

矢量量化在多媒体信号处理中的应用

Applications of Vector Quantization in Multimedia Signal Processing

陆哲明 郑为民 孙圣和 著

 科学出版社

矢量量化在多媒体信号 处理中的应用

陆哲明 郑为民 孙圣和 著

科学出版社

北京

内 容 简 介

本书是矢量量化理论及应用的专著。书中首先介绍矢量量化的基本原理和应用现状,接着介绍各种矢量量化器的结构和基本思想,然后介绍基本矢量量化器的三大关键技术——码书设计、码字搜索和码字索引分配的各种有效算法。最后着重介绍矢量量化技术在多媒体信号处理领域中的应用,包括图像图形处理、语音处理、数字水印和基于内容的多媒体检索领域的应用。

本书取材广泛,内容全面、新颖,充分反映了近几年来矢量量化技术在各领域应用的最新研究动态,并总结了作者近10年的研究成果。

本书可供从事计算机通信,数字图像、音频、三维模型和视频信号处理,模式识别,数字产品版权保护,优化设计,多媒体检索等领域的科技人员与教学人员阅读和参考,并可以作为上述领域研究生课的教材。

图书在版编目(CIP)数据

矢量量化在多媒体信号处理中的应用/陆哲明,郑为民,孙圣和著. —北京:科学出版社,2011

ISBN 978-7-03-031359-1

I. ①矢… II. ①陆…②郑…③孙… III. ①矢量-量化-应用-多媒体-信号处理 IV. ①TP37

中国版本图书馆CIP数据核字(2011)第104821号

责任编辑:鞠丽娜 / 责任校对:马英菊
责任印制:吕春珉 / 封面设计:三函设计

科学出版社出版

北京东黄城根北街16号
邮政编码:100717

<http://www.sciencep.com>

双青印刷厂印刷

科学出版社发行 各地新华书店经销

*

2011年7月第一版 开本:B5(720×1000)

2011年7月第一次印刷 印张:20

印数:1—2 500 字数:400 000

定价:50.00元

(如有印装质量问题,我社负责调换〈双青〉)

销售部电话 010-62134988 编辑部电话 010-62138978-8002

版权所有,侵权必究

举报电话:010-64030229 010-64034315; 13501151303

前 言

随着计算机和大规模集成电路的飞速发展,数字信号分析和处理技术得到很大发展,并已经广泛用于通信、雷达和自动化等领域。数字信号的明显优点是便于传输、存储、交换、加密和处理等。在数字通信中,码速率高不仅影响传输效率,而且增加了存储和处理的负担。因此,在数字通信中通常对数字信号进行信源编码。数据压缩是信源编码的目的和手段。数据压缩的主要目的是为了降低码速率或减少存储空间。量化是有损数据压缩中常用的技术,基本上可以分为三种,即标量量化、矢量量化和序列量化。最基本的标量量化每次只量化一个采样,并对所有采样都采用相同特性的量化器进行量化,而且每个采样的量化都和其他采样无关。矢量量化和序列量化则利用相邻采样之间的相关性。矢量量化(vector quantization, VQ)的理论基础是香农的速率失真理论,其基本原理是用码书中与输入矢量最匹配的码字的索引代替输入矢量进行传输和存储,而解码时只需简单的查表操作。矢量量化作为一种有效的有损压缩技术,其突出优点是压缩比大以及解码算法简单。矢量量化技术涉及多学科领域的理论和技术,如信息论、编码理论、通信原理、保密技术、信号处理、优化理论、模糊集合论、矩阵分析、神经网络、小波变换、视觉模型、拓扑学、随机概率理论、预测技术、模式识别等。矢量量化技术于20世纪80年代开始得到学者们的普遍关注,刚开始应用在语音编码和图像压缩中。VQ技术在20世纪90年代后期逐渐淡出数据压缩领域的原因在于:①出现了基于DCT和DWT的JPEG、MPEG、JPEG2000、H. 26X等图像、视频压缩标准;②VQ的码书标准化问题和VQ编码复杂度问题。但矢量量化压缩比大、解码简单的特点使得其在语音编码、纹理压缩、视频编码等标准中仍占有一席之地。到了21世纪,仍有不少学者坚守基于矢量量化的压缩技术研究,主要从如下几个角度作努力:①通过提高码字搜索速度来降低矢量量化编码复杂度,解决矢量量化的实时性问题,以期使矢量量化的研究重新热起来或余温不减。②通过构造低复杂度高质量的矢量量化器,以期使矢量量化编码性能高于JPEG和JPEG2000,使学者们重新审视矢量量化的压缩编码潜力。③将矢量量化与其他技术相结合。截至目前,矢量量化在语音编码、图像编码、视频编码、三维模型压缩等领域仍有不少新方法和新论文出现。

