

多媒体数据库与 内容检索

DMT

DUOMEITI SHUJUKU YU
NEIRONG JIANSUO

马修军 著



北京大学出版社
PEKING UNIVERSITY PRESS

多媒体数据库与内容检索

马修军 著



北京大学出版社
PEKING UNIVERSITY PRESS

图书在版编目(CIP)数据

多媒体数据库与内容检索/马修军著. —北京:北京大学出版社, 2007. 7

ISBN 978-7-301-09920-9

I. 多… II. 马… III. 计算机网络—情报检索—教材 IV. G354.4

中国版本图书馆 CIP 数据核字(2007)第 125634 号

书 名: 多媒体数据库与内容检索

著作责任者: 马修军 著

责任编辑: 王树通

标准书号: ISBN 978-7-301-09920-9/TP · 0823

出版发行: 北京大学出版社

地 址: 北京市海淀区成府路 205 号 100871

网 址: <http://www.pup.cn> 电子信箱: zpup@pup.pku.edu.cn

电 话: 邮购部 62752015 发行部 62750672 编辑部 62752038

出版部 62754962

印 刷 者: 世界知识印刷厂

经 销 者: 新华书店

787 毫米×1092 毫米 16 开本 21.25 印张 530 千字

2007 年 7 月第 1 版 2007 年 7 月第 1 次印刷

定 价: 32.00 元

未经许可,不得以任何方式复制或抄袭本书之部分或全部内容。

版权所有,侵权必究

举报电话: (010)62752024 电子信箱: fd@pup.pku.edu.cn

前 言

多媒体(multimedia)内容与高速网络、智能数码设备的融合已经成为现实,个性化的视频点播、互联网信息检索、移动商务、高级的协同工作(视频会议、网络聊天室等)、个性化的内容索引(图片和家庭视频),已经渗入到了我们的工作、学习和生活方式中。数码设备(智能手机、数码相机、个人数字助理(personal digital assistant,简称 PDA 等)的集成和普及,交互电视、互联网和移动通讯带来的数字化多媒体内容的爆炸,进一步推动了多媒体内容管理和检索技术的发展。

在多媒体领域,过去的研究主要集中在多媒体通信和多媒体制作与表现工具方面。但随着数字媒体信息获取和制作技术越来越普及,近年来,多媒体的研究也转到了高效存储管理和多媒体信息检索(information retrieval,简称 IR)领域。同样的转变发生在 30 年前,当结构化信息越来越多时,促进了数据库的发展。然而,由于多媒体信息特征与结构化信息的特征和需求不同,传统的数据库并不能有效地处理多媒体信息,因此多媒体信息管理需要有新的数据管理和信息检索技术。

多媒体信息内容管理和信息检索是一项综合性非常强的技术,它涉及多媒体信号处理、计算机视觉、语音识别、图像处理、模式识别、数据库、计算机网络、人机交互、认知科学等许多研究领域。然而,上述领域的学者只是把多媒体内容管理和信息检索作为一个应用方向,关注各自的技术细节研究,却忽略了多媒体内容管理和信息检索技术综合性要求的理论研究探索。

我们发现,目前多媒体方面的教材和著作大都针对某类技术或某种媒体类型,内容深浅不一,而且大多局限于多媒体技术、多媒体著作和多媒体通信方面;关于多媒体内容检索的文献大多侧重于研究特定的问题,多媒体内容管理和信息检索的教材很少。另外,多媒体数据库的内容大多在数据库的教材的部分章节进行简单介绍,没有涉及多媒体内容检索技术,只是从数据模型和查询语言方面简单论述。其他针对各种媒体类型的内容管理和信息检索的著作和教材,往往也是各成体系,例如针对文本的信息检索与 web^① 搜索引擎,图像检索、音频和视频检索。许多想从事这方面研究的学者和教师很难找到合适的教材,通常要花费很长时间从不同的领域查阅相关文献。因此,一本全面的、综合性的、覆盖当前多媒体内容管理和信息检索的教材是非常急需的。

本教材的目的是综合性地讲述当前多媒体信息检索与管理的最新技术和趋势,内容覆盖多媒体数据编码及其标准、针对文本的信息检索与 web 搜索引擎,图像检索、视频检索与结构化,语音识别、音频和音乐检索,多媒体数据库,多维特征相似性匹配技术,数字图书馆和多媒体信息安全等 12 章内容。

