



多媒体 通信协议 与应用



[美] F.Kuo,
W.Effelsberg, et al. 著
龙晓苑 译
吴良芝 审校

IN913.04
K99

版
社



清华大学出版社

TN915.04

K99

432138

Prentice Hall PTR

北京科海培训中心

多媒体通信协议及应用

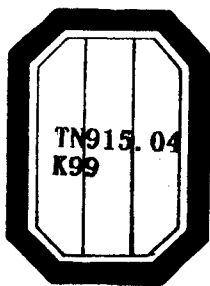
[美] F. Kuo, W. Effelsberg, et al. 著

龙晓苑

译

吴良芝

审校



清华大学出版社

(京)新登字 158 号

著作权合同登记号:01-98-0888

DX12/04

内 容 提 要

这是一本有关网络多媒体规划及方案的技术指南,它涵盖了设计和管理多媒体通信的工程师及网络专业人员所必须了解的最新内容。

本书从详细介绍网络多媒体所涉及的技术及性能需求,阐述可用来最大限度地提高网络多媒体应用系统性能的技术手段及方法;内容包括:对压缩方法的比较,子波压缩及分形压缩技术、子网技术、ATM 网上 IP 多站点技术;同时介绍新一代多媒体的网络层、传输层协议以及视频服务器系统、推技术、视频会议、网络多媒体教学、多选手游戏系统的应用等。

Multimedia Communications: protocols and applications

Copyright © 1997 by Prentice Hall PTR

All rights reserved. No part of this book shall be reproduced, stored in a retrieval system, or transmitted by any means, electronic, mechanical, photocopying, recording, or otherwise, without written permission from the publisher.

本书中文简体字版由美国西蒙与舒斯特公司授权科海培训中心和清华大学出版社。未经出版者书面允许,不得以任何方式复制或抄袭本书内容。

版权所有,盗版必究。

本书封面贴有 PRENTICE HALL 激光防伪标志,无标志者不得进入各书店。

书 名: 多媒体通信协议及应用

原著者: Franklin Kuo, Wolfgang Effelsberg

译 者: 龙晓苑

出版者: 清华大学出版社(北京清华大学校内,邮编 100084)

印刷者: 北京市朝阳区科普印刷厂

发 行: 新华书店总店北京科技发行所

开 本: 11.5 印张: 11.5 字数: 278 千字

版 次: 1998 年 7 月第 1 版 1998 年 7 月第 1 次印刷

印 数: 00001~5000

书 号: ISBN 7-302-03068-5/TP · 1231

定 价: 19.00 元

致 谢

我们十分感谢各位作者,他们的合作精神及敬业态度令人敬佩。我们还要特别感谢 Mannheim 大学的 Wieland Holfelder 先生,他作为本书的主要技术编辑,工作成绩突出。另外,本书作者之一,UCSC 大学的 Brian Levine 先生,给 Kuo 先生及 Garcia—Luna—Aceves 先生提供了特别的帮助,我们也要特别感谢。我们也要感谢 Prentice Hall 出版社的执行编辑 Mary Franz 女士的建议、帮助,尤其是她的耐心。最后我们要感谢德国 Bonn Alexander von Humboldt 基金会的慷慨赞助,如果他们不给 Kuo 先生提供研究经费,就没有这本书的诞生。

Franklin Kuo ,
Wolfgang Effelsberg
J. J. Garcia—Luna—Aceves
1997 年

前 言

在 21 世纪即将到来之际,一场融合了通信与计算机技术的信息革命也正在悄然兴起。当今的 INTERNET 就是这场革命的先兆。即将出现的新一代技术,是一种把商业、教育、企业及远程通信系统联合在一起的技术。新的消费产品形式正在出现,它综合了电话、PC 机及电视等多种功能。目前正在开发极具创新性的远程通信系统,这种系统能够在语音、数据、图像、视频等新一代个人信息终端之间自由传输各种多媒体信息。这种新型远程通信系统需要把交换式公用电话网、广播网及有线电视(CATV)网、局域及广域数据网联系在一起。这场由远程通信系统引发的信息革命,其主要技术推动力是多媒体技术。

本书适用范围

本书可用作电气工程师和计算机科研人员的专业参考书,也可用作远程通信专业和计算机专业的高年级学生及研究生的教科书。

本书覆盖的范围

本书讲述了多媒体技术的基本概念及多媒体网络环境下的通信原理。其中,介绍了计算机通信系统的系统特性,着重介绍实现多媒体通信系统所必须的网络协议,其覆盖范围从最底层(物理层)协议到最高层(应用层)协议。

