



Visual C++

视频会议开发技术与实例

吴志军 马 兰 沈笑云 编著

关键技术篇

基础篇

- 视频会议系统简介
- 视频会议系统的详细设计

- 视频会议系统的视频编码标准
- 视频会议系统的视音频编码分析
- 视频会议系统的文件传输
- 视频会议系统的文字讨论
- 视频会议系统的安全保密

案例篇

- 基于实时通信 (RTC) API 的视频会议系统
- 基于VFW的视频会议系统
- 基于VIS H.323 DLL Library的视频会议系统

Visual C++

视频会议开发技术与实例

吴志军 马 兰 沈笑云 编著

 人民邮电出版社
OSTS & TELECOM PRESS

图书在版编目 (CIP) 数据

Visual C++视频会议开发技术与实例/吴志军, 马兰, 沈笑云编著.

—北京: 人民邮电出版社, 2006.1

ISBN 7-115-14257-2

I. V... II. ①吴...②马...③沈... III. C 语言—程序设计 IV. TP312

中国版本图书馆 CIP 数据核字 (2005) 第 157753 号

内 容 提 要

本书全面、详细地介绍了 Visual C++视频会议开发的基本原理和编程知识, 并结合实际视频会议开发的工程实际, 给出了丰富的实例和大量的 Visual C++源代码文件。

全书共分 3 篇 10 章, “基础篇”(第 1 章和第 2 章)介绍了如何对视频会议系统进行相关的需求分析, 以及视频会议的相关标准和体系结构等; “关键技术篇”(第 3 章至第 7 章)介绍了视频会议系统的视频编码标准、视音频编码分析、文件传输、文字讨论和安全保密技术等; “案例篇”(第 8 章至第 10 章)给出了 3 个综合的视频会议系统, 以供读者边学习边实践。

本书内容由浅入深, 从最基本的视频会议相关标准一直到综合实例的实现, 都有详细的讲解。配套光盘中还提供了书中实例的全部源代码, 方便读者举一反三, 开发出适合自己的工程项目。

本书适合从事视频会议开发的人员进行学习和参考, 同时, 可作为高等院校相关专业课程设计和毕业设计的学习和参考用书。

Visual C++视频会议开发技术与实例

◆ 编 著 吴志军 马 兰 沈笑云

责任编辑 汤 倩

◆ 人民邮电出版社出版发行 北京市崇文区夕照寺街 14 号

邮编 100061 电子函件 315@ptpress.com.cn

网址 <http://www.ptpress.com.cn>

北京鸿佳印刷厂印刷

新华书店总店北京发行所经销

◆ 开本: 787×1092 1/16

印张: 22

字数: 538 千字

印数: 1-5 000 册

2006 年 1 月第 1 版

2006 年 1 月北京第 1 次印刷

ISBN 7-115-14257-2/TP · 5128

定价: 39.00 元 (附光盘)

读者服务热线: (010) 67132692 印装质量热线: (010) 67129223

前言

Foreword

视频会议系统是集计算机技术、通信技术和多媒体技术于一体的远程、异地通信方式。就像 19 世纪的电报和 20 世纪的电话一样，视频会议是 21 世纪多媒体通信时代的典型代表。

视频会议通过网络把两个或多个地点的多媒体会议终端连接起来，在其间传送各种图像、话音和数据信号，使出席会议者有亲临现场的感觉。视频会议能为用户提供直接、全面的沟通交流，并能节约时间、降低成本、提高生产率。目前，国内外很多科研机构和厂商都在进行多媒体多点通信的研究，并推出了各自的视频会议系统。在研究各种视频会议系统的基础上，国际电信联盟（ITU-T）形成了视听多媒体通信系统的国际标准，使不同厂商的多媒体通信产品能够互通，推动了多媒体通信技术标准化的进程。

● 为什么编写本书

目前，作为新型技术的视频会议，已经成为开发和研究的热点。许多企业的技术人员正在考虑本企业视频会议系统的具体方案；计算机和通信等专业的师生，也将视频会议相关技术的学习和研究列为课程设计和毕业设计的内容。

本书旨在帮助从事视频会议开发的人员，全面、深入地学习和掌握视频会议的相关标准、技术要点和开发方法等，并通过案例的学习，开发出符合实际需要的视频会议系统。

● 本书特色

- ✓ 本书共分 3 篇 10 章，全面介绍了 Visual C++ 视频会议开发的相关知识，包括了基础知识、关键技术和综合案例等，方便读者快速地找到自己感兴趣的和对自己有帮助的内容。
- ✓ 内容的安排上，依据视频会议系统的组成结构和功能划分，逐层深入。特别对 MPEG-4 编解码、H.263 和 H.264 标准的实现等视频会议中的关键技术，进行了较为详细的讲解。
- ✓ 光盘中提供了书中涉及实例的完整源代码，方便读者举一反三，开发出适合自己的工程项目

