

新一代 视频压缩编码标准 ——H.264/AVC

(第二版)

毕厚杰 王 健 ■ 主编

 人民邮电出版社
POSTS & TELECOM PRESS

新一代 视频压缩编码标准 ——H.264/AVC

(第二版)

毕厚杰 王 健 ■ 主编

人 民 邮 电 出 版 社
北 京

图书在版编目 (C I P) 数据

新一代视频压缩编码标准——H.264/AVC / 毕厚杰,
王健主编. — 2版. — 北京: 人民邮电出版社,
2009.11

ISBN 978-7-115-21436-2

I. ①新… II. ①毕… ②王… III. ①视频信号—频
率压缩—图象编码—国际标准 IV. ①TN941.1-65

中国版本图书馆CIP数据核字(2009)第164506号

内 容 提 要

H.264 视频编码标准被称为新一代视频编码标准。本书在介绍数字视频和视频编码基本原理的基础上,系统阐述了 H.264/AVC 标准的特点、编码器原理、解码器原理以及码流的句法和语义。根据音视频编码技术和应用发展的特点,还分别论述了 H.264 视频编码传输在移动通信中的应用、H.264 的高保真范围扩展 FExt 和 H.264 的可伸缩编码 SVC。

本书取材新颖、内容全面,重点论述了 H.264 视频编码标准,还介绍了已有的若干视频编码国际标准。本书可作为高等院校通信、广播电视专业本科生教材或教学参考书,也可供该领域的硕士生、博士生深入研究参考,还可作为视频技术、视频服务领域广大技术人员的必备参考书。

新一代视频压缩编码标准——H.264/AVC (第二版)

- ◆ 主 编 毕厚杰 王 健
责任编辑 杨 凌
- ◆ 人民邮电出版社出版发行 北京市崇文区夕照寺街 14 号
邮编 100061 电子函件 315@ptpress.com.cn
网址 <http://www.ptpress.com.cn>
北京隆昌伟业印刷有限公司印刷
- ◆ 开本: 787×1092 1/16
印张: 18.75
字数: 462 千字
印数: 9 401—13 400 册
- 2009 年 11 月第 2 版
2009 年 11 月北京第 1 次印刷

ISBN 978-7-115-21436-2

定价 49.00 元

读者服务热线: (010)67129264 印装质量热线: (010)67129223
反盗版热线: (010)67171154

前言（第二版）

2003年3月，ITU-T/ISO的联合项目组JVT发布了视频编码标准H.264/AVC，由于其优越的压缩比以及较强的网络适应能力，H.264/AVC获得了日益广泛的应用。为了进一步扩大H.264标准的应用范围，使其适合应用于高保真视频压缩，JVT于2004年7月增加了重要扩展FRExt部分；为了更好地适应各种不同的网络环境和用户终端的网络业务应用，JVT最终于2007年6月确定了基于H.264架构扩展的可伸缩编码SVC标准。

本书（第二版）正是在前版的基础上修订而成，根据视频编码技术和业务发展的要求，在修订过程中主要考虑如下：完善和优化H.264/AVC标准的基本技术的部分内容，增加H.264视频编码传输及其在移动通信中的应用，增加H.264/AVC的新扩展部分FRExt和SVC。

本书（第二版）的第8章为新增加的内容，主要论述H.264视频编码传输及其在移动通信中的应用，讨论了H.264在无线传输中的应用和它的容错技术，并以中国移动正在运营的“万花筒”业务为例，增加了H.264应用于移动互联网的多媒体业务平台的设计以及移动无线信道中的精确码率控制算法，这一章由马国强、王健编写。

由于个人视频应用、广播通信、高清电视的应用、数字存储、远程医疗等的需求增多，人们期望将彩色格式由目前的4:2:0提升到4:4:4，以提高色彩的逼真度，对数字高清电视的清晰度期望提高到4k（3840×2160@60），甚至8k（7680×4320@60），比特深度期望由8bpp提升到12bpp，因而视频信号的传输码率将大幅度增加。于是人们希望进一步提高视频编码压缩比，及网络适应能力。这样一来就提出了H.265视频编码标准问题（其压缩比至少比H.264提高了1倍），为了适应这种形势，ITU-T JVT在2003年5月完成H.264标准基本部分的基础上，于2004年7月完成了高清扩展部分FRExt，它包括了HP、Hi10P、Hi422P、Hi444P四个档次。因此本书（第二版）中增加了第9章高质量视频编码，由毕厚杰编写。

