



普通高等教育“十一五”国家级规划教材
21世纪高等学校计算机规划教材
21st Century University Planned Textbooks of Computer Science

多媒体 通信网络

Multimedia Communication Networks

李国辉 涂丹 张军 编著

- 从问题和需求入手，引出多媒体通信网络的解决方案
- 介绍最新技术发展的同时，注重阐述技术背后的原理
- 按网络体系结构自底向上组织内容，符合读者知识建立的习惯
- 融合多媒体与通信网络的知识，突出内容的关联，又自成体系



高校系列



人民邮电出版社
POSTS & TELECOM PRESS



普通高等教育“十一五”国家级规划教材

21世纪高等学校计算机规划教材

21st Century University Planned Textbooks of Computer Science

多媒体 通信网络

Multimedia Communication Networks

李国辉 涂丹 张军 编著



高校系列

人民邮电出版社

北京

图书在版编目 (CIP) 数据

多媒体通信网络 / 李国辉, 涂丹, 张军编著. — 北京: 人民邮电出版社, 2010.2

21世纪高等学校计算机规划教材
ISBN 978-7-115-21600-7

I. ①多… II. ①李… ②涂… ③张… III. ①多媒体—计算机通信网—高等学校—教材 IV. ①TN919.85

中国版本图书馆CIP数据核字(2009)第224513号

内 容 提 要

我们处于数字信息汇聚的网络时代, 数字信息汇聚产生了多媒体, 多媒体与通信网络结合产生了交叉的技术领域——多媒体通信网络。本书的特点是: 问题引导技术, 讲清楚来龙去脉, 围绕通信网络协议及其技术如何有效支持连续多媒体的传送构建全书的内容, 注重概念和原理的阐述, 同时力求反映多媒体通信网络技术的新发展。

全书分为7章。第1章给出多媒体通信网络的基本概念和全书的概貌; 第2章从网络多媒体应用出发, 得出分布多媒体应用对通信网络的性能需求; 第3章介绍各种通信子网技术如何支持多媒体通信; 第4章介绍通信网络高层协议中如何支持多媒体传送的问题, 包括网络层和传送层的多媒体通信协议及其QoS机制; 第5章介绍分布多媒体应用系统及其协议; 第6章介绍多媒体压缩编码原理及其传输技术; 第7章是多媒体同步的概念和模型。

本书可以作为高等学校教材, 用于通信工程、计算机网络、信息系统工程、信息系统、电气信息和计算机类各相关学科和领域的本科高年级或研究生教学; 也可供从事通信和网络工程、流媒体、视频会议系统、IP电话、Web应用、数字媒体处理系统和技术的研究、设计和开发人员学习参考。

普通高等教育“十一五”国家级规划教材

21世纪高等学校计算机规划教材

多媒体通信网络

◆ 编 著 李国辉 涂 丹 张 军

责任编辑 邹文波

◆ 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>

北京艺辉印刷有限公司印刷

◆ 开本: 787×1092 1/16

印张: 24.75

字数: 655千字

印数: 1—3 000册

2010年2月第1版

2010年2月北京第1次印刷

ISBN 978-7-115-21600-7

定价: 39.00元

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

随着多媒体技术的发展和广泛应用，多媒体通信网络成为重要的专业基础知识。通信网络包括了计算机网络、无线网络、电话网络和电视网络，多媒体通信网络是信息网络时代数字汇聚的产物，超出了传统计算机网络的范畴。我们处于数字信息和网络的时代，数字信息越来越多地用多媒体来承载，而多媒体是如何通过通信网络有效地传送呢？本书就是针对这个问题展开的。

多媒体通信网络是一门交叉技术，涉及多媒体和通信网络两个领域的知识。从事多媒体技术工作的，通信网络方面知识欠缺；而从事通信网络工作的，对多媒体技术又理解不透。作者执教过多媒体技术、计算机通信和计算机网络课程 10 多年，希望能够结合两个专业领域的内容，撰写出一本有特色的教材。

本书注重问题引导：多媒体及其应用给通信网络传输带来了什么问题？然后，讨论多媒体通信网络中的协议设计问题，即根据多媒体通信网络中的问题，逐个地阐述针对这些问题所采用的方法和措施，由此构建本书的核心内容。本书内容重点涉及多媒体传输的时延特性、抖动特性、差错控制、多播、带宽及其协议的设计问题。

书中注重阐述技术和方法的来龙去脉。把多媒体和通信网络技术产生和发展相关的资料融入到内容中，让读者了解并感受到技术的创新和发展过程。多媒体通信网络技术的发展非常快，一本书不可能包含所有收集到的资料，我们注重从繁杂的内容中精选出多媒体通信网络知识的主线，重点介绍与多媒体通信网络及其应用相关的基本概念、原理、思路、应用和技术发展。

许多通信和计算机网络方面的书籍是从通信和网络的协议角度来描写多媒体通信网络技术的，通信和网络方面的技术性强，但是多媒体特性体现得不够。本书将平衡多媒体和通信网络方面的内容，注重描述多媒体概念及其对通信网络的影响，从而让读者更好地理解为什么多媒体通信网络要如此设计和实现。

本书的内容组织

多媒体通信网络是一门新兴的交叉领域，技术和理论发展快，内容多而繁杂。我们按照网络体系结构来组织本书，采用大章方式组织，与计算机网络课程的组织方式对应，这样使得全书的结构清晰明了，读者易于学习和掌握。

