

何书前 陆 娜 蒋文娟 吴淑雷 编著

现代 网络视频编码技术

XIANDAI WANGLUO SHIPIN BIANMA JISHU

湖北科学技术出版社

此书受海南师范大学学术专著出版资金支助

现代 网络视频 编码技术

何书前 陆娜 蒋文娟 吴淑雷 编著

XIANDAI WANGLUO
SHIPIN BIANMA JISHU

湖北科学技术出版社

图书在版编目(CIP)数据

现代网络视频编码技术 / 何书前等编著. —武汉: 湖北科学技术出版社, 2009.11

ISBN 978-7-5352-4430-7

I. 现… II. 何… III. 计算机网络—视频系统—图像编码 IV. TN94 TN919.81

中国版本图书馆 CIP 数据核字 (2009) 第 168433 号

责任编辑: 唐洁 刘志敏

封面设计: 戴旻

出版发行: 湖北科学技术出版社

电话: 027-87679468

地址: 武汉市雄楚大街 268 号

邮编: 430070

(湖北出版文化城 B 座 12-13 层)

网址: <http://www.hbstp.com.cn>

印刷: 武汉科利德印务有限公司

邮编: 430071

880×1230 1/32

11.25 印张

280 千字

2009 年 11 月第 1 版

2009 年 11 月第 1 次印刷

定价: 18.00 元

本书如有印装质量问题 可找本社市场部更换

内 容 提 要

本书针对目前网络视频编码的研究背景和发展现状,全面系统地介绍了数字视频数据编码的基础理论知识和国际视频编码标准的形成与发展,重点分析了先进的视频压缩编码标准 H. 264 的各项关键技术,包括预测编码、变换编码和熵编码的基本原理和性能分析;对数字视频可分级编码与传输进行了论述,最后结合数字水印技术的发展,介绍数字视频数据的安全性研究的进展。

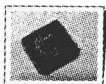
本书作者结合多年的科研工作体会,阅读了最新的视频压缩编码相关文献、资料、学位论文,编写此书时,详细地描述了现有视频编码最新的研究进展,并引入了一定的图表说明,实验数据分析,供读者查阅参考。

本书适用于视频技术、视频服务和多媒体网络的工程技术人员与管理人员阅读,可作为计算机应用技术、通信与信息系统、信号与信息处理和数字多媒体等专业的高年级本科、研究生、研究人员和高校教师的参考书。

前 言

进入信息时代以来,随着通信技术与计算机技术的发展,推动了多媒体在处理和传输领域的迅猛发展。通信与计算机的结合产生了计算机网络,从而促使多媒体应用的更新换代,从模拟到数字化,从表示到传输,从电视台到家庭,无不进入多媒体网络时代。作为多媒体核心地位的视频处理起到了举足轻重的推动作用。首先,数字视频编码与传输领域取得了突破,使得在目前的异构网络共存的低码率带宽环境中传输高清晰质量的视频成为可能;其中,包括了流媒体技术的出现,无线视频传输,可分级编码技术等。其次,超大规模集成电路技术的发展,为复杂的视频压缩算法提供了硬件平台,保证了给定失真的情况下可获得最优的压缩效果。最后,数字水印技术的发展,使得混乱的网络数字媒体环境得到了净化,为网络数字视频提供安全保障。

另外,信息编码理论领域不断取得新的进展,为数字视频编码提供了新的信息编码理论指导;因此,研究有效的数据压缩和解压缩技术成为了重要的、关键的研究方向,率失真理论也随着数据压缩技术的发展得到了解决,从而促进了一大批新的数字视频压缩与通信算法的提出,大大地提高了视频压缩和传输的效率。同时,ISO/MPEG 和 ITU-T 最终走到一起,实现了强强联合,共同制定了最新的国际视频压缩标准 H. 264,使数字视频压缩标准实现了单一化,即从低码率到高码率,从存储到传输,从媒体服务到视频消费,H. 264 都具有普遍适用性,为无限宽带、全连接性、高移动性和海量存储的多媒体应用提供了支持。特别是无线网络技术的爆炸性发展改变了多媒体网络应用的基础结构,无线多媒体网