本书还强调了多媒体的通信需求,尤其强调了设计这类网络协议时所蕴含的需求。

除介绍协议之外,我们还用了一整章篇幅综述了与多媒体通信系统有关的网络基础技术,包括相当重要的异步传输模式(ATM)技术。另外,我们还介绍了与之相关的系统议题,但没有讨论这类技术的硬、软件基础方面的话题。

最后,我们分析了几种重要的多媒体通信应用系统,并提出了新的系统方法,以尽可能有效而又开销不大地支持这些通信系统。

因此,本书所介绍的各种应用既涉及到多媒体也涉及到通信技术。

本书未覆盖的内容

本书所介绍的应用系统均未涉及到多媒体信息的表示手段。比如多媒体的编辑方法及多媒体数据库的设计等,因为多媒体表示并不涉及通信系统的特性。

本书未介绍的议题还有:计算机网络基本原理,比如 TCP/IP 技术,路由技术及局域网技术等。本书假定读者已具有基本的计算机网络的知识背景,知道如何使用 INTERNET 及如何浏览 WWW 信息。

本书的由来

编写本书的最初想法来源于德国 Mannheim 大学的 Wolfgang Effelsberg 与 Franklin Kuo 这两个编者之间的一系列讨论。Kuo 先生是 1995~1996 学年间 Mannheim 大学的访问学者,其研究项目得到 Alexander von Humboldt 基金会的赞助。他的导师 Effelsberg 教授是 Mannheim 大学计算机科学系主任。Kuo 先生不仅对 Mannheim 中心展示的多媒体通信系统的技术上的高复杂度印象极深,而且对德国其它研究机构,包括位于 Heidelberg 的欧洲 IBM 网络中心(距 Mannheim 大学约二十公里)所展示的多媒体通信系统的技术复杂性印象也极深。因此,Kuo 先生与 Effelsberg 先生决定共同编辑一本有关多媒体通信的书籍,专门介绍该领域内欧洲计算机科学家的专长及技术洞察力。

由于多媒体技术的研究是国际性的,本书的第三位编者,来自美国 UCSC(University of California, Santa Cruz)大学的 J. J. Garcia-Luna-Aceves 也参加了这个编写小组。他的专长是对网际连接技术及可靠的多站点传送技术的研究。本书收录这两方面的介绍内容,使得全书更为完整并且更能反映当今的发展潮流。

目 录

第 1 章 多媒体介绍	(1)
1.1 因特网与多媒体通讯	(1)
1.2 连续媒体与离散媒体	(1)
1.3 数字信号	(2)
1.3.1 采样	(2)
1.3.2 量化与编码	(3)
1.3.3 比特率	(3)
1.4 静态图像	(4)
1.5 文本和图形	(5)
1.6 动态图形与动态图像	(5)
1.7 编码和译码	(6)
1.8 带宽与压缩处理	(7)
1.9 远程教育计划	(8)
1.9.1 背景和动机	(9)
1.9.2 远程教学方案	(9)
1.9.3 多媒体教学材料	(11)
1.9.4 工业合作伙伴	(12)
1.9.5 可对比的项目	(12)
参考文献	(12)
第 2 章 多媒体网络需求与性能	(14)
2.1 分布式多媒体应用	(14)
2.2 对等(peer-to-peer)及不对等(multipeer)通讯	(14)
2.3 适用于多媒体的网络性能参数	(15)
2.3.1 吞吐量	(16)
2.3.2 差错率	(16)
2.3.3 延迟(delay)	(16)
2.3.4 往返(round-trip)延迟	(17)
2.3.5 延迟波动(delay jitter)与延迟变差(delay variation)	(17)
2.4 多媒体通信源的特点	(17)
2.4.1 时变吞吐(throughput variation with time)	(18)
2.4.2 时间依赖(time dependency)	(18)
2.4.3 双向对称(bidirectional symmetry)	(18)
2.5 影响网络性能的因素	(19)
2.5.1 吞吐性能因素	(19)
2.5.2 网络差错性能问题	(20)
2.5.3 网络延迟性能问题	(21)
2.6 多媒体传输的网络需求	(22)
2.6.1 吞吐需求	(22)
2.6.2 可靠性(差错控制)需求	(22)
2.6.3 延迟需求	(22)
2.7 服务质量	(23)
参考文献	(24)