本书按照通信网络的层次体系结构来写，从低到高，但不拘泥于标准的 7 层或 5 层结构。随着通信网络技术的发展，层次结构及其功能趋于融合和交叉。本书参考标准的 OSI/RM 和 TCP/IP 体系结构，分三大层次介绍多媒体通信网络技术：通信子网、网络和传送协议、多媒体应用。把网络层以下的通信网络技术作为一章来介绍。把通信子网作为一章来集中分析各种通信子网对多媒体传输的支持，其中包括了物理层、数据链路层协议及其通信子网技术。这个观点也与 Internet 网络互连的思想一致，Internet 是网络的网络，是通信子网的互连。在其之上，就是通信网络的协议，包括了网络层和传送层的多媒体传输协议，以及 QoS 机制。最高层

是多媒体应用部分，分布多媒体应用突出了多媒体技术的重要性，流媒体、多媒体会议系统、视频点播、交互电视、IP 电话是典型的多媒体应用系统。

然而，仅仅从通信网络角度理解多媒体通信网络知识是不够的，还需要融合相关的多媒体技术。因此，本书用 3 个章节分别介绍了多媒体通信网络的需求、多媒体编码与传输机制、多媒体同步。第 1 章作为全书的引入，概述了全书的内容，为读者提供多媒体通信网络知识的整体视图。本书内容与“计算机网络”、“数据通信”和“多媒体技术”等课程的内容有一定联系，但是内容相对独立和完整。具有一定通信网络知识的读者，就可以阅读和学习本书的内容。

本书由李国辉撰写了第 1 章~第 5 章、第 7 章，并负责全书的统稿。涂丹和张军撰写了第 6 章。研究生柳伟、冯明月、贾立、范新胜、尹逊亮等同学撰写了部分小节内容，并帮助绘制了部分的图表。

本书的读者对象

虽然本书是多个作者共同编写的，但是本书的结构和内容是经过精心设计和编写的。各章内容可以剪裁使用，既可以作为大学本科高年级和研究生的教科书，又可供该领域或相关领域的研究开发人员参考。

致谢

在本书的写作期间，许多人曾给予我们帮助和支持。感谢国防科技大学信息系统与管理学院的领导和专家、教师、科研和教学管理人员和研究生们。同时我们要给予我们的家人以特别的感谢。

在本书的撰写过程中，我们参考了计算机网络和多媒体技术领域的许多著名著作（文献 [1]~[9]，[36]），受到很大启发和影响，同时还引用了许多前沿的研究资料和技术文献（虽然在参考文献中列出，但是由于内容分散，有可能没有在具体的内容中给出确切的引用），特向这些著作和资料的作者表示衷心感谢。

虽然编者尽力所为，但书中难免还存在一些缺点和错误，敬请广大读者批评指正。

编 者

于国防科技大学

2009 年 8 月

目 录

第 1 章 概论	1	2.1.1 听觉感知	33
1.1 什么是多媒体	1	2.1.2 数字音频	35
1.1.1 多媒体的概念	1	2.1.3 视觉感知	36
1.1.2 媒体的分类	3	2.1.4 数字视频	38
1.1.3 多媒体数据的基本类型	4	2.2 典型的分布多媒体应用需求	40
1.1.4 多媒体数据的时空特性	6	2.2.1 音频流和视频流	40
1.2 多媒体与通信网络	7	2.2.2 对话及其多方通信	41
1.2.1 什么是通信网络	8	2.2.3 点到点通信	42
1.2.2 常用的通信网络	8	2.2.4 Web 中的多媒体	43
1.2.3 通信网络的数字汇聚	13	2.2.5 分布多媒体应用的需求	43
1.3 媒体及其流特性	14	2.3 性能需求	44
1.3.1 传输的时间约束	14	2.3.1 面向连续比特流的通信	44
1.3.2 连续媒体流的特性	15	2.3.2 典型的性能度量指标	45
1.3.3 媒体与逻辑数据单元	18	2.3.3 带宽	46
1.4 多媒体与分层的体系结构	18	2.3.4 时延	47
1.4.1 网络体系结构的概念	18	2.3.5 差错控制	50
1.4.2 OSI 参考模型	21	2.3.6 多点通信	51
1.4.3 TCP/IP 参考模型	22	2.4 服务质量	52
1.4.4 多媒体通信的需求	24	2.4.1 QoS 的概念	52
1.4.5 多媒体与层功能	25	2.4.2 QoS 参数	52
1.5 多媒体标准	27	2.4.3 编码与 QoS	53
1.5.1 ITU-T 视听通信标准	27	2.4.4 协议层中的 QoS	53
1.5.2 ISO/IEC/JTC1 标准	27	本章小结	54
1.5.3 Internet 多媒体标准	28	思考与练习	54
1.5.4 其他标准相关的论坛和组织	29	第 3 章 通信子网	56
1.6 本书的内容结构	30	3.1 多媒体应用对通信子网的需求回顾	56
本章小结	32	3.2 电信网络技术与多媒体传输	57
思考与练习	32	3.2.1 电话网	57
第 2 章 多媒体通信网络的需求	33	3.2.2 ISDN 网络	58
2.1 音频和视频基础	33	3.2.3 SONET/SDH	59
2.1.1 听觉感知	33	3.3 传统面向连接的分组交换网络与多媒体传输	63
2.1.2 数字音频	35		
2.1.3 视觉感知	36		
2.1.4 数字视频	38		
2.2 典型的分布多媒体应用需求	40		
2.2.1 音频流和视频流	40		
2.2.2 对话及其多方通信	41		
2.2.3 点到点通信	42		
2.2.4 Web 中的多媒体	43		
2.2.5 分布多媒体应用的需求	43		
2.3 性能需求	44		
2.3.1 面向连续比特流的通信	44		
2.3.2 典型的性能度量指标	45		
2.3.3 带宽	46		
2.3.4 时延	47		
2.3.5 差错控制	50		
2.3.6 多点通信	51		
2.4 服务质量	52		
2.4.1 QoS 的概念	52		
2.4.2 QoS 参数	52		
2.4.3 编码与 QoS	53		
2.4.4 协议层中的 QoS	53		
本章小结	54		
思考与练习	54		

