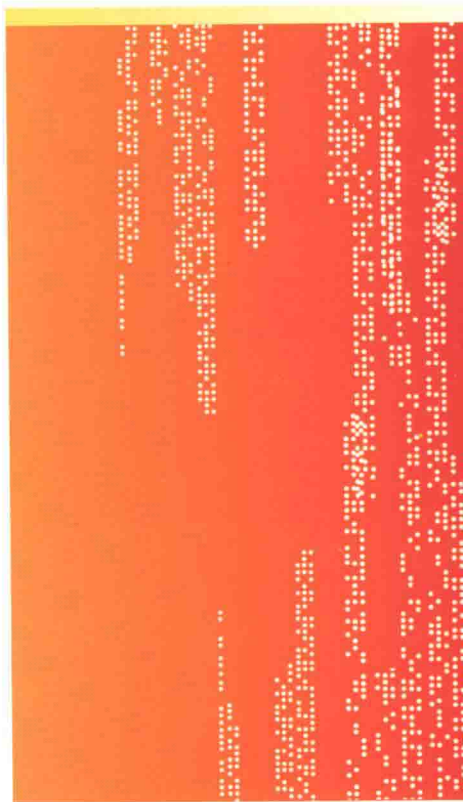




高端图像与视频新技术丛书

新一代高效视频编码 H.265/HEVC: 原理、标准与实现

► 万帅 杨付正 编著



H.265/HEVC是目前最新的视频编码国际标准，包含了最先进的视频编码技术。本书从原理、标准与实现三个角度出发，系统全面地介绍了H.265/HEVC视频编码标准。



电子工业出版社
PUBLISHING HOUSE OF ELECTRONICS INDUSTRY

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

新一代高效视频编码 H.265/HEVC: 原理、标准与实现

万 帅 杨付正 编著

電子工業出版社
Publishing House of Electronics Industry
北京 • BEIJING

内 容 简 介

本书系统全面地介绍了新一代视频编码标准 H.265/HEVC, 深入讲解了相关编码原理和实现方法。全书分为 12 章。第 1 章概述了视频编码国际标准和 H.265/HEVC 的发展历程, 介绍了 H.265/HEVC 的特色技术。第 2 章讨论了数字视频格式和 H.265/HEVC 的编码视频格式。第 3 章详细解析了 H.265/HEVC 编码结构以及参数集, 并介绍了 H.265/HEVC 的档次、层和级别。第 4~8 章为 H.265/HEVC 编码技术的分模块论述和语法语义解析, 包括预测编码、变换编码、量化编码、环路后处理和熵编码。为方便读者理解, 每个模块都包含了相应的背景知识、语法语义、实现方式等。第 9~12 章针对 H.265/HEVC 的实现和应用, 详细介绍了 H.265/HEVC 的网络适配层、并行处理技术、率失真优化和编码速率控制。

本书可作为电子信息类和广播电视类本科高年级学生和研究生等相关课程教材, 也可供视频技术研究领域内的研究生、教师、工程师参考, 并适合相关技术人员作为解读 H.265/HEVC 标准的参考。

