# H.264/ AVC Video Coding Standard

An overview of the H.264/ AVC Standard

Kevin Loken, P.Eng.

# Background

- MPEG-1
  - Spatial and temporal video coding
  - Predictive, bi-directional
- MPEG-2
  - Added support for interlaced video
  - Widely used for transmission of video, both in standard definition and high definition

# Background

- More high definition TV and lower bandwidth channels require higher coding efficiency
  - Cable modem, DSL have lower bandwidth than broadcast
  - Higher coding efficiency means either higher quality or more channels

# Background

- In 1998 a call for proposals on H.26L was issued
  - Desired to double coding efficiency of existing video codecs
  - That is, half the bit rate for the same fidelity
- Aimed to finalize draft of H.264/ AVC in March 2003

# The Standard

- Only the decoder is standardized
- Imposes restrictions on bitstream and syntax
  - Defines the HRD (hypothetical reference decoder)
- Permits maximum freedom to optimize implementations
  - Doesn't guarantee end-to-end reproduction quality
  - Poor encoding can still be conformant

# Enhancements

- Variable block-size motion compensation
  - Use blocks as small as 4x4
  - MPEG-1 had fixed 8x8 block sizes
- Quarter sample motion compensation
  - Prior standards enabled half-sample motion
  - Reduced complexity from MPEG-4 Visual

# Enhancements

- Motion vectors over picture boundaries
- Multiple reference picture motion compensation
  - Don't have to just reference previous I frame (from H.263)
  - P frames in MPEG-2 used only one previous frame
  - An encoder can select from a large number of stored frames in the decoder

# Enhancements

- Decoupling of referencing order from display order
  - Encoder can choose ordering of pictures for referencing bounded only by available memory
- Decouple picture representation from picture reference capability
  - Not restricted to using I-frames for reference … allows closer temporal frames to be used

# Enhancements

- Weighted prediction
  - Allows motion-compensated prediction signal to be weighted and offset
  - Improves coding efficiency for effects like fades
- In-the-loop deblocking filtering
  - Blocking artifacts can be removed with the application of this filter

# Coding Efficiency

- Small block-size transform
  - Utilize blocks as small as 4x4
- Hierarchical block transform
  - Flexible format allows for longer basis functions, including 8x8 blocks for chroma information
- Short word-length transform
  - Only requires 16-bit arithmetic, not 32

# Coding Efficiency

- Exact match inverse transform
  - Previous standards only provided inverse transform within an error bound
  - H.264/AVC provides exact inverse transform, so all compliant decoders produce exactly the same output
- Arithmetic entropy coding
  - Better compression than Huffman
- Context-adaptive entropy coding

# Robustness

- Parameter set structure
  - Header information handled separately
- NAL unit syntax structure
  - Each structure is placed into a logical data packet (Network Abstraction Layer)
  - Allows NAL unit syntax allows for customization of the method of carrying the signal (depending on network)

# Robustness

- Flexible slice size
  - MPEG 2 had rigid slice size
  - H.264/AVC is flexible (like MPEG-1)
- Flexible macroblock ordering (FMO)
  - Picture can be partitioned into regions (slice)
  - Each slice can be independently decoded

# Robustness

- Arbitrary slice ordering (ASO)
  - Since each slice is independently decodable (approximately), slices can be sent and received out of order
  - This can improve end-to-end delay time on certain networks (like the Internet)
- Redundant pictures
  - An encoder can send redundant regions of pictures (typically lower resolution) for use during loss of data

# Robustness

- Data Partitioning
  - Each region allows for up to three different partitions for transmission
- SP/ SI synchronization/ switching
  - New picture types allow exact synchronization of the decoding process without penalizing other decoders
  - Avoids penalizing some decoders when sending an I picture

# Network Abstraction Layer

- NAL Units
  - Each unit is effectively a packet that contains an integer number of bytes.
  - First byte is header byte
  - Payload data is interleaved with emulation prevention bytes to avoid start code prefix

# Network Abstraction Layer

- NAL Units in Byte Stream
  - Some transport mechanisms are streams of bytes
  - Use start code prefix to identify beginning of NAL unit
  - A small amount of data allows for byte alignment as necessary

