

SZSYPIJSYY

数字视音频技术应用

SHUZI SHIYINPIN JISHU YINGYONG

刘毓敏 等编著

数字视音频技术应用

刘毓敏 等编著



机械工业出版社

本书在对数字视音频存储/传输系统的4项关键技术——信源编码技术、信道编码与调制技术、存储技术及网络传输技术进行系统介绍的基础上,对数字视音频技术在消费电子领域、广播电视领域、多媒体计算机领域及多媒体通信领域的实际应用进行了较全面的介绍,使读者对数字视音频技术有系统全面的了解。本书可作为广播电视、家用电器、计算机、通信等领域的从业人员使用,也可供中职中专相关专业师生参考。

图书在版编目(CIP)数据

数字视音频技术应用/刘毓敏等编著. —北京:机械工业出版社,2003.7
ISBN 7-111-12452-9

I. 数… II. 刘… III. ① 视频信号-数字技术 ② 音频设备-数字技术 IV. ① TN941.3 ② TN912.271

中国版本图书馆CIP数据核字(2003)第049392号

机械工业出版社(北京市百万庄大街22号 邮政编码100037)
责任编辑:贾玉兰 版式设计:冉晓华 责任校对:李汝庚
封面设计:陈沛 责任印制:闫焱
北京京丰印刷厂印刷·新华书店北京发行所发行
2003年9月第1版第1次印刷
787mm×1092mm^{1/16}·15.75印张·388千字
0 001—4 000册
定价:25.00元

凡购本书,如有缺页、倒页、脱页,由本社发行部调换
本社购书热线电话(010) 68993821、88379646
封面无防伪标均为盗版

前 言

数字视音频技术是电子信息领域职业技术教育的主要内容,这是因为,一方面,在电子信息领域,由于数字化技术迅猛发展,使得原来界限分明的专业领域,如广播电视、家用电器、计算机、通信等的界限越来越模糊,无论是技术层面,还是业务层面都出现了交叉和融合的发展趋势。作为电子信息数字化核心内容的数字视音频技术,正是这种交叉与融合的技术基础。数字视音频技术知识已经成为上述专业领域技术能力的有机组成部分。另一方面,作者注意到近年来与数字视音频技术教育有关的两个发展趋势:一是,大家都看好的,以计算机、通信、广播电视、家用电器为主体的信息技术产业的良好发展势头;二是,我国在经济发达地区近年来大力普及高等职业教育。基于上述两方面的认识,我们编写这本书,供广播电视、家用电器、计算机、通信等领域的从业人员使用,也可供中职中专相关专业师生参考。

我们采用了“从具体到抽象再到具体”的结构。即先对目前已经实用化的各类数字视音频系统进行抽象,并在此基础上形成了“数字存储/传输系统”模型(这主要是第一章第一节的内容)。再从这4项关键技术(尤其是信源编/解码技术)在各种实用化的数字视音频系统中的具体表现(即个性)中,进一步深化其技术原理的认识(这主要是第二、三、四、五和六章的内容)。

数字视音频技术在不同的应用领域中的具体表现形式可以说是丰富多彩、五花八门、令人眼花缭乱的。本书采用上述编写结构的优点是,有助于初学者从数字视音频技术的五花八门的各种具体应用形态中解脱出来,迅速深入到该技术的核心来把握该技术的本质特征,然后再从该技术的各种具体表现形态的比较中进一步深化认识。当然,这仅是作者的主观愿望,能否达到预期的效果还有待广大读者和同行在阅读中检验。

本书由刘毓敏策划,参加编写的有刘毓敏(第一章、第四章以及第二章、第三章、第五章和第六章的部分内容)、黄燕(第二章和第三章的部分内容)、杜嵩榕(第五章和第六章的部分内容)。全书由刘毓敏统稿。作者对在本书编写过程中所参考主要文献均附于书后,在此向有关作者致以衷心感谢,如有疏漏敬请原谅。