络将彻底融入人们的工作和生活,怎么解决数字视频巨大的数据量和无线网络带宽的限制之间的矛盾,H.264可分级编码扩展标准给出了最佳的解决方案。随着多媒体研究领域(包括多媒体计算机、多媒体通信、多媒体表现)成为计算机和通信发展中重要的研究热点,逐步融合了图形图像、计算机视觉、计算机网络和数据库等领域,形成了一个新的多学科交叉的研究领域。本书介绍了数字视频编码的基础理论,重点突出最新的视频压缩标准H.264的发展,详细描述了视频编码和视频通信最新的研究进展和成果。

本书在内容编排上,力求由浅入深、循序渐进,合理地安排章节,既突出严密的基础理论推理,又不乏实际的应用背景分析。全书共有7章,各章内容如下。

第1章 视频编码概述。介绍人眼视觉系统的特性,给出了各种描述数字视频信号的视觉模型;在进行视频处理之前,如何对连续视频信号进行采样,以保证视频应用系统的处理参数,以及如何转化视频格式,方便系统采用的视频压缩标准处理。针对人眼视觉系统的特性,介绍了几种常用的视频质量评价方法。

第2章 信息论基础。介绍视频编码基本原理之前,首先要很好地了解一些信息理论知识,本章结合视频编码涉及的基础知识,重点描述了信息论基本概念、信源编码的基本原理、信息压缩方法和率失真理论。

第3章 视频数据压缩基本原理。介绍了视频数据的特点,压缩的基本原理,重点包括了变换编码、熵编码、量化编码等。

第4章 视频压缩编码标准。介绍了MPEG-1,MPEG-2,MPEG-4,MPEG-7,H.261,H.263,H.263+视频编码标准的基本框架,分析了各种标准的实际应用和相关性能。

第5章 H.264压缩视频编码标准。对比分析了H.264与以前视频编码标准的不同,系统地介绍了H.264视频编码标准的一些新的关键技术和基本原理,包括比特流组织、预测编码、可变

块大小、多精度运动估计、多参考帧选择、去块滤波、整数变换和熵编码。

第6章 视频运动估计编码技术。运动估计作为视频编码框架结构中最重要模块,本章主要介绍了运动估计的研究背景、基本原理和最新的运动估计算法。

第7章 视频可伸缩编码与网络传输。主要介绍了 H. 264 可伸缩编码的基本原理(包括空域可分级、时域可分级、质量可分级和 MCTF)、分类和实验分析,结合视频数据传输所涉及的网络协议标准,描述了可伸缩编码的差错隐藏原理。

第8章 视频水印技术。主要介绍了视频水印的基本原理、应用前景和发展现状以及最新的视频水印算法与性能评价。

第9章 率失真优化与视频编码。描述的率失真理论,介绍了经典的视频率失真模型,分析了各种压缩标准的率失真码率控制算法。

前言和第5、7章由何书前同志执笔,第1、2、6章由陆娜同志执笔,第3、4、8章由蒋文娟同志执笔,第9章由吴淑雷同志执笔。全书由何书前同志负责本版书稿的修订和审校工作。另外,本书是海南省自然科学基金项目(批准号:807063,807062)、海南省教育厅项目(批准号:Hjkj2008—36, Hjkj2009—35, Hj2009—99, Hj2009—94)和海南师范大学青年教师资助项目(批准号:HSQN0610)的成果总结。

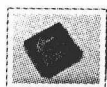
本书在编写过程中,参阅了大量的国内外经典文献、著作和学位论文,并列于每章的参考文献中,在此,谨向这些作者表示衷心的感谢。由于时间仓促,水平有限,错误和不足指出在所难免,敬请广大读者多多批评指正,以便我们进一步更正和完善。

编者

2009年5月

目 录

第一章 视频编码概述	1
1.1 人眼视觉特性与视觉模型	5
1.2 视觉模型	13
1.3 视频采集	17
1.4 视频格式	21
1.5 视频质量评价	24
参考文献	33
第二章 信息论基础	36
2.1 信息的概念	36
2.2 信源熵	43
2.3 信源编码	50
2.4 信道编码	52
2.5 率失真理论	53
参考文献	55
第三章 视频数据压缩基本原理	56
3.1 视频数据压缩的条件	56
3.2 视频数据冗余	57
3.3 数据压缩算法	58
参考文献	93
第四章 视频压缩编码标准	95
4.1 视频压缩编码标准	95
4.2 H. 26X 系列标准	98
4.3 MPEG 系列标准	100



