



数字技术：

数字视频和音频压缩

〔美〕 Stephen J. Solari 著

陈河南 吴少波 高胜友 等译

Digital Video and Audio
Compression



电子工业出版社

PUBLISHING HOUSE OF ELECTRONICS INDUSTRY

URL: <http://www.phei.com.cn>

数字技术:数字视频和音频压缩

Digital Video and Audio Compression

[美] Stephen J. Solari 著

陈河南 吴少波 高胜友 等译

电子工业出版社

Publishing House of Electronics Industry

北京·BEIJING

Original edition Copyright © 1997 by The McGraw-Hill Companies, Inc.

All rights reserved.

Chinese translation edition Copyright © 2000 by Publishing House of Electronics Industry.

All rights reserved.

本书中文简体专有翻译出版权由美国 McGraw-Hill, Inc. 授予电子工业出版社。该专有出版权受法律保护。

图书在版编目(CIP)数据

数字技术:数字视频与音频压缩/(美)索拉里(Solari, S.J.)著;陈河南等译. - 北京:电子工业出版社, 2000.1

书名原文: Digital Video and Audio Compression

ISBN 7-5053-5694-1

I. 数… II. ①索… ②陈… III. ①数字信号: 视频信号-频率压缩 ②语音数据处理-频率压缩
IV. TN941.3

中国版本图书馆 CIP 数据核字(1999)第 67388 号

书 名: 数字技术: 数字视频和音频压缩

原 书 名: Digital Video and Audio Compression

著 者: [美] Stephen J. Solari

译 者: 陈河南 吴少波 高胜友 等

策 划: 电子工业出版社外版书编辑部

责任编辑: 窦 昊

特约编辑: 胡万三

排版制作: 电子工业出版社计算机排版室监制

印 刷 者: 北京天竺颖华印刷厂

出版发行: 电子工业出版社 URL: <http://www.phei.com.cn>

北京市海淀区万寿路 173 信箱 邮编 100036

经 销: 各地新华书店

开 本: 787×1092 1/16 印张: 13.5 字数: 337 千字

版 次: 2000 年 1 月第 1 版 2000 年 3 月第 2 次印刷

书 号: ISBN 7-5053-5694-1
TP·2931

印 数: 2000 册 定价: 25.00 元

版权贸易合同登记号 图字: 01-1999-3359

凡购买电子工业出版社的图书, 如有缺页、倒页、脱页、所附磁盘或光盘有问题者, 请向购买书店调换。
若书店售缺, 请与本社发行部联系调换。电话 68279077

11/16/08

译者的话

信息时代社会的飞速发展,改变了人们的生活方式。信息高速公路、电子邮件、因特网、网络电话、视频点播……给我们的生活带来了极大的方便,但新技术让我们感觉到压力,信息更新太快,知识更新太快,甚至有大学学历的知识分子若不更新知识也会被信息社会关在门外。

数字信号技术是信息技术的基础,从模拟信号产品到数字信号产品的转变过程标志信息革命的进行过程。数字信号的高品质、高容量让人们惊喜和狂热。本书是让你狂热的时候清醒,看看是什么具有这么大的魔力。

本书主要研究声音、图片、动画等具有大数据容量的信号,它们的压缩、传输、噪音的理论和前沿处理技术。本书的内容既是理论,又是资料。作者从压缩和解压缩技术、传输和噪音、声音、听觉和视觉原理讲到这些技术的应用,全书结构严密,自成体系,是音频和视频技术的很好的理论教材。书中涉及的大量材料来自世界上各领域最高水平的专家在专著及国际会议宣读的论文,具有权威性,这些资料在书中有注明。读者自己若想查这么多资料,恐怕得花几个月的时间。

这本书由清华大学几位相关专业的教师翻译。第1章由高胜友老师翻译,陈河南老师审阅,第2章到第5章由王雷老师翻译,贺军老师审阅,第6章到最后由吴少波老师翻译,崔新宇老师审阅。全书最后由龚亚萍老师统稿。潇湘工作室的蒋方、李伟、李争胜等人也做了大量的工作。

