



经典译丛

Springer



信息与通信技术

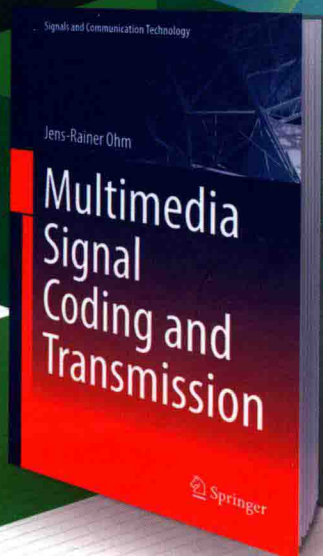
Multimedia Signal Coding and Transmission

多媒体信号编码与传输

Multimedia Signal Coding and Transmission

【德】 Jens-Rainer Ohm 著

卢鑫 金雪松 顾谦 译
赵志杰 审校



中国工信出版集团



电子工业出版社
PUBLISHING HOUSE OF ELECTRONICS INDUSTRY
<http://www.phei.com.cn>

经典译丛·信息与通信技术

多媒体信号编码与传输

Multimedia Signal Coding and Transmission

[德] Jens-Rainer Ohm 著

卢鑫 金雪松 顾谦 译

赵志杰 审校



电子工业出版社

Publishing House of Electronics Industry

北京·BEIJING

内 容 简 介

本书主要内容包括多媒体基础知识,多媒体计算机及多媒体音频/视频数据处理,多媒体的关键技术,超文本、超媒体及多媒体数据库技术等。全书共9章,包括:概论、基础知识、感知与质量、量化和编码、信号压缩方法、帧内编码、帧间编码、语音与音频编码、多媒体数据的传输与存储等。全书涵盖了图像、视频和音频压缩及表示的理论背景和多方面应用,并通过举例,使读者的理解更加深入。

本书深入浅出,且多数章节辅以练习题,适合作为电子信息、通信工程、计算机等相关专业的本科生或研究生的多媒体通信系统课程的教材,也可供多媒体信号处理的初学者或爱好者学习。本书还包括视频处理的前沿动态与应用,所以也可以为多媒体通信系统的研究人员和开发人员提供参考。

Translation from the English language edition:

Multimedia Signal Coding and Transmission by Jens-Rainer Ohm

Copyright © Springer-Verlag Berlin Heidelberg 2015

This work is published by Springer Nature

The registered company is Springer-Verlag GmbH

Authorized Simplified Chinese Language edition Copyright © 2018 Publishing House of Electronics Industry.

All Rights Reserved.

本书中文简体字版专有出版权由 Springer Nature 授予电子工业出版社。未经出版者预先书面许可,不得以任何方式复制或抄袭本书的任何部分。

版权贸易合同登记号 图字: 01-2017-7611

图书在版编目(CIP)数据

多媒体信号编码与传输 / (德) 延斯·赖纳·奥姆著; 卢鑫, 金雪松, 顾谦译. — 北京: 电子工业出版社, 2018.11
(经典译丛·信息与通信技术)

书名原文: Multimedia Signal Coding and Transmission

ISBN 978-7-121-32964-7

I. ①多… II. ①延… ②卢… ③金… ④顾… III. ①信号处理—编码技术 ②信号传输 IV. ①TN911.21
②TN911.7

中国版本图书馆 CIP 数据核字 (2017) 第 258411 号

策划编辑: 杨 博

责任编辑: 杨 博 特约编辑: 张传福

印 刷: 三河市良远印务有限公司

装 订: 三河市良远印务有限公司

出版发行: 电子工业出版社

北京市海淀区万寿路 173 信箱 邮编 100036

开 本: 787×1092 1/16 印张: 27.5 字数: 704 千字

版 次: 2018 年 11 月第 1 版

印 次: 2018 年 11 月第 1 次印刷

定 价: 109.00 元

凡所购买电子工业出版社图书有缺损问题, 请向购买书店调换。若书店售缺, 请与本社发行部联系, 联系及邮购电话: (010) 88254888, 88258888。

质量投诉请发邮件至 zlts@phei.com.cn, 盗版侵权举报请发邮件至 dbqq@phei.com.cn。

本书咨询联系方式: yangbo2@phei.com.cn。

译者序

随着信息处理技术、计算机技术和通信技术的迅猛发展,多媒体技术应运而生,并已经成为一项最具活力、发展速度很快的新兴技术之一。多媒体技术突破了计算机、通信、电视等传统产业间相对独立发展的界限,将多种前沿技术有机融合在一起。多媒体技术的广泛应用已经对人们的生产、生活产生了深刻的影响,可以预见,它必将成为科技进步、经济发展、人们生活质量提高的主要推动力之一。

原著 *Multimedia Signal Coding and Transmission* 是 Springer 出版社“信号与通信技术”系列丛书中的一部优秀著作。它由国际著名学者 Jens-Rainer Ohm 教授积多年科研和教学经验编撰而成。Ohm 教授是被誉为欧洲“麻省理工学院”的德国亚琛工业大学(RWTH Aachen)通信工程研究所的所长,H.264/AVC、H.265/HEVC 标准的主要制定者之一,国际电信联盟(ITU)联合视频协作组(JVC-VC)主席,德国工程师协会信息技术组发言人,国际视频编码标准制定过程中的领军人物及视频信号处理领域的权威学者。