3.3.1 X.25	63	3.9.7 3G 的多媒体业务	111
3.3.2 帧中继	64	本章小结	113
3.4 Internet 与多媒体传输	65	思考与练习	114
3.4.1 Internet 的多媒体能力	65	第 4 章 多媒体网络与传送协议	116
3.4.2 多媒体电子邮件	66	4.1 传统网络协议	116
3.4.3 多播与 Mbone	68	4.1.1 协议的功能	116
3.4.4 Internet 2	70	4.1.2 传统协议的问题	119
3.5 局域网与多媒体传输	71	4.2 IPv6 协议	121
3.5.1 标准以太网	71	4.2.1 IPv6 协议	121
3.5.2 交换以太网	72	4.2.2 数据报格式	121
3.5.3 高速以太网	73	4.2.3 IPv6 对多媒体的支持	124
3.5.4 无线局域网	74	4.3 网络层服务质量	126
3.6 ATM 网络与多媒体传输	76	4.3.1 提高服务质量的技术	126
3.6.1 ATM 的概念和特性	76	4.3.2 综合服务	134
3.6.2 ATM 协议参考模型	79	4.3.3 差别服务	135
3.6.3 ATM 连接和交换	81	4.3.4 标签交换和 MPLS	138
3.6.4 ATM 提供的服务与 QoS	83	4.3.5 服务质量问题的讨论	146
3.7 接入网络与多媒体传输	84	4.4 资源预留协议	146
3.7.1 信息使用的对称与非对称模式	84	4.4.1 RSVP 基础	147
3.7.2 电话线接入网 ADSL	85	4.4.2 RSVP 预留方式	148
3.7.3 有线电视接入网 HFC	87	4.5 IP 多播协议	152
3.7.4 光纤接入网	90	4.5.1 广播与多播的概念	152
3.8 WiMAX 技术	91	4.5.2 广播路由选择	153
3.8.1 什么是 WiMAX	91	4.5.3 多播路由选择	155
3.8.2 802.16 协议栈	91	4.5.4 IP 多播及其协议	157
3.8.3 无线传输方式	92	4.5.5 IP 多播编程	160
3.8.4 MAC 控制	93	4.6 实时传送协议	164
3.8.5 帧结构	94	4.6.1 RTP	165
3.9 3G 移动多媒体通信	95	4.6.2 RTCP	167
3.9.1 移动通信的发展	95	本章小结	169
3.9.2 3G 体制种类	99	思考与练习	170
3.9.3 WCDMA 技术	100	第 5 章 分布多媒体应用及协议	171
3.9.4 TD-SCDMA 技术	102	5.1 多媒体应用	171
3.9.5 CDMA2000 技术	105		
3.9.6 移动管理	108		

5.1.1 多媒体应用的发展	171	5.8.1 交互式视频服务的概念	226
5.1.2 多媒体应用的类型	172	5.8.2 纯视频点播与准视频点播	228
5.1.3 ITU 应用	175	5.8.3 视频服务与管理	230
5.1.4 MPEG 应用	178	5.8.4 机顶盒	232
5.2 应用级成帧	182	5.8.5 传送网络	235
5.2.1 ALF 的概念	182	5.8.6 视频服务器	237
5.2.2 ALF 与 RTP	182	5.8.7 一种视频点播系统实例	241
5.3 Internet 上的实时语音通信应用	183	5.9 P2P 视频服务	243
5.3.1 IP 尽力服务的问题	183	5.9.1 P2P 概述	243
5.3.2 平滑抖动和时延限制	184	5.9.2 P2P 系统框架	246
5.3.3 数据包丢失处理	187	5.9.3 基于 P2P 的视频传输方式	249
5.3.4 VoIP	188	5.9.4 典型 P2P 视频直播系统和 软件实例	250
5.4 流媒体	193	5.10 多媒体内容分发	251
5.4.1 流媒体的概念	193	5.10.1 什么是 CDN	251
5.4.2 流媒体服务结构	194	5.10.2 CDN 实例	252
5.4.3 实时流协议	196	本章小结	253
5.4.4 流媒体技术实例	197	思考与练习	254
5.5 基于分组的多媒体通信系统 H.323	198	第 6 章 多媒体压缩编码及 传输技术	256
5.5.1 H.323 的基本概念	198	6.1 多媒体传输与压缩编码基础	256
5.5.2 系统组成	199	6.1.1 多媒体传输对压缩编码的需求	256
5.5.3 终端构成	202	6.1.2 多媒体压缩基础	257
5.5.4 呼叫过程	204	6.1.3 变换压缩编码	261
5.6 会话发起协议	206	6.1.4 无损编码方法	270
5.6.1 SIP 的基本概念	206	6.2 图像压缩编码技术	271
5.6.2 SIP 的组成	207	6.2.1 图像传输对编码的需求	272
5.6.3 SIP 的会话过程	208	6.2.2 JPEG 图像压缩标准	272
5.6.4 SIP 的报文格式	209	6.2.3 JPEG 2000 图像压缩标准	280
5.6.5 SIP 与 H.323 比较	211	6.2.4 基于对象的编码方法	283
5.7 多媒体会议系统	212	6.3 视频压缩编码技术	287
5.7.1 什么是多媒体会议系统	212	6.3.1 视频传输对编码的要求	287
5.7.2 多媒体会议系统的结构	216	6.3.2 H.261 编码标准	287
5.7.3 多点控制单元	218	6.3.3 H.263 系列编码标准	288
5.7.4 会议系统的数据应用	219	6.3.4 MPEG-1 编码标准	289
5.7.5 会议控制	223		
5.8 交互电视与视频点播	225		

