Video Coding (esp. ITU & ISO/IEC Standards)

Gary J. Sullivan, Ph.D.

ITU-T VCEG Rapporteur | Chair
ISO/IEC MPEG Video Rapporteur | Co-Chair
ITU/ISO/IEC JVT Rapporteur | Co-Chair
IEEE Fellow (2006)

Microsoft Corporation Video Architect

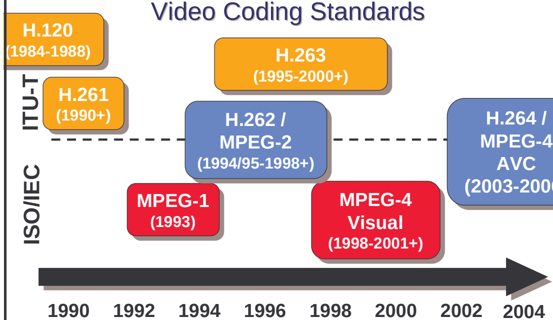
March 2006

Video Coding Standardization Organizations

§ Two organizations have historically dominated general-purpose video compression standardization:
- ITU-T Video Coding Experts Group (VCEG)
  - International Telecommunications Union – Telecommunications Standardization Sector (ITU-T, a United Nations Organization, formerly CCITT), Study Group 16, Question 6
- ISO/IEC Moving Picture Experts Group (MPEG)
  - International Standardization Organization and International Electrotechnical Commission, Joint Technical Committee Number 1, Subcommittee 29, Working Group 11

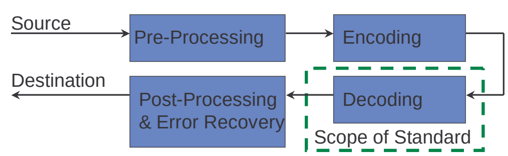
§ Recently, the Society for Motion Picture and Television Engineers (SMPTE) has also entered with "VC-1", based on Microsoft's WMV 9 *but this talk covers only the ITU and ISO/IEC work*.

Chronology of International Video Coding Standards



The *Scope* of Picture and Video Coding Standardization

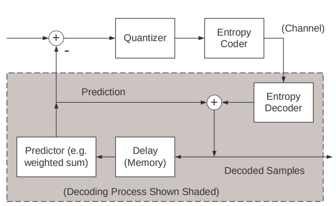
§ Only the *Syntax* and *Decoder* are standardized:
- Permits optimization beyond the obvious
- Permits complexity reduction for implementability
- Provides *no* guarantees of Quality



Predictive Coding and DPCM

§ Separate quantization of each sample is known as pulse-code modulation (PCM)

§ *Predictive Coding or Differential PCM:* Generate an estimate for the value of the input data, and encode only the remaining difference.



H.120 : The First Digital Video Coding Standard

§ ITU-T (ex-CCITT) Rec. H.120: The first digital video coding standard (1984)
- v1 (1984) had conditional replenishment, DPCM, scalar quantization, variable-length coding, switch for quincunx sampling
- v2 (1988) added motion compensation and background prediction
- Operated at 1544 (NTSC) and 2048 (PAL) kbps
- Few units made, essentially not in use today

“Intra” Picture Coding by DCT

Basic “intra” image representation: Discrete Cosine Transform (DCT) (early ‘70s, ITU+ISO JPEG approved ‘92):

- Analyze 8x8 blocks of image according to DCT frequency content (images tend to be smooth)
- Find magnitude of each discrete frequency within the block
- Round off (“quantize”) the amounts to scaled integer values (‘50s, ‘60s, ...)
- Send integer approximations to decoder using “Huffman” variable-length codes (VLC, early ‘50s)

The Discrete Cosine Transform

§ The DCT (unitary type II DCT):

$$F_{m,n}(u,v) = \sum_{x=0}^{M-1} \sum_{y=0}^{N-1} \left(c_u \sqrt{\frac{2}{M}}\right)\left(c_v \sqrt{\frac{2}{N}}\right) f(mM + x, nN + y) \cdot \cos\left[\frac{(2x+1)u\pi}{2M}\right] \cdot \cos\left[\frac{(2y+1)v\pi}{2N}\right]$$

