

Signals and Communication Technology

K. R. Rao
Do Nyeon Kim
Jae Jeong Hwang

Video Coding Standards

AVS China, H.264/MPEG-4 PART 10,
HEVC, VP6, DIRAC and VC-1

 Springer

Signals and Communication Technology

For further volumes:
<http://www.springer.com/series/4748>

K. R. Rao · Do Nyeon Kim
Jae Jeong Hwang

Video Coding Standards

AVS China, H.264/MPEG-4 PART 10,
HEVC, VP6, DIRAC and VC-1

 Springer

K. R. Rao
Electrical Engineering, Nedderman Hall
University of Texas at Arlington
Arlington, TX
USA

Jae Jeong Hwang
Radio Communication Engineering
Kunsan National University
Kunsan
Republic of Korea (South Korea)

Do Nyeon Kim
Barun Technologies Corporation
Seoul
Republic of Korea (South Korea)

ISSN 1860-4862

ISBN 978-94-007-6741-6

ISBN 978-94-007-6742-3 (eBook)

DOI 10.1007/978-94-007-6742-3

Springer Dordrecht Heidelberg New York London

Library of Congress Control Number: 2013939570

© Springer Science+Business Media Dordrecht 2014

This work is subject to copyright. All rights are reserved by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed. Exempted from this legal reservation are brief excerpts in connection with reviews or scholarly analysis or material supplied specifically for the purpose of being entered and executed on a computer system, for exclusive use by the purchaser of the work. Duplication of this publication or parts thereof is permitted only under the provisions of the Copyright Law of the Publisher's location, in its current version, and permission for use must always be obtained from Springer. Permissions for use may be obtained through RightsLink at the Copyright Clearance Center. Violations are liable to prosecution under the respective Copyright Law. The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

While the advice and information in this book are believed to be true and accurate at the date of publication, neither the authors nor the editors nor the publisher can accept any legal responsibility for any errors or omissions that may be made. The publisher makes no warranty, express or implied, with respect to the material contained herein.

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

Preface

To our knowledge this book is unique and unusual in the sense that it consolidates all the recent major video coding standards: AVS China, H.264/MPEG-4 Part 10 (here after referred to as H.264/AVC), VP6 (now VP10), DIRAC, VC-1/2 and HEVC/NGVC (high efficiency video coding/next generation video coding). All these standards with the exception of the later have been adopted by the industry (at various levels) developing consumer electronics products for mass markets (set top boxes, search engines, lap tops, PCs, tablets, net books, digital cameras, DVDs, Blu-ray discs, smart phones, various hand held/mobile devices, VOD, video streaming, etc.). Products based on HEVC/NGVC are under various stages of development as can be observed from various papers (conference proceedings, journal papers, tutorials, keynote speeches—see [Chap. 5](#)). The focus of this book has been mainly on the basic functionalities, tools, techniques and operations inherent in these standards leading to compression/coding at various bit rates, quality levels and applications. Intentionally, detailed step-by-step process for their implementation is left out. However, thorough understanding of the standards, besides their performance comparison and limitations are presented. With on line resources such as web/ftp sites, standards documents, open source software, review papers (both journal publications and conference presentations), keynote speeches, tutorials, reflectors and related resources, the reader is well equipped to implement the encoders and decoders at various profiles and levels and evaluate their performances.

With insight and ingenuity, the reader can improve the performance, reduce the complexity and explore additional applications. Other than [Chaps. 1](#) and [2](#), the remaining chapters end with projects (can be tailored to M.S. theses and some of them even to Ph.D. dissertations) presented as thought provoking items. Added to this is the prospect of transcoding from one standard to another (not encoder/decoder of one standard followed by encoder/decoder of another standard). This is to be accomplished by using parts of the encoded bit stream of one standard effectively/efficiently in the other standard thus significantly reducing the overall complexity of the transcoder. This by itself is no easy task. However, many challenges have been overcome i.e. MPEG-2 to H.264/AVC transcoder and vice versa and low complexity H.264/AVC to VC-1 transcoding.

HEVC/NGVC (See [Chap. 5](#)) is the latest standard that is aimed at reducing the bit rate by nearly 50 % at the same visual quality compared with H.264/AVC. It has come out as a final draft international standard (FDIS) in January 2013 (main, 10 bit depth and all intra profiles) with the working draft updated periodically. This has been approved by ITU as ITU-T H.265/ISO/IEC 23008-2 HEVC in January 2013. It is projected that the HEVC-based decoders will reach 2 billion by end of 2016. This is followed by extensions such as scalable video coding (SVC) and multi-view coding (3D video, free view point video, stereoscopic video, etc.) scheduled for standardization in 2014. These developments provide a fertile ground for R&D besides the transcoder from an established standard such as H.264/AVC to HEVC and vice versa (See projects at the end). The main focus of this book is on video with little description on audio. However, some theses/projects describe in detail the coding and multiplexing of the video/audio bit streams at the encoder followed by demultiplexing and decoding of these bit streams while maintaining the lip sync between the video and audio. The reader is referred to Dr. Rao's website <http://www-ee.uta.edu/dip> (click on courses and then click on EE5359 multimedia processing, scroll down to access theses/projects/pp slides/papers/proposals, etc.) i.e. multiplexing/de-multiplexing AVS China video with AAC audio bit streams achieving lip sync and multiplexing/de-multiplexing H.264/AVC video with HE-AAC audio bit streams achieving lip sync.