第 3 章 压缩方法	(25)
3.1 压缩方法介绍	(25)
3.2 基本编码方法	(28)
3.2.1 行程编码(Run-length Coding)	(28)
3.2.2 哈夫曼(Huffman)编码	(28)
3.2.3 算术编码(Arithmetic Coding)	(29)
3.2.4 离散余弦变换	(30)
3.2.5 差分脉冲码调制	(32)
3.2.6 运动补偿预测	(33)
3.3 视频压缩	(33)
3.3.1 非标准化技术	(34)
3.3.2 JPEG 方法	(35)
3.3.3 H. 261 方法	(38)
3.3.4 MPEG-1 视频	(41)
3.3.5 MPEG-2 视频	(44)
3.3.6 MPEG-4 视频	(45)
3.3.7 H. 263	(45)
3.3.8 小波图像压缩	(46)
3.3.9 分形图像压缩	(48)
3.4 音频压缩	(50)
3.4.1 脉码调制的变体	(50)
3.4.2 MPEG-1 音频	(50)
3.4.3 音感编码器(Perceptual Audio Coder)	(53)
3.5 更多的有关压缩方法的信息	(53)
参考文献	(54)
第 4 章 子网技术	(59)
4.1 多媒体应用的网络需求	(59)
4.1.1 吞吐量	(59)
4.1.2 端到端延迟	(59)
4.1.3 多点通信	(60)
4.1.4 可靠性(reliability)	(60)
4.1.5 信道同步性	(61)
4.2 建网技术	(61)
4.2.1 相关参数	(61)
4.2.2 以太网(Ethernet)	(61)
4.2.3 100Base-T(快速以太网)	(62)
4.2.4 等时以太网	(62)
4.2.5 令牌环	(63)
4.2.6 需求优先(Demand Priority)	(64)
4.2.7 FDDI	(65)
4.2.8 FDDI II	(66)
4.2.9 DQDB	(67)
4.2.10 X. 25 分组交换	(68)
4.2.11 帧中继(FR)	(69)
4.2.12 IP 分组交换网	(70)
4.2.13 ISDN	(70)
4.2.14 ATM	(72)
4.2.15 IP 交换技术	(77)

4.2.16 网络性能小结	(79)
4.3 网络化基本层结构的发展	(80)
4.3.1 网络配置	(81)
4.3.2 WAN 发展	(81)
4.3.3 LAN 拓扑技术发展	(83)
4.3.4 将 ATM 并入 LAN 底层结构	(85)
4.3.5 ATM 上的传统 IP(RFC 1577)	(85)
4.3.6 ATM 上的 IP 多站点传输	(87)
4.3.7 ATM 上的 LAN 仿真服务	(89)
4.3.8 ATM 上的本地多媒体通讯	(90)
4.4 小结	(91)
参考文献	(91)
第 5 章 多媒体的网络层与传输层协议	(94)
5.1 传统协议的原理和算法	(94)
5.1.1 路由选择	(94)
5.1.2 可靠性	(101)
5.1.3 多站点传送	(102)
5.2 传统协议带来的问题	(104)
5.2.1 传统协议不能承诺服务质量	(105)
5.2.2 传统协议不支持多站点传送	(108)
5.3 新一代多媒体协议	(111)
5.3.1 ST2,数据流协议版本 2	(112)
5.3.2 tenet 协议	(113)
5.3.3 多站点 IP 和 MBone	(114)
5.3.4 IP 版本 6	(115)
5.3.5 RSVP,INTERNET 的资源预订协议	(117)
5.3.6 实时传输协议 RTP	(119)
5.4 媒体过滤,媒体缩放及自适应性应用	(120)
5.4.1 媒体过滤	(120)
5.4.2 媒体缩放(MEDIA SCALING)	(121)
5.4.3 自适应性应用系统	(122)
5.5 小结	(122)
参考文献	(123)
第 6 章 端到端可靠的多站点传输	(127)
6.1 定义端到端可靠性	(127)
6.2 可靠多站点协议的分类	(128)
6.2.1 发送方触发协议	(128)
6.2.2 接收方触发协议	(129)
6.2.3 基于树的协议	(131)
6.2.4 基于环的协议	(132)
6.3 可靠协议的最大吞吐量	(133)
6.4 协议实现	(134)
6.4.1 可缩放的可靠多站点方式(SRM)	(134)
6.4.2 基于树的协议	(136)
6.5 缩放和效率	(138)
6.5.1 重新分配内存	(138)
6.5.2 公用 ACK 树	(140)
6.5.3 有效的 ACK 树构造	(140)