● 技术支持

吴志军负责了全书的策划和统筹，并负责编写了第 1 章、第 4 章和第 7 章的内容，还对全书进行了统稿；马兰编写了第 5 章、第 8 章至第 10 章的内容；沈笑云编写了第 2 章至 3 章和第 6 章的内容，并负责最终修订；武朝东、陈志峰、刘颖、姚彬和左倩倩等人，不辞辛苦地收集和整理了大量的资料，提供了很好的素材，并参与了部分章节的编写。

另外，感谢万康视通的大力支持，为本书提供了珍贵的资料和程序代码。感谢“115 软件工厂”（115Studio）钱斌和蒲宇提供的 VIS H.323 DLL Library V2.3 Release FOR Windows

98/Me/2000/XP 的资料和代码。感谢章泽昂提供的用 MFC 实现局域网内点对点的大文件传输的代码。感谢王剑对本书初稿的审阅，并提出了宝贵意见。

在编写过程中，我们力求精益求精，但由于视频会议系统涉及的研究领域广泛，内容新颖，技术难点多，难免存在一些不足之处，恳请广大读者批评指正，并请发 E-mail 至 tangqian@ptpress.cn 与我们联系。

编 者

目 录

Contents

第一篇 基 础 篇

第 1 章 视频会议系统简介	2
1.1 视频会议系统的现状	2
1.1.1 视频会议系统的概念	2
1.1.2 视频会议的发展	2
1.1.3 视频会议系统的类型	3
1.1.4 视频会议系统的组成	4
1.2 视频会议系统的相关标准	4
1.2.1 H.320 和 H.323 技术标准比较	4
1.2.2 SIP 和 H.323 技术标准比较	5
1.2.3 视频会议系统的视频编码技术	7
1.2.4 视频会议系统的音频编码技术	8
1.2.5 视频会议采用的框架标准	9
第 2 章 视频会议系统的详细设计	11
2.1 视频会议系统设计原则	11
2.2 视频会议系统的需求分析	12
2.2.1 承载网络分析	12
2.2.2 终端功能分析	12
2.2.3 数据管理功能分析	13
2.2.4 实例——“远程教学/会诊视频会议系统”需求分析	14
2.3 视频会议系统的体系结构	15
2.3.1 基本框架结构	15
2.3.2 H.323 会议的组织形式	18
2.3.3 系统硬件配置	19
2.3.4 系统功能设计	21
2.3.5 会议功能设计	22
2.3.6 会议安全性设计	23
2.3.7 会议管理功能设计	24
2.3.8 实例——基于 Web 管理的视频会议功能分析	24

第二篇 关键技术篇

第 3 章 视频会议系统的视频编码标准	36
3.1 H.261 视频编码标准	36
3.2 H.263 视频编码标准	36
3.2.1 概述	37
3.2.2 信源编码器	38
3.2.3 信源编码算法	40
3.2.4 信源编码控制	41
3.3 H.264 视频编码层次结构	42
3.3.1 H.264 的组成	42
3.3.2 网络提取层 (NAL)	42
3.3.3 视频编码层 (VCL)	45
3.4 H.264 的核心技术	48
3.4.1 类和级及其应用	48
3.4.2 H.264 的优化	49
3.5 H.264 在视频会议中的应用	72
3.5.1 在 H.323 系统中实现 H.264	72
3.5.2 H.264 编码实现	73
3.5.3 H.264 解码实现	92
第 4 章 视频会议系统的视音频编码分析	97
4.1 MPEG-4 编码技术	97
4.1.1 MPEG 系列标准	97
4.1.2 MPEG-4 编码源代码详细分析	98
4.2 MPEG-4 解码技术	125
4.2.1 MPEG-4 的数据结构定义	125
4.2.2 解码器数据结构的定义	127
4.2.3 解码器的初始化	130
4.2.4 帧解码	139
4.3 视频会议系统的音频编码	151
4.3.1 CELP 标准简介	151
4.3.2 G.721 代码实现	151
4.3.3 G.723 代码实现	153
4.3.4 G.728 (LD-CELP) 代码实现	156
4.3.5 G.729 (CS-ACELP) 代码实现	181