§ The Inverse DCT (unitary type III DCT):

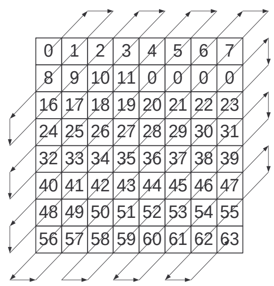
$$\hat{f}(mM + x, nN + y) = \sum_{u=0}^{M-1} \sum_{v=0}^{N-1} \left(c_u \sqrt{\frac{2}{M}}\right)\left(c_v \sqrt{\frac{2}{N}}\right) \hat{F}_{m,n}(u,v) \cdot \cos\left[\frac{(2x+1)u\pi}{2M}\right] \cdot \cos\left[\frac{(2y+1)v\pi}{2N}\right]$$

§ Definition of Constants

$c_u = 1/\sqrt{2}$ for $u = 0$, otherwise 1. $M = 8$ in current visual standards

$c_v = 1/\sqrt{2}$ for $v = 0$, otherwise 1. $N = 8$ in current visual standards

Coefficient Scan Order: The Zig-Zag Scan



Interframe Motion Prediction

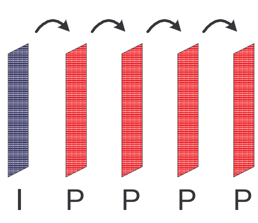
§ Large areas of images stay the same from frame to frame, changing mostly due to motion

§ *Conditional Replenishment:* Can signal to leave a block area of the image unchanged, or replace it with new data

§ *Interframe Difference Coding:* Could encode a refinement to the value of an area

§ *Displaced Frame Difference Coding:* Can predict an image area by copying some nearby part of the previous image (motion compensation) and optionally adding some refinement

P-Picture Predictive Coding



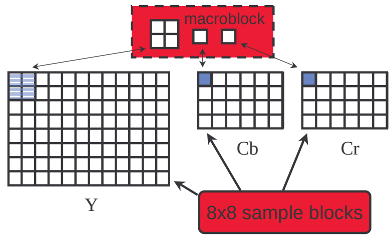
H.261: The Basis of Modern Video Compression

§ ITU-T (ex-CCITT) Rec. H.261: The first widespread practical success

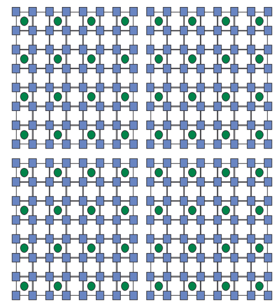
- First design (late ‘90) embodying typical structure dominating today:
  - **16x16 macroblock motion compensation,**
  - **8x8 DCT,**
  - **scalar quantization,**
  - **zig-zag scan, and**
  - **run-length**
  - **variable-length coding**
- Key aspects later dropped by other standards: loop filter, integer motion comp., 2-D VLC, header overhead
- v2 (early ‘93) added a backward-compatible high-resolution graphics trick mode
- Operated at 64-2048 kbps
- Still in use, although mostly as a backward-compatibility feature – overtaken by H.263

Blocks and Macroblocks

The luma and chroma planes are divided into blocks. Luma blocks are associated with Cb and Cr blocks to create a <u>macroblock</u>.



H.261&3 Macroblock Structure



= luma sample

= chroma sample

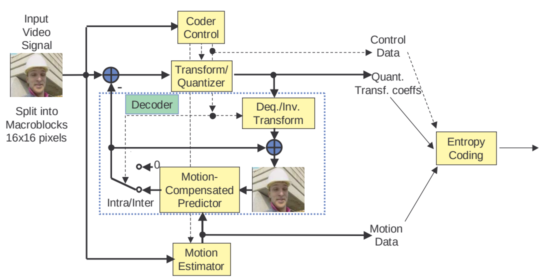
(two chroma fields)