未经许可, 不得以任何方式复制或抄袭本书之部分或全部内容。

版权所有, 侵权必究。

图书在版编目 (CIP) 数据

新一代高效视频编码 H.265/HEVC: 原理、标准与实现 / 万帅, 杨付正编著. —北京: 电子工业出版社, 2014.12

ISBN 978-7-121-24699-9

I. ①新… II. ①万… ②杨… III. ①视频编码 IV. ①TN762

中国版本图书馆 CIP 数据核字 (2014) 第 257744 号

责任编辑: 董亚峰 文字编辑: 王 纲 吴长莘

印 刷: 北京季蜂印刷有限公司

装 订: 北京季蜂印刷有限公司

出版发行: 电子工业出版社

北京市海淀区万寿路 173 信箱 邮编 100036

开 本: 720×1 000 1/16 印张: 25 字数: 409 千字

版 次: 2014 年 12 月第 1 版

印 次: 2014 年 12 月第 1 次印刷

定 价: 65.00 元



凡所购买电子工业出版社图书有缺损问题, 请向购买书店调换。若书店售缺, 请与本社发行部联系, 联系及邮购电话: (010) 88254888。

质量投诉请发邮件至 zltz@phei.com.cn, 盗版侵权举报请发邮件至 dbqq@phei.com.cn。

服务热线: (010) 88258888。

前 言

视觉是人类感知世界最重要的方式。中文说“眼见为实”，英文强调“Seeing is believing”，人们总是更乐于接受所能看到的信息，因此人类的科技一直不懈地致力于为人们提供更多、更好的用于“看”的信息。黑白电视的发明曾经带给人们巨大的震撼，很快，人们对于色彩的天然渴望，又促使彩色电视出现。进入数字时代之后，数字视频更是紧随着 IT 技术的浪潮，获得了非常迅速的发展。追求更高的清晰度，是数字视频技术领域从未停止的步伐。如今，各式各样的视频应用已经渗透到人类社会的各个领域，可以说，视频应用是现代人类社会运转的重要组成部分。

作为一种数据量非常巨大的信息载体，视频若想获得实际应用，必须采取高效的数据压缩和编码。自 20 世纪 80 年代以来，国际标准化组织一直在持续研究视频编码方法，并根据当时整体技术水平，制定相应的视频编码国际标准。每一次视频编码国际标准的颁布，都会从很大程度上促进视频技术的发展，催生更多的视频应用。相应地，视频应用的不断涌现也为视频编码提出了更高的要求，进而推动着视频编码标准向更高的压缩效率不断挺进。

近年来，智能移动终端的普及使得视频应用越来越多样化，涉及的视频内容丰富多样，立体视频应用也开始获得人们的关注。此外，高清晰度的显示设备越来越常见，人们对视频分辨率的要求也随之水涨船高。虽然网络和存储技术也在快速发展，但是视频数据量增长的速度更为惊人，大有“道高一尺魔高一丈”的态势。尽管现有的视频编码标准特别是 H.264/AVC 具有出色的压缩性能，但是已明显不能满足新型视频应用的需求。在这样的背景下，

国际电信联盟 ITU—T 与国际标准化组织 ISO/IEC 再次通力合作,于 2013 年发布了新一代的高效视频编码标准 (High Efficiency Video Coding, HEVC, H.265)。H.265/HEVC 包含着最新的视频编码技术,与它的前代 H.264/AVC 相比, H.265/HEVC 在同样的编码质量下能够节约 50%左右的码率。除了出色的压缩性能, H.265/HEVC 还包含特有的并行处理设计,其软硬件实现将具有更好的实用性。因此, H.265/HEVC 将于短期内迅速取代 H.264/AVC, 在各类视频业务中获得广泛的应用。

由于 H.265/HEVC 采用了大量编码新技术,相关标准的语法语义解读比较困难。本书作者根据自己长期研究视频编码的经验,对本书的内容进行了模块化的编排。在给出 H.265/HEVC 整体编码框架和其中的关键技术之后,针对 H.265/HEVC 的各个编码模块分别进行详尽分析。为方便读者理解,每个模块都包含了相应的背景知识、语法语义、实现方式等。全书分为 12 章。第 1 章概述了视频编码国际标准和 H.265/HEVC 的发展历程,介绍了 H.265/HEVC 的特色技术。第 2 章讨论了数字视频格式和 H.265/HEVC 的编码视频格式。第 3 章详细解析了 H.265/HEVC 编码结构以及参数集,并介绍了 H.265/HEVC 的档次、层和级别。第 4~8 章为 H.265/HEVC 编码技术的分模块论述和语法语义解析,包括预测编码、变换编码、量化编码、环路后处理和熵编码。第 9~12 章针对 H.265/HEVC 的实现和应用,详细介绍了 H.265/HEVC 的网络适配层、并行处理技术、率失真优化和编码速率控制。

