



● **新世纪** 网络课程教材

现代多媒体技术 应用教程

罗万伯 主编

陈 炜 罗霄峰 徐群英 等编



高等教育出版社

<http://www.hep.com.cn>
<http://www.hep.edu.cn>

新世纪网络课程教材

现代多媒体技术应用教程

罗万伯 主编

陈 炜 罗霄峰 徐群英 等编

高等教育出版社

内 容 提 要

本书介绍现代多媒体技术及应用,内容包括多媒体技术概述、知觉基础、数据编码、音频处理、图像技术、图像设备及工具、视频技术、多媒体集成交互及开发应用、多媒体文档、多媒体存储技术、多媒体网络及应用、多媒体内容处理及多媒体信息安全,覆盖了多媒体技术的绝大多数知识点。本书内容完整、详细、全面,选材新颖,如 MPEG 标准最新进展、基于内容的检索技术、流媒体技术、多媒体信息安全等方面都体现了多媒体领域的最新研究成果。

本教材配有教学课件和相关教学材料,可供教师授课选用。

本书可作为计算机、电子、通信和信息技术等专业本科生教材,对上述专业的研究生、从事多媒体技术应用及开发的人员也具有参考价值。

图书在版编目(CIP)数据

现代多媒体技术应用教程/罗万伯主编. —北京: 高等教育出版社, 2004.11

ISBN 7-04-015129-4

I. 现... II. 罗... III. 多媒体技术—教材
IV. TP37

中国版本图书馆 CIP 数据核字 (2004) 第 113755 号

策划编辑 董建波 责任编辑 刘 英 市场策划 陈 振
封面设计 王凌波 责任印制 孔 源

出版发行 高等教育出版社
社 址 北京市西城区德外大街 4 号
邮政编码 100011
总 机 010-58581000

购书热线 010-64054588
免费咨询 800-810-0598
网 址 <http://www.hep.edu.cn>
<http://www.hep.com.cn>

经 销 新华书店北京发行所
印 刷 北京星月印刷厂

开 本 787×1092 1/16
印 张 40.75
字 数 850 000

版 次 2004 年 11 月第 1 版
印 次 2004 年 11 月第 1 次印刷
定 价 38.50 元

本书如有缺页、倒页、脱页等质量问题,请到所购图书销售部门联系调换。

版权所有 侵权必究

物料号: 15129-00

前 言

多媒体技术是一种覆盖面很宽的技术,是多种技术特别是计算、通信和广播电视技术发展、融合、渗透的结果。多媒体技术又是世界上发展最快的技术之一,该技术领域文献、技术和标准更新迅速。

作者在跟踪、研究 ACM 和 IEEE/CS 制订的《Computing Curricula 2001》基础上,结合十多年来在多媒体技术研究与开发的经验和教学实践,力求全面系统地介绍现代多媒体技术的原理及应用,既注重理论、方法和标准的结合,又兼顾实际应用技术,努力将最新的多媒体技术材料组织成教材。例如,图像中的 SVG、JBIG2、JPEG-LS、JPEG 2000 等,运动图像的 MPEG-4,超媒体的 MHEG-5、SMIL,已成为标准的 MPEG-7,正在开发的 MPEG-21 等。而多媒体网络技术中的多播、流媒体技术,IP 网络的服务质量,以及成为热点的多媒体信息检索、多媒体信息安全等则是现有教材中较少或没有涉及的。

全书共 13 章。其中第 1~9 章讲述最基本的内容,第 10、11 章侧重于多媒体开发应用,第 12、13 章侧重于对多媒体技术和应用研究与开发。读者可以根据情况有选择地学习。每章的习题可以帮助读者巩固基本概念和知识。建议在学习本教材的同时,进行适当的练习和上机操作。附录 B 中的建议,可供选用本教材的教师参考。本教材还配有教学课件和相关教学材料,可供教师授课选用。