本书紧密联系现代多媒体系统的发展需要,将多媒体信息处理(编码)与无线及有线通信传输融为一体,系统地介绍了多媒体信号与信息处理的基础理论以及研发、应用的成果。全书共 9 章,涵盖了图像、视频、语音和音频的压缩及表示的理论背景和多方面应用,并通过举例,使读者的理解更加深入。内容深入浅出,适合视频信号处理的初学者阅读;大多数章节都辅以练习题,适合学生进行课程学习;书中包括视频信号处理的前沿动态与应用,可为工业多媒体通信系统的研究人员和开发人员提供参考。

第 1 章概论:主要引入和介绍基本概念。第 2 章基础知识:介绍信号与系统的基本知识、傅里叶变换、多媒体信号的采样、离散信号处理、统计分析、线性预测、线性块变换与滤波器组变换等内容。第 3 章感知与质量:介绍人类的视觉/听觉特性,引入信号质量评价标准。第 4 章量化与编码:介绍标量量化和脉冲编码调制、编码理论、量化器的率失真优化、熵编码、向量量化等知识。第 5 章信号压缩方法:包括行程编码、预测编码、变换编码、可伸缩编码、多描述编码和分布式信源编码等。第 6 章帧内编码:包括二值图像编码、多幅值图像编码、无损和近无损图像编码、三维图像编码、重构滤波与静止图像编码标准等。第 7 章帧间编码:介绍混合视频编码、时空变换编码、可伸缩视频编码、多视图编码及视频编码标准等内容。第 8 章语音与音频编码:包括语音、音频、音乐和声音编码的方法与标准等。第 9 章多媒体数据的传输与储存:包括数字多媒体服务、网络接口、媒体储存与传输的相关知识等。

本书由卢鑫、金雪松、顾谦翻译,其中卢鑫博士翻译前言、第 1 章、第 2 章、第 4 章、第 6 章、第 7 章、第 8 章、第 9 章,金雪松博士翻译第 3 章,顾谦硕士翻译第 5 章、附录 A、附录 B、附录 C、附录 E。顾谦硕士、余唱硕士、周必兴硕士协助校对了全书的图表和公式。赵志杰教授负责全书的审校。

本书在翻译过程中,得到了多位老师和研究生的帮助与参与。特别需要感谢英国华威大

学(University of Warwick)的 Graham R. Martin 教授为本书的翻译提供的宝贵意见和建议。感谢同学们的辛苦校对，感谢译者家人对本书翻译工作的支持与帮助，感谢杨博编辑在本书出版过程中付出的辛勤劳动。

本书的出版得到了国家自然科学基金重点国际合作研究项目(61720106002)、国家自然科学基金青年基金项目(61401123)、中央高校基本科研业务费专项资金(HIT.NSRIF.201617)、哈尔滨市科技创新人才项目(优秀学科带头人 2017RAXXJ055)等的支持，在此表示衷心感谢。

由于时间仓促，译者水平有限，译文中如有不妥之处，恳请广大读者谅解，并提出改进意见，不胜感激！

卢 鑫

于哈尔滨

2017 年 7 月

前言^①

近几十年来,数字视听信息在专业领域及日常生活中可谓无处不在。视听媒体已经深刻地改变了人们沟通合作、信息获取、学习、工作以及交互的方式。与此同时,传统的分发和访问方式已被互联网、移动网络和数字存储逐渐取代。在此背景下,多媒体通信构建了人与/或机器之间进行通信的全新模式,其特点包括普遍性、综合性、交互性以及智能性。

多媒体通信系统课程在亚琛工业大学已开设多年,本书连同 *Multimedia Content Analysis*^② 正是以该课程的讲义为基础,是 2004 年出版的教材《多媒体通信技术》(*Multimedia Communication Technology*)的全面升级版。多媒体信号编码和传输是本书讨论的主要内容,而另一本书《多媒体数据内容分析》(*Multimedia Content Analysis*, MCA)则主要围绕多媒体信号的识别展开。这两本书中的内容(以及作为这两本书基础的课程讲义)都是自成体系的,因此不能将其理解为一套图书的上部和下部。但是,编码和内容分析之间存在着许多共性问题(两者都建立在信号处理和信息论的思想基础上),因此,读者经常会发现两书之间存在交叉引用(包括指向章节号的索引)的情况。关于信号处理(Signal Processing, SP)和信息论(Information Theory, IT)的基本知识有助于对本书内容的理解,因此,第 2 章和第 4 章针对相关的基本概念进行了总结。

自《多媒体通信技术》(2004 版)出版以来,视听数据的压缩技术又取得了惊人的进展。因此,本书体现了该领域的最新进展,其中包括关于高效视频编码(High Efficiency Video Coding, HEVC)标准所运用的编码思想的更深层次理解,以及三维视频和音频数据的压缩方法,同时本书还简述了未来可能的发展趋势。但是,本书无意阐述现有标准的具体实现细节,而是侧重于为读者提供关于基本概念的深入理解,最终为读者成为掌握该领域新兴技术的设计人员提供支持。