本书可作为电子信息类和广播电视类本科高年级学生和研究生相关课程教材,也可供视频技术研究领域内的研究生、教师、工程师参考,并适合相关技术人员作为解读 H.265/HEVC 标准的参考。

本书的撰写受国家自然科学基金“基于感知失真度量的高效视频编码 (HEVC) 率失真优化研究”(批准号 61371089)的支持。感谢贺竞博士、李维博士、张泰、张新在本书撰写过程中在资料收集和 content 整理方面做出的辛勤工作。由于时间有限,书中的论述难免出现疏漏,恳请广大读者批评指正。作者联系邮箱: swan@nwpu.edu.cn。

作 者

2014 年 10 月 10 日

目 录

第1章 绪论	1
1.1 视频压缩与编码概述	2
1.1.1 视频	2
1.1.2 视频压缩与编码	3
1.2 视频编码标准	5
1.2.1 什么是视频编码标准	5
1.2.2 视频编码国际标准的发展	7
1.3 H.265/HEVC 简介	12
1.3.1 标准化历程	12
1.3.2 编码框架	13
1.3.3 特色编码技术	15
参考文献	18
第2章 数字视频格式	22
2.1 数字视频	22
2.1.1 颜色空间	23
2.1.2 量化深度	26
2.1.3 空间分辨率	27
2.1.4 时间分辨率	28
2.2 数字视频格式	29
2.2.1 色度亚采样	29
2.2.2 数字视频格式的规范标准	31

2.3 H.265/HEVC 编码视频格式	37
2.3.1 编码图像格式	37
2.3.2 解码图像格式	40
参考文献	46
第3章 编码结构	48
3.1 编码结构概述	48
3.2 视频参数集	51
3.2.1 视频层描述	51
3.2.2 视频层参数集	52
3.3 序列参数集	55
3.4 图像参数集	61
3.5 片段层	67
3.5.1 片与片段	67
3.5.2 片头语法语义	69
3.6 Tile 单元	75
3.6.1 Tile 单元描述	75
3.6.2 Slice 与 Tile	76
3.7 树形编码块	77
3.7.1 编码单元	78
3.7.2 预测单元	83
3.7.3 变换单元	86
3.8 档次、层和级别	88
3.8.1 档次	88
3.8.2 层和级别	90
参考文献	93
第4章 预测编码	94
4.1 视频预测编码技术	94
4.1.1 预测编码的原理	94
4.1.2 帧内预测编码	97
4.1.3 帧间预测编码	101
4.2 帧内预测	111

4.2.1	帧内预测模式	111
4.2.2	帧内预测过程	117
4.3	帧间预测	124
4.3.1	运动估计	124
4.3.2	MV 预测技术	130
4.3.3	加权预测	135
4.4	PCM 模式及相关语法语义	137
4.4.1	PCM 模式	137
4.4.2	相关语法语义	138
	参考文献	138
第 5 章	变换编码	142
5.1	离散余弦变换	143
5.1.1	DCT 原理及特点	143
5.1.2	整数 DCT	148
5.1.3	H.265/HEVC 中的整数 DCT	152
5.2	离散正弦变换	154
5.2.1	DST 原理	154
5.2.2	H.265/HEVC 中的整数 DST	156
5.3	变换结构	158
5.3.1	变换单元	158
5.3.2	相关语法语义	159
5.4	哈达玛变换	160
5.4.1	原理及特点	160
5.4.2	哈达玛变换的应用	161
	参考文献	163
第 6 章	量化	166
6.1	标量量化	167
6.1.1	基本原理	167
6.1.2	均匀量化	168
6.1.3	Lloyd-Max 量化器	170
6.1.4	熵编码量化器	171