本教材主要适用对象是计算机、电子、通信和信息技术等专业本科生。对上述专业的研究生、从事多媒体技术应用及开发的人员,本教材也有很大的参考价值。

本教材在编写过程中参考、学习了有关的多媒体教材、文献和网站,并引用了一些材料,在此对这些作者表示衷心的感谢。

本教材由罗万伯主编,陈炜、罗霄峰、徐群英、彭舰、冉蜀阳、董毅和胡晓等参与编写。有很多同事和研究生为本书的撰写工作提供了帮助,在此一并表示感谢。

由于书稿涉及许多新内容和研究课题,尽管作者已尽了最大努力,囿于学识和水平,仍自感疏漏难免。诚望读者不吝赐教斧正,以利再版修订,至臻完善。

作 者

E-mail:luowanbo@cs.scu.edu.cn

2004 年 6 月 8 日

目 录

第1章 多媒体技术概述	(1)	思考与练习题	(55)
1.1 媒体和多媒体的概念	(1)	第3章 数据编码	(56)
1.1.1 媒体	(1)	3.1 数据压缩概述	(56)
1.1.2 多媒体	(2)	3.1.1 为什么要压缩	(56)
1.1.3 多媒体分类和描述	(2)	3.1.2 数据的冗余性	(57)
1.1.4 数字媒体	(5)	3.1.3 压缩和解压缩过程	(58)
1.2 多媒体标准和系统	(7)	3.1.4 压缩评价和分类	(59)
1.2.1 多媒体标准	(7)	3.2 赫夫曼编码	(61)
1.2.2 多媒体系统的概念	(8)	3.3 算术编码	(63)
参考文献	(12)	3.4 行程长度编码	(66)
思考与练习题	(12)	3.5 词典编码	(68)
第2章 知觉基础	(13)	3.5.1 词典编码分类	(68)
2.1 听觉	(13)	3.5.2 LZ77 算法	(69)
2.1.1 发音器官	(13)	3.5.3 LZSS 算法	(72)
2.1.2 听觉特性	(13)	3.5.4 LZ78 算法	(73)
2.1.3 听觉心理	(15)	3.5.5 LZW 算法	(75)
2.1.4 立体声效	(20)	3.6 预测编码	(79)
2.1.5 语音评测	(21)	3.6.1 PCM	(79)
2.2 视觉	(24)	3.6.2 DPCM	(81)
2.2.1 电磁波和人眼	(24)	3.6.3 ADPCM	(83)
2.2.2 视觉心理学	(25)	3.6.4 帧间预测编码	(84)
2.2.3 视觉特性度量	(30)	3.6.5 运动图像帧间内插	(86)
2.2.4 色空间	(33)	3.7 变换编码	(87)
2.2.5 计算机视频基础	(38)	3.7.1 变换的基本原理	(87)
2.2.6 图像质量评价	(51)	3.7.2 离散傅里叶变换	(88)
2.3 其他知觉	(52)	3.7.3 离散余弦变换	(89)
2.3.1 皮肤生理	(52)	3.7.4 小波变换	(91)
2.3.2 触觉	(52)	3.8 基于模型编码	(96)
2.3.3 动觉	(53)	3.8.1 基于语义编码	(97)
2.3.4 平衡感觉	(53)	3.8.2 基于物体编码	(98)
2.4 知觉的量度和复合作用	(54)	3.9 分形编码	(99)
参考文献	(54)	3.9.1 分形编码的思路	(99)

3.9.2 分形编码方法和步骤	(100)	4.7 语音识别	(155)
3.10 其他压缩编码	(101)	4.7.1 语音识别发展历史	(155)
3.10.1 子带编码	(101)	4.7.2 语音识别技术	(156)
3.10.2 向量量化编码	(103)	4.7.3 困难与对策	(158)
3.10.3 感知编码	(105)	4.7.4 语音识别的应用	(159)
3.11 测错和纠错编码	(105)	4.8 3D 音效	(160)
3.11.1 CRC	(105)	4.8.1 人类的听觉和 HRTF	(161)
3.11.2 RS 编码	(106)	4.8.2 3D 音效的分类	(162)
3.11.3 CIRC 编码	(107)	4.9 音频应用	(163)
3.11.4 RSPC 编码	(107)	4.9.1 音频卡的应用	(163)
3.11.5 EFM 编码	(107)	4.9.2 音频文件格式	(169)
参考文献	(108)	参考文献	(170)
思考与练习题	(108)	思考与练习题	(171)
第 4 章 音频处理	(111)	第 5 章 图像技术	(172)
4.1 音频信号	(111)	5.1 数字图像	(172)
4.1.1 声音的特征指标	(111)	5.1.1 向量图	(172)
4.1.2 数字音频技术指标	(112)	5.1.2 位图	(173)
4.2 音频数字压缩	(113)	5.1.3 图像的数据容量和压缩	(173)
4.2.1 音频压缩编码技术	(113)	5.2 二值图像压缩	(174)
4.2.2 音频编码技术标准	(115)	5.2.1 3 类传真	(175)
4.3 MPEG-1 音频	(117)	5.2.2 4 类传真	(175)
4.3.1 MPEG-1 音频标准	(117)	5.2.3 JBIG 传真标准	(176)
4.3.2 声音编码系统基本结构	(118)	5.3 JPEG 标准	(179)
4.3.3 滤波器组	(119)	5.3.1 概述	(179)
4.3.4 通用编码概念	(119)	5.3.2 基于 DCT 的顺序模式	(182)
4.3.5 三层音频系统	(120)	5.3.3 基于 DCT 的渐进模式	(186)
4.4 MPEG-2 音频	(127)	5.3.4 无损模式	(188)
4.4.1 MPEG-2 音频特点	(127)	5.3.5 分层模型	(189)
4.4.2 MPEG-2 AAC	(128)	5.4 JPEG-LS	(190)
4.4.3 AC-3 编码	(131)	5.5 JPEG 2000	(191)
4.5 MPEG-4 音频	(133)	5.5.1 概述	(191)
4.5.1 自然音频合成	(135)	5.5.2 JPEG 2000 编码解码器	(191)
4.5.2 合成声音	(142)	5.5.3 文件格式	(197)
4.5.3 AudioBIFS	(146)	5.5.4 小波变换	(198)
4.5.4 音频轮廓	(149)	5.6 GIF	(201)
4.6 电子音乐合成与 MIDI	(150)	5.7 PNG	(203)
4.6.1 电子音乐合成	(150)	5.7.1 简介	(203)
4.6.2 电子乐器数字接口	(152)	5.7.2 概念	(203)