6.5.4 树维护	(141)
6.6 小结	(142)
参考文献	(142)
第7章 网络上的多媒体应用	(145)
7.1 引言	(145)
7.2 应用层组帧技术(ALF)	(145)
7.3 音频/视频会议工具	(146)
7.3.1 会议目录	(147)
7.3.2 音频/视频会议工具	(147)
7.3.3 自适应性应用	(150)
7.3.4 接收方的异质性(异机种性)	(153)
7.3.5 备有资源预订的实时应用	(155)
7.4 视频服务器	(157)
7.4.1 视频服务器系统的结构	(157)
7.4.2 MBone VCR;MBone 的视频服务器	(159)
7.5 需要可靠多站点传送的应用	(160)
7.5.1 白板	(161)
7.5.2 用于共享正文编辑的网络正文编辑器	(161)
7.5.3 MultiTalk	(162)
7.5.4 多站点文件传输	(162)
7.6 万维网上的多媒体应用	(163)
7.6.1 多站点 WEB 网页共享技术	(163)
7.6.2 万维网上的音频/视频(A/V)流	(165)
7.6.3 适用于会议的 Java 小应用程序	(167)
7.7 交互式多选手游戏	(167)
7.7.1 在线纸牌游戏	(167)
7.7.2 MiMaze	(167)
7.8 小结	(169)
参考文献	(169)

第1章 多媒体介绍

在多数人看来,多媒体指的是文本、声音、视频等多类信息的表示。事实上,多媒体的含义远非如此。多媒体的含义不仅在于它能够表示多种信息,而且还含有借助于计算机对这些信息进行操作和控制的含义,以及通过远程通信信道传输这些信息的能力。因此,我们对多媒体定义如下:

- 多媒体指的是多种混合信息(文本、数据、图像、声音、视频等数字信号)的表示形式。
- 多媒体通信指的是通过网络通信信道操作、传输和控制这些视听信号的技术。

1.1 因特网与多媒体通信

凡是使用过因特网(Internet)的万维网浏览器的人已经有多媒体通信的体验。网上大多数 HTML 文件既涉及到正文也涉及到图像。某些应用还用到语音、音频和视频。如果你有网上万维网的使用经历,对网上多媒体通信的不足之处也一定深有体会。从网上下载一幅图像与仅从网上下载一篇正文比较,前者所花的时间要多许多。

例如,Mannheim 大学的主页信息显示有该大学主建筑的一幅图像,同时附有有关该大学的文字说明(图 1.1)。对于与该大学同处一地的用户来说,假定他使用 Netscape 之类的浏览器进行浏览,上述图像和正文说明部分可以迅速且同时在屏幕上显示出来。而对于德国其他城市的用户来说,正文部分的出现要比图形部分的出现稍快一些。在某些采用低速链路连接到因特网的第三世界国家,网上图像传输所花的时间太长,使这种图形图像的传输失去了实际意义。

为什么会出现上述情况呢?简单回答是因为与正文比较,对图形图像的描述需要占用相当大的数据空间。描述动态图像更是耗费空间。在万维网出现之前,多数因特网应用仅仅涉及到正文或数据,比如 E-mail(电子邮件)或 FTP(文件传输)之类应用。主要考虑不是实时通信,图形和图像的下载也不多见。再者,先前的 Internet 几乎不用于语音通信。

今天的万维网改变了这种状况。今天的 Internet 就好似一条破旧的、坑洼不平的旧道路,远不能满足现代高速轿车需要了。如今,大多数国家正在计划为未来的国际网络建立新的信息高速公路。在新一代的高速公路上,多媒体通信迅速而高效。然而,这类信息高速公路的设计是复杂的。它不仅涉及到建立宽带网络的问题,还涉及到访问、控制、管理网络信息等一系列问题。这些议题正是本书的中心议题。

1.2 连续媒体与离散媒体

本书将媒体分为两大类:连续的与离散的。连续媒体(又称时间媒体)随时间变化而变化。离散媒体则是与时间无关的。连续媒体的例子有音频与视频。常见的离散媒体例子有文本(格式与非格式化的),静态图像和图形。