6.2 H.265/HEVC 中的量化	173
6.2.1 量化	173
6.2.2 反量化	175
6.2.3 RDOQ	176
6.2.4 量化参数	178
6.2.5 量化矩阵	182
参考文献	184
第 7 章 环路后处理	187
7.1 去方块滤波	188
7.1.1 去方块滤波技术	190
7.1.2 去方块滤波的实现方法	199
7.1.3 语法语义	204
7.2 样点自适应补偿	206
7.2.1 SAO 技术	207
7.2.2 SAO 的实现方法	210
7.2.3 语法语义	217
参考文献	223
第 8 章 熵编码	226
8.1 熵编码基本原理	226
8.1.1 熵	226
8.1.2 变长编码	229
8.1.3 指数哥伦布编码	230
8.1.4 算术编码	232
8.2 零阶指数哥伦布编码及其应用	235
8.2.1 零阶指数哥伦布编码	235
8.2.2 H.265/HEVC 中的应用	237
8.3 CABAC	240
8.3.1 CABAC 的原理	240
8.3.2 H.265/HEVC 中的 CABAC	242
8.3.3 CABAC 在 H.265/HEVC 中的应用	249
8.4 变换系数熵编码	259

8.4.1	变换系数扫描	259
8.4.2	非零系数位置信息编码	261
8.4.3	非零系数幅值信息编码	263
	参考文献	265
第 9 章	网络适配层	269
9.1	分层结构	269
9.2	图像类型	272
9.3	网络适配层单元	276
9.3.1	NALU 载荷	277
9.3.2	NALU 头	279
9.4	视频比特流中的 NALU	282
9.4.1	介入单元	282
9.4.2	参数集	285
9.5	网络适配层单元的应用	286
9.5.1	字节流应用	286
9.5.2	分组流应用	287
	参考文献	291
第 10 章	编解码并行处理	293
10.1	视频编解码并行处理技术	294
10.1.1	并行处理基本概念	294
10.1.2	H.265/HEVC 编解码并行处理方式	296
10.1.3	H.265/HEVC 编码单元数据依赖关系	297
10.2	H.265/HEVC 标准中并行处理新技术	298
10.2.1	Tile	299
10.2.2	波前并行处理	301
10.2.3	依赖片	303
10.3	H.265/HEVC 编解码并行策略	303
10.3.1	GOP 级并行	304
10.3.2	图像级并行	305
10.3.3	Slice 级并行	306

10.3.4	Tile 级并行	306
10.3.5	CTB 级并行	308
	参考文献	310
第 11 章	率失真优化	312
11.1	率失真优化技术	313
11.1.1	率失真理论	313
11.1.2	视频编码中的率失真优化	322
11.2	H.265/HEVC 编码器中的率失真优化方法	327
11.2.1	视频图像组的率失真优化	328
11.2.2	片层的率失真优化	331
11.2.3	CTU 层率失真优化	332
11.2.4	CU 层率失真优化	333
11.2.5	PU 层率失真优化	334
11.2.6	编码参考模型 HM 中的率失真优化方法	335
	参考文献	342
第 12 章	速率控制	344
12.1	视频编码速率控制	344
12.1.1	速率控制的基本原理	344
12.1.2	缓冲机制	347
12.1.3	速率控制技术	348
12.2	H.265/HEVC 速率控制	355
12.2.1	目标比特分配	356
12.2.2	量化参数确定	359
	参考文献	361
附录 A	术语及英文解释	365
附录 B	缩写	384

第 1 章

绪 论

视觉是人们感知和认知外部世界的主要途径。实验心理学家赤瑞特拉通过大量实验证实人类获得信息的 80% 以上都来自视觉。所以中文谚语讲“眼见为实”，英文谚语也强调“**Seeing is believing**”，都是符合科学事实的。正因为如此，在这个信息化的时代，与视觉相关的应用往往受到用户极大地青睐。通信、娱乐、军事侦察、抗震救灾等，人们总是希望看到相应的动态影像，即视频信息。数字视频在这些应用中扮演着关键性的角色。

然而，人们在实际应用中接触到的视频，都是压缩过的视频。这是因为未经压缩的原始视频其数据量是非常惊人的，根本无法直接用于实际的传输或存储。因此，视频应用的一项关键技术就是视频编码（**Video Coding**），也称视频压缩，其目的是尽可能去除视频数据中的冗余成分，减少表征视频的数据量。本章首先剖析视频压缩与编码的基本概念，进而简要介绍视频编码标准的发展历程。本章最后一部分针对本书所关注的新一代高效视频编码 **H.265/HEVC**，从整体理念出发，给出其编码的基本架构和特征。