本书的特色是用一个统一理论框架为主线,系统地阐述各种媒体的内容检索技术,并通过这样一个框架均衡各种媒体检索技术的深度和广度,在内容检索的技术背景下,讲述多媒体数据库的关键技术,便于读者理解和掌握多媒体数据库的理论、设计需求和最新技术进展以及当前有影响的、有代表性的多媒体内容检索系统和商业多媒体数据库系统。本书内容覆盖了多媒体数据

① “web”是指万维网(world wide web)中的“web”。“web 搜索引擎”这种提法现在被广泛接受。

编码、基于内容多媒体检索、多媒体数据库等,内容全面,框架清晰;每章末都提供习题,便于读者检验学习效果。

多媒体内容管理和信息检索已经成为当前信息和智能科学技术的一个重要的研究方向。本书的适用对象还包括:

(1) 相关专业的高年级本科生和研究生。

(2) 多媒体内容管理和检索系统的开发人员。众多著名的 IT 公司(如微软、Google、百度、IBM 等)都在开发多媒体内容管理和检索的系统,本书可作为技术人员了解相关技术问题的参考书。

(3) 多媒体内容管理和信息检索研究人员。本书可帮助他们了解当前研究方向和发展趋势。

此外,本书还适用于任何想了解多媒体内容管理和信息检索相关技术问题和发展现状的其他读者。

在本书的编写过程中,作者得到了许多人的帮助,在此深表谢意。特别感谢北京大学信息科学技术学院选修“多媒体信息系统”课程的研究生,他们做了很多工作(包括文献调研),并对各种方法的见解和讨论以及文献材料整理提出了有益的建议。尤其感谢直接参加本书相关章节编写的同学:金星星、余晋、刘杰、孙怡舟、韩亮、李晨煜、陈薇、徐丹、帅猛、张静肖、张月祥;另外,陈冠华参加了书稿的校订工作。陈霄、胡子敬、邱宝军、范裕等同学也对本书的资料汇编付出了努力,一并表示感谢。

衷心感谢北京大学信息科学技术学院基础教育部甘学温教授的热心帮助,并感谢北京大学出版社的支持。

由于作者水平有限,时间紧迫,再加上多媒体内容管理和信息检索是当前众多技术的交叉领域,发展迅速,书中难免有疏漏之处,敬请读者批评指正,以便日后予以更正。

马修军

于北京大学

2005 年 9 月

目 录

第一章 绪论	(1)
§ 1.1 引言	(1)
§ 1.2 概念和术语	(2)
1.2.1 多媒体技术的有关概念	(2)
1.2.2 特征抽取,内容表示和索引	(3)
1.2.3 数据检索与信息检索	(4)
1.2.4 用户任务	(5)
§ 1.3 多媒体信息管理的技术需求	(6)
1.3.1 多媒体信息管理的需求	(6)
1.3.2 多媒体信息分类	(7)
1.3.3 多媒体信息管理功能需求	(8)
§ 1.4 相关技术	(10)
1.4.1 多媒体技术	(10)
1.4.2 数据库管理系统	(11)
1.4.3 信息检索技术	(12)
1.4.4 模式识别技术	(13)
§ 1.5 发展趋势和未来的研究问题	(14)
§ 1.6 本书组织	(18)
§ 1.7 参考文献	(19)
§ 1.8 习题	(19)
第二章 多媒体数据类型与编码	(21)
§ 2.1 引言	(21)
§ 2.2 文本	(21)
2.2.1 简单文本	(21)
2.2.2 结构化文本	(22)
2.2.3 文本压缩	(22)
§ 2.3 数字图像和图形	(25)
2.3.1 数字图像和图形表示	(25)
2.3.2 图像压缩原理	(25)
2.3.3 静态图像压缩标准——JPEG	(26)
2.3.4 其他图像标准	(28)
§ 2.4 声音和音频	(29)