[Chapters 6–8](#) address the functionalities involved in VP6 (originally developed by On2 Technologies—subsequently acquired by Google) with websites related to VP9/VP10, DIRAC developed by BBC, which uses wavelets rather than the traditional DCT/INT DCT and VC-1 based on Windows Media Video 9 of Microsoft, respectively. Apart from [Chaps. 1 and 2](#) all the others are supplemented by projects many of which are at the M.S. and Ph.D. levels. Appendices A thru H (based on former graduate students in the University of Texas at Arlington except Appendices B and H) provide additional resources that complement the projects many of which are at the M.S. and Ph.D. levels. Bibliography provides a plethora of references including web/ftp sites, review papers, standards documents, keynote speeches, special issues, open source software, etc., related to video coding standards presented in [Chaps. 3–8](#).

It is hoped that this book and all the resources outlined can provide the reader the prospects for understanding and implementation of these standards. With foresight the reader can propose changes/additions/extensions/modifications to these standards leading to improved performance, reduced complexity and additional applications. Purposely, the book is not aimed as a text for any specific course. On the contrary it is projected to be a reference at the graduate/research levels for academia, research institutes and industry.

Acknowledgments

This book is the result of long-term association of the three authors, K. R. Rao, D. N. Kim and J. J. Hwang. Special thanks go to their respective families for their support, perseverance and understanding. Both Dr. Kim and Dr. Hwang were visiting professors in the multimedia processing lab (MPL) at the University of Texas at Arlington (UTA) whose significant and sustained contributions made this book possible. The first author likes to acknowledge the support provided in various forms by Dr. Jean-Pierre Bardet, Dean, College of Engineering, Dr. J. W. Bredow, Chair, Department of Electrical Engineering, and colleagues all in UTA. Dr. G. J Sullivan, Microsoft, Dr. Nam Ling, University of Santa Clara, Dr. Ankur Saxena and Dr. Zhan Ma (both from Samsung Research labs), Dr. Wen Gao, Peking University, Dr. M. Budagavi, TI, Dr. M. T. Sun, University of Washington, Dr. H. Lakshman, Fraunhofer HHI, Dr. T. Borer, BBC, Dr. Deshpande, Sharp Labs, Dr. J. Bankoski and Dr. D. Mukherjee (both from Google) Dr. Y. Reznik, Interdigital, Dr. H. Kalva, Florida Atlantic University, and Dr. E. Izquierdo, Queen Mary University of London for their unwavering help. Ms. Priyadarshini Anjanappa alumnae at UTA contributed countless hours in tying up all loose ends (references/copy right releases, proof reading and million other details). The graduate students and alumnae in multimedia processing lab (MPL) at UTA in various ways have made constructive comments, especially [Chap. 5](#), High Efficiency Video Coding (HEVC) leading to innumerable revisions and updates. All the appendices A thru H (except for Appendices B and H) are essentially the project reports or theses by former students in MPL at UTA. They are B. Shrestha, C. Chukka, P. Anjanappa, S. Muniyappa, R. Veerla and M. Krishnan. The authors sincerely appreciate their contributions.

Contents

1	Introduction	1
1.1	Popular Video and Audio Standards	1
1.2	Digital Representation of Video	3
1.3	Basic Structure of Video Codec	18
1.4	Performance Comparison Metrics for Video Codec	18
1.5	Digital Representation of Audio	28
1.6	Basic Structure of Perceptual Audio Coding	29
1.7	Performance Comparison Metrics for Audio Codec	31
1.8	Summary	36
2	Video Coding Standards and Video Formats	37
2.1	Introduction	37
2.2	Complexity Reduction	40
2.3	Video Coding Standards	40
2.4	MPEG and H.26x	41
2.4.1	H.120	41
2.4.2	H.261	41
2.4.3	MPEG-1	42
2.4.4	H.262/MPEG-2	42
2.4.5	H.263, H.263+ and H.263++	43
2.4.6	MPEG-4	43
2.4.7	H.264/MPEG-4 Part 10/AVC	44
2.4.8	H.265/HEVC	45
2.5	Video Formats and Quality	45
2.5.1	Frames and Fields	45
2.5.2	Color Spaces	46
2.5.3	Video Formats	49
2.5.4	Quality	49
2.6	Summary	50

3 AVS China	51
3.1 AVS China	51
3.2 AVS China Profiles and Levels	52
3.2.1 AVS-Video Jizhun (Base) Profile	52
3.2.2 AVS-Video Jiben (Basic) Profile	53
3.2.3 AVS-Video Shenzhen (Extended) Profile	53
3.2.4 AVS-Video Jiaqiang (Enhanced) Profile	55
3.3 Data Formats Used in AVS	55
3.3.1 AVS Video Layered Structure	56
3.4 AVS Video Encoder	59
3.4.1 Encoder Process Outline	60
3.4.2 Coding Tools Used in AVS Video Coder	61
3.5 AVS Video Decoder	69
3.6 AVS Video Bit Stream	69
3.6.1 Start Code	70
3.6.2 Start Code Value	70
3.6.3 Picture_coding_type	71
3.7 NAL Unit for AVS Video Stream	72
3.7.1 NAL Unit Mapping with AVS Video Stream	72
3.7.2 NAL Unit Header Description	72
3.8 Introduction to AVS-M (AVS Part 7)	74
3.8.1 Data Structure of AVS-M	75
3.8.2 Embodiment of AVS-M	78
3.8.3 Various Levels in Jiben Profile	80
3.9 Block Mode Prediction Modes	81
3.9.1 Intra Prediction	81
3.9.2 Inter Prediction	83
3.9.3 Skip Mode Prediction	84
3.9.4 RD Optimization	85
3.10 Transform, Quantization and Entropy Coding	85
3.10.1 Transform	85
3.10.2 Quantization	86
3.10.3 Entropy Coding	86
3.10.4 Simplified Deblocking Filter	87
3.11 AVS Part-1: System	89
3.11.1 Program Stream	91
3.11.2 Transport Stream	92
3.12 IEEE AVS	96
3.12.1 Applications	97
3.12.2 Profiles and Levels	97
3.12.3 Overview of the Design Characteristics	97
3.13 Summary	98
3.14 Projects	98