第 5 章 视频会议系统的文件传输	183
5.1 FTP 文件传输	183
5.1.1 FTP 概述	183
5.1.2 实例——基于 FTP 实现网络文件传输	184
5.2 P2P 文件传输	208
5.2.1 P2P 概述	208
5.2.2 实例——基于 P2P 实现网络文件传输	209
第 6 章 视频会议系统的文字讨论	213
6.1 系统设计	213
6.1.1 需求分析	213
6.1.2 总体设计	213
6.1.3 通信流程	213
6.2 私人文字讨论功能模块	215
6.2.1 私人文字讨论效果设计	215
6.2.2 私人文字讨论功能实现	216
6.3 多人文字讨论功能模块	225
6.3.1 多人文字讨论效果设计	225
6.3.2 多人文字讨论功能实现	229
第 7 章 视频会议系统的安全保密	237
7.1 视频会议的安全机制	237
7.1.1 密码系统概要	237
7.1.2 IP 多播视频会议的安全机制	238
7.1.3 基于 SIP 的视频会议安全机制	241
7.2 视频会议系统的安全保密方法	243
7.3 视频会议中的加密算法	246
7.3.1 数据加密标准 (DES)	246
7.3.2 RSA 公钥密码体制	253
7.3.3 MD5 算法及其实现	260

第三篇 案 例 篇

第 8 章 基于实时通信 (RTC) API 的视频会议系统	276
8.1 视频通信实现步骤	277
8.1.1 建立实时通信客户端接口	278
8.1.2 选择通信类型	278

8.1.3 调整通信设备	279
8.2 应用程序的初始化操作	280
8.3 实时通信事件处理	292
8.3.1 通信会话创建	293
8.3.2 实时流会话事件	293
第 9 章 基于 VFW 的视频会议系统	297
9.1 VFW 简介	297
9.2 视频实现	299
9.2.1 Video Capturer 的主要结构	299
9.2.2 基本的捕获设置	300
9.2.3 捕获窗口	302
9.2.4 视频捕获驱动	303
9.3 录音实现	307
9.4 程序运行演示	314
第 10 章 基于 VIS H.323 DLL Library 的视频会议系统	316
10.1 接口函数	316
10.1.1 CaudioMixer 类函数	316
10.1.2 CaudioPlayer 类函数	320
10.1.3 CaudioRecorder 类函数	321
10.1.4 CvideoMixer 类函数	322
10.2 客户端代码分析	327

第一篇

基础篇

- 视频会议系统的现状
- 视频会议系统的相关标准
- 视频会议系统设计原则
- 视频会议系统的需求分析
- 视频会议系统的体系结构

第 1 章 视频会议系统简介

1.1 视频会议系统的现状

随着人类经济、政治及文化活动的不断发展和科学技术的进步，人与人之间的信息交流愈来愈频繁。为此人们对通信方式提出了越来越多的需求。声音和图像具有直观性强、信息内容丰富等特点。在人类的日常生活中，大部分信息是通过听觉和视觉来获取的。因此，通信中引入了声音和图像信息的传输，特别是把图像、声音等信息综合起来，不仅大大丰富了通信的内容，而且也更适合人们获取外部信息的特点。也正因为如此，在诸多的通信业务中，视频会议是近年来发展较快的一项新业务。视频会议是网络时代出现的新型的会议方式，它的数据和图像传送功能是传统的会议方式做不到的。

在视频会议技术出现前，人们要实现异地召开会议，只能利用会议电话技术，参会人员彼此只能听见声音。而人们彼此通信交流有 3 个基本要素：内容、表述方式和视频。其中视频因素是最有力的，任何人类交流的应用都因引入视觉的因素而使效果大大提高。视频会议正好满足了这方面的要求。

1.1.1 视频会议系统的概念

视频会议系统（Video Conference）是指两个或两个以上不同地方的个人或群体通过传输线路及多媒体设备，将声音、影像及文件资料互传，达到即时且互动的沟通。

视频会议系统是集通信、计算机技术、多媒体技术于一体的远程异地通信方式。在召开视频会议时，处于两地或多个不同地点的与会代表，既可以听到对方的声音，又能看到对方的形象，同时还能看到对方会议室的场景，以及会议中展示的实物、图片、表格、文件等，与真实的会议无异，使每个与会者有身临其境之感。在实际应用中，视频会议能够提高工作效率，降低远距离会议的费用。

