频会议电视编码的标准在 H 的子集里,如 H.261,H.262 和 H.263.ISO/IEC 的标准按序号排列,如 MPEG-1 相对应的是 11172,MPEG-2 相对应的是 13818.MPEG-4 相对应的是 14496 等.ITU-T 的建议标准主要用于实时视频通信,如视频会议、可视电话等.而 MPEG 标准主要用于广播电视、DVD 和视频流媒体.大多数情况下,这两个标准组织独立制定不同的标准,但在许多方面也有共同之处,例如 H.262 标准和 MPEG-2 的视频编码标准基本上就是同一个标准.

自 1997 年以来,ITU-T 的视频编码专家组 VCEG(video coding experts group)一直致力于研究一种新的视频编码标准——H.26L^[1].H.26L 是一种高效的压缩方法,从核心技术特征角度上讲,与 H.263 v2(H.263+)^[2]或 MPEG-4 Simple Profile 标准^[3]相比,H.26L 具有如下特性:① 在使用多数编码方法类似的最佳编码器时,可节省超过 10%的码率.② 对包括低比特率在内的所有比特率,H.26L 都能持续提供较高的视频质量和基于不同的约束条件来完成不同性质的任务.其中低延时模式包括实时通信的视频会议等,没有延时限制的应用包括视频存储和以服务器为基础的视频流式应用等.③ 强大的容错处理能力.H.26L 不仅提供在包传输网中处理包丢失所需的工具,而且在易产生误码的无线网中处理比特误码的工具.④ 网络友好.它将网络层和编码层从概念上进行分离,便于对数据进行打包处理和优先控制好信息,实现所编码的视频流在当前所有的协议和复合结构中能够无缝播放并容易组合.在 2001 年末,ISO 的运动图像编解码组 MPEG(moving picture encoder group)中的视频编解码组和 VCEG 决定在当前的 H.26L 编码草案基础上,联合制定一种新的视频标准——JVT(joint video team)标准(也可以写为 JVT/H.26L)^[4].图 1 概括了 ITU-T 和 ISO-MPEG 标准的研究发展阶段和不同标准之间的关系.

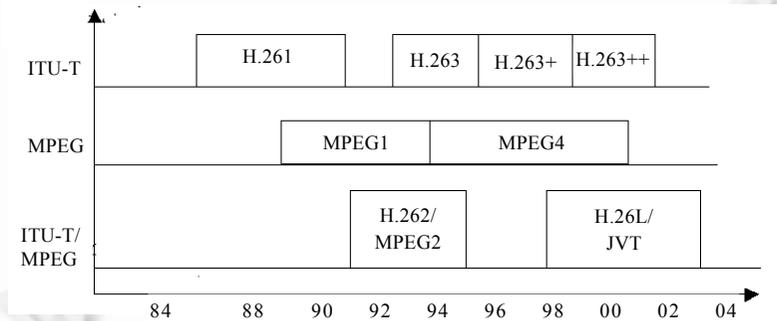


Fig.1 The history of ITU-T and MPEG
图 1 ITU-T 和 ISO-MPEG 标准发展史

作为一个代表未来视频流技术发展方向的国际标准,JVT 标准的作用是对“对话方式”(视频会议)或“非对话方式”(存储、广播或流)等应用提供一个简单和直接的视频编码规范来增强压缩的效率并提供网络友好的视频表达.它可以作为 ITU-T 未来所推荐 H.26L 的重要组成部分(例如 H.264)和 MPEG-4 的一个新的组成部分(例如,Part10).这是由于 JVT 视频编码方案主要由下面两个部分组成:研究编码效率的视频编码层 VCL(video coding layer)和支持视频在不同网络之间透明传输的网络抽象层 NAL(network abstraction layer,在一些文献(例如文献[5])中使用 Network Abstraction Layer,而在另一些文献(例如文献[10])中使用 Network Adaptation Layer 来表示 NAL.在这里,我们根据草案文献使用 Network Abstraction Layer 来表示 NAL).其中,VCL 主要集中于研

究编码效率,是视频内容的核心压缩内容的表述;NAL 是对特定网络类型传输数据的打包表述,它提供在多种网络上无缝传输图像数据的工具.这样的结构便于信息的封装和对信息进行更好的优先级控制.

从技术上讲,JVT 是基于现有的技术标准(包括 H.263 和 MPEG4)之上的一个新标准,其主要特性包含了 H.26L 的重要组成部分,对不同的视频应用程序而言这些特征可以转换成不同的优势.本文主要研究 JVT 的系统结构以及与组成 JVT 标准的 VCL 和 NAL 的核心技术,并描述了 JVT 的应用前景.

1 JVT 的系统组织结构

在 JVT 标准中,VCL 主要研究基于常规的运动补偿、系数变换编码、熵编码等编码技术来提高视频信号编码效率,其使用最基本的表示法是切片(slice)技术,它将一个待处理的图像分割成许多宏块来进行处理.一般而言,VCL 主要包括 4 个重要方面的技术:

- ① 降低对视频帧的处理时间:通过将每个视频帧划分成由点阵组成的块,将对视频帧的处理降到对块的处理.
- ② 降低视频帧在空间的冗余量:通过变换、量化和熵编码(或者变长编码)方式对原始块进行编码.
- ③ 在有效帧的宏块之间进行时间独立性分析,采用运动估计和补偿的技术只对那些在有效帧之间变化的数据进行编码.
- ④ 通过对剩余块的编码来实现对视频帧中所存在和保留的任何空间冗余进行利用.

