

| “十三五”国家重点出版物出版规划项目

| 虚拟现实与数字媒体技术前沿

# 数字视频 编码技术原理 (第二版)

高 文 赵德斌 马思伟 著

**PRINCIPLES OF  
DIGITAL VIDEO  
CODING  
TECHNOLOGY  
(Second Edition)**



科学出版社

“十三五”国家重点出版物出版规划项目  
虚拟现实与数字媒体技术前沿

# 数字视频编码 技术原理

(第二版)

Principles of Digital Video  
Coding Technology  
(Second Edition)

● 高文 赵德斌 马思伟 著



科学出版社

北京

## 内 容 简 介

本书力图沿着技术与标准发展的主线,向读者介绍视频编码各主要方面的技术原理和算法,引导读者了解和掌握视频编码技术的核心,为读者进一步的研究提供参考。主要内容包括数字视频的基本概念及编码基础,预测、变换、量化、熵编码等核心编码技术,MPEG、H.26x 和 AVS 系列标准简介,可伸缩编码、视频转码、三维视频编码、分布式编码及基于视觉特性的编码等新型编码方法、编码标准以及质量评价技术等。

本书可供计算机、通信专业高校师生参考,也可作为从事数字视频编码研究工作的科研人员的参考书。

---

### 图书在版编目(CIP)数据

数字视频编码技术原理 / 高文, 赵德斌, 马思伟著. —2 版. —北京: 科学出版社, 2018.3

(虚拟现实与数字媒体技术前沿)

“十三五”国家重点出版物出版规划项目

ISBN 978-7-03-056805-2

I. ①数… II. ①高… ②赵… ③马… III. ①数字视频系统—数字信号处理 IV. ①TN941.3

中国版本图书馆 CIP 数据核字(2018)第 048412 号

---

责任编辑: 任 静 / 责任校对: 郭瑞芝

责任印制: 张克忠 / 封面设计: 迷底书装

科 学 出 版 社 出 版

北京东黄城根北街 16 号

邮政编码: 100717

<http://www.sciencep.com>

中国科学院印刷厂 印刷

科学出版社发行 各地新华书店经销

\*

2010 年 10 月第 一 版 开本: 720×1 000 1/16

2018 年 3 月第 二 版 印张: 24

2018 年 3 月第一次印刷 字数: 484 000

定价: 145.00 元

(如有印装质量问题, 我社负责调换)

# 序

视频已与人们的生活密不可分，是人类获得信息的重要途径之一，因而与视频应用相关的各种技术问题都得到了广泛关注。在视频技术的发展过程中，从模拟到数字的转换是一次伟大的技术革新，但由于数字视频的原始数据量之大是存储空间和传输带宽都难以承受的，因此数字视频压缩成为数字视频领域的重要研究问题之一。

视频编码始于 20 世纪 60 年代。第一次学术讨论会于 1969 年在美国波士顿举行，会议名称叫 PCS (Picture Coding Symposium)。

对于视频压缩，如何消除视频数据中的冗余，成为视频压缩的关键所在。从 20 世纪 50 年代至今，视频编码技术不断发展，形成了以预测、变换和熵编码为核心的主流视频编码框架，并在此基础上制定了 MPEG、H.26X 等一系列编码标准。近几年，视频编码技术又有新的进展，自 H.264/AVC 以及 AVS 等新一代编码标准发布之后，新的编码技术能够再次大幅度提高编码效率，与此同时分布式视频、超高清视频、立体视频等新型应用也逐渐展开，并对视频技术提出了新的需求。这给我们提出许多问题，如视频压缩的潜力挖掘有无止境？对于新型应用如何设计高效的视频编码技术？对于这些问题，我们需要从原理上深入探究，进而找到解决问题的关键技术。这些问题都可以在该书中找到答案。