Intra/Inter Decisions:
16x16 macroblock
DCT of 8x8 blocks

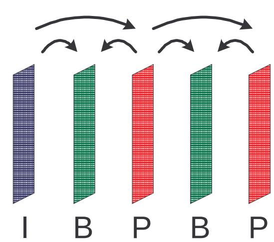
H.261:
16x16 1-pel motion

H.263:
16x16 1/2-pel motion
or (AP mode)
8x8 1/2-pel motion
with overlapping

Basic Hybrid Structure of H.261, etc. (late '90)



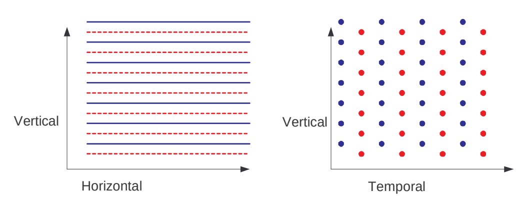
Predictive Coding with (old-fashioned) B Pictures



MPEG-1: Practical Video at Higher Rates than H.261

§ Formally ISO/IEC 11172-2 ('93), developed by ISO/IEC JTC1 SC29 WG11 (MPEG) – use is fairly widespread (esp. Video CD in Asia), but mostly overtaken by MPEG-2
- Superior quality to H.261 when operated a higher bit rates (≥ 1 Mbps for CIF 352x288 resolution)
- Can provide approximately VHS quality between 1-2 Mbps using SIF 352x240/288 resolution
- Technical features inherited from H.261
  - **16x16 macroblocks**
  - **16x16 motion compensation,**
  - **8x8 DCT,**
  - **scalar quantization,**
  - **zig-zag scan, and**
  - **run-length**
  - **variable-length coding**
- Technical features added:
  - **Bi-directional motion prediction**
  - **Half-pixel motion**
  - **Slice-structured coding**
  - **DC-only "D" pictures**
  - **Quantization weighting matrices**

Interlaced Video (Welcome to the 1940 Analog World)



MPEG-2/H.262: Even Higher Bit Rates and Interlace

§ Formally ISO/IEC 13818-2 & ITU-T H.262, developed ('94) jointly by ITU-T and ISO/IEC SC29 WG11 (MPEG) – Now in very wide use for DVD and standard and high-definition DTV (the most commonly used video coding standard)
- Primary new technical features:
  - **Support for interlaced-scan pictures**
  - **Increased DC quantization precision**
- Also
  - **Various forms of scalability (SNR, Spatial, breakpoint)**
  - **I-picture concealment motion vectors**
- Essentially the same as MPEG-1 for progressive-scan pictures, and MPEG-1 forward compatibility required
- Not especially useful below 2-3 Mbps (range of use normally 2-5 Mbps SDTV broadcast, 6-8 DVD, 20 HDTV)
- Essentially fixed frame rate

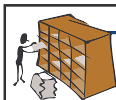
H.263: The Next Generation

§ ITU-T Rec. H.263 (v1: 1995): The next generation of video coding performance, developed by ITU-T – the current premier ITU-T video standard (has overtaken H.261 as dominant videoconferencing codec)
- **Superior quality to prior standards at all bit rates (except perhaps for interlaced video)**
- **Better by a factor of two at very low rates**
- **Versions 2 (late 1997/early 1998) & v3 (2000) later developed with a large number of new features**
- **Profiles defined early 2001**
- A somewhat tangled relationship with MPEG-4

What Was in H.263 Version 1?

§ "Baseline" Algorithm Features beat H.261
- Half-pel motion compensation (also in MPEG-1)
- 3-D variable length coding of DCT coefficients
- Median motion vector prediction
- More efficient coding pattern signaling (?)
- Deletable GOB header overhead (also in MPEG-1, but not 2?)