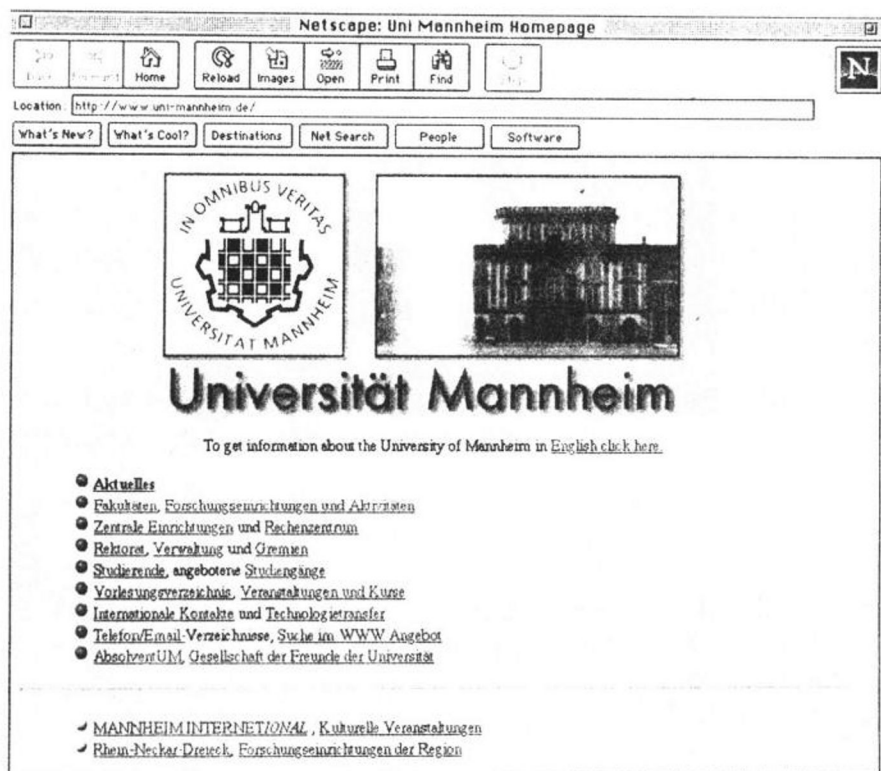


图 1.1 Mannheim 大学主页

本书所提及的媒体指的就是这两类媒体。处理离散媒体通信相对简单。由于文本文件不随时间变化,当前时刻往网上传送一份文本文件与隔 10min 后再传送这份文本文件,效果是相同的。但是,对连续媒体说来,情况就不同了。实时传送实况电视节目,其中含视频与音频媒体信息,两者都是与时间有关的。视频数据流与音频数据流必须同步传送,以确保这两种数据流能同时到达接收端。倘若通信信道允许音频信息流比视频信息流更快地通过信道,那么接收端的接收效果看起来一定很别扭。

无论是离散媒体的通信还是连续媒体的通信,对用来传输这两类信息的远程通信系统都有特定的要求,我们必须详细审视这些要求。

1.3 数字信号

传统的远程通信系统中,所传送的信息一般是以模拟形式出现的,表现为一种时间连续的信号。而在计算机通信系统中,视听信号的传送方式更乐于采用数字信号方式。这些数字信号是通过对最初模拟信号的采样、量化及编码变换而来。

1.3.1 采样

对信号的时间采样指的是在有规律的时间间隔 T 内对模拟信号进行检测和测量。由此,某个连续信号 $S(t)$ 的采样样本可以表示为:

$$S(t) = \{S(T), S(2T), S(3T) \dots S(nT)\}$$

其中, T 为采样周期, 而采样频率可以表示成 $f = 1/T$ 。

根据 Nyquist(奈奎斯特)的采样理论^[1], 如果某个信号的频率组成中, 最高频率为 f_0 , 那么为了精确地表示该信号, 最小频率取值应为 $2f_0$ 。例如, 如果某种声音信号的最大频率值为 3KHz, 那么其最小采样频率至少应取为 6KHz。即, 每秒必须对该语音信号进行至少 6 000 次采样, 才能做到准确地表示原始的模拟信号。

1.3.2 量化与编码

在完成对模拟信号的采样后, 再对采样值量化并编码成二进制位。量化过程就是把采样值表示成一个离散型振幅值的集合。量化/编码过程的精确性是由采样值经量化/编码后所得的比特数来决定的。例如, 如果某采样值 $S(iT)$ 由一个具有 3 比特数的码值量化, 则 $S(iT)$ 可以假定, 仅有 8 个不同的离散值, 由 000 变化到 111。其中还可将每一离散值进一步编码以提高数字化信息传输或存储的效益。对照图 1.2, 我们可看到在时间间隔为 nT 的各离散区间上对模拟信号 $S(t)$ 进行采样, 然后将采样值量化并编码, 每一采样周期上的结果用垂直粗线条表示。