6.3.5 MPEG-2 编码标准	290	7.3.1 自然同步	340
6.3.6 MPEG-4 编码标准	297	7.3.2 合成同步	342
6.3.7 H.264/MPEG-4 AVC 编码标准	302	7.4 分层的同步系统结构	342
6.3.8 可伸缩视频编码	308	7.4.1 分层的同步机制	342
6.4 音频压缩编码技术	309	7.4.2 网络环境下的同步控制	344
6.4.1 音频传输对编码的要求	309	7.5 时间与同步	347
6.4.2 G 系列音频编码	310	7.5.1 时间性	347
6.4.3 MPEG 系列音频编码	312	7.5.2 相对时间与绝对时间	347
6.5 实时媒体流的差错控制技术	321	7.5.3 时间抽象表示	350
6.5.1 实时媒体流的差错控制技术基础	321	7.6 同步模型及其表示	351
6.5.2 基于编码端的差错恢复编码技术	323	7.6.1 基于时间段的同步模型	351
6.5.3 基于解码端的差错隐藏技术	325	7.6.2 基于图形的同步模型	353
6.5.4 基于编解码端交互的差错控制技术	327	7.6.3 基于脚本的同步模型	355
6.5.5 联合信道编码技术	328	7.6.4 同步中的交互描述	359
6.5.6 基于内容的差错控制技术	328	7.7 同步多媒体集成语言	360
6.6 视频编码传输中的控制技术	329	7.7.1 什么是 SMIL	360
6.6.1 编码速率控制技术	330	7.7.2 SMIL 基本标记	361
6.6.2 计算复杂度控制技术	332	7.7.3 多媒体片段的顺序和并行组织	362
本章小结	333	7.7.4 时间控制	363
思考与练习	333	7.7.5 空间布局	366
第 7 章 同步	335	7.7.6 链接	368
7.1 多媒体同步的概念	335	7.7.7 动画描述	371
7.1.1 什么是多媒体同步	335	7.7.8 转场效果	372
7.1.2 同步关系	336	7.8 同步中的其他问题	373
7.2 媒体间与媒体内同步	337	7.8.1 同步容忍度	373
7.2.1 媒体间和媒体内的同步概念	337	7.8.2 系统支持的有关问题	376
7.2.2 时基逻辑数据单元	338	本章小结	377
7.2.3 时基逻辑数据单元实例	339	思考与练习	377
7.3 自然与合成同步	340	附录 A 英文缩写对照	378
		参考文献	386

第 1 章

概论

本章作为全书的入门，为读者阐述多媒体与通信网络的关系，从而引出多媒体通信网络的概念。了解多媒体及其媒体流的特性是理解多媒体通信网络技术的基础。有了这些概念基础之后，本章按照网络体系结构的理论框架分别介绍每一层与多媒体传输的关系，每一层需要哪些特殊的功能及其协议，以满足多媒体传输的需要。

1.1 什么是多媒体

多媒体是数据和信息的载体，多种媒体复合具有更丰富的信息表现力。多媒体数据具有与常规字符数值数据不同的特性，不同的媒体类型又具有不同的特性，其中时间特性对多媒体传输的影响最大。

1.1.1 多媒体的概念

1. 多媒体的含义

什么是多媒体？从字面上看，多媒体的英文“multimedia”由 multiple 和 media 两个词组合而成，表示多种媒体的综合和利用。经过 20 多年的发展，多媒体的概念远远超出了“多种”媒体的含义，已经形成了一个研究和应用的专门领域。

注意，这里的“媒体”，从普通意义上看，是**数据和信息的载体**，是信息的传递和表现物质。文本、图像、视频、语音、空气、水、大气等是媒体，它们可以是数据和信息的载体，如空气和水是声波的载体，报纸上的文本是新闻的载体，图像和视频是视觉信息的载体。

然而，多媒体并没有正式的定义。多媒体有多种多样的解释，例如：

“计算机提交的电子系统，允许用户控制、组合和操纵不同类型的媒体，如文本、声音、视频、计算机图形和动画”；

“一种综合文本、音频、视频和动画图形的计算机技术”；

“方便访问的大量信息，如工业培训手册、词典和百科全书”；

“通常采用小的存储设备，如光盘”；

“在 20 世纪 90 年代对计算机和通信业造成革命性的影响”……

从用户角度看，计算机信息可以通过音频、视频、文本、图像、图形和/或动画表现出来，这些媒体的组合就称为**多媒体**。

从技术角度看，还要考虑到媒体之间的时间关系、集成特性和用户交互能力。要处理好其中的时间、集成和交互特性，才是多媒体技术面临的真正挑战。

在多媒体通信系统中，真正的难点还是在连续媒体的传输方面。所谓**连续**（continuous）媒体，是指数字连续媒体，如果它的基本数据单元（如帧和数据包）是以**周期性**序列形式表现，那么就称为连续媒体。典型的连续媒体是数字视频和音频。而与连续媒体对应的是**离散**（discrete）媒体，其包含的基本数据单元之间没有时间约束，如文本、图像和图形。图像的像素在表现时，像素之间没有时间关系。同样，图形元素之间在表现时，也没有严格的时间约束关系。文本也类似。