除了研究提高编码效率和简化语法形式之外,JVT 的另一个主要设计目的是支持“网络友好”.由于网络传输协议在可信度、服务品质保证、数据封装和时间调度等方面都具有异构性,同时,不同的传输系统提供的内在建立和配置协议也有区别.为了实现“网络友好”和无缝传输 VCL 数据,NAL 将 VCL 从具体的传输层中抽象出来,定义了一个基本的和独立于网络的表示法.虽然 JVT 标准中并没有提供将 JVT 编码数据封装传输协议的技术,但是,NAL 概念对将 JVT 编码数据组合到现在的网络中提供了更高的灵活性,同时,它也提供了对不同传输层之间的网关设计提供了足够的公共基础.其特点主要表现在如下几个方面:

- ① NAL 对 JVT 标准所传输的所有信息提供了一个抽象的描述.
- ② 它通过合适的语法规则来完成支持 JVT 的各种必须的网络功能(例如,H.320 系统没有提供封装机制,而基于 IP 的网络系统完全是基于包技术等).
- ③ 传输系统可以使用不同 QoS 在不同的逻辑通道中传输不同的 NAL 单元类型.所使用的配置协议也是相关的(例如,在基于 IP 应用程序中使用的 SIP/SDP 协议以及在 H.324 中使用的 H.245 协议等).

作为 JVT 标准的组成部分,VCL 和 NAL 之间的关系非常密切.例如:VCL 和 NAL 都与媒质有关,它们都知道网络低层的属性和约束条件,包括包丢失率、MTU 长度、传输延迟瓶颈等;同时,VCL 和 NAL 都具有自适应特性,特别是 VCL 能够自动通过调整宏块内的速率和切图的尺寸来响应错误恢复.但是 VCL 和 NAL 之间还存在很大的差异,这主要表现在 JVT 编/解码器^[5]的设计在 VCL 和 NAL 两个不同层的意义是完全不同的,不同层的接口需要使用不同的传输协议.在 JVT 编码标准中定义了 VCL 和 NAL 编/解码器的接口标准,并从概念上对 VCL 和 NAL 之间的接口进行了合适的区分,它有助于描述 VCL 和 NAL 不同的任务并且从结构上将它们区分开来.更进一步讲,在不同的传输系统(例如 H.320,MPEG-2 系统以及 RTP/IP 等)中对 JVT 数据的封装和传输都不是 JVT 标准的组成部分.图 2 描述了在传输环境中的 JVT 编/解码工作流程.

为了支持上述的系统结构,JVT 规定了一种类似于 MPEG-4 文件格式^[6]的一种新的文件格式(关于 MPEG-4 的文件格式的深入分析,我们在另一篇文章中将重点加以介绍).一个 JVT 文件是由‘clump’组成,它的功能与符合 MPEG-4 标准的文件(.mp4)格式中的‘box’以及 QuickTime 中的对象功能相类似.

在 JVT 文件规范中通常包括下面的内容:

- ① 采用 ISO/IEC 提供的规范描述语言 SDL(syntactic description language)^[7]来定义文件格式.
- ② 一个‘clump’可以包含其他的‘clump’.
- ③ 多数‘clump’可以有自己的成员属性.
- ④ 一些‘clump’需要按顺序包含索引值.

- ⑤ 通常采用网络字节顺序或‘big-endian’格式等大字节来保存‘clump’中的属性值。
 - ⑥ 采用类似 MPEG-4 规范文件格式中的轨道(‘trak’)技术来实现对相应‘clump’或媒体数据的调用。
- 关于 JVT 具体的文件格式描述可参阅文献[8].

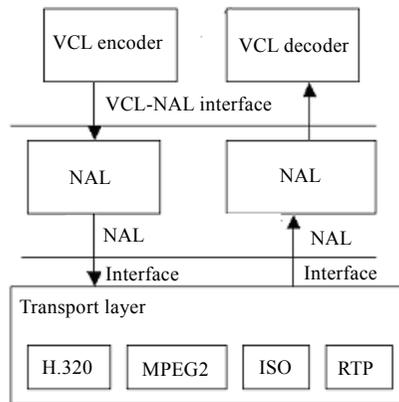


Fig.2 JVT encoding/decoding in network transport environment
图 2 在网络传输环境中的 JVT 编码

为了管理大量的编码工具和适应支持灵活的格式与广泛的比特率的需要,与 MPEG-4 类似^[9],JVT 也使用 profile 和 level 来定义一些一致点(conformance point)集合.其中一致点是指在各种应用程序中体现的功能是非常相近的.Profile 定义用来产生一个合适比特流的编码工具或算法工具集;Level 是对比特流中某些关键参数定义约束条件.在这里,Profile 主要包括 Baseline,Main 和 X 这 3 种工具,用于表达适用于不同应用环境的不同工具集.Baseline Profile 由低复杂度和低隐含技术特征所组成,主要适用于交互应用程序,特别适合容易出现错误的环境中;Baseline Profile 支持的工可以适用于所有的 Profile;Main Profile 主要用于那些需要高效编码的环境中,例如广播应用程序;X Profile 主要用于流媒体,包括在比特流之间进行灵活的错误隐藏和错误恢复技术.表 1 是 JVT 对 Profile 的定义,详细描述参见文献[10].