5.7.3 数据块	(207)	参考文献与参考网站	(246)
5.8 动画	(209)	思考与练习题	(247)
5.9 SVG	(211)	第7章 视频技术	(248)
5.9.1 概述	(211)	7.1 MPEG-1 和 MPEG-2 视频	
5.9.2 模块化	(212)	标准	(248)
5.9.3 基本概念	(213)	7.1.1 MPEG 视频算法	(248)
5.9.4 SVG 图像示例	(213)	7.1.2 MPEG 系统	(261)
5.9.5 Web 页使用 SVG	(215)	7.2 MPEG-4 视频	(265)
参考文献与参考网站	(215)	7.2.1 MPEG-4 系统概述	(265)
思考与练习题	(216)	7.2.2 视频编码	(270)
第6章 图像设备及工具	(218)	7.2.3 视频轮廓	(278)
6.1 显示器	(218)	7.2.4 MPEG-4 文件格式	(280)
6.1.1 显示器概述	(218)	7.3 H.26x 标准系列	(281)
6.1.2 CRT 显示器	(218)	7.3.1 H.261 和 H.263	(281)
6.1.3 LCD	(220)	7.3.2 H.264	(283)
6.1.4 等离子显示板	(222)	7.4 M-JPEG	(287)
6.1.5 背投显示技术	(223)	7.4.1 Motion JPEG 2000 概述	(288)
6.1.6 抖动和 HAD 技术	(224)	7.4.2 核心概念	(288)
6.2 图像卡	(224)	7.4.3 媒体的物理结构	(289)
6.2.1 图像卡的组成部件	(225)	7.4.4 媒体的时域结构	(289)
6.2.2 图像卡标准	(226)	7.4.5 交错	(289)
6.2.3 3D 加速	(227)	7.4.6 组合	(290)
6.2.4 API	(228)	7.4.7 文件组织	(290)
6.2.5 硬件接口	(228)	参考文献	(290)
6.3 图像打印设备	(229)	思考与练习题	(291)
6.3.1 激光打印机	(229)	第8章 多媒体集成交互及开发应用	(292)
6.3.2 页描述语言	(231)	8.1 服务质量	(292)
6.3.3 喷墨打印机	(231)	8.2 多媒体同步	(294)
6.3.4 其他类型的打印机	(232)	8.2.1 什么是同步	(294)
6.4 图像输入设备	(232)	8.2.2 时间同步的说明	(297)
6.4.1 信号拾取器材和接口	(233)	8.2.3 SMIL	(303)
6.4.2 图像扫描仪	(234)	8.3 唇读	(305)
6.4.3 OCR	(235)	8.4 数字电视	(308)
6.4.4 数码相机	(235)	8.4.1 数字电视系统的结构	(309)
6.4.5 摄像头和摄像机	(237)	8.4.2 数字电视方案	(310)
6.4.6 触摸屏	(238)	8.4.3 ITV 和 VOD	(319)
6.5 颜色管理	(241)	8.4.4 多媒体家庭平台	(323)
6.6 常见图像和视频文件扩展名	(243)	8.5 3DAV	(324)

8.6 虚拟现实	(327)	10.1 半导体存储技术	(384)
8.6.1 虚拟现实的特征	(327)	10.1.1 RAM	(384)
8.6.2 虚拟现实的系统构成	(328)	10.1.2 ROM	(386)
8.6.3 实现虚拟现实的关键技术	(329)	10.2 磁盘存储系统	(387)
8.6.4 VRML 简介	(330)	10.2.1 硬盘性能	(388)
8.7 多媒体应用开发	(332)	10.2.2 RAID 技术	(389)
8.7.1 多媒体应用软件的开发 过程	(332)	10.3 光碟存储系统	(391)
8.7.2 多媒体著作工具	(334)	10.3.1 概述	(391)
8.7.3 单质媒体工具	(338)	10.3.2 CD	(392)
8.7.4 多媒体非线性编辑	(339)	10.3.3 DVD	(409)
参考文献与参考网站	(342)	10.4 存储策略	(419)
思考与练习题	(342)	10.4.1 直接相连存储	(420)
第9章 多媒体文档	(343)	10.4.2 NAS	(421)
9.1 文档和超文本	(343)	10.4.3 SAN	(423)
9.1.1 超文本和超媒体概念	(343)	10.4.4 IP 存储协议	(424)
9.1.2 超媒体系统的基本元素	(345)	参考文献与参考网站	(425)
9.1.3 超媒体结构模型	(349)	思考与练习题	(426)
9.2 超文本的文档模型	(357)	第11章 多媒体网络及应用	(427)
9.2.1 ODA 模型	(358)	11.1 多媒体通信概述	(427)
9.2.2 SGML 模型	(358)	11.1.1 多媒体通信的特点和对网络 要求	(427)
9.2.3 HyTime 模型	(360)	11.1.2 当前网络对多媒体通信的 支持	(430)
9.3 Internet 上的超媒体系统 WWW	(361)	11.2 多媒体 Internet 多播及流技术	(431)
9.3.1 概述	(361)	11.2.1 多播	(431)
9.3.2 HTTP	(361)	11.2.2 RTP 和 RTCP	(436)
9.3.3 HTML	(362)	11.2.3 RTSP	(440)
9.3.4 XML	(366)	11.2.4 音视频流在 Internet 上的 传输	(441)
9.3.5 动态网页生成技术	(369)	11.3 Internet 服务改进	(443)
9.3.6 XHTML	(370)	11.3.1 调度和管制机制	(443)
9.3.7 Web 系统的关键技术	(371)	11.3.2 综合服务与资源预留协议	(446)
9.3.8 语义 Web	(373)	11.3.3 区分服务	(450)
9.4 MHEG 标准	(375)	11.3.4 IPv6 的改进	(452)
9.4.1 MHEG 概述	(375)	11.4 多协议标记交换	(453)
9.4.2 MHEG 的对象和类	(376)	11.4.1 基本工作过程	(453)
9.4.3 MHEG-5	(381)	11.4.2 转发等价类	(455)
参考文献与参考网站	(382)	11.4.3 标记栈	(456)
思考与练习题	(383)		
第10章 多媒体存储技术	(384)		