1.1.2 视频会议的发展

视频会议大致经历了以下几个发展阶段。

（1）20 世纪 60 年代至 80 年代的模拟技术视频会议。早在 20 世纪 60 年代，就已经有了视频会议系统的研究，当时美国电报电话公司（AT&T）曾推出模拟可视电话系统（Picture Phone）。1964 年，世界上最早的模拟技术可视电话——Picture Phone MOD-I 诞生于美国贝尔实验室，采用了图像和话音分时传输的方法，传输带宽为 300~3400Hz。后来经过改进成为 Picture Phone MOD-II，带宽高达 1MHz，并用于匹兹堡和芝加哥之间的商业通信，但因为这种模拟传输的可视电话系统费用昂贵，商用业务最终被迫停止。

进入 20 世纪 70 年代，由于相关技术领域的不断进步，最主要是数字式传输的出现，传

统视频会议系统所用模拟信号的采样和变换方法得到极大改善,数字信号处理技术逐渐成熟。对数据压缩的研究生成了视频会议产品的核心技术—CODEC(Coder/Decoder)。

进入 20 世纪 80 年代中期,由于 CODEC 制造技术的提高,体积在急剧减小,与此同时,数字式网络发展也非常迅速,开始出现更低速率的网络服务。另外,技术的进步也使得 CODEC 价格大幅下降。视频会议系统逐步进入市场。但当时只限于高档的会议室视频会议系统的应用。

(2) 20 世纪 90 年代初至 1995 年基于 ISDN 的数字视频会议。这一阶段的前期是基于 SDH、DDN 网视频会议,后期主要是基于 ISDN 网视频会议。20 世纪 90 年代初期,第一套国际标准 H.320 获得通过,不同品牌之间的兼容性问题得到了解决。视频会议系统朝着小型化发展,移动型视频会议系统成为视频会议应用中的主流产品。PC 应用于视频会议领域,使得贵族化的技术具有了平民化的价格。

(3) 1995 年以后基于 IP 网的数字视频会议。20 世纪 90 年代后期,随着 PC 的快速升级,数据存储和处理能力越来越强,信息压缩技术发展快速,进一步推动了视频会议系统的发展。目前视频会议技术正在向 Internet 网络上的多媒体会议方向发展。

随着通信网络运营商宽带业务的拓展以及在保证网络服务质量情况下对核心宽带骨干网技术的应用,可以预见,网络传输能力将进一步提高,从而促进视频会议系统的快速发展。视频会议系统未来的发展趋势有以下几点。

(1) 编解码方式由硬件向软件转化。由于计算机处理速度和附属板卡的处理速度提高,许多需要专用设备进行的数据处理过程可以交由计算机及其内置的通用板卡来完成,而在效果上没有太大的区别。

(2) 协议类型从 H.320 向 H.323 转化。在初期,由于 IP 协议应用不是很普遍,基于 ISDN 线路,符合 H.320 协议的产品占主要比例。随着近几年 IP 协议的普遍应用和互联网的高速发展,特别是对于宽带网络的应用,符合 H.323 协议的产品在整体市场中占有的比例急剧上升。

(3) 市场重心向低价位普通用户市场转移。由于网络条件、技术水平、生产工艺等因素的变化,视频会议系统已经从高价位专用市场向低价位普通用户市场转移。现在包括 POLYCOM、VCON 等大型公司都有相当的产品定位于中小企业用户市场中。

(4) 流媒体广播技术越来越多的应用于视频会议系统。大部分广播会议,如网上路演、采访、现场会等,能够为用户传输多媒体信息,是对传统视频会议功能的扩展。

(5) 虚拟合作应用服务成为视频会议面向企业的核心功能之一。大多数跨国集团和分支机构较多的企业在使用视频会议系统时,除了会议功能外,使用最为频繁的是虚拟合作,如远程的项目管理、客户服务、技术培训等,这种应用在知识型企业中尤为明显。

1.1.3 视频会议系统的类型

根据运行环境和支持标准等视频会议系统可以划分为以下不同的类型。

(1) 根据会议节点数目不同,视频会议系统分为点对点视频会议系统和多点视频会议系统。点对点视频会议系统应用于两个通信节点间。多点视频会议系统应用于两个以上节点之间的通信。

(2) 根据运行的通信网络不同,视频会议系统分为数字数据网(DDN)或其他专用网型、局域网/广域网型(LAN/WAN)和公共交换电话网型(PSTN)3种。使用 DDN 或专用网,在 384~2048kbit/s 速率下,可提供 25~30 帧/s 的 CIF 或 QCIF 图像;在 LAN/WAN 环境中,

运行在 384kbit/s 速率下, 可提供每秒 15~20 帧/s 图像; 而在 PSTN 上, 运行在 28.8kbit/s 或 33.6kbit/s 等速率下, 只能达到 5~10 帧/s。