2.4.1 声音的物理特性	(29)
2.4.2 声音的数字表示	(29)
2.4.3 电子乐器数字接口	(31)
2.4.4 音频压缩	(31)
2.4.5 其他音频压缩标准	(37)
§ 2.5 视频和动画	(37)
2.5.1 模拟视频	(38)
2.5.2 数字视频	(38)
2.5.3 视频压缩	(39)
2.5.4 其他视频标准	(45)
§ 2.6 多媒体表示标准	(45)
2.6.1 同步多媒体集成语言	(45)
2.6.2 MHEG 标准	(46)
§ 2.7 多媒体元数据标准	(46)
2.7.1 多媒体元数据概述	(47)
2.7.2 XML 技术	(48)
2.7.3 语义 web 与互操作	(49)
2.7.4 多媒体元数据标准化框架	(51)
§ 2.8 小结	(52)
§ 2.9 参考文献	(53)
§ 2.10 习题	(54)
第三章 文本处理与信息检索	(55)
§ 3.1 引言	(55)
§ 3.2 信息检索模型	(56)
3.2.1 信息检索模型分类	(56)
3.2.2 经典检索模型	(56)
3.2.3 扩展经典检索模型	(61)
§ 3.3 文本处理	(63)
3.3.1 文本预处理	(63)
3.3.2 文本特性	(67)
3.3.3 文本聚类	(69)
§ 3.4 文本索引	(71)
3.4.1 倒排文件表	(72)
3.4.2 后缀树和后缀数组	(75)
3.4.3 签名文件	(79)
§ 3.5 相关反馈和查询扩展	(81)
§ 3.6 检索评测	(83)
3.6.1 性能评估	(83)
3.6.2 测试文档集	(88)

§ 3.7 小结	(89)
§ 3.8 参考文献	(90)
§ 3.9 习题	(91)
第四章 Web 信息检索	(92)
§ 4.1 引言	(92)
§ 4.2 Web 信息检索的挑战	(93)
§ 4.3 Web 搜索引擎	(94)
4.3.1 搜索引擎的体系结构	(94)
4.3.2 爬虫设计	(97)
4.3.3 排序算法	(99)
4.3.4 交互界面设计	(101)
§ 4.4 Web 分类索引	(102)
4.4.1 Web 目录	(102)
4.4.2 分类索引与搜索引擎的合并使用	(105)
§ 4.5 元搜索	(105)
4.5.1 元搜索的起源	(106)
4.5.2 元搜索组织结构	(106)
4.5.3 元搜索的分类	(107)
4.5.4 元搜索主要的性能指标	(108)
4.5.5 元搜索的特点	(109)
§ 4.6 Web 挖掘	(110)
4.6.1 Web 挖掘简介	(110)
4.6.2 Web 内容挖掘	(111)
4.6.3 Web 结构挖掘	(112)
4.6.4 Web 日志挖掘	(112)
§ 4.7 小结	(113)
§ 4.8 参考文献	(114)
§ 4.9 习题	(114)
第五章 基于内容的音频检索	(115)
§ 5.1 引言	(115)
§ 5.2 音频主要属性和特征	(115)
5.2.1 时域特征	(116)
5.2.2 频域特征	(117)
5.2.3 声谱图	(120)
5.2.4 主观特征	(120)
§ 5.3 通用的音频内容检索过程	(121)
5.3.1 音频分段与分类	(121)
5.3.2 通用音频内容检索过程	(121)
§ 5.4 音频的分段和分类	(122)

5.4.1	音频例子的特征提取	(123)
5.4.2	逐步判断分类	(123)
5.4.3	特征向量分类	(124)
§ 5.5	语音识别与检索	(125)
5.5.1	语音识别简介	(125)
5.5.2	语音识别过程	(126)
5.5.3	语音识别系统框架	(127)
5.5.4	语音识别常用的技术	(128)
5.5.5	语音识别系统评估	(131)
5.5.6	说话者识别	(132)
5.5.7	小结	(133)
§ 5.6	音乐的索引与检索	(133)
5.6.1	音乐的存储类型	(133)
5.6.2	音乐的索引	(134)
5.6.3	基于内容的音乐检索	(135)
§ 5.7	小结	(137)
§ 5.8	参考文献	(138)
§ 5.9	习题	(139)
§ 5.10	附录: HMM	(139)
第六章	基于内容的图像检索	(143)
§ 6.1	引言	(143)
§ 6.2	CBIR 系统的框架	(143)
§ 6.3	基于颜色特征的图像检索	(144)
6.3.1	颜色空间模型	(144)
6.3.2	颜色直方图	(145)
6.3.3	颜色矩	(147)
6.3.4	颜色聚合向量	(147)
6.3.5	颜色相关图	(148)
§ 6.4	基于纹理特征的图像检索	(148)
6.4.1	Tamura 纹理特征	(148)
6.4.2	灰度直方图的矩	(150)
6.4.3	基于共生矩阵的纹理描述方法	(150)
6.4.4	分形模型	(151)
6.4.5	自回归纹理模型	(152)
6.4.6	基于小波变换的纹理描述	(153)
6.4.7	基于 Gabor 变换的纹理描述	(154)
§ 6.5	基于形状特征的图像检索	(155)
6.5.1	基于轮廓的形状描述方法	(155)
6.5.2	基于区域的形状描述方法	(157)