11.4.4 标记对换	(457)	12.4.7 BiM	(567)
11.5 居民接入网	(458)	12.4.8 MPEG-7 终端	(567)
11.5.1 xDSL 技术	(458)	12.5 基于内容的检索技术	(568)
11.5.2 光纤同轴混合网	(461)	12.5.1 基于内容的多媒体信息	
11.5.3 FTTx 技术	(463)	检索	(569)
11.5.4 以太网接入	(464)	12.5.2 网络多媒体信息查询	(574)
11.5.5 无线接入	(464)	参考文献与参考网站	(579)
11.6 分布式多媒体应用	(465)	思考与练习题	(580)
11.6.1 分布式多媒体系统基本		第 13 章 多媒体信息安全	(581)
概念	(465)	13.1 概述	(581)
11.6.2 MPEG-21 多媒体框架	(466)	13.1.1 多媒体信息的威胁和攻击	(581)
11.6.3 计算机支持的协同工作	(478)	13.1.2 多媒体信息安全的要素	(583)
11.6.4 通用数字视听系统	(481)	13.1.3 多媒体安全服务及安全	
11.6.5 多媒体会议系统	(483)	机制	(583)
11.6.6 其他网络多媒体应用	(493)	13.1.4 当前的主要问题和解决	
参考文献与参考网站	(497)	方案	(588)
思考与练习题	(498)	13.2 音频视频信息的保护	(588)
第 12 章 多媒体内容处理	(499)	13.2.1 一般保护技术	(588)
12.1 多媒体数据管理	(499)	13.2.2 信息隐藏	(589)
12.1.1 多媒体数据管理环境	(499)	13.2.3 数字水印	(593)
12.1.2 多媒体数据库	(501)	13.3 数字权限管理	(601)
12.2 多媒体内容分析	(505)	13.3.1 知识产权	(601)
12.2.1 多媒体数据及内容的基本		13.3.2 条件访问	(602)
概念	(505)	13.3.3 DRM 概述	(603)
12.2.2 音频分析	(508)	13.3.4 DRM 系统	(603)
12.2.3 视频分析	(515)	13.4 MPEG 的 IPMP	(606)
12.3 元数据	(536)	13.4.1 MPEG 的 IPMP 概述	(607)
12.3.1 概述	(536)	13.4.2 MPEG-2 IPMP	(609)
12.3.2 DCMI 元数据	(537)	13.4.3 MPEG-4 IPMP	(611)
12.3.3 HTML 元数据元素	(538)	13.4.4 MPEG-21 IPMP	(614)
12.4 MPEG-7	(540)	13.5 其他的内容保护方案	(620)
12.4.1 MPEG-7 的目标和范围	(540)	参考文献与参考网站	(623)
12.4.2 MPEG-7 的主要元素	(542)	思考与练习题	(623)
12.4.3 描述定义语言	(544)	附录	(624)
12.4.4 MPEG-7 多媒体描述		附录 A 多媒体研究资源汇总	(624)
方案	(545)	附录 B 综合练习	(626)
12.4.5 MPEG-7 视频	(549)	附录 C 多媒体缩略语	(627)
12.4.6 MPEG-7 音频	(559)		

第1章 多媒体技术概述

多媒体是科学技术发展,特别是计算技术、通信技术和电视技术发展的必然结果。现代多媒体技术形成于20世纪80年代,是计算机、广播电视和通信这三大原来各自独立的领域相互渗透、相互融合,进而迅速发展起来的一门新兴技术。多媒体技术的一个例子就是多媒体计算机,它一出现,很快在世界范围内,在家庭教育和娱乐方面得到广泛的应用,并由此引发了小型激光视盘(VCD和DVD)的诞生,促进了数字电视和高清晰度电视(HDTV)的迅速发展。

但是,什么是多媒体?什么是多媒体技术?多媒体技术有哪些特点?本章将对有关媒体、多媒体、多媒体技术等基本概念进行简要介绍,对多媒体技术的发展进行回顾和展望。

1.1 媒体和多媒体的概念

1.1.1 媒体

“媒体(medium)”的原意是“things in the middle”,又称载体或介质,指信息传送和存储的最基本的技术和手段。根据信息被人们感知、加以表示、使之呈现、实现存储或进行传送的载体的不同,基本媒体可分为感知媒体(perception medium)、表示媒体(representation medium)、表现媒体(presentation medium)、存储媒体(storage medium)和传输媒体(transmission medium)五类。

① 感知媒体,指人的感觉器官所能感觉到的信息的自然种类,如声音、图形、图像、动画、电影和文本等。人的感觉包括视觉、听觉、触觉、嗅觉、味觉等。感知媒体帮助人类感知环境的信息。目前,人类主要靠视觉和听觉来感知环境的信息。触觉作为一种感知方式也慢慢引入到计算机中。