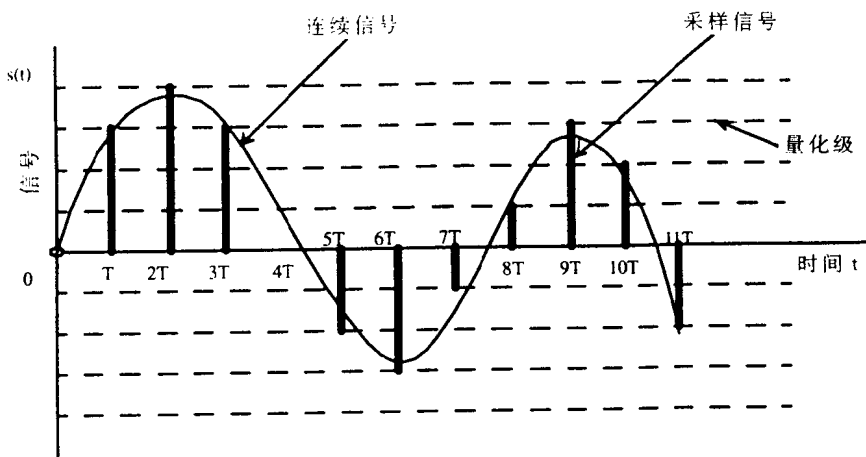


图 1.2 经采样及量化的模拟信号

电话通信应用中, 语音的数字化标准目前是采用 16 个比特数来表示每一声音样本。这样一来即可表示 2^{16} 或 65 536 种不同的振幅值。在另一些语音压缩应用中, 可能只需 8 个量化位, 这样我们就只能区分 256 种不同的振幅值。

1.3.3 比特率

比特率是采样率和量化过程中使用的比特数的产物。例如, 电话通信中, 语音信号的带宽大约为 3kHz, 根据 Nyquist 理论, 这意味着采样率(频率)至少应为 6kHz。如果我们保守一点, 把采样率定为 8kHz, 假定我们使用的是一个 8bit 的量化器, 那么该电话通信所要求的比特率应为: $8000 \times 8 = 64\text{kbps}$ 。

再看另一类应用。例如,音频光盘(CD)。对于高保真音频,通常其带宽应取为 20kHz,因而其最小的采样频率应为 40kHz。采用 44.1kHz 的工业标准采样率和十六位比特数的量化器,那么对于立体声(双声道)应用,我们得到该音频光盘的比特率为: $44100 \times 2 \times 16 = 1\,410$ kbps。

比特率是数据通信的一个重要参数。公用数据网的信道传输能力通常是以每秒传送多少 kb 数或多少 Gb 数来衡量的。对于综合业务数字网(ISDN)来说,语音通信中标准比特率定为 64kbps。

表 1.1 给出电话通信、远程会议通信(对音质要求更高)、音频光盘(CD)及数字音频带(DAT)等几类应用的比特率的相关比较。

表 1.1 数字音频格式^[6]

应用类型	采样速率 (kHz)	带宽 (kHz)	频带 (Hz)	比特率 (kbps)
电话	8.0	3.0	200~3 200	64
远程会议	16.0	7.0	50~7 000	256
光盘	44.1	20.0	20~20 000	1 410
数字音频带	48.0	20.0	20~20 000	1 536

1.4 静态图像

前面我们提到这样一个事实:静态图像的传送与文本传输比起来,前者所花的时间要多许多。为什么会出现这种情况呢?最根本的原因是,图像是由像素组成的,每一幅图中都包含了相当多数量的像素。像素是什么?在多数一般性字典里是找不到对它的定义的。某些好的技术字典是这样定义像素的:

像素:(名词)图像的最小的不可分单元或点,该单元或点的颜色或亮度可以控制。

如果我们在放大镜下观看报纸上的一张相片,就能发现相片是由一些十分精细的网格点构成的。其中每一具体的点都具有特定的灰度,由黑变到白。这些点即被称为像素点。我们认为这类像素都具有具体的值,就报纸上的相片来说,这个具体的值就是灰度。

对于彩色相片或彩色显示屏来说,像素值还与颜色及亮度有关。

对像素的一个更好的定义是:

像素:图形元素,是图像的最小解析单元。该单元既可显示在屏幕上也可以存储在内存中。单色图的像素有着自身的亮度值,由 0(代表黑)变化到某个最大值(代表白),对于八位像素来说,其最大值为 255。对彩色图来说,每一个像素都具有自己的亮度及色彩值。

计算机图像是由像素构成的位图。对普通分辨率的计算机显示屏来说,有 768 行,每行由 1024 个像素点组成。拿彩显方式来说,假设每一像素的色彩值及亮度值均由 24 个比特数来表示(每单位像素内的比特数,或称 bpp),那么每帧计算机图像所占的总比特数为:

$$\begin{aligned}\text{总比特数} &= 1024 \times 768 \times 24 \\ &= 18.874\text{Gb}\end{aligned}$$

假定我们在 14.4kbps 的调制解调器上传送这幅图,那么所花的传输时间应为:
 $18874000/14400=1310s$ 或 21.84min。

这个传送时间对于计算机实时用户来说显然太长了。那么有没有改进的方法呢?下面给出了四种基本的改进办法:

1. 使用更快速率的信道来传送图像。诸如比特率为 1.544Mbps 的 T1 载波线路。
2. 减少每单位像素上的比特数。即允许更少数目的亮度及色彩配置。
3. 降低显示器的分辨率(每行允许更少的像素及每帧图用更少的行来表示)。
4. 消除显示器的冗余。这里指的是消除那些多余的表示同一对象的像素。

图像压缩技术把方法 2、3、4 组合在一起。

1.5 文本和图形

相对于图像来说,普通文本或格式化文本对信道传输能力的要求就相对低了。普通文本字符由 8 个 bit(也称一字节)来表示。格式化文本字符由二字节表示。对于单页正文说来,由 64 行组成,每行有 80 个字符。因此,一整页正文所占的比特数为: $80 \times 64 \times 2 \times 8 = 82kbits$ 。其传送起来只需 5.7s(仍然使用 14.4kbps 的调制解调器)。与图像相比,图形指的是由人或计算机绘制的图形,空间上由行构成。我们可以把图形看成是表示信息的对象组合。这些对象可以用数学方法去描述,也可以通过计算机命令去描述。比如,通过指定两端点得到一条直线段,通过指定圆心位置及半径可得到一个圆,等等。图形是可以修改和编辑的,而图像则不行。与位映像图比较起来,图形所需的内存空间要少许多。再就是网上传送图形所花的传输时间也比位映像图所花的时间要少得多。

多媒体通信中,我们既要处理图像,也要处理图形。

1.6 动态图形与动态图像

动画是由图片的时序序列构成的。每一幅这样的图我们称之为帧。帧与帧之间的图形对象间只有细微差别。帧序列的播放是按时间顺序进行的,这样人眼看起来就有运动的效果。每秒所播放的帧的数目称之为播放速度,又称为帧速率。如果帧速率太慢,则对象运动的效果看起来就很不连贯。电影胶片的帧速率通常介于每秒 25 帧至 30 帧之间(25~30fps)经验告诉我们,常用每秒 16 帧或更高速率来表现均匀运动。

动画和摄影机表现的是真实世界中运动的对象。每一个摄影镜头就是一幅独立的帧图。这些随时间变化的镜头序列放映出来就表现出了对象的运动过程。

对于视频显示来说,帧速率就是时序序列的播放速率。通常其帧速率在 25 至 30fps 之间。

一秒的视频里究竟能容纳多少数据?我们以公用中间格式(CIF)为例(本书后面章节要进一步介绍 CIF)。假定帧速率为 30fps,每行有 360 个像素点,每幅图由 288 行组成,每单位像素采用 24 个比特数来存放亮度及颜色值。则每秒 CIF 视频中的比特数为:

$$360 \times 288 \times 24 \times 30 = 74.65Mbps$$

当今社会,最常用的多媒体设备就是PC机。PC机中最常见的通信设备为调制解调器,它是用来联接普通电话线的一类设备。由于目前调制解调器的速率为28.8kbps或更低,则传送一秒CIF视频所花的时间为2592s或43min。这意味着实际上PC机根本不可能接收视频信息。同时,也隐含了如果PC机希望做到接收视频数据并存储起来以便日后播放,则PC机本身必须有足够巨大的视频存储空间。

1.7 编码和译码

人类只能感知视听模拟信号。对于数字通信系统来说,这类模拟信号必须经编码器转换成数字信号,即执行模数(analog to digital:A到D)转换、量化、压缩等处理。当今的许多电话网,比如综合业务数字网(ISDN)就是完全数字化的,因此它既可用于数字传输又可用于数字转换。这类电话系统的终端用户可以是人,也可以是机器(计算机)。如果发送端和接收端的用户均是人,那么初始信号和终端信号都必须为模拟信号,必须进行模数转换和数模转换。参见图1.3。相反,如果终端用户是计算机,那么就不必进行这两类转换了。除了数模和模数转换外,编码器/译码器还必须对数据进行压缩处理。数据压缩指的是使用某种紧缩格式来表示数字信号的技术。压缩方法可分为两类:不失真和失真压缩方法。对于非失真压缩说来,编码/译码完成后,接收端得到的信号与初始的发送信号是完全吻合的。但对失真压缩来说,接收端接收的信号与发送端最初发送的信号之间是存在某些差异的,这种差异即称之畸变。压缩算法的目标一方面是为了得到对原始信号高度压缩的版本,而另一方面是把畸变控制在尽可能小的范围内。许多情况下,压缩信号是发送端信号的低比特率版。