本书多数章节都辅以习题加以补充,访问网址 <http://www.ient.rwth-aachen.de> 可以获得参考答案^③。

本书的出版离不开学术和标准化机构中广大师生、科技人员及同事们的贡献,以及 25 年来在图像、视频、音频的处理、编码和识别领域与本人开展合作的其他人员的努力。为简洁起见,不进行一一致谢,在此向所有人一并表示感谢。

Jens-Rainer Ohm

于德国亚琛

2014 年 12 月 21 日

① 中译本的一些图示、参考文献、符号及其正斜体形式等沿用了英文原著的表示方式,特此说明。

② 该书也将由电子工业出版社引进出版。正文中引用该书时用 MCA 表示。

③ 也可以向 te_service@phei.com.cn 发送邮件申请。

目 录

第 1 章 概论	1
1.1 概念和术语	1
1.1.1 信源编码的信号表示	2
1.1.2 传输的优化	4
1.2 信号源与信号的获取	5
1.3 多媒体信号的数字表示	7
1.3.1 图像与视频信号	8
1.3.2 语音和音频信号	13
1.3.3 压缩技术的必要性	13
1.4 习题	13
第 2 章 基础知识	15
2.1 信号与系统	15
2.1.1 基本信号	15
2.1.2 系统运算	17
2.2 信号与傅里叶频谱	19
2.2.1 二维和多维坐标上的频谱	19
2.2.2 时空信号	21
2.3 多媒体信号的采样	23
2.3.1 可分离的二维采样	24
2.3.2 不可分离的二维采样	26
2.3.3 视频信号的采样	30
2.4 离散信号处理	34
2.4.1 线性移不变系统	34
2.4.2 离散傅里叶变换	34
2.4.3 z 变换	35
2.4.4 多维线性移不变系统	37
2.5 统计分析	39
2.5.1 样点的统计特性	39
2.5.2 联合统计特性	43
2.5.3 随机信号的频谱特性	47
2.5.4 马尔可夫链模型	48
2.5.5 信息论的统计学基础	50
2.6 线性预测	54
2.6.1 自回归模型	54

2.6.2	线性预测	60
2.7	线性块变换	70
2.7.1	正交基函数	70
2.7.2	正交变换的类型	73
2.7.3	变换的效率	84
2.7.4	重叠块变换	84
2.8	滤波器组变换	86
2.8.1	抽取和插值	87
2.8.2	子带滤波器的性质	91
2.8.3	滤波器组结构的实现	96
2.8.4	小波变换	101
2.8.5	二维与多维滤波器组	106
2.8.6	金字塔分解	109
2.9	习题	112
第3章	感知与质量	118
3.1	视觉特性	118
3.1.1	眼生理学	118
3.1.2	灵敏度函数	119
3.1.3	彩色视觉	122
3.1.4	双目视觉与重现	122
3.2	听觉特性	124
3.2.1	耳生理学	124
3.2.2	灵敏度函数	124
3.3	质量测度	127
3.3.1	客观信号质量测度	128
3.3.2	主观评价	133
3.4	习题	137
第4章	量化与编码	138
4.1	标量量化与脉冲编码调制	138
4.2	编码理论	143
4.2.1	信源编码定理与率失真函数	143
4.2.2	相关信号的率失真函数	144
4.2.3	多维信号的率失真函数	147
4.3	量化器的率失真最优化	148
4.4	熵编码	152
4.4.1	变长码字的特性	152
4.4.2	哈夫曼码的设计	154
4.4.3	系统变长码	155

4.4.4	算术编码	159
4.4.5	自适应与基于上下文的熵编码	163
4.4.6	熵编码与传输误差	166
4.4.7	伦佩尔-齐夫编码	167
4.5	向量量化	169
4.5.1	向量量化的基本原理	169
4.5.2	使用均匀码书的向量量化	172
4.5.3	使用非均匀码书的向量量化	174
4.5.4	结构化码书	176
4.5.5	自适应向量量化	180
4.5.6	码率约束向量量化	182
4.6	网格编码量化	183
4.7	习题	186
第 5 章	信号压缩方法	188
5.1	行程编码	188
5.2	预测编码	190
5.2.1	开环和闭环预测系统	190
5.2.2	非线性移变预测	192
5.2.3	传输损耗的影响	193
5.2.4	向量预测	193
5.2.5	延迟决策预测编码	196
5.2.6	多分辨率金字塔中的预测	197
5.3	变换编码	200
5.3.1	离散变换编码的增益	201
5.3.2	变换系数的量化	203
5.3.3	变换系数的编码	206
5.3.4	有损传输下的变换编码	208
5.4	具备多重解码能力的码流	208
5.4.1	联播与转码	209
5.4.2	可伸缩编码	209
5.4.3	多描述编码	210
5.5	分布式信源编码	212
5.6	习题	213
第 6 章	帧内编码	215
6.1	二值图像压缩	215
6.1.1	二级图像的压缩	215
6.1.2	二值形状编码	218
6.1.3	轮廓编码	220