- 4 H.264/MPEG-4 Advanced Video Coding 99**
 - 4.1 Introduction. 99
 - 4.2 Profiles and Levels of H.264. 100
 - 4.2.1 Profiles in H.264 100
 - 4.2.2 Levels in H.264 105
 - 4.3 H.264 Encoder. 105
 - 4.4 Intra-Prediction 107
 - 4.5 Inter-Prediction 108
 - 4.6 Inter Prediction of Macroblocks in P-Slices. 108
 - 4.7 Sub-Pixel Motion Vectors 109
 - 4.8 Transform and Quantization 112
 - 4.9 In-Loop Deblocking Filter. 112
 - 4.9.1 Filter Strength 114
 - 4.10 B-Slices and Adaptive Weighted Prediction 116
 - 4.11 Entropy Coding 117
 - 4.12 H.264 Decoder 119
 - 4.13 Some Applications of H.264 120
 - 4.14 Summary 121
 - 4.15 Projects. 121

- 5 High Efficiency Video Coding (HEVC) 125**
 - 5.1 Introduction. 125
 - 5.2 Joint Collaborative Team on Video Coding. 125
 - 5.3 Analysis of Coding Tools in HEVC Test Model,
 - HM 1.0: Intra Prediction. 132
 - 5.4 HEVC Encoder 132
 - 5.4.1 Intra Prediction 135
 - 5.4.2 Transform Coefficient Scanning. 136
 - 5.4.3 Luma and Chroma Fractional Pixel Interpolation. 137
 - 5.4.4 Comparison of Coding Tools of HM1
 - and HEVC Draft 9. 137
 - 5.5 Extensions to HEVC 138
 - 5.6 Profiles and Levels. 140
 - 5.7 Performance and Computational Complexity
 - of HEVC Encoders 143
 - 5.8 System Layer Integration of HEVC 144
 - 5.9 HEVC Lossless Coding and Improvements 144
 - 5.10 Summary 146
 - 5.11 Projects. 148

- 6 VP6 Video Coding Standard. 159**
 - 6.1 Introduction. 159
 - 6.2 Comparison with Previous Flash Codec MX 160
 - 6.3 VP6 Algorithm Fundamentals 165

6.4	Coding Profiles in VP6.	166
6.5	Types of Frames	167
6.5.1	Golden Frames	168
6.6	MB Modes	168
6.6.1	MB Modes in I-Frames (Intra-Mode)	168
6.6.2	MB Modes in P-Frames (Inter-Modes and Intra-Mode)	169
6.7	Nearest and Near Blocks.	170
6.8	Motion Vectors	171
6.8.1	Encoding.	172
6.8.2	Prediction Loop Filtering	172
6.9	Filtering for Fractional Pixel Motion Compensation.	172
6.9.1	Bilinear Filtering	173
6.9.2	Bicubic Filtering	173
6.10	Support for Unrestricted Motion Vectors	174
6.11	Prediction Loop Filtering	174
6.12	DCT, Scan Orders and Coefficient Token Set	174
6.12.1	Scan Orders.	180
6.12.2	DCT Coding and Coefficient Token Set	181
6.13	Quantization	186
6.14	Entropy Coding	187
6.14.1	Use of Context Information.	188
6.14.2	Huffman Coder	188
6.14.3	BoolCoder.	189
6.15	An Overview on VP6 Coding	189
6.16	Performance of VP6 Coding	190
6.17	VP6 Golden Frames	191
6.18	Background/Foreground Segmentation	191
6.19	Context Predictive Entropy Encoding	192
6.20	Bitstream Partitions	192
6.21	Dual Mode Arithmetic and VLC Encoding	194
6.22	Adaptive Sub-Pixel Motion Estimation	194
6.23	VP6-E and VP6-S Encoder Profiles	194
6.24	Device Ports and Hardware Implementations.	195
6.25	Summary	197
6.26	Projects.	197
7	Performance Analysis and Comparison of the Dirac Video Codec with H.264/MPEG-4, Part 10	199
7.1	Introduction.	199
7.2	Dirac Architecture	200
7.2.1	Dirac Encoder	200
7.2.2	Dirac Decoder	201
7.3	Stages of Encoding and Decoding in Dirac	202