Table 1 Profile definition

表 1 Profile 定义

| Coding tools | | Baseline profile | Main profile |
|-----------------------|-------------------------------------|---------------------|---------------------|
| Picture formats | Progressive pictures | X | X |
| | Interlaced pictures | Level 2.1 and above | Level 2.1 and above |
| Slice/Picture types | I and P coding types | X | X |
| | B coding types | | X |
| Macroblock prediction | SI and SP coding types | | |
| | Tree-Structured motion compensation | X | X |
| | Intra blocks on 8x8 basis | | |
| Motion compensation | Multi-Picture | X | X |
| | 1/4-pel accurate | X | X |
| | 1/8-pel accurate | | |
| Transform coding | Adaptive block size transform | | X |
| Entropy coding | VLC-Based entropy coding | X | X |
| | CABAC | | X |
| In-Loop filtering | In-Loop deblocking filter | X | X |

2 VCL 技术

VCL 技术研究的主要目的是希望基于兼容当前视频编码的流行标准 H.263 和 MPEG-4 技术找到一种新的标准,用于取代现有任何实现高视频质量的视频编码标准.与以前的编码方案相类似,VCL 先将每个视频帧分成像素块,实现以像素块为单位进行处理视频帧方式;接着通过对一些没有帧间相关性的像素块进行变换、量化和熵编码(即变字长编码),以去除其空间相关性;然后使用运动估值和运动补偿,去除存在于相邻帧相应像素块

之间时间方向的相关性,使得只有相邻帧间两个相应像素块的变化部分才需要编码.一般而言,VCL 与其他的编码标准的差异主要表现在:① VCL 使用不同尺寸的块和形状,高分辨率的亚像素运动估计和选择多个参考帧来实现运动估计和运动补偿技术.② 与其他标准采用的 DCT 变换不同的是,VCL 采用基于整数的变换,其优势在于反变换时不存在误匹配的问题.③ VCL 中采用的熵编码技术包括一致变长编码 UVLC(universal variable length codes)^[11]或基于内容匹配的自适应二进制算术编码 CABAC(context-based adaptive binary arithmetic coding)^[12].

图 3 给出 VCL 编/解码器的主要工作原理.它主要包括 Intra/Inter 预测和编码、变换、量化处理和与熵编码等关键技术(实际上,我们可以利用运动估计和补偿技术对视频流 I/P/B 帧处理的关系来理解它).从编码效率来讲,在这些关键技术中,使用 7 种不同的尺寸和形状的帧间预测可以节省 15%的码率;使用亚像素空间精度比使用整数像素空间精度可以节省 20%的码率;使用 5 个参考帧进行预测比只使用一个参考帧可以节省 5%~10%的码率;使用基于内容的二进制自适应算术编码可以节省 10%的码率.下面我们简单叙述这些关键技术^[13~16].

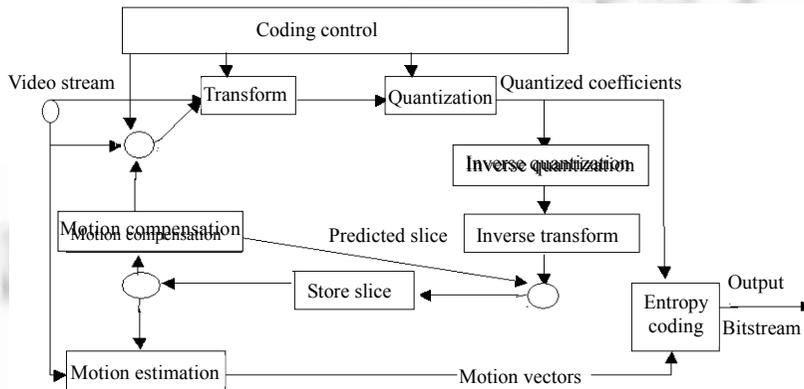


Fig.3 The mechanism of VCL encoder and decoder

图 3 VCL 编/解码器工作原理

2.1 整数变换

虽然 VCL 采用的变换类似于 DCT 的变换,但是它使用的是 4×4 的整数块,而 DCT 使用的是 8×8 的浮点块,这样,由于它使用的是以整数为基础的空间变换,因此其反变换不存在取舍误差的问题,同时能够解决编码器和使用反变换的解码器之间的误匹配问题.此外,采用小的形状块有助于降低块效应和明显的人工处理痕迹.

2.2 量化处理

量化步长是对宏块数据压缩的一个重要组成部分.类似于 H.263 使用 31 个不同的量化步长,VCL 提供了 32 个不同的量化步长,此外,VCL 使用非固定宽度的尺度量化方法对变换系数进行量化,这些步长的增加按 12.5% 的混合速率增加,通过使用精确的量化步长,更能改进色度部分的精度.

量化变换的系数对应不同的频率,一个对应 DC 值,其余的分别对应不同的频率值,VCL 将所有的变换系数放在一个数组中,通过之字形(zigzag)扫描和双扫描对数组中的数据进行读取.双扫描只用于使用较小量化级的块内,由于它的步长小,从而有助于提高编码效率.