6.2	图像的向量量化	221
6.3	图像的预测编码	223
6.3.1	二维预测	223
6.3.2	二维块预测	224
6.3.3	预测误差的量化与编码	226
6.3.4	二维 DPCM 中的误差扩散	227
6.3.5	二维预测编码中边信息的编码	228
6.4	图像的变换编码	230
6.4.1	块变换编码	230
6.4.2	变换系数的量化与编码	232
6.4.3	重叠块变换编码	240
6.4.4	子带和小波变换编码	241
6.4.5	针对信号特性的变换基局部自适应	246
6.5	无损和近无损图像编码	248
6.6	基于合成的图像编码	249
6.6.1	基于区域的编码	250
6.6.2	颜色和纹理合成	252
6.6.3	照明效果的合成	254
6.7	分形图像编码	255
6.8	三维图像编码	257
6.9	重构滤波	258
6.10	静止图像编码标准	261
6.11	习题	264
第 7 章	帧间编码	268
7.1	帧内补偿编码	268
7.2	混合视频编码	269
7.2.1	运动补偿混合编码器	269
7.2.2	帧内预测误差信号的特性	272
7.2.3	量化误差反馈和误差扩散	274
7.2.4	运动补偿预测中的参考图像	276
7.2.5	运动补偿的精度	287
7.2.6	隔行视频信号的混合编码	292
7.2.7	混合编码器的优化	294
7.2.8	采用子带/小波变换的混合编码	296
7.3	时空变换编码	298
7.3.1	帧间变换与子带编码	298
7.3.2	运动补偿时域滤波	300
7.3.3	基于 MCTF 表示的量化与编码	304
7.4	边信息的编码(运动、模式、划分方式)	306

7.5	可伸缩视频编码	312
7.5.1	混合视频编码中的可伸缩性	313
7.5.2	可伸缩视频编码的标准化	319
7.6	多视图视频编码	322
7.7	基于合成的视频编码	328
7.7.1	基于区域的视频编码	329
7.7.2	超分辨率合成	331
7.7.3	动态纹理合成	331
7.8	视频编码标准	333
7.9	习题	340
第 8 章	语音与音频编码	342
8.1	语音信号编码	342
8.1.1	线性预测编码	342
8.1.2	参数(合成)编码	345
8.1.3	语音编码标准	346
8.2	一般音频、音乐与声音编码	347
8.2.1	音频信号的变换编码	347
8.2.2	基于合成的音频与声音信号编码	350
8.2.3	立体声与多声道音频信号的编码	352
8.2.4	音乐与声音编码标准	358
第 9 章	多媒体数据的传输与存储	362
9.1	数字多媒体服务	362
9.2	网络接口	363
9.3	对信道特性的适应性	366
9.3.1	码率与传输控制	367
9.3.2	错误控制	370
9.4	媒体传送、存储和再现	374
9.4.1	广播应用	375
9.4.2	通信服务	378
9.4.3	文件存储	379
9.4.4	互联网和移动流媒体	380
附录 A	向量与矩阵代数	385
附录 B	符号和变量	391
附录 C	缩略语	395
附录 D	参考文献	403
附录 E	词汇表	418

第1章 概 论

多媒体通信是指通常可由人类感官感知的多种类型信息的结合。在此背景下,视听信息(语音/音频/声音/图像/视频/图形)显得尤为重要,但是就信号处理的复杂度而言,对视听信息的处理又面临很大挑战。多媒体信息的数字表示为数据的交换、分发、获取提供了更大的自由度,其中的通信包括人与人之间、人与设备之间,或设备与设备之间的信息交换。对视听数据进行再现,总是希望获得良好的质量,但是,用于数据传输的比特数量通常是有限的,因此压缩技术成为一项关键技术。本章将介绍多媒体信号压缩和传输中所涉及的基本概念和术语,并对视听信号源数字表示的获取、常见格式和数据量进行概述。

1.1 概念和术语

通信系统设计的基本原则是,以尽可能低的代价传递信息,同时为用户提供足够高的质量,使用户最大限度地受益。多媒体通信中,由于数据量大而对传输速率和处理速度要求较高的信息类型是最为重要的,包括音频(audio)、图像(image)、视频(video)和图形(graphics)等信源。对于这些类型的数据,通信系统(网络和设备)往往会在极限状态下工作。因此,在开发具有更强处理能力和更大存储空间的设备以及具有更高容量的网络时,视听信号的处理、存储和传输已经成为最为重要的驱动因素之一。多媒体信源有的是由传感器采集产生的(比如,能够完成模数转换的相机、话筒等),有的是通过合成获得的(比如,图形、合成的语音或声音等)。为了将信息再现以供人们使用,通常需要进行其他转换,将信息搭载到相应的物理媒介上(比如,通过扬声器产生声波)。