# Network Abstraction Layer

- NAL Units in packet transport systems
  - For internet protocol and other packet networks, don't require start code prefix, since packet defines the unit
- VCL and Non-VCL units
  - Some units are video data
  - Other units are parameter sets, data that pertains to a large amount of VCL data

# Network Abstraction Layer

- Parameter Sets
  - Sequence parameter sets apply to a series of consecutive coded video pictures
  - Picture parameter sets apply to the decoding of one or more individual pictures
  - VCL has a reference to the appropriate parameter set
  - Example: video format, entropy encoding

# Network Abstraction Layer

- Access Units
  - A set of NAL units in a specified form
  - Contains VCL NAL units for the primary picture
  - May also contain additional VCL for redundant pictures
  - Supplemental enhancement information (like timing)

# Video Coding Layer

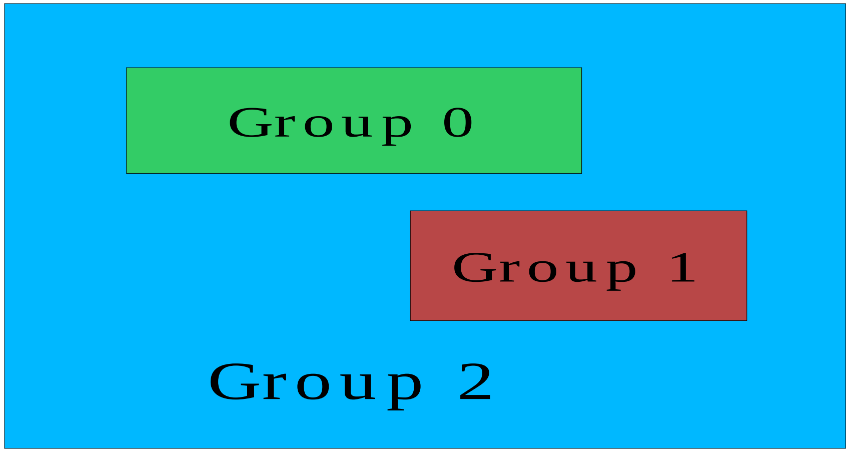
- Pictures, Frames and Fields
  - A video sequence is a set of coded pictures
  - A picture can be either a frame or a field (for interlaced video)
- Macroblocks
  - A picture is partitioned into fixed size blocks
  - 16x16 samples for luma, 8x8 for each of the two chroma channels

# Video Coding Layer

- Colour space
  - YCbCr colour space
  - 4:2:0 sampling
  - Y – luma, CbCr – chroma
  - Humans are more sensitive to luma, so it has four times the samples of the chroma channels.

# Video Coding Layer

- Slices and Slice Groups
  - A picture in H.264/AVC can be a collection of one or several slices
  - Each slice is a series of macroblock



# Video Coding Layer

- Adaptive Frame/Field Coding
  - For interlaced (as opposed to progressive) video, H.264/AVC allows for separate encoding of each field.
  - Frame/field encoding decision can be made independently for each vertical pair of macroblocks (16x32 luma region).

# Video Coding Layer

- Intra-Frame Prediction
  - Each macroblock can be transmitted in a number of different coding schemes
  - Intra_4x4
    - Good for detail
  - Intra_16x16
    - Good for smooth areas
  - I_PCM
    - Transmit samples directly (no prediction)

# Video Coding Layer

- Inter-frame Prediction
  - Partitions of luma block sizes 16x16, 16x8, 8x16, and 8x8 are supported
  - Accuracy of motion compensation is in units of one quarter of the distance between luma samples
  - Allows motion compensation over picture boundaries
  - Doesn't allow motion compensation over slice boundaries

# Summary

- Separation of standard into NAL and VCL
- Added robustness through ASO and FMO
- Improved coding efficiency
  - Small blocks, 16-bit arithmetic
- Improved quality
  - Better predictive coding, smaller blocks, in-the-loop deblocking

# Discussion

- Good points?
- Bad points?