§ 6.6 基于空间关系的图像检索	(159)
6.6.1 图像分割方法	(159)
6.6.2 拓扑关系	(160)
6.6.3 方位关系	(160)
6.6.4 距离关系	(161)
6.6.5 属性关系图	(161)
6.6.6 基于空间关系特征检索的步骤	(162)
§ 6.7 不同特征的比较与综合特征检索	(162)
6.7.1 不同图像特征的特点总结	(162)
6.7.2 不同特征的比较	(163)
6.7.3 特征的综合使用	(163)
§ 6.8 图像检索过程中的相关反馈技术	(164)
6.8.1 查询向量相关反馈	(164)
6.8.2 特征权重相关反馈	(166)
6.8.3 其他图像相关反馈技术	(169)
§ 6.9 基于内容的图像检索系统实例	(170)
§ 6.10 小结	(171)
§ 6.11 参考文献	(173)
§ 6.12 习题	(174)
第七章 视频索引、检索与结构化	(175)
§ 7.1 引言	(175)
§ 7.2 基于镜头的视频索引和检索	(176)
7.2.1 视频结构化	(176)
7.2.2 基于镜头的视频索引和检索	(177)
§ 7.3 镜头检测和分段	(179)
7.3.1 镜头切换和运动	(179)
7.3.2 突变镜头检测	(180)
7.3.3 渐变镜头检测	(182)
§ 7.4 视频索引和检索	(184)
7.4.1 关键帧提取	(184)
7.4.2 运动特征提取与索引	(185)
7.4.3 视频对象特征提取和视频分割技术	(189)
7.4.4 视频检索的相似性度量	(190)
§ 7.5 视频表现和抽象	(191)
7.5.1 视频摘要	(191)
7.5.2 基于图像的视频摘要	(192)
7.5.3 基于内容的视频摘要	(194)
7.5.4 基于结构的视频摘要	(194)
§ 7.6 TRECVID 及 IBM 参赛视频检索系统	(195)

7.6.1 镜头边界检测	(196)
7.6.2 高层特征检测	(197)
7.6.3 检索	(199)
§ 7.7 小结	(201)
§ 7.8 参考文献	(202)
§ 7.9 习题	(203)
第八章 多媒体数据库	(204)
§ 8.1 引言	(204)
§ 8.2 多媒体数据库技术的发展	(205)
8.2.1 第一阶段	(205)
8.2.2 第二阶段	(205)
8.2.3 第三阶段	(207)
§ 8.3 数据模型	(208)
8.3.1 概述	(208)
8.3.2 数据模型的需求	(209)
8.3.3 商业数据库的扩展	(210)
8.3.4 数据模型实例	(210)
§ 8.4 多媒体查询语言 SQL/MM	(212)
8.4.1 引言	(212)
8.4.2 全文	(212)
8.4.3 空间数据	(213)
8.4.4 静态图像	(215)
§ 8.5 设计问题	(216)
8.5.1 体系结构	(216)
8.5.2 界面设计	(217)
8.5.3 信息空间表现	(218)
8.5.4 可视推理	(219)
8.5.5 可视化查询范型分类	(220)
8.5.6 多媒体信息系统交互与相关反馈技术	(221)
8.5.7 多媒体数据库系统实例	(223)
§ 8.6 MPEG-7 与多媒体数据库系统	(227)
§ 8.7 其他问题	(228)
§ 8.8 参考文献	(229)
§ 8.9 习题	(229)
第九章 多媒体数据库中高维特征的索引和检索技术	(231)
§ 9.1 引言	(231)
9.1.1 基于特征的相似性匹配的基本形式	(231)
9.1.2 基于特征的相似性匹配过程	(232)
§ 9.2 高维索引方法原理	(233)

