

错误多发异质网络环境下视频编码系统的总体设计

沈未名¹ 朱立¹ 李国宽¹

(1 武汉大学多媒体网络通信工程湖北省重点实验室, 武汉市珞喻路 129 号, 430079)

摘要: 描述了针对错误多发异质的网络环境进行视频编码系统总体设计的全过程。分析了外部需求, 提出了改进视频表现、增强容错能力和提供可分级编码能力的系统设计目标。在简要介绍 H. 263+ 协议的全部 16 个可选的高级编码模式的基础上, 重点分析了设计中对系统是否采用某个高级编码模式进行取舍的理由。对视频数据流的打包、错误掩蔽和错误跟踪的机制等方面进行了讨论。实际应用表明, 根据本文设计进行的系统实现完全达到了原定的系统设计目标。

关键词: 视频编码; H. 263+ 协议; 错误多发; 异质网络
中图法分类号: TN919.81; TP311.12

公众需求、技术发展和工业界的努力正推动多媒体通信渐成网络通信业务的主流。实现多媒体通信系统的关键技术问题之一是如何对海量的数字化视频信息进行数据压缩和必要的容错处理, 使之得以在有限带宽的通信网络中正确传输, 即所谓视频编码问题。国际电信联盟标准委员会 (ITU-T) 继 1993 年推出了视频编码领域世界上第一个国际标准 H. 261 协议之后, 又分别在 1996 年和 1998 年推出了第二代视频编码国际标准 H. 263 协议和 H. 263+ 协议 (H. 263 协议第二版)^[1]。视频编码国际协议的制定旨在促使多媒体通信系统的不同实现遵循同样的视频编码规范, 从而保证多媒体通信系统在全球范围内的互通性。H. 263+ 协议适用于多样化的通信网络环境, 提供了多种可选的编码模式供实现者根据系统的目标运行环境选用, 同时在编码策略、算法、前后端处理等方面保留了相当的灵活性, 以鼓励不同实现者之间的竞争和技术创新。因此, 同样是遵循 H. 263+ 协议的视频编码系统, 其性能表现有可能迥异, 有针对性的优化视频编码系统的总体设计就显得尤其重要。本文面向错误多发异质的网络环境 (如 IP 广域网或互联网), 基于 H. 263+ 协议对视频编码系统涉及的各个方面进行了总体设计。

1 视频编码系统的设计目标和需求分析

本文涉及的视频编码系统预期作为遵循 ITU-T H. 323 协议的网络视讯会议系统的一部分, 运行在计算资源受限的计算机平台上和错误多发异质的 IP 网络环境中。笔者根据视频编码系统的工作环境作出系统需求分析, 并将系统设计目标分解为以下 3 个子目标。

1) 在网络带宽和计算资源的双重限制下提供清晰流畅的视频表现。

获得清晰流畅的视频表现的主要途径在于提高编码的效率 (压缩比), 这始终是视频编码问题的核心, 在实用系统的设计中往往还需要考虑对编码算法计算复杂度的限制, 这增加了系统设计的难度。编码端和解码端的滤波处理常常也能够一定程度上改善视频输出的主观表现。

2) 面向错误多发的网络环境提供容错编码、错误掩蔽和错误恢复的机制。

IP 网络上的数据错误主要表现为数据包丢失或乱序, 由于多媒体通信系统对实时性有很高的要求, 这类错误通常无法通过利用包的重传来克服。因此, 必须在编码端进行特殊的容错处理, 尽可能地消除包与包之间的数据相关性, 限制错误的扩散, 在解码端进行错误掩蔽以改善视频

表现,同时通过编码端和解码端的协同提供从错误中恢复的能力。

3) 面向充分异质的网络环境提供视频比特流的可分级能力。

现在的 IP 网络环境(广域网或互联网)是充分异质的,人们可以通过 PSTN、ISDN、DSL、FD-DI 或 ATM 等各种方式接入,各接入端点可以获得的网络 QoS(特别是可用带宽)有很大的差别。因此,一个异质网络环境下的多点视讯系统或视频流广播系统必须提供视频比特流的可分级能力,从而保证以相对较低带宽接入的终端能够获得基本的视频表现质量,同时不妨碍以相对较高带宽接入的终端获得更高质量的视频表现。

2 ITU-T H.263+ 协议内容及其基本原理简介

H.263+ 是 ITU-T 在 1998 年 2 月通过的关于窄带视频编码的国际协议。该协议所规定的视频编码器,其主体与 H.261、H.263 协议所规定的相同,是一个基于运动补偿的帧间预测与变换编码技术的混合体。在作为主体的基本编码模式之外,H.263+ 提供了多达 16 个可选的高级编码模式(包括 H.263 协议已经提供的 5 个高级编码模式),它们分别适用于不同的应用环境,对改进视频编码系统的性能至关重要。下面对 H.263+ 新增的各个高级编码模式的内容、特点和主要功效作一个简要的介绍。

