

中华博士 园地

这是本刊特为海内外正在就读和学成立业的博士、博士后青年学者们开辟的一片科普园地. 深学浅著是一门德识、慧学、素质修养的学问. 你们的新知识、新调研、新观察、新目光、新展望, 能够用尽可能深入浅出、通俗流畅的语言, 汇报给祖国人民、家乡父老子弟乡亲们吗? 中华博士园地, 乃耕耘忠孝之地, 科教兴国、民族昌盛之地. 要用慈母听得懂的语言, 写出你们的心声!

中图法分类号: TN919.81 文章编号: 1006-8961(2001)06-0608-05

面向 Internet 的视频编码技术: 精细可分级性(FGS)的实现

陈志波 何 芸

(清华大学电子工程系微波与数字通信国家重点实验室, 北京 100084)

0 引 言

Internet 自 60 年代以来得到了迅猛发展, 近几年更以惊人的速度增长, 联网主机量每年翻一番, 万维网站点每半年翻一番, 同时随着音频、视频等相关多媒体技术的发展, 基于 Internet 的多媒体应用层出不穷, Internet 上多媒体信息的流量也与日剧增, 而且 Internet 已经逐步由单一的数据传输网向数据、语音、图象等多媒体信息的综合业务传输网演化, 而视频流式传输(video streaming)则是其中一

项正在迅速发展的技术. 所谓视频流式传输, 首先是一个动态的概念, 即从服务器端连续发送视频数据经过网络到达用户端, 这一连续的随时间变化的码流被称为流(stream); 另外从用户端的角度来看, 不需要下载全部码流就可以开始解码播放, 而且码流的到达是连续无间隔的. 视频流式传输在视频点播、网上购物、电视会议、远程教育, 以及实况转播等方面均有着广泛的应用前景.

视频流式传输具有连续传输、数据量大、且对实时和可靠性有一定要求等特点, 但是由于以下一些原因: ①基于无连接传输的 Internet 网络不是为连续媒体的传输而设计的, 而是一个与时间和空间的变化都相关的共享通信信道; ②网络通信中同一群组中的各个用户不仅可能在机器类型、操作系统、外设性能(如显示分辨率和颜色)、CPU 处理能力等方面均存在着巨大差异, 而且连接各用户的各段网络也会在带宽、延时和误码率等方面存在差异, 此即所谓的网络异质特性(heterogeneity)结构(如图 1 所

陈志波 1998 年毕业于清华大学电子工程系, 获学士学位, 现为清华大学电子工程系博士研究生, 研究方向为视频通信、多媒体传输;

示);③基于公平竞争原则的 Internet 网络结构无法提供多媒体业务所需的 QoS 保证,因此所有这些都

对 Internet 的多媒体应用提出了巨大的挑战.