9.2.1	高维空间中的一些效应	(233)
9.2.2	基本概念和定义	(235)
9.2.3	高维索引结构	(237)
9.2.4	高维索引管理与维护	(238)
9.2.5	区域	(239)
§ 9.3	基本算法	(239)
9.3.1	插入、删除和更新	(239)
9.3.2	精确匹配查询	(241)
9.3.3	范围查询	(241)
9.3.4	最近邻居查询和 k 近邻查询	(242)
§ 9.4	代价模型	(246)
§ 9.5	高维索引方法	(247)
9.5.1	数据驱动的索引结构	(248)
9.5.2	空间驱动的索引结构	(252)
9.5.3	降维的方法	(255)
9.5.4	高维索引方法比较	(258)
§ 9.6	小结	(259)
§ 9.7	参考文献	(260)
§ 9.8	习题	(262)
§ 9.9	附录: GiST	(263)
第十章	多媒体通信与分布式多媒体数据库系统	(265)
§ 10.1	引言	(265)
§ 10.2	多媒体通信网络与传输协议	(265)
10.2.1	多媒体通信网	(265)
10.2.2	异步传输模式	(266)
10.2.3	多媒体传输协议	(268)
§ 10.3	多媒体服务质量管理	(272)
10.3.1	QoS 定义	(272)
10.3.2	通用的 QoS 框架	(273)
10.3.3	QoS 规范	(273)
10.3.4	QoS 机制	(274)
10.3.5	QoS 体系结构	(275)
10.3.6	QoS 处理实例	(276)
§ 10.4	多媒体同步	(276)
10.4.1	多媒体同步规范	(276)
10.4.2	多媒体不同步原因分析	(277)
10.4.3	保证多媒体同步机制	(278)
10.4.4	基于 QoS 框架的解决方案	(279)
§ 10.5	分布式多媒体数据库系统	(280)

§ 10.6	小结	(284)
§ 10.7	参考文献	(284)
§ 10.8	习题	(285)
第十一章	数字图书馆	(286)
§ 11.1	引言	(286)
11.1.1	数字图书馆的背景	(286)
11.1.2	数字图书馆的定义	(287)
11.1.3	数字图书馆的特征	(288)
§ 11.2	数字图书馆中多媒体资源存储	(289)
11.2.1	数字对象	(289)
11.2.2	数字对象逻辑模型	(290)
11.2.3	数字对象命名规范	(291)
11.2.4	元数据	(291)
§ 11.3	数字图书馆中多媒体信息检索	(292)
11.3.1	检索语言	(293)
11.3.2	基于关键词的检索	(294)
11.3.3	基于内容的图像检索	(295)
11.3.4	基于内容的视频检索	(296)
11.3.5	基于内容的音频检索	(297)
§ 11.4	标准	(297)
11.4.1	内容编码的标准规范	(297)
11.4.2	数字对象逻辑模型的标准规范	(297)
11.4.3	资源格式的标准规范	(298)
11.4.4	目录模式的标准规范	(298)
11.4.5	检索条件的标准规范	(299)
§ 11.5	原型、项目与界面	(299)
§ 11.6	小结	(302)
§ 11.7	参考文献	(303)
§ 11.8	习题	(303)
第十二章	多媒体信息安全	(304)
§ 12.1	引言	(304)
12.1.1	信息安全的基本概念	(304)
12.1.2	多媒体信息安全的特殊性	(306)
§ 12.2	常用信息安全技术	(306)
12.2.1	密钥加密	(306)
12.2.2	公开密钥加密	(308)
12.2.3	单向散列算法	(309)
12.2.4	密钥管理	(309)
12.2.5	数字签名、数字证书和 SSL	(310)

§ 12.3	信息伪装、数字水印和多媒体鉴定	(312)
12.3.1	信息伪装	(312)
12.3.2	数字水印	(314)
12.3.3	多媒体鉴定方案	(315)
§ 12.4	数字版权管理	(317)
12.4.1	基本概念	(317)
12.4.2	DRM 系统整体功能结构	(318)
12.4.3	版权模型	(319)
12.4.4	用户认证	(321)
12.4.5	内容管理	(321)
§ 12.5	趋势和研究问题	(323)
§ 12.6	参考文献	(323)
§ 12.7	习题	(325)