矢量量化作为一种聚类工具也在模式识别和机器学习领域得到了应用。第一个比较成功的应用是在语音识别领域。第二个比较成功的应用是多媒体检索领域。多媒体检索可以归为模式识别的问题,其实质是在给定模式库中寻找与给定

模式最相似的前几个模式。基于内容的多媒体检索研究于 20 世纪 90 年代初兴起。所谓基于内容的信息检索是对文本、图像、音频、视频等媒体对象进行内容语义的分析和特征的提取,并基于这些特征进行相似性匹配的信息检索技术。从 2005 年起,本书作者一直继续将矢量化技术应用到图像检索和视频检索领域,主要是利用快速码字搜索来提高检索速度以及采用基于码字索引的特征来提高检索精度,为压缩域视觉信息检索研究领域提供一种新思路。第三个方面的应用是人脸检测。2005 年,本书作者成功地将矢量化用到了人脸检测领域,该应用得到其他学者的跟进研究,为人脸肤色的聚类判别提供新思路。第四方面的应用是火灾识别:2010 年,本书作者成功地将矢量化技术用到了基于视频的火灾识别中,并取得了良好效果。

数字水印技术是将一些标识信息(即数字水印)直接嵌入数字媒体中而不影响原媒体的使用价值,也不容易被探知和再次修改,但它可以被生产方识别和辨认。通过这些隐藏在载体中的信息,可以达到确认内容创建者、购买者、传送隐秘信息或者判断载体是否被篡改等目的。数字水印是信息隐藏技术的一个重要研究方向。自 2000 年以来,在本书作者等人的努力下,矢量化技术开始扩展,并应用到了数字水印领域,从而引导多个学者继续深入研究矢量化数字水印算法。这不仅拓展了矢量化应用领域,也开辟了新的数字水印嵌入域。随后,包括本书作者在内的许多学者逐步把矢量化用于多功能水印、可逆水印、音频水印、视频水印、三维模型水印等分支领域中。除了上述领域,矢量化技术还成功地应用到其他一些多媒体处理领域,包括去噪、彩色化等领域。作者坚信,随着矢量化技术和多媒体信号处理领域的进一步发展,矢量化技术必定在更多的领域加以应用。

本书旨在为从事多媒体信号处理的研究人员介绍矢量化关键技术研究现状和矢量化技术的应用现状。目的是推出一本较新、较全的矢量化著作,使研究人员能够对矢量化技术及其应用有全面了解,并推动国内对矢量化技术的深入研究。本书共七章。前两章从标量化技术的介绍入手,引出矢量化的概念、原理、关键技术和特点,介绍 30 年来学者们提出的各种矢量化器的基本结构和基本原理。第三章重点介绍矢量化三大关键技术,首先介绍各种码书设计算法,包括基于神经网络的矢量化码书设计算法、基于全局寻优技术的码书设计算法和基于模糊理论的矢量化码书设计算法;接着介绍各种快速码字搜索算法,包括基于不等式判据的、基于变换域的、基于金字塔结构的等;最后介绍各种码字索引分配算法。本书从第四章到第七章详细介绍矢量化在各个多媒体信号处理领域的应用,其中第四章介绍矢量化在图像图形处理中的应用,包括图像编码、彩色化、人脸检测和三维模型压缩;第五章介绍矢量化技术在语音编码和语音识别中的应用;第六章介绍了矢量化技术在数字水印中的应用,包括图像水印、音频水印、视频水印和三维模型水印等领域;第七章介绍矢量化在基于内容的多媒体信