2.1 改良 PB 帧模式

这个模式是 H.263+ 对 H.263 中已经包含的 PB 帧模式的改良。主要的不同是,增加了两种编码模式供选择,允许基于一个独立的运动矢量进行前向预测或进行零运动矢量后向预测。这样的改良增进了 PB 帧模式的编码效率,尤其适用于当前面的 P 帧与当前的 PB 帧之间存在场景切换的时候。

2.2 高级帧内编码模式

帧内块的变换系数通过相邻的帧内块来进行预测后仅对差值编码,对这些差值的编码使用独立的变长编码(VLC)码表。这个模式极大地提高了帧内编码的效率。

2.3 去块斑滤波器模式

使用一种跨越块的边缘的滤波器来削弱块斑效应。这个滤波器作用在被用作后续图像的预测参照的重建帧上,因而被包含在运动预测的环路中。一个有益的附加约定是,在这种模式下同样

允许运动矢量指向图像边缘之外,而且图像的某些宏块可以使用 4 个对应于块的不同运动矢量。

2.4 片层结构模式

在比特流语法结构中块组层被片(slice)层替代。如同块组层,片层同样是位于图像层和宏块层之间的中间层,但是它允许对图像进行不依照自然的、空间顺序更为复杂的分割和排序。这种模式产生的比特流更适宜于底层的包交换传输机制,有助于提供增强的容错能力和减小视频时延。

2.5 参照帧选择模式

允许使用非最邻近传输帧作为参照帧来进行时域上的预测。通过反向通道提供的状态消息,编码器能够获知远端的解码器是否已经正确接收了包含某帧图像的比特流,从而选择合适的编码预测方式,增进了实时视频系统在错误高发信道上的表现。

2.6 时间域、信噪比和空间域上可分级性模式

支持时间域、信噪比和空间域上的可分级性。可分级性意味着比特流由一个基本层和一个或多个相关的增强层组成。基本层是一个可以独立解码的比特流。增强层则必须与基本层相结合进行联合解码以增进在帧率、图像质量或图像大小方面的视频表现。其中,时间域上的可分级性的实现基于所谓的 B 帧,它由参照层上的两帧图像通过双向预测而得。这个高级模式主要用于内含不同带宽能力的异质网络,并且可以与纠错机制相结合。

2.7 参照帧重采样模式

支持对参照帧的重采样以形成当前帧的预测器。这个模式允许在视频编码端高效地动态切换图像分辨率,而且也支持用作全局运动补偿器或特效生成器的图像弯折处理。

2.8 降低分辨率刷新模式

允许以降低的分辨率对具有较高分辨率的参照帧进行刷新。这个模式被设计用于高度动态场景的编码,允许编码器提高图像帧率以表达场景的运动部分,同时保持场景中静态部分的高分辨率表示。

2.9 独立分段解码模式

图像被分段为由彼此数据无相互依赖关系的部分构建,这些部分可以是块组、多块组图像分段或片层。这个模式通过阻止错误数据跨越图像分段区域边界的传播,提供了对数据错误的稳健性。

2.10 选择性帧间变长编码模式

通过允许使用为帧内编码设计的变长编码码表,提高帧间变化显著时帧间编码的效率。

2.11 改进量化模式

这个模式能够提供改进的码率控制能力, 减少色度分量的量化误差, 扩展可表示 DCT 系数的范围, 并对系数的取值加以某些限制(因而增进错误检测的性能表现并且降低解码器的复杂度)。

3 视频编码系统的总体设计

错误多发异质网络环境下的视频编码系统的设计目标主要涉及增进视频表现、提高丢包容错能力和支持可分级编码三个方面。以下介绍了针对这些目标在 H. 263+ 高级编码模式的选择、视频数据流 RTP 打包机制、错误掩蔽和错误跟踪策略等方面进行的设计并阐述其理由。

3.1 H. 263+ 高级编码模式的选择

H. 263+ 协议提供相当多的可选高级编码模式以提供不同应用场合下视频编码器实现的灵活性。在 H. 263+ 的某个实现中提供对所有高级编码模式的支持是不必要的, 并且从计算复杂度的角度来看也是不现实的。因此在具体的协议实现中必须对 H. 263+ 各个可选高级编码模式的取舍作出决定, 决定的主要依据一般包括: 对主观视频质量的改进程度; 对时延的影响; 复杂性(包括计算负载、数据依赖关系、实现难度等)^[2]。此外, 系统的设计目标已经决定了它必须支持两个重要的特性: 包交换网络环境下的充分的丢包容错能力; 分层编码能力, 以支持不同带宽网络之间的透明无失真的视频数据共享。