1.1 视频压缩与编码概述

1.1.1 视频

最初的视频信号是模拟的,最早是基于光电管及阴极射线管的电视系统产生的。但是模拟时代早已过去,如今我们所说的视频通常是指数字视频,从本质上讲是一系列内容连续的数字图像,按时间顺序排列而成。由于人眼有视觉暂留机理,连续播放的图像会形成平滑连续的视觉效果,当播放速度足够快时,人眼不再分辨出每一幅图像,而是在脑海中形成连续的视频。因此,图像是视频信号的基本单位。为了与静止图像相区别,视频中完整图像通常被称为帧(Frame),由许多帧按照时间顺序组成的视频也被称为视频序列(Video Sequence)。

视频序列中的每一幅图像,都是由 $N \times M$ 个像素(Pixel)组成的,每个像素都有具体的数值。因此,视频序列可以表示为三维矩阵,其中 $N \times M$ 这两个维度表示每幅图像的像素值,形成视频的空间域;另一个维度代表视频的时间域,如图 1.1 所示。这里需要注意的是,彩色的视频需要三个这样的矩阵,分别代表三个基本的色彩分量,或亮度和色度分量。此外,每秒播放的帧数目叫作帧率(Frame Rate),单位为 fps。为了使人眼能够有平滑连续的感受,视频的帧率需要达到 25~30fps 以上。这部分内容将会在第 2 章进行详细介绍。这部分内容属于视频技术的基础,请参见文献[1]。

从上文的分析可以看出,原始视频的数据量是非常巨大的。以大家常见的标清电影视频格式(720P)为例,假设 3 个色彩分量的每个像素均以 8 比特表示,帧率为 30fps,这样的每一秒的视频数据量达到 $1280 \times 720 \times 3 \times 8 \times 30 = 6.64 \times 10^8$ 比特。再大的硬盘,也存不了几部电影,更无法将原始视频数据在网络上进行传输。因此,视频应用的一项关键技术就是视频编码,也称视频压缩,其目的是尽可能去除视频数据中的冗

余成分, 减少压缩或编码后的数据量。

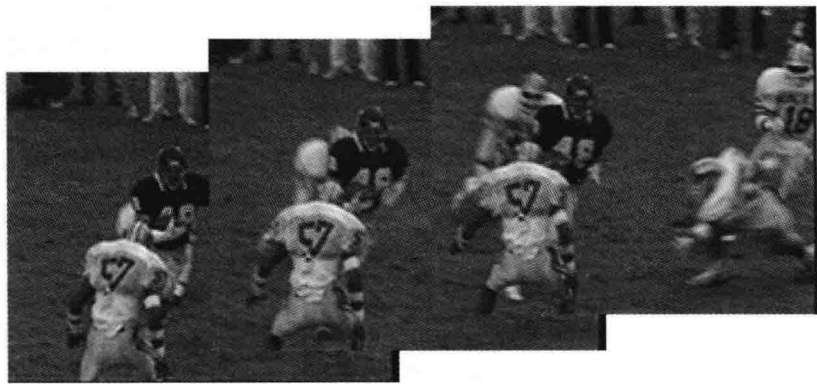


图 1.1 视频 Football

1.1.2 视频压缩与编码

视频压缩是一类特殊的数据压缩方法。数据是信息的载体, 对于定量的信息, 设法减少表达这些信息所用的数据量称为数据压缩。数据压缩通常分为无损压缩和有损压缩两大类。其中无损压缩是指数据经过压缩后, 所携带的信息并没有损失, 通过重建可以完全恢复原来的数据。无损压缩适用于数据需要严格完全重建的情形, 常用于对文本文件、程序文件等进行压缩 (例如压缩成的.zip 或.rar 的文本文件, 在解压重建后与原文件应是完全相同的)。在某些特殊应用场合, 也可以对音频或图像进行无损压缩, 例如需要完美音质的音乐制作, 用于精确诊断的医学图像, 来之不易的遥感图像等。然而, 受信源熵的限制, 无损压缩的压缩率普遍不高, 通常都在 5:1 以下, 对于图像的无损压缩来说, 压缩率以 3:1 左右最为常见^[2,3]。对于海量的原始视频数据来说, 这样的压缩率是远远不够的, 因此在绝大多数情况下, 视频压缩都采用有损压缩的方式。