② 表示媒体,指被交换的数据类型,它们定义信息的特性。表示媒体的特性用信息的计算机内部编码表示,例如语音PCM编码、图像JPEG编码、文本ASCII编码和乐谱等。

③ 表现媒体,指为人们再现信息的物理工具和设备(输出设备),或者指获取信息的工具和设备(输入设备),如显示器、扬声器、打印机等输出类表现媒体,以及键盘、鼠标器、扫描器等输入类表现媒体。

④ 存储媒体,指存储数据的物理介质,如纸张、磁盘、光碟等。

⑤ 传输媒体,指传输数据的物理媒介,如双绞线、同轴电缆、光缆、无线电链路等传输媒体。

基本媒体可能组合起来。例如,交换媒体,是系统之间交换信息的手段和技术,可以是存储媒体、传输媒体,或两者的结合。

1.1.2 多媒体

“多媒体(multimedia)”一词大约在1960年至1965年开始使用,其中“multi”是“许多”的意思。术语多媒体的原意是指“通过计算机、电视、电话等散布和表现用图形、动画、音频、视频、文本等格式编码的信息”。也就是说,多媒体是由多种媒体数据以及这些数据的一系列交互组成的,即

多媒体 = 多种媒体数据 + 一系列交互

所谓多种媒体数据是多种源的、多种类型的和多种格式的数据(multisource, multitype and multiformat data)的集合。多媒体成分之间的交互则由复杂的关系组成,如果没有这些关系,多媒体只能是音频数据、视频数据与其他数据的简单集合。

交互是人类获取信息、掌握知识的重要途径。研究表明:看见的信息人们只能记住20%;听见的信息人们只能记住30%;既看见同时又听见的信息人们能记住50%;既看见同时又听见并且也参与的信息人们能记住80%。

多媒体技术涉及许多方面,其中最重要的有多媒体信号处理技术、多媒体数据存储技术、传输技术、表现技术、安全技术等。通常将多媒体信号处理定义为使用信号处理工具对多媒体数据进行表示、解释、编码和解码。多媒体信号处理的目的是为各种多媒体应用提供实际和有效的多媒体内容存取、操作、交换和存储。在现代社会中,多媒体的采集或生成、处理、存储、传送和表现等过程,是离不开计算机的。现代社会广泛使用的计算机是数字计算机,在数字计算机中的信息都是数字化的信息。多媒体信息当然也不例外。

需要指出的是,一般所说的“多媒体”,不仅指多种媒体信息本身,而且指处理和应用多媒体信息的相应技术,因此“多媒体”常被当做“多媒体技术”的同义词。

与多媒体有关的术语还有多媒体应用和多媒体系统。关于多媒体应用和多媒体系统有很多定义。区分一个应用或系统是否是多媒体应用或多媒体系统的三条主要标准是:媒体的数目、支持多媒体的类型和媒体的集成化程度。应用中的媒体数目是最基本的一条标准。支持处理多种媒体,其中至少包含一种时间相关媒体,以及支持多种类型媒体集成化处理,是衡量是否属多媒体应用和系统的标准。

1.1.3 多媒体分类和描述

从人类感知角度划分,媒体可以分为听觉媒体、视觉媒体、触觉媒体、嗅觉媒体和味觉媒体等5大类。人类从外部世界获取信息的70%~80%从视觉获得,10%从听觉获得,其余的

10%从触觉、嗅觉和味觉获得。听觉媒体又称为可聆听(audio)媒体,视觉媒体又称为可视(visual)媒体。现在多媒体包含的主要是可听和可视两大类。

可听媒体的例子如人说话的话语声(speech)、乐器演奏的音乐声(music)、动物声音、风声等。在不强调声源时,一般统称声音(sound或voice)。声音可以是自然形成的(natural),也可以是人工合成的(synthesis)。

可视媒体的例子如文本(text)、图形(graphics)、图像(picture或image)、运动图像(motion picture)或视频(video)。可视多媒体可以是自然形成的,也可以是人工合成的(例如动画)。

不同媒体形式的信息其表示一般也是不同的。比如文本通过由一系列字符组成的句子来表达信息,语音以声波的形式表达信息。它们都是用值(Value)来表示的。某些媒体的表达值是用所谓的自包含的量或物理量来描述的,例如温度、味觉、嗅觉,其他的媒体需要先定义一套符号(或达成协议),人们必须通过学习来理解这些信息。这类媒体包括文本、语音和手势。

表达值可以是一个连续值序列或者一个单值序列。声波是压力波的波动引起的,它不会产生离散的值。可被人眼察觉的电磁波没有时间的间隔,这意味着它是连续的值。一段文本的特征以及一段音频的采样值是由一系列单一的值组成的。

媒体需要在一定的表现域中才能表现(presentation)。表现域又称为表现空间,可以是时间域、空间域或时间-空间域。纸张和计算机显示器都是视觉表现空间的例子。计算机幻灯片表现时在整个放映屏,放映一屏的内容也是一个可视化的表现空间。立体声和四声道定义的是声音的表现空间。压力手套(pressure gloves)是一个触觉媒体的表现空间,气动仿真器(pneumatic simulator)是一个平衡感官的表现空间等。表现空间是上述用来输出信息的显示媒体的一部分。