息检索中的应用,包括图像检索、视频检索和三维模型检索等领域。

本书可供从事计算机通信,数字图像、音频、三维模型和视频信号处理,模式识别,数字产品版权保护,优化设计,多媒体检索等领域的科技人员与教学人员阅读和参考,并可以作为上述领域研究生课的教材。本书的第一章到第六章由陆哲明教授执笔,第七章由郑为民教授执笔,全书由孙圣和教授审定。在本书的撰写过程中得到了浙江大学航天电子工程研究所多位教师、博士生和硕士生的协助,在此表示衷心地感谢。作者在本书中述及的研究工作得到国家自然科学基金(项目号:60272074)和全国优秀博士论文作者专项基金(项目号:2003027)资助。作者之一郑为民教授谨以此书纪念其祖父郑一萍百岁冥诞。

限于作者水平,书中难免有错误与不妥之处,恳请读者批评指正。

作 者

2010年10月

于浙江大学航空航天学院航天电子工程研究所

目 录

前言

第一章 绪论	1
1.1 通信系统和多媒体信号处理	1
1.1.1 消息、信号、信息和数据	1
1.1.2 通信系统	2
1.1.3 多媒体信号处理	4
1.2 标量量化	7
1.2.1 标量量化的基本原理	7
1.2.2 均匀量化	9
1.2.3 最佳标量量化	11
1.2.4 最佳标量量化器设计算法	13
1.3 预测量化	15
1.3.1 差值量化与预测量化	16
1.3.2 差分脉码调制	17
1.3.3 增量调制	20
1.4 矢量量化	23
1.4.1 矢量量化概念的引入	23
1.4.2 矢量量化的理论基础	24
1.4.3 矢量量化的定义和基本原理	25
1.4.4 矢量量化器的结构	27
1.4.5 矢量量化器的性能指标	31
1.4.6 矢量量化关键技术	35
1.5 矢量量化与标量量化的比较	36
1.6 矢量量化的应用现状	38
第二章 矢量量化器	40
2.1 引言	40
2.2 穷尽搜索矢量量化器	40
2.3 约束矢量量化器	44
2.3.1 树型矢量量化器	44
2.3.2 分类矢量量化器	46

2.3.3	变换域矢量量化器	48
2.3.4	乘积码矢量量化器	49
2.3.5	多级矢量量化器	50
2.3.6	分层矢量量化器	53
2.3.7	格型矢量量化器和格型码书矢量量化器	55
2.4	预测矢量量化器	57
2.4.1	有记忆矢量量化器的分类	57
2.4.2	预测矢量量化器	58
2.5	有限状态矢量量化器	61
2.5.1	反馈矢量量化器	61
2.5.2	有限状态矢量量化器	64
2.6	自适应矢量量化器	67
2.6.1	均值自适应矢量量化器	67
2.6.2	增益自适应矢量量化器	70
2.6.3	开关码书自适应矢量量化器	72
2.6.4	地址矢量量化器	74
2.7	变速率矢量量化器	76
2.7.1	变速率编码	76
2.7.2	变维矢量量化器	77
2.7.3	其他变速率矢量量化器	79
第三章 矢量量化关键技术		81
3.1	矢量量化码书设计	81
3.1.1	概述	81
3.1.2	传统码书设计算法	82
3.1.3	基于神经网络的码书设计算法	85
3.1.4	随机松弛码书设计算法	87
3.1.5	遗传码书设计算法	88
3.1.6	禁止搜索码书设计算法	91
3.1.7	基于模糊集合理论的码书设计算法	94
3.2	矢量量化码字搜索	96
3.2.1	概述	96
3.2.2	部分失真搜索算法	97
3.2.3	基于绝对误差不等式的快速码字搜索算法	98
3.2.4	基于三角不等式的快速码字搜索算法	100
3.2.5	基于均值不等式的最近邻搜索算法	102