§ Optional Enhanced Modes
- Increased motion vector range with picture extrapolation
- Variable-size, overlapped motion with picture extrapolation
- PB-frames (bi-directional prediction)
- Arithmetic entropy coding
- Continuous-presence multipoint / video mux



H.263+ Feature Categories

§ Error resilience

§ Improved compression efficiency (e.g., 15-25% overall improvement over H.263v1)

§ Custom and Flexible Video Formats

§ Scalability for resilience and multipoint

§ Supplemental enhancement information

H.263++ Version 3 Features

§ **Annex U**: Fidelity enhancement by macroblock and block-level reference picture selection – a significant improvement in picture quality

§ **Annex V**: Packet Loss & Error Resilience using data partitioning with reversible VLCs (roughly similar to MPEG-4 data partitioning, but improved by using reversible coding of motion vectors rather than coefficients)

§ **Annex W**:Additional Supplemental Enhancement Information
- IDCT Mismatch Elimination (specific fixed-point fast IDCT)
- Arbitrary binary user data
- Text messages (arbitrary, copyright, caption, video description, and URI)
- Error Resilience:
  - Picture header repetition (current, previous, next+TR, next-TR)
  - Spare reference pictures for error concealment
- Interlaced field indications (top & bottom)

MPEG-4 "**Visual**": Baseline H.263 and Many Creative Extras

§ MPEG-4 **part 2** (v1: early 1999), formally ISO/IEC 14496-2

§ Contains the H.263 baseline design
- **coding efficiency enhancements (esp. at low rates)**

§ Adds many creative new extras:
- **more coding efficiency enhancements**
- **error resilience / packet loss enhancements**
- **segmented coding of shapes**
- **zero-tree wavelet coding of still textures**
- **coding of synthetic and semi-synthetic content,**
- **10 & 12-bit sampling,**
- **more**
- **v2 (early 2000) & v3 (early 2001) & …**

MPEG-4 **Visual** Focus: Simple Profile

- The most basic video coding profile of MPEG-4
- No shape coding
- Progressive-scan video only
- Most popular in low cost / low rate / low resolution apps (e.g., mobile) – top bit rate & resolution limited
- Basic contents
  - H.263 baseline
  - Motion vectors over picture boundaries
  - Variable block-size motion compensation
  - Intra DCT coefficient prediction
  - Handling of four streams in most levels
  - Error / packet-loss features – data partitioning, RVLC

MPEG-4 **Visual** Focus: Advanced Simple Profile

- Target goal: General rectangular video with improved coding efficiency
- Progressive-scan and interlaced video support
- Up to SDTV resolution
- Basic contents
  - All of Simple profile
  - B pictures
  - Global motion compensation
  - Quarter-sample motion compensation
  - Interlace handling

MPEG-4 **Visual** Focus: Studio Profile

- Target goal: studio & professional use
- Progressive-scan and interlaced video support
- Up to very high resolution and bit rate
- Basic contents
  - Enhanced-accuracy IDCT
  - B pictures
  - 10 & 12 bit sample accuracy
  - 4:2:2 & 4:4:4 chroma sampling structures

The H.264/MPEG-4 Advanced Video Coding (AVC) Standard

Gary J. Sullivan, Ph.D.

ITU-T VCEG Rapporteur | Chair
ISO/IEC MPEG Video Rapporteur | Co-Chair
ITU/ISO/IEC JVT Rapporteur | Co-Chair

Microsoft Corporation Video Architect

March 2006

The Advanced Video Coding Project AVC = ITU-T H.264 / MPEG-4 part 10

- History: ITU-T Q.6/SG16 (**VCEG - Video Coding Experts Group**) "H.26L" standardization activity (where the "L" stood for "long-term")
- **Aug 1999**: 1st test model (TML-1)
- **July 2001:** MPEG open call for technology: H.26L demo'ed
- **Dec 2001**: Formation of the **Joint Video Team (JVT)** between VCEG and MPEG to finalize H.26L as a new joint project (similar to MPEG-2/H.262)
- **July 2002**: Final Committee Draft status in MPEG
- **Dec '02** technical freeze, FCD ballot approved
- **May '03** completed in both orgs
- **July '04** Fidelity Range Extensions (FRExt) completed
- **Jan '05** Scalable Video Coding launched

H.264/MPEG-4 AVC Objectives

- **Primary technical objectives:**
  - Significant improvement in coding efficiency
  - High loss/error robustness
  - "Network Friendliness" (carry it well on MPEG-2 or RTP or H.32x or in MPEG-4 file format or MPEG-4 systems or …)
  - Low latency capability (better quality for higher latency)
  - Exact match decoding
- **Additional version 2 objectives (in FRExt):**
  - Professional applications (more than 8 bits per sample, 4:4:4 color sampling, etc.)
  - Higher-quality high-resolution video
  - Alpha plane support (a degree of "object" functionality)