该书作者高文教授多年来一直从事视频编码领域教学和科研工作，成果很多，造诣很深。他还是 AVS 标准的主要缔造者，在算法研究、标准制定以及产业化发展方面都积累了丰富的经验。此次他和他的学生们将这些研究成果及个人心得在该书中结集出版，对于本领域的研究人员和工程师将大有裨益。该书内容丰富，文献翔实，将理论与技术密切结合起来，讨论问题具体深入，既可以给初学者以基础，又可以给专业人员以技能。在此特推荐该书作为高校本科生和研究生的教学用书，或研究人员和工程师的工具参考书。

黄熙韬

于伊利诺伊大学香槟分校

# 前言

自本书第一版出版至今大约六年的时间里,视频编码领域又有很多值得关注的重要进展,如国际上公布了新一代的 H.265/HEVC 标准,国内在 2012 年发布了 AVS+广播影视行业标准后,近期又发布了第二代 AVS2 广播影视行业标准,编码效率比第一代 AVS 标准再次提升一倍,所有这些进展都推动了本书的再版。

我们说视频在大数据中可谓“大中之大”,这不仅是因为视频处理数据量大、难度高,还因为在人的感知过程中,视觉感知对整体感知的影响最大。很多论文提到,在人的感知系统所获取的信息中,视觉信息大约占到 80%~85%。

在商业领域,“视频为王”的观点已经被普遍接受。但是在技术领域,视频处理距离理想境地还有很长的路要走。视频编码、视频处理、视频检索等技术还在不断地进化中。

视频数据的利用需要经过视频获取、视频传输或存储、视频显示几个环节。为了获得高清晰度的视频,需要高精度的视频采集设备。但是高精度的采集会产生大的数据量,这会给传输带来压力。经过分析不难发现,视频采集过程带来的数据冗余有这样或那样的规律可循,因此有可能利用这些规律的编码处理实现压缩。这个技术领域就被称为视频编码(video coding),其最有影响的技术产物就是视频编码标准。

视频编码技术与视频编码标准是包括数字电视、网络视频、手机电视、MP3 等音视频产业的技术基础。如果没有 MPEG-1 标准,就不会有 VCD 和 MP3;如果没有 MPEG-2 标准,就不会有 DVD、数字电视机顶盒等。像其他年轻的技术领域一样,视频编码技术也在不断进步。为了回答视频编码是否也像半导体领域一样存在类似的摩尔定律等问题,本书作者之一高文,曾在很多场合预言:视频编码标准按照每十年效率提高一倍的规律发展。我们不妨将这个规律称为“视频编码标准进化定律”。这个定律在过去二十年中已不断被验证,比如 1995 年的 MPEG-2 视频编码标准可以将一路高清压缩至 20Mbps,2005 年的第二代 H.264/AVC, AVS1 等视频编码标准可以将一路高清压缩至 10Mbps,2015 年的 H.265/HEVC、AVS2 视频编码标准已可以将一路高清码流压缩至 5Mbps。按此规律发展,一路高清电视的码率 2025 年需要 2.5Mbps,2035 年需要 1.25Mbps。

可能有读者要问,视频编码的摩尔定律能够持续多少年?换句话说,视频编码的上限是什么?要回答这个问题,首先要了解为什么视频可以被压缩。

数字视频因为存在数据冗余,所以可以被压缩。

数字视频由于在数字化时采用了帧内与帧间均匀采样,并由 RGB 三个分量均匀表达采样量化后的数据,从而带来了空间冗余、时间冗余和编码冗余。正是由于这三种冗余的存在,才使得视频可以被压缩。很显然,视频编码的上限就是当冗余不再存在时的数据表达。

香农的信息论不仅是通信理论的基础,同时也是视频编码理论的基础。六十多年来,视频编码领域的科学家们利用各种可能的技术努力去除视频中的空间冗余、时间冗余以及编码冗余。