现代网络视频编码技术 2

XIANDAIWANGLUOSHIPINBIANMAJISHU

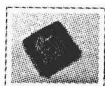
4.4	H. 261、H. 263 和 MPEG-1	105
4.5	H. 261 的码流格式	109
4.6	H. 263 编码标准基本原理	112
4.7	MPEG-4 编码标准基本原理 ^[17-20]	126
	参考文献	139
第五章	H. 264/AVC 压缩视频编码标准	142
5.1	H. 264 视频压缩编码标准主要技术	143
5.2	H. 264 预测编码	154
5.3	变换编码	177
5.4	熵编码	187
	参考文献	198
第六章	视频运动估计编码技术	202
6.1	二维运动估计原理	203
6.2	块匹配运动估计算法	209
6.3	块匹配运动估计的主要技术	211
6.4	全搜索运动估计	217
6.5	快速运动估计算法 ^{[11][12]}	218
6.6	运动补偿技术	227
6.7	一种基于 H. 264 的高效率快速块匹配运动 估计算法	236
	参考文献	254
第七章	视频可伸缩编码与网络传输	260
7.1	可伸缩视频编码技术	261
7.2	基于 H. 264 的可伸缩编码技术	263
7.3	基于小波变换的可伸缩编码	273
7.4	可伸缩视频流网络传输	280
7.5	H. 264-SVC 研究实验分析	299
7.6	总结与展望	306
	参考文献	306

第八章 视频水印技术	311
8.1 数字视频水印应用	311
8.2 数字视频水印特点	312
8.3 数字水印基本原理	313
8.4 视频水印的分类	322
8.5 数字视频水印发展	325
参考文献	331
第九章 率失真优化与视频编码	335
9.1 率失真理论	335
9.2 经典率失真模型	337
9.3 优化位分配	339
9.4 宏块编码参数选择	340
9.5 码率控制	341
参考文献	347

第一章 视频编码概述

进入 21 世纪以来,随着通信技术和计算机技术的发展,数字多媒体的应用和互联网的推广,使得家家户户在家里就能享受数字媒体带来的娱乐生活和网络时代带来的变化;而随着下一代 Internet 技术、NGN 技术和 3G 技术的发展,为交互式高清晰数字视频的应用提供了坚强的基础支撑,创造了非常有利的发展环境。目前,在我国,电话网、电视网和互联网开始融合,3G 移动通信网络已经基本成型,各种数字家庭技术也开始深入家庭生活和国民经济各方面;数字视频技术作为重要的内容服务基础,促进了各项新技术的推广。同时,随着国民经济和社会生活各方面的发展,新技术的应用,又会不断对视频编码和通信提出各种类型的应用要求,预计在未来一二十年,视频通信将成为现代通信的主要内容之一。

在传统的通信系统中,一般将通信的制式分为模拟制和数字制两大类。模拟制所处理的通信信号是时间和幅度连续的模拟信号,如传统的语音电话系统;而数字制涉及的通信信号是离散的数字信号,如计算机终端数据通信与移频制的无线电传真,都是数字制通信的早期典型例子。传统的通信系统都是从模拟制发展起来的,如电话、语音广播、电视广播等各类通信系统,到了 20 世纪 70 年代,随着数字集成电路器件的发展,数字通信才由理论研究转入实际应用。由于数字系统的抗干扰性、灵活性和低成本等优势,使得数字通信系统逐步取代了传统的模拟通信技术。到目前为止,各种各样的通信应用系统(如光纤通信、移动通信、卫星通信和图像通信)都逐步向数字制式转变,而广播电视也刚刚完成高清晰数