第一章 绪 论

§ 1.1 引 言

随着人类社会由工业社会发展到信息社会,数字化信息正在以惊人的速度迅速增长,这就是所谓的“信息爆炸”。根据美国加利福尼亚大学伯克利分校信息管理和系统学院行业分析家的估计,2001~2003年间产生的数据比记录历史的全部数据的总和还要多。他们经过研究发现,全球每年产生的不重复信息量在1~20 PB^①之间,也就是说,全球每人(包括小孩)年均产生约250 MB^②的数据。在这些日益膨胀的信息中,多媒体信息(包括文本文件、扫描图像、视频剪辑、音频等)是信息爆炸中的重要组成部分,而且所占比重越来越大。目前,诸如图像、音频和视频等多媒体信息在web中占据15%,且该比重还在飞速增长。种种证据表明,多媒体信息越来越丰富。

多媒体领域的研究过去主要集中在多媒体通信和多媒体制作与表现工具方面。但随着数字媒体信息获取和制作技术越来越普及,近年来,多媒体的研究也转到高效存储管理和多媒体信息检索方面。同样的转变发生在30年前,越来越多的结构化信息促进了数据库的发展。然而,由于多媒体信息特征与结构化信息的特征和需求不同,传统的数据库并不能有效地处理多媒体信息,因此,多媒体信息的管理需要有新的数据管理和信息检索技术。

本书的主要目的是介绍与多媒体数据管理和内容检索有关的技术和研究问题以及最新的研究趋势。为了解决多媒体数据管理的问题,人们首先想到的方法就是利用关系数据库管理系统(relational database management system,简称RDBMS)来管理多媒体数据。对于图像来说,实际上早在20世纪70年代人们就开展了对图像数据库的研究,其解决方法通常是利用人工输入图像的各种属性,建立图像的元数据库来支持查询。在70年代末80年代初,这些系统经历了短暂的兴盛之后就衰落了。90年代以来,随着多媒体技术的发展,可获取的图像和其他多媒体数据越来越多,数据库容量不断增大,这种用人工输入属性和注释的方法就暴露出了它的缺点:第一个缺点是人工注释需要大量的人力,尤其是对于大型的多媒体信息库,如web网络资源、数字图书馆等。在这样的信息环境中,每天都有大量的新资料出现,需要及时把这些资料归档。只用人工注释,没有计算机的自动或辅助处理,资料的更新周期就不能满足用户的需要。第二个缺点是人工注释难以解决蕴藏在多媒体数据中丰富的内容以及对内容感知描述的主观性的问题。人们常说,一幅图胜过千言万语,而音频、视频(AV)等媒体包含了更丰富的内容,这些内容很难用文字来描述清楚。第三个缺点就是对于实时广播流媒体,手工处理是根本不可行的,必须用计算机进行实时的内容分析。由此,基于内容的多媒体信息检索研究应运而生。

① 1 PB=10¹⁵ B。“B”是字节(byte)的单位符号。

② 1 MB=10⁶ B。

本章将首先介绍多媒体数据管理和内容检索相关的一些基本概念和术语;然后对多媒体信息系统的需求进行分析;接着对其他相关的学科和技术进行总结回顾;最后总结、展望多媒体数据管理和内容检索的当前研究焦点和发展趋势。

§ 1.2 概念和术语

1.2.1 多媒体技术的有关概念

1. 媒体类型和多媒体

媒体(media)又称媒介,是承载、传输和表现信息的手段。按照国际电信联盟(International Telecommunication Union,简称ITU)的定义,媒体有以下五种:感觉媒体、表示媒体、显示媒体、存储媒体和传输媒体。感觉媒体指的是用户接触信息的感受形式,如视觉、听觉和触觉等;表示媒体指的是信息的表示和表现形式,如图形、声音和视频等;显示媒体是表现和获取信息的物理设备,如显示器、打印机、扬声器、键盘和摄像机等;存储媒体是存储数据的物理设备,如磁盘、光盘、磁带等;传输媒体是传输数据的物理设备,如电缆、光缆等。

对多媒体信息管理技术来说,我们关心的是表示媒体。表示媒体又可以根据表示值和表示空间进行分类。表示空间是信息输出表现的媒介,例如纸和计算机屏幕可以作为图形图像的可视表示空间,立体声和四轨录音/放音是声音的表示空间。表示值决定不同媒体的信息表示,例如文本可以用可视的方式表示为一串字符组成的句子,也可以用语音媒体表示。

根据表示值的不同,媒体可以分为离散媒体和连续媒体。

(1) 离散媒体。

离散媒体中的信息是由一组不随时间变化而变化的独立元素组成的。例如文本、图像、图形等,都属于离散媒体。

(2) 连续媒体。

连续媒体的信息表示与时间有关,随时间的变化而变化。时间或时序关系是信息的一部分。如果时序发生了变化,或者媒体中项的次序发生了变化,那么信息的含义也会发生变化。音频、视频和动画都属于连续媒体。