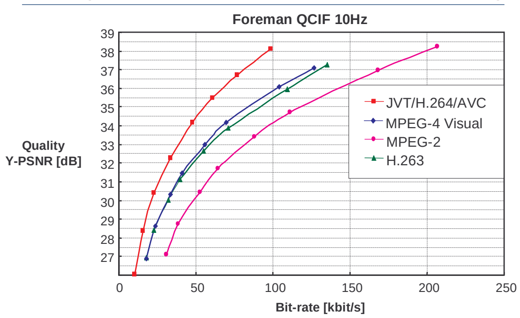
Relating to Other ITU & MPEG Standards

- **Same design to be approved in both ITU-T VCEG and ISO/IEC MPEG**
- **In ITU-T VCEG this is a new & separate standard**
  - ITU-T Recommendation H.264
  - ITU-T Systems (H.32x) support it
- **In ISO/IEC MPEG this is a new "part" in the MPEG-4 suite**
  - Separate codec design from prior MPEG-4 visual
  - New part 10 called "Advanced Video Coding" (AVC – similar to "AAC" position in MPEG-2 as separate codec)
  - Not backward or forward compatible with prior standards (incl. the prior MPEG-4 visual spec. – core technology is different)
  - MPEG-4 Systems / File Format supports it
- **H.222.0 | MPEG-2 Systems also supports it**

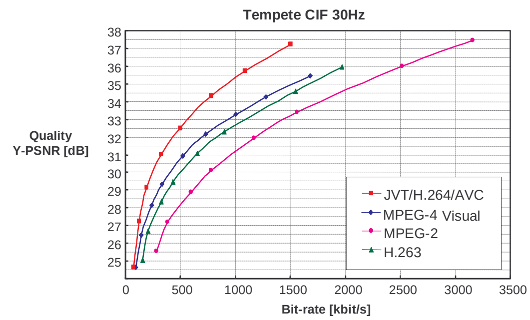
A Comparison of Performance

- Test of different standards (ICIP 2002 study)
- Using same rate-distortion optimization techniques for all codecs
- Streaming test: High-latency (included B frames)
  - Four QCIF sequences coded at 10 Hz and 15 Hz (Foreman, Container, News, Tempete) and
  - Four CIF sequences coded at 15 Hz and 30 Hz (Bus, Flower Garden, Mobile and Calendar, and Tempete)
- Real-time conversation test: No B frames
  - Four QCIF sequences encoded at 10Hz and 15Hz (Akiyo, Foreman, Mother and Daughter, and Silent Voice)
  - Four CIF sequences encoded at 15Hz and 30Hz (Carphone, Foreman, Paris, and Sean)
- Compare four codecs using PSNR measure:
  - **MPEG-2** (in high-latency/streaming test only)
  - **H.263** (high-latency profile, conversational high-compression profile, baseline profile)
  - **MPEG-4 Visual** (simple and advanced simple profiles with & without B pictures)
  - **H.264/AVC** (with & without B pictures)
- Note: These test results are from a private study and are not an endorsed report of the JVT, VCEG or MPEG organizations.

Comparison to MPEG-2, H.263, MPEG-4p2



Comparison to MPEG-2, H.263, MPEG-4p2



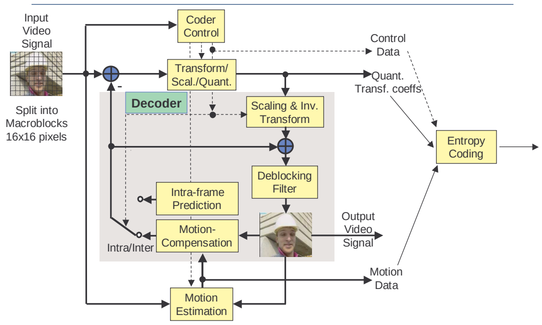
Caution: Your Mileage **Will** Vary

- This encoding software may not represent **implementation quality**
- These tests **only up to CIF** (quarter-standard-definition) resolution
- **Interlace, SDTV, and HDTV not tested** in this test
- Test sequences **may not be representative** of the variety of content encountered by applications
- These tests so far **not aligned** with profile designs
- This study reports PSNR, but **perceptual** quality is what matters