3.2.6	基于变换域的快速码字搜索算法	108
3.2.7	基于金字塔结构的快速码字搜索算法	111
3.3	矢量量化码字索引分配	114
3.3.1	概述	114
3.3.2	传统码字索引分配算法	115
3.3.3	禁止搜索码字索引分配算法	119
3.3.4	基于能量分配的码字索引传输	120
第四章	矢量量化在图像图形处理中的应用	123
4.1	基于矢量量化的图像编码	123
4.1.1	图像编码技术概述	123
4.1.2	快速相关矢量量化图像编码算法	124
4.1.3	基于边缘匹配矢量量化的图像编码	129
4.1.4	基于预测矢量量化的图像编码算法	139
4.2	基于矢量量化的彩色化	143
4.2.1	图像彩色化的研究意义	143
4.2.2	图像彩色化的研究现状	144
4.2.3	基于矢量量化的彩色化	146
4.3	基于矢量量化的人脸检测	149
4.3.1	概述	149
4.3.2	肤色提取	150
4.3.3	人脸检测方案	151
4.3.4	实验结果	152
4.4	基于矢量量化的三维模型压缩	154
4.4.1	概述	154
4.4.2	三维模型压缩研究现状	154
4.4.3	基于扩展码书矢量量化的三维模型压缩算法	155
4.4.4	基于多环预测矢量量化的三维模型压缩算法	161
4.4.5	小结	163
第五章	矢量量化在语音信号处理中的应用	165
5.1	语音信号分析与处理简介	165
5.1.1	语音信号	165
5.1.2	语音信号的时域分析	167
5.1.3	语音信号的频域分析	169
5.1.4	语音信号的同态分析	170
5.1.5	语音信号的线性预测分析	172

5.2	矢量量化在语音编码中的应用	175
5.2.1	概述	175
5.2.2	语音信号的波形编码	176
5.2.3	语音信号的参数编码	180
5.2.4	基于矢量量化的语音/音频压缩标准	187
5.3	矢量量化在语音识别中的应用	191
5.3.1	概述	191
5.3.2	基于动态时间规整的语音识别	194
5.3.3	基于矢量量化的语音识别	197
5.4	矢量量化在说话人识别中的应用	199
5.4.1	概述	199
5.4.2	说话人识别系统的结构和原理	199
5.4.3	基于动态时间规整的说话人识别	201
5.4.4	基于矢量量化的说话人识别	202
第六章	矢量量化在数字水印中的应用	205
6.1	引言	205
6.2	基于矢量量化的数字图像水印算法	207
6.2.1	基于基本 VQ 的鲁棒数字图像水印算法	208
6.2.2	基于预测或分类 VQ 的鲁棒数字图像水印算法	215
6.2.3	基于索引受限矢量量化的脆弱数字图像水印算法	220
6.2.4	基于约束 VQ 的多功能图像水印算法	222
6.3	基于矢量量化的数字视频水印算法	225
6.3.1	概述	225
6.3.2	基于空域矢量量化的多功能视频水印算法	227
6.3.3	基于变换域矢量量化的数字视频水印算法	230
6.4	基于矢量量化的数字音频水印算法	234
6.4.1	概述	234
6.4.2	基于矢量量化技术的鲁棒音频数字水印算法	235
6.4.3	基于矢量量化技术的多功能音频数字水印算法	240
6.5	基于矢量量化的三维模型可逆水印算法	243
6.5.1	概述	243
6.5.2	基于预测矢量量化的三维模型可逆水印算法	245
第七章	矢量量化在多媒体检索中的应用	248
7.1	多媒体检索概述	248
7.1.1	背景与概念	248

7.1.2 基于内容的多媒体检索研究现状	249
7.1.3 多媒体检索的主要功能模块	251
7.1.4 性能评价	254
7.2 矢量量化在视觉信息检索中的应用现状	255
7.2.1 特征提取	255
7.2.2 离线的索引或特征空间聚类	256
7.2.3 快速匹配	256
7.2.4 语义检索	257
7.2.5 存在的问题	257
7.3 基于矢量量化的图像特征提取	258
7.3.1 基于约束 VQ 的特征提取	258
7.3.2 基于距离加权边沿预测矢量量化的特征提取	260
7.3.3 仿真实验	261
7.4 基于矢量量化的快速图像检索算法	263
7.4.1 基于等均值 K 最近邻的快速检索	263
7.4.2 等均值等方差 K 最近邻快速渐进图像检索方法	268
7.4.3 基于分类矢量量化的快速匹配算法	272
7.5 基于矢量量化的实时图像语义检索	276
7.5.1 基于语义的图像检索研究现状	276
7.5.2 基于矢量量化的图像语义标注与检索	277
7.6 基于矢量量化和全局运动的视频检索	281
7.6.1 概述	281
7.6.2 基于矢量量化的图像全局运动矢量的提取	282
7.6.3 全局运动检索方案	283
7.6.4 仿真实验	284
7.7 基于矢量量化的三维模型检索	285
7.7.1 概述	285
7.7.2 基于矢量量化索引直方图的三维模型检索	286
7.7.3 仿真实验	288
参考文献	290