(3) 根据技术支持类型的不同, 视频会议系统可分为基于线路的视频会议系统和基于分组的视频会议系统, 但是现在两者间的界限已经越来越模糊。基于线路的视频会议系统, 也称为常规视频编解码系统, 依照专用线路提供一个确定的比特率, 诸如租用线路或公用线路交换服务。基于分组的视频会议系统是从分组视频通话系统演化而来的, 其基本原理是相同的, 即利用桌面计算机支持视频会议的服务。

(4) 根据所选用的终端类型不同, 视频会议系统又可分为桌面视频会议系统 (Desktop Video Conference)、会议室型视频会议系统 (Room/ Roll-about Video Conference) 和可视电话系统。

桌面型视频会议系统是在普通计算机上增加一些附加设备, 主要利用计算机软件完成会议功能。桌面视频会议日益受到青睐, 因为它有效地利用了现有的资源。

会议室视频会议系统在带有环境控制设备的专用会议房间里装置一个或多个大屏幕, 系统由屏幕、摄像机、麦克风和辅助设备等组成。这些配置是永久性的, 它们不能移到别的房间或大楼内, 但可以提供高质量的视频和同步音频。

可视电话系统用于点到点通信, 它满足了在电话上进行视频会议传输的需求。系统组成包括一个小屏幕、内部摄像机、视频编解码器、音频系统和键盘。

1.1.4 视频会议系统的组成

典型的多媒体视频会议系统由终端设备、通信链路、多点控制单元 (MCU, Multi-point Control Unit) 及相应的软件部分组成。

(1) 终端设备。终端设备不仅要完成各自的数据处理任务, 还要并行完成多媒体通信协议的处理、音视频信号的接收、存储与播放, 并记录和检索大量与会议相关的数据与文件。终端设备的硬件配置包括音、视频信号处理器, 压缩与解压缩卡, 以及摄像机、话筒、扬声器、电子书写板、图像扫描仪和通信网卡等。

(2) 通信链路。通信链路的选择有很多种, 包括 PSTN、LAN、WAN、N-ISDN、Frame Relay 或者 B-ISDN、ATM 等。

(3) MCU。MCU 是视频会议系统的核心设备, 它是一个数字处理单元, 通常设在网络节点 (汇接局) 处, 用于处理多个地点同时进行通信, 其主要功能是将各终端送来的信号进行分离, 抽取出音频、视频、数据和信令信号, 分别送到相应的处理单元, 进行音频混合或切换、数据广播和确定路由选择、定时和处理会议控制等。

(4) 软件部分。软件部分包括协议处理、会议服务、音频与视频信号处理、协同工作管理和图形用户接口等。国外常见的视频会议软件有 CU-SeeMe (由美国 Cornell 大学开发)、IVS (INRIA Videoconferencing System, 由法国 INRIA Sophia Antipolis 开发) 和 ShowMe (由 Sun Microsystems 公司开发) 等。

1.2 视频会议系统的相关标准

1.2.1 H.320 和 H.323 技术标准比较

1997 年 3 月 ITU-T (国际电信联盟电信标准化部门) 发布了用于局域网上的视频会议标

准协议——H.323, 为与 Internet 和 Intranet 相连的视频会议系统提供了互通的标准, 各厂商也纷纷推出符合该标准的视频会议产品。在此以前, 用于 ISDN 上的视频会议标准协议——H.320 一直主导着视频会议领域的技术和产品发展。两者的不同点主要在以下几方面。

(1) 组网结构。H.323 是总线型网络结构, 不会因为某一个终端出现临时故障而影响整个会议和网络。H.320 是主从星形汇接结构, 可能因为单点临时故障, 而又没有重要节点上的容错备份机制, 导致运行不正常。

(2) 业务发展。H.320 仅对基于电路交换的电视会议系统进行了定义, 因此, 仅能在传输网络平台上开展标准的电视会议应用, 而不能扩展为一个多媒体、多应用平台。

在网络上, H.323 技术可以开发出许多与底层网络传输无关的多媒体应用, 如多媒体视频会议、多媒体监控、多媒体生产调度指挥、远程企业培训和教育、多媒体呼叫中心、网络 IP 电话、网络 IP 传真、网络视频点播和广播等, 可以利用 H.323 技术将多种应用和业务叠加到同一个网络传输平台上。

(3) 性能价格比。H.320 由于受传统电视系统会议技术体制的限制, 不具备灵活性和丰富的功能。而且, 无论是 H.320 的终端还是 H.320 MCU, 它的用户单机成本和用户线路使用费用都较高。H.323 采用了 TCP/IP 技术, 在提供相同性能和更多功能的同时, 大大降低了用户终端的成本以及用户线路使用费用, 具有较高的性能价格比。