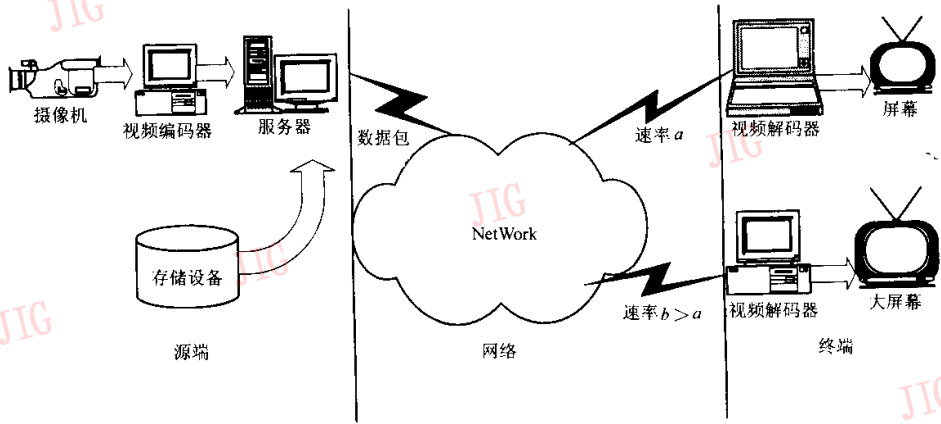


图 1 Internet 上不同的用户群和网络带宽

针对上述问题,已有的众多解决方案主要从如下两个角度考虑问题:其一是从信号处理的角度考虑,即针对 Internet 网络的异质特性,在信源端采用分级(scalable)编码、多速率编码、速率适配或在网络传输节点采用转换编码器(transcoder)等;其二是从网络层的角度考虑,即在原有 Internet 网络结构的基础上进行改进,如增加带宽、采用组播(multicast)服务、增强路由器和服务器功能,以及使用数据缓存(data caching)和边缘服务器(edge server)等等;或者是在此基础上,对现有传输控制协议进行改进.

虽然网络的发展可使有效带宽变得越来越宽,但事实上由于网络上用户数量的增长速度比有效带宽的增长速度快得多,因此图象和视频信号的有效压缩始终是一个好的视频传输系统所必须解决的问题.目前 ITU-T, ISO 等国际标准化组织都已经展开了研究,并致力于制定下一代适应于网络应用的视频编码协议,其中, MPEG-4 的第 4 版即针对 Internet 视频流式传输的应用,定义了视频图象编码的精细可分级性 FGS(Fine Granularity Scalability)以及实现工具,因而是一个全新的编码体系.由于 FGS 编码提出了一个由网络或者接收端来控制其所接收的码流的概念,从而使视频流发送端同时面向各种网络和更多的用户,并为视频流在异质特性的网络上向大量用户传输提供了理想的码流结构.

1 传统的分级编码技术

在视频的分级编码技术中,视频信息被分成多

个重要性不同的层,其中基本层包含了视频最重要的基本信息,以此可以保证一个最基本的图象质量;增强层的作用是在基本层的基础上进一步提高图象质量.文献[1]中已证明:对信源 X ,先以失真为 D_1 、码率为 R_1 的粗描述 \hat{X}_1 ,失真为 D_2 ($D_2 \leq D_1$)、码率为 R_2 的细描述 \hat{X}_2 来逐次精细化过程,当且仅当 $X \rightarrow \hat{X}_2 \rightarrow \hat{X}_1$ 为马尔可夫链时,才有 $R_1 + R_2 = R(D_2)$,其中, $R(\cdot)$ 为率失真函数.一般来说,实际的分级编码中,马尔可夫链的条件是不能严格满足的,这也从理论上说明分级编码的编码效率相对来说会有所下降.

在传输过程中,基本层往往被赋予较高的传输优先权(例如在 ATM 的信元头中有 1bit 标识了信元的优先级),或者有更多的纠错保护;而增强层则赋予较低的传输优先权.这样在网络发生拥塞而丢包时,就可以先丢弃优先级较低的增强层,或者说,由于基本层和增强层的不同保护机制,使得基本层发生丢包或误码的概率比增强层小,从而保证了在发生丢包或误码的情况下,重建图象仍然有一个可以令人接受的质量.除此之外,分级编码还可以应用于事先无法知道信道带宽和接收端的处理能力的情况,通过分级传输信息来达到适应多种速率的信道,以及适应多种接收终端的目的.这样分级编码就可以应用在广播或组播传输方式中,所以,在 MPEG-2、MPEG-4 和 H.263+ 中都采纳了分级编码的技术,而 JPEG 中采用的渐进编码(progressive coding)模式也是一种分级编码.