可伸缩编码（SVC），多年来一直受到广泛关注。SVC可以实现一次压缩后，视频能以不同码率、不同帧率、不同空间分辨率、不同SNR进行解码，这种SVC适用于不同用户、不同信道的不同需求，具有相当大的灵活性。JVT于2007年6月确定了最终的SVC标准，为此本书（第二版）增加了第10章H.264可伸缩编码，由王健、马国强编写。

相对于第一版的内容而言，在本书（第二版）的第4章视频编码标准简介中，增加了MPEG-4、JPEG2000以及AVS标准的简要介绍，删除了第一版中第5章MPEG-4压缩编码标准。针对第一版中的一些错误和缺陷也进行了更正和补充，这些工作主要由王健承担。

全书由毕厚杰作了校对和审核。

鉴于上述工作，本书（第二版）的主编为毕厚杰和王健。马国强博士参与了部分内容撰写。

本书编写过程中，得到了南京大学季晓勇教授的大力支持，多媒体通信与信息安全教研室的杨晓琴、高民芳等研究生参与了基本部分的错误更正，特此表示衷心感谢。欢迎广大读者进一步提出修改、补充意见。

编者

2009年6月25日

前 言

数字视频技术在通信和广播领域获得了日益广泛的应用，特别是 20 世纪 90 年代以来，随着 Internet 和移动通信的迅猛发展，视频信息和多媒体信息在 Internet 和移动网络中的处理和传输技术成为了当前我国信息化中的热点。

众所周知，视频信息具有一系列优点，如直观性、确切性、高效性、广泛性等等。但是视频信息的信息量太大，要使视频得到有效的应用，必须首先解决视频压缩编码问题，其次解决压缩后视频质量保证的问题。这两者是相互矛盾的，是矛盾的两个方面。我们的任务是既要有一定的压缩比，又要保证一定的视频质量。

为此，人们付出了巨大的辛勤的劳动，现已结出丰硕的成果。从 1984 年 CCITT 公布第一个视频编码国际标准以来，至今已有 20 年了。ITU-T 等国际标准化组织陆续颁布了接近十个视频编码国际标准，大大推动了视频通信和数字电视广播的发展，这也是有目共睹的事实。但是严格地讲，视频通信和数字电视广播这两大领域至今的发展仍不能令人满意，总起来讲，应用的范围不广，主要是视频压缩与质量之间的矛盾不能很好解决。例如，可视电话一直被认为是一种理想的通信设备，可近 30 年来至今未能普及，就是因为性能价格比不高。

2003 年 3 月，ITU-T/ISO 正式公布了 H.264 视频压缩标准，由于其具有比以往标准更出色的性能，被人们称为新一代视频编码标准。具体讲，与 H.263 或 MPEG-4 相比，在同样质量下，其数码率能降低一半左右；或者说在同样码率下，其信噪比明显提高。这样一来，H.264 标准在国际上受到了广泛地重视和欢迎。在这样的背景下，我们编写了这一本书，对其进行详细介绍。

本书的特点是取材新颖、内容全面。它不仅重点论述了 H.264，而且首先介绍了数字视频和视频编码的基础知识，介绍了已有的若干视频编码国际标准（特别是 MPEG-4），以便为进一步学习 H.264 打下良好的基础。

全书共 9 章，在 H.264 部分（第 6~9 章）详尽地论述了 H.264 的特点、编码器原理、解码器原理、编解码器的实现方案。为了更好地理解 H.264 编解码器的原理及其实现方案，第 7 章详细介绍了 H.264 码流的句法和语义。最后对 H.264 视频编码传输的 QoS 进行了专门地论述。

本书可作为通信、广播电视专业高校本科生教材，并可供该领域的硕士生、博士生深入研究用，也可供广大的视频技术、视频服务领域的技术人员参考用。

本书由毕厚杰教授主编，撰写了第 1~4 章及第 6 章前 3 节。左雯撰写了第 5 章。马国强、徐苏珊撰写了第 7 章。焦良葆、王健撰写了第 9 章。其余部分由方晖、焦良葆、王健、马国强、左雯、李涛、徐苏珊、鹿宝生等人共同编写完成。全书由毕厚杰负责审稿，左雯负责校对。