(4) 数据功能。H.320 系统运用 T.120 标准来实现数据会议功能(如电子白板、文件传送、应用共享), 其数据应用是包含在 H.320/H.221 复用信道中的。H.320 系统的 T.120 数据信道最高达到 64kbit/s。在传送大容量文件、高质量图文时, 由于视频信道被抢占, 会出现图像质量下降和图像暂停现象。

H.323 标准沿用 T.120 体系下的数据会议标准来实现数据应用功能, 但它的数应用是独立于 H.323 会话进程的, 其数据信道不需要经过复用过程, 直接在 TCP 或 UDP(广播时) 开立单独的 T.120 数据信道, 此信道带宽可以从几 kbit/s 到数 Mbit/s 或数十 Mbit/s 可调, 表现了很大的优越性和灵活性。

(5) 多点广播。H.323 是基于 TCP/IP 的, 而 IP 具备多点广播功能——IP Multicast (RFC1112), 能够在网上轻松实现多媒体广播业务, 如视频广播。

H.320 本身不具备多点广播功能, 而且没有有效的下层协议支持。所以, H.320 系统不具备多点广播功能, 即建立广播频道。它可能会借助 MCU 用交互多点实现准广播功能, 而非广播频道。

尽管 H.320 和 H.323 系统采用的是同一种图像和语音的编解码技术, 在图像和声音的表现质量和还原上本质是相同的, 但从以上的诸多优势对比不难看出, H.323 协议标准是代表多媒体视频会议以及其他网上多媒体应用的发展方向和潮流的。H.320 作为传统的技术标准, 由于本身固有的局限性和高成本, 已经逐渐退出历史舞台。

1.2.2 SIP 和 H.323 技术标准比较

1. SIP 简介

SIP (Session Initiation Protocol) 称为会话发起协议, 是由 IETF (Internet Engineering Task Force) 于 1999 年提出的在基于 IP 的网络中, 特别是在 Internet 结构的网络环境中, 实现实时通信应用的一种信令协议。而所谓的会话 (Session), 就是指用户之间的数据交换。在基于 SIP 的应用中, 每一个会话可以是各种不同的数据, 可以是普通的文本数据, 也可以是经

过数字化处理的音频、视频数据，还可以是诸如游戏等应用程序的数据。

2. SIP 的基本组成

按逻辑功能区分，SIP 系统包括 4 方面：用户代理、SIP 代理服务器、重定向服务器以及 SIP 注册服务器。

(1) SIP 用户代理，又称为 SIP 终端，是 SIP 系统中的端用户，在 RFC3261 中将它们定义为一个应用。根据在会话中扮演的角色的不同，又可分为用户代理客户机（UAC）和用户代理服务器（UAS）两种。前者用于发起呼叫请求，后者用于响应呼叫请求。

(2) SIP 代理服务器（SIP Proxy Server），是一个中间元素，它既是一个客户机又是一个服务器，具有解析名字的能力，能够代理前面的用户向下一跳服务器发出呼叫请求。然后服务器决定下一跳的地址。

(3) 重定向服务器（Redirect Server），是一个规划 SIP 呼叫路径的服务器，在获得了下一跳的地址后，立刻告诉前面的用户，让用户直接向下一跳地址发出请求而自己则退出对这个呼叫的控制。

(4) SIP 注册服务器（SIP Register Server），用来完成对 UAS 的登录，在 SIP 系统的网元中，所有 UAS 都要在某个登录服务器中登录，以便 UAC 通过服务器能找到它们。

3. SIP 和 H.323 的比较

同 SIP 的应用目的类似，H.323 同样对基于包交换的网络上的多媒体通信应用做出了相关定义。由于 H.323 和 SIP 标准最初是由两个面对行业领域有所不同的标准组织所提出的，虽然目的都是为实现多媒体通信，但由于不同的应用目标，决定了这两个标准有其各自不同的特点，不能简单地将二者对立起来，指出哪个标准更先进、更科学。

下面针对它们的应用目标、体系结构、系统组成以及系统实现的难易程度等几个方面进行简单分析。

(1) 应用目标。H.323 的应用目标是，在基于 IP 的网络环境中，实现可靠的面向音视频和数据的实时应用。SIP 的应用目标是在基于 Internet 的环境中，实现数据、音视频实时通信，特别是通过 Internet 将视频通信这种应用大众化，引入到千家万户。相对于 H.323 而言，SIP 协议简单、自由，使用相对小的成本就可以构造满足应用的系统。例如仅仅使用微软基于 SIP 的 MSN 和 RTC，就可以构造一个简单的，基于 Internet 应用环境的视频通信环境。