2.3 Intra 预测和编码

Intra 编码研究利用一个视频图像中的空间冗余来提高编码效率的问题.基于 16×16 像素宏块的 MPEG-4 Intra 预测编码技术^[17]所不同的是,VCL 的 Intra 预测技术是利用相邻像素块的相关性,先将 QCIF(176×144)格式的图像分成 99 个 16×16 宏块,然后分成 4×4 像素块执行预测编码.对每个 4×4 像素块,提供 6 种预测 4×4 相邻亮度块的模式,其中模式 0 是直流 DC 预测模式和模式 1~5 是邻近像素的方向预测模式.同时,VCL 能够基于对块周围编码模式的检测规则找到最常用的模式,然后通过对这些模式采用最短的符号表示,这种方式可以提高

对每个块的预测模式编码效率。

2.4 Inter预测和编码

Inter 预测编码是使用运动估计和运动补偿技术,通过利用有效帧之间的时间冗余来提高编码效率。除了支持其他的视频标准通常使用的 I-帧、P-帧和 B-帧以外,VCL 还支持一种新的码流间可转换帧——SP 帧。通过使用一个比特流中的 SP 帧,VCL 可以在那些有类似内容但有不同码率的码流之间快速切换,并同时支持随机访问和快速回放模式。此外,VCL 所支持的运动估计策略具有更多的关键特征,能够增加帧间预测模式的数量和精度,以及通过增加运动估计的灵活性和功能来改进运动估计的效率。具体内容包括:① VCL 采用不同尺寸和形状的宏块分割方法形成 7 种不同的帧间预测模式,它比使用一个 16×16 的块尺寸可以降低一定的比特率和改进预测效果,特别是能够处理运动细节和视觉质量。② 在使用运动补偿算法中,VCL 采用 $1/4$ 或 $1/8$ 点亚像素。与使用原始像素和 $1/2$ 亚像素的标准以及使用 $1/4$ 亚像素的 MPEG-4 标准相比,这样可以降低一定的比特率,运动估值精度也比使用整数点的精度要高,从而能够更有效地增加编码的效率并提供高质量的视频分辨率。③ 对于任何给定的像素块,VCL 在预测中使用以前的 5 幅不同的参考帧去寻找匹配像素块,并给出相应的运动矢量作为编码和解码的参考像素块。与以前只使用一幅参考帧的编码方法相比,这样可以更有效地提高视频质量和实现更好的码流误码恢复。同时,降低比特率。当然,采用这种方法要求系统对编/解码器提供更高的处理能力并增加额外的延时和存储容量。④ VCL 在预测循环中对块边缘使用基于 4×4 的自适应 Deblock 过滤器的目的是,删除由于块预测错误引起的块效应,减少方块效应,而且有助于恢复真实数据。但是该算法非常复杂。

2.5 熵编码

视频编码的最后一步是熵编码。其中,基于 VLC 的熵编码是一个应用最广的方法,它主要应用在对量化变换系数、自适应块变换 ABT(adaptive block transform)、运动向量和其他编码信息的压缩技术中;而基于内容的自适应二进制算术编码主要是使编码和解码两边都能使用所有句法元素(变换系数、运动矢量)的概率模型。在具体应用上,VCL 使用下面 UVLC 和 CABAC 两种编码方式:① 在使用 UVLC 编码中,无论符号表述什么类型的数据,VCL 都使用统一变长编码表。其优点是简单,缺点是由于码表从概率统计分布模型得出,因此没有考虑编码符号之间的相关性,在中高码率时效果不是很好。与 H.263 等标准中根据要编码的数据类型(如变换系数和运动矢量等)采用不同的 VLC 码表技术相比,使用 VLC 的 UVLC 可以节约一定的码率。此外,文献[11]介绍了一种针对 ABT 支持 3D UVLC 的编码模型,它可以比 2D 的编码模型有更低的熵,从而能够节约更多的码率。② 在使用 CABAC 编码中,为了提高算术编码的效率,VCL 通过一种称为内容建模的过程,使基本概率模型能够适应随视频帧而改变的统计特性。内容建模提供了编码符号的条件概率估计,利用合适的内容模型,存在于符号间的相关性可以通过选择目前要编码符号邻近的已编码符号的相应概率模型来去除,不同的句法元素通常保持不同的模型。对于给定的符号为非二进制值,VCL 将使用 UVLC 二进制树来使模型较好地保持对实际统计特性的跟踪。

2.6 Profile技术

事实上,一个 Baseline 解码器的处理对象主要是 intra(I)切图,它只包含 intra 帧宏块,而 inter(P)切图是由时间预测宏块和 intra 帧宏块组成,为了对运动视频中的图像预测更加准确,VCL 技术通常将一个普通解码器使用的宏块在一维或二维中将它们切为二等分,采用 $1/4$ 亚像素和 7 种不同的帧间预测模式的运动估计技术在预测中依赖所选择的 level,最多可使用到 15 个不同的参考帧,需要指出的是,在 Baseline 技术中,允许切图以任意顺序 ASO(arbitrary slice order)到达解码器,并通过 FMO(flexible macroblock ordering)策略来决定切图位置映射关系,同时由于在切图之间没有存在空间预测依赖关系,该技术通过允许一个切图的数据在比特流中可以重复出现(我们将这些冗余切图称为 RS(redundant slice))的策略来避免错误数据的传播,从而增强比特流的容错处理能力。与此对应,一个 Main 解码器除了具备 Baseline 解码器的不包括 FMO,ASO 和 RS 之外的所有算法之外,还增加了一些重要的特性,它主要用于处理任意数量的 B 切图信息(介于 I 和 P 切图之间),所采用的重要方法是自适应块变换 ABT 技术,它可以增强处理结果的信噪比 PSNR。文献[16]通过对 CCIR 测试序列 football 使用