综合以上考虑, 本文涉及的视频编码系统被设计为优先支持以下几种高级编码模式。

3.1.1 高级帧内编码模式

选用这个模式可以显著地提高帧内宏块的编码效率, 而增加的计算要求在编码端和解码端都微乎其微(在解码端对于每个 8×8 的块只需要增加最多 8 个加法/减法运算和使用一个不同但近似的变长编码码表)。

3.1.2 去块斑滤波器模式

因为去块斑滤波器对改进图像主观质量的显著作用, 它早已被广泛用作一种对解码图像的后处理方法。如今, 这个高级模式使去块斑滤波器从一种后处理方法转变为置于编码环路中的基本操作单元。这样的转变使滤波器的实现更为容易(减少了存储要求), 并且在某种程度上改进了编码的表现。如同高级预测模式, 这个模式同样包括了一个宏块对应于四个基于块的运动矢量和以运动补偿为目的的图像边缘外推这两个特性, 因

而能够进一步增进编码效率。实现去块斑滤波器的计算要求大概是每个编码宏块需几百次运算, 但内存操作和计算依赖性并不复杂。正是最后这一点使我们倾向于选择去块斑滤波器模式而不是高级预测模式。

3.1.3 改进量化模式

这个模式包括了扩展的 DCT 系数取值范围、改进的 DQUANT 语法和改进的色度空间量化步长等特性。前两个特性为编码器提供了决定量化系数的更多的机动性并且实际上降低了编码器的计算开销(通过消除当系数饱和发生时必须对宏块进行重新编码的可能性)。第三个特性显著地改进了色彩的逼真度, 而代价只是稍微增加了比特率, 计算负载则基本不变。在解码端, 惟一值得一提的计算负担是需要提供解释某些新的比特流语义符号的能力。

3.1.4 片层结构模式

这个模式提供了对包交换传输层的兼容性, 并且有可能增进在网络丢包的情况下进行容错处理的能力。我们支持这个模式的所有子模式, 包括矩形片层子模式和任意片层次序子模式。增加的计算负担相当小, 主要限于比特流生成和解析。

3.1.5 时间域、信噪比和空间域上的可分级性模式

这个模式特别适用于具有不同带宽能力的异质网络, 并且与纠错方案相结合能够为视频数据流提供分级的容错保护能力。

本文涉及的视频编码系统被设计为不支持其他高级编码模式, 理由如下:

1) 不必支持多点连续呈现和视频复用模式, 因为可以通过建立多个独立的视频逻辑通道来传送多路视频或通过 MCU 内置的 MP 单元对多路视频进行集中处理(如四路视频的矩阵式复合)来提供多点之间的多画面呈现能力, 而且另一方面, ITU-T H. 323 框架协议明确规定在包交换网络多媒体系统中不宜采用这个高级编码模式;

2) 不必支持非受限运动矢量模式和高级预测模式, 因为这两个模式中被实践证明有效的要素已经在优先采用的去块斑滤波器模式中得到了支持;

3) 不必支持基于语法的算术编码模式, 因为该模式带来的编码效率的提高十分有限, 而算术编码码流的强烈的上下文关联性严重限制了视频数据容错打包机制的灵活性;

4) 不必支持 PB 帧模式、改良 PB 帧模式和参照帧选择模式, 因为这几个模式所能够提供的

时域上的可分级编码能力已经由优先采用的时间域、信噪比和空间域上的可分级性模式提供;

5) 不必支持独立分段解码模式, 因为该模式以相当大程度地降低编码效率来换取错误高发网络环境下的较强的容错能力, 而本文涉及的系统的目标运行网络环境下的数据丢包率一般并不非常高, 因此无须如此严重地牺牲编码效率;

6) 不必支持参照帧重采样模式、降低分辨率刷新模式和选择性帧间变长编码模式, 因为这几个模式只有在特定的场景(镜头摇转、强烈运动等)下才能够表现出其在提高编码效率方面的优越性, 而视讯会议系统所面向的场景基本上还是以头肩像的适度运动为主, 不必采用计算复杂度过高的选项。

3.2 H.263+协议以外相关处理的考虑

3.2.1 视频数据流 RTP 打包机制

ITU-T H.323 协议规定, 包交换网络多媒体系统的视频数据应该以 RTP 协议方式打包, 并且通过独立的数据逻辑通道进行传输, 这些视频逻辑通道应利用 H.245 信令机制来建立, 并且如果采用 H.263+ 可分级编码模式, 则对应于每个层次的视频码流必须分别建立独立的逻辑通道^[3]。