(2) 体系结构。H.323 不是一个单一标准，而是一个关于在 IP 环境中实时多媒体应用的完整标准族，对于呼叫的建立、管理以及所传输媒体格式等各个方面都有完善而严格的规定。SIP 是一个实现实时多媒体应用的信令标准，由于它采用了基于文本的编码方式，使得在点到点的应用环境中，具有极大的灵活性、扩充性以及跨平台使用的兼容性。但是 SIP 自身不支持多点的会议功能以及管理和控制功能，而是要依赖于别的协议，影响了系统的完备性，特别是对于需要多点通信的要求，应用单纯的 SIP 系统难以实现。

(3) 系统的组成。SIP 的 UA 等价于一个 H.323 的终端，实现呼叫的发起和接收，并完成所传输媒体的编解码应用；SIP 代理服务器、重定向服务器以及注册服务器的功能则等价于 H.323 的 Gatekeeper，实现了终端的注册、呼叫地址的解析以及路由。

虽然在呼叫信令和控制的具体实现上不同，但一个基于 SIP 的呼叫流程与 H.323 的 Q931 相类似，SIP 所采用的会话描述协议（SDP）则类似于 H.323 中的呼叫控制协议 H.245。

(4) 系统实现的难易性。H.323 标准的信令信息是采用符合 ASN.1per 的二进制编码，并且在连接实现全过程都要严格标准的定义，系统的自由度小，如要实现大规模的应用，需要

对整个网络的各个环节进行规划。SIP 标准的信令信息是基于文本的,采用符合 ISO10646 的 UTF-8 编码,并且全系统的构造结构相对灵活,终端和服务器的实现相对容易,成本也较低,从网络运营商的角度考虑,构造一个大规模视频通信网络,采用 SIP 系统的成本要廉价许多,而且也更具有可实现性。

通过对 SIP 和 H.323 协议之间进行比较,不难看出, H.323 和 SIP 之间不是对立的关系,而是在不同应用环境中的相互补充。SIP 作为以 Internet 应用为背景的通信标准,是将视频通信大众化。而 H.323 系统和 SIP 系统有机结合,又确保了用户可以在构造相对廉价的 SIP 视频系统的基础上,实现多方会议等多样化的功能。

1.2.3 视频会议系统的视频编码技术

信源的压缩始终是通信领域研究的核心问题之一。对于视频业务而言,压缩技术显得尤其重要,因为视频信号的数据量很大,标准的 PAL 信号的速率为 216Mbit/s。如果不压缩基本上不可能传输,压缩到 2Mbit/s 时也只能在通信干线上传输,不可能扩展到终端用户。因此需要提高压缩效率,一方面通过图像质量的降级,但主要还是提高编码效率。

国际上有两个负责音视频编码的标准化组织,一个是 MPEG (Motion Picture Expert Group), 国际标准化组织下的运动图像专家组。另一个是 VCEG (Video Code Expert Group), 是国际电信联合会下的视频编码专家组。以上两个标准化组织制定的相关编码标准都获得了广泛的应用。VCEG 制定的标准有 H.261、H.262、H.263 等, 这些标准成为视频会议的视频压缩标准, 最新推出的 H.264 是为新一代交互视频通信制定的标准。MPEG 制定的标准有 MPEG-1、MPEG-2、MPEG-4、MPEG-7 和 MPEG-21。MPEG-1 即俗称的 VCD, MPEG-2 则为 DVD 所采用, MPEG-4 是为交互式多媒体通信制定的压缩标准, MPEG-7 是为互联网视频检索制定的压缩标准。

1. MPEG-4 协议

MPEG 组织于 1999 年 2 月正式公布了 MPEG-4 (ISO/IEC14496) 标准第一版本。同年年底 MPEG-4 第二版发布, 且于 2000 年年初正式成为国际标准。

MPEG-4 与 MPEG-1 和 MPEG-2 有很大的不同。MPEG-4 不只是具体压缩算法, 它是针对数字电视、交互式绘图应用 (影音合成内容)、交互式多媒体 (WWW、资料摄取与分散) 等整合及压缩技术的需求而制定的国际标准。MPEG-4 标准将众多的多媒体应用集成于一个完整的框架内, 旨在为多媒体通信及应用环境提供标准的算法及工具, 从而建立起一种能被多媒体传输、存储、检索等应用领域普遍采用的统一数据格式。

