

精通

Visual C++

视频 / 音频编解码技术

马兰 沈笑云 万棣 编著

基础篇：详细介绍多媒体视/音频通信的基础知识和系统架构

视频篇：重点讲解MPEG-2、H.263、MPEG-4、H.264等
视频标准的编码、解码技术

音频篇：重点讲解G.711、G.721、G.723、GSM、G.728和
G.729等语音编码的原理和实现方法

案例篇：通过两个综合案例，展示了视频/音频技术在实际
工程中的应用



人民邮电出版社
POSTS & TELECOM PRESS



精通

Visual C++

视频/音频编解码技术

马兰 沈笑云 万棣 编著

人民邮电出版社

北京

图书在版编目 (CIP) 数据

精通 Visual C++ 视频 / 音频编解码技术 / 马兰, 沈笑云, 万棣编著. —北京: 人民邮电出版社, 2008.7
ISBN 978-7-115-18118-3

I. 精… II. ①马…②沈…③万… III. C 语言—程序设计 IV. TP312

中国版本图书馆 CIP 数据核字 (2008) 第 069036 号

内 容 提 要

本书全面、详细地介绍了 Visual C++ 视频/音频编解码技术的基本原理和编程知识, 并结合工程实际, 给出了丰富的实例和大量的 Visual C++ 源代码文件。

全书共分为 4 篇 16 章, 其中“基础篇”(第 1 章和第 2 章)介绍多媒体视频通信系统的基础知识, 并通过一个案例讲解了系统架构的方法;“视频篇”(第 3 章至第 7 章)重点讲解 MPEG-2、H.263、MPEG-4、H.264 等视频标准的编码、解码技术;“音频篇”(第 8 章至第 14 章), 重点讲解 G711、G721、G723、GSM、G728 和 G729 等语音编码的原理和实现方法;“案例篇”(第 15 章和第 16 章)通过两个综合性较强的视频/音频编解码案例, “基于 SIP 的视频会议系统”和“Skype 网络语音通信系统 API 分析”, 全面展示了视频/音频技术在实际工程中的应用。

本书内容由浅入深, 从最基本的视频/音频的标准一直到编解码技术的实现, 都有详细的讲解。在随书附带的光盘中还给出了书中涉及实例的源代码, 方便读者举一反三, 开发出适合自己的工程项目。

本书适合计算机或通信专业的师生, 以及相关领域的工程技术人员阅读和参考。

精通 Visual C++ 视频/音频编解码技术

- ◆ 编 著 马 兰 沈笑云 万 棣
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
- ◆ 开本: 787×1092 1/16
印张: 31.75
字数: 843 千字
印数: 1—4 000 册
- 2008 年 7 月第 1 版
2008 年 7 月北京第 1 次印刷

ISBN 978-7-115-18118-3/TP

定价: 65.00 元 (附光盘)

读者服务热线: (010)67170985 印装质量热线: (010)67129223

反盗版热线: (010)67171154

前言

随着人类社会步入信息化时代,信息的内涵和种类不断丰富,信息传递的形式也发生了巨大变化。人们的交流已经不仅限于笔、纸和电话……取而代之的是图、文、声并存的多媒体通信方式。

个人计算机的普及、微电子技术和多媒体技术的飞速发展、综合业务数字网的建立及宽带综合业务数字网的研究进展,都有力地推动了多媒体通信的发展,同时出现了许多新的多媒体通信方式。

多媒体通信是指在一次呼叫过程中能同时提供多种媒体信息,如声音、图像、图形、数据、文本等。它是通信技术和计算机技术相结合发展的产物。

为什么要写本书

目前,多媒体技术已经成为人们开发和研究的热点。许多企业的技术人员正在考虑利用多媒体技术实现视频会议系统等具体的方案。计算机和通信等相关专业的师生,也将多媒体相关技术的学习和研究列为课程设计或毕业设计的内容。

本书旨在帮助从事多媒体开发的人员,全面、深入地学习多媒体技术中的核心内容,即视频/音频编解码技术;掌握其相关标准、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系统。