序 言

是什么这样激动人心？为什么电话公司涉足娱乐界？为什么有线电视产业期望对基础设施进行大量的投资？为什么其他领域的老总们在喋喋不休地谈论即将来临的通信技术焦点？

当前，人们对数字媒体的狂热也许有些过分，但这很容易解释。其原因在于人们对音频和视频的成本低廉的数字处理技术满怀期望。

模拟信号是信息密集型的，需要高带宽，并且还要维护传输。在模拟领域控制信号（创建、编辑或者操纵视觉或声音）是可能的，但是很麻烦，因而限制了可以和这些媒体交流的人数。另一方面，通过模拟电视传输而提供的被动的娱乐活动已经可以在 98% 的美国家庭中享用。

集成电路复杂程度的不断增长和成本的不断降低，使得对于大多数应用而言，数据压缩的成本已经非常低。

视频和音频已经融入到计算机上的其他种类的数据当中，它们提供了多媒体计算机。视频在电话会议产品中已经和声音相互融合。美国已经为高清晰度电视采用了一种数字系统。广播公司、有线电视公司和电话公司正在相互竞争，目的是为家庭提供新的服务，包括点播视频和交互电视。

只有大公司才可以承受实现这样的系统所必须的资金投资。然而，新的技术有可能改变这个领域。正如计算机已经动摇了公司管理信息系统部门和大型机生产商的地位一样，因特网已经为个人提供了访问各种信息的途径，广泛应用的数字媒体最终可以为个人提供更大的选择和功能。

为帮助评价在我们的生活中引入这些技术的影响，本书介绍了数字音频和视频压缩的基本技术、基础结构、成本和应用。

什么是压缩？即使是对技术了解甚少的读者在打开本书时，也会知道什么是压缩。它就是将某些东西挤压到一起或以强制性的手段减少它的空间，这就像你的废纸篓压缩器对昨天晚上的剩余物所做的事情。这种技巧的另一方面就是解压缩，即按照可以接受的可能损失要求重新建立原始的东西。

压缩和解压缩的原因是可以更好地使用可用的空间或资源。在电子世界中，这意味着更加有效地利用传输带宽或存储介质。压缩可以在代表模拟或数字领域的信号上进行。本书着重介绍今天在数字领域中普遍实现的技术，尽管有些原则对于广泛应用的低廉的数字处理技术有些超前。

使用本书

可以将本书视为那些希望使用数字媒体的人们的初级烹调书。传统的烹调书指导一个人如何做出菜肴以满足味觉和嗅觉的需要，而本书讨论了视觉和声音的成分、处理和表达。

本书和当今许多流行的烹调书不同，它们要依赖于晦涩难懂的食物和工具，本书讲述了基本内容。尽管具有一定的物理或工程背景可以帮助读者理解某些技术内容，但任何有兴趣的读者都可以学习本书，并加深对压缩术语、技巧和技术的理解。

烹调要选择味道，在这个方面，压缩和烹调极为相似。烹调依赖于 Terry 所推行的技术，

在这个方面,压缩和烹调也总是相似的。Terry 是 Fawlty Towers 电视节目的厨师,他说:“眼睛看不见的,厨师就忽略了”。

本书分为 4 个综合的部分:

- 第 1 章,压缩的原因和方式。第 1 章说明了在电子世界中展现我们所看见和听见的真实世界信息的各种不同方法。这部分内容解释了为什么视频和音频压缩是可能的,并介绍了源编码的原则。然后本章介绍了视频和音频压缩在各种不同的候选应用中的要求和特点。
- 第 2 章到第 7 章,视觉。这里首先回顾了人类如何捕获和处理视觉和运动信息,提出了现存的模拟广播标准,然后研究了各种减少与视觉信息相关的数据的技术。无论是由标准化委员会(例如 JPEG, MPEG),还是各个公司(例如 DigiDipher, Indeo)提出的压缩策略,都使用不同压缩技术的组合。例如,正是在这一点上,MPEG-2 可以与其他提议具有相同的成分。
- 第 8 章,声音。第 8 章首先研究了如何听,然后研究了声音的记录、传输、重现和音频压缩方法。
- 第 9 章到第 11 章,综合压缩应用。现在我们已经讨论了各种不同的策略和技术,我们可以了解不同的工业在各种不同的应用和服务中是如何组合这些技术的。

随着技术的迅速进步,紧跟技术的发展对于商业上的成功是非常重要的。本书的目的就是为读者提供当前最热门的领域即视频和音频压缩领域的有价值的信息。

致 谢

感谢 Michael Jacobsen 对本书所做的贡献,感谢 Rolf-dieter Gutschman 对第 4 章的编写所做的贡献,感谢 Leo Warmuth 对讲授和翻译所做的贡献。