本书力图沿着技术与标准发展的主线,向读者介绍视频编码各主要方面的技术原理和算法,引导读者了解和掌握视频编码技术的核心,为读者进一步的研究提供参考。

本书第 1 章为数字视频编码概论;第 2 章对视频基础进行介绍;第 3~6 章,对预测、变换、量化、熵编码等关键技术进行详细介绍;第 7~10 章分别对 MPEG、H.26x 和 AVS 系列标准进行介绍;第 11 章和第 12 章对视频编码应用过程中的可伸缩编码、转码等技术进行介绍;第 13~15 章对三维视频编码、分布式编码以及基于视觉的编码等新型编码方向进行介绍;第 16 章对图像/视频质量评价技术进行介绍。相比第一版,主要修改体现在第 7 章增加了关于 AVS+标准的内容,增添了第 8 章、第 9 章分别对 H.265/AVC、AVS2 标准的介绍,第 13 章对于三维视频编码增加了最新的 MPEG 3DV 标准方面的内容。此外,其他章节也结合最新进展做了内容修改或补充。

本书的主要内容是高文、赵德斌和马思伟多年从事本领域教学与科研的一些成果和心得集合。这些集合中包含了很多本实验室已经毕业博士研究生的研究成果,包括:张莉和王强(第 6 章),季向阳和孙俊(第 11 章)、袁禄军和张鹏(第 12 章)、刘延伟(第 13 章)、国玫和郭响(第 14 章)等。而且,许多已经毕业或在读博士生还直接参与了本书部分章节内容的讨论与图文编辑工作,包括:张莉、王苦社、王强、齐洪刚、季向阳、孙俊、刘延伟、国玫、张永兵、黄倩、王诗淇、苏荔、王志航、张凯、李勇鹏等,他们的无私奉献对于本书完成至关重要,我们在此深表谢意。作者高文早期的学生吴枫博士极力建议写作本书;孙惠方博士、熊瑞勤博士、范晓鹏博士、蒋婷婷博士等作者的朋友和同事通读了书稿,提出了很多宝贵的修改意见;作者在本书相关内容研究与写作过程中还曾与李卫平教授、陈学敏博士、张亚勤博士、王瑶教授、陈长汶教授、熊子祥教授、柯瑞德(Cliff

Reader) 博士、李世鹏博士、沈向洋博士、张宏江博士、芮勇博士、陈宏铭教授、雷少民博士、郭宗杰教授等进行过很多有益的讨论, 在此一并表示感谢, 感谢他们对本书最终完成所起的推动作用。本书在成稿过程中, 有关视觉心理学、计算机视觉、视频分析与检索等方面得到了作者高文牵头的 973 项目“基于视觉特性的视频编码理论与方法研究”组中各位同仁的很大帮助, 在此我们特别表达对濮鸣亮教授、李量教授、吕宝梁教授、陈熙霖教授、山世光博士、王亦洲博士、黄铁军教授、黄庆明教授、郭宗明博士等的衷心谢意。

本书的完成与作者家人的支持是分不开的, 没有他们的支持, 很难想象我们能有足够的时间和精力完成文字工作。值此工作完成之际, 作者愿意把这本著作作为送给我们三位所有家人的一个答谢, 感谢他们常年无怨无悔的支持。

本书的出版得到了国家重大基础研究计划项目“基于视觉特性的视频编码理论与方法研究”(2015CB351800)、国家自然科学基金重点项目(61322106 和 61632001)、国家科学技术学术著作出版基金等的支持, 在此表示衷心感谢。