鉴于作者水平有限,书中错漏在所难免,敬请读者和同行批评指正。

作者

目 录

前言

第一章 数字视音频技术基础 1

第一节 数字视音频应用系统技术

概况 1

一、系统模式 1

二、系统性能 2

(一) 有效性 2

(二) 可靠性 2

(三) 保密性 2

(四) 便利性 2

第二节 数字视音频信源码技术及其发展 2

一、概述 2

(一) 信源数字化技术基础 2

(二) 信源高效编码技术基础 5

二、视频数据压缩编码技术基础 5

(一) 视频数据存在的信息冗余为

压缩编码提供了可能 5

(二) 视频数据压缩的基本方法 6

三、音频数据压缩编码技术基础 6

(一) 音频数据存在的信息冗余为

压缩编码提供了可能 6

(二) 音频数据压缩的基本方法 7

四、数字视音频编码技术标准 7

(一) MPEG 7

(二) 多媒体通信视音频压缩编

技术标准 13

第三节 数字视音频信道码技术及其发展 13

一、纠错编码 14

(一) 概述 14

(二) 数字视音频传输/存储系统中

常用的纠错码 14

二、调制 15

(一) 概述 15

(二) 脉冲调制方式 16

(三) PCM 信号的再调制 16

三、编码与调制的结合——格形编码调

制 (TCM) 17

(一) 格形编码调制 17

(二) 格形编码调制的应用 17

第四节 数字视音频存储技术及其

发展 17

一、磁存储技术 17

(一) 磁存储技术的发展 17

(二) 常见的设备形式 20

二、光盘存储技术 21

(一) 光盘存储系统 21

(二) 常见的设备形式 24

第五节 数字视音频传输技术及其

发展 26

一、概述 26

(一) 数字信号传输的基本方式 26

(二) 同步技术 29

(三) 信道共享技术 31

二、网络传输技术 32

(一) 网络技术的基本概念 32

(二) 数字视音频网络传输技术 35

第二章 消费类电子领域的实用数字

视音频技术 41

第一节 数字音频光盘 41

一、数字音频光盘 (CD-DA) 41

二、数字音频光盘 (MD) 43

三、高新格式数字音频光盘 (DVD-

Audio) 46

四、高新格式数字音频光盘 (超级音

频 CD) 47

第二节 数字视盘机 48

一、VCD 与超级 VCD 机 48

(一) VCD、超级 VCD 信号的记录

格式 49

(二) VCD、超级 VCD 机的基本结构 和功能	52	(一) 数字电视演播室的设备配置 及其发展	101
二、DVD 机	54	(二) 数字电视演播室技术参数	103
(一) DVD 信号记录格式	54	(三) 数字电视演播室配置	105
(二) DVD 机的基本结构和功能	59	二、数字摄像机	107
第三节 数字磁带录像机	62	三、数字磁带录像机	107
一、DV 格式	62	(一) D-1 格式	107
(一) DV 信号记录格式	63	(二) D-2 格式	108
(二) DV 录像机的基本结构和原理	70	(三) D-3 格式	108
二、DVCPRO 格式	73	(四) 数字 Betacam 格式	110
三、DVCAM 格式	73	(五) Betacam SX 格式	112
第三章 广播电视领域的实用数字 视音频技术	75	(六) Digital-S 格式	115
第一节 数字电视与 HDTV	75	四、数字视频特技机	115
一、数字电视发展概况	75	五、非线性编辑系统	116
(一) 日本 (从 MUSE 到 ISDB)	75	(一) 非线性编辑技术概况	116
(二) 欧洲 (从 MAC 到 DVB)	76	(二) 非线性编辑系统的硬件体系 结构	118
(三) 美国 (从六雄争霸到 GA HDTV)	78	(三) 非线性编辑系统的软件体系 结构	120
(四) 我国 HDTV 发展概况	78	第四章 交互式电视 (ITV) 与点播 电视 (VOD)	123
二、美国的 ATSC 数字电视标准	79	第一节 ITV 与 VOD 技术发展概况	123
(一) 系统概况	79	一、ITV 与 VOD	123
(二) 视频子系统	80	二、VOD 系统的体系结构	123
(三) 音频子系统	82	(一) 对系统的要求	123
(四) 业务复用与传输子系统	82	(二) VOD 系统的常见体系结构	124
(五) 射频/传输子系统	82	第二节 视频服务器技术	126
(六) 接收机	83	一、视频服务器技术概况	126
三、欧洲的 DVB 数字电视标准	83	(一) 视频服务器的基本类型	126
(一) DVB 数字电视标准	83	(二) 视频服务器的技术特点	126
(二) DVB-S	85	二、基于通用计算机的视频服务器	128
(三) DVB-C	87	三、基于高级工作站的视频服务器	128
(四) DVB-T	87	四、基于专用硬件平台的视频服务器	129
四、我国 HDTV 的技术方案	88	五、分布式层次结构视频服务器	130
(一) 信源码方案	88	(一) 分布式视频服务器	130
(二) 传输方案	89	(二) 分布式层次结构视频服务器	131
五、数字电视的接收技术	95	第三节 用户接入设备技术	131
(一) 数字卫星电视接收机的基本 结构	95	一、用户接入设备技术概况	131
(二) 服务信息	97	(一) 用户接入设备的基本类型	131
(三) 机顶盒	101	(二) 用户接入设备的技术特点	133
第二节 数字电视演播设备	101	二、常见的用户接入设备方案	137
一、数字电视演播室技术概况	101		

(一) 摩托罗拉的机顶盒方案	137	(三) MIDI 接口	187
(二) ST 的数字机顶盒方案	139	(四) MIDI 消息	187
(三) 开放网络机顶盒方案	139	(五) MIDI 合成器原理	188
第四节 VOD 组网技术	140	三、MIDI 系统	189
一、VOD 组网技术概况	140	(一) 基于 MIDI 设备的 MIDI 系统	189
(一) 高速网络的技术特点	140	(二) 基于 PC 平台的 MIDI 系统	190
(二) SDH 网的技术特点	143	第四节 桌面数字视音频制作	191
(三) ATM 网的技术特点	145	一、桌面数字视音频制作技术及其发	
(四) HFC 网的技术特点	149	展	191
二、VOD 组网的主要技术问题	149	(一) 桌面数字视音频技术发展概	
(一) 传输网络的总体结构	149	况	191
(二) ATM 交换机	149	(二) 多媒体计算机系统数字	
(三) HFC 系统中上、下行信道频道		视频技术标准	192
的划分及回传信道协议	150	二、桌面数字视音频制作技术	193
第五节 实用 VOD 系统	153	(一) 基于视频捕捉卡的桌面制作	
一、时代-华纳的 VOD 方案	153	技术	193
(一) 时代-华纳的全业务网	153	(二) 基于 1394 卡的桌面 DV 制作	
(二) 时代-华纳的柏伽索斯方案	161	技术	197
二、欧洲 DVB 标准的 ITV 模型	166	三、基于通用数字视音频工具的数字	
第五章 多媒体计算机领域的实用数字		视音频制作技术	200
视音频技术	168	(一) 常见通用数字视音频工具	200
第一节 概述	168	(二) 常见通用数字视音频工具的	
一、多媒体计算机	168	功能和使用方法简介	203
二、MPC 系统	168	第六章 多媒体通信领域的实用数字	
(一) MPC	168	视音频技术	210
(二) MPC 技术标准	168	第一节 会议电视与可视电话	210
第二节 MPC 系统的数字视音频		一、会议电视与可视电话技术概况	210
技术	170	(一) 会议电视与可视电话技术发展	
一、常用视音频文件格式	170	概况	210
(一) 常用视频文件格式	170	(二) 数字会议电视系统	213
(二) 常用音频文件格式	172	二、H.320 会议电视系统	221
二、常见视音频卡	173	(一) H.320 会议电视技术标准	221
(一) 常见视频卡	173	(二) H.320 会议电视系统的典型	
(二) 常见音频卡	176	结构	222
第三节 MIDI 技术	184	三、H.323 会议电视系统	222
一、MIDI 技术概况	184	(一) H.323 会议电视技术标准	222
(一) 关于乐音的基本概念	185	(二) H.323 会议电视系统的典型	
(二) 概况	185	结构	223
二、MIDI 技术基础	186	四、H.324 可视电话系统	223
(一) 有关 MIDI 的若干术语	186	五、会议电视新技术	224
(二) MIDI 技术规范	186	(一) 基于 ATM 的会议电视系统	224
		(二) 基于 IP 的会议电视系统	225