特别感谢 Hans Van Weersch 对审阅、编辑和校对本书的主要部分所做的贡献。

Stephen J. Solari

目 录

第 1 章 理论和实践	(1)
1.1 基本原理	(1)
1.1.1 源编码	(1)
1.1.2 冗余和不相关性	(1)
1.1.3 无损和有损源编码	(1)
1.1.4 编码和传输	(2)
1.2 傅立叶变换	(2)
1.3 Z 变换	(3)
1.4 应用要求	(3)
1.4.1 点对点与广播	(3)
1.4.2 实时与离线	(3)
1.4.3 保真度	(3)
1.4.4 质量	(3)
1.4.5 错误的敏感度	(4)
1.4.6 可伸缩性	(4)
1.4.7 专用硬件	(4)
1.4.8 压缩率	(4)
1.4.9 成本	(4)
1.5 数字域中的信号	(5)
1.5.1 合成源	(5)
1.5.2 模拟信号的数字化	(5)
1.6 时间采样	(6)
1.6.1 采样幅值	(8)
1.6.2 脉冲编码调制	(9)
1.6.3 数字处理	(9)
1.6.4 模拟再现的准备	(10)
1.7 人类视觉系统	(10)
1.7.1 图片质量	(11)
1.7.2 调制传输功能	(13)
1.7.3 倾斜效应:水平和垂直感觉	(13)
1.7.4 马赫效应和轮廓感觉	(13)
1.7.5 屏蔽	(14)
1.7.6 移动对象分辨率的损失	(16)
1.7.7 采样运动的插值	(18)
1.7.8 移动和静态图片传输编码	(19)
1.8 参考文献	(21)

第 2 章 视频压缩处理	(22)
2.1 扫描	(22)
2.1.1 同步	(22)
2.1.2 隔行扫描	(22)
2.1.3 分辨率	(23)
2.2 色彩基础	(24)
2.2.1 兼容性	(24)
2.3 颜色传输标准	(25)
2.3.1 NTSC 编码器	(26)
2.3.2 NTSC 解码器	(27)
2.3.3 PAL	(28)
2.3.4 PAL 编码器	(28)
2.3.5 PAL 解码器	(29)
2.3.6 SECAM	(29)
2.3.7 SECAM 编码器	(30)
2.3.8 SECAM 解码器	(31)
2.3.9 3 个主要标准的比较	(31)
2.4 在视频信号数字处理过程中的特殊考虑	(31)
2.4.1 噪声	(32)
2.4.2 特殊滤波技术	(34)
2.4.3 空间滤波	(34)
2.4.4 时间滤波	(34)
 第 3 章 视频压缩技术:编码	(36)
3.1 图片作为消息	(36)
3.2 编解码器参考模型	(36)
3.2.1 第 1 个图片模型和它的编码	(37)
3.2.2 特定源的码字分配	(39)
3.2.3 第 2 个图片模型:源存储器	(41)
3.3 数字图片编码:解相关和量化	(43)
3.3.1 解相关:源失去记忆	(44)
3.3.2 预测编码	(45)
3.3.3 转换编码基础	(46)
3.3.4 解相关信号的量化	(47)
3.3.5 量化器的设计步骤	(48)
3.3.6 非线性量化器	(49)
3.3.7 最大量化器,劳埃德算法	(50)
3.3.8 线性量化器	(51)