对于丢包多发的广域网络环境下的视频系统而言, 必须要求系统提供对每个数据包独立于其他数据包的容错解码能力。笔者根据 H.263+ 编码方案的特性, 参考文献[4]中的方法, 对视频数据流的 RTP 打包机制进行了设计, 以满足上述关于数据包独立解码的要求。这个打包机制的核心原则是, 在数据包中对包内视频数据涉及的来自其他包的信息进行冗余的编码, 在一定程度上以牺牲编码效率换取面向丢包的容错能力。

3.2.2 错误掩蔽和基于 RTCP 反馈的错误跟踪

当丢包错误确实发生时, 图像的某个部分(某个块组、片层或图像一行宏块的一部分)可能无法通过正常的方法在解码端重建, 但仍然有可能通过基于简单推理的空间预测得到粗略的恢复, 这就是所谓的错误掩蔽^[5]。

另一方面, 通过 RTCP 反馈通道, 编码端能够从解码端获得有关网络丢包目前情况的准确信息, 视频编码器可以据此调整其行为, 以便在当前网络状况下提供尽可能好的视频表现, 同时帮助解码器尽快从错误中恢复。

一种简单的调整策略是, 可以根据这些反馈信息计算评估网络当前的丢包率, 并据此调整视频输出带宽和宏块帧内编码的刷新频率。采用一种更为精细的调整策略, 可以将帧内编码模式的选择精确定位到确实受丢包错误影响而必须进行帧内编码的宏块, 这里涉及到对丢包造成的数据错误在帧间传播的准确分析^[6]。

4 结 论

本文针对错误多发异质的网络环境, 基于 H.263+ 协议对视频编码系统涉及的各个方面进行了总体设计, 这些工作基本上是原创性的。这个总体设计在保持系统对协议的符合性的基础上, 充分利用了协议的灵活性以满足实际应用系统的需求。本文设计的视频编码系统是自主开发的网络视讯会议终端的子系统。实际应用表明, 该系统的实现完全达到了系统设计目标; 在网络带宽和计算资源的双重限制下, 能够提供清晰流畅的视频表现; 面向错误多发的网络环境能够提供相当强大的容错编解码能力; 面向充分异质的网络环境能够提供视频比特流的可分级传输能力。

参 考 文 献

- 1 ITU-T Recommendation, Video Coding for Low Bit Rate Communication, 1982
- 2 ITU-T Recommendation H.263 Appendix II. Recommended Optional Enhancement, 1998
- 3 ITU-T Recommendation H.223 Annex B. Procedures for Layered Video Codecs, 1998
- 4 RTP Payload Format for the 1998 Version of ITU-T Rec. H.263 Video (H.263+), RFC 2429. IETF 技术文档, 1998
- 5 Video Codec Test Model Near-Term, Version 11, Document Q15-D-65d1, Study Group 16 (Video Coding Experts Group), ITU-T 第 15 专家组技术文档, 1998
- 6 ITU-T Recommendation H.263 Appendix I. Error Tracking, 1997

作者简介: 沈未名, 教授。现主要从事图像与视频数据压缩和编码、计算机立体视觉等方面的研究开发工作。代表成果: 可视电话视频编码子系统的设计和实现; 基于小波变换和神经网络的遥感影像数据压缩; 基于神经网络智能整体影像匹配, 等。

E-mail: wmshe@public.wh.hb.cn

Framework Design on Video Coding System for Error-prone Heterogenous Network

SHEN Weiming¹ ZHU Li¹ LI Guokuan¹

(1 The Key Laboratory of Multimedia and Network Communications Engineering, Hubei Province, Wuhan University, 129 Luoyu Road, Wuhan, China 430079)

Abstract: This paper describes the complete procedure of framework design on a video coding system for error-prone heterogenous network environment. First of all, the requirements of the system are analyzed and the targets of the system design are emphasized on three categories: ①improving video performance under the constraints of network bandwidth and computational complexity, ②providing error robust mechanism when packet loss occur, ③adding capability of layered video coding for heterogenous network transmission. Then, ITU-T H. 263+ recommendation is introduced, especially about its 16 optional enhanced coding modes: feature, effect and benefit. The system design on mode selection is somewhat a trade-off. Five enhanced coding modes are preferred for our purpose: advanced INTRA coding mode, deblocking filter mode, modified quantization mode, slice structured mode and temporal, SNR and spatial scalability mode. At last, some useful and important system elements beyond H. 263+ recommendation are discussed: RTP packetization, error concealment and error tracking. Application shows that all the targets are easily touched by the implementation based on this design.

Key words: video coding; H. 263+; error-prone; heterogenous network

About the author: SHEN Weiming, professor. His research interests include image and video coding, stereo vision etc. His articles include Design and implementation of a video codec in videophone. Data compression on remote sensed image based on wavelet and neural network. Intelligent global image matching based on neural network, etc.

E-mail: wmshen@public.wh.hb.cn