多媒体是集成多种媒体的一种媒体形态。为了更好地理解多媒体，可以从其具备的特性去分析，看其特性与相关的概念有什么区别。多媒体与其他相关概念的区别在于以下几个特性。

（1）**媒体组合**。连续媒体和离散媒体都包含的多媒体系统，才是**严格意义**上的多媒体系统。也就是说，只有包含了音频或视频的系统，才能真正算是多媒体系统。如果只是包含图像、图形等多种媒体，还不能算是多媒体系统，因为这样的系统早就有了，如图形工作站、图像处理系统等。

（2）**集成性**。多媒体的组合不是松散耦合的，而是集成在一起的，是计算机可控的。各个媒体之间具有时间同步关系、空间关系及其语义等关系。媒体的集成使得综合的信息量大于单个媒体信息量的总和。

（3）**可操纵性**。多媒体是计算机控制的数字媒体，是可以交互的。这个特性区别于传统的电视媒体。传统电视媒体可以交互吗？可以，但它只是在换频道的水平，交互的颗粒度太大，可控性弱。数字电视出现后，交互电视才属于多媒体的范畴。交互性使得人们获取信息的方式从被动变为主动，操纵信息的颗粒度更细，由此使得用户掌握和利用信息的能力更强。

从**严格意义**上看，只有包含、处理和应用连续媒体的计算机、设备、技术和系统才是多媒体的。只有连续媒体数字化后进入到计算机，能够被计算机处理和管理，多媒体的时代才真正到来。但是从**广义**上看，人们把任何能够处理多种媒体（可能不包含连续媒体）的计算机、设备、计算和系统都看成是多媒体的。

2. 多媒体的出现

多媒体的出现是计算机、家用电器和通信业发展的汇聚，又称为数字汇聚。最早研究和提出多媒体系统的分别是计算机工业的代表 IBM、Intel、Apple 及 Commodore 公司，家用电器公司的代表 Philips、Sony 等。他们推出了能够交互式综合处理多媒体信息的设备或系统。

Apple 公司是多媒体产业的先锋和时尚引导者，在 IBM PC 只能用单调的合成声音发声的时候，就把数字音乐引入到台式计算机。IBM 和 Intel 联合推出的 DVI（Digital Video Interactive）系统，使计算机能够处理影像信息。这项技术是多媒体真正引人注目的里程碑技术。Microsoft 公司等一大批软件开发商以多媒体应用为契机，推出的各类多媒体软件和 CD 光盘，造就了一大批多媒体应用及用户。

以 Philips 和 Sony 公司为代表的家用电器工业，早期推出了可交互的家电娱乐和阅读装置 CD-I。CD-I 属于 CD 光盘系列的绿皮书标准，它在 CD-ROM 标准基础上补充了音频、视频和计算机程序方面的规定。

多媒体与通信网络尤其是与 Internet 的结合，使得多媒体技术在人们的生活和工作中广泛应用。目前，数字电视工业的兴起，就是多媒体、娱乐和广播通信技术结合的典型产物，今后将向交互式数字电视的方向发展。Web 已经成为最普遍的多媒体应用平台。可视电话、视频会议、视频点播、流媒体、3G、4G 移动通信技术等都是通信业在多媒体技术上的新发展。

3. 多媒体不仅仅是图文并茂

人们对多媒体的认识,往往停留在“图文并茂”上。其实,多媒体的技术和应用不仅仅局限在图文并茂上。从上面的描述看,多媒体技术和应用的范围很广,涉及媒体通信、存储、表现、表示、感知等多个方面。

多媒体既不是计算机的升级换代,也不是电视或其他什么设备的变种,更不仅仅是多媒体课件 ppt、图像编辑 Photoshop、动画制作 3D Max 等,而是在一个新的环境下产生的一个新的技术时代和技术领域。

多媒体也不只是一种人机界面技术。用人机界面来描述多媒体,只反映了它的一个侧面。多媒体的内涵将比人机界面更加丰富。

多媒体是在新的层次上将各种技术进行了集成和发展,在数据库领域、通信网络领域、信息处理领域、人机接口与交互领域、系统部件和硬件设计领域都需要研究、开发和应用的新技术、新概念和新方法。本书侧重于通信网络领域的多媒体技术及其应用。

1.1.2 媒体的分类

根据 ISO 对媒体的分类,把媒体分为^[1]:

- (1) 感知媒体;
- (2) 表示媒体;
- (3) 表现媒体;
- (4) 存储媒体;
- (5) 传输媒体。

还有一种媒体称为信息交换媒体,它实质上是由存储媒体和传输媒体构成的,并强调信息交换的标准化和规范化。

感知 (perception) 媒体用于人对环境的感知,即在计算机环境下如何感知信息。人通过眼睛、耳朵等 (还有触觉、嗅觉等) 来感知环境和信息。视觉媒体模拟看的通道,如文本、图像、视频等媒体。听觉媒体模拟听的通道,如语音、音乐、声响等。

表示 (representation) 媒体用于信息在计算机内部的表示,即如何对计算机信息进行编码。在计算机中,用各种格式表示媒体信息,即媒体的数字化表示。例如,Text-ASCII、EBCDIC、Audio-PCM、Image-JPEG、Video-MPEG 等。

表现 (presentation) 媒体用于信息的表现,信息通过输入和输出的工具和设备表现出来,即计算机通过什么媒体获取和提交信息。例如,提交信息的媒体设备可以是纸 (文本)、屏幕 (图像、图形和视频)、扬声器 (音频)、头盔 (各种媒体) 等,而获取信息可以通过键盘 (文本)、鼠标 (交互)、照相机 (图像)、麦克风 (音频)、摄像机 (视频) 等。