3.3.9 量化之前解相关的原因	(51)
3.4 参考文献	(52)
第4章 视频压缩技术:差分脉码调制技术	(54)
4.1 作为简单的解相关方法使用 DPCM	(54)
4.1.1 误差影响的 DPCM	(54)
4.1.2 从预测值中获取的信息	(55)
4.1.3 电视图片的预测器	(56)
4.1.4 内部预测	(58)
4.1.5 自适应预测器	(60)
4.1.6 通道的传输错误	(61)
4.1.7 预测误差的量化器	(62)
4.1.8 自适应量化器	(66)
4.2 图像序列的帧间编码和运动补偿	(66)
4.2.1 DPCM 的预测器	(66)
4.2.2 运动补偿的原因	(67)
4.2.3 带有运动补偿的帧间 DPCM	(67)
4.2.4 通过块匹配估计运动的方法	(67)
4.2.5 使用块匹配技术的运动估计器	(69)
4.2.6 块匹配所需的工作量	(71)
4.2.7 削减计算要求的块匹配技术	(72)
4.2.8 运动估计的其他方法	(74)
4.2.9 干扰对测量的影响	(75)
4.2.10 一般应用的运动向量	(76)
4.3 参考文献	(76)
第5章 视频压缩技术:变换编码	(78)
5.1 变换编码的重要性	(78)
5.1.1 频率域中的编码技术:变换编码	(78)
5.1.2 变换的一般解释	(78)
5.1.3 系数量化和编码优于像素的原因	(81)
5.1.4 哪种变换最好	(83)
5.1.5 平面变换	(84)
5.1.6 系数的量化过程	(87)
5.2 实现考虑	(88)
5.2.1 帧内变换编码	(90)
5.2.2 系数编码	(93)
5.2.3 其他考虑因素	(96)
5.2.4 帧间变换编码	(96)

5.2.5 帧间 DCT 或混合 DCT	(97)
5.2.6 量化和传输引起的典型编码错误	(98)
5.2.7 实践考虑	(98)
5.3 参考文献	(100)
第 6 章 视频压缩技术:向量量化	(102)
6.1 引言	(102)
6.1.1 概述:标量量化器	(102)
6.2 向量量化器	(103)
6.2.1 随机 VQ:设计代码簿	(104)
6.2.2 最相似码向量的搜索策略	(106)
6.2.3 晶格 VQ	(107)
6.2.4 哪种晶格最好	(108)
6.2.5 工作量更少:金字塔 VQ	(109)
6.2.6 晶格 VQ 还是随机 VQ	(109)
6.3 应用于图片编码的 VQ	(111)
6.3.1 分类的 VQ(CVQ)	(111)
6.3.2 高通和带通信号的应用	(112)
6.3.3 预先可视化 VQ	(113)
6.3.4 向量量化和防错	(113)
6.4 参考文献	(116)
第 7 章 子带编码	(118)
7.1 从声音到视觉	(118)
7.2 解相关原则	(118)
7.2.1 基本框图	(119)
7.2.2 增益	(120)
7.2.3 子带编码和 DPCM	(123)
7.2.4 频带分割的应用	(123)
7.2.5 一个简单的例子	(124)
7.2.6 变换也是滤波器组	(126)
7.2.7 子带编码和变换编码	(126)
7.2.8 QMF 组	(127)
7.2.9 图片频带分割的 3 种结构	(128)
7.2.10 量化	(130)
7.2.11 能消除所有依赖吗	(131)
7.2.12 应用考虑	(131)
7.2.13 运动图片的子带编码	(132)
7.2.14 子带编码:成果和前景	(133)

7.3 参考文献	(133)
第8章 声音和音频	(135)
8.1 概述	(135)
8.1.1 听觉屏蔽	(135)
8.1.2 频谱屏蔽	(136)
8.1.3 瞬态屏蔽	(136)
8.1.4 失真	(137)
8.1.5 声道	(138)
8.2 模拟压缩技术	(141)
8.2.1 预加重	(141)
8.2.2 宽带压缩扩展	(141)
8.3 数字音频压缩	(143)
8.4 数字压缩技术	(144)
8.4.1 子带 ADPCM 编码	(145)
8.4.2 线性预测	(146)
8.4.3 自适应量化	(147)
8.4.4 apt-X100 编码延时	(148)
8.4.5 位错误响应	(148)
8.4.6 apt-X100 的实现过程	(149)
8.4.7 自同步	(149)
8.4.8 子带 APCM 编码	(149)
8.4.9 ISO/MPEG 第2层编码	(150)
8.4.10 比例系数	(152)
8.4.11 音频数据位分配	(152)
8.4.12 取消错误	(152)
8.4.13 编辑压缩数据	(152)
8.4.14 应用方面的考虑	(152)
8.5 参考文献	(153)
第9章 视频压缩标准	(154)
9.1 设置标准	(154)
9.2 视频压缩系统	(154)
9.2.1 JPEG	(154)
9.2.2 基线编码器	(155)
9.2.3 MPEG	(155)
9.2.4 MPEG-1	(157)
9.2.5 MPEG-2	(159)
9.2.6 视频源代码	(159)