第二节 流媒体技术及其应用	226
一、流媒体技术概况	226
(一) 流媒体技术的产生与发展	226
(二) 流媒体技术应用概况	227
二、流媒体技术原理	227
(一) 流媒体系统	227
(二) 三大主流流媒体技术解决方案	229

三、典型流媒体技术应用系统	233
(一) 系统结构与组成	233
(二) 系统工作原理	235
(三) 系统特点	236

附录 英文缩略语含义英汉对照表	238
-----------------------	-----

参考文献	244
------------	-----

第一章 数字视音频技术基础

第一节 数字视音频应用系统技术概况

一、系统模型

从信息运动过程的本质来看，任何信息系统中所发生的信息运动都不外乎涉及以下三类物理过程：① 将信息从“这里”传输到“那里”的信息通信过程；② 将信息从“现在”传输到“未来”的信息存储过程；③ 将信息从“这样”变为“那样”的信息处理过程。

我们把系统中发生的信息运动以第①类过程为主的信息系统称为通信系统；以第②类过程为主的信息系统称为存储系统；以第③类过程为主的信息系统称为信息处理系统。

考虑到对第①、②类系统的信息运动过程都是以信息的“传输”为本质特征。因此，我们可将这两类系统称为信息“传输系统”。数字传输系统模型如图 1-1 所示。其中的信源编码和信源解码统称为信源编/解码；信道编码和信道解码统称为信道编/解码，格式编码和格式解码统称为格式编/解码。其中，信源编码主要解决有效性问题，通过对信源的压缩、扰乱、加密等一系列处理，力求用最少的数码传递最大的信息量，使信号更适宜传输、存储；信道编码主要解决可靠性问题，即尽量使处理的信号在传输/存储过程中不出错或少出错，即使出了错也要能自动检错和尽量纠错。信道编码通常包括调制和纠错编码，前者主要用于解决码间干扰产生的错误，后者主要用于解决“噪声”引起的突发性错误（如光盘刮伤、污迹等）；格式编码主要解决运用的高效性问题，即通过对所存储/传输数据的组织达到提高数据存取速度的目的。

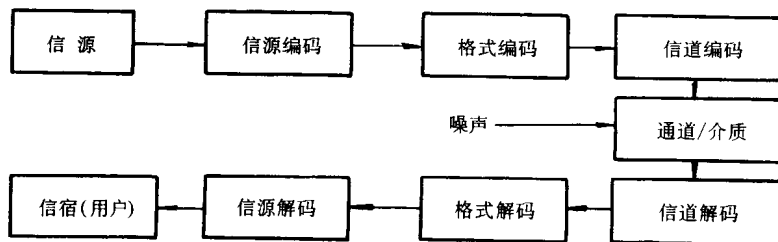


图 1-1 数字传输系统模型

数字信息传输系统的主要优点有：抗干扰能力强，特别是在中继传输时更为明显；可以进行差错控制，因而提高了信息传输存储的可靠性；便于使用现代计算机技术，对信号进行处理、存储和变换，从而提高了信息传输/存储的灵活性；便于加密，对所传输/存储的信息进行保密或版权保护；易于与其它系统配合使用，构成一个灵活、通用、多功能的综合业务信息传输网；易于集成化和大规模生产，其性能一致性好，且成本低。

11/11/2018



二、系统性能

信息传输的根本任务是传输信息,人们总是希望被传输的信息愈多愈好,失真愈小愈好,安全性愈高愈好,使用愈方便愈好。因此,衡量一个信息传输/存储系统优劣程度的最重要的性能指标有有效性、可靠性、保密性和便利性。

(一) 有效性

数字传输/存储系统的有效性是指在给定信道内能够传输的信息的多少,通常用码元传输速率或信息传输速率来进行量度。

码元传输速率是指单位时间(通常为s,下同)内传输的码元数目,记为 R_b ,其单位为波特(B)。码元传输速率又称为传码率或波特率。需要注意的是:这里所说的码元并没有限定它取何种进制。信息传输速率是指单位时间内传输的信息量,记为 R_b ,其单位为比特/秒(bit/s)。信息传输速率又称为传信率、码率或比特率。

(二) 可靠性

数字传输/存储系统的可靠性是指信宿所接收到的信息的准确程度,通常用码元差错概率或信息差错概率来进行度量。

码元差错概率是指所传输的码元总数目中发生差错的码元数目所占的比值(取统计平均值,下同),记为 P_e 。码元差错概率简称误码率。信息差错概率是指所传输的信息比特总数目中发生差错的信息比特数目所占的比值,记为 P_b 。信息差错概率又称为比特差错概率或误比特率。在二进制情况下,码元差错概率和信息差错概率在数值上相等,即 $P_e = P_b$,但含义不同。