本书编写过程中，得到了南京大学章德教授、陈启美教授的多方支持，特此深表谢意。本书的编写也引用了不少参考文献，特向这些作者表示诚挚的谢意。

由于时间仓促及水平有限，书中难免出现不当之处，恳请广大读者批评指正，以便再版时进一步修正。

编者

目 录

第 1 章 绪论	1
1.1 信息化与视频通信	1
1.1.1 什么是信息	1
1.1.2 什么是信息化	1
1.1.3 我国的信息化和视频通信	2
1.2 视频信息和信号的特点	2
1.2.1 直观性	2
1.2.2 确定性	2
1.2.3 高效性	2
1.2.4 广泛性	3
1.2.5 视频信号的高带宽性	3
1.3 视频压缩编码的要求和可能性	3
1.3.1 视频压缩编码的目标	3
1.3.2 视频压缩的可能性	3
1.4 视频压缩编码技术综述	4
1.4.1 基本结构	4
1.4.2 基于波形的编码	5
1.4.3 基于内容的编码	5
1.4.4 立体(三维)视频编码	6
参考文献	6
第 2 章 数字视频	7
2.1 数字电视的基本概念	7
2.1.1 数字电视的优越性	7
2.1.2 数字电视的 PCM 原理	8
2.2 数字电视信号	11
2.2.1 电视信号的时间和空间取样	11
2.2.2 彩色空间	12
2.2.3 彩色电视取样格式	12
2.2.4 数字电视信号的编码参数	13
2.3 视频信号的预处理	14
2.3.1 色彩插值(Color Interpolation)	14
2.3.2 色彩校正(Color Correction)	15
2.3.3 伽马校正(Gamma Correction)	15
2.3.4 图像增强(Image Enhancement)	17

2.3.5 白平衡 (White Balance)	20
2.4 视频质量	22
2.4.1 主观质量的评定	22
2.4.2 客观质量的测量	22
参考文献	23
第3章 视频压缩编码的基本原理	24
3.1 预测编码	24
3.1.1 预测编码的基本概念	24
3.1.2 帧内预测编码	25
3.1.3 帧间预测编码	29
3.1.4 运动估计	32
3.2 变换编码	43
3.2.1 变换编码的基本概念	43
3.2.2 K-L 变换	44
3.2.3 离散余弦变换 (DCT)	44
3.2.4 锯齿形扫描和游程编码	46
3.3 变换编码与预测编码的比较	47
3.4 熵编码	47
3.4.1 变长编码	47
3.4.2 算术编码	48
参考文献	49
第4章 视频编码标准简介	51
4.1 视频编码发展简史	51
4.2 H.261 标准	51
4.2.1 图像格式	52
4.2.2 H.261 视频编解码器	53
4.3 H.263 标准	56
4.3.1 H.263 标准图像格式	56
4.3.2 H.263 视频信源编码算法	57
4.4 MPEG-1 标准	57
4.4.1 功能	57
4.4.2 图像类型和编码结构	57
4.5 MPEG-2 标准	58
4.5.1 MPEG-2 编码复用系统	58
4.5.2 MPEG-2 档次和级别	58
4.5.3 MPEG-2 视频编码器	59
4.6 MPEG-4 标准	61
4.6.1 MPEG-4 标准构成	61
4.6.2 基于 VOP 的视频编码	63
4.6.3 MPEG-4 档次和级别	66