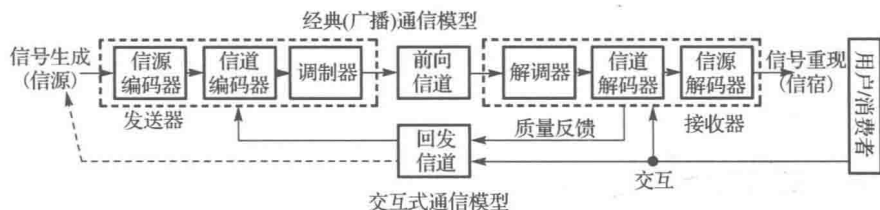


图 1.1 经典通信系统与反馈/交互信道的概念

单向通信系统的经典概念如图 1.1 的上半部分所示。该模型适用于不存在回发信道的应用环境,比如广播。左侧的设备(发射器)与右侧的设备(接收器)是相辅相成的。信源编码器与解码器、信道编码器与解码器,采用协议规范互相配合工作,设法补偿信道损失和信道容量的局限。这种系统的目标是以尽可能高的质量传递尽可能多的信息。而这是通过优化信源编码器、信道编码器和调制器实现的:信源编码器将信源压缩,以尽可能少的比特数表示信源;信道编码器保护码流免于可能产生的损失;调制器以最合适的方式在物理信道中发送符号载体。考虑到信道特性(比如,由干扰和噪声造成的信道损耗)以及可用的物理带宽,有必要对上述各模块进行联合优化。信道具有某一容量(capacity),表示该信道能在发射器和接收器之

间进行无损失传输时的最大信息速率。在这种意义上,信源编码的主要任务是,将通常存在于多媒体信源原始信息中的冗余去除(remove redundancy),并以较低的码率将其压缩。信道编码加入冗余(add redundancy),这是为了当出现信道损失时,为码流提供差错保护。信源产生的码率不可以超过信道容量,如果出现这种情况,要么会产生信息损失,要么需要采用更有效的方法对信源进行压缩(见 1.1.1 节)。而后者可能是更好的选择,因为它受到编码器的控制,即便在引入较多失真时,也可以产生比较平稳的降质,而不会造成随机的信道损失。

在通信的经典概念中,接收器会被动地、确定性地解码一切经信道接收到的信号。该模型适用于广播系统,这种系统没有必要对某一用户的干预做出反应,但是,在点对点通信系统中,情况便不同了。经典的通信系统模型通过回发信道和交互组件加以扩展,如图 1.1 的下半部分所示。用户本身能够与系统链路中的任何环节进行交互并产生影响,其中包括信号的产生环节。而且,接收器的任何一个环节都可以自动地产生反馈,例如,请求重传丢失的信息,或者通知发射器当前信道的状况等。更“智能”的系统甚至包括信号内容分析等器件,用来协助用户完成交互以及搜索多媒体数据。

除传统的广播信道外,如今往往需要通过特征差异非常大的异构信道(heterogeneous channels)完成多媒体数据流的传输(例如,基于互联网的物理传输)。尤其在无线和移动传输中,由于信道特征的瞬时波动,需要频繁而且适当地对传输进行调整。从广义的角度来看,信道可以是网络或者存储设备,其中不仅可以在传输链路的末端(服务器和客户端)加入存储设备,还可以在代理服务器中设置临时存储设备,用于快速访问可能会被多次存取的内容。当出现传输错误时,在客户端进行信息恢复,要远比传统的信道和信源解码复杂得多,比如,需要采用包括内容分析在内的先进的差错隐藏方法。

归根到底,可以根据为用户带来的价值,即质量与成本之比(质量/成本)来评价一个多媒体通信系统的性能。服务质量(Quality of Service, QoS),这里也可以将其理解为感知质量,得到了广泛的使用。传输速率、延迟和损耗等网络参数,信源编码/解码的压缩性能,以及与信道特性的相互关系都会影响 QoS。向用户收取的费用取决于完成这些任务的设备成本。因此,在满足用户需求的前提下,最好采用复杂度较低的系统。

本书以多媒体信号的编码表示(coded representation)和传输(transmission)为重点进行讨论。与此相关的概念和术语将在本节的余下部分进行简要的介绍。

1.1.1 信源编码的信号表示

通过多媒体信号压缩(multimedia signal compression),在获得尽可能高的视觉质量的条件下,可以实现最简洁的表示。采集到信号以后,将其转换成具有有限样点个数与幅度等级的数字表示。这一步骤将会对最终的质量产生影响。如果在获取信号时,并不知道某个候选信道能够传输的码率范围或某种应用所需的分辨率,则可取的做法是,以尽可能高的质量采集信号,必要时可在以后对信号进行下采样,以获得质量较差的信号。

冗余(redundancy)(信号样点之间存在的相似性,或某些信源状态出现得较为频繁)是信号本身的属性,它使信源编码器(source coder)能在不造成信息损失的条件下,降低码率。如果针对的用户是人,考虑到感知特性(perceptual properties),由于为用户提供的质量没有必要比其所能(或所希望)感知到的更加精细,因此可以进一步对信源编码方法进行调整。某些发生在信源编码和解码过程中的失真可能不会被感知到,或者失真很小而能够容忍,或者需要进

行修复(比如,去除传感器噪声)。这些是原始信号中无关紧要的(irrelevance)部分,编码时最好能将其去除^①。然而,如果信道容量很小,仅去除冗余或不重要部分可能不足以降低码率。那就难免会引起其他失真。这些失真在本质上是不同的,例如,空间或时间分辨率的降低,或者编码/量化噪声。还可以考虑与内容相关的(content-related)属性。例如,对用户来说更加重要的部分或片段,可以为其提供较好的质量。