第一章 绪 论

1.1 通信系统和多媒体信号处理

矢量量化最早是作为一种信源编码技术提出来的,主要用于信号压缩。随着应用的扩展,矢量量化开始应用于其他信号处理分支。因此,这里有必要从通信系统和多媒体信号处理角度来引入矢量量化技术。本节首先介绍消息、信号、信息和数据的基本概念,然后介绍通信系统模型,最后简要介绍与本书有关的多媒体信号处理的各个研究领域。

1.1.1 消息、信号、信息和数据

在日常生活和科学研究中,经常碰到“消息”、“信号”、“信息”和“数据”这四个近义的相关词汇,这里有必要简要介绍它们的区别和联系。人们常把外界的各种报道称为消息(message),它反映知识状态的改变。我国古代就把客观世界的变化,把它们的发生、发展和结局,把它们 的枯荣、聚散、沉浮、升降、兴衰、动静、得失等变化中的事实称为“消息”。只是到了近代才逐渐成为一种固定新闻载体,所以“消息”又称新闻。

信号(signal)是用来传递消息或命令的光、电波、声音、动作等的统称。古代中国利用烽火传送边疆警报,希腊人借助火炬位置表示字母符号,这就是最原始的光信号通信。利用击鼓鸣金来报送时刻或传达命令,这就是声信号通信。19世纪以来,人们开始采用电信号传送消息,如1837年莫尔斯发明了电报,用点、划、空的适当组合表示数字和字母,1876年贝尔发明了电话,直接将声信号转变为电信号沿导线传输。19世纪末,赫兹等人致力于用电磁波传送无线电信号,从而使电信号的通信方式得到广泛应用和迅速发展。如今,人们可方便地利用电话、寻呼机、移动电话以至卫星和计算机互联网进行语音、图像和数据等各种信号的传输。由此可见,信号是运载消息的工具,是消息的载体,人们必须通过对光、声、电信号进行接收,才知道对方要表达的消息。消息和信号是内容和形式的关系。信号按数学关系、取值特征、能量功率、处理分析、所具有的时间函数特性、取值是否为实数等,可分为确定性信号和非确定性信号、连续信号和离散信号^[1,2]、能量和功率信号、时域和频域信号、时限和频限信号、实信号和复信号等。

信息(information)是信息论的一个术语,常把消息中有意义的内容称为信息。

1948年,美国数学家、信息论创始人香农在题为“通信的数学理论”的论文中指出:“信息是用来消除随机不定性的东西”。这说明,信息是使概率分布发生变动的东西,信息是确定性的增加,即肯定性的确认。信息就是对消息接收方来说不确定的和有意义的那部分消息内容。显然,信息还是通过信号来传输,信息与信号是内容与形式的关系。

在文献中,还常常会碰到“数据”(data)这个词^[1]。前面提到的信号是指大小随时间或空间变化的函数^[1],其中时间和空间变量的取值可以是连续的或离散的,相应的函数值也可以是连续的和离散的,且函数值可以是标量或矢量。数据经常作为信号的同义词,但通常它是指一个数列或矢量序列,常常认为数据是离散时间信号。但是,近年来,数据这个词在大多数文献中常常与数字信号联系在一起,即离散时间离散幅度的信号。

1.1.2 通信系统

通信系统是实现信息传递所需的一切技术设备和传输媒质的总和,一般模型如图 1.1 所示。图中,信源把待传输的消息转换成原始电信号,常称为基带信号。基带信号可分为数字基带信号和模拟基带信号,相应地,信源也分为数字信源和模拟信源。发送设备的基本功能是将信源产生的基带信号变换成适合在信道中传输的信号,调制是最常见的变换方式;对传输数字信号来说,发送设备又常常包含信源编码和信道编码等。信道是指信号传输的通道,可以是有线的,也可以是无线的,甚至还可以包含某些设备。图中的噪声源,是信道中的所有噪声以及分散在通信系统中其他各处噪声的集合。在接收端,接收设备的功能与发送设备相反,通过解调、译码、解码等操作从带有干扰的接收信号中恢复出相应的原始电信号来。信宿的作用是将复原的电信号转换成相应的消息。