4.7 JPEG 标准	67
4.8 JPEG2000 标准	69
4.8.1 JPEG2000 标准的组成和特点	69
4.8.2 JPEG2000 核心编解码器	70
4.9 AVS 标准	72
4.9.1 AVS 标准进展情况	72
4.9.2 AVS 视频编码及关键技术	72
参考文献	75
第 5 章 H.264/AVC 编码器原理	76
5.1 概述	76
5.2 H.264/AVC 编解码器	77
5.2.1 H.264 编解码器特点	77
5.2.2 H.264 编码器	78
5.2.3 H.264 解码器	78
5.3 H.264/AVC 的结构	78
5.3.1 名词解释	78
5.3.2 档次和级	79
5.3.3 编码数据格式	79
5.3.4 参数图像	81
5.3.5 片和片组	81
5.4 帧内预测	83
5.4.1 4×4 亮度预测模式	84
5.4.2 16×16 亮度预测模式	86
5.4.3 8×8 色度块预测模式	86
5.4.4 帧内预测模式的选择和编码	87
5.5 帧间预测	89
5.5.1 可变尺寸块运动补偿	89
5.5.2 运动矢量	90
5.5.3 MV 预测	91
5.5.4 B 片预测	92
5.5.5 加权预测	95
5.6 H.264 的 SP/SI 帧技术	95
5.6.1 SP/SI 帧的应用	96
5.6.2 SP/SI 帧的基本原理	98
5.6.3 实验结果和性能分析	101
5.7 整数变换与量化	103
5.7.1 整数变换	103
5.7.2 量化	106
5.7.3 DCT 直流系数的变换量化	108
5.8 CAVLC (基于上下文自适应的可变长编码)	109

5.8.1	熵编码的基本原理	109
5.8.2	CAVLC 的基本原理	110
5.8.3	CAVLC 的上下文模型	110
5.8.4	CAVLC 的编码过程	110
5.8.5	CAVLC 的解码过程	112
5.8.6	CAVLC 编解码过程实例	114
5.8.7	CAVLC 与 UVLC 比较	115
5.9	CABAC (基于上下文的自适应二进制算术熵编码)	117
5.9.1	自适应算术编码	117
5.9.2	上下文模型	120
5.9.3	对输入流预编码	122
5.9.4	初始化	123
5.9.5	结论	123
5.10	码率控制	124
5.10.1	基于 Lagrangian 优化算法的 H.264 编码控制模型	124
5.10.2	实验结果和性能分析	126
5.11	去方块滤波	129
5.11.1	去方块滤波基本概念	129
5.11.2	边界分析	130
5.11.3	滤波过程	132
5.12	其余特征	135
5.12.1	参考图像的管理	135
5.12.2	重排序	135
5.12.3	隔行视频	135
5.12.4	数据分割片	136
5.12.5	H.264 传输	136
	参考文献	137
第 6 章	H.264 的句法和语义	140
6.1	句法	140
6.1.1	句法元素的分层结构	140
6.1.2	句法的表示方法	143
6.2	句法表	145
6.3	语义	158
6.3.1	NAL 层语义	158
6.3.2	序列参数集语义	160
6.3.3	图像参数集语义	162
6.3.4	片头语义	164
6.3.5	参考图像序列重排序的语义	168
6.3.6	加权预测的语义	169
6.3.7	参考图像序列标记 (Marking) 操作的语义	169

6.3.8 片数据的语义	170
6.3.9 宏块层的语义	171
6.3.10 宏块预测的语义	175
6.3.11 子宏块预测的语义	175
6.3.12 用 CAVLC 方式编码的残差数据的语义	176
6.3.13 用 CABAC 方式编码的残差数据的语义	176
6.4 总结	177
参考文献	177
第 7 章 H.264/AVC 解码器的原理和实现	179
7.1 解码器原理	179
7.2 NAL 单元	180
7.2.1 NAL 单元结构	180
7.2.2 NAL 单元解码过程	180
7.3 图像序列号的计算	181
7.3.1 图像序列号 (POC)	181
7.3.2 POC 类型为 0 的 POC 计算	183
7.3.3 POC 类型为 1 的 POC 计算	183
7.3.4 POC 类型为 2 的 POC 计算	185
7.4 宏块片组映射图的产生	186
7.5 片数据分割的解码	188
7.6 参考图像列表的初始化	189
7.6.1 图像序号的计算	189
7.6.2 参考图像列表的初始化	189
7.6.3 参考帧列表的重排序	193
7.7 解码的参考图像的标记过程	195
7.7.1 frame_num 不连续的解码过程	196
7.7.2 参考图像滑窗标记过程	196
7.7.3 参考图像的自适应内存控制标记过程	197
7.8 帧内预测	199
7.8.1 4×4 亮度块预测方式的推导	199
7.8.2 4×4 亮度块的帧内预测编码方式	200
7.8.3 16×16 亮度块的帧内预测方式	204
7.8.4 8×8 色度块的帧内预测方式	205
7.9 帧间预测解码处理	207
7.9.1 MV 分量及参考索引获取	208
7.9.2 帧间预测像素解码处理	210
7.10 变换系数解码	216
7.10.1 变换系数逆扫描过程	217
7.10.2 DCT 变换系数中直流系数的逆变换量化	217
7.10.3 残差变换系数的反量化	219