7.3.1	Wavelet Transform.	202
7.3.2	Scaling and Quantization	204
7.3.3	Entropy Coding	205
7.3.4	Motion Estimation	206
7.3.5	Motion Compensation.	207
7.3.6	Decoder	208
7.4	Implementation	209
7.4.1	Code Structure Overview	209
7.4.2	Simplicity and Relative Speed of Encoding.	209
7.5	Results	211
7.5.1	Compression Ratio Test	211
7.5.2	SSIM Test.	213
7.5.3	PSNR Test	214
7.6	Conclusions.	217
7.7	Future Research.	218
7.8	Summary	218
7.9	Projects.	218
8	The VC-1 Video Coding.	221
8.1	The VC-1 Structure	221
8.2	Integer Transform Coding	222
8.2.1	Inverse Transform	222
8.2.2	Forward Transform	224
8.3	Motion Estimation/Compensation.	226
8.3.1	Loop Filter	227
8.3.2	Complexity	228
8.3.3	Profiles and Levels.	229
8.4	The Simple Profile.	232
8.4.1	Bitstream Structure.	232
8.4.2	Baseline Intra-Frame Compression.	233
8.4.3	Variable-Size Transform Specifications	234
8.4.4	Overlapped Transform	236
8.4.5	4MV per MB.	238
8.4.6	Quarter-pel MC for Y.	240
8.5	The Main Profile	244
8.5.1	Quarter-pel MC for C_bC_r	244
8.5.2	Start Codes	244
8.5.3	Extended MV	246
8.5.4	Loop Filter	246
8.5.5	Dynamic Resolution Change	247
8.5.6	B Frames	249
8.5.7	Adaptive MB Quantization	250
8.5.8	Intensity Compensation.	253
8.5.9	Range Adjustment	254

- 8.6 The Advanced Profile 254
 - 8.6.1 Bitstream Structure 254
 - 8.6.2 Interlace 255
 - 8.6.3 Sequence Level User Data 257
 - 8.6.4 Entry Point Layer 258
 - 8.6.5 Display Metadata 258
- 8.7 The H.264 to VC-1 Transcoding 259
 - 8.7.1 Intra MB Mode Mapping 260
 - 8.7.2 Inter MB Mode Mapping 261
 - 8.7.3 Motion Vector Mapping 262
 - 8.7.4 Reference Pictures 263
 - 8.7.5 Skipped MB 263
- 8.8 Transport of VC-1 263
 - 8.8.1 Encapsulation of VC-1 in TS 265
 - 8.8.2 Encapsulation of VC-1 in PS 265
- 8.9 VC-2 Video Compression 267
 - 8.9.1 Introduction 267
 - 8.9.2 Scope 268
- 8.10 Summary 269
- 8.11 Projects 269

Appendix A: Investigation of Image Quality of Dirac, H.264 and H.265 271

- A.1 Introduction 271
- A.2 H.265 271
- A.3 Image Quality Assessment Using SSIM and FSIM 272
- A.4 Results 277
 - A.4.1 Results using Foreman QCIF Sequence 277
 - A.4.2 Results using Foreman CIF Sequence 277
 - A.4.3 Results using container QCIF Sequence 277
 - A.4.4 Results using container CIF Sequence 277
- A.5 Conclusions 277
- A.6 Projects 294

Appendix B: PSNR Average for AVSNR Software 295

Appendix C: A Universal Image Quality Index and SSIM Comparison 297

- C.1 Introduction 297
- C.2 Universal Image Quality Index [Q8] 302
- C.3 Structural Similarity Index [Q13] 304
- C.4 Images with Distortions [G11] 311
- C.5 Results 315
- C.6 Conclusions 318

C.7 Project 319

C.8 JVT Document on Video Quality Metrics in the H.264 Reference Software 320

Appendix D: Implementation of Mode Dependent

DCT/DST in H.264 325

D.1 Introduction. 325

D.2 Transform Implementation in the Reference Software 326

D.3 Proposed Scheme. 327

D.3.1 Mapping from Intra Prediction Modes to DCT/DST. . . 327

D.3.2 Obtaining DST Matrices for H.264 327

D.3.3 Implementation of DCT/DST in the Reference Software for H.264/AVC 329

D.4 Calculation of BD-PSNR and BD-Bit rate. 330

D.5 Performance Analysis. 331

D.5.1 Results for WQVGA (416×240) Sequences 331

D.5.2 Results for WVGA (832×480) Sequences. 332

D.5.3 Results for HD (1920×1080) Sequences. 332

D.5.4 Results for HD (1080×720) Sequences. 335

D.5.5 Results for different combinations of DCT/DST applied to RaceHorses Sequences. 335

D.6 Conclusions and Future Work 341

Appendix E: Performance Analysis and Comparison

of JM, Intel IPP and X264 for H.264 Softwares 347

E.1 H.264 347

E.2 JM Software [H30] 349

E.3 X264 [X1] 349

E.4 Intel IPP [X3] 350

E.5 JM (17.2) Performance Analysis 351

E.6 X264 Performance Analysis 355

E.7 Intel IPP Performance Analysis 356

E.8 Comparison of SSIM for JM, X264 and Intel IPP Softwares in Baseline, Main and High Profiles. 360

E.9 Comparison of PSNR for JM, X264 and Intel IPP Softwares in Baseline, Main and High Profiles. 362

E.10 Comparison of Encoding Time for JM, X264 and Intel IPP Softwares in Baseline, Main and High Profiles 364

E.11 Comparison of Compression Ratio for JM, X264 and Intel IPP Softwares in Baseline, Main and High Profiles 366

E.12 Conclusions. 368

E.13 Future Work 368

- Appendix F: Implementation of AIC Based on I-Frame Only Coding in H.264 and Comparison with Other Still Frame Image Coding Standards Such as JPEG, JPEG 2000, JPEG-LS and JPEG-XR** 369
 - F.1 Introduction. 369
 - F.2 Advanced Image Coding. 370
 - F.3 Modified AIC 375
 - F.4 H.264 Standard 377
 - F.5 JPEG 380
 - F.6 JPEG2000 381
 - F.7 JPEG XR 383
 - F.8 JPEG-LS. 384
 - F.9 JPEG-LS Algorithm 385
 - F.10 Main Differences [AC1, H11, J22, JX3, JL2, JL4]. 387
 - F.11 Evaluation Methodology 388
 - F.12 Conclusions and Future Work 392