存储 (storage) 媒体是存储信息的载体,即信息存放在哪里。例如,常用的数字存储设备有硬盘、光盘、半导体存储器、磁盘阵列等。

传输 (transmission) 媒体是信息传输的载体,即信息通过什么进行传送。例如,通过光纤网络、无线局域网、3G 移动网络传输等。

从上面的概念定义看,我们说的多媒体实际上指的是感知媒体和表现媒体。而表示媒体的含义实际上是媒体编码。存储媒体和传输媒体实际的含义是存储介质和传输介质。这些媒体的结合、综合使用,就构成了所谓的多媒体。

1.1.3 多媒体数据的基本类型

从广义的角度看,多媒体数据类型包括所有的数据类型,但是典型的多媒体数据是视频、音频、图像、图形、动画等。我们一般把字符数值型数据划分为传统数据。字符数值型数据记录的是事物的简单属性(如性别)、数值属性(如人数),或是抽象属性(如事物所属类别)。这种数据具有简单、规范和结构化的特点,因而易于管理。传统信息系统管理和传输的主要是这类数据类型,这类数据类型中的数据单元之间不具备时间关系。但是在实际的内容管理上,往往把文本数据也作为重要的内容来管理。因此,在多媒体数据类型中,可以考虑文本数据。

1. 文本数据

文本数据是各种格式和各种语言编码的**文本(text)**或复合文本。文本是最常见的媒体形式,各种书籍、文献、档案等无不是由文本数据为主构成的。在计算机内,文本数据是由具有特定意义的字符串表示的。文本是非结构化的数据。

在 Web 上的文本是用超文本标记语言(HyperText Markup Language, HTML)和可扩展的标记语言(eXtensible Markup Language, XML)描述的。通过 HTML 或 XML 表示后,文本就是半结构化的了,便于管理和处理。

超文本是一种非线性结构组织的文本,其结点和链是其基本结构。结点是文本内容单元,通过链把这些内容单元关联起来,如 Web 的结构。这种结构区别于常规的顺序文本结构,因此称为超文本(hypertext)。

HTML 网页通过超文本传送协议(HyperText Transfer Protocol, HTTP)从 Web 服务器传送到浏览器端,显示出来给用户。HTML 网页中可以包含和引用各类多媒体数据,如视频、音频、图像等。

2. 声音数据

按照声音(sound)的特性,声音可以分为波形音频、音乐和语音。音频是通过录音设备录制的。音乐和语音也可以用音频来记录和表示。

音乐(music)数据在计算机里可以用乐器数字接口(Music Instrument Digital Interface, MIDI)符号表示,因而数据量很小,播放时通过音乐合成器再现出来。语音(voice)也可以通过语音合成器发音,或是设置一个字词语音库,根据文本字符来索引语音库,调出与字词对应的语音数据,拼接成一段语音后播放。但是这种方式合成的音乐和语音效果不太好,质量不高,除了特殊的应用之外(如自动语音报号、自动语音信息播报等),应用不多。

声音数据是以数字化的波形**音频(audio)**为主,这样存储空间就比较大。数字化音频可以表示音乐、语音和其他所有声音。声音是一种典型的听觉媒体,是一种具有时间特性的媒体。没有时间,就没有声音。因此实时传输音频数据的时候,需要考虑其时间特性。

3. 图形数据

图形(graphics)由矢量元素组成,如绘制图、素描、图表、地图、三维真实感图等。图形数据的数据库管理已有一些成功的应用范例,如地理信息系统、工业图纸管理系统、建筑 CAD 数据库等。图形数据可以分解为点、线、弧等基本图形元素。描述图形数据的关键是要有可以描述层次结构的数据模型。

对图形数据来说,最大的问题就是如何对数据进行图形化表示,这又与应用密切相关。一般来说,由于图形是用符号或特定的数据结构表示的,更接近于计算的形式,因此还是易于管理的。图形数据是空间数据,图形元素之间没有时间约束关系。

4. 图像数据

图像 (image) 数据是指位图图像, 由像素元素组成。图像是物体和场景的空间表示。可以是彩色、二值、黑白灰度图像; 或可见光、波谱、红外、雷达图像等。图像数据是应用最普遍的一种媒体形式。

在通信网络中传输的是数字图像。数字图像是由像素矩阵表示的, 像素代表一个量化的强度值。如果用 8 比特表示一个像素值, 那么它的取值范围是 0~255。

图像的基本单元是像素。在压缩编码时图像可以分块, 如 16×16 像素块。一幅图像可以在逻辑上分为子图像, 如前景子图像和背景子图像。彩色图像数据的组织可能是分平面的, 如红、绿、蓝分量的像素平面, 或亮度、色差分量的像素平面。这些就是图像的基本逻辑单元。这些逻辑单元之间一般没有时间关系。

但是随着图像传感器分辨率的提高, 如 1 000 万像素的照相机已经普及了, 拍摄下来的图像数据量非常大, 一幅图像在网络上传输时, 可能需要几秒钟 (甚至更长) 的时间, 用户在线观看图像时, 等待的时间长。一种解决办法就是渐进的图像传输机制, 先传输图像的基本信息, 然后逐渐把图像的细节信息传输过去, 最后在接收端获得清晰的图像。接收端的用户不需要等待很长时间, 就可以尽快地看到图像的概貌, 在感觉上好多了。