每一个表现空间有一个或多于一个的表现维(presentation dimension)。一个计算机的显示器是二维空间,全息摄影和立体声需要三维空间。超文本(hypertext)的表现空间是网络空间或赛伯空间(cyberspace)。立体电影需要四维时空域,即 X 、 Y 和 Z 轴,再加上时间轴。

如果媒体的显示域是一个连续域,这种媒体一般称为模拟媒体(analog medium)或连续媒体(continuous medium)。声音是一个时间域的模拟媒体示例,黑白照片则是一个二维空间域的模拟媒体示例,运动图像则是时间-空间域的模拟媒体。

如果媒体的显示域是一些离散的点,这种媒体一般称为离散媒体(discrete medium)。

在显示域里的时间维对多媒体系统来说很关键。按照显示域里的时间维,媒体被分为非时间媒体或时间离散媒体与时间媒体或时间连续媒体两类。

1. 非时间媒体或时间离散媒体

文本、图形和图像等属于此类(图1.1.1),它们由不依赖于时间的信息组成。实际上,它们可以根据各种定时甚至排序来显示其信息。人们常说时间不是离散媒体语义的一部

分。由于现代基于计算机的文本和图形显示经常是数值离散、时间连续的,因而“离散”这个术语有点模糊。例如,一本书的文本是离散的媒体。处理离散媒体的各种方法应该是尽可能快的。另一方面,时间不是重要因素,因为数据的有效性(也就是正确性)并不依靠于时间状况(至少不在几秒或更少的时间帧内)。

2. 时间媒体或时间连续媒体

声音和运动图像等属于此类,其显示要求随时间变化连续地播放(图 1.1.2)。换句话说,时间或者更确切地说信息单元之间的时间依赖性信息本身的一部分。如果定时信息或者信息单元的顺序发生变化,其意义就会改变。时间是连续媒体语义的一部分,连续媒体也被称为时间媒体。当处理连续媒体时,从技术上来说另一个结果就是连续媒体的传输网络同样需要支持这种时间依赖性。



图 1.1.1 时间离散媒体

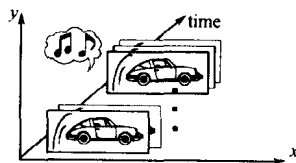


图 1.1.2 时间连续媒体

因为这些媒体处理的数据的有效性(正确性)取决于时间条件,所以时间很关键。例如,如果一个音频采样值传输得太晚就可能无效或者出错,跟在这个值后的音频数据可能已被扩音器播完。在音频和视频数据中,表现值形成一个连续序列,其中视频意味着纯粹的运动图像。电视或电影中声音和运动图像联合,与视频这一术语不同义。由于这一原因,它们被称为连续媒体。当区分非周期出现的依赖时间的表现值时,经常不把它们放在连续媒体的范畴。对于一个多媒体系统,也必须考虑这些表现值的非连续序列。当信息被指示器(如鼠标)捕捉住并在具有共同显示窗口的协作应用中传输,就会出现这种显示值序列。这时,连续媒体和依赖时间的媒体是同义的。这种定义中,连续媒体是自然或人工起源的视频(运动图像),通常被存储为数字化压力波采样值序列的音频和来自各种传感器(如空气压力、温度、湿度、压力)或放射性传感器的信号。

时间媒体这一术语描述一种时间上离散或连续的媒体,但并没涉及内部数据表示。例如,按这种方式展示媒体指的是观察者或听众得到的印象。电影的例子显示出连续媒体数据经常由离散值序列组成,这些离散值在表示空间中作为时间的函数一个接着一个。在此例中由于人眼的视觉机制(视觉暂留),每秒至少 16 幅的图像序列则给人以连续的印象。

本书在此后的叙述中,如果不特别说明,连续媒体均指时间连续媒体,离散媒体均指时间离散媒体。

从网络连接的角度,媒体可以分为非实时媒体和实时媒体两大类。根据应用的不同,上述时间离散媒体既可能是非实时的,也可能是实时的(图 1.1.3)。

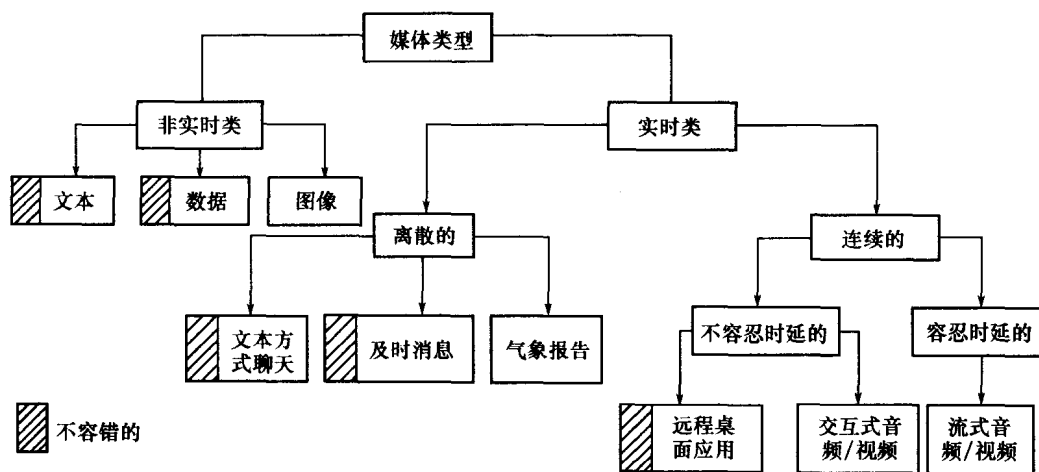


图 1.1.3 媒体的网络分类

1.1.4 数字媒体

无论是连续媒体,还是离散媒体,其表达值可能是连续范围内的任意数值,也可能是一些离散的数值。如果一个离散媒体的表达值只能是离散的数值系列,则其称为数字媒体(digital medium)。例如,定期人口统计报告就是一个数字媒体示例。