9.2.7 空间和 SNR 可量测性	(160)
9.3 使用 MPEG-2 传输流进行定时和同步	(162)
9.3.1 MPEG-2 视频压缩	(163)
9.3.2 MPEG-2 传输流	(164)
9.3.3 系统时间戳	(165)
9.3.4 参考时钟同步	(167)
9.4 参考文献	(168)
第 10 章 压缩的应用	(169)
10.1 视频会议	(169)
10.1.1 H.261	(169)
10.1.2 H.261 编码	(172)
10.2 数字视频转换	(172)
10.2.1 TV 接收模式	(172)
10.3 CD 应用	(174)
10.3.1 CD 物理规范	(174)
10.3.2 CD 系统标准	(175)
10.3.3 红皮书:音频 CD	(176)
10.3.4 红皮书重放	(177)
10.3.5 黄皮书:CD-ROM	(178)
10.3.6 ISO 9660 格式	(179)
10.3.7 绿皮书(CD-I)和黄皮书	(180)
10.3.8 CD-I 中的 ADPCM 音频压缩	(180)
10.3.9 CDI 就绪光盘	(181)
10.3.10 增强 CD(CD+)	(181)
10.3.11 CD 桥光盘	(181)
10.3.12 相片 CD	(182)
10.3.13 CD-I 全动和视频 CD	(183)
10.3.14 可录制 CD	(183)
10.3.15 CD 视频	(184)
10.3.16 MMCD	(184)
10.4 参考文献	(185)
第 11 章 计算机的压缩应用	(186)
11.1 简介	(186)
11.1.1 TV 音频和视频信号	(186)
11.1.2 PC 总线和视频捕获结构	(186)
11.1.3 模拟 MUX 结构	(188)
11.1.4 数字 MUX 结构	(189)

11.1.5	DAC 附加结构	(189)
11.1.6	共享帧缓冲区结构	(190)
11.1.7	外设元件接口(PCI)结构	(190)
11.1.8	PCI 多媒体接口	(190)
11.1.9	立体声解码器	(191)
11.1.10	视频换算器	(192)
11.1.11	闭路字幕和图文电视解码器	(192)
11.1.12	MPEG-1 音频和视频解码器	(192)
11.1.13	软件解码	(193)
11.2	只用软件压缩的算法	(196)
11.2.1	向量量化	(196)
11.2.2	小波变换	(196)
11.2.3	分形	(196)
11.3	PC 应用	(197)
11.3.1	PCI 总线	(197)
11.3.2	PCI 视频使新应用得以实现	(198)
11.3.3	PCI 多媒体桥 IC 中的 SAA714X 家族	(199)

第1章 理论和实践

1.1 基本原理

我们的眼睛和耳朵以及大脑具有不可思议的获取和处理信息的能力。尽管这种能力很强大,但是它仍然是有限的,而且大多数的声音、图像和文本资源并没有以最有效的方式展现其中的信息。经济学中的基本概念是资源的稀缺性。存储介质的使用,例如磁盘和光盘,以及传输通道,例如电话线或通信带宽,是有限的。通过源编码,有可能更加有效地使用这些有价值的资源。

1.1.1 源编码

源编码(source coding)听起来很神奇,但是实际上很普通。任何有效的展现、传输或存储信息的策略都属于这个类别。源编码的例子包括电报中的莫尔斯代码、电话线上的有限带宽以及在电影院中利用静态图像的序列投影来制造动态错觉。传真使用源编码技术。压缩是源编码的子集。

1.1.2 冗余和不相关性

冗余和不相关性并不是生活的本质,它们是两种基本的现象,正是它们使得在源编码中减少数据成为可能。例如,一些源可能以每秒钟很多位的速度来发出一些信号,而实际上严格来说并不需要这么多位。在信息理论中,这称为冗余。源编码可以减少这种冗余,并用每秒钟更少的位数来产生信息的新表现形式。因为仅仅去除了冗余信息,并没有丢失任何“真实”的信息。

很多信息通常都是相关的。电话簿中的名称是按照阿拉伯字母顺序排列的。以英语开头的句子通常不会以德语结束。视频通常在几个方面都是相关的。当一个人沿着图片的某行扫描时,颜色通常不会突然变化;因而,它们是空间相关的。与此相反,这在由文本构成的页面中通常是不正确的,在这里黑白之间的迅速变化构成了字母。然而,由文本构成的页面的空间相关允许在传真机上进行更加有效的源编码。