5. 视频数据

视频 (video) 是由图像序列 (帧序列) 组成的, 通过录像设备摄制。动态视频要复杂得多, 在通信网络上传输存在新的问题。

视频除了具有空间特性 (视频是连续的图像, 图像具有空间特性) 之外, 还具有时间特性。视频记录的事件需要一个过程才能表现出来, 这就是时间特性。

视频在时间上具有多层次的颗粒度 (即逻辑单元), 如场景、镜头单元等。在通信网络中传输, 这些逻辑单元之间具有时间关系。对于基于时间的媒体来说, 为了真实地再现就必须做到实时, 而且需要考虑视频与其他媒体的复合与同步。例如, 给一段视频配音, 声音与图像必须配合得恰到好处。复合和同步涉及采集、传输协议、媒体同步表现、数据压缩等诸多方面。

6. 动画数据

动画 (animation) 是图形绘制的图像序列, 或者是计算机可解释运行的动画描述。动画可以转换为视频输出, 成为视频信号, 这样可以当做视频数据来管理。

动画分为二维和三维的。输入和编辑关键帧, 或定义对象的运动路径, 就可以让计算机生成二维动画的中间帧 (根据关键帧和关键点插值)。二维动画可以有立体感, 是三维在二维上的投射。三维动画景物有正面、侧面和背面。它基于三维模型及其光照, 通过三维造型运动、纹理映射和绘制, 形成三维动画。

动画的主要特性是其结构化的元素, 如三维对象、二维对象、场景对象等, 这些动画逻辑单元之间具有一定的时间关系。但是与视频不同的是, 这些时间关系是人为可控的。对于视频来说, 其时间关系在采集时就自然确定了, 在播放的时候要求保持这种时间关系, 就可以正确地再现记录的信息或事实。而动画中的角色和场景是人工编排的, 其运动是人工控制的, 因此在通信网络中的同步机制与视频不同。当动画录制为视频形式之后, 其在通信网络上的传输要求就和视频一样了。

7. 复合媒体

复合媒体是由以上所述的两种或两种以上数据类型合成的媒体, 如有语音说明的幻灯片、

语音伴随的视频等。为什么需要多媒体？其实多媒体的一种重要特性就是反映在其多种媒体提交信息的融合、互补和增强上，达到 1 加 1 大于 2 的效果，也就是说复合媒体携带的信息更丰富。

复合媒体在通信网络上传输的最大问题是同步。如果在发送端就把多个媒体复合为一个媒体，就容易保持媒体之间的同步。但是，如果多个媒体来自于不同的服务器，那么就必须在中结点或接收端进行复合。这时，由于每个媒体网络上传输的时延是不同的，在接收端同步非常困难。

请读者特别注意视频与动画、图像与图形、声音中的音频、音乐和语音之间的差别和不同的特性。

1.1.4 多媒体数据的时空特性

1. 媒体的不同特性

没有任何一种媒体在所有场合都是最优的。每一种媒体都有其各自擅长的特定范围，在使用时必须根据具体的信息内容、上下文和使用目的，来选择相应的媒体。人在问题求解过程中的不同阶段对信息媒体有不同的需要。

相对来说，能提供具体信息的感知媒体适用于最初的探索阶段，能描述抽象概念的文本媒体适用于最后的分析阶段。一般说来，文本信息擅长表现概念和刻划细节；图形信息擅长表达思想的轮廓以及那些蕴含于大量数值数据内的趋向性信息；图像信息逼真地表现现场的场景和人物；视频媒体则适合于表现活动的场景和事件；音频媒体是听觉感知的记录。

听觉与视觉信息可以共同出现，往往适用于作说明和示意，进行效果的渲染和烘托。同样，运动媒体则反映了用户直接的交互意图和系统所做出的反应。

在信息表达方面，媒体数据具有以下性质。

(1) 媒体的编码性。媒体是经过编码的，也就是说，只有对编码进行解释，才能使用这种媒体。

(2) 媒体的内容性。不同媒体表达信息的特点和程度各不相同，愈接近原始媒体形式，信息量愈大；愈是抽象，信息量愈小但愈精确。从原始视听内容角度看待多媒体数据的传输，是区别于传统字符数值数据传输的一个重要出发点。

(3) 媒体之间可以相互转换。如文本转换为语音、图像转换为图形、视频转换为图像、动画转换为视频。但可能会丢失部分原始信息，或增加一些伪信息。在通信网络上传输多媒体数据，可能在发送端、中间结点或接收端对媒体进行转换。

(4) 媒体之间的关系也具有丰富的信息。因此，在媒体传输过程中，保持媒体的同步关系是非常重要的。

2. 媒体的空间性质

图形、图像、视频数据就其内容而言具有一定的空间特性。多媒体信息的空间性质具有如下含义。

(1) 表现空间。尤其是指显示空间的安排。其中包括每种可视媒体在显示器上的显示位置、显示形式、先后关系等。对于声音媒体则安排它在听觉空间中表现。这实际是媒体内的空间关系，具有媒体对象的空间相邻、方位、叠加等关系。

(2) 空间的上下文关系。把环境中各种表达信息的媒体按相互的空间关系进行组织，整体地反映信息的空间结构，而不仅仅是零散的信息片段。这种空间实际上是由系统通过显示器和其他

设备给出一个观察世界的窗口,并将环境的媒体信息进行空间的组织,反映出媒体信息的空间结构。例如,在网上虚拟博物馆系统中,对于一幅博物馆中雕塑的照片,可能会使人联想起这座雕塑的侧面、后面、上面、下面等,也就要有相应的图像衔接这一幅照片的周围。随着用户的移动,他可以观察到所有的信息。这种根据媒体内容的空间关系,实际上是将信息在空间上进行了有序的组织。这就是空间“上下文”关系。