# 目 录

序  
前言

## ► 第 1 章 概论 / 1

### 1.1 视觉感知 / 2

1.1.1 色彩原理 / 3

1.1.2 视觉系统的颜色感知 / 4

1.1.3 视觉系统的光强感知 / 6

1.1.4 视觉感知的高级阶段 / 8

### 1.2 数字视频 / 10

1.2.1 视频采集 / 10

1.2.2 色彩空间 / 13

1.2.3 视频格式 / 15

### 1.3 视频数据冗余 / 17

1.3.1 数据冗余类型 / 17

1.3.2 去除数据冗余的方法 / 18

### 1.4 视频编码标准 / 19

习题 / 21

## ► 第 2 章 视频编码基础 / 22

### 2.1 信息论基础 / 22

### 2.2 香农编码定理 / 24

### 2.3 视频编码技术 / 25

2.3.1 预测 / 26

2.3.2 变换 / 29

- 2.3.3 量化 / 31
- 2.3.4 熵编码 / 32
- 2.3.5 环路滤波 / 34
- 2.4 视频图像质量评价 / 35
  - 2.4.1 主观质量评价 / 35
  - 2.4.2 客观质量评价 / 36
- 习题 / 38

### ▶ 第3章 预测编码 / 39

- 3.1 预测编码概述 / 39
  - 3.1.1 帧内预测 / 40
  - 3.1.2 帧间预测 / 43
- 3.2 运动的表示及估计 / 44
  - 3.2.1 基于像素的运动估计 / 44
  - 3.2.2 基于块的运动估计 / 46
  - 3.2.3 全局运动估计 / 48
- 3.3 子像素插值 / 53
  - 3.3.1 子像素精度与插值滤波器 / 53
  - 3.3.2 自适应插值技术 / 55
- 3.4 运动矢量预测 / 56
- 习题 / 57

### ▶ 第4章 变换编码 / 58

- 4.1 K-L 变换 / 58
- 4.2 DCT 变换 / 60
  - 4.2.1 一维 DCT 变换 / 60
  - 4.2.2 二维 DCT 变换 / 61
  - 4.2.3 DCT 的编码性能 / 62
  - 4.2.4 DCT 的重要性质 / 64

- 4.2.5 快速 DCT 变换 / 65
- 4.2.6 类 DCT 整数变换 / 68
- 4.3 DST 变换 / 71
- 4.4 小波变换 / 72
  - 4.4.1 一维小波变换与二维小波变换 / 72
  - 4.4.2 离散小波变换 / 73
  - 4.4.3 提升小波 / 75
  - 4.4.4 快速小波变换 / 77
  - 4.4.5 小波编码方法 / 78
- 4.5 变换编码增益 / 78
- 习题 / 79

## ► 第 5 章 量化 / 81

- 5.1 量化基本原理 / 81
  - 5.1.1 基本概念 / 81
  - 5.1.2 标量量化 / 84
  - 5.1.3 矢量量化 / 87
  - 5.1.4 自适应量化 / 89
  - 5.1.5 视觉量化器设计 / 91
- 5.2 量化与码率控制 / 93
  - 5.2.1 率失真理论 / 93
  - 5.2.2 率失真理论与优化编码参数选择 / 95
  - 5.2.3 率失真模型与码率控制 / 96
- 习题 / 107