- Appendix G: Higher Order 2-D ICTs for HD Video Coding.** 421
 - G.1 Discrete Cosine Transform and Video Compression. 421
 - G.2 Integer Cosine Transforms 423
 - G.3 Simple 2-D Order 16 ICT. 425
 - G.4 Modified 2-D Order 16 ICT 429
 - G.5 2-D Order 16 binDCT Based on Loeffler’s Factorization 433
 - G.6 Transform Coding Gain 435
 - G.7 Implementation in H.264/AVC and Performance Analysis 437
 - G.8 Implementation in AVS Video and Performance Analysis. 439
 - G.9 Conclusions and Future Work 442

- Appendix H: Comparison of H.264 Codecs.** 443

- Bibliography** 449

- Index** 479

Abbreviations

2D-VLC	Two-dimensional variable length coding
AAC	Advanced audio coder
AAD	Adaptive arithmetic decoder
ACM	Association for computing machinery
ACR	Absolute category rating
ADPCM	Adaptive DPCM
AES	Audio engineering society
AIC	Advanced image coding
AIF	Adaptive interpolation filter
ALF	Adaptive loop filter
AMVP	Adaptive motion vector prediction
ANSI	American national standards institute
APIF	Adaptive pre-interpolation filter
APSIPA	Asia pacific signal and image processing association
ARM	Advanced RISC microprocessor
ASIC	Application specific integrated circuits
ASO	Arbitrary slice order
ASSP	Acoustics, speech and signal processing
ATSC	Advanced television system committee
AU	Access unit
AVC	Advanced video coding
AVS	Audio and video coding standard of China
AVS-M	Audio video standard for mobile
AWGN	Additive white Gaussian noise
B slice	Bi-directionally predictive slice
BBC	British broadcasting corporation
BD	Bjøntegaard delta
B-Frame	Interpolated frame
BMP	Bit map format
bps	Bits per second
BSTM	Butterfly style transform matrices
BT	Broadcasting service (television)
C2DVLC	The context-based 2D variable length coding

CABAC	Context-based adaptive binary arithmetic coding
CAFI	Combined adaptive-fixed interpolation with multi-directional filters
CAVLC	Context-adaptive variable length coding
CB	Coding block
CBAC	Context-based binary arithmetic coding
CBP	Coded block pattern
CBR	Constant bit rate
CC	Color conversion
CCD	Charge coupled devices
CD	Committee draft
CE	Core experiment
CI	Confidence interval
CIF	Common intermediate format
CODEC	Coder and decoder
CRA	Clean random access
CSIQ	Categorical subjective image quality
CSNR	Channel signal to noise ratio
CSVT	Circuits and systems for video technology
CTB	Coding tree block
CTU	Coding tree unit
CU	Coding unit
dB	Decibel
DCC	Data Compression Conference
DCT	Discrete cosine transform
DDCT	Directional DCT
DIP	Direct intra prediction, digital image processing
DMB	Digital multimedia broadcasting
DMOS	Differential mean opinion score
DPB	Decoded picture buffer
DPCM	Differential pulse code modulation
DRM	Digital rights management
DSCQS	Double stimulus continuous quality scale
DSIS	Double stimulus impairment scale
DSP	Digital signal processing
DTS	Decoding time stamp
DVB	Digital video broadcasting
DVD	Digital versatile disk
DWT	Discrete wavelet transform
EBCOT	Embedded block coding with optimized truncation
EBU	European broadcasting union
EE	Electrical engineering
EEQM	Energy equalization quality metric
EI	Electronic imaging
EOB	End of block

ES	Elementary stream
EZW	Embedded zero-tree wavelet coding
FAAC	Free advanced audio coder
FDCT	Forward DCT
FDIS	Final draft international standard
FFT	Fast Fourier transform
FLAC	Free lossless audio codec
FMO	Flexible macroblock order
fps	Frames per second
FR	Frame rate, full refetence
FRExt	Fidelity range extensions
FSIM	Feature similarity index
GIF	Graphics interchange format
GM	Gradient magnitude
GOP	Group of picture(s)
HD	High definition
HD-photo	High-definition photo
HDTV	High definition television
HE	High efficiency
HEVC	High efficiency video coding
HHR	Half-horizontal resolution, horizontal high resolution
HM	HEVC test model
HP	High profile
HR	High resolution
HVR	Half-vertical resolution
HVS	Human visual system
I slice	Intra slice
IASTED	International Association of Science and Technology for Development
ICALIP	International Conference on Audio, Language and Image Processing
ICC	Inverse CC
ICIP	International Conference on Image Processing
ICT	Integer cosine transform
IDR	Instantaneous decoding refresh
IEC	International electrotechnical commission
IEEE	Institute of electrical and electronics engineers
I-Frame	Intra frame
IJWMIP	International Journal on Wavelets, Multimedia and Image Processing
ILF	In-loop filtering
IMS	IP multimedia subsystem
IPR	Intellectual Property Rights
IPTV	Internet Protocol TV
IS	International Standard