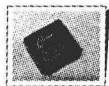
字电视的过渡。

在过去的 20 年中,通用的、无缝的、高质量的数字视频一直是企业、研究机构、标准机构和国家的目标。数字视频在传统的多媒体系统(例如广播电视系统和用户视频传输系统)已经占据了主要的位置,由于计算机网络技术的发展也推动了网络视频服务的发展,传统的电话视频会议,多媒体邮件,实时通信业务都开始向数字视频方向转变,在这中间,产生了巨大的商业利益。正是因为数字视频产业的不断演变带动了技术的革新和商业模式的转变,在这些变化过程中,商业价值也相应体现;反过来,商业利益的存在,也促使商家不惜任何代价大力发展相关的新技术,融合新的功能,不断地推出新的通信和娱乐产品,以取得先机。因此,数字视频技术在这种相互的作用力推动下,不断地发展壮大。近年来,随着无线通信技术的成熟,使得无线视频通信技术成为新的研究热点;相应的,数据安全也开始向数字视频数据扩展,视频水印和视频隐藏技术的应用也开始提上了日程。技术动力来源于处理性能的不断提高,促进了大容量存储和传输机制的出现以及视频和图像处理技术的研究和发展;当然,本书也主要围绕数字视频领域新的研究热点进行讲述。

数字视频通信主要的目的是把数字视频数据从发送端准确无误地传到接收端,在传输过程中包括了数字视频数据编码输出、网络传输、接收解码 3 个步骤,其中最重要的问题就是压缩和解压缩问题,这是由视频数据的海量性和网络环境的局限性所决定的,压缩编码是为了使得视频数据消耗的网络带宽或设备处理资源在可控制的范围内,以便网络传输和存储,然后重建并显示,达到最终的通信目的。因此,高效率的压缩处理和解压缩处理是数字视频通信的主要研究内容,是体现高性能的数字视频系统的核心内容,是目前这个领域研究结构和商业机构不断探索新方向的突破口,同时,也体现了一个数字视频产品的技术和商业优势所在,如可以提供更好的视频解码质量、更高的压缩比和灵活性。

数字视频工业发展的初期,数字视频技术在 VCD 视频和模拟视频系统的竞争中不断发展,逐步扩展了数字视频技术在多媒体领域的影响力,但是,并没有完全取代模拟技术。随着数字技术的不断革新,MPEG-2 标准的出现,完全确立了数字视频工业的统治地位,主要是广播数字电视和 DVD 视频的成功,该标准属于国际标准化组织 ISO/IEC 13818^[1](MPEG 这个名字起源于发展这个标准的工作组,即 Moving Picture Expert Group 运动图像专家组)。后来,随着网络技术的发展,无线网络的出现,广域网的高速化以及面向对象视频编码的出现,推动了新的更好的压缩工具应用,产生了两个更新的视频压缩标准,即 ISO/IEC 14496 Part 2 (MPEG-4 视频)^[3]和 ITU-T Recommendation H. 264/ISO/IEC 14496 Part 10(H. 264)^[4]。MPEG-4 视频的目的在于去掉矩形视频图像的限制,提供一个开放的、灵活的、基于对象的处理过程,属于下一代视频编码技术,但是还不够成熟。而 H. 264 有一个实际的目标,即提出一种覆盖所有应用类型的分块视频压缩标准,不仅适合与高码率的应用也能满足低码率的通信解码需求,同时编解码框架结构沿用了传统的混合编码架构,能够继承旧标准的优点,并引入了很多新的基于块编码成果,大大提高了编码效率,力求做到更有效、更有鲁棒性、更实际,支持在市场上变得越来越广泛的各种应用(例如交互式、存储和流式传输)。经过测试研究,MPEG-4 视频和 H. 264 有共同的制定目标和编码工具,具有非常相似的特征,于 2002 年走到了一起,共同制定了新的 H. 264 JVT 标准。本书主要内容是以 H. 264 为技术标准基础,介绍目前新的数字视频通信的新技术和新研究成果。

网络技术的不断发展,网络(包括了局域网和广域网)带宽也不断增加,无线网络也已经普及,家庭宽带网络也非常普遍;高容量存储技术、高集成电路和光技术的发展,使得现有的存储介质和传输介质的存储传输性价比比以往任何时候都高。虽然数字视频压缩的必要性在这个时期没有直接体现,但是,人们还是一直在努