本书主要内容

全书共分为4篇16章,主要内容如下所述。

第一篇“基础篇”,包括第1章和第2章,主要介绍多媒体视频通信系统的基础知识,并通过一个案例讲解了系统架构的方法。

第二篇“视频篇”,包括第3章至第7章,重点讲解MPEG-2、H.263、MPEG-4、H.264等视频标准的编码、解码技术。

第三篇“音频篇”,包括第8章至第14章,重点讲解G.711、G.721、G.723、GSM、G.728和G.729等语音编码的原理和实现方法。

第四篇“案例篇”，包括第 15 章和第 16 章，通过两个综合性较强的视频/音频编解码案例，“基于 SIP 的视频会议系统”和“Skype 网络语音通信系统 API 分析”，全面展示了视频/音频技术在实际工程中的应用。

本书附带的光盘中给出了书中涉及实例的源代码，方便读者举一反三，开发出适合自己的工程项目。

读者对象

本书适合计算机或通信专业的师生，以及相关领域的工程技术人员阅读和使用。

关于作者

马兰负责全书的内容规划和统筹，并负责编写第 4 章、第 5 章、第 7 章、第 8 章、第 10 章至 16 章的内容，并对全书进行了统稿；沈笑云负责编写第 1 章至第 3 章，以及第 6 章的内容，并负责最终修订；万棣负责编写第 9 章的内容；杨伟为本书的编写提供了大量的文字和代码。感谢史荣、阚洪涛、张文涛、潘雯、徐正、胡杨坤和左倩倩等在本书的编写过程中，不辞辛苦地帮助收集和整理了大量的资料，提供了很好的素材。

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 软件工厂提供的相关代码。由于本书涉及资料众多，收集难度较大，许多资料的来源在参考文献中已列出。

在本书编写过程中，我们力求精益求精，但由于多媒体技术涉及的研究领域广泛，内容新颖，技术难点多，本书难免有错误、不足和疏漏之处，恳请广大读者批评指正。如果读者使用本书时遇到问题，请发 E-mail 至 computerbook@126.com 与我们联系。

编者

容内要主书本

光盘说明

CD-ROM

为了方便读者学习，本书附带了一张光盘，下面对光盘内容及使用方法进行简要的介绍。

1. 光盘的运行环境

硬件环境：CPU 的主频在 1.0GHz 以上、内存在 256MB 以上。

软件平台：操作系统为 Windows 98/Me/NT/2000/XP（推荐使用 Windows 2000/XP），调试环境为 Visual C++ 6.0 及其以上版本。

2. 光盘目录

MPEG-2 编码实现	第 4 章\ MPEG-2(h.262)编码
MPEG-2 解码实现	第 4 章\ MPEG-2(h.262)解码
H.263 编码技术	第 5 章\ H.263 编码技术
H.263 解码技术	第 5 章\ H.263 解码技术
MPEG-4 编码实现	第 6 章\ MPEG-4 编码
MPEG-4 解码实现	第 6 章\ MPEG-4 解码
H.264 编码技术	第 7 章\ H.264 编码技术
H.264 解码技术	第 7 章\ H.264 解码技术
音频编码实现	第 8 章\ 音频编码实现
音频解码实现	第 8 章\ 音频解码实现
G.711 编码实现	第 9 章\ G.711 编码
G.711 解码实现	第 9 章\ G.711 解码
G.721 编码实现	第 10 章\ G.721 编码
G.721 解码实现	第 10 章\ G.721 解码
G.723 编码实现	第 11 章\ G.723 编码
G.723 解码实现	第 11 章\ G.723 解码
GSM 编码实现	第 12 章\ GSM 编码

GSM 解码实现	第 12 章 GSM 解码
G.728 编码实现	第 13 章 G.728 编码
G.728 解码实现	第 13 章 G.728 解码
G.729 编码实现	第 14 章 G.729 编码
G.729 解码实现	第 14 章 G.729 解码
SIP 代码实现	第 15 章 SIP 代码实现
Skype 代码实现	第 16 章 Skype 代码实现

3. 光盘使用方法

将光盘中的源代码拷入硬盘中，去掉其只读属性（否则可能无法正常使用这些源程序），用 Visual C++ 打开项目文件，即可编译运行。如遇到缺少链接库的情况，请自行到网上下载。