ISCAS	International Symposium on Circuits and Systems
ISDB	Integrated Services Digital Broadcasting
ISO	International Standardization Organization
IST	Integer sine transform
ITU-T	The Telecommunication Standardization Sector of the International Telecommunication Union
IVC	Images and Video Communications
IZZ	Inverse zig-zag scan
JAES	Journal of AES
JM	Joint model
JPEG	Joint photographic experts group
JPEG-LS	JPEG lossless coding
JPEG-XR	JPEG extended range
JVCIR	Journal of Visual Communication and Image Representation
JVT	Joint video team
kbps	kilo bps
KLТ	Karhunen-Loeve transform
KTA	Key technical areas
LAN	Local area network
LBT	Lapped bi-orthogonal transform
LIVE	Laboratory for Image and Video Engineering
LOCO-I	Low complexity lossless compression for images
LR	Low resolution
LZW	Lempel-Ziv-Welch
M-AIC	Modified advanced image coding
MB	Macroblock
MBAFF	Macroblock-level adaptive frame/field coding
Mbps	Mega bps
MC	Motion compensation
MCPE	Motion compensated prediction error
MDCT	Modified DCT
MDDT	Mode dependent directional transform
ME	Motion estimation
MICT	Modified ICT
MICTL	Media Information and Communication Technology Laboratory
MMS	Multimedia messaging service
MMSE	Minimum MSE, multimedia signal processing
MOMS	Maximal-order interpolation with minimal support
MOS	Mean opinion score
MOV _s	Mean output variables
MP	Main profile
MPEG	Moving picture experts group
MPM	Most probable mode
MS SSIM	Multi scale structural similarity
MSE	Mean square error

MSSIM	Mean structural similarity
MSU	Moscow State University
MUSHRA	Multiple stimulus with hidden reference and anchors
MV	Motion vector
NAB	National Association of Broadcasters
NAL	Network abstraction layer
NMSE	Normalized mean square error
NO	No reference
NORM	No-reference video quality monitoring algorithm
NTSC	National Television System Committee
OBMC	Overlapped block-based motion compensation
ODG	Overall difference grade
OLT	Overlapped transform
P slice	Predictive slice
PAFF	Picture-level adaptive frame/field coding
PAL	Phase alternating line
PC	Phase congruency
PCM	Pulse code modulation, Pacific-Rim Conference on Multimedia
PCS	Picture Coding Symposium
PDMOSCL	Predicted degradation of differential mean opinion score due to coding-artifacts and packet losses
PE	Prediction error
PEAQ	Perceptual evaluation of audio quality
PES	Packetized elementary stream
P-Frame	Predicted frame
PGM	Portable gray map
PID	Packet identifier
PIT	Prescaled integer transform
PMT	Program map table
PNM	Portable any map
PPM	Portable pixel map
PPS	Picture parameter set
PS	Program stream
PSI	Program specific information
PSNR	Peak-to-peak signal-to-noise ratio
PU	Prediction unit
PVR	Personal video recorder
Q	Quantization
QCIF	Quarter CIF
QF	Quality factor
QM	Quality metric
QMF	Quadrature mirror filter
QP	Quantization parameter
RAM	Random access memory
R-QF	Rate QF

RD Cost	Rate distortion cost
RDO	Rate distortion optimization
RGB	Red, green and blue
RL	Run-length
RMS	Root mean square
RMSE	Root mean square error
ROI	Region of interest
ROT	Rotational transform
RR	Reduced reference
RV	Real video
RVC	Reconfigurable video coding
SAD	Sum of absolute differences
SATD	Sum of absolute transform differences
SBR	Spectral band replication
SD	Standard definition
SDO	Standards development organizations
SEI	Supplemental enhancement information
SI	Switched intra
SICT	Simple order 16 ICT
SIF	Source input format
SIMD	Single instruction multiple data
SMPTE	Society of Motion Picture and Television Engineers
SNR	Signal to noise ratio
SoC	System on chip
SP	Switching prediction, switched predictive
SPS	Sequence parameter set
SSCQE	Single stimulus continuous quality evaluation
SSE	Streaming SIMD extensions
SSIM	Structural similarity
SVC	Scalable video coding
SSVC	Spatially SVC
TB	Transform block
TE	Tool experiments
TF-MSSIM	Time-frequency MSSIM
TID2008	Tampere image database 2008
T-MSSIM	Temporal MSSIM
TS	Transport stream
T-STD	Transport system target decoder
TU	Transform unit
UIQI	Universal image quality index
UTA	University of Texas at Arlington
UVLC	Universal variable length coding
VBR	Variable bit rate
VC-1	Video coder 1 by SMPTE
VCEG	Video coding experts group

VCIP	Visual communications and image processing
VCIR	Visual communications and image representation
VCL	Video coding layer
VLC	Variable length coding, variable length codes
VLD	Variable length decoder
VLSI	Very large scale integration
VOD	Video on demand
VQ	Vector quantization
VQEG FR-TV	Video quality experts group full reference television
VQM	Video quality model, video quality metric
WD	Working draft
WMA	(Microsoft) Windows media audio
WQVGA	Wide quarter video graphics array
WVGA	Wide video graphics array
YUV	Luminance and chrominance color components
ZZ	Zig-zag scan

Chapter 1

Introduction

Abstract A brief description (comparison) of various video/image coding standards such as JPEG, MPEG and H.26x series besides, DIRAC (Chap. 7), AVS China (Chap. 3) and VC-1 (Chap. 8) are presented. Similarly brief description on audio coding followed by comparison metrics concludes the chapter.