## ► 第 6 章 熵编码 / 108

- 6.1 基本原理 / 108
  - 6.1.1 信源的数学描述及信息熵测度 / 109
  - 6.1.2 变长编码 / 113

6.1.3 算术编码 / 116

6.2 变换系数分布特性 / 117

6.2.1 DCT 系数的统计特性 / 118

6.2.2 小波变换系数的统计特性 / 119

6.3 典型的图像熵编码方法 / 119

6.4 典型的视频熵编码方法 / 126

6.4.1 基于上下文的变长编码方法 / 127

6.4.2 基于上下文的算术编码 / 130

6.4.3 位平面编码 / 133

6.4.4 零树编码 / 135

习题 / 136

## ► 第7章 第一代视频编码标准 / 137

7.1 H.261 / 137

7.1.1 码流结构 / 138

7.1.2 解码过程 / 140

7.2 H.263, H.263+及 H.263++ / 141

7.2.1 码流结构 / 142

7.2.2 运动矢量预测 / 143

7.2.3 PB 帧 / 143

7.2.4 子像素插值 / 144

7.2.5 重叠块运动补偿 / 145

7.3 MPEG-1 / 146

7.4 MPEG-2 / 146

7.5 MPEG-4 / 147

习题 / 148

## ► 第8章 第二代视频编码标准 / 149

8.1 H.264/AVC / 149

- 8.1.1 帧内预测 / 150
- 8.1.2 帧间预测 / 152
- 8.1.3 整数变换与量化 / 152
- 8.1.4 熵编码 / 156
- 8.1.5 环路滤波技术 / 159
- 8.2 AVS/AVS+ / 161
  - 8.2.1 帧内预测 / 162
  - 8.2.2 帧间预测 / 164
  - 8.2.3 B 帧预测编码模式 / 166
  - 8.2.4 整数变换与量化 / 170
  - 8.2.5 熵编码 / 171
  - 8.2.6 环路滤波技术 / 174
- 习题 / 176

## ► 第 9 章 新一代 HEVC 视频编码标准 / 177

- 9.1 HEVC 简介 / 177
- 9.2 帧内预测 / 179
- 9.3 帧间预测 / 181
  - 9.3.1 高级运动矢量预测技术 / 181
  - 9.3.2 运动合并模式 / 182
  - 9.3.3 子像素插值 / 182
- 9.4 变换 / 185
- 9.5 量化 / 185
- 9.6 熵编码 / 186
- 9.7 环路滤波 / 187
  - 9.7.1 去块效应滤波器 / 187
  - 9.7.2 样本自适应偏移滤波器 / 188
- 习题 / 189

## ▶ 第 10 章 新一代 AVS2 视频编码标准 / 190

- 10.1 AVS2 简介 / 190
- 10.2 帧内预测 / 191
- 10.3 帧间预测 / 193
  - 10.3.1 预测划分 / 193
  - 10.3.2 运动矢量预测 / 194
  - 10.3.3 子像素插值 / 195
- 10.4 变换与量化 / 196
  - 10.4.1 基于四叉树的非方形变换 / 196
  - 10.4.2 帧内预测残差的二次变换 / 197
- 10.5 熵编码 / 198
  - 10.5.1 两层变换系数编码 / 198
  - 10.5.2 上下文自适应算术编码 / 198
- 10.6 环路滤波 / 199
  - 10.6.1 去块效应滤波 / 199
  - 10.6.2 样本偏置滤波 / 199
  - 10.6.3 样本补偿滤波 / 200
- 10.7 场景视频编码 / 200
  - 10.7.1 背景建模 / 202
  - 10.7.2 帧级背景建模 / 204
  - 10.7.3 块级背景建模 / 205
- 习题 / 206

## ▶ 第 11 章 可伸缩视频编码 / 207

- 11.1 传统的分层视频编码方法 / 207
- 11.2 精细粒度可伸缩视频编码方法 / 208
- 11.3 可伸缩小波视频编码方法 / 210
- 11.4 层间预测编码 / 212

习题 / 212

## ► 第 12 章 视频转码 / 213

- 12.1 码率缩减的视频转码 / 215
- 12.2 分辨率缩减的视频转码 / 218
- 12.3 帧率缩减的视频转码 / 220
- 12.4 其他类型的转换 / 222
- 习题 / 223

## ► 第 13 章 三维视频编码 / 224

- 13.1 三维视觉原理 / 226
  - 13.1.1 立体视觉 / 226
  - 13.1.2 双目视差 / 228
  - 13.1.3 立体视觉模型 / 229
  - 13.1.4 三维视频获取 / 230
  - 13.1.5 立体显示 / 232
- 13.2 三维场景表示 / 234
  - 13.2.1 典型的三维场景表示技术 / 234
  - 13.2.2 二进制场景描述语言 / 236
- 13.3 立体视频编码 / 238
  - 13.3.1 立体视频的生成 / 238
  - 13.3.2 传统的立体视频编码 / 242
  - 13.3.3 纹理+深度格式的立体视频编码 / 244
- 13.4 多视点视频编码 / 245
  - 13.4.1 基于 3D 模型的多视点视频编码 / 246
  - 13.4.2 基于混合编码框架的多视点视频编码 / 246
  - 13.4.3 基于四维小波的多视点视频编码 / 258
- 13.5 MPEG 3DV 标准 / 260
  - 13.5.1 依赖视点纹理编码 / 263