编码信息通常以二进制数[比特(bits)]的形式进行表示。码率的单位是比特/样点(bit/sample)^②,或比特每秒(bit per second)[比特/秒(b/s)],其中后者是由比特/样点(bit/sample)的值乘以样点/秒(samples/s)(采样率)的值得到的。压缩率(compression ratio)是评价信源编码方案性能的一项重要指标。它表示的是原始未压缩信源和压缩后的信源所需码率之比。比如,对数字高清电视而言,如果原始未压缩的信源需要至少 800 Mb/s^③的码率,若压缩后的码率须达到 4 Mb/s,那么所需的压缩率就是 $800:4=200$ 。如果压缩的信号流以文件的形式存储在计算机磁盘上,可以根据文件的大小,评价压缩性能。当将文件大小转换为码率时,需要注意的是,通常文件的大小是以 K(千)字节(Kbyte, KB), M(兆)字节(MByte, MB)等为单位的,其中 1 字节(Byte)由 8 比特(bit)组成,1K 字节(KByte)=1 024 字节(Byte),1M 字节(MByte)=1 024 K 字节(KByte)。

典型多媒体信号编码和解码系统的基本框图,如图 1.2 所示。

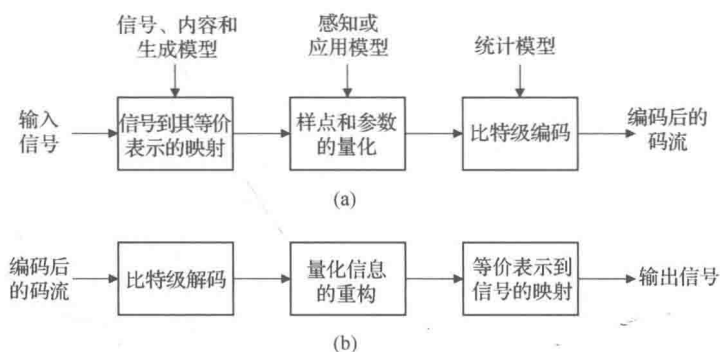


图 1.2 多媒体信号编码系统的基本原理。(a) 编码器; (b) 解码器

- 第一步是信号样点到其等价表示的映射,理想情况下,通过反向映射可以完全恢复信号样点。一般来说,在等价表示空间中的表示要比在原始信号空间中的表示更加稀疏,所以后续步骤可以变得更简单并能更好地对信号进行优化。稀疏性指的是等价表示需要更少的样点,或者样点的幅度集中在较少的幅值上。预测(prediction)和变换(transform)是实现这种映射的重要方法。当采用线性变换完成映射时,映射又被称为去相关(de-correlation),这是因为,映射通常可以去除在等价表示空间中样点之间的线性统计相关性。可以采用统计模型对这种系统进行优化,同时还需要考虑信号的生

① 当针对的用户是机器时,信号中哪些部分是无关紧要的,由信号的使用目的决定。例如,信源编码引起的失真可能会导致对内容的分析失败,这是无法接受的。

② 图像信号还使用比特/像素(bit/pixel, b/p),其中“像素”作为图像元素(picture element)的缩写被广泛使用,即图像样点。为了在用词上对各种类型的信号没有偏向性,对于图像和视频,我们也使用“样点”一词。

③ Mb/s: 10^6 比特每秒, kb/s: 10^3 比特每秒。

成(例如, 获取过程的特性)和语义的内容。如果映射函数是可变的(正如具有局部变化特征的多媒体信号通常所需的), 则可能需要产生额外的边信息(side information), 在解码时需要这些边信息来完成逆映射。

- 第二步是**量化(quantization)**, 它将信号、信号的等价表示或边信息参数映射成离散字符集中的值(如果之前的值并不是离散字符集中的值, 或者需要进行压缩效果更明显的有损压缩)。在量化时, 还应考虑观察者的感知局限或其他非相关性准则(与预期的应用相关), 这样利用给定数量的离散(量化值)状态便可以尽可能多地保持重要信息。一般来说, 量化值状态的个数较少, 自然会使每个样点需要的比特数也较少, 但是也会使失真增加。
- 最后一步是**比特级编码(bit-level encoding)**, 这一步的目的是以更低的码率表示离散信息状态, 而通常并不改变这些信息状态。编码的优化主要是以统计原则为基础进行的, 比如各个状态出现的频率。各个状态实现唯一性表示(比如, 有限离散字符集中的字符)所能达到的最低码率称为**熵(entropy)**。

除**码率(rate)**和**失真(distortion)**外, 其他调整信源编码算法的参数还包括**延迟(latency)**和**复杂度(complexity)**。这4个参数是相互影响的。码率与失真之间的关系是由**率失真函数(rate distortion function)**决定的(见2.5.5节和4.2.1节), 当给定失真的上限时, 利用率失真函数可以得到码率的理论下限值。增加编码/解码算法的复杂度往往能够使率失真性能得到提高(就是说, 在保持失真不变的条件下, 获得更高的压缩率), 例如, 根据各种信源的不同统计特性, 采用更复杂的方式实现编解码算法。另外, 增加延迟也可以提高压缩性能, 比如, 如果编码器能够预知当前的决策对后续编码过程所产生的影响, 则将有利于提高编码器的性能。