(三) 保密性

数字传输/存储系统的保密性是指所传输/存储系统通过条件接收技术,以防范在系统中传输/存储信息被非授权信宿终端所接收、正确解码和利用的有效程度。

(四) 便利性

数字传输/存储系统的便利性是指信宿终端用户按需从传输信道/存储介质中检索出目标信息的简捷程度。

第二节 数字视音频信源码技术及其发展

一、概述

信源码应包括数字化编码和高效编码两部分。

(一) 信源数字化技术基础

PCM是“数字化”的最基本的技术,模拟信号的数字化如图1-2所示。模拟信号正是通过PCM而变换成数字信号的。PCM是Pulse Code Modulation(意为脉冲编码调制,简称脉码调制)的缩略语,其具体操作:通过取样、量化和编码三个步骤,用若干代码表示模拟形式的信息信号(如图像、声音信号),再用脉冲信号表示这些代码来进行传输/存储。

所谓“代码”是指表示数值的一组二进制或多进制的数字符号,例如,表示数值“六”

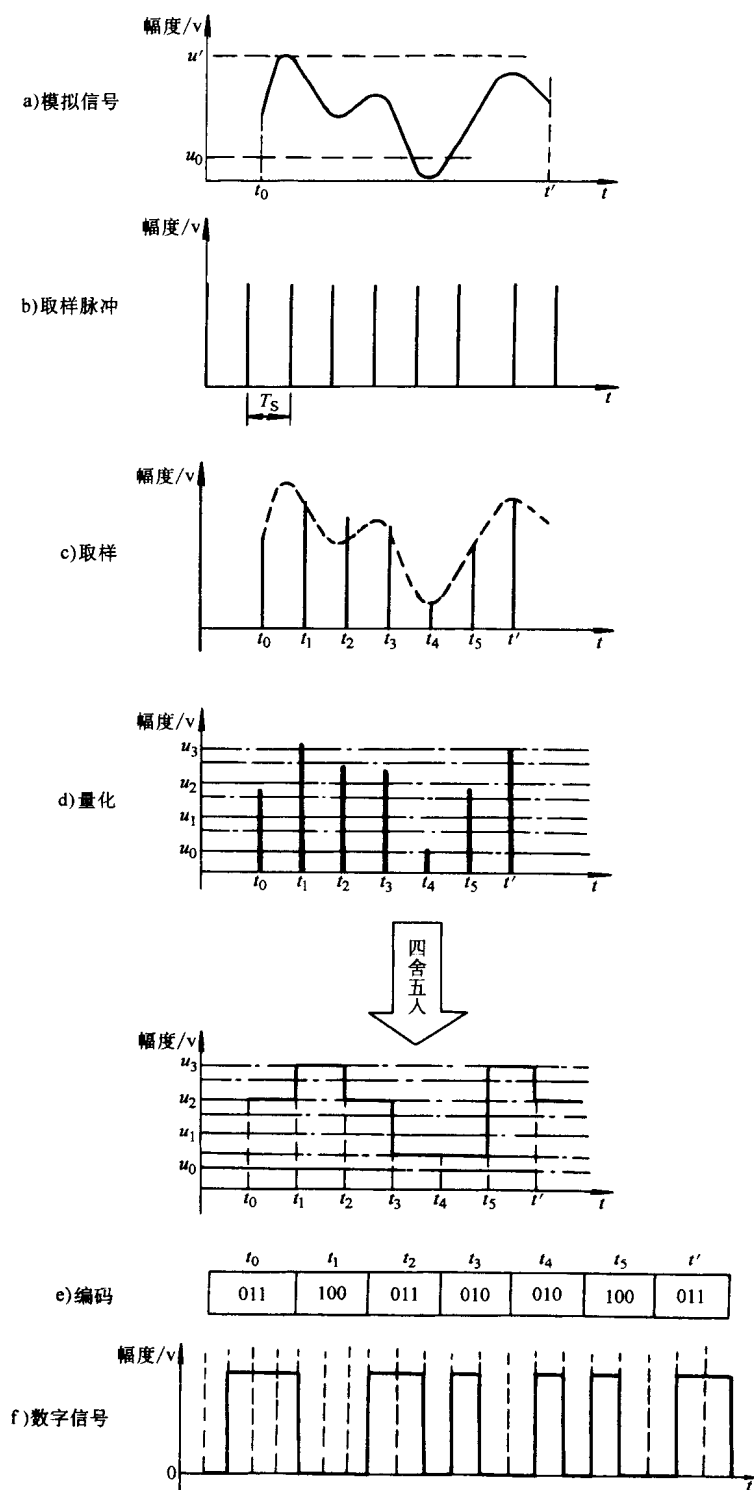


图 1-2 模拟信号的数字化

的十进制代码是“6”，二进制代码是“110”。PCM 技术中，通常用二进制代码。

这些抽象的代码是怎样表示一个信息信号的呢？我们知道，图 1-2a 所示的模拟式信息信号的基本特征是连续性。因为它是连续的，所以在它出现的时域 $t_0 \sim t'$ 内，任何一个时刻 t 都对应一个信号幅值 $u(t)$ （唯一值）。若用一个代码表示一个信号幅值，则因为在 $t_0 \sim t'$ 的有限时段内存在无数个时刻而需无数个代码才能将原信号表示出来；而一方面，因为原信号是连续的，所以其幅值是变化的动态范围 $u_0 \sim u'$ 内的（用 A 表示动态范围，即 $A = u' \sim u_0$ ）的任一实数值。而即使在 $u_0 \sim u'$ 这一有限区间内，也存在无数个不同的实数值。若用一个 n 位代码表示一个信号幅值，则需要 $n \rightarrow \infty$ 位代码才能将这些实数值表示出来。显然，沿这一思路来“用代码表示模拟信号”在技术上是不可行的，因为任何技术都无法在有限的时间内处理无数个代码，每个代码又是无穷位。