根据分级编码中对视频信息的分割方式不同,分级编码可以具有不同的实现方式,主要有时域分

级(Temporal Scalability)、空域分级(Spatial Scalability)和质量分级或又称 SNR 分级(SNR Scalability)3 种分级方式. 图 2 给出了一个采用空域分级的编码系统, 其中, D 表示解码器. 由于时域分级中信息的分割是在时域上进行, 因此基本层视频信息的帧率较低, 若加上增强层, 则可以得到一个较高的帧率; 而在空域分级编码中, 基本层和增强层具有不同的空间分辨率, 因此基本层是对原始图象的亚采样进行编码, 而增强层则是对多余的信息进行编码. SNR 分级是在频率域上进行分级, 其中一种分级是直接对频率域的系数分割成高频和低频部分, 其分别对应为基本层和增强层, 这在 MPEG-2 中称为数据分割(Data Partitioning); 另一种分级则是首先在基本层对频域系数进行粗量化, 然后在增强层对上述频域系数的量化误差进行细量化. 如今这 3 种分级编码方法在 ITU 的 H. 263、MPEG 的 MPEG-2 和 MPEG-4 中都被采纳. 由于各种不同的分级方式, 其算法复杂度和开销是不同的, 因此在实际应用中需根据不同的具体要求来进行选择.

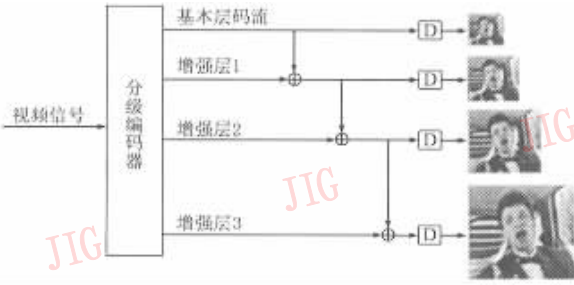


图 2 分级视频编码系统示意图(空域分级)

2 精细可分级编码 FGS (Fine Granularity Scalability)

传统的分级编码方法并不能有效满足网络传输视频信号的应用, 因为这种传统的分级步长相对于网络的变化来说太粗糙了, 且这个步长往往是整个增强层的比特数, 而网络的状况则要求服务器能随着网络真实带宽的变化来进行确实有效地速率调整, 由此可见, 如果能够实现精细的视频分级(Fine Granularity Video Scalability), 那么码流结构就可以达到这个目标.

2.1 FGS 编码器

FGS 的基本编码结构如图 3 所示, 其中, 虚线框中是 FGS 增强层. FGS 编码时, 基本层采用的是

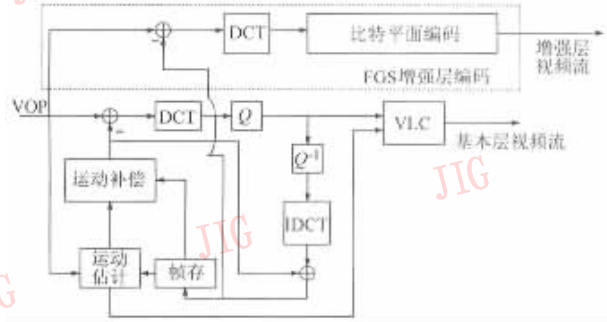


图 3 基本 FGS 编码器框图

MPEG-4 基本模式编码, 并输出基本视频流, 增强层的输入信号是原始 VOP 和重建的 VOP, 编码时, 增强层编码器首先将二者相减, 并对其差值进行 8×8 的 DCT, 然后再对 DCT 系数进行比特平面编码^[2].

FGS 增强层中的比特平面编码包括比特平面移位, 搜索最大值(MAX), 比特平面变长编码(VLC)3 部分. 其中, 比特平面移位操作是将 64 个 DCT 系数进行加权运算, 因为各个部分的 DCT 系数的重要性不一样, 所以可以通过移位来增加感兴趣区域的权值, 使其处于码流中前面的位置而不易被丢弃; MAX 运算则是对加权了的 DCT 系数求最大值, 从而确定比特平面的层数; 比特平面操作则是对每一层比特平面进行变长编码.

在 FGS 增强层的码流结构中, 根据比特平面的重要性, 将最重要的比特放在码流的前端, 比如每一个块的 MSB, 而将次要的比特放在后端, 比如每一个块的 LSB, 这样在网络发生拥塞时就可以先丢弃部分不重要的比特, 以减缓网络负载, 但解码端接收的码流仍然可以解码, 只是相应的重建图象质量会有所下降.