(3) 多种媒体的空间关系。视觉媒体空间和听觉空间既相互独立又需要相互结合。视觉空间的内容通过各种显示器、摄像机采集和表现;听觉空间通过麦克风、扬声器等进行获取和再现。两个空间相互结合,就可以构成多媒体下的虚拟空间信息环境。这是媒体间的空间关系,是多个媒体间媒体单元或对象的关系。

3. 媒体的时间性质

有些多媒体数据还具有时间需求,如视频、音频和动画。在其存储、传输、操作和显示时需要考虑时间性,即把从不同信息源获取的各种媒体数据在规定时间内同步显示出来。媒体的时间性质有两种含义。

(1) 表现所需的时间。这是所有媒体都需要的。对于图像、文字等静态媒体来说,它至少需要一定的表现时间,接收者也需要一定的接收时间去接收理解它。对声音来说,没有时间也就没有了声音,声音总是完全依赖于时间的变化。视频信息虽然也要依赖于时间的变化,但它的每一帧都可以单独存在(也就是图像),并且可以表现。

(2) 媒体单元之间的时间关系。媒体的时间也可以包含媒体单元在时间坐标轴上的相互关系。例如,同一个地点的照片,由于时间的不同,表现出来的空间效果也不同。这种时间关系可以是周期性的(如春夏秋冬),也可以是非周期性的。时间关系还存在于同步、实时等许多方面。媒体单元之间的时间关系有两种:一个是描述单个媒体内部单元之间的时间结构,另一个是描述多种媒体单元之间的时间结构。我们称具有这种时间关系的媒体为**时基媒体**,而其他媒体为**非时基媒体**,如图1.1所示。在第2章将详细讨论时基媒体的时间关系。

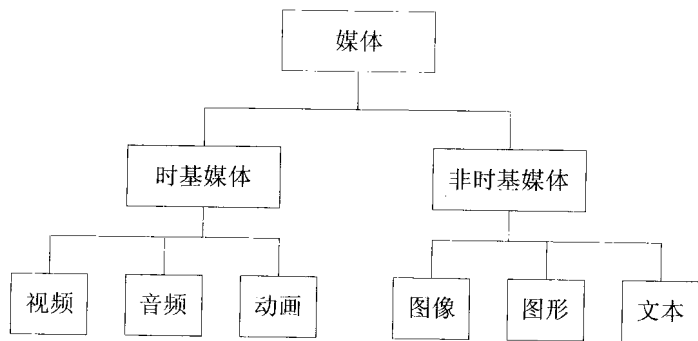


图 1.1 按时间特性划分媒体类型

1.2 多媒体与通信网络

我们在生活和工作中无处不在使用电话网络、计算机网络、电视网络和无线网络,它们是Web信息和各种信息网络的运载基础设施,我们把这些网络统称为通信网络。现在通信网络上传递的主要数据是多媒体数据。

1.2.1 什么是通信网络

过去,人们把“通信网络”这个术语看成是计算机通信的一个阶段,是从通信发展到计算机网络的中间阶段,即有了点到点的计算机通信,然后实现多台计算机的通信,但是还没有网络操作系统的支持,不能实现网络资源的共享。这个阶段称为计算机通信网络。这是一些传统的计算机网络教科书中阐述的观点。

但是在本书中,从数字媒体汇聚及其计算机网络技术发展的角度看,把**通信网络**(communication networks)看成是信息交换和共享的各种通信和网络系统的统称,而不是计算机网络的极低阶段。

通信网络传递的信息可以是数值数据、文本、语音、视频、图像、电子邮件、计算机文件等。用户可以通过固定电话、移动电话、电视机机顶盒、个人数字助理(PDA)、计算机等设备访问通信网络,而不是局限于只用计算机联网。因此,从广义的角度看网络,称为通信网络。按照物理上相对独立构建的方式,可把通信网络分为4大类:

- (1) 电话网络;
- (2) 计算机网络;
- (3) 有线电视网络;
- (4) 无线网络。

1.2.2 常用的通信网络

1. 电话网络

电话网络是最早的通信网络,专门为语音通信而设计。我们现在使用的电话网络称为**公共交换电话网络**(Public Switched Telephone Network, PSTN),主要提供远程的语音通信能力。随着光纤技术及其数字技术的发展,电话网络也可以承载远程数据通信的任务。

现在电话网络中的关键技术和概念包括以下几个方面。

(1) 电路交换方式。

(2) 数字化技术。包括语音脉冲编码调制(PCM),对语音进行数字化和编码,并以数字方式在电话网络中传送。数字载波系统(Digital Carrier System, DCS)和数字信号(Digital Signal, DS)系列是电话网络数字时分多路复用的规范,如我国和欧洲的E1数字接口线路,北美的T1数字接口线。

(3) 公共信道信令。公共信道信令(Common Channel Signaling, CCS)就是把控制信令用共用的网络传输,与语音传输分离,这样便于计算机控制。

(4) 光纤链路。20世纪80年代开始部署光纤链路,采用数字时分多路复用方式支持电话网络的语音和数据传输。著名的光纤链路规范有北美的同步光纤网络(Synchronous Optical Network, SONET)和欧洲及其我国的同步数字系列(Synchronous Digital Hierarchy, SDH)。

(5) ISDN。电话网络的设计者也想为电话网络赋予更多的能力,不仅仅提供语音通信,还可以提供计算机数据通信和视频通信的能力,这就是综合业务数字网络(Integrated Services Digital Network, ISDN)。

图1.2、图1.3和图1.4所示为电话网络发展的简要过程。图1.2所示为早期的电话网络,终端是电话机,所有的线路都是模拟线路。早期的电路交换是人工插接的,即由话务员人工插线实现交换。“信令”形式表现为用户与话务员之间的对话,一个用户要求接通某某用户,话务员就人