4. 技术支持

本书附带的源代码均是作者编写和测试过的，仅提供读者学习时使用，不能用于其他商业用途。

读者使用光盘时如果发现差错或者遇到问题，敬请批评指正，并请发 E-mail 至 tangqian@ptpress.com.cn 与我们联系。

目 录

第一篇 基础篇

第1章 多媒体通信系统简介.....2	2.2 多媒体通信系统的需求分析.....13
1.1 多媒体通信的基本概念.....2	2.2.1 多媒体传输网络性能分析.....13
1.1.1 信息的定义.....2	2.2.2 媒体终端功能分析.....13
1.1.2 多媒体的定义.....3	2.2.3 多媒体数据管理功能分析.....15
1.1.3 多媒体通信的概念.....4	2.2.4 实例——“远程教学/会诊视频 会议系统”需求分析.....16
1.2 多媒体通信的特征.....4	2.3 视频会议系统的体系结构.....17
1.3 多媒体通信的关键技术.....5	2.3.1 基本框架结构.....17
1.3.1 多媒体数据的压缩编码.....5	2.3.2 H.323 会议的组织形式.....20
1.3.2 多媒体数据的同步.....6	2.3.3 系统硬件配置说明.....21
1.3.3 多媒体数据库.....7	2.3.4 系统功能设计.....22
1.3.4 多媒体通信网.....8	2.3.5 会议功能设计.....23
1.4 多媒体通信的应用.....9	2.3.6 安全性设计.....24
第2章 多媒体通信系统的详细设计.....12	2.3.7 会议管理功能设计.....25
2.1 多媒体通信系统设计原则.....12	

第二篇 视频篇

第3章 视频编码标准.....28	3.2.3 量化与重排序.....51
3.1 视频编码标准概述.....28	3.2.4 熵编码.....54
3.1.1 视频编码技术的目的及其标准.....29	3.3 视频编码标准.....60
3.1.2 视频编解码基本原理.....31	3.3.1 静止图像压缩编码标准 JPEG.....60
3.2 视频编解码器核心技术.....34	3.3.2 视听会议压缩编码标准 H.261.....63
3.2.1 运动估计与补偿.....34	3.3.3 数字声像存储压缩编码 标准 MPEG-1.....64
3.2.2 变换编码.....44	

3.3.4 通用视频图像压缩编码 标准 MPEG-2 (H.262)	66
3.3.5 低比特率视听会议压缩编码 标准 H.263	72
3.3.6 低比特率音频与视频对象 压缩编码标准 MPEG-4	77
3.3.7 极低码率视频编码 H.264	85
第 4 章 通用视频图像压缩编码 标准 MPEG-2 (H.262)	93
4.1 MPEG-2 编码技术	93
4.1.1 MPEG-2 编码结构说明	93
4.1.2 MPEG-2 编码源代码详细分析	94
4.2 MPEG-2 解码技术	105
4.2.1 MPEG-2 的数据结构定义	105
4.2.2 解码器数据结构的定义	106
4.2.3 解码器的初始化	107
4.2.4 MPEG-2 解码核心源程序分析	111
第 5 章 低比特率视听会议压缩编码 标准 H.263	117
5.1 H.263 编码技术	117
5.1.1 H.263 编码结构说明	117
5.1.2 H.263 编码源代码详细分析	121
5.2 H.263 解码技术	143
5.2.1 H.263 数据结构定义	143
5.2.2 H.263 解码器数据结构的 定义	144
5.2.3 H.263 解码器的初始化	144
5.2.4 H.263 帧解码	146
5.2.5 H.263 解码源代码详细分析	148
第 6 章 视频压缩编码标准 MPEG-4	156
6.1 MPEG-4 编码技术	156
6.1.1 MPEG-4 基本介绍	156
6.1.2 MPEG-4 编码结构说明	159
6.1.3 MPEG-4 编码源代码详细 分析	163
6.2 MPEG-4 解码技术	179
6.2.1 MPEG-4 的数据结构定义	180
6.2.2 MPEG-4 解码器数据 结构的定义	181
6.2.3 MPEG-4 解码器的初始化	184
6.2.4 MPEG-4 帧解码	190
6.2.5 MPEG-4 解码源代码详细 分析	192
第 7 章 低比特率音频与视频对象压缩 编码标准 H.264	201
7.1 H.264 编码技术	201
7.1.1 H.264 编码结构说明	201
7.1.2 H.264 编码源代码详细分析	202
7.2 H.264 解码技术	232
7.2.1 H.264 的数据结构定义	232
7.2.2 H.264 解码器数据结构的 定义	239
7.2.3 H.264 解码器的初始化	242
7.2.4 H.264 解码源代码详细分析	243