力发展压缩技术,主要的原因如下。首先,压缩技术使得数字视频的传输和存储应用范围不断扩大,是原始非压缩视频所无法比拟的;当视频终端之间的网络带宽不能保证充足的资源传输非压缩的视频数据流的时候,丢包会造成接收端无法解码,即使在很低的视频帧率和很小的视频空间分辨率的情况下;在没有压缩的情况下,一张 DVD 光盘存储的电视质量分辨率和帧率的视频数据只能达到几秒钟,即使存储技术的发展也无法适应非压缩视频数据的海量性。其次,视频压缩技术能够提高传输和存储资源的利用率,降低相应的运营成本;在相同的传输信道条件下,传输多个同分辨率的压缩视频数据比单独传输一个同分辨率非压缩视频的吸引力更大,或者是剩下的资源可以用于其他用途,成本更低。因此,虽然存储和传输容量不断提高,但是在未来的很长时间内,压缩仍然是数字视频技术发展的核心内容。

压缩是指在去除携带信息信号的冗余,从而得到小数据量的携带信息的信号。目前的数字视频压缩方法主要分为两种:无损压缩和有损压缩。其中,无损压缩所取得的压缩效率比较低,压缩比只能达到 2~5 的范围内,起不到应有的应用效果,很少使用,它的原理是去除视频数据的时空域冗余(主要是统计冗余),在解码端可以完整地重建原始信号。因此,为了获得更高的压缩比,提高压缩编码的效率,多数实用的视频压缩技术是基于有损压缩的,有损压缩的原理是牺牲一定的信号质量,也就是允许重建信号有一定的失真,解码信号与原始信号不会完全一样,该失真在应用所能接受的范围内提高压缩比,达到传输和存储的目的。当前数字视频压缩的研究热点问题就是视频压缩算法在获得有效的压缩效果的同时,使得压缩过程中引起的失真最小。现有的视频压缩算法是通过去除时间、空间和频率的冗余来实现的。经过 20 多年的发展,图像和视频压缩成为一个非常活跃的研究和发展领域,期间许多不同的压缩/解压缩系统和算法被提出来。为了鼓励协调工作,竞争和增加机会,有必要定义标准的压缩编码和解码算法,使得来

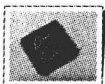
自不同厂商的产品能够有效地互联互通,相互兼容。这种需求促进了一系列图像/视频压缩国际标准的发展,针对不同的应用,出现了 JPEG、MPEG 和 H. 26X 序列标准。

1.1 人眼视觉特性与视觉模型

不管视频系统是模拟制式,还是数字制式,其处理与通信的目的地均是以人为信宿,因此,视频系统的输出显示必须根据人的接收系统——眼睛的物理特性来设计,这些物理特性包括了光感、频率、时间和结合作用,依据人眼对这些物理特性的独立或组合的视觉效应,建立合适的视觉模型,以利于不同应用领域的视频压缩编码处理。

1.1.1 人类视觉系统的频率响应

研究人眼视觉系统(HVS; Human Visual System)的物理特性,如何对视频信号的时间和频率作出响应是及其重要的。因为,作为视频系统最终的观看者,在特定的阈值条件下,人眼视觉系统存在对视频信号的不可见性,这些特性对视频压缩处理起到了至关重要的作用。在这里,我们主要研究人眼对图像亮度的时间和空间变化响应。Kelly 等人已经通过实验验证,人眼视觉系统对图像的敏感度主要是由图像的时间和空间频率所决定的,在两大参数下的不可感知性我们称为视觉隐盖特性。其中,视觉敏感度在空间和时间频率上的效应曲线基本一致,开始的阶段是随着时空域的上升而线性提高,在某些范围内视觉敏感度达到最高值,然后是一个单调迅速下降的过程,最后在某些频率段上退隐消失。这些隐退频率段我们称为截止频率,它是表示人眼视觉系统在空间和时间变化的分界线,在该分界线以上,对于人眼是不可见的。视觉截止频率响应的不可见特性在设计数字视频系统中是非常重要的。例如,时间和空间截止频率是设计视频数据采集和图像信号显示的帧率和行率的基础条件,同时,时空截止频率又决定了物理