为解决这两个“无穷”的问题。PCM 技术采用了“取样”和“量化”两项措施。

取样就是在信息信号出现的时域 $t_0 \sim t'$ 内，用间隔为 T_s 的 $t_0, t_1, t_2, t_3, t_4, t_5, t'$ 等 7 个时刻所对应的信号幅值，近似地代表原信号在 $t_0 \sim t'$ 内的无数个幅值。其具体实现方式：用原信号对周期为 T_s （即频率为 f_s ）、脉宽 $\Delta \rightarrow 0$ 的脉冲序列（见图 1-2b，称为取样脉冲，其频率 f_s 称为取样频率）进行幅度调制，结果如图 1-2c 所示。因为取样脉宽无限窄，故调制后每一脉冲的幅值等于其出现时刻所对应的原信号幅值。我们把这些用于代表原信号无数个幅值的有限个幅值称为取样值，简称样值。

图 1-2c 所标出的各样值是实测值，它们是实数，这就势必存在前述的第二个“无穷”的问题。为此，PCM 技术采取了第二个措施——量化。其具体操作：将信号幅值变化的动态范围 A ，人为地划分为若干等级 u_i （ $i=0, 1, 2, \dots, n$ ），图 1-2d 中划分为 $u_0 = 1V, u_1 = 2V, u_2 = 3V, u_3 = 4V$ 等 4 个等级电平；然后，用“四舍五入”的方式将各样值（有无限个可能值的实数）变换成有限的 n 个（上例是 4 个）量化等级电平值，这样就可以用有限位代码完全表示这有限个等级电平值。我们将相邻两个量化等级电平的差值称为量化步长 ΔA （上例的 $\Delta A = 1V$ ，当然也可以选择任意值的电平差，如 $0.1V$ 或 $0.5V$ 等）。

编码就是用 N （bit）二进制代码表示各样值的量化等级电平值，在图 1-2c 中，可用 3bit 二进制代码就完全可以穷尽表示 $1V, 2V, 3V$ 和 $4V$ 等 4 个量化电平值，如图 1-2d 所示。将代码的位数 N 称为量化位数或量化比特数。

显然，原信号的动态范围 A 一定时，量化步长 ΔA 愈小，则量化等级数 n 愈大，且 $n = A/\Delta A$ ；量化等级数 n 愈大，所需的量化比特数 N 愈大，且 $n = 2^N$ 。

就这样，PCM 技术通过取样和量化，实现了用有限个代码，每个代码只用有限位就将原来具有无限个信号幅值的连续模拟信号近似地表示出来了，如图 1-2e 所示。

但这些抽象的数字符号代码还不能供机器处理、传输和存储，它们必须转换成物理信号形式，通常 PCM 技术用低电平代表“0”，高电平代表“1”的脉冲信号表示这些代码（见图 1-2f）。这些脉冲信号称为脉码调制（PCM）信号，它就是数字信号。众所周知，这种信号形式正是现代电子计算机能够直接识别和处理的。正因为数字化的信息能直接由电子计算机处理，所以人们可以在各种信息系统中引入电子计算机，从而使信息系统的各种信息处理自动化、智能化成为可能。

(二) 信源高效编码技术基础

在数字视音频存储/传输系统中, 为了达到令人满意的视听效果, 必须解决视音频数据的大容量存储和实时传输问题。但数字化的视音频数据往往数据量巨大、传码率高, 这就对媒体的存储容量和信道带宽提出了高要求。可幸的是, 我们习惯上称为“数据”的仅仅是代表样值的 N (bit) 二进制代码, 它与信息不完全是一回事。而决定视听效果的只是信息, 并非数据。信息是对发生事件的抽象描述, 而数据是在确定了描述方法后对事件的具体描述记录。显然, 对同一个信息, 若使用不同的描述方法, 则形成记录的数据量可能完全不同。因此, 使用更高效的编码方式用尽可能少的数据来表示信息, 从而在保证一定视听效果的前提下, 节省传输和存储的开销。同时也使计算机实时处理和播放视音频信息成为可能。通常, 可用压缩比、图像和声音质量、编/解码速度等指标来衡量某种压缩编码方法的优劣, 除此以外, 还要考虑压缩算法所需要的软、硬件开销。

二、视频数据压缩编码技术基础

(一) 视频数据存在的信息冗余为压缩编码提供了可能

视频信号是由一系列在时间上分立的图像序列构成的动态图像信号。研究表明, 视频数据表示中存在着大量的冗余, 即图像的取样点之间在亮度和色度信息方面存在极强的相关性。利用这些相关性, 一部分取样点的参数可以由另一部分取样点的参数推导出来, 从而使原始的视频数据量极大地减少, 有利于传输/存储。视频数据主要存在以下形式的冗余:

1. 空间冗余 视频图像序列中的某帧画面的某一取样点的亮度和色度信息, 与其在同一帧画面的相邻取样点之间所存在极强的相关性。

2. 时间冗余 视频图像序列中的某帧图像的某一取样点的亮度和色度信息, 与其在时间轴上相邻帧的相应位置上的取样点存在极强的相关性。

3. 结构冗余 在有些图像的纹理区, 图像的像素值存在着明显的分布模式。

4. 知识冗余 有些图像的理解与某些知识有相当大的相关性。

5. 视觉冗余 实验发现人眼具有以下视觉非均匀特性: ① 对图像的亮度敏感, 而对色度的敏感性远不如前者。② 随着亮度的增加, 对量化误差的敏感度降低。这是由于人眼的辨别能力与物体周围的背景亮度成反比。③ 把图像的边缘和非边缘区域分开处理。这是将图像分成非边缘区域和边缘区域, 分别进行编码的主要依据。这里的边缘是指灰度值发生剧烈变化的地方, 非边缘区域是指除边缘之外的图像其它任何部分。④ 把视网膜上的图像分解成若干个空间有向的频率通道后再进一步处理。在编码时, 若把图像分解成符合这一视觉内在特性的频率通道, 则可能获得较大的压缩比。在记录原始的图像数据时, 通常假定视觉系统是线性的和均匀的, 对视觉敏感和不敏感的部分同等对待, 从而产生了比理想编码 (即把视觉敏感和不敏感的部分区分开来编码) 更多的数据, 这就是视觉冗余。后面提到的小波编码就是在一定程度上利用了这一特性。

6. 图像区域的相同性冗余 在图像中的两个或多个区域所对应的所有像素值相同或相近, 从而产生的数据重复性存储, 这就是图像区域的相同性冗余。在以上的情况下, 记录了一个区域中各像素的颜色值, 则与其相同或相近的其它区域就不再需要记录其中各像素的值。后面提到的矢量量化方法就是针对这种冗余性的图像压缩解码方法。

7. 纹理的统计冗余 有些图像纹理尽管不严格服从某一分布规律, 但是它在统计的意

义上服从该规律。利用这种性质也可以减少表示图像的数据量，所以我们称之为纹理的统计冗余。

(二) 视频数据压缩的基本方法

视频数据压缩方法根据不同的依据可产生不同的分类。最常见的是，根据质量有无损失可分为有失真压缩编码和无失真压缩编码，压缩编码的基本方法如图 1-3 所示。按照压缩冗余信息的机理不同，目前的图像压缩编码方法大致可以分成三类：一是，着眼于图像信源的统计特征的压缩方式，主要采用统计编码法、预测编码法、变换编码法、矢量量化编码法、神经网络编码法等；二是，着眼于人眼视觉特性的压缩方式，主要采用基于方向滤波的图像编码法、基于图像轮廓-纹理的编码法；三是，着眼于图像传递的景物（内容）特征的压缩方式，主要采用基于模型的（分形）编码法。

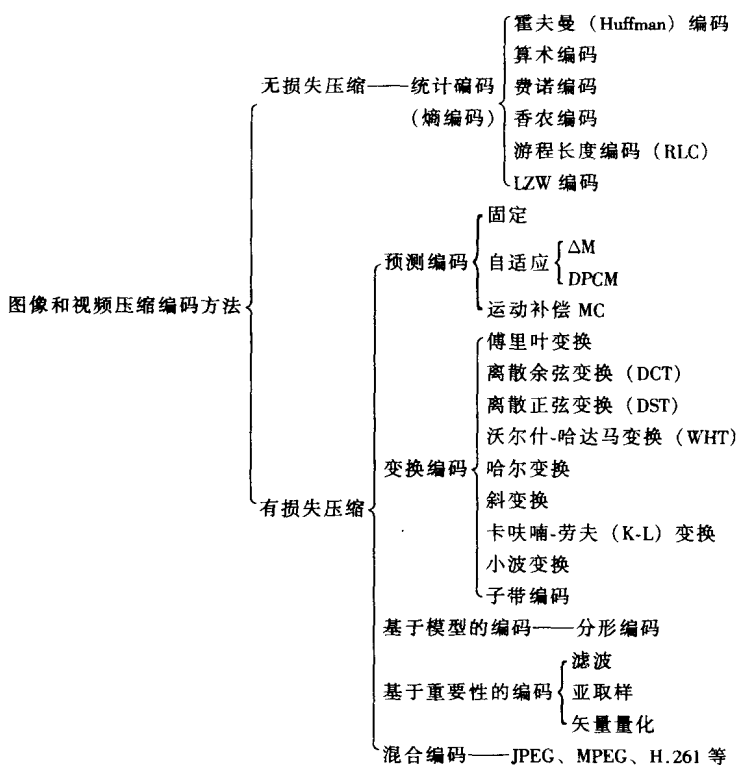


图 1-3 压缩编码的基本方法

此外，还可按照其作用域在空间域或频率域上分为：空间方法、变换方法和混合方法，或根据是否自适应分为自适应性编码和非适应性编码。一般来说，每一个编码方法都有其相应的自适应方法。

三、音频数据压缩编码技术基础

(一) 音频数据存在的信息冗余为压缩编码提供了可能

1. 音频信号 声音按其频率可分可听声、次声和超声三种。其中：频率低于 20Hz 的声波为次声，频率高于 20kHz 的声波为超声频率在 20Hz ~ 20kHz 之间的声波为可听声。前两者

是人耳听不见的。

音频信息主要是指可听声,通常将其分为:①语音(200Hz~3.4kHz):由口腔发出的声波;②音乐声:是由各种乐器产生的,音频范围内都可以存在;③效果声:大自然物理现象产生的,如风声、雨声、雷声等;还有一些人工产生的,如爆破声等。

2. 音频信号的冗余 音频信号主要有如下几方面冗余:

(1) 时域冗余 音频信号在时域上的冗余主要表现为:①幅度分布的非均匀性;②样值间的相关性;③周期之间的相关性;④基音之间的相关性;⑤静止系数;⑥长时自相关函数。