在多媒体技术中,绝大多数数字媒体是对模拟媒体进行适当处理而得到的(图 1.1.4)。处理大致分为两步:

第一步,对模拟媒体在显示域进行采样(sampling),形成离散媒体。

第二步,对离散媒体的表达值(即第一步得到的样本值)进行量化(quantization)。

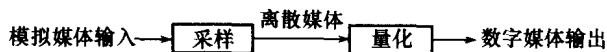


图 1.1.4 模拟-数字转换

上述过程又称为模拟-数字转换(Analog-Digital Convert, ADC)。下面以声音信号为例,说明此过程。

在某些特定的时刻对声音模拟信号进行采样。采样得到的幅值是无穷多个实数值中的一个,因此幅度还是连续的。然后把信号幅度取值的数目加以限定,例如,假设输入电压的范围是 $0 \sim 0.7 \text{ V}$,并假设它的取值只限定在 $0, 0.1, 0.2, \dots, 0.7$ 共 8 个值。如果采样得到的幅度值是 0.123 V ,它的取值就应算做 0.1 V ,如果采样得到的幅度值是 0.26 V ,它的取值就应算做 0.3 V (假定采用四舍五入)等。这种过程就是“量化”。

如果对模拟信号,每隔相等的一小段时间采样一次,这种采样称为均匀采样(uniform

sampling); 否则称为非均匀采样。通常采用均匀采样。图 1.1.5 所示为一维模拟信号的采样。连续幅度的离散化通过量化来实现,就是把信号的强度划分成多个小段,并用某种准则(例如四舍五入)将其数值表示出来。如果幅度的划分是等间隔的,就称为线性量化,否则称为非线性量化。图 1.1.6 所示为线性量化的示例。

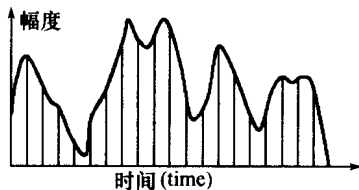


图 1.1.5 一维模拟信号的采样

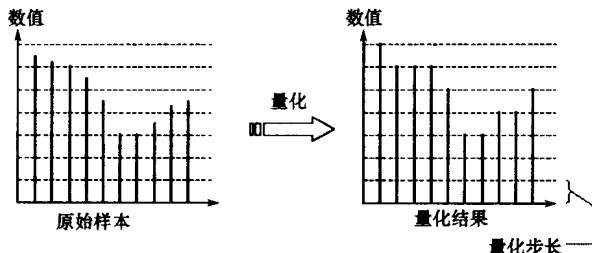


图 1.1.6 线性量化的示例

模拟-数字转换需要解决两个问题:每秒需要采集多少个信号样本,也就是采样频率(f_s)是多少;每个信号样本的比特数 b/s (bit per sample) 应该是多少,也就是量化精度。

根据奈奎斯特理论 (Nyquist theory), 采样频率是由模拟信号本身的最高频率决定的。奈奎斯特理论指出,采样频率不应低于模拟信号最高频率的两倍,这样就能把以数字表达的信号还原成原来的信号,这叫做无损数字化 (lossless digitization)。公式表示为

$$f_s \geq 2f \text{ 或者 } T_s \leq T/2$$

式中, f 为被采样信号的最高频率, T 为被采样信号的最低周期, f_s 称为采样频率, T_s 为采样间隔。

可以这样来理解奈奎斯特理论:模拟信号可看成是由许许多多正弦波组成的,一个振幅为 A 、频率为 f 的正弦波至少需要两个采样样本表示,因此,如果一个信号中的最高频率为 $f_{\max}$, 采样频率最低要选择 $2f_{\max}$ 。例如,电话语音的信号频率约为 3.4 kHz, 采样频率通常选为 8 kHz。

模拟运动图像的数字化要复杂一些。首先是在时间维上采样,得到静止图片;然后对每幅图片在水平方向采样 (X 轴), 得到行图像;然后对每个行图像采样 (Y 轴);最后对得到的每个样值进行量化。

如果要在模拟设施上表现数字媒体,则需要将数字媒体首先转换成模拟媒体。数字媒体转换成模拟媒体的过程称为数模转换 (Digital-Analog Convert, DAC)。

数字多媒体涉及的关键问题很多,既涉及性能方面,又涉及质量方面。性能方面的关键问题如带宽、存储容量、处理能力等,质量方面包括实时性 (real time)、容错 (error tolerance) 和同步 (synchronization) 等。