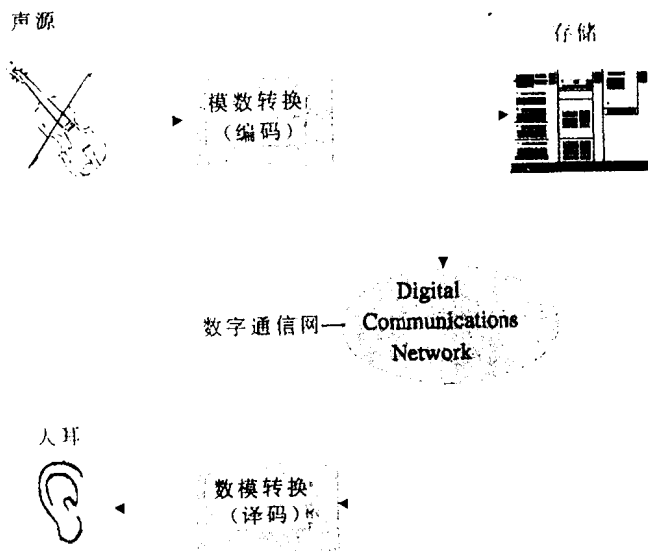


图 1.3 数字传输中的编码/译码

不同种类的媒体使用不同的压缩算法。语音压缩算法与视频压缩算法就截然不同。压缩方法优劣的评价与人的视听感觉有关,同时与信号本身的保真度也有关系。例如,对高保真音频设备来说,通常频率的范围在20~20 000Hz之间。而根据常识,我们知道现实生活中绝

大多数人不可能听到频率高于 16 000Hz 或低于 40Hz 的声音。设计压缩算法时应考虑这类实际情况,生成可接受的逼真音频信号。对视频信号来讲,在设计压缩算法时应减少单位像素内的比特数,如将 24 位减为 16 位。或者将那些不需迅速变化的可视画面的帧速率从 30fps 降低到 16fps。

现代许多压缩方法都是源于常见的人类感性认识。这些方法也被看成是感性编码。音频压缩技术中的微感畸变(JND)^[7]概念就是这方面的一个例子。它是指初始信号与编码信号之间细小差别的畸变级可以被听者感觉出来。对高保真音频设备来说,JND 值与频率可以被测量出来或者模型化。有了 JND 曲线的知识背景,在进行信号压缩时就可以把位于曲线下方的信号忽略掉而不会对人的听力效果产生影响。

1.8 带宽与压缩处理

如上所述,信号压缩技术对多媒体通信来说是重要的。如今远程通信环境下,压缩技术及高速网环境已经取得了巨大的进步。随着实际带宽的不断增长和局域网、广域网成本的不断降低,对信号的压缩需求也在不断减少。但是,由于越来越多的用户发送和接收多媒体数据,尽管实际带宽在不断增加,仍然需要使用压缩技术。

压缩技术在不断发展。例如,电话通信系统中,1972 年网络通信的标准比特率为 64kbps。而到 1984 年,新标准比特率定为 32kbps。到 1992 年,更新的标准比特率出现了,为 16kbps。目前正在做的工作是进一步降低标准比特率以期达到 8kbps。这种改变应归功于语音压缩算法的发展以及功能日益增强的用于语音压缩的集成电路的发展。

目前国际压缩标准有不少,它们分别适用于电话网络、音频、图像和视频信息的传输。表 1.2 中列举了其中一些标准。

表 1.2 电话、音频及视频的国际标准

图标	比特率	应用程序
G. 721	32 kbps	电话
G. 728	16 kbps	电话
G. 722	48~64kbps	电视会议
MPEG-1(audio)	128~384 kbps	双声道音频
MPEG-2(audio)	320 kbps	五声道音频
JBIG	0.05~0.10 bpp	二进制图像
JPEG	0.25~8.0 bpp	静态图像
MPEG-1,2(video)	1~8 Mbps	视频
Px64	64~1 536 kbps	视频会议
HDTV	17 Mbps	高清晰度 TV

另一个压缩技术的例子是针对高保真 CD 音频的。我们已经知道,在频率为 44.1kHz,单位像素容纳 16 比特数的条件下,对频率为 20kHz 的声音进行采样,可以生成比特率为 1.412Mbps 的立体声。使用音感编码器(PAC),AT&T 贝尔实验室的科学家们通过实验证明了以 64kbps^[9]速率播放 CD 音质的音乐的可能性。这表明 CD 音质的音频可以通过基率 ISDN 信道发送出去。图 1.4 中给出了当前信号压缩的范畴。