(2) 频域冗余 音频信号在频域的冗余主要表现为:①长时功率谱密度的非均匀性;②语音特有的短时功率谱密度。

(3) 听觉冗余 音频信号的最终用户是人,因此,要充分利用人类听觉的生理-心理特性对于音频信号感知的影响,以免做“即使记录了,人耳也听不见”的无用功。

(二) 音频数据压缩的基本方法

在音频数据的存储或传输中,数据压缩是必需的。通常数据压缩造成音频质量的下降、计算量的增加。因此,人们在实施数据压缩时,要在音频质量、数据量、计算复杂度三方面进行综合考虑。目前,常用的音频压缩编码主要有如下几类:

1. 基于音频数据统计特性进行数据压缩的编码方法 波形编码是基于音频数据的统计特性进行编码的典型技术,其目标是使重建音频波形保持原波形的形状。PCM(脉冲编码调制)是最简单最基本的编码方法。它直接赋予样值一个代码,没有进行压缩,因而所需的存储空间较大。为了减少存储空间,人们寻求压缩编码技术。利用音频样值的幅度分布规律和相邻样值具有相关性的特点,提出了差值量化(DPCM)、自适应脉码调制(APCM)和自适应差分脉码调制(ADPCM)等算法,实现了数据的压缩。波形编码适应性强,音频质量好,但压缩比不大,因而数据的码率较高。

2. 基于音频的声学参数进行数据压缩的编码方法 基于音频的声学参数进行参数编码,可进一步降低数据率。其目标是使重建音频保持原音频的特性。常用的音频参数有共振峰、线性预测系数、滤波器组等。这种编码技术的优点是数据的码率低,但重建音频信号的质量较差,自然度低。

3. 混合编码方法 将上述两种编码算法很好地结合起来,采用混合编码的方法。这样就能在较低的码率上得到较高的音质。如代码激励线性预测(CELP)编码、多脉冲激励线性预测编码(MPLPC)等。

4. 基于听觉特性进行数据压缩的编码方法 基于人的听觉特性进行编码主要是从人的听觉系统出发,利用掩蔽效应,设计心理声学模型,从而实现更高效率的数字音频的压缩。其中以国际活动画面专家组(MPEG)标准中的高频编码和Dolby AC-3最有影响。

四、数字视音频编码技术标准

当前压缩编码技术发展的一个重要方向就是综合现有的压缩编码技术,制定全球的统一标准,使信息管理系统具有普遍的互操作性并确保了未来的兼容性。

(一) MPEG

MPEG是Moving Picture Expert Group(运动图像专家组)的缩略语,该专家组成立于1988

年,在国际标准化组织(ISO)和国际电工委员会(IEC)辖下,有约300名专家分11个小组进行工作,现已公布和将要公布的标准有:

MPEG-1 (ISO/IEC 11172):1993年8月公布,其全称:《适于约1.5Mbit/s以下数字存储媒体的运动图像及伴音的编码》。所谓数字存储媒体(DSM)是指常见的数字存储设备,如CD-ROM、DAT、硬盘、可写光盘等。该标准也适于远程通信,如综合业务数字网(ISDN)、局域网(LAN)等。

MPEG-2 (ISO/IEC 13818):1994年11月公布,其全称:《运动图像及其伴音通用编码》,适用于多媒体计算机、多媒体数据库、多媒体通信、常规电视数字化、高清晰度电视(HDTV)及交互式电视(ITV)等领域。

MPEG-4 (ISO/IEC 14496):1999年1月公布版本1(V1.0),同年12月公布版本2(V2.0),该标准的初衷主要是面向会议电视、可视电话等超低码率的压缩编码需求,在制订过程中,MPEG组织深深感受到人们对媒体信息,特别是对视频信息的需求由播放型转向基于内容的访问、检索和操作。MPEG-4与JPEG、MPEG-1和MPEG-2等有很大差异,它为多媒体数据压缩编码提供了更为广阔的平台,它定义的是一种格式、一种框架,而非具体算法,它希望建立一种更自由的通信与开发环境。于是MPEG-4的新目标就锁定为:支持多种多媒体的应用,特别是多媒体信息基于内容的访问和检索,可根据不同的应用需求,现场配置解码器。编码系统也是开放的,可随时加入新的有效的算法模块。该标准适用于多媒体因特网、视频会议和视频电话等个人通信,交互式视频游戏和多媒体邮件、基于网络的数据服务、光盘等交互式存储媒体、远程紧急事件系统、远程的视频监视及无线多媒体通信等。

MPEG-7:全称为《多媒体内容描述接口》,它将确定各种类型的多媒体信息标准的描述方法。

2000年3月,MPEG会议成立MPEG-21工作组,进行新一代MPEG标准的研究开发。

1. MPEG-1 MPEG-1(ISO/IEC 11172)标准由系统、视频和音频等三部分构成。

(1) MPEG-1系统(ISO/IEC 11172-1) 这部分是关于同步和多路复用的技术,目标是要将经压缩编码的图像和伴音数据复合成数据传输率在1.5Mbit/s以下的单一数据位流。MPEG-1编/解码器模型如图1-4所示。

MPEG-1编码系统对视/音频PCM位流的编码,包括压缩编码和系统编码两个层次:压缩层编码就是在MPEG-1视/音频编码器中,执行MPEG-1视/音频所描述算法,对视/音频PCM位流进行压缩编码;系统编码就是将经压缩编码的MPEG-1视/音频位流,分成若干视/音频单元并按包(pack)和小包(packet)两层次进行的格式编码(称“打包”)。