CABAC 编码对此技术的性能进行了比较(如图 4 所示)。

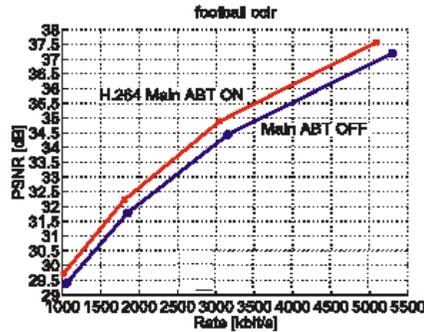


Fig.4 PSNR of ABT and non-ABT coding in main profile on the test sequence football, CCIR 601 format

图4 在 Main profile 中针对测试序列 football,CCIR601 格式采用 ABT 前后的 PSNR

2.7 JVT与JPEG/JPEG2000/MJ2技术比较

文献[13,16]分别基于测试文件 Container(QCIF),News(QCIF),Tempete(CIF)和 Mobile(CCIR601)给出了 JVT 与 JPEG/JPEG2000/MJPEG 的性能比较结果,文献[15]给出了 AVC 与 MPEG-4 编解码器在运行中所占用的系统资源和处理时间与效果.受篇幅限制,我们这里仅给出文献[16]中 Container(QCIF)和 Mobile(CCIR601)的实验结果(其中 H.264 是 JVT 视频压缩内容部分的标准),如图 5 所示.

| Sequence (Size) | Av.Q (JPEG) | QP (H.264) | Bit rate (av.) | PSNR | | | Difference JM to | | PSNR | | Difference |
|-------------------|----------------|---------------|-------------------|-------|----------|-------|------------------|----------|-------|-------|------------|
| | | | | JPEG | JPEG2000 | H.264 | JPEG | JPEG2000 | MJ2 | H.264 | |
| Container (QCIF) | 82 | 20 | 1.715 | 37.26 | 43.28 | 43.28 | 6.02 | 0.87 | 48.24 | 46.15 | -2.09 |
| | 43 | 28 | 0.892 | 31.64 | 35.63 | 37.24 | 5.60 | 1.61 | 41.53 | 40.52 | -1.01 |
| | 7 | 36 | 0.405 | 25.26 | 29.28 | 31.55 | 6.29 | 2.27 | 37.32 | 36.63 | -0.69 |
| | 0 | 44 | 0.153 | 18.42 | 23.68 | 26.05 | 7.63 | 2.37 | 33.87 | 34.81 | 0.94 |
| Mobile (CCIR 601) | 86 | 20 | 2.632 | 36.16 | 42.05 | 42.68 | 6.52 | 0.63 | 46.34 | 43.27 | -3.07 |
| | 63 | 28 | 1.530 | 30.21 | 34.94 | 36.02 | 5.81 | 1.08 | 39.94 | 37.51 | -2.43 |
| | 21 | 36 | 0.724 | 25.66 | 28.19 | 29.53 | 3.87 | 1.34 | 35.47 | 33.73 | -1.74 |
| | 4 | 44 | 0.253 | 21.01 | 23.03 | 23.96 | 2.95 | 0.93 | 32.11 | 31.17 | -0.94 |
| Total | | | | | | | 5.29 | 1.12 | | | -1.42 |

I picture coding performance chrominance coding performance.

Fig.5 Comparison JVT intra coding to JPEG-2000, JPEG, MJ2

图5 JVT 与 JPEG2000/JPEG/MJ2 技术应用结果比较

从上述实验结果中我们可以看出,在对图像处理过程中,在 intra 模式下采用 JVT 技术得到的效果非常好.由于 JPEG2000 是使用小波变换对整个切图信息进行系数扫描、量化和熵编码,而 JPEG 是依次使用 8×8 整数 DCT 变换进行系数扫描、量化和熵编码,同时 JPEG2000 和 JPEG 都没有使用 JVT 所采用的 intra 帧预测模式,因此在任何情况下(包括图像的尺寸和比特率),采用 JVT 技术都比 JPEG 技术所得到的结果要好得多,平均差值达到 5.29dB.同时,采用 JVT 和 JPEG2000 两个标准分别得到的处理结果之间的差异随着图像尺寸的降低而变得非常小,特别是在小图像和低比特率(对应于高量化参数 QP(quantization parameter))的条件下,采用 JVT 技术得到的处理效果要优越于采用 JPEG2000 得到的处理结果.在运动系统中,从整体上看,MJ2 对色度的处理结果要强于 JVT 对色度的处理.可能的原因是,JVT 对色度的处理比对亮度的处理其精确度要低,例如 2×2 的整数变换,未来 JVT 标准将可能通过对颜色进行补偿和对色度进行高精度采样等技术来改进处理效果.此外,需要指出的是,由于 JPEG2000 采用非预测的小波分解、尺度量化和算术编码等技术,因此对大图像尺寸的处理结果可能出现一些变形的效果(参见文献[16]中的图 5).

3 NAL 技术

NAL 用于定义适合传输层或存储介质需要的数据格式,同时提供头信息,从而提供视频编码与外部世界的接口.它是以 NALU(NAL unit)^[18]为单元来支持编码数据在大多数基于包交换技术网络中的传输的.每个单元都

包含一个整数的字节数.一个 NAL 单元定义了可用于基于包和基于比特流系统的基本格式.区别这两种格式的标准在于每个比特流传输层都有一个起始代码.在 NAL 解码器接口,它假定按传输顺序传递 NALU,同时,如果在有效载荷中包含位错误,则在 NALU 的头部设置标识接收正确的、丢失的或错误的位数据的标识位.如果在对应的 NALU 检测到传输错误,那么任何与媒体有关的网络元都可以改变这个标志.因此,通过网络可以传输带有错误的有效载荷,解码器和网关都可以决定是否解码或处理那些错误的 NALU.