第三篇 音频篇

第 8 章 音频编码标准	258
8.1 音频编码标准概述	258
8.1.1 信源编码	261
8.1.2 信源编码算法	262
8.1.3 信源编码控制	264
8.2 典型的音频编码器和解码器	268
8.2.1 编码器与解码器结构	270
8.2.2 编解码过程	274
8.2.3 视频缓存证实器	277
8.3 常用多媒体通信的音频编码标准	278
8.3.1 码激励线形预测编码 (CELP) 标准简介	283
8.3.2 G.711	285
8.3.3 G.721 (ADPCM)	286
8.3.4 G.723	290
8.3.5 GSM (RPE-LTP)	292
8.3.6 G.728 (LD-CELP)	297
8.3.7 G.729 (CS-ACELP)	299
第 9 章 G.711 语音编码标准	302
9.1 G.711 语音编码标准原理结构	303

9.1.1 G.711 语音编码标准说明	303	12.1.2 GSM 语音编/解码标准算法	369
9.1.2 G.711 语音编码标准算法	307	12.2 GSM 语音编码标准实现	372
9.2 G.711 语音编码标准实现	310	12.2.1 GSM 语音编码实现代码分析	372
9.2.1 G.711 语音编码实现代码分析	311	12.2.2 GSM 语音解码实现代码分析	377
9.2.2 G.711 语音解码实现代码分析	314		
第 10 章 G.721 语音编码标准	318	第 13 章 G.728 语音编码标准	380
10.1 G.721 语音编码标准原理结构	318	13.1 G.728 语音编码标准原理结构	380
10.1.1 G.721 语音编码标准说明	319	13.1.1 G.728 语音编码标准说明	381
10.1.2 G.721 语音编码标准算法	319	13.1.2 G.728 语音编码标准算法	383
10.2 G.721 语音编码标准实现	330	13.2 G.728 语音编码标准实现	387
10.2.1 G.721 语音编码实现代码分析	331	13.2.1 G.728 语音编码实现代码分析	392
10.2.2 G.721 语音解码实现代码分析	344	13.2.2 G.728 语音解码实现代码分析	399
第 11 章 G.723 语音编码标准	348	第 14 章 G.729 语音编码标准	409
11.1 G.723 语音编码标准原理结构	348	14.1 G.729 语音编码标准原理结构	409
11.1.1 G.723 语音编码标准说明	349	14.1.1 G.729 语音编码标准说明	410
11.1.2 G.723 语音编码标准算法	357	14.1.2 G.729 语音编码标准算法	413
11.2 G.723 语音编码标准实现	361	14.1.3 G.729 编码器	414
11.2.1 G.723 语音编码实现代码分析	361	14.1.4 G.729 解码器	415
11.2.2 G.723 语音解码实现代码分析	365	14.2 G.729 语音编码标准实现	416
第 12 章 GSM 语音编码标准	368	14.2.1 G.729 语音编码实现代码分析	416
12.1 GSM 语音编码标准原理结构	368	14.2.2 G.729 语音解码实现代码分析	429
12.1.1 GSM 语音编码标准说明	368		

第四篇 案例篇

第 15 章 基于 SIP 的视频会议系统	434	15.4.3 使用的相关参数	437
15.1 SIP 简介	434	15.4.4 编码接收	437
15.2 SIP 和 H.323 的对比	435	15.4.5 载入配置文件和日志	438
15.3 基于 SIP 的 IP 电话系统基本工作流程	435	15.4.6 初始化会话描述协议	438
15.3.1 用户注册	435	15.4.7 配置事务层	439
15.3.2 会话建立	435	15.4.8 用户代理客户端回叫	440
15.3.3 通话过程	436	15.4.9 用户代理服务器端回叫	440
15.4 SIP 软件电话源码分析	436	15.4.10 网络端口	443
15.4.1 退出会话	436	15.4.11 文件接收	443
15.4.2 响应消息的处理	436	15.4.12 码流	445
		15.4.13 TCP 传输层的描述	446