有损压缩以引入一定失真为代价, 换取更高的压缩比。能够应用有损压缩的条件是人们对于引入的失真“无法察觉”或者“可以接受”。有损压缩的典型应用对象就是用于人类认知的音频、图像和视频。这是由于对于人耳或者人眼来说, 丢掉某些信息是很难察觉的。例如, 图像中

往往包含着许多细节, 这些细节在频域里表现为大量的高频信息。而人眼对于细节或者高频信息并不敏感, 在压缩时丢掉部分高频信息可能并不会被人眼察觉。以图 1.2 所示的图像为例, 图 1.2 (a) 经过压缩比为 8:1 的有损压缩获得图 1.2 (b), 这二者在视觉上差别较小。此外, 即使压缩产生的失真能够被人感知到, 但是如果不会影响人们对视频内容的理解, 人们也通常愿意接受质量稍差的音、视频或者图像, 以获取较高的压缩比。例如, 对比图 1.2 (d) 与图 1.2 (a), 能够明显看出字迹的模糊, 但并不影响人们对于图像内容的理解; 而此时, 我们能够获得更大的压缩比 (64:1)。在音视频和图像压缩算法中, 大量利用了人类的感知特性, 尽可能使压缩产生的失真发生在人不容易察觉到的地方。总的来说, 有损压缩能够获得比无损压缩高出许多的压缩比, 然而, 世上没有免费的午餐, 其代价就是必然在质量上产生损失。

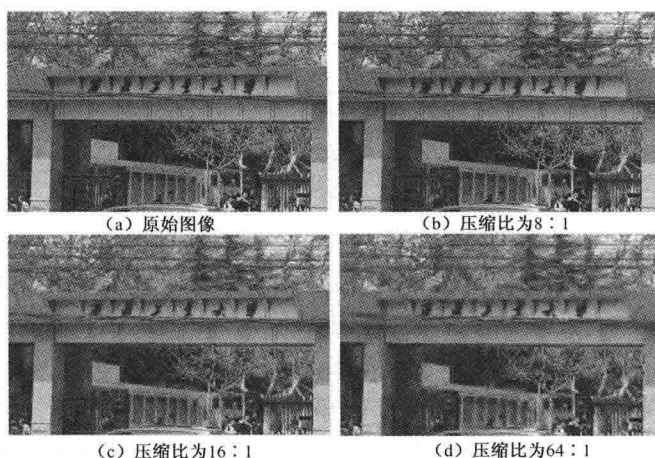


图 1.2 图像的有损压缩 (JPEG 2000)

为了获得较高的压缩率, 视频通常采用有损压缩, 即以损失一定质量的代价获取高压缩比。此时衡量压缩算法性能优劣与两个参数有关: 码率和失真。有损压缩追求的是, 在质量损失一定下获得最高的压缩比 (最低的码率); 或者在码率一定的条件下, 视频的质量最好。对于视频来说, 还应当考虑视频在时间域的质量, 也就是帧率的变化。视频的时间域失真常见于网络视频传输的场景, 传输中的视频数据遇到带宽变化,

容易在接收端产生停顿等令人观看不适的现象。

虽然近年来信息技术发展非常迅速,有线与无线网络的带宽都在不断提高,各类存储器的容量也在不断增长,但是与此同时,人们对于视频源保真度的要求也越来越高。如今,高清视频日渐普及,存储容量与网络带宽的增长始终无法满足人们对于存储和传输高分辨率视频的要求。因此,视频压缩与编码技术的进步和革新始终没有停歇。