Keywords Standards · Video codec · Audio codec · Color space · Subjective quality · Objective quality · Metrics · Audio quality · Video quality · Performance

1.1 Popular Video and Audio Standards

The combination of multiple sources of video, audio, image and text is usually known as multimedia. The requirements for multimedia communications have increased rapidly in the last two decades in broad areas such as television, entertainment, interactive services, telecommunications, conference, Internet, consumer electronic devices, medicine, security, business, traffic, defense and banking. Usually, video and audio data have to be compressed before storage or transmission because the original data volume is too huge, and the compressed data should be decoded before display or for further processing. Compression is also referred to as encoding or coding, and decompression as decoding. Therefore the software or device to compress and decompress video/audio data is called as video/audio encoder (coder) and decoder respectively. Encoder and decoder are abbreviated as codec for convenience. Although plenty of video and audio coding algorithms have been developed, it is the video and audio coding standards [B7], which guarantee interoperability between software and hardware provided by multiple vendors, that make multimedia communications practical. Series of video and audio coding standards have been developed by Standards Development Organizations (SDO), including ISO/IEC (the International Standardization

Organization and the International Electrotechnical Commission) [H53] [H54], ITU-T (the Telecommunication Standardization Sector of the International Telecommunication Union, formerly CCITT) [H55], SMPTE (Society of Motion Picture and Television Engineers) [C31], AVS China (the Audio and Video coding Standard of China) [A51], DIRAC by BBC [D1] [D5], and well known companies, including Microsoft [C32], Real Networks [R3] and On 2 Technologies [P11].

ISO/IEC has developed several video and audio standards including MPEG-1 (ISO/IEC 11172) [S2], MPEG-2 (ISO/IEC 13818) [S3] and MPEG-4 (ISO/IEC 14496) [S8]. ITU-T has also developed several standards, but unlike MPEG, the video standards and audio standards are separate. ITU-T H.261 [S4], H.262 [S3],¹ H.263 [S5], H.263+ (H.263 Version 2) [S6] [S7], H.26L [PC1], and H.264 [S10] [H23] are designed for video, while ITU-T G.723.1 [O14] and G.729 [O15] are for audio. Besides these standards, Video Coder 1 (VC-1) [C11] [C14] and Video Coder 2 (VC-2) [C6] by SMPTE, Windows Media Video 9 [C1] [C2] by Microsoft, VP 6 [P3] [P5] [P6] and VP7 [P4] by On 2 Technologies, Dirac [D1] [D5] by BBC, Real Video 9 and Real Video 10 [R1] [R2] by Real Networks are also popularly used video standards on Internet and personal computers. In recent years, AVS China [A2, A10, A59–A66] has attracted great deal of attention from industries related to television, multimedia communications and even chip manufacturing from around the world. This new standard includes four main technical areas, which are systems, video, audio and digital rights management (DRM), and some supporting documents such as consistency verification. The second part of the standard known as AVS1-P2 [A2] (video–Jizhun) was approved as the national standard of China in 2006, and several final drafts of the standard have been completed, including AVS1-P1 (systems) [A1], AVS1-P2 (video–Jiaqiang) [A3], AVS1-P3 (audio) [A4], and AVS1-P7 (mobile video) [A74].

AVS China provides optimized coding performance with the lowest total cost, including transmission and storage cost, implementation cost and intellectual property rights (IPR) cost because the AVS Working Group took the technical IPR cost into account from the beginning of setting up the standard. There are two aspects of the IPR cost. One is the IPR for the content, which is out of the scope of audio and video coding standards. The other aspect is the IPR cost for the technologies used in the audio and video coding standards.

Some recent research results on AVS China are reported in a special issue of the journal “Signal Processing: Image Communication” in several papers [A58–A66]. An overview of the video part of AVS is given in [A59], which also describes the coding tools available and gives examples of the application-driven profiles defined in AVS. In [A60] two context-based entropy coding schemes for AVS video coding standard are presented. One is the context-based 2D variable length coding (C2DVLC) as a low-complexity entropy coding scheme for AVS Part-2 Jizhun (base) profile. The other is the context-based binary arithmetic coding (CBAC) as an enhanced entropy coding scheme for AVS Part-2 Jiaqiang (enhanced) profile.

¹ H.262 has also audio coding among the several parts.

In [A61], a sub-pixel interpolation filter known as combined adaptive-fixed interpolation (CAFI) with multi-directional filters is proposed to obtain good coding efficiency with low computational complexity. In addition, implementations [A64] [A65], reconfigurable video coding (RVC) framework [A62], trick modes [A63] and robust dual watermarking algorithm [A66] are also discussed in this issue.

Popular video standards are listed in Table 1.1, and the algorithmic element comparisons of these video standards are listed in Table 1.2. Table 1.1 is adopted from T. Ebrahimi and M. Kunt, “Visual data compression for multimedia applications”, Proc. IEEE, vol. 86, pp. 1109–1125, June 1998 [G1]. Some minor changes have been made. Recent standards such as H.264/MPEG-4 Part 10, Dirac, AVS China, JPEG-LS, JPEG-XR, JBIG, VC-1 (SMPTE), VC-2, HEVC/NGVC and VP6 (now VC10) are now added.

1.2 Digital Representation of Video

As video is used to record and/or show moving objects, it is composed of a sequence of pictures taken at regular temporal intervals. The number of frames (pictures) per second is called as the frame rate. Frame rates below 10 frames per second (fps) are sometimes used for very low bit-rate (below 64 kbps) video communications. Between 10 and 20 fps is more typical for low bit-rate video communications. Sampling at 25 or 30 frames per second is standard for television pictures; 50 or 60 frames per second produces smooth apparent motion [B8].