15.4.14	UDP 传输层的描述	449
15.4.15	事务处理	454

第 16 章 Skype 网络语音通信系统

API 分析	456
16.1 Skype 简介	456
16.1.1 Skype 与 P2P	456
16.1.2 Skype 通信原理	457
16.2 Skype API 详解	459
16.2.1 Skype API 分类	459
16.2.2 Skype Access API 术语	460

16.2.3 Skype Access API 对象及属性	461
16.2.4 管理对象属性	468
16.2.5 Skype Access API 命令的使用	468
16.3 Skype Access API 应用实例	492
16.3.1 实例说明	492
16.3.2 实现步骤	493
参考文献	498

设计

第 1 章

多媒体通信系统简介

信息社会 (Information Society) 高速发展的明显特征是媒体信息的丰富多彩和信息传输的多样化。因此,人们普遍把“信息通信”(Information Communication)作为信息社会的主要标志。信息表示和传输的载体是媒体。多媒体数据是指多种式样信息的载体,如文本、图形、图像、声音等数据。所以,在信息社会,信息通信就是多媒体通信。多媒体通信提供综合了文本、声音、图像和视频信息的服务。而提供这些服务的方法应保留易用性和交互性。

1.1 多媒体通信的基本概念

多媒体通信是指这样一个领域:表示、存储、检索和传播机器可处理的信息,而这些信息表达为诸如文本、图像、图形、音频、视频、动画、手迹和数据文件之类的多种媒体。

多媒体通信的基础是信息。对于信息的定义呈现出多定义而又无定论的局面,这一局面的形成是由于观察事物的多维视野造成的。多维视野是现代自然科学、社会科学、人文科学以及横向科学研究的一个显著特点,因此,不同学科有不同的信息定义,即使是同一学科也可能出现差异很大的命题。在信息定义这一问题上,也是如此,自然科学、信息科学、管理科学中所说的信息常常不是一回事,前者多指数据、指令,后者多指消息、情报,但即便如此,同属社会科学的消息、情报与信息也有一定的差距。

1.1.1 信息的定义

信息化的创始人香农和韦弗 1948 年在《通信的数学理论》中指出:“凡是在一种情况下能减少不确定性的任何事物都叫做信息 (Information)。”这一定义是从通信科学的角度来探讨信息概念的,排除了信息的语义因素的定义,不考虑信息的含义价值,特指一种形式信息和统计概率信息,因此不能直接用于研究人类的信息传播。但这一定义指出了信息的一个价值——减少不确定性,这是信息的认知知识功能,即当一个信息为人们所感知和确认后,这一信息就成为一定意义上的知识,形成后的知识又可以作为信息来传递(冗余信息);而尚未被认识的信息,则成为人们努力探讨的对象,在尚未认识它们之前还不是知识(未知信息)。

控制论的创始人维纳对信息的定义是:“信息就是我们在适应外部世界,并且使这种适应作

用于外部世界的过程中,同外部世界进行交换的内容的名称。”这一定义深入到了人与世界的交换关系,涉及交换内容,因此比前一个定义更适合用于研究探索人类的信息传播。

我国国家标准 GB489885《情报与文献工作词汇基本术语》中,关于“信息”的解释是:“Information,物质存在的一种方式、形态或运动状态,也是事物的一种普遍属性,一般指数据、消息中所包含的意义,可以使消息中所描述事件的不定性减少。”

这个定义首先明确了信息的本质是物质的属性,而不是物质实体本身。客观存在的一切事物,包括自然界、人体本身和人类社会,都是在不断运动着的,运动的物质必然会产生相互作用和影响,从而引起物质结构、数量等多方面的变化,事物的这些变化,便成为信息产生的物质基础。因此,信息不是事物本身,而是由事物发出的数据、消息中所包含的意义。

这个定义中首先弥补了信息论中对信息定义的特定指向性,即只描述了非人类、非语义层的信息概念,把信息这一概念放到人类社会以及人类交往中考察,也纠正了控制论信息定义中对信息概念的泛化倾向,继而明确指出信息是物质的属性,而不是事物本身,是由事物发出的消息、指令、数据、信号等所包含的内容,是数据、消息中的意义。