码率和延迟能够直接利用许多参数进行定量评价, 比如, 均值、峰值和变化量等, 而对失真的定量评价将需要较复杂的度量方式, 其中需要考虑用户的感知, 包括均方误差、信噪比(SNR)等参数被广泛采用, 同时这些参数也较容易进行计算(见3.3节)。对复杂度的定量评价也是个挑战, 因为同一编/解码器的各种组件可能会很容易地在不同的平台上实现, 但也可能会很困难, 并且复杂度的判断需要考虑多种因素, 比如, 逻辑门的数量、芯片尺寸、功耗、内存使用情况、软件/硬件友好性、并行运算能力等。

1.1.2 传输的优化

信源编码器与信道之间的关系对整体感知服务质量也是至关重要的。信源编码器去除信号中的冗余, 而信道编码器向码流中加入冗余(add redundancy), 以达到在有损环境下保护和恢复码流的目的。在接收端, 信道解码器去除由信道编码器加入的冗余, 并利用这些冗余进行纠错, 而信源解码器补充由信源编码器去除的冗余, 利用接收到的信息尽可能好地对信源进行重构。由此看来, 信源编码/信道解码与信道编码/信源解码的作用是相似的。实际上, 最复杂的部分通常处于去除冗余的一端, 因为这部分的目标是对完备表示中的**重要信息(relevant information)**进行估计。事实上, 为了获得最优性能, 需要对信源编码和信道编码进行联合优化。比如, 利用信道编码, 向码流中对用户来说不太重要的部分加入冗余, 是没有什么意义的。信道上的传输还包括调制, 在通信技术中, 为了逼近信道容量的极限, 通常会联合使用信道编码和物理层调制^[Proakis, Salehi, 2007]。

在设计多媒体系统时, 利用某一模型把信道看成一个“黑匣子”, 往往是比较有利的。这与误差/损失特性、传输速率、延迟(延时)等对网络质量来说最重要的参数是密切相关的。

当可以保证网络的最低质量时,则几乎能够以最优的方式,使信源编码和网络传输相匹配。这通常需要通过协商协议来实现。如果网络的质量无法保证,可以引入专门的机制使服务器端和客户端相互匹配。这包括基于网络质量评估的专用差错保护机制或重传协议。为了提高传输质量,引入时延也是一种切实可行的方法,例如,优化传输调度,在信息传输到接收端之前采用信息临时缓冲,或者当传输中可能发生突发丢包时,对码流进行置乱/交错(见 9.3 节)。在极端情况下,接收器能够请求重传丢失或已损坏的信息,直到最终收到该信息,但这样做将会增加传输延迟。信道编码部分也同样如此,如果允许采用具有较长时延(higher latency)或较高复杂度(higher complexity)的系统,则预计将会获得更好的质量。然而,这并不是对所有应用都适用。比如,实时会话服务(比如,视频电话)只允许低时延,对于需要快速反应的所有交互应用,也同样如此。对于移动设备,其电池的容量是最为关键的,因此,一般应维持较低的整体系统复杂度。

从物理上来说,抽象的信道通常是由若干具有不同特性的网络路径组成的链路。此时,整个链路中最薄弱的环节决定了信道的性能。在异构网络中,媒体流具有适应网络特性变化的能力,是非常重要的。在某些情况下,可能需要将码流转码(transcode)成一种不同的形式,使其能够更好地适应特定的网络特性。从这个意义上说,能够产生具有很强适应能力并可以独立于编码过程而被截断的可伸缩码流(scalable streams)的信源编码方法是非常具有优势的(见 5.4.2 节)。另外一种方法是基于内容多描述(multiple descriptions)的编码方法(见 5.4.3 节),这种方法把内容的多个描述经过并行的传输路径传送给接收器。

1.2 信号源与信号的获取

多媒体系统通常处理的是数字表示的信号,而自然信号(natural signals)的获取和产生可能并不是由数字设备直接完成的,例如,转换中可能会涉及电磁(话筒)、光学(镜头)、化学(胶片)媒介。模数转换本身由采样和量化两个步骤组成,采样是将空间-时间连续的信号映射为离散的样点,量化是将幅度连续的信号映射为数值。采集装置,比如配备 CCD 或 CMOS 传感器的数码相机,在输出端本身就能提供采样的信号和量化的表示。多媒体信号也可以通过合成产生,比如,图形、合成声音、人造音效等,还可以通过将自然声音和合成声音混合而产生。即使采集到了自然信号,原本存在于外部(三维)世界的部分信息也会丢失,这是由于:

- 采集设备的频率带宽或分辨率是有限的;
- 采集设备的“非遍及性”,采集设备位于三维外部世界的某一位置,所以只能在这一特定的观察点或收听点获取信号,即使使用多个相机或话筒,采集到的三维空间信息也是不完整的(incomplete),这是因为传感器不可能布置得无限密集。