7.10.4	残差变换系数的逆 DCT 变换	219
7.10.5	去方块滤波前的图像恢复与重建	219
7.11	SP 片中的 P 宏块和 SI 片中的 SI 宏块的解码过程	220
7.11.1	主 SP 片中 P 宏块的解码过程	220
7.11.2	辅 SP/SI 片的解码过程	221
	参考文献	223
第 8 章	H.264 视频编码传输及其在移动通信中的应用	224
8.1	概述	224
8.2	H.264/AVC 在无线通信中的传输	224
8.2.1	H.264/AVC 在无线通信中的应用	224
8.2.2	H.264 分层传输结构	225
8.2.3	H.264 分组交换业务传输协议栈	226
8.3	H.264 视频编码的容错技术	227
8.3.1	H.264 的视频编码层的错误恢复	228
8.3.2	H.264 的网络提取层的错误恢复	229
8.4	移动无线信道中的精确码率控制算法	232
8.4.1	宏块能量的提取	232
8.4.2	与宏块能量相结合, 在宏块级别分配码率	234
8.4.3	计算编码量化参数 Qp	234
8.5	移动互联网中多媒体业务平台的设计	235
8.5.1	移动互联网的多媒体业务核心平台	235
8.5.2	编转码系统	236
8.5.3	设计流分发服务器	236
8.5.4	客户端	238
8.5.5	系统小结	238
	参考文献	238
第 9 章	高质量视频编码	241
9.1	H.264+与 H.265	241
9.1.1	视频编码的需求	241
9.1.2	视频编码标准及其应用简史	241
9.1.3	H.264+和 H.265 最近研究方向	241
9.1.4	H.265 主要特征与 KTA	242
9.1.5	H.265 的几个方向性课题	244
9.2	H.264 的高保真范围扩展 FRExt	244
9.2.1	引言	244
9.2.2	自适应分块	245
9.2.3	空间域的残差变换	246
9.2.4	帧内预测方法	246
9.2.5	基于视觉的量化缩放矩阵	247
9.2.6	色度空间的残差色度变换	247

9.2.7 无失真宏块模式	247
9.2.8 FRExt 的测试结果	247
9.3 本章结论	248
参考文献	248
第 10 章 H.264 的可伸缩编码 SVC	249
10.1 概述	249
10.2 SVC 的基本原理	250
10.2.1 传统可伸缩编码	251
10.2.2 精细粒度可伸缩编码	253
10.2.3 渐进精细粒度可伸缩编码	253
10.3 H.264 的 SVC	254
10.3.1 空域可伸缩性	255
10.3.2 时域可伸缩性	257
10.3.3 质量可伸缩性	259
10.3.4 H.264 的扩展	260
10.4 性能与分析	262
参考文献	263
缩略语	266
附录一 CAVLC 相关码表	270
附录二 CABAC 相关码表	275
附录三 H.264 档次和级	283

第 1 章 绪 论

1.1 信息化与视频通信

本书在讨论视频编码之前，先简要介绍一些与信息化有关的问题，研讨当前信息社会背景，然后讨论信息化与视频通信的关系。

1.1.1 什么是信息

众所周知，人类社会的三大支柱是物质、能量和信息。具体而言，农业现代化的支柱是物质，工业现代化的支柱是能量，而信息化的支柱是信息。

广义地讲，信息就是客观世界的描述和分析，它无所不在、无时不在，具有广泛性和通用性，这是信息的一个特性。信息没有重量，没有长度，具有抽象性，但它确实存在，这是信息的另一个特性。

信息的第三个特性是无限性。例如，关于物质的信息，物质具有无限的不可分性，物质由分子组成，分子由原子组成，下面还有中子、质子、电子、中微子等；关于通信网络的信息，为了增加通信容量，最初一对电话线只能通一路电话，后来利用 N-ISDN 技术，可在一对电话线上同时通两路电话，以后又发展 B-ISDN、ATM、IP、MPLS，直到今天，通信网络技术仍在不断地发展，应该说，它具有无限性。