同时,这一定义明确了信息的认知知识的功能,即能减少不确定性的能力,可以说,信息是知识的源泉,知识是对获得信息进行处理并使之系统化的结果。这一功能是信息的基本功能,是人类解释客观世界发展规律的重要途径,知识的积累、科技的发展进步、经济文化的繁荣,都离不开信息的这一功能,经过大脑对信息的鉴别、筛选、归纳、提炼和存储,人类对客观世界的认识逐步深入,人类逐步进化、进步、发展。

其次,这一定义明确了信息是指数据与消息中所包含的意义,是数据与消息这样的讯息中所包含的内容,区分了信息与讯息,从结构上使信息的概念更加准确。

1.1.2 多媒体的定义

媒体是用于信息表示和传输的载体,是一个重要的概念。国际电联(ITU)电信标准部(TSS)的ITU-T I.374建议将媒体划分为感觉媒体、表示媒体、显示媒体、存储媒体和传输媒体5类。

感觉媒体(Perception Medium):指那些能直接作用于人的感觉器官,使人产生直观感觉的媒体。如引起人的听觉反应的语言、音乐及自然界的其他声音,引起视觉反应的文字、图形、图像及自然景象等。

表示媒体(Representation Medium):指传输感觉媒体的中介媒体,如声音编码、图像编码等,它定义了信息的特征。

显示媒体(Presentation Medium):指电信号和感觉媒体之间的转换媒体,包括键盘、光笔、扫描仪、摄像机、话筒等输入显示媒体和显示器、喇叭、打印机、拷贝机等输出显示媒体。

存储媒体(Storage Medium):指存储表示媒体的物理设备,如磁盘、磁带、光盘和IC存储器等。

传输媒体(Transmission Medium):指传输表示媒体的物理介质,如电缆、光缆、电磁波和红外线等。

有关ITU对多媒体的含义,主要是对能处理多种表示媒体而言。可以将多媒体定义为:能同时处理多种信息,使信息之间建立某种逻辑联系,集成为一个交互性系统的技术。或者定义为:多媒体是使声音、文字、图形和图像等媒体有机结合成信息的表示形式。

多媒体数据是指多种式样信息的载体,如文本、图形、图像、声音等数据。其特点主要有以下几点。

(1) 多媒体数据种类繁多(大多是非结构化数据),不同来源的媒体,具有完全不同的形式和格式。

(2) 多媒体数据量庞大。

(3) 多媒体数据具有时间特性和版本概念,如在视频点播系统中,必须考虑到媒体间以及媒体内部在时间上的同步关系。

在多媒体通信系统中,另一个常出现的词汇是“超媒体”。在出版物中经常会出现表示注解意思的“注”字,由“注”可以找到与之相关的一段文字或一篇文章。这种由“注”而链接到一段文字或一篇文章的链即称为超链接,同理,超级链也可以将若干不同媒体链接起来,其集合便称为“超媒体”。

1.1.3 多媒体通信的概念

多媒体通信是 20 世纪 90 年代特别是近几年来通信领域的一种新的通信业务,各种通信网络都在研究和开发支持多媒体通信业务的技术。多媒体通信不同于多媒体。所谓的多媒体通信,是指同时支持动态图像、语音、数据、传真等多种业务的通信。多媒体通信对网络的要求主要有两点:一是传输速率要高,排除编码技术因素,一般地,速率越高,图像的清晰度越好;二是信息传送要实时,必须保证图像的连续性和动感。

多媒体通信中的“多媒体”一词,指的是由在内容上相互关联的文本、图形、图像、音频和视频等媒体数据构成的一种复合信息实体。计算机以数字化的方式对任何一种媒体进行表示、存储、传输和处理,并且将这些不同类型的媒体数据有机地合成在一起,形成多媒体数据,这就是多媒体计算机技术。多媒体计算机技术是以计算机为核心的集图、文、声、像处理技术为一体的综合性处理技术。