(2) MPEG-1视频(ISO/IEC 11172-2) 这部分是关于电视图像的压缩技术,目标是要把分辨率为 352×240 ,帧频为30Hz的电视图像压缩成数据传输率为1.2Mbit/s的编码图像。所推荐的MPEG算法采用了两种基本压缩技术:为减少时间冗余,采用 16×16 个像素组成的图像块的运动补偿;为减少空间冗余,采用 8×8 图像块的DCT技术。MPEG算法必须满足:随机存取、快速搜索、逆向播放、编辑功能、编码与解码延时、视音频同步、容错能力以及格式的灵活性等。

为获得高压缩率而又不使图像质量太差,同时还要支持上面提到的几个特性,因此MPEG-1采用了以下三类压缩技术:

1) 频带压缩技术 利用人的视觉系统对亮度敏感而对色度不敏感的特性,对Y像素值

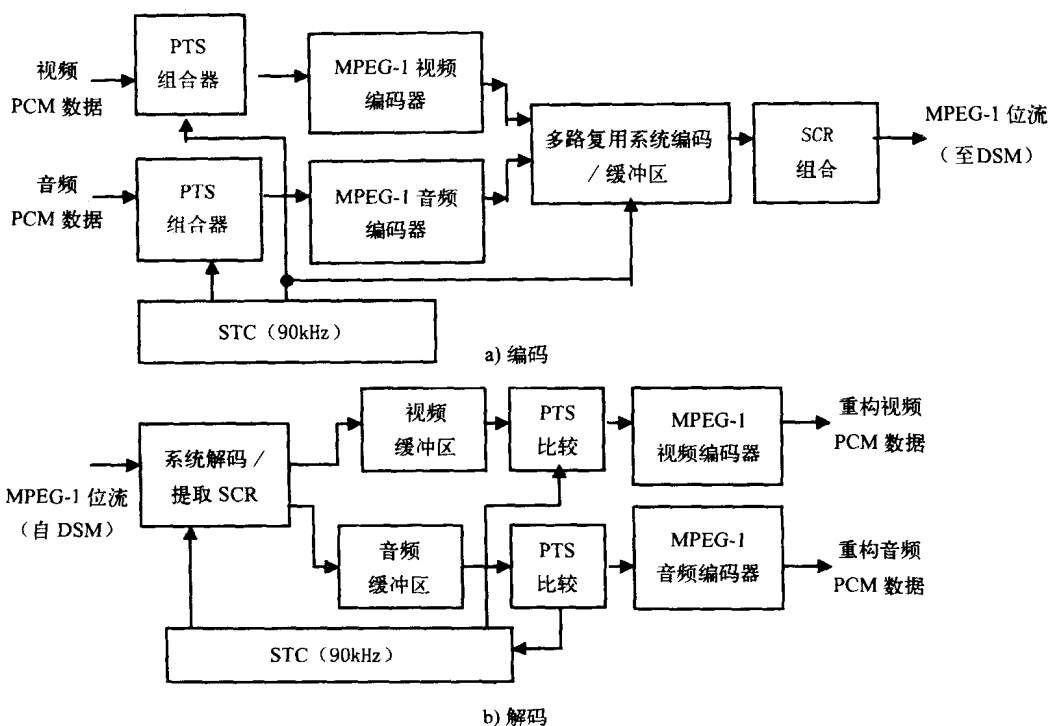


图 1-4 MPEG-1 编/解码器模型

以全分辨率进行取样，而 C_R 、 C_B 像素值进行亚取样，即采用类似于 4:2:2 标准，从而降低编码图像的数量或数据传输率（即频带）。

2) 有损压缩技术 采用有损压缩算法以消除视频信号的时间冗余和空间冗余。主要采用运动估值与补偿、帧内插、DCT、自适应量化以及 Zig-Zag 排序技术。

3) 无失真压缩技术 无失真压缩算法采用了基于统计分析的最佳熵编码，主要采用了可变长度编码和游程编码技术。

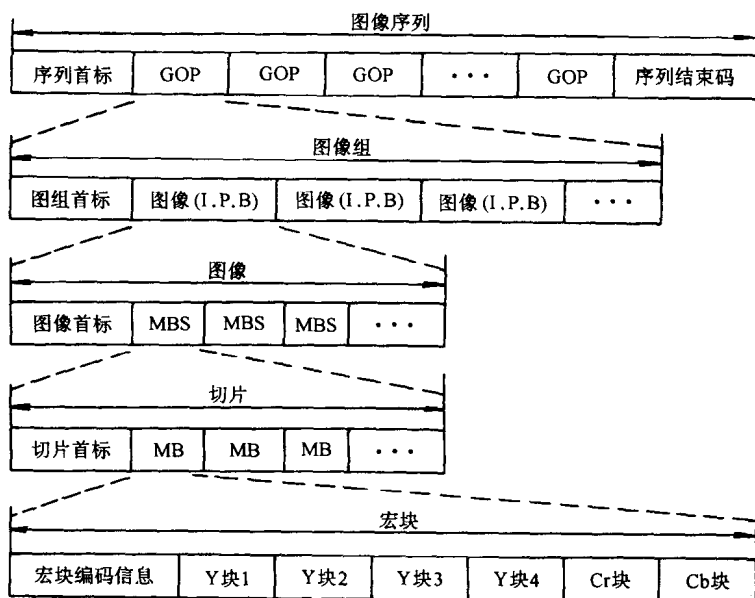


图 1-5 MPEG-1 视频位流的结构

按 MPEG-1 视频规定，MPEG-1 视频编码器输出的编码视频位流的结构如图 1-5 所示，它是由图像序列、图像组 (GOP)、图像、切片、宏块和块等 6 个语义层构成的复合结构。其中块是基本编码单元，是最小的 DCT 单位，各层的首标用来唯一地说明后面跟着的数据。