总之，信息具有通用性、抽象性、无限性三个特征，其中，无限性最重要。

从以上描述可知，信息是很有用的，它是客观世界中最本质的客观规律描述和分析，是人类社会可用的重要资源。信息资源如果能被充分开发和利用，人类社会的物质和精神文明水准将大大提高。

地球上的物质资源是有限的。石油、煤等在一定期限内总会开采完，但信息资源是无限的，对客观规律性的探讨是无尽的。而信息是要通过不断研究、不断分析，通过科学研究、反复实践才能掌握的。

1.1.2 什么是信息化

信息化是这样一个过程：“在现代信息技术广为普及的基础上，通过开发和利用信息资源，各种社会和经济活动的功能和效益得以显著提高。人类社会的物质和精神文明达到了一个新的水平。”

可见，信息化的目的是提高人类的社会效益和经济效益，而实现信息化的关键在于开发和利用各种信息资源。

信息化的实现依赖于图 1.1 所示的信息系统来构建。

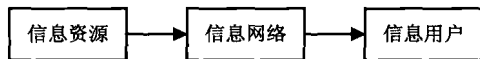


图 1.1 信息系统

就规模而言，我国已成为世界第一的通信大国。截止到 2007 年年底，固定电话的普及率已达到 27.8%，移动电话的普及率已达到 41.6%。宽带网络正在不断建设中，八纵八横的光纤构成了我国骨干网。以 MSTP（多业务传输平台）技术为主的城域网正在大力建设，以 ADSL 为主的宽带接入网也在迅速发展。应该指出，我国信息化的瓶颈是信息资源的开发和利用。

信息资源中, 视频信息的开发、利用更具有重要的理论意义和应用价值。

20 世纪 90 年代 Internet 的迅猛发展, 推动了 IP 技术应用普及到各个方面, 网上教育、远程医疗、电子商务、电子政务、电子游戏、网上证券等如雨后春笋。一句话, IP 视频通信已十分引人注目。

为了更深入地理解视频通信，有必要对它的特点进一步探讨。

利用人的视觉获取的信息称为视频信息,它具有直观性的特点。话音信息则是利用人的听觉获取的信息。两者相比,视频信息给人印象深刻、具体,话音信息则相对较浅显。从信息交流的客观效果讲,视频信息的效果更好。

视频信息直观具体，不易和其他内容相混淆，能保证信息的准确性。而话音信息则会由于地方口音的不同产生歧义，导致不必要的损失。

利用视觉，人们可以并行地观察一幅图像的各个像素，因而获取视频信息的效率比音频信息高得多。例如，通过一幅电机构建的图，人们可以很快搞清楚定子、转子及其相关位置，从而很快弄清电机的结构及其原理；如果人们没有这样的图，只是一味地听讲，通过音频信息去反复理解电机结构，仍可能搞不清楚，其接受的效率要低得多。

1.2.4 广泛性

据统计,人们每天通过视觉获取的信息占外界信息总量的70%左右。也就是说,人们每天获得的信息大部分是视觉信息。

1.2.5 视频信号的高带宽性

视频信息的表示形式是视频信号,通常为视频的电信号。视频信号通过网络传送至终端用户,并在屏幕上显示。

视频信号所包含的信息量大,其内容可以是活动的,也可以是静止的;可以是彩色的,也可以是黑白的;有时变化多、细节多,有时十分平坦。一般而言,视频信号信息量大,传输网络所需要的带宽相对较宽。例如,一路可视电话或会议电视信号,由于其活动内容较少,所需带宽较窄,但要达到良好的质量,不压缩需若干兆比特每秒,压缩后需要384kbit/s。又如,一路高清晰度电视信号(HDTV),由于其信息量相当大,不压缩需1Gbit/s,利用MPEG-2压缩后,尚需20Mbit/s。可见,视频信息虽然具有直观性、确定性、高效性等优越性能,但要传送包含视频信息的信号却需要较高的网络带宽,这就是为获得视频信息所需付出的代价。