与此类似,除非场景发生变化,在视频通信中后续的图片与前面的图片之间不会有太多的变化。这称为时间相关。

1.1.3 无损和有损源编码

在无损编码中,原始的信息或信号可以以完全相同的方式重新构造。因而,信息可以用这种方式处理很多遍,而不会有任何退化。这是因为在无损编码中,仅仅去除了冗余信息。

然而,完美的重新构造输入信号在很多情况下是没有必要的。在音频和视频数据中,有一些信息人类很难或者不可能听见或看见。这样的信息应该被认为是不相干的,而且应该在编码过程中删除(丢失)。这称为有损源编码。然而,数据会丢失某些完整性,这将在信号中产生

失真。

1.1.4 编码和传输

有损编码的一个例子是将一个数字“字”分配给相应的模拟值。这可以用多种方法完成,其中的某些比其他的方法有效得多。例如,我们可以为那些经常出现的值指定短一些的字。

人们期望对某段信号进行一种计算,以便在另一个领域或范围中进行分析。例如,我们可能从时域中映射一个信号到频域中,在这里可以对信号采用压缩技术(变换),如图 1-1 所示。

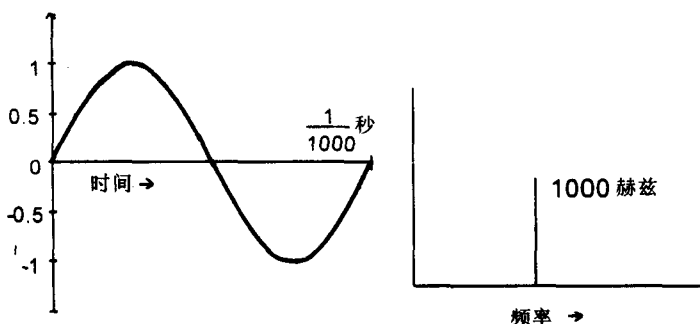


图 1-1 时域和频域中的正弦波形

对于为信息内容建立模型和应用压缩原则,变换都是非常强大的工具。在音频世界中,经过变换,可以在音频频谱上查看内容。在视频领域中,变换可以帮助人们分析某张图片中的空间相关性(或细节部分);它也可以用于图片高度和宽度的处理。视频压缩的最著名的变换是离散余弦变换(DCT),这将在下一章中进行讨论。DCT 是傅立叶变换(Fourier transform)的特殊情况。

1.2 傅立叶变换

傅立叶分析建立在这样的前提基础之上,即任何周期性波形可以分解为一系列谐波相关的幅值和相位不同的正弦信号。相反,通过适当的重新组合这些正弦波形可重新建立相同的信号。这种观点允许我们在时域采集一段看起来非常复杂的波形,然后在频域中查看它的每个组成部分。

在过去(15年前),这样的变换通常是在模拟域内完成的,方法是在连续周期性音调上使用滤波器去除某些频率。随着数字信号处理的普遍发展,现在这样的工作是在数字领域内完成的。

傅立叶变换像统计一样,它是一种强大的工具。和统计类似的是,如果没有正确的理解和使用,它可能导致错误的结果。要记住,音频和视频并不是仅仅由连续的周期音调所组成的,我们需要在时间或空间上定义一个窗口,在这个窗口上进行变换。而且,这种工具用于评价前置滤波器和后置滤波器以及其他的信号处理器。如果进行数字处理,则假设数字化过程没有增加人工噪声。

1.3 Z 变换

另一个有帮助的工具是 Z 变换,可以由复杂的频谱 $X(jf)$ 乘以 $\exp(j \times 2 \times \pi \times f \times T_a)$ 得到,它等于具有周期 T_a 依赖时间的信号的时间历程。由于延迟 T_a 是一个基本的运算,引入公式:

$$z = \exp(j \times 2 \times \pi \times f \times T_a)$$

Z 变换的结果,可以用于分析输出数字序列 $y(n)$ 和输入数字序列 $x(n)$ 在时间点 $n \times T_a$ 上的关系。在上下文中,Z 变换经常以 $z[\exp(-1)]$ 或 $z[\exp(-n)]$ 的形式出现, $z[\exp(-1)]$ 是信号延迟一个时钟周期 T_a , $z[\exp(-n)]$ 是延迟 n 个时钟周期。