下面以相机系统为例,讨论外部世界与信号的不完全投影之间的关系。用“世界坐标系”表示位置 $\mathbf{W}=[W_1, W_2, W_3]^T$, 其中,在某一特定位置 $\mathbf{W}$, 某一特定时间 t 的成像特性[又称为全视函数(plenoptic function)] $I(\mathbf{W}, \Theta, t)$, 由来自角度方向 $\Theta=[\theta_1 \ \theta_2 \ \theta_3]^T$ 的光线颜色和强度给出[Mcmillan, Bishop, 1995]。这可以通过位于位置 $\mathbf{W}$ 、光轴(观察线)方向为 Θ 的相机获取。对于这种“光场”表示,一台相机只能在位置 $(\mathbf{W}, \Theta)$ 处获得一个样点,若有其他相机,则能够使这个样点变得更加完整。另外,其他视图还可以通过内插或者外推产生(见 MCA^①, 第 7 章)。

① MCA 表示 Multimedia Content Analysis 一书, 后文同。——编者注

针孔相机(pin-hole camera)模型如图 1.3 所示。这里,世界坐标系以相机所在的位置为原点,其中,相机的光轴沿 W_3 (深度)轴方向,垂直于图像平面。相机的二维图像平面(例如,胶片或电子传感器)在时间 t 接收到由三维外部世界投射到图像平面的投影。在针孔相机模型中,可以保证只有一束(one)来自外部世界的光线(light ray)可以到达图像平面中的一个点,这样便能够保证图像是聚焦的^①。外部世界坐标系中, $\mathbf{W}_P = [W_{1,P}, W_{2,P}, W_{3,P}]^T$ 位置处的一点 P 与其在图像平面中的投影——点 $\mathbf{t}_P = [t_{1,P}, t_{2,P}]^T$ 之间的关系,由中心投影方程(central projection equation)确定,其中 F 是相机的焦距,

$$t_{i,P} = F \frac{W_{i,P}}{W_{3,P}}, \quad i=1,2 \quad (1.1)$$

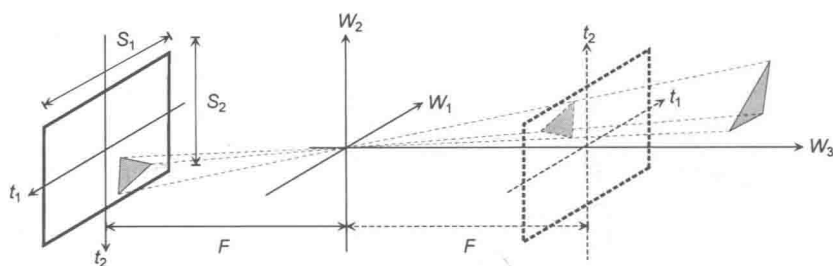


图 1.3 针孔相机模型

世界坐标系的原点位于“针孔”的位置,即相机的焦点。图像平面本身位于 $W_3 = -F$, 其中心位于 $W_1 = 0$ 、 $W_2 = 0$ 处; $W_2 = 0$ 处的指向为 t_1 轴方向的图像平面,与 $-W_1$ 平行; $W_1 = 0$ 处的指向为 t_2 轴方向的图像平面,与 $-W_2$ 平行。当光线穿过位于 $W_3 = F$ 的镜像平面时,同样的投影方程依然成立,因此,两个坐标轴 t_i 不需要关于相应的 W_i 镜像对称。图像平面本身的宽度为 S_1 , 高度为 S_2 。关于光轴对称的水平视角和垂直视角为

$$\varphi_i = \pm \arctan \frac{S_i}{2F}; \quad i=1,2 \quad (1.2)$$

在电子相机和扫描仪中,二维图像平面通常是按照由左及右/从上到下的顺序逐行进行扫描和采样的。然后,宽度为 S_1 、高度为 S_2 的图像平面被映射入间距为 $T_i = S_i/N_i$ 的 $N_1 \times N_2$ 个离散采样位置。这种采样过的并且空间上有界的图像可以用矩阵表示。一般情况下,将图像左上角的样点作为坐标原点(0,0),这个样点也是矩阵左上角的元素^②。

可以用一个包含 $N_1 \times N_2$ 个样点的矩阵对此进行说明,如图 1.4 所示,

$$\mathbf{S} = \begin{bmatrix} s(0,0) & s(1,0) & s(2,0) & \cdots & \cdots & s(N_1-1,0) \\ s(0,1) & s(1,1) & & & & s(N_1-1,1) \\ s(0,2) & & \ddots & \ddots & & \vdots \\ \vdots & & & & & s(N_1-1,N_2-2) \\ s(0,N_2-1) & s(1,N_2-1) & \cdots & \cdots & s(N_1-2,N_2-1) & s(N_1-1,N_2-1) \end{bmatrix} \quad (1.3)$$

① 从物理上来说,这是不可能做到的,因为小孔本身必然具有一定直径。在带有镜头的相机中,只有在镜头聚焦时,来自外部世界同一点的多条光线才能会聚于图像平面中的某一点。

② 需要注意的是,中心投影方程中使用的二维坐标系,其原点位于光轴穿过图像平面的位置(可能位于现有样点之间的位置),相对于这一坐标系,此坐标系存在一个偏移量。