1.3 视频压缩编码的要求和可能性

1.3.1 视频压缩编码的目标

如上所述,视频信号由于信息量大,传输网络带宽要求高,就像一辆体型巨大的货车只有在宽阔的马路上才能行驶一样。于是出现一个问题:能否将视频信号在传送前先进行压缩编码,即进行视频源压缩编码,然后在网络上进行传送,以便节省传送带宽和存储空间。视频信号压缩编码有两个要求:

- (1) 必须压缩在一定的带宽内,即视频编码器应具有足够的压缩比;
- (2) 视频信号压缩之后,经解压重建应保持一定的视频质量。

这个视频质量有两个标准:一个为主观质量,由人从视觉上进行评定;一个为客观质量,通常用信噪比(S/N)表示。

如果不考虑质量,一味地压缩,虽然压缩比很高,但压缩后信息严重失真,显然达不到要求;反之,如果只考虑质量,压缩比太小也不符合传送要求。

当然,在以上两个要求下,视频编码器的设计应力求简单、易实现、成本低、可靠性高,这也是其基本的要求。

1.3.2 视频压缩的可能性

1. 预测编码

前面讨论了视频信息的优越性、视频信号压缩的必要性,也提出了视频压缩的目标和要求,但实现这些目标的可能性如何?

众所周知,一幅图像由许多个称为像素的点组成,如图1.2中的“o”表示一个像素。大量的统计表明,同一幅图像的像素之间具有较强的相关性,两个像素之间的距离越短,其相关性越强,通俗地讲,即两个像素的值越接近。换言之,两个相邻像素的值发生突变的概率

极小, 相等、相似或缓变的概率则极大。

于是, 人们可利用这种像素间的相关性进行压缩编码。例如, 当前像素 X (设为立即传送的像素) 可用前一个像素 a 、 b 或 c , 或三者的线性加权来预测。 a 、 b 、 c 被称为参考像素。实际传送时, 把实际像素 X (当前值) 和参考像素 (预测值) 相减, 只传送 $X-a$, 到了接收端再把 $(X-a)+a$ 恢复成 X 。由于 a 是已传送的 (在接收端被存储), 于是得到当前值。由于 X 与 a 相似, $(X-a)$ 值很小, 视频信号被压缩, 这种压缩方式称为帧内预测编码。

不仅如此, 还可利用图 1.3 所示的帧间相关性进行压缩编码。由于邻近帧之间的相关性一般比帧内像素间的相关性更强, 因此压缩比也更大。

由此可见, 利用像素之间 (帧内) 的相关性和帧间的相关性, 找到相应的参考像素或参考帧作为预测值, 可以实现视频压缩编码。

2. 变换编码

大量统计表明, 视频信号中包含着能量上占大部分的直流和低频成分 (图中的平坦部分), 还有少量的高频成分 (图像的细节部分)。因此, 可以用另一种方法进行视频编码, 将图像经过某种数学变换后, 得到变换域中的图像 (如图 1.4 所示), 其中 u 、 v 分别是空间频率坐标轴。在图 1.4 中, 用 “o” 表示的低频和直流占图像能量中的大部分; 而用 “x” 表示的高频成分则是少量的; 其余均是零值, 用 “O” 表示。于是可用较少的码表示直流、低频以及高频, 而 “O” 则不必用码表示, 由此可完成压缩编码。