设备时空显示的代价关系。下面,我们首先描述人眼视觉系统的空间和时间频率响应的视觉阈值,并介绍了当人眼跟踪运动物体时视觉频率的响应问题。

1.1.2 时间频率响应的视觉阈值

人眼视觉系统的时间频率响应是指在不同频率的时间变化图像的视觉敏感度。在不少文献中已经叙述了有关时间频率响应的影响因素,包括了观测距离、显示亮度、运动状态、背景复杂度和环境亮度等,并已经进行了很多实验来确定人眼视觉系统的时间频率响应。如图 1.1 所示,该图为 Kelly^[5]所做的实验得到的结果,为观测者提供了一个平面屏幕。通过该实验结果分析,可以把时间频率响应的亮度参数表示为以下形式:

$$\psi(t) = L(1 + m\cos 2\pi ft) \quad (1.1.1)$$

该式表明亮度值是与正弦变化成正比关系,其中 L 为图像平均亮度级,一般为常量, f 为闪烁频率, m 为调制参数,为变化参数。在实验中,我们要确定一个最低调制参数 $m_{\min}$,在 $m_{\min}$ 作用下,屏幕亮度的时间变化(即闪烁频率)对人眼刚好变得不可觉察。同时,我们把 $m_{\min}$ 的倒数表示为人眼对于给定闪烁频率 f 条件下的时间变化的敏感度, $1/m_{\min}$ 也称为对比敏感度,用于表示人眼视觉系统时间频率响应的视觉敏感度。

由图 1.1 的人眼视觉系统的时间频率响应曲线图,我们可以看到,引入图像的时间变化激励时,人眼视觉系统的响应相当于一个带通滤波器的响应,在低频段开始缓慢上升,在某个中间频率点达到峰值,在高频率段响应迅速下降,最后在截止频段隐退。其中,带通滤波器的峰值随着图像的平均亮度而增加,在实验中已经画出了响应的变化曲线,我们分析两个极端的现象,当平均亮度为最小时即 $L=0.06$,峰值响应出现在大约 5Hz 的频段,截止频率相应的也落在比较小的频段范围,大约是 10~20Hz。在最高的平均亮度时即 $L=9300$ 等阑时,响应大约在 20Hz 左右最高,大约在 78Hz 的频段退隐。

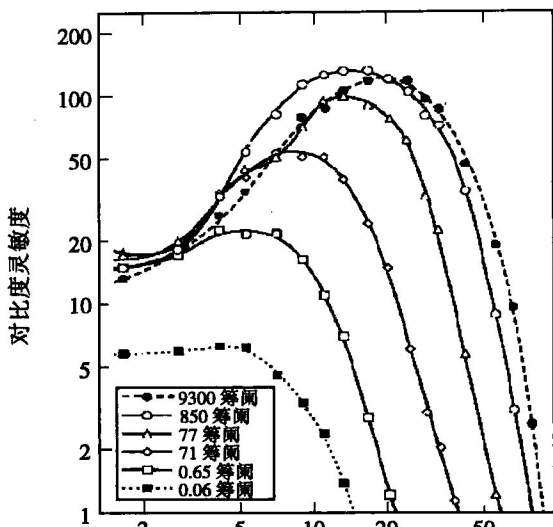


图 1.1 由视觉实验获得的 HVS 的时间频率响应

从人眼视觉系统的带通滤波器的时间响应曲线可以看出,时间频率响应的对比敏感度会在一定的频率下隐退,那么这个隐退的频率我们称为截止频率。不同的平均亮度值对应于不同的时间截止频率。从人眼视觉物理意义来分析,该时间截止频率决定了人眼对图像感知的暂留效应,即人眼在较高的时间频率处对比敏感度隐退的一个原因是人眼对图像的感知能保持一定时间距离,即实际图像已经移出视线范围,而人眼感觉图像还在,这种现象称为视觉暂留。也就是说,当图像的时间频率比人眼视觉系统的刷新率大时,会引起被观察图像在时间轴上的模糊不清晰;相反,当图像的时间频率比人眼视觉系统的刷新率小的时候,视觉暂留使视频信号显示为一个连续的帧序列。从图像序列的播放时间周期来对比,只要帧的时间间隔比视觉暂留周期短,则人眼视觉系统感觉到的是一个连续变化的图像序列;当帧的时间间隔比视觉暂留周期长时,人眼就会觉察到图像帧的闪烁,这种现象是指不连续帧