MPEG-4 标准同以前标准的最显著的差别在于采用基于对象的编码理念, 即在编码时将一幅景物分成若干在时间和空间上相互联系的视频音频对象, 分别编码后, 再经过复用传输到接收端, 然后对不同的对象分别解码, 从而组合成所需要的视频和音频。这样既方便对不同的对象采用不同的编码方法和表示方法, 又有利于不同数据类型间的融合。例如, 可以将一个卡通人物放在真实的场景中, 或者将真人置于一个虚拟的演播室里, 还可以在互联网上方便的实现交互, 根据自己的需要有选择的组合各种视频音频以及图形文本对象。

2. H.264 协议

H.264 是 ITU-T 的 VCEG (视频编码专家组) 和 ISO/IEC 的 MPEG (运动图像专家组) 的联合视频组 (JVT, Joint Video Team) 开发的新的数字视频编码标准, 它既是 ITU-T 的 H.264, 又是 ISO/IEC 的 MPEG-4 的第 10 部分。1998 年 1 月份开始草案征集, 1999 年 9 月, 完成第

一个草案,2001年5月制定了其测试模式 TML-8,2002年6月的 JVT 第5次会议通过了 H.264 的 FCD 板。

H.264 和以前的标准一样,也是 DPCM 加变换编码的混合编码模式。但它采用“回归基本”的简洁设计,不用众多的选项,获得了比 H.263+好得多的压缩性能;加强了对各种信道的适应能力,采用“网络友好”的结构和语法,有利于对误码和丢包的处理;应用目标范围较宽,以满足不同速率、不同解析度以及不同传输(存储)场合的需求;它的基本系统是开放的,使用无需版权。

H.264 的算法在概念上可以分为两层:视频编码层 VCL (Video Coding Layer) 负责高效的视频内容表示,网络提取层 NAL (Network Abstraction Layer) 负责以网络所要求的恰当的方式对数据进行打包和传送。H.264 支持 1/4 或 1/8 像素精度的运动矢量。在 1/4 像素精度时可使用 6 抽头滤波器来减少高频噪声,对于 1/8 像素精度的运动矢量,可使用更为复杂的 8 抽头的滤波器。在进行运动估计时,编码器还可选择“增强”内插滤波器来提高预测的效果。H.264 草案中包含了用于差错消除的工具,便于压缩视频在误码、丢包多发环境中传输,如移动信道或 IP 信道中传输的健壮性。

H.264 具有广阔的应用前景,例如实时视频通信、因特网视频传输、视频流媒体服务、异构网上的多点通信、压缩视频存储、视频数据库等。H.264 优越性能的获得不是没有代价的,其代价是计算复杂度的大大增加,据估计,编码的计算复杂度大约相当于 H.263 的 3 倍,解码复杂度大约相当于 H.263 的 2 倍。

H.264 的技术特点可以归纳为 3 个方面,一是注重实用,采用成熟的技术,追求更高的编码效率,简洁的表现形式;二是注重对移动和 IP 网络的适应,采用分层技术,从形式上将编码和信道隔离开来,实质上是在源编码器算法中更多地考虑到信道的特点;三是在混合编码器的基本框架下,对其关键部件做了重大改进,如多模式运动估计、帧内预测、多帧预测、统一 VCL、4×4 二维整数变换等。

1.2.4 视频会议系统的音频编码技术

H.323 框架协议下的 G 系列标准协议 G.711、G.722、G.723、G.728 等,是 ITU 为语音通信推荐的系列协议标准,音频占用 64kHz 以下带宽,音质相当于电话效果(G.722 能达到调频广播的音质)。

目前,大多数视频会议产品都采用 G 系列标准协议。但国内有部分厂家针对国内市场的需求,为提高视频会议的声音质量,采用了 MP3 音频编码。相比较而言,G 系列标准是主流的视频会议音频编码标准,但是声音质量比 MP3 略差。在传递发言时差距不明显,但是用音乐 CD 测试时,差距就比较大了。

1. H.323 会议系统中的语音编码标准

H.323 会议系统中的音频(语音)编码方式主要有 6 种:G.711、G.722、G.723.1、G.728、G.729 和 MPEG Audio。其中 G.711 是必备的,其他为可选项。除了上述 6 种编码方式外,也可以通过能力协商来采用其他方法。

H.323 会议系统中主要采用 G.723.1 和 G.729 两种语音方式。

(1) G.723.1 语音编码。G.723.1 是一个双速率的语音编码器,它的两个编码速率分别为 6.3kbit/s 和 5.3kbit/s。高速率采用多脉冲激励最大似然量化(MP_MLQ)算法,低速率采用代数码本激励线性预测(ACELP)算法。这两种算法具有相同的理论基础,都是基于线性预