1.4 应用要求

音频和视频都用于不同情况的不同目的中。这样就不可能有一种统一的、用于一般目的的压缩策略。在下面讨论的应用中,某些参数可以具有很大的差别。

1.4.1 点对点与广播

在电话会议这个例子中,两个地点同时向另一方发送和从另一方接收数据。在电视这个例子中,一个源可以向千百万接收者发送数据。从这两个例子中,我们可以看见编码器和解码器的比例差别非常大。如果编码器明显少于解码器,可以期望在编码器中完成更加复杂的工作(因而提高成本)。

1.4.2 实时与离线

一个新的报告是这样的一个例子,一些信息需要实时,也就是在发生的时候进行传送。另一方面,电影可以用很多天进行压缩,只要它可以连续地解压缩以供观看。这种信息传输方式是离线模式。

1.4.3 保真度

解压缩的图像需要与原始资料保持多大的真实程度?例如,如果图像用于医疗诊断,我们可以合理地要求具有较高的保真度。

1.4.4 质量

质量是人们如何接收资料的量度。对于视频,这包括图片大小、分辨率、帧速率和可能的颜色数目。对于声音,质量包括动态范围、频率响应和是否失真。前面已经说明,对于不同的用途可以接受不同的质量要求。一部引起轰动的电影的视觉和声音质量可以极大地增加感性体验。另一方面,许多人从卡通,例如 The Simpsons 中找寻乐趣,在这些卡通中,编写和风格弥补了颜色的不足。在过去的 15 年中,孩子和成人花费了成百上千万的小时用相对低质量的音频和视频玩视频游戏。

1.4.5 错误的敏感度

一些传输通道或存储媒体比其他的要可靠得多。当有些信号开始衰减时,有可能完全丢失声音或图片;其他的信号衰减很慢,正如模拟信号那样。例如,远离传送者的电视信号接收器收到的信号部分衰减,可以产生噪声,但是声音和图像是可以理解的。

1.4.6 可伸缩性

是否仅仅只有一个应用的实现?或者某些系统是否比其他系统具有更大的解压缩能力?可以期望创建这样一种编码信号,它对于那些愿意为高性能付出代价的人进行非常好的解压缩,而仍然为那些硬件功能较差的用户提供某些种类的图片。

1.4.7 专用硬件

我们可能希望在已经安装的设备基础上提供音频和视频服务,例如在个人计算机上,而不需要人们外出以购买另外的专用硬件。在这种情况下,我们选择对目标硬件平台合适的软件算法。

1.4.8 压缩率

考虑到可以使用的带宽或存储能力,以及期望的分辨率(在视频情况下是空间和时间分辨率,在音频情况下是频率和动态范围),将导致人们所期望的压缩率。虽然某些压缩技术比其他技术更加适合高压缩率的情况,但是要记住,低压缩率将产生高保真度和低人为噪声。

1.4.9 成本

一些应用比其他应用要对价格更加敏感。虽然很容易想象,一个视频产品工作室可能愿意为压缩设备付出成百上千万,但是通常假设,很难在硬件成本上付出 300 美元,以使大多数家庭中购买数字视频解码器。图 1-2 说明了在流行的应用中编码器和解码器的成本是如何分配的。下面建立了 4 种不同应用的模型,每一个都具有不同的要求:

- 用户广播。用户广播的例子包括现存的陆地通信、计划中的高清晰度陆地通信以及通过卫星直接面对家庭的通信。在这些例子中,每个传输者具有成千上万的接收者。因为接收者是消费者,所以对价格很敏感。对于现场直播,例如体育报道或新闻,编码过程必须实时进行。
- 用户回放。用户回放系统的例子包括各种不同的基于压缩磁盘的产品,例如那些从 Sega、Philips 和 3DO 得到的产品。作为消费产品,它们的价格定位在 300 美元左右。信号编码并不需要实时进行。
- 桌面视频。这个术语包括相当广泛的情况。正如桌面出版的迅速发展增加了创建时事通信和邀请函的人数一样,桌面视频拓展了视频和多媒体生产者的基础。那些正在创建压缩磁盘只读存储器(CD-ROM)的人可能使用音频和视频编码。希望欣赏这些 CD-ROM 的人仅仅需要解码器,他们可以有或者没有解压缩的专用硬件。在 1994 年,卖出了 1800 万个 CD-ROM 驱动器用于个人计算机的配置中,其平均售价为 1700 美元。