多媒体通信是多媒体技术和通信技术的有机结合,突破了计算机、通信、电视等传统产业间相对独立发展的界限,是计算机、通信和电视领域的一次革命。它在计算机的控制中,对独立的信息进行集成的产生、处理、表现、存储和传输。这些信息中至少包含一个时间相关媒体(如连续的视频信号)或一个时间无关媒体(如离散的数据信号)。

多媒体通信是多媒体、通信、计算机和网络等相互渗透和发展的产物。多媒体通信系统的出现,大大缩短了计算机、通信和电视之间的距离,将计算机的交互性、通信的分布性和电视的真实性完美地结合在一起,向人们提供全新的信息服务。多媒体通信的广泛应用将会极大地提高人们的工作效率,减轻社会的交通运输负担,改变人们的教育和娱乐方式。多媒体通信将成为 21 世纪人们通信的基本方式。

1.2 多媒体通信的特征

顾名思义,多媒体通信是指在一次呼叫过程中能同时提供多种媒体信息:声音、图像、图形、数据、文本等的新型通信方式。它是通信技术和计算机技术相结合的产物。

多媒体通信技术的发展打破了传统通信的单一媒体、单一电信业务的通信系统格局,反映了通信向高层次发展的一种趋势,是人们对未来社会工作和生活方式的向往。和电话、电报、传真、计算机通信等传统的单一媒体通信方式比较,利用多媒体通信,相隔万里的用户不仅能声像图文并茂地交流信息,分布在不同地点的多媒体信息,还能步调一致地作为一个完整的信息呈现在用户面前,而且用户对通信全过程具有完备的交互控制能力。

多媒体通信技术是一种综合技术,涉及多媒体技术、计算机技术、通信技术等各个领域。多媒体通信系统必须同时兼有集成性、交互性、同步性3个主要特征。

1. 集成性

多媒体通信系统的集成性是指能对内容数据信息、多媒体和超媒体信息、脚本信息和特定的应用信息4类信息进行存储、传输、处理和显示的能力。

(1) 内容数据信息。信息是以某一种结构的形式存在的。典型的结构有两种:一种是对象结构,其中可处理的最小单元为对象(Object);另一种是文件结构,其中可处理的最小单元为文件(File)。

(2) 多媒体和超媒体信息。多媒体和超媒体信息与单媒体信息不一样,它们是结构化的信息,由结构框架和内容数据两部分组成。多媒体和超媒体信息的最小表达形式有两类,一类称为对象,另一类称为文件。

(3) 脚本信息。脚本信息是一组特定的用语意关系联系起来的、结构化的多媒体和超媒体信息,需要提供表示这一组多媒体信息的运作过程和与外部处理模块间的关系。

(4) 特定的应用信息。上述3类信息都是低层信息,可以由标准来定义和表示。特定的应用信息是高层信息,是与应用密切相关的,将随应用场合的不同而不同,它的表示方法是基于上述3类信息的基础之上的。

2. 交互性

交互性是指在通信系统中人与系统之间的相互控制能力。在多媒体通信系统中,交互性有两个方面的内容。一是人机接口,也就是人在使用系统的终端时,用户终端向用户提供的操作界面;二是用户终端与系统之间的应用层通信协议。

多媒体通信终端的用户对通信的全过程有完备的交互控制能力,这是多媒体通信系统的一个主要特征,也是区别多媒体通信系统与非多媒体通信系统的一个主要准则。

3. 同步性

同步性是指在多媒体通信终端上显现的图像、声音和文字均以同步方式工作。如用户要检索一个重要的历史事件的片断,该事件的活动图像或静止图像存放在图像数据库中,其文字叙述和语言说明则放在其他数据库中。多媒体通信终端通过不同的传输途径将所需要的信息从不同的数据库中提取出来,并将这些图像、声音、文字同步起来,构成一个整体的信息呈现在用户面前。

多媒体通信系统中的同步性是多媒体通信系统最主要的特征之一,信息的同步与否决定了系统是多媒体系统还是非多媒体系统。同步可在链路层、表示层和应用层3个层面上实现。

1.3 多媒体通信的关键技术

1.3.1 多媒体数据的压缩编码

多媒体系统中需要对多媒体数据进行捕获、存储、传输和播放等处理工作,数据压缩技术是多媒体通信技术的核心技术之一。先进的数据压缩技术尤其是视频压缩技术可实现较低的时延和高的压缩比,达到较好的图像质量,这正是多媒体视听业务能被广泛接受的重要因素之一。

