

基于可伸缩编码的视频会议系统

叶晓蕾¹, 梁金明², 邓云²

(1. 四川省盐业学校计算机中心, 四川 自贡 643000; 2. 四川理工学院计算机科学学院, 四川 自贡 643000)

摘要:针对传统视频会议需要在同型终端和同型网络中使用的不足,提出一种基于可伸缩性视频编码技术和有线/无线网络交互式技术的新型的视频会议系统。该系统通过可伸缩性编码生成在时域和空域上具有不同分辨率的视频数据,使不同终端可根据自己的终端处理能力和网络带宽,选择合适的视频码率进行交互,从而不受时间、空间、终端、带宽等限制开展视频会议,扩大了视频会议系统的应用范围,也为 3G 网络提供了更丰富的应用。

关键词:可伸缩视频编码;视频会议系统;异构网络;异型终端

中图分类号:TP393.09

文献标识码:A

引言

早期的视频会议系统基于固定的专用网络和硬件设施,使用者需要为此支付高昂的网络租赁和硬件维护费用。随着互联网的普及,现在的会议系统大量采用了 IP 网络技术,不再需要长期租用网络专线,极大地降低了成本,推动了视频会议系统的发展^[1]。

然而,不论是基于硬件还是基于软件的视频会议方案,对系统中多个会议终端的定义都是相同的,即假定其终端都是具有相同计算性能、相同显示性能以及相近网络带宽的设备,比较常见的是,所有终端都是互联网上相同类型的 PC 机(可能存在一些具体配置上的差异)。这种相同终端对视频会议系统在实际使用时带来了以下两个明显的不便之处:

第一,在召开视频会议时,需要所有与会人员具有差别不大的网络状况,否则的话,由于个别人员的网络状况过差,将导致整个会议系统变得不流畅;或者系统为了适应最差的网络情况,将浪费掉大部分人员良好的网络带宽,使整个会议的视频质量下降。第二,召开视频会议需要提前预约,不预约而临时召开视频会议的话,极易遇到与会人员不在自己的终端设备前,从而无法参加会议。

实际上,随着手持式设备性能的不断改善,手机、

IPAD 等设备已经具有了强大的多媒体处理能力,并得到了广泛的应用,如:手机电视、IPAD 上的流媒体视频点播等。3G 网络的普及也为手持式设备提供了更高的带宽,进一步促进了网络多媒体应用的发展^[2]。因此,如果能提供多种网络接入方式,允许用户使用不同的终端设备参加视频会议,将极大地改善会议系统的便利性。

第二,如果采用可伸缩视频编码(SVC)技术应用于视频会议系统,将能根据用户的终端设备性能以及接入网络的带宽自动为用户选择合适的视频质量、帧率、分辨率^[3]。如:高带宽的用户可以使用更高的帧率,分辨率,以及更好的画面质量,而低带宽的用户则使用较低的帧率、分辨率和低画面质量,从而在确保整个系统流畅运行的同时,充分利用了终端设备的性能和网络资源,给予用户不同的体验。

本文即是针对传统视频会议固有的缺陷,设计一种基于可伸缩视频编码的多终端视频会议系统,通过提供灵活方便的参与方式和适应不同的网络带宽,完全消除用户参加会议的工具、地点和场所限制。这意味着无论是公司、企业,还是事业单位,都能及时召开视频会议、实时处理问题,从而赢得宝贵的时间。

1 可伸缩视频编码技术 SVC

SVC(Scalable Video CODEC),即可伸缩视频编码技

术,该技术可通过一次编码生成具有不同时空分辨率以及不同质量的码流,在传输时动态地根据终端处理能力和网络状况,从原始码流中提取出部分码流进行传输,接收端可从收到的不完整码流中正确解码出特定质量的视频图像。最新的 SVC 技术定义于 H. 264/AVC 标准的附录 G,是 H. 264/AVC 标准的一个扩展。一个 SVC 码流由完全兼容于 H. 264/AVC 语法的基本层和附录 G 中定义的增强层构成^[4]。其中,基本层可独立传输,不能被截断,提供了可被接受的最低质量的视频流;增强层需要在基本层的基础上解码,可以在任意位置被截断,根据截断后实际传输的部分,可在时域,空域分辨率或图像质量上对基本层重建图像进行增强^[5]。相对于不可伸缩的视频编码,这种可在任意位置被截断的特点,使得 SVC 具有以下优点:(1)能够动态适应网络带宽的变化;(2)具有抵抗数据丢失的鲁棒性。(3)能够同时满足具有不同处理能力的终端用户的需求。