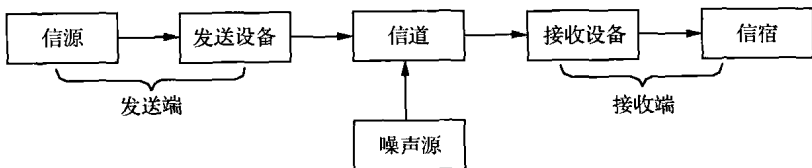


图 1.1 通信系统的一般模型

人们把信道中传输模拟信号的系统称为模拟通信系统。模拟通信系统的组成可由一般通信系统模型略加改变而成,如图 1.2 所示。对于模拟通信系统,它主要包含两种重要变换。一是把连续消息变换成电信号(由信源完成)和把电信号恢复成连续消息(由信宿完成)。由于从信源输出的电信号(基带信号)具有频率较低的频谱分量,一般不能直接作为传输信号而送到信道中去。因此,模拟通信系统里常

有第二种变换,即将基带信号转换成其适合信道传输的信号,这一变换由调制器完成;在接收端同样需经相反变换,它由解调器完成。必须指出,从消息的发送到消息的恢复,事实上并非仅有以上两种变换,通常在一个通信系统里可能还有滤波、放大、天线辐射与接收、控制等过程。

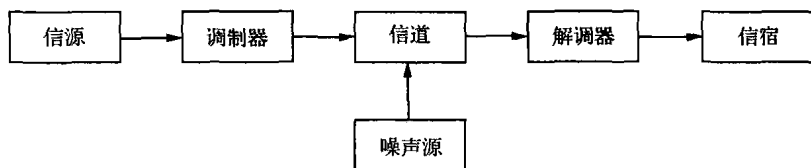


图 1.2 模拟通信系统模型

信道中传输数字信号的系统,称为数字通信系统^[3],如图 1.3 所示。在模拟通信中强调已调参量与代表消息的基带信号之间的比例特性;而在数字通信中,则强调已调参量与代表消息的数字信号之间的一一对应关系。在数字通信系统中,信道噪声或干扰所造成的差错,原则上是可以通过差错控制编码来控制的。当需要实现保密通信时,可对数字基带信号进行加密,此时在接收端就必须进行解密。数字通信中还必须收发“同步”,包括“码元同步”和“帧同步”。需要说明的是,调制器/解调器、加密器/解密器和编码器/译码器等环节,在通信系统中是否全部采用,这要取决于具体设计条件和要求。

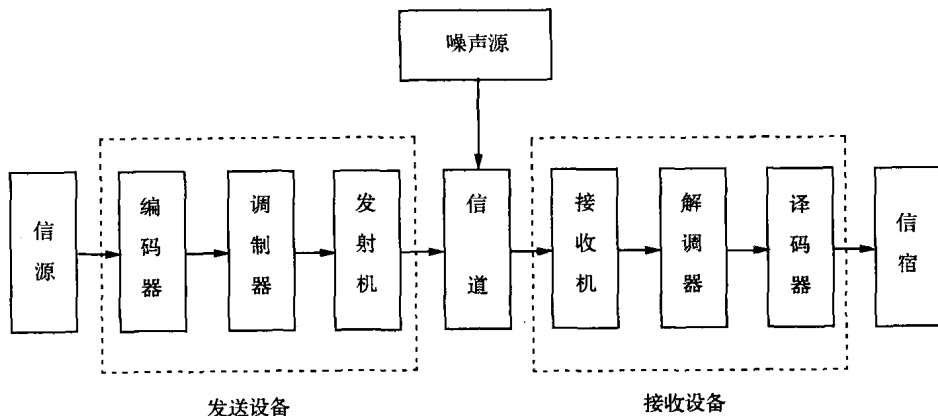


图 1.3 数字通信系统模型

数字通信的主要优点是:抗干扰能力强;差错可控;易加密;易于与现代技术相结合。数字通信主要缺点是:频带利用率不高;接收端需要严格的同步系统;设备比较复杂。不过,随着新的宽带传输信道的采用、窄带调制技术和超大规模集成电路的发展,数字通信的这些缺点已经弱化。随着微电子技术和计算机技术的迅猛