在中文里,“视频压缩”和“视频编码”两个词常常被认为是等同的,被广泛交替使用。无论是视频压缩还是视频编码,通常都是指采用预测、变换、量化和熵编码等方式,尽可能减少视频数据中的冗余,使用尽可能少的数据来表征视频。但是从严格意义上讲,二者存在着细微的差别。视频压缩是“目的”,视频编码则更强调“手段”和“方法”。因此在讨论视频压缩的方法时,国际上通常采用“视频编码”这一说法,相应的标准也被称为“视频编码标准”。因此按照国际惯例,本书中均使用“视频编码”的说法。

1.2 视频编码标准

各式各样的视频应用从一开始就催生了多种视频编码方法。为了使编码后的码流能够在大范围内互通和规范解码,从20世纪80年代起,国际组织开始对视频编码建立国际标准。视频编码的国际标准通常代表着同时代最先进的视频编码技术。2006年,我国也形成了具有自主知识产权的视频编码标准AVS(Audio Video coding Standard)。目前国际上最新的视频编码标准就是本书所介绍的H.265/HEVC。

1.2.1 什么是视频编码标准

值得注意的是,视频编码标准只是规定了编码码流的语法规义和解码器,只要求视频编码后的码流符合标准的语法结构,解码器就可以根据码流的语法规义进行正常解码。因此,符合某个视频编码标准的编码器是有很大的自由度的,只要编码后的码流符合标准的规定即可。

在编码器输出的码流中,数据的基本单位是语法元素,每个语法元素由若干比特组成,它表征了某个特定的物理意义,如预测类型、量化参数等。视频编码标准的语法规定了各个语法元素的组织结构,而语义则阐述了语法元素的具体含义。编码器输出的比特码流中,每比特都隶属于某个语法元素,每个语法元素在标准中都有相应的解释。可见,视频编码标准规定了编码后码流的语法语义,也就阐明了从比特流提取语法元素并进行解释的方法,也就是视频的解码过程。

然而,在编码标准的制定过程当中,为了确定如何对语法元素进行合理的设计,首先要明确该标准所支持的编码方式,以及相应可能出现的编码方法。在标准的制定过程中,标准化组织会向业界广泛征集各类提案。这些提案当中,包含了大量编码新技术的设计,并会逐渐形成标准组织发布的参考软件(Reference Software),这些参考软件通常包含一整套标准的编解码器。由于参考软件的开发凝聚了广大科研人员的新思路,并且经过标准提案的多种性能测试,参考软件中的编码方法往往代表了当时先进的编码技术。因此,标准发布的参考软件不仅可用于标准开发过程中的测试和研究,也常常被科研人员作为研究先进视频编码的方法和平台,甚至作为商业开发的基础和参考。

目前,国际上制定视频编码标准的两大组织分别是国际电信联盟电信标准化部门(International Telecommunication Union-Telecommunication Standardization Sector, ITU-T^[4])与国际标准化组织(International Organization for Standardization, ISO)/国际电工委员会(International Electrotechnical Commission, IEC)。ITU-T 制定的视频编码标准通常被称为 H.26X 系列,包括 H.261、H.263 (H.263+、H.263++)等,这些标准被广泛应用于基于网络传输的视频通信,例如可视电话、会议电视等。ISO/IEC 的动态图像专家组(Moving Picture Experts Group, MPEG)制定了大名鼎鼎的 MPEG 系列视频编码标准,主要应用于视频存储(例如 VCD/DVD)、广播电视、网络流媒体等。值得一提的是,这两个组织曾经有过两次非常成功的合作。ITU-T 与 ISO/IEC 在视频编码标准中的首次合作形成了 H.262/MPEG—2 标准,成为风靡一时的 DVD 的核心技术。2003 年二者再次携手,开发了 H.264/AVC 视频编码标准,涵盖了包括视频广播、视频存储、交互式视频等各式各样的视频应用。新一代视频编