1.2 多媒体标准和系统

1.2.1 多媒体标准

多媒体技术的迅猛发展,究其原因,除了有关技术的进步和社会的需求外,也和相关标准的开发密切联系。

什么叫做标准?所谓标准是记录在案的一致意见(documented agreements),包含技术规范或其他被一致地用做规则(rules)、指南或角色定义的精确的准则,以便保证材料、产品、过程和服务合乎各自的目的。

有很多组织机构开发和发布与多媒体技术有关的协议,包括国际组织、区域组织、政府机构、企业事业和研究单位等。下面对影响最大的几个国际组织以及中国的标准化组织进行简单介绍。

与多媒体标准开发密切有关的国际组织很多。国际标准化组织(International Organization for Standardization, ISO)是其中最重要的国际组织之一。ISO是一个涉及范围很广的开发标准的国际机构。与多媒体有关的很多ISO标准实际上是ISO与国际电工委员会(International Electrotechnical Commission, IEC)合作工作的结果(主要工作组是JTC 1)。JTC 1中的SC29与多媒体有关。ISO/IEC JTC1/SC29有3个工作组从事多媒体系统交换格式的标准化,即WG1(JPEG和JBIG)、WG11(MPEG)和WG12(MHEG)。他们开发的标准MPEG系列(包括MPEG-1、MPEG-2、MPEG-4、MPEG-7和MPEG-21)、MHEG、JPEG系列(包括JPEG-LS、JPEG2000)、JBIG、SGML、HyTime、ODA/HyperODA、AVIS、PREMO、ODP/ANSA、CGM等标准都是与多媒体有关的ISO标准。

另一个重要的国际组织是ITU-T(International Telecommunication Union-Telecommunication Standards Sector,国际电信联盟-电信标准部),它是国际电信联盟(ITU)的永久成员机构。音频标准G. 711、G. 721、G. 722、G. 723、G. 728等,视频标准H. 261、H. 263和H. 264,以及视频会议标准H. 320、H. 323、H. 324等就是它开发的。ITU-T在2001年—2004年期间的研究小组中,SG 9(Integrated broadband cable networks and television and sound transmission,集成宽带电缆网络及电视和声音传输)和SG 16(Multimedia services, systems and terminals,多媒体服务、系统和终端)直接与多媒体有关。

另外两个重要的国际组织是IEFT(Internet Engineering Task Force,因特网工程任务组)和W3C(The World Wide Web Consortium,万维网联盟),它们的目标之一是发展因特网上的多媒体应用。著名的RSVP、DiffServ以及直接支持多媒体的IPv6等都是IEFT开发的。W3C开发了HTML、XML、XHTML、PNG、SMIL、SVG等标准。

中国国家标准化管理委员会(Standardization Administration of the People's Republic of

China, SAC) 是中国的官方标准化组织。与多媒体有关的标准主要由信息技术标准化技术委员会(对口 ISO/IEC/JTC 1)归口管理。ISO/IEC 的许多多媒体标准已经或正在被接纳为中国国家标准(GB 或 GB/T)。

1.2.2 多媒体系统的概念

1.2.2.1 多媒体系统及特性

多媒体系统是一种支持多种来源、多种类型和多种格式的多媒体数据的交互式信息系统。

显然,多媒体系统必须是计算机控制的。这样在将信息显示给用户时,至少有一台计算机是集成化的,也就是说,使用最小数目的不同设备。例如使用计算机显示屏显示各种视觉信息,它们必须支持媒体独立性。最后,它们能处理离散和连续媒体。下面将描述这些关键特性。

1. 离散和连续媒体

并不是媒体的任意合并就可以称为多媒体。由于处理嵌入式图形的简单的字处理器可以处理两种媒体,许多人不恰当地将它称为多媒体应用。根据一般的定义,如果一种应用同时使用离散和连续媒体,才可以称为多媒体。这意味着一个多媒体应用至少处理一种离散和一种连续媒体。根据这一定义,处理嵌入式图形的字处理器并不是一个多媒体应用。

2. 独立媒体

多媒体系统中所操作的媒体应该是独立的,这一点很重要。尽管计算机控制的视频录像机处理声音和运动图像信息,但在声音部分和视频部分之间具有时间依赖性。与之相对照,一个系统将 DAT(数字音频带)录音机中的信号和计算机中的文本合并来创建一个表示,这符合独立性的标准。另一个例子则是合并的文本和图形块,它们能以任意的空间顺序关联。

3. 计算机控制的系统

媒体独立性创建了一种媒体以任意形式联合起来显示的方式。基于此目的,计算机是理想的工具。也就是说,需要一个能够通过计算机控制处理媒体的系统,该系统能够由系统程序员或用户(在某些限制内)进行可选的编程。系统中各种媒体的简单记录或播放,以及像视频录像机之类,并不足以满足计算机控制的标准。

4. 集成

可以将计算机控制的独立媒体集成,形成一个综合系统,一起提供某种功能。为此,在它们之间创建时间、空间和内容的同步关系。这样,如果一个支持文本、电子表格和图形的字处理器在数据之间不允许可编程的引用,它就不满足集成的标准。当相关电子表格的内容变化时,如果应用能够自动地更新图形和文本项,那就获得了更高级的集成。

这种灵活的媒体处理方式,即使在许多贴着多媒体系统标签的产品中,也不是一件想当然的事情。当谈论到集成的多媒体系统时,这一特性才是重要的。简单地说,这种系统应允