3.1 NAL单元的基本语法结构

NALU 由一个字节的头部和一个位字符串组成.后面的位字符串代表切图的宏块.头部字节由一个前面所提到的错误标志、一个处理的 NALU 标志和 NALU 类型组成.处理标志用于标识出 NAL 有效载荷的内容是否属于一个没有存储在多图像缓冲中的图,以及允许服务器、网络和网关在没有介绍错误属性的情况下丢弃 NALU.NAL 有效载荷类型包括指出 VCL 是切图类型(I 切图、P 切图或 B 切图),或者随机访问点、参数设置信息、附加的增强信息等.表 2 是一个 NAL 单元的语法,其中 EBSP 是 Encapsulated Byte Sequence Payload 的省略写法.

Table 2 An NALU syntax
表 2 一个 NAL 单元的基本语法结构

| nal_unit (EBSPsize){ | Descriptor | Nodes |
|--------------------------------|---|-------|
| Error_Flag | Signals whether a bit error exists in the NALU | u(1) |
| Nal_Unit_Type | Indicates the type of element contained in the NALU | u(5) |
| Picture_Header_Flag | Indicates that the EBSP contains a picture header that precedes any other data in the EBSP | u(1) |
| Disposable_Flag | Signals whether the content of the EBSP belongs to a picture that is not stored in the multi-picture buffer | u(1) |
| For ($i=0; i<EBSPsize; i++$) | | |
| ebsp[i] | Contains the NALU data in the encapsulated byte sequence payload (EBSP) format | b(8) |
| } | | |

3.2 比特流技术

使用包的网路系统可以通过将 NALU 映射为 RTP 包^[19]作为 H.223 AL3 SDU 或 RTP 包的 payload^[20]直接使用 NALU,而在一些基于流媒体的系统中(例如,针对视频会议的 H.320 或适用于数字电视的 MPEG-2 系统以及 MPEG-4 系统)要求采用比特/字节流格式^[21].因此,JVT 定义了一种将 NALU 转换成这种流格式的变换机制:使用符合常规视频编码标准的头编码(start code)将 NALU 封装起来,头编码是 16 或 24 位,其长度依赖于 NALU 的 payload 重要性,此外,头编码必须出现在字节对齐的位置.解码器扫描头编码,然后通过一种基于字节的内存复制操作提取 NALU.

为了解决在某些情况下头编码仿真(头编码与字节流格式一致)的问题,多数视频编码标准都采用严格的熵编码设计标准.由于 JVT 标准中包含两种不同的熵编码模式,很难实现对头编码仿真的自由处理,因此,JVT 采用字节填充机制,即在出现头编码仿真的位置采用字节填充机制将一些非零字节插入到 NALU 比特流中.为了实现网关设计,甚至在一些不必要的环境中也使用该机制,特别是在包环境中,注意,VCL-NAL 接口只是一个概念,可以将避免头编码仿真的实现技术作为 VCL 熵编码中的一个组成部分.

3.3 复合包

此外,JVT 中涉及到另一个概念:复合包^[21].它通过在每个 NALU 的后面增加一个 16 位的信息区域的方式将多个 NALU 放到一个 RTP 包中.该技术主要用在具有不同 MTU 尺寸的网关设计中,这样它可以节省在 IP/UDP/RTP 上包头的比特数.现在,JVT 组织正在研究下面两种形式的复合包:一种形式是复合具有相同 RTP 时间戳(属于同一个帧)的包;另一种是增加一个包含管理信息的整数,用于对 NALU 中 RTP 时间戳的差异进行编码.对媒体流而言,特别是低比特率环境中,一般需要将多帧的信息放在一个包中,由于 JVT 的比特流不带有自己的时间信息,这就需要使使用 RTP 的时间戳来处理这样的信息.

3.4 参数集概念

在包丢失环境中,解决视频传输的关键问题之一起源于当前所有视频编码标准的层次特征.切图/图像/GOP/序列的头信息对重建整个切图/图像/GOP/序列来讲是必需的.但是,为了降低比特数,常规编码只处理一次切图/图像/GOP/序列的头信息,这样,包含头部信息包的丢失将会导致依赖该头信息的数据不能被使用.为了解决这个问题,特别是在包丢失率超过 5%的环境中,通常使用的包括 MPEG-4 HEC,H.263 Annex W 或 RFC2429(RTP 专用的 payload)在内的视频编码技术都采用冗余编码技术.冗余编码技术增加了比特率,在有些情况下,所增加的比特数使容错视频缺乏实用性,同时也增加了设计的复杂度.所有这些技术的一个主要缺陷是无法预测解码器是否能够访问头部数据.在每个传输包中重复使用整个头部信息将会大量增加比特数.

JVT 提供了一种新的解决方案:同步和实时媒体传输所使用的完整的包是一个完全自包含的数据包.也就是说,每个包在重建时都可以不需要依赖其他包的信息.由于我们可以将切图的尺寸调整到符合大多数系统所支持的 MTU,因此,切图层是一个最小的自包含单元(如果没有采用数据重复机制).文献[21]详细描述了参数集概念(如图 6 所示).