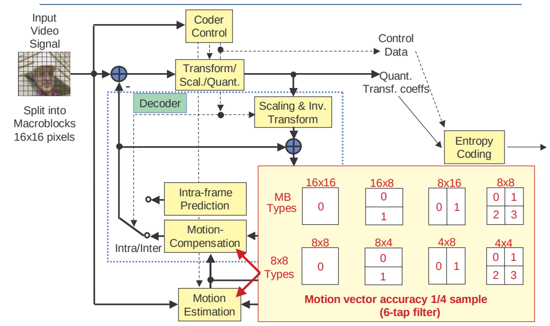
Computing Resources for the New Design

- New design includes relaxation of traditional bounds on computing resources – rough guess 2-3x the MIPS, ROM & RAM requirements of MPEG-2 for decoding, 3-4x for encoding
- Particularly an issue for low-power (e.g., mobile) devices
- Problem areas:
  - Smaller block sizes for motion compensation (cache access issues)
  - Longer filters for motion compensation (more memory access)
  - Multi-frame motion compensation (more memory for reference frame storage)
  - In-loop deblocking filter (more processing & memory access)
  - More segmentations of macroblock to choose from (more searching in the encoder)
  - More methods of predicting intra data (more searching)
  - Arithmetic coding (adaptivity, computation on output bits)

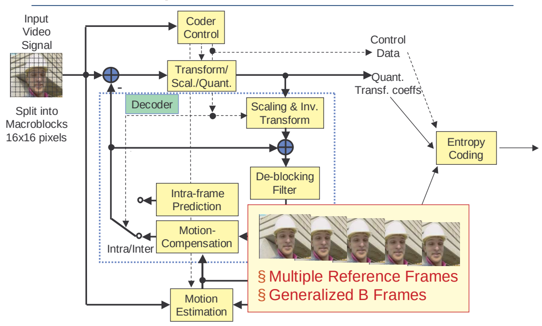
H.264/MPEG-4 AVC Structure



Motion Compensation Accuracy

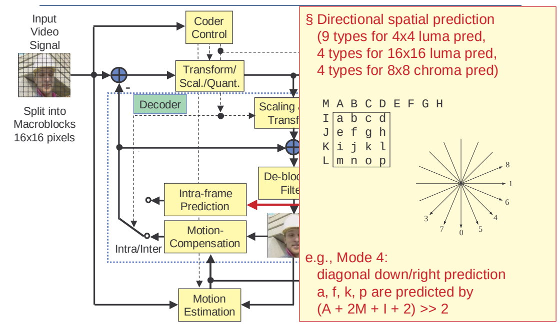


Multiple Reference Frames



- Multiple Reference Frames
- Generalized B Frames

Intra Prediction



- Directional spatial prediction (9 types for 4x4 luma pred, 4 types for 16x16 luma pred, 4 types for 8x8 chroma pred)

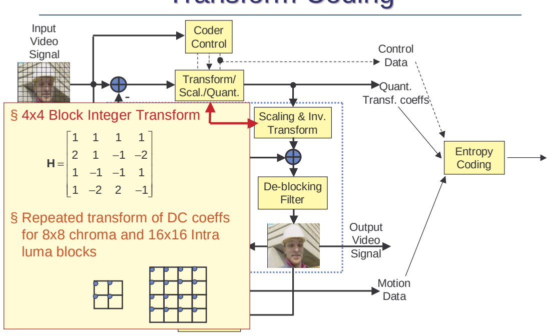
e.g., Mode 4: diagonal down/right prediction a, f, k, p are predicted by (A + 2M + I + 2) >> 2