从字面上来看,“多媒体”(multimedia)即为“多”(multiple)和“媒体”(media)的复合词。因此,广义上讲,多媒体就是多种信息表示媒体的组合。但严格地说,必须既有连续媒体,又有离散媒体,才能称为多媒体。

2. 多媒体技术

多媒体技术是由计算机平台、通信网络、人机接口以及相应媒体数据组成的系统技术,注重改善信息的表示方式、技术的集成性和实时交互性,在系统级别层次上面向用户交互,不仅提高了系统的性能,还促进了用户对信息的获取和控制。多媒体技术具有集成性和交互性。

多媒体技术促进了通信、娱乐和计算机的融合。因此,喜欢文字游戏的人参照著名的爱因斯坦(Einstein)能量公式 $E=mc^2$ (E 为能量, m 为质量, c 为光速),将现在的信息环境表示成

$$E=mc^2=m \cdot c \cdot c,$$

其中 E 表示信息环境, m 表示多媒体, c 依次表示计算机和通信。

由此可见多媒体对于信息社会之重要性。随着技术的进步和市场前景的明朗,多媒体技术

的研究与应用已在世界各地如火如荼地展开,出现了一些有巨大市场影响力的战略型产品。

从多媒体研究的发展来看,目前只是走过了多媒体概念认识的“启蒙”阶段。这一阶段最典型的应用是“多媒体演示系统”,尽管许多人称其为多媒体信息管理系统,但离真正的多媒体信息管理还有很大的距离。这种演示系统对用户的概念教育是直观有效的,因此,这一阶段的工作大多也是按照类似概念演示的思路进行的,例如多媒体硬件接口、用户界面多媒体化、多媒体编辑创作、多媒体通信等。经过这个启蒙阶段,规范化的多媒体研究体系和重要的研究领域已经初步形成,需要研究的重点问题也已出露端倪。多媒体技术的研究内容主要有以下几方面:

- (1) 多媒体信息特性与建模;
- (2) 多媒体信息的组织与管理;
- (3) 多媒体信息表现与交互;
- (4) 实时性;
- (5) 多媒体通信与分布式处理;
- (6) 虚拟现实和多媒体协同工作环境。

1.2.2 特征抽取,内容表示和索引

多媒体内容检索技术最重要的问题是多媒体数据的特征提取和内容表示。这对大规模的多媒体数据管理是非常有价值的。我们需要区分几个概念:

- (1) 多媒体数据(data):即多媒体数据资料,包括图形、静止图像、视频、影片、音乐、语音、声音、文本和其他相关的音频、视频媒体等。
- (2) 特征(feature):指多媒体数据的某种特性。特征本身不能比较,而要用有意义的特征表示(描述子)和它的实例(描述值)进行比较。例如图像的颜色、语音的声调、音频的旋律等。
- (3) 内容描述子(descriptor):是特征的表示。它定义特征表示的句法和语义,可以赋予描述值。一个特征可能有多个描述子。例如颜色特征可能的描述子有颜色直方图、频率分量的平均值、运动的场描述、标题文本等。
- (4) 内容描述值(descriptor value):是描述子的实例,例如具体的颜色直方图。

多媒体特征抽取或内容处理是一个自动化或半自动化的从多媒体数据中抽取内容的过程,即先对原始媒体进行特征提取,提取内容,然后用标准形式对它们进行描述,以支持各种基于内容的查询检索。这个过程可分为三大部分:特征提取、内容描述和内容检索,也可将其看成是内容处理的三个步骤,如图 1-1 所示。

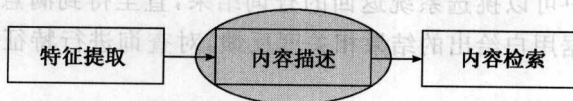


图 1-1 多媒体内容处理流程

多媒体数据中蕴涵的丰富内容决定了多媒体特征的多维性。从空间特征上看,有对象的纹理和形状特征以及对象的空间关系等;从时间特征上看,有对象随时间变化的轨迹,如音乐片段的持续时间。特征表示是多层的:

- (1) 客观特征:反映多媒体数据本身具备的特性,如对象的颜色、形状、纹理、音频频率