因此,使用 SVC 作为视频会议系统的核心,在保证视频质量的同时,可以提供多种灵活有效的伸缩方式,将极大提高视频会议参与用户的带宽灵活性,解决传统视频会议系统要求同型网络的缺点。

2 系统总体架构

基于可伸缩视频编码技术的多终端视频会议系统是一个综合软件系统,既涉及音视频编码实现方式,又涉及相关传输和控制技术,还必须对音视频的播放和管理做相应处理。为简化设计,系统采用以下三层体系架构,并分别对各层进行设计。

2.1 系统的三层架构体系

为保证底层核心技术的稳定,并可在不影响应用层模块的情况下进行关键技术升级,整个系统按三层设计,并采用模块化方式进行构架,结构如图 1 所示。

系统最底层为音视频编码层,包含 SVC 模块和音频编码模块,分别处理视频和音频的编码。

系统中间层为网络层,包含网络传输模块和网络监控模块,主要解决视频会议中对各终端网络状况的监控和实现音视频信息的传输。

系统最高层为应用层,主要面向用户提供友好界面和提供各种管理功能。

2.2 系统各层设计

系统音视频编码层、网络层、应用层具体设计如下:

2.2.1 音视频编码层

该层的 SVC 模块负责视频数据的编码、截取以及解码,音频编码模块使用 AAC 或 AC3 标准对声音数据进行编码和解码。这两个模块是完全独立的模块,对上层

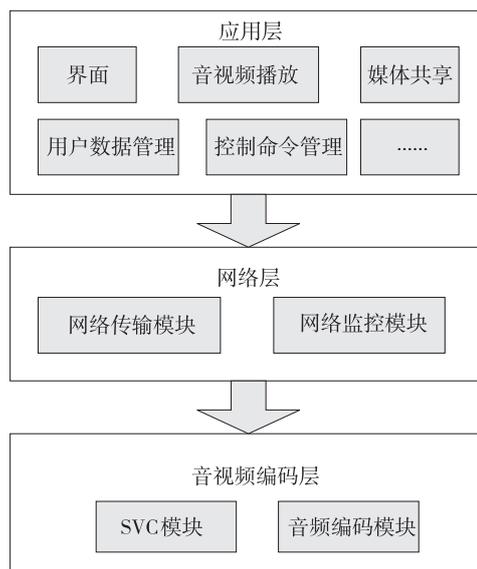


图 1 模块化方式进行构架

提供规范的调用接口,在改进各种核心算法后可直接进行升级替换。

SVC 模块是整个系统最核心和重要的模块,需要具有极高的编码效率和适应性,能在各种设备上完成实时编码^[6]。由于 SVC 技术的新颖性,目前尚无完全成熟的方案,需要独立进行技术研发以解决各种技术难题。有鉴于此,系统的解决方案是从两方面进行优化研究:一方面,在算法上参考 JVT 标准组织提供的 JSVM 参考模型深入掌握各种算法,同时,自主研究各种快速编码技术,如快速空域层编码方式技术,快速时域层运动矢量预测技术等,来提高编码效率,降低需要的复杂度;另一方面,参考一些高效的开源视频编解码方案,针对 X86, ARM 等平台充分使用多种多媒体汇编指令,如 MMX/SSE、NEON 等,发挥不同硬件平台的特性,进一步提高模块性能。

音频编码模块与 SVC 处于同一层面,但由于相对开销较小可直接采纳成熟的开源产品的解决方案,通过具体实现上的高度优化来满足性能上的需求,可不再进行更深入的算法级优化。