Transform Coding

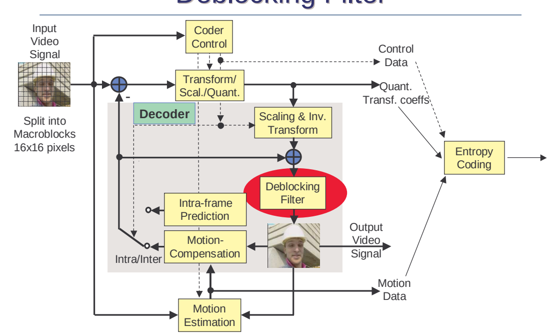
- 4x4 Block Integer Transform

$$\mathbf{H} = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 2 & 1 & -1 & -2 \\ 1 & -1 & -1 & 1 \\ 1 & -2 & 2 & -1 \end{bmatrix}$$

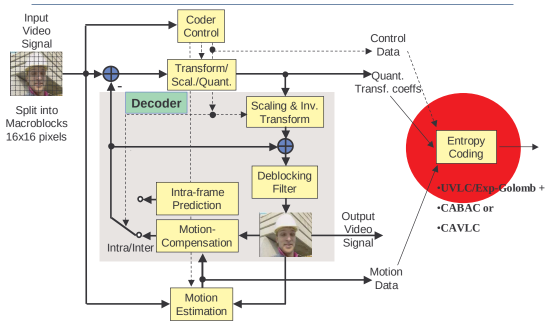
- Repeated transform of DC coeffs for 8x8 chroma and 16x16 Intra luma blocks



Deblocking Filter



Entropy Coding



AVC Version 1 Profiles

- **Three profiles in version 1: Baseline, Main, and Extended**
- **Baseline (esp. Videoconferencing & Wireless)**
  - I and P progressive-scan picture coding (not B)
  - In-loop deblocking filter
  - 1/4-sample motion compensation
  - Tree-structured motion segmentation down to 4x4 block size
  - VLC-based entropy coding
  - Some enhanced error resilience features
    - Flexible macroblock ordering/arbitrary slice ordering
    - Redundant slices

Non-Baseline AVC Version 1 Profiles

- **Main Profile (esp. Broadcast)**
  - All Baseline features **except** enhanced error resilience features
  - Interlaced video handling
  - Generalized B pictures
  - Adaptive weighting for B and P picture prediction
  - ***CABAC (arithmetic entropy coding)***
- **Extended Profile (esp. Streaming)**
  - All Baseline features
  - Interlaced video handling
  - Generalized B pictures
  - Adaptive weighting for B and P picture prediction
  - ***More error resilience: Data partitioning***
  - ***SP/SI switching pictures***

Amendment 1: Fidelity-Range Extensions

- AVC standard finished 2003
  - ITU-T/H.264 finalized May, 2003
  - MPEG-4 AVC finalized July, 2003 (same thing)
  - Only corrigenda (bug fixes) since then
- Fidelity-Range Extensions (FRExt)
  - New work item initiated in July 2003
  - More than 8 bits, color other than 4:2:0
  - Alpha coding
  - More coding efficiency capability
  - Also new supplemental information

FRExt Finished July 04

- Project initiated July 2003
  - Motivations
    - Higher quality, higher rates
    - 4:4:4, 4:2:2, and also 4:2:0
    - 8, 10, or 12 bits (14 bits considered and not included)
    - Lossless
    - Stereo
- Finished in one year! (July 04)

New Things in FRExt – Part 1

- Larger transforms
  - 8x8 transform (again!)
  - Drop 4x8, 8x4, or larger, 16-point…
- Filtered intra prediction modes for 8x8 block size
- Quantization matrix
  - 4x4, 8x8, intra, inter trans. coefficients weighted differently
  - Old idea, dating to JPEG and before (circa 1986?)
  - Full capabilities not yet explored (visual weighting)
- Coding in various color spaces
  - 4:4:4, 4:2:2, 4:2:0, Monochrome, with/without Alpha
  - New integer color transform (a VUI-message item)

New Things in FRExt – Part 2

- Efficient lossless interframe coding
- Film grain characterization for analysis/synthesis representation
- Stereo-view video support
- Deblocking filter display preference