13.5.2 深度图编码 / 269

13.6 全景视频编码 / 274

习题 / 277

## ► 第 14 章 分布式视频编码 / 278

14.1 分布式信源编码理论与技术 / 279

14.1.1 相关信号的无损压缩——Slepian-Wolf 理论 / 279

14.1.2 基于辅助信息解码的有损压缩——Wyner-Ziv 理论 / 280

14.1.3 分布式信源编码 / 281

14.1.4 信道编码 / 283

14.2 分布式视频编码技术 / 285

14.2.1 视频辅助信息的生成 / 285

14.2.2 频域变换与系数组织 / 287

14.2.3 分布式视频编码框架 / 288

14.3 分布式编码技术应用 / 289

14.3.1 分布式多视视频编码 / 289

14.3.2 基于分布式编码的可伸缩视频编码 / 293

14.3.3 基于分布式编码的多码率码流切换 / 293

习题 / 295

## ► 第 15 章 视觉编码 / 296

15.1 人眼视觉系统 / 296

15.1.1 视觉通路 / 296

15.1.2 神经细胞感受野 / 298

15.1.3 人类视觉系统特性 / 299

15.2 Foveation 视觉模型及应用 / 301

15.2.1 几何方法 / 303

15.2.2 基于滤波的方法 / 304

15.2.3 多分辨率分析方法 / 305

习题 / 306

## ► 第 16 章 图像质量评价 / 307

16.1 图像质量评价概述 / 307

16.1.1 图像质量含义 / 307

16.1.2 图像质量评价的应用 / 308

16.1.3 质量评价方法分类 / 309

16.1.4 性能指标和评价准则 / 309

16.2 图像质量评价方法 / 310

16.2.1 主观质量评价 / 311

16.2.2 客观质量评价 / 312

16.2.3 JND 模型 / 316

16.3 人工效应分析 / 319

16.3.1 块效应及其评价方法 / 319

16.3.2 模糊、振铃效应及其评价方法 / 321

习题 / 323

## ► 结束语 / 325

## ► 参考文献 / 326

## ► 词汇表索引 / 362

# 第1章

## 概 论

自从有了电视，视频就以各种各样的方式影响和改变着社会，不论是政治、经济、体育、文化、教育，还是娱乐。自从数字视频开始取代模拟视频，特别是有了互联网以后，数字视频应用就成为信息产业发展最重要的驱动力。可以预见，数字视频在今后若干年，仍然将是多媒体领域的热点，不论是在研究领域还是在产业界。

在数字视频研究的众多分支领域中，视频编码和视频理解是最具有挑战性的两个分支。视频编码的研究动机，是去除数字视频中的冗余，节省传输和存储开销。视频理解的研究动机，是利用人工智能的方法实现视频的识别与理解。本书论述重点是前者，即视频编码。对于视频编码领域，新入门读者可能会提出如下问题：

(1) 为什么数字视频是可以压缩的，数字视频中的冗余是如何产生的？

(2) 数字视频的采集与处理，与人的视觉系统工作机理是否一致？哪一个效率更高？

一般而言，图像是由摄影空间中物体对光源的反射和折射在摄影平面上投影产生的。视频是由按一定时间间隔获取的图像序列，序列中的一幅图像也被称为一帧图像。视频在获取或显示时，每秒钟按逐行扫描处理的图像数称为帧率，每秒钟按隔行扫描处理的图像数称为场率，例如每秒 25 帧或 50 场等。由于光是连续的，投影后的图像在空间是连续的，也即原始视频在空间上是连续的。

为了处理、传输和存储方便，需要将图像在空间上的连续模拟量进行采样和