国际标准化组织 (ISO)、国际电工委员会 (IEC)、国际电信联盟 (ITU) 制定了一系列视频压缩编码标准, 主要标准如下所示。

(1) ISO 10918 (JPEG) 是用于连续色调静止图像压缩编码的标准。

(2) H.261 适用的速率范围是 $p \times 64\text{ kbit/s}$ ($p=1 \sim 30$), 即 $64 \sim 1920\text{ kbit/s}$, 主要用于可视电话和会议电视系统。

(3) H.263 是以 H.261 为基础改进而来的, 可以获得更高的压缩比和较高的图像质量。

(4) ISO 11172 (MPEG1) 主要用于 CIF (公共中间格式) 格式的图像分辨率和大约 1.5 Mbit/s 的数码率, 适用于 VCD。

(5) ISO/IEC 13818 (MPEG2) 基于 $3 \sim 4\text{ Mbit/s}$ 或 4 Mbit/s 以上速率的压缩存储视频, 图像质量可达到高清晰度电视水平, 主要适用于 DVD、数字电视、视频点播和数字视频广播 (DVB) 系统。

(6) MPEG4 比 MPEG2 的应用范围更广, 其压缩方法不再是限定的某种算法, 而是可以根据不同的应用进行系统裁剪, 选用不同的算法。MPEG4 中引入的最重要, 也最引人注目的新概念是视频对象平面 (VOP)。这一概念直接导致了基于内容的压缩, 为提供更高的压缩比打下了基础, 同时也将传统的基于帧的时空可分级性扩展到基于图像内容的时空可分级性。

(7) MPEG7 是有关多媒体内容描述接口标准, 适用于基于视频和音频内容的多媒体检索业务。

有关音频压缩编码的国际标准有 ITU-T 的 G.711、G.721、G.722、G.728、G.723 标准以及 MPEG1 音频编码标准 (ISO 11172-3)、MPEG2 音频编码标准 (ISO 13818-3) 和 AC3 音频编码系统。

随着多媒体业务的不断发展, 新的视频和音频压缩编码标准不断涌现, 适用范围逐渐扩大, 压缩效率大大提高, 从而使多媒体通信逐渐应用到日常生活的各个领域。

1.3.2 多媒体数据的同步

在多媒体通信技术领域, 同步技术十分重要。目前, 多媒体技术可以处理视觉、听觉甚至触觉信息, 但支持的媒体越多, 计算机系统的相应处理子系统也越多, 处理这些媒体之间的同步问题也就越复杂。

分布式多媒体系统中的同步要求主要分为多媒体通信同步、多媒体表现同步及多媒体交互同步等。这些同步功能表现为多媒体同步体系结构中的不同层次的同步要求。多媒体通信的同步属于中层同步, 即合成同步。它的作用就是将不同媒体的数据流按照一定的时间关系进行合成, 一些要求精度较高的连续同步就属于这一类。多媒体通信的同步要求是分布式处理系统同步的最基本要求, 是其他同步功能的基础, 它和其他同步要求相互影响、相互制约。

目前解决多媒体通信中同步信息的方法很多, 下列为 3 种基本的方法。

(1) 时间戳法。这种方法既可用于多媒体通信, 也可用于多媒体数据的存取。在发送或存储时, 设想将各个媒体都按时间顺序分成若干小段, 放在同一时间轴上, 每个单元都做一个时间记号, 即时间戳。处于同一时标的各个媒体单元具有相同的时间戳。这样, 各个媒体到达接收端或取出时, 具有相同时间戳的媒体单元同时进行表现, 由此达到媒体之间同步的目的。

用时间戳同步法传输时, 不需要改变数据流, 也不需要附加同步信息, 因此其应用范围非常广泛。其缺点是选择相对时标和确定时间戳的操作较为复杂, 需要一定的比特开销用于同步。此外, 在主媒体失步或丢失的情况下都会引起其他媒体的失步或丢失。

(2) 同步标记法。发送时在媒体流中插入同步标记, 接收时按收到的同步标记来对各个媒体