Video can be divided into analog and digital types. Analog video is represented with an analog signal, which is captured by progressive or interlaced scanning using an analog camera. An example of analog video is the signal used in analog television systems such as PAL [G12] and NTSC [G13] [G14]. Digital video is often captured with a digital camera, although it can also be converted from analog video signal. The natural scene is projected onto a sensor, such as an array of charge coupled devices (CCD array) [G15], in a digital camera, which converts the brightness or color of the scene into digital data. Each image (picture) in a video sequence includes M by N picture elements (pixels), where M is the number of rows and N the number of columns. For color images, each pixel is usually composed of three color components, which are Red (R), Green (G) and Blue (B) (abbreviated as RGB). Each color component is separately filtered and projected onto a CCD array. Any color can be created by combining R, G and B in varying proportions. With all possible values of R, G and B, vectors of (R, G, B) form a space called as RGB color space. Each color component is represented with a K -bit integer. For ordinary use, 8 is the sufficient value of K . Larger K is needed for more accurate uses such as medical, broadcast, surveillance, and studio editing, etc.

Although the RGB color space is well-suited to capture and display color images, the YC_bC_r space, formed by vectors of (Y, C_b , C_r), is more efficient for compression, where Y represents luminance (brightness) of a pixel, C_b and C_r are the pixel’s chrominance components proportional to the color differences of B–Y

Table 1.1 Popular video standards [G1] © IEEE 1998

Standard	Standardization body	Main target bitrate	Main compression technologies	Main target applications
JPEG	ISO/IEC	Compression ratio 2-30	DCT Perceptual quantization Zig zag reordering Huffman coding Arithmetic coding Wavelets EBCOT	Internet imaging Digital photography Image and video editing
JPEG2000	ISO/IEC	Compression ratio 2-50		Internet imaging Digital photography Image and video editing Printing Medical imaging Mobile applications Color fax Satellite imaging Region of interest coding, photo and art digital libraries, progressive transmission by pixel accuracy or resolution, laser print rendering, pre-press imagery Lossless and near lossless coding of continuous tone still images
JPEG-LS Part1 and Part2	ISO/IEC		Part1 Context modeling Prediction Golomb codes Part2 Arithmetic coding More effective prediction Modified Golomb coding	Lossless and near lossless coding of continuous tone still images

(continued)

Table 1.1 (continued)

Standard	Standardization body	Main target bitrate	Main compression technologies	Main target applications
JPEG-XR (AIC)	ISO/IEC		Based on HD photo of microsoft (windows media photo)	Storage and interchange of continuous tone photographic content (lossless and lossy)
MPEG-1	ISO/IEC	Bitrates up to about 1.5 Mb/s	DCT Perceptual quantization Adaptive quantization Zig zag reordering Predictive motion compensation Bi-directional motion compensation Half-sample accuracy motion estimation Huffman coding Arithmetic coding	Storage on CD-ROM Consumer video Digital camera
MPEG-2	ISO/IEC	Bitrates 1.5 Mb/s up to about 35 Mb/s	DCT Perceptual quantization Adaptive quantization Zig zag reordering Predictive motion compensation Bi-directional motion compensation Frame/field based motion compensation Half pixel accuracy motion estimation Spatial scalability Quality scalability Huffman coding Arithmetic coding Error resilient coding	Digital TV Digital HDTV High quality video Satellite TV Cable TV Terrestrial broadcast Video editing Video storage Digital camera

(continued)

Table 1.1 (continued)

Standard	Standardization body	Main target bitrate	Main compression technologies	Main target applications
MPEG-4 Part-2 visual	ISO/IEC	Bitrates 8 Kb/s up to about 35 Mb/s	DCT Wavelet Perceptual quantization Adaptive quantization Zig zag reordering Zero tree reordering Predictive motion compensation Bi-directional motion compensation Frame/field based motion compensation Half-sample accuracy motion estimation Advanced motion estimation Overlapping motion compensation Spatial scalability Temporal scalability Quality scalability View dependent scalability Bitmap shape coding Sprite coding Face animation Dynamic mesh coding Huffman coding Arithmetic coding Error resilient coding	Internet Interactive video Visual editing Content manipulation Consumer video Professional video 2D/3D computer graphics Mobile
H.261	ITU-T	Bitrates $p \times 64$ kb/s ($p: 1-31$)	DCT Adaptive quantization Zig zag reordering Predictive motion compensation Integer-sample accuracy motion estimation Huffman coding Error resilient coding	ISDN video-conferencing

(continued)

Table 1.1 (continued)

Standard	Standardization body	Main target bitrate	Main compression technologies	Main target applications
H.262	ITU-T	Bitrates 1.5 Mb/s up to about 35 Mb/s	DCT Perceptual quantization Adaptive quantization Zig zag reordering Predictive motion compensation Bi-directional motion compensation Frame/field based motion compensation Half pixel accuracy motion estimation Spatial scalability Quality scalability Huffman coding Arithmetic coding Error resilient coding	Digital TV Digital HDTV High quality video Satellite TV Cable TV Terrestrial broadcast Video editing Video storage Digital camera
H.263 H.263+ and H.263++	ITU-T	Bitrates 8 Kb/s up to about 1.5 Mb/s	DCT Adaptive quantization Zig zag reordering Predictive motion compensation Bi-directional motion compensation Half-sample accuracy motion estimation Advanced motion estimation Overlapping motion compensation Huffman coding Arithmetic coding Error resilient coding	POTS video-telephony Desktop video telephony Mobile video telephony Video conferencing

(continued)