2.2 FGS 解码器

FGS 解码器是编码器的逆过程, 图 4 是 FGS 解码器基本框图, 其虚线框中为 FGS 增强层的解码器. FGS 解码器的输入是增强层和基本层的码流, 增强层的 VOP 码流首先经过比特平面 VLD, 即对

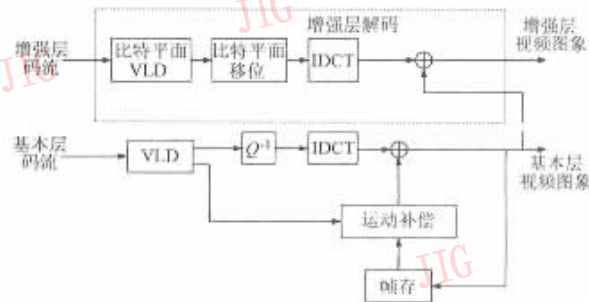


图 4 基本 FGS 解码器框图

每一比特平面进行变长解码、DCT 系数合成及移位还原, 然后进行反 DCT 变换, 最后与基本层视频图象相加, 以输出包含增强层信息的视频图象, 而基本层的视频图象则作为选项输出. 在进行增强层解码时, 如果部分低优先级的比特平面被丢弃, 或同一比特平面部分优先级低的比特被丢弃, 则只能使增强层图象重建的精度降低, 而不会影响码流解码的有效性. 由于数据丢弃是以 bit 为单位, 所以码率可以以 bit 为单位进行调整, 重建图象质量也可以以 bit 为单位变化. 这也正是 FGS 精细可分级的原理和优点所在.

2.3 对 FGS 编码效率的改进

FGS 的优势在于精细的比特分级, 而且其增强层的码流还可以实现任意比特的截取, 因此相应解码端的图象重建质量将随着获取增强层的比特数的增多而增加. 另外由于 FGS 增强层的预测始终只是在重建质量很低的基本层上进行的, 而且没有利用同层的信息预测, 因此 FGS 的编码效率显然不高, 如今也有许多研究试图进一步提高 FGS 的编码效率, 如文献 [3] 提出的 PFGS (Progressive Fine Granularity Scalable) 的原则就是首先尽可能地使用同层的预测来提高编码效率, 同时保证预测的路径总是从参考帧的低层开始, 来进行误差恢复和信道的适配, 虽然其代价是需要更多的帧存, 但在同等码率下采用 PFGS 可以使重建图形的 PSNR 提高近 2dB.

目前 ITU 正在制定的 H. 26L 标准中也采纳了 FGS 的技术, 它基本与 MPEG-4 的 FGS 技术相类似, 其基本层采用了 H. 26L 的编码方法 (如基于 4×4 的整形变换等), 而增强层则只对原始图象与基本层解码图象之间的残差进行编码, 而且也采用了基于 4×4 的整形变换, 并利用整形变换的特点, 将量化后的系数分成 16 个子带进行比特平面编码, 而熵编码则采用基于上下文的算术编码^[4].

3 FGS 与组播技术的结合

Internet 上数据传输的方式通常分为单播 (unicast)、广播 (broadcast)、组播 (multicast). 其中, 单播本质上是点对点的传输, 即源节点向每一个目的节点独立地传送一份数据流, 虽然目前 Internet 上大部分应用的传输都采用这一方式, 但其缺点是网络资源利用率低, 以及可扩展性差; 广播是一对全部

的传输, 即由一个源节点发送的一份数据, 可被网络中的所有节点同时接收, 无论用户是否需要或者发送端是否想让某些用户接收, 其缺点是会造成无效耗费资源和安全问题; 组播是一种将一组数据发送到某个组播组内的多个用户端的传输方式, 经过多年地讨论和研究, 这一技术已得到了越来越多的重视, 其相对于传统的基于 IP 的单播、广播传输方式来说, 更加有效、经济和安全, 且占用更少的带宽和处理器资源, 规模易于扩展, 而且不会轻易发生网络拥塞. 大公司可以将更新的软件和数据通过组播传到分公司, 并在全世界范围进行存储, 而许多媒体公司也可以将实时的或存储的视、音频信息, 通过组播实时地传到自己的用户那里, 一些 Internet 服务提供商 (ISP) 们已经开始利用组播技术将内容传到一些分布式的服务器上. 虽然组播具有很多优点, 但是组播传输同样面临着众多的问题, 其中一个就是如何在网络的异质特性下, 能够保证每个组播组内的不同用户可以得到与其带宽或处理能力相符的最佳图象重建质量, 而 FGS 的精细分级特性已可使这种实现成为可能.