发展和广泛应用,数字通信在今后的通信方式中必将逐步取代模拟通信而占主导地位。

本书研究对象是多媒体通信系统。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,包括声音、图像、图形、数据、文本等。它是多媒体技术、通信技术和计算机技术相结合的产物。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,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。

1.1.3 多媒体信号处理

与信号有关的理论或数学过程有:信号的发生、信号的传送、信号的接收、信号的分析、信号的处理、信号的存储、信号的检测与控制等,也可以把这些与信号有关的过程统称为信号处理。严格地讲,信号处理是在事件变化过程中抽取特征信号,经去干扰、分析、综合、变换和运算等处理,从而得到反映事件变化本质或处理者感兴趣的信息的过程。本书的处理对象是多媒体信号,下面首先介绍多媒体概念及相关技术,然后介绍与本书有关的几个多媒体信号处理领域。

1. 多媒体的概念和相关技术

“多媒体”一词译自英文“multimedia”,而该词又是由 multiple 和 media 复合而成的。媒体(medium)有两重含义,一是指存储信息的实体,如磁盘、光盘、磁带、半导体存储器等介质;二是指传递信息的载体,如数字、文字、声音、图形等媒介。与多媒体对应的一词是单媒体(monomedia),从字面上看,多媒体就是由单媒体复合而成。多媒体计算机是一组硬件和软件设备,它结合各种视觉和听觉媒体,能够产生令人印象深刻的视听效果。视觉媒体包括图形、动画、图像和文字等媒体,而听觉媒体则包括语言、立体声响和音乐等媒体。用户可以从多媒体计算机同时接触到各种各样的媒体来源。总之,多媒体技术就是具有集成性、实时性和交互性的计算机综合处理声文图信息的技术。

多媒体的传统关键技术主要集中在以下四类:数据压缩技术、大规模集成电路制造技术、大容量的光盘存储器、实时多任务操作系统。因为这些技术取得了突破性的进展,多媒体技术才得以迅速的发展而成为像今天这样具有强大的处理声音、文字、图像等媒体信息的能力的高科技技术。针对互联网络的多媒体关键技术,有些专家却认为可以按层次分为媒体处理与编码技术、多媒体系统技术、多媒体信息组织与管理技术、多媒体通信网络技术、多媒体人机接口与虚拟现实技术,以及多

媒体应用技术这六个方面,而且还应该包括多媒体同步技术、多媒体操作系统技术、多媒体中间件技术、多媒体交换技术、多媒体数据库技术、超媒体技术、基于内容检索技术、多媒体通信中的 QoS 管理技术、多媒体会议系统技术、多媒体视频点播与交互电视技术、虚拟实景空间技术等。

2. 多媒体压缩

把源符号集转换成另一个符号集的特定方案、规则或映射称为编码或编码系统^[1]。在通信系统中,把原始信号转换为适合在信道中传输的格式的过程称为信源编码。如果在传输时需要降低数据率,则编码过程可称为“信号压缩”或“数据压缩”。因此,编码概念大于压缩概念。在实际系统中,编码过程往往包含了压缩,所以编码和压缩的概念常常可混用^[1]。数据压缩是信源编码的目的和手段^[4],主要目标是尽可能减小传输或存储所需的比特率或者说是给定通信系统的带宽和存储空间。数据压缩是为了在现有系统特性(如频带限制)条件下满足工作的要求,或者在新系统的设计中降低成本^[3]。