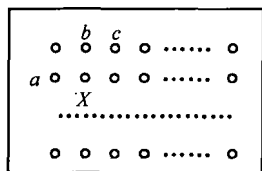


图 1.2 像素间相关性解释

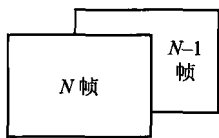


图 1.3 帧间相关性解释

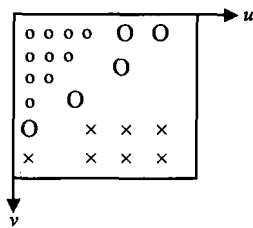


图 1.4 变换域图像

1.4 视频压缩编码技术综述

1.4.1 基本结构

视频编码系统的基本结构如图 1.5 所示。

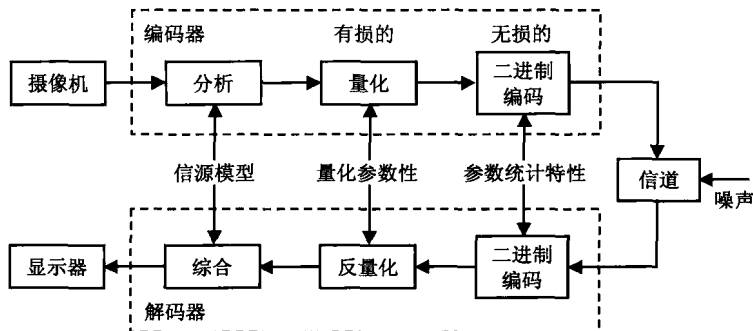


图 1.5 视频编码系统

由图 1.5 可见, 视频编码方法与可采用的信源模型有关。如果采用“一幅图像由许多像素构成”的信源模型, 这种信源模型的参数就是每个像素的亮度和色度的幅度值, 对这些参数进行压缩编码的技术称为基于波形的编码。如果采用“一个分量由几个物体构成”的信源模型, 这种信源模型的参数就是各个物体的形状、纹理和运动, 对这些参数进行压缩编码的技术称为基于内容的编码。

由此可见, 根据采用信源模型, 视频编码可以分为两大类: 基于波形的编码和基于内容的编码。它们利用不同的压缩编码方法, 得到相应的量化前的参数; 再对这些参数进行量化, 用二进制码表示其量化值; 最后, 进行无损熵编码进一步压缩码率。解码则为编码的逆过程。

1.4.2 基于波形的编码

如上所述, 利用像素间的空间相关性和帧间的时间相关性, 采用预测编码和变换编码技术可大大减少视频信号的相关性, 从而显著降低视频序列的码率, 实现压缩编码的目标。

基于波形的编码采用了把预测编码和变换编码组合起来的基于块的混合编码方法。

为了减少编码的复杂性, 使视频编码操作易于执行, 采用混合编码方法时, 首先把一幅图像分成固定大小的块, 例如块 8×8 (即每块 8 行, 每行 8 个像素)、块 16×16 (每块 16 行, 每行 16 个像素) 等, 然后对块进行压缩编码处理。

自 1989 年 ITU-T 发布第一个数字视频编码标准——H.261 以来, 已陆续发布了 H.263 等视频编码标准及 H.320、H.323 等多媒体终端标准。ISO 下属的运动图像专家组 (MPEG) 定义了 MPEG-1、MPEG-2、MPEG-4 等娱乐和数字电视压缩编码国际标准。

2003 年 3 月, ITU-T 颁布了 H.264 视频编码标准的基本部分。为适应专业级视频应用的需求, ITU-T 开始对 H.264 进行一系列的扩展。2004 年完成了保真度范围扩展 FRExt, 2006 年完成了可伸缩编码扩展 SVC。视频压缩与以往标准相比有了明显提高, 还具有良好的网络亲和性, 特别是对 IP 互联网、无线移动网等易误码、易阻塞、QoS 不易保证的网络视频传输性能有明显的改善。本书的第 1~8 章介绍 H.264 的基本部分, 第 9 章叙述 FRExt, 第 10 章介绍 SVC。

所有这些视频编码都采用了基于块的混合编码法, 都属于基于波形的编码。

1.4.3 基于内容的编码

如上所述, 基于块的编码易于操作, 但由于人为地把一幅图像划分成许多固定大小的块, 当包含边界的块属于不同物体时, 它们分别具有不同的运动方向, 不能用同一个运动矢量表示该边界块的运动状态。如果强制划分成固定大小的块, 这种边界块必然会产生高的预测误差和失真, 严重影响了压缩编码信号的质量。

于是产生了基于内容的编码技术。先把视频帧分成对应于不同物体的区域, 然后分别对其进行编码。具体说来, 就是对不同物体的形状、运动和纹理进行编码。在最简单的情况下, 利用二维轮廓描述物体的形状, 利用运动矢量描述其运动状态, 而纹理则用颜色的波形进行描述。

当视频序列中的物体种类已知时, 可采用基于知识或基于模型的编码。例如, 对人的脸部, 已开发了一些预定义的线框对脸的特征进行编码, 这种编码效率很高, 只需少数比特就能描述其特征。

对于人脸的表情 (如生气、高兴等), 可能的行为可用语义编码, 由于物体可能的行为