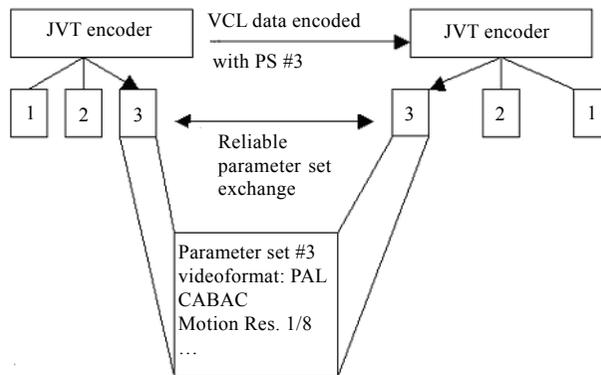


Fig.6 Parameter set

图 6 参数集

在高层的所有信息都是异步传输,JVT 将经常改变的参数增加到切图层,其余的参数放在参数集中,编/解码器共同维护着同步的参数集.VCL 数据的切图头部定义了解码器所使用的用于重建编码切图的参数集.JVT 标准里面定义了一个参数集信息包,它可以按照一种指定的 NAL 单元类型转换参数集,其对应语法正在修订中.不过,参数集的概念允许使用不同的逻辑通道或硬件流控制协议(out_of_band)将参数集从编码器传输到解码器,软件流(in_band)参数集信息一般不和硬件流控制协议组合在一起使用.包含在参数集中的高级信息一般只与系统有关,而与内容无关.因此,它非常适合网络传输.在引用参数集的同步视频流之前,都可以在进程配置或异步进程期间可靠传输每个参数集.

3.5 基于IP的实时流媒体实现技术

虽然 H.261,MPEG-2 和 H.263 等视频编码标准可以使用不同的传输协议,但是它们主要针对专用的应用程序和传输协议来设计,并且用于电路交换和基于比特流的环境^[22].JVT 在视频编/解码方面的研究从一开始针对的就是基于包的网路,除了 IP 网路之外,它还考虑到在第三代移动系统中基于 IP 的视频传输^[23],所针对的应用程序是流和常规的服务,包括视频电话和视频会议.

基于 IP 的实时流媒体实现方法是在网路层使用 IP 协议、传输层上使用 UDP 协议,同时在应用层上使用 RTP 和与 RTP 同时使用的 payload 规范.其中 RTP 的 payload 类型包括了 payload 使用的媒体编码信息.类似于 H.263 的 RFC2429 和 MPEG-4 的 RFC3016,payload 规范描述了打包和解包进程的全部操作(包括将一个比特流映射到信息包边界的分割规则)、时间戳和一些必要的头冗余编码技术.JVT 的一个重要功能是设计一种简单编码方案,通过将该编码的封装数据以一种非常简单的方式在网路中传输.根据参数集概念,由于在切图之上没

有头部数据,因此 JVT 在将 NALU 直接作为 RTP payload 进行传输时,不需要研究冗余的头数据的编码技术。

JVT 通过采用一些对基于 IP 的固定或移动传输的公共测试条件来接受基于 IP 的固定或移动传输思想.这些测试条件允许选择合适的编码特征、测试和评估容错特征并产生可预测的结果.公共定义的测试示例包括固定的 Internet 常规服务:包交换常规服务和 3G 移动网络的包交换流服务,也包括一些离线的网络仿真软件:在理想传输条件下捕获错误、特定的视频序列、合适的比特速率和评估标准等。

4 总 结

作为一个代表未来视频流技术发展方向的国际标准,JVT 可以广泛地应用在不同码率的环境中,它的 VCL 层所提供的编码策略能够有效地降低比特率,提高对数据的压缩效率;它的 NAL 层提出的复合包和参数集等概念有助于在不同的网络环境中采用不同的数据传输方式实现“网络友好”,具有非常广阔的应用前景.当然,作为一种草案,本研究也指出在 JVT 标准中目前还存在许多有待解决的问题,如 Deblock 过滤器该算法复杂度过高、在中高码率时如何改进统一变长子长编码表、参数集的完整语法结构等等.这些问题的研究将有助于对 JVT 标准的补充和完善,关于 JVT 草案详细征集的内容和重点研究的范围,有兴趣的研究人员可参见文献[24].

致谢 在此,我们向对本文的工作给予支持和建议的同行,尤其是北京大学计算机科学技术研究所的老师和同学以及方正研究院多媒体组的同事表示感谢。

References:

- [1] ITU-T H.26L Standardisation (ITU-T Q6/16, VCEG). 2002. <http://www.itu.int/ITU-T/news/jvtpro.html>
- [2] Hwang JS, Kuo TY, Shen MY, Kim JW, C.-C Jay K. Enhancement of H.263+ video compression. 2002. <http://viola.usc.edu/extranet/Projects/network-h26x/default.htm>
- [3] Koenen R. Profiles and levels in MPEG-4: approach and overview. 2002. http://mpeg.telecomitalia.com/events&tutorials/mpeg-4_si/11-Profiles_paper/11-Profiles_paper.htm
- [4] VCEG requirements for proposed JVT codec. 2001. http://standards.pictel.com/ftp/video-site/0112_Pat/JVT-A004.doc
- [5] Wiegand T. Joint video team (JVT) of ISO/IEC MPEG and ITU-T VCEG. Joint Committee Draft (CD), JVT-C167, 2002.
- [6] Singer D, Belknap W, Franceschini G. ISO media file format specification. ISO/IEC JTC1/SC29/WG11 MPEG01/N4270-1, 2001.
- [7] Specification and description language (SDL). ITU-T Z.100, 2002.
- [8] Hannuksela MM. H.26L file format. ITU—Telecommunications Standardization Sector, VCEG-O44, 2001.
- [9] Luthra A, Gandhi R, Panusopone K. Performance of MPEG-4 profiles used for streaming video. In: MPEG-4 Proc. of the Workshop and Exhibition. San Jose, 2001. 103~106.
- [10] Schwarz H, Wiegand T. The emerging JVTH.26L video coding standard. In: IEEE Int'l. Conf. on Image Processing. NY, ICIP 2002.9. <http://www.icip2002.com/Tutorial1.asp>
- [11] Kerofsky L. Core experiment on adaptive block transform. ITU—Telecommunications Standardization Sector, VCEG-M68.
- [12] Marpe D, Blättermann G. Video compression using context-based adaptive binary arithmetic coding. 2002. <http://bs.hhi.de/~marpe/cabac.html>
- [13] Halbach T. Performance comparison: H.26L intra coding vs. JPEG2000. Joint Video Team (JVT) of ISO/IEC MPEG & ITU-T VCEG, JVT-D039. 2002.
- [14] Emerging H.26L Standard: Overview and TMS320C64x Digital Media Platform Implementation White Paper. 2002. http://www.ubvideo.com/public/h26l-white_paper.pdf
- [15] Saponara S, Blanch C, Denolf K, Bormans J. The JVT advanced video coding standard: Complexity and performance analysis on a tool-by-tool basis. 2003. <http://www.polytech.univ-nantes.fr/pv2003/papers/pv/papers/cr1008.pdf>
- [16] Halbach T, Wien M. Concepts and performance of next-generation video compression standardization. 2002. <http://www.norsig.no/norsig2002/Proceedings/papers/cr1121.pdf>
- [17] Ebrahimi T, Horne C. MPEG-4 natural video coding—An overview. Signal Processing: Image Communication, 2000,15:365~385.

- [18] Wiegand T. Editor's Proposed Draft Text Modifications for Joint Video Specification (ITU-T Rec. H.264|ISO/IEC 14496-10 AVC), Draft 7, JVT-E022d7, 2002.
- [19] RFC 1889. 1996. <http://www.faqs.org/rfcs/rfc1889.html>
- [20] Wenger S, Stockhammer T, Hannuksela M. RTP payload Format for JVT Video. draft-wenger-avt-rtp-jvt-00.txt, 2002.
- [21] 2002. <http://www.ietf.org/internet-drafts/draft-ietf-avt-rtp-h264-00.txt>
- [22] Zheng X. Research on building a VOD system with MPEG4 technology. In: Int'l. Symp. DCABES 2002. China, 2002. 377~380.
- [23] Hannuksela M, Stockhammer T, Wenge S. H.26L/JVT coding network abstraction layer and IP-based transport. In: IEEE Int'l. Conf. on Image Processing. NY, 2002. <http://www.icip2002.com/Papers/viewpapers.asp?paperum=2738>
- [24] Thailand P. Requirements for JVT codec. Joint Video Team (JVT) of ISO/IEC MPEG and ITU-T VCEG.JVT-A004, 2001.

第5届中国计算机图形学大会(Chinagraph 2004)

征文通知

由国家自然科学基金委员会、中国计算机学会、中国自动化学会、中国工程图学学会、中国图象图形学会、中国系统仿真学会、香港多媒体及图象计算学会联合主办的第五届中国计算机图形学大会(Chinagraph 2004)将于2004年9月23日~26日在西安举行。中国计算机图形学大会是华语学者计算机图形学的重要论坛。大会内容包括大会学术报告,图形学热点问题专题研讨,图形学最新成果和应用软件演示,并邀请国内外著名图形学专家到会作特邀报告。大会录用的论文将收入会议论文集出版,其中优秀的论文将直接推荐到《计算机学报》、《计算机辅助设计与图形学学报》、《软件学报》、《工程图学学报》的2004年9月份相应的期刊上发表,《系统仿真学报》和《中国图象图形学报》也将刊登部分优秀的论文。

中国计算机图形学大会设立中国计算机图形学贡献奖和中国计算机图形杰出奖,每届会议各奖一名。此外大会还将设立最佳论文奖。本届会议由西北工业大学承办。热忱欢迎一切从事计算机图形学研究、应用及软件开发的海内外所有华语专家、学者和专业技术人员踊跃投稿。

一、征文内容

几何造型、计算机辅助几何设计、图形学基础理论与算法、虚拟设计与制造中的图形/图像技术、科学计算可视化、虚拟现实、计算机视景与图形仿真、多媒体技术、计算机动画、真实感图形、工程图形及应用、基于图像的图形技术、与计算机图形学相关的其他领域。

二、论文格式及投寄地址

所投寄论文应在国内外未曾公开发表过,在理论或应用方面有创见。

论文包括:中英文的题目、摘要、关键词、作者姓名、单位名称与地址,中文的论文正文,参考文献。

每篇论文应提供第一作者的简介、通讯地址、电话及Email地址。

论文最好以Email方式投稿至: cg2004@nwpu.edu.cn

如电子投稿有困难,请将论文一式二份与存有电子文档的计算机软盘一份寄至:

西安市 西北工业大学 324 信箱 谢琴(收), 邮编: 710072(请在信封后注明: chinagraph 2004 稿件)

联系电话: 029-88491576(兼传真)

三、重要日期

截稿日期: 2004年4月5日

录用通知: 2004年5月15日

修改稿截稿日期: 2004年6月15日

四、Chinagraph 2004 网址

<http://chinagraph2004.nwpu.edu.cn>

欢迎上网查询大会各项文件和最新通知