图 5 是一个可能的 FGS 视频流组播传输的例子, 其中, 假设用户 1 通过 10M 以太网接入网络, 用户 2 通过 E1 的 ISDN 网接入, 用户 3 用 56kModem, 而用户 4 采用 28kModem. 这种组播的特点是服务器将码流根据不同的 QoS 要求打包, 当拥塞发生或带宽限制时, 则由路由器根据网络连接和用户的情况有选择的发送数据包, 如图 5 中二元组表示 (容量, 请求), 其中, 容量表示实际路由器或终端用户系统的有效资源, 而请求则表示路由器或用户要求获得的视频资源, 且每个路由器的请求是下一级所有路由器的要求和连接到此路由器的用户请求中的最大值, 而对终端用户来说, 其请求参数则由用户的选择和交互行为决定. 路由器 3 先根据用户 3 和用户 4 的请求发出上一级路由器的请求, 路由器 1 根据下级路由器 2、3 的

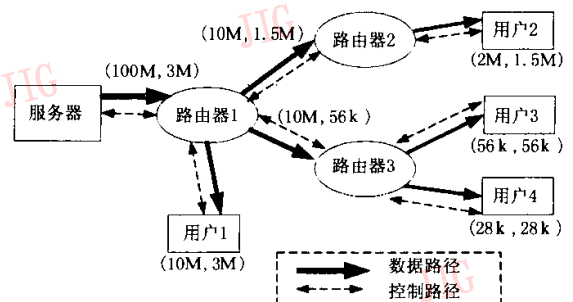


图 5 FGS 的组播传输方案

请求和直接联到本路由器的用户 1 的请求向服务器发出请求, 每级路由器根据下级请求转发相应的数据包, 当拥塞发生时, 路由器可以通过控制丢弃低优先权的包(如 FGS 中最不重要的比特平面的数据)来降低流量, 以减缓拥塞。

由图 5 可以看出, 这种组播传输方案对路由器的功能也提出了新的要求, 如进行包的有选择拷贝, 优先权控制等。这也正如引言所说的, 只有从信号处理和网络层两方面来进行综合考虑才有可能促进多媒体在 Internet 上的飞速发展。

4 总 结

Internet 是一个具有高度异质特性的系统, 这对 Internet 上进行可靠的流式视频传输提出了巨大的挑战。虽然分级编码技术就是针对这一问题提出的, 但是传统的分级编码不能提供码率可以连续分级的码流, 而 FGS 编码从根本上解决了这一问题, 再结合组播传输技术, 则可以很好地适应 Internet 的异质特性。如果在服务器端能够发送基本层速率

符合最低网络带宽条件, 增强层能满足最高质量要求的 FGS 视频流, 然后由网络和接收端根据不同用户需要和网络情况对增强层码流进行截取, 以便与设备和网络状态相匹配, 那么这样就可以保证每个组内的不同用户能够得到与其带宽或处理能力相符的最佳图象重建质量。

可以预见, FGS 结合组播(Multicast)传输, 将在未来网络多媒体的应用中起重要的作用。

参 考 文 献

- 1 William H R Equitz, Thomas M Cover. Successive refinement of information. IEEE Trans. Inform. Theory, 1991, 37(2):
- 2 FGS verification model version 4. 0. ISO/IEC JTC1/SC29/WG11, April, 2000, N3317
- 3 Shipeng Li, Feng Wu, Ya-Qin Zhang. Study of a new approach to improve FGS video coding efficiency. ISO/IEC JTC1/SC29/WG11, Dec 1999, M5583.
- 4 Marpe D, Blättermann G, Heising G. Technical description of a quality-scalable coder for H. 26L. ITU/SG16 Q15-K-12, Aug, 2000.