2.2.2 网络层

网络传输模块负责对音视频数据进行 RTP 协议的打包,将打包后的数据使用组播或点播方式传输给在线用户;同时,负责传输系统中的各种控制命令,对网络差错进行处理,并在用户之间反馈各种状态信息。RTP 方式的打包需要对协议进行充分研究,并针对会议系统在实际应用上的特点进行优化处理。此部分内容同样采用两种方式协同进行:一方面研究和分析协议本身;一方面通过对 live555 以及类似的一些解决方案进行研究,吸

收多种方案的优点。

网络监视模块负责监视各种异构情况的实时网络状况,需要针对各种可能的网络进行独立的协议研究,并将监控数据实时反馈到中央服务器或控制台以及用户终端,使得网络传输模块可以根据实际情况传输合适的数据。

2.2.3 应用层

应用层完成各种具体与用户交互的功能,并完成界面管理和各种数据处理功能。

为了提供美观且人性化的用户界面,方案使用微软提供的专业界面编写工具来编写界面。

在视频播放上,为了提高性能,使用 direct 3D 技术进行显示,发挥显卡的硬件性能,减小 CPU 开销。对音频播放,同样使用 direct sound 进行播放,通过声卡进行硬件处理。

对媒体共享功能,因为需要保持数据的完整性,使用 TCP 协议进行传输,并针对不同的媒体,在 TCP 协议上再制定私有的应用层协议。

对用户数据,如用户 IP 和端口,通过中央服务器进行集中管理,并使用定时广播的方式向所有在线用户进行广播。用户 IP 和端口无需固定,在每次用户登录时向服务器进行注册,使得用户可以不受时间和地点的限制使用会议系统。

3 关键技术设计思路

本系统的关键技术主要有:SVC 设计、自适应网络传输设计、音视频同步播放设计三方面,系统采用的设计思路如下:

3.1 可伸缩视频编解码设计

本系统在视频编码格式上以最新的可伸缩性视频编码标准——H.264/AVC 附录 G 定义的 SVC(Scalable Video Coding)作为核心,提供时域、空域和质量层次上的三层可伸缩性。

首先,在时域分级上,考虑到视频会议通常使用的帧率为 15fps^[6],为了尽量减小网络延时,本系统在基本层使用 7.5fps 的帧率,且在基本层只使用 I 帧和 P 帧,不使用 B 帧,以保证解码的零延时。在时域增强层,通过非参考的 B 帧提供 15fps 的帧率,既提高了时域分辨率,也由于 B 帧的非参考性,使得 B 帧中发生的错误不会扩散到基本层中,影响基本层的视频感受。终端对自己进行编码时,同时生成基本层和增强层码率,在传输时,则根据连接方对时域流畅度的要求选择不同层次的码流数据进行传输。在空域分级上,以 QCIF 分辨率为基本层,同时提供 CIF 格式的增强层分辨率。为了减

小计算的复杂度并增加抗差错能力,对时域基本层中的空域增强层,只使用空域层间预测,不使用帧间预测,即增强层图像只从基本层图像预测而来,不从前一帧增强层图像进行预测;对时域增强层中的空域增强层图像,则可自适应地进行时域预测或空域层间预测。在传输时,不同终端可根据自己的显示设备的分辨率选择不同层次的码流数据进行传输。在质量层次上,通过 MGS 方式提供中等颗粒度的可伸缩性,使得对特定的空域和时域分辨率,可进一步根据网络带宽对其质量增强层进行数据截断传输,以在可用的网络带宽下获得最好的图像质量。整个编码结构的预测方式如图 2 所示。

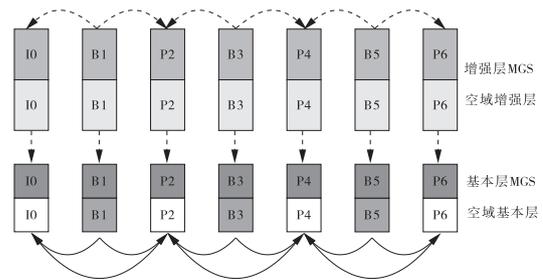


图 2 SVC 分层编码预测结构图

3.2 自适应网络传输设计

自适应网络传输从以下三个方面进行设计:

(1) 网络状况监控:在视频会议进行中,各终端通过 SNMP, QoS 等协议对自己的网络情况进行实时监控,并反馈到控制中心和其它终端。各终端根据自身的网络情况,实时地调整 SVC 编码参数,以生成不同码率范围内的基本层和增强层数据,如当有低带宽的用户加入会议时,降低基本层码率,当有高带宽用户加入会议时,加大最高增强层码率;控制中心则根据各终端的带宽情况,从每个视频流中提取特定部分的数据发送到各个终端,如,当用户在不其它网络应用时,可将全部带宽用于会议系统,控制中心从原始流中提取更多增强层数据进行发送;当用户打开其它网络应用消耗了部分带宽时,系统将减小增强层数据的发送或仅发送基本层数据以保证在带宽下降时也有流畅的效果,不影响会议进度。整个过程由系统在后台自主完成,无需用户参与。

(2) 音视频打包:系统根据 H.323 协议对提取的视频数据和编码生成的音频数据进行打包发送,并监控用户的参与情况。

(3) 传输与反馈:对打包后的数据按组播或点播方式进行传输,对其它需要保证正确的数据和命令进行可靠传输,并在系统所有用户间反馈网络状况。

3.3 音视频同步播放设计

平台使用摄像头采集视频数据,使用麦克风采集音

频数据。系统对采集到的音视频数据进行编码,对收到的音视频混合流进行解码,在不同平台上使用硬件设备进行图像和声音的显示和播放。

系统使用 RTP 协议将音视频流打包到一个传输流中进行传输,通过 RTP 协议对音视频数据进行同步。同时,在系统所有终端中使用私有协议对参会用户进行时钟同步,从而保证所有用户之间的数据能够同步。对音视频数据进行同步播放,对系统所有用户进行时间同步,保证发言和观看的同步。

4 结束语

本文采用先进的可伸缩编码、有线与无线网络交互技术,基于互联网构建新型的视频会议系统。本系统由于具有能按网络带宽调整视频编码和整合不同视频终端的特点,将彻底改变视频会议的传统模式,使得视频会议可以借助互联网不限时间、空间的开展,必将推动

视频会议系统的应用发展和创造较佳的经济效益。

参考文献:

- [1] 曹洁,马晓昱,龚光武.网络视频会议系统研究与设计[J].计算机与现代化,2010,181(9):66-72.
- [2] 吉宁.3G 可视终端在视频会议系统中的应用研究[J].江苏通信,2011(8):48-51.
- [3] 郭春辉.基于 H.264 SVC 的可分级视频会议系统的实现[J].电视技术,2011,35(7):111-113.
- [4] 向友君,张克新.完全可伸缩视频编码的实现和改进[J].计算机应用研究,2009,26(9):3578-3580.
- [5] 毛年胜,卓力.基于 H.264 SVC 的 IP 网络视频传输系统的实现[J].测控技术,2010,29(5):5-8.
- [6] 任勇.SVC 技术在网络摄像机的应用与分析[J].中国安防,2009,46(12):45-48.

Video Conference System of Multiple Terminal Based on SVC

YE Xiao-lei¹, LIANG Jin-ming², DENG Yun²

(1. Computer Center, Sichuan Salt Industry School, Zigong 643000, China;

2. School of Computer Science, Sichuan University of Science & Engineering, Zigong 643000, China)

Abstract: According to defects of the traditional video conference with requirements of the same type of terminal and network, a video conference framework based on the scalability video codec technology and cable and wireless network interactive technology is proposed. The proposed system uses SVC to generate video stream with scalable temporal and spatial resolution, which enable heterogeneous terminals to choose difference bit-rate to join the conference according to their process ability and bandwidth, not restricted by time and locations. The system can promote the video conference system to be widely used, also provides more rich applications for 3G network.

Key words: scalable video codec; video conference system; multiple terminal