8x8 16-Bit (Bossen) Transform

$$\begin{bmatrix} 8 & 8 & 8 & 8 & 8 & 8 & 8 & 8 \\ 12 & 10 & 6 & 3 & -3 & -6 & -10 & -12 \\ 8 & 4 & -4 & -8 & -8 & -4 & 4 & 8 \\ 10 & -3 & -12 & -6 & 6 & 12 & 3 & -10 \\ 8 & -8 & -8 & 8 & 8 & -8 & -8 & 8 \\ 6 & -12 & 3 & 10 & -10 & -3 & 12 & -6 \\ 4 & -8 & 8 & -4 & -4 & 8 & -8 & 4 \\ 3 & -6 & 10 & -12 & 12 & -10 & 6 & -3 \end{bmatrix}$$

8x8 Transform Advantage (JVT-K028, IBBP coding, prog. scan)

Sequence	% BD bit-rate reduction
Movie 1	11.59
Movie 2	12.71
Movie 3	12.01
Movie 4	11.06
Movie 5	13.46
Crawford	10.93
Riverbed	15.65
Average	**12.48**

Quantization Matrix

- Similar concept to MPEG-2 design
- Vary step size based on frequency
- Adapted to modified transform structure
- More efficient representation of weights
- Eight downloadable matrices (at least 4:2:0)
  - Intra 4x4 Y, Cb, Cr
  - Intra 8x8 Y
  - Inter 4x4 Y, Cb, Cr
  - Inter 8x8 Y

New Profiles Created by FRExt

- 4:2:0, 8-bit: "High" (HP)
- 4:2:0, 10-bit: "High 10" (Hi10)
- 4:2:2, 10-bit: "High 4:2:2" (Hi422)
- 4:4:4, 12-bit: "High 4:4:4" (Hi444)

- Effectively the same tools, but acting on different input data (with a couple of exceptions in the 4:4:4 profile)

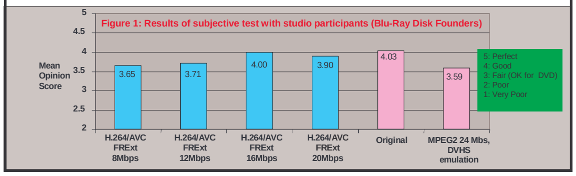
Some Notes on Quality Testing

- Use appropriate "High" profile (incl. adaptive transform)
- If testing for PSNR, use "flat" quant matrices
- Otherwise, use "non-flat" quant matrices
- Use more than 1 or 2 reference pictures
- Use hierarchical reference frames coding structure
- Use CABAC entropy coding
- If testing high-quality PSNR, use adaptive quantization*
- Use rate-distortion optimization in encoder
- Use large-range good-quality motion search
- Use bi-predictive search optimization (see JVT-N014)

*** = See G. Sullivan & S. Sun, "On Dead-Zone…", VCIP 2005/JVT-N011**

A Performance Test for High Profile (from JVT-L033 - Panasonic)

§ Subjective tests by Blu-Ray Disk Founders of FRExt HP
- 4:2:0/8 (HP) 1920x1080x24p (1080p), 3 clips.
- Nominal 3:1 advantage to MPEG-2
  - 8 Mbps HP scored better than 24 Mbps MPEG-2
- Apparent **transparency** at 16 Mbps



For Further Information

§ JVT, MPEG, and VCEG management team members:
- Gary Sullivan (garysull@microsoft.com)
- Jens Ohm (ohm@ient.rwth-aachen.de)
- Ajay Luthra (aluthra@motorola.com)
- Thomas Wiegand (wiegand@hhi.de)

§ IEEE Transactions on Circuits and Systems for Video Technology Special Issue on H.264/AVC (July 2003)

§ Paper in Proceedings of IEEE January 2005 (Sullivan & Wiegand)

§ I. Richardson, *H.264 and MPEG-4 Video Compression,* 2003

§ Overview incl. FRExt: SPIE Aug 2004 (Sullivan, Topiwala, & Luthra)

§ Paper at VCIP 2005: Meta-overview and deployment

§ Wikipedia H.264 page