在多媒体计算系统中,若要表示、传输和处理大量数字化的声音/图片/视频等信息,数据量是非常大的。例如,一幅具有中等分辨率(640 像素×480 像素)真彩色图像(24 位/像素),它的数据量约每帧 7.37Mb。若要达到每秒 25 帧的全动态显示要求,每秒所需的数据量为 184Mb,而且要求系统的数据传输速率必须达到 184Mb/s,这在目前是无法达到的。因此,在多媒体计算机系统中,为了达到令人满意的图像、视频画面质量和听觉效果,必须解决视频、图像、音频信号数据的大容量存储和实时传输问题。解决的方法,除了提高计算机本身的性能及通信信道的带宽外,更重要的是对多媒体进行有效的压缩。多媒体数据之所以能够压缩,是因为多媒体信息存在许多数据冗余^[5]。例如,一幅图像中的静止建筑背景、蓝天和绿地,其中许多像素是相同的,如果逐点存储,就会浪费许多空间,这称为空间冗余。又如,在电视和动画的相邻序列中,只有运动物体有少许变化,仅存储差异部分即可,这称为时间冗余。此外还有结构冗余、视觉冗余等。

数据压缩可分成两种类型,一种称为无损压缩,另一种称为有损压缩。无损压缩是指使用压缩后的数据进行重构,重构后的数据与原来的数据完全相同;无损压缩用于要求重构的信号与原始信号完全一致的场合。一个很常见的例子是磁盘文件的压缩。根据目前的技术水平,无损压缩算法一般可以把普通文件的数据压缩到原来的 1/2~1/4。常用的无损压缩算法有霍夫曼算法和 LZW 压缩算法。有损压缩是指使用压缩后的数据进行重构,重构的数据与原来的数据有所不同,但不影响人对原始资料表达的信息造成误解。有损压缩适用于重构信号不一定非要和原始信号完全相同的场合。图像和声音的压缩就可采用有损压缩,因为其中包含的数据往往多于视觉和听觉系统所能接收的信息,丢掉一些数据而不至于对声音或

者图像所表达的意思产生误解,但可大大提高压缩比。

3. 多媒体数字水印

数字水印技术^[6]是将一些标识信息(即数字水印)直接嵌入数字媒体中而不影响原媒体的使用价值,也不容易被探知和再次修改,但它可以被生产方识别和辨认。通过这些隐藏在载体中的信息,可以达到确认内容创建者、购买者、传送隐秘信息或者判断载体是否被篡改等目的。数字水印是信息隐藏技术的一个重要研究方向。

按水印的特性可以将数字水印分为鲁棒数字水印和易损数字水印两类。鲁棒水印主要用于在数字作品中标识著作权信息。利用这种水印技术在多媒体内容的数据中嵌入创建者、所有者的标示信息,或者嵌入购买者的标示(即序列号)。在发生版权纠纷时,创建者或所有者的信息用于标示数据的版权所有,而序列号用于追踪违反协议而为盗版提供多媒体数据的用户。用于版权保护的数字水印要求有很强的鲁棒性和安全性,除了要求在一般图像处理(如:滤波、加噪声、替换、压缩等)中生存外,还需能抵抗一些恶意攻击。易损水印与鲁棒水印的要求相反,它主要用于完整性保护。这种水印同样是在内容数据中嵌入不可见的信息。当内容发生改变时,这些水印信息会发生相应的改变,从而可以鉴定原始数据是否被篡改。易损水印应对一般图像处理(如:滤波、加噪声、压缩等)有较强的鲁棒性,同时又要求有较强的敏感性,也就是说既允许一定程度的失真,又要能将失真情况探测出来。按水印所附载的媒体,可将数字水印分为图像水印、音频水印、视频水印、文本水印以及用于三维网格模型的网格水印等。随着数字技术的发展,会有更多种类的数字媒体出现,同时也会产生相应的水印技术。

4. 基于内容的多媒体检索

在数字图书馆环境下,以多媒体、超文本为主要存储形式的海量数字化信息资源,关键词已经难以足够形象和准确地描述多媒体信息所呈现的视觉或听觉感知,致使适用于文本信息资源的关键词检索方式显得相形见绌,而基于内容的多媒体信息检索技术在数字图书馆建设中逐渐显示出无比的优越性。

所谓基于内容的信息检索^[7]是对文本、图像、音频、视频等媒体对象进行内容语义的分析和特征的提取,并基于这些特征进行相似性匹配的信息检索技术。它与传统数据库基于关键词的检索方式相比,具有如下特点:

1) 突破了关键词检索基于文本特征的局限,直接从媒体内容中提取特征线索,使检索更加接近媒体对象。

2) 提取特征的方法多种多样。例如,可以提取图像的形状特征、颜色特征、纹理特征,视频的动态特征,音频的音调特征等。