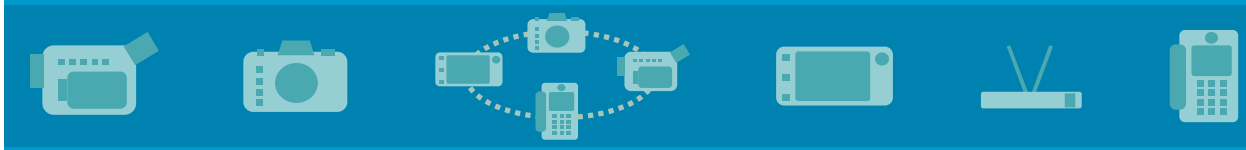
Ittiam

i think therefore i am

Tutorial:
The H.264 Advanced Video Compression Standard

By:
Sriram Sethuraman
Ittiam Systems (Pvt.) Ltd., Bangalore

IEEE Multimedia Compression Workshop
October 27, 2005
Bangalore



world's most preferred DSP IP supplier 2004

DSP Professionals Survey by Forward Concepts

Overview

- Motivation – comparison against other standards
- AVC in the market
- Standards history & Tools progression
- AVC Coding Tools – how they work
- Fidelity Range Extension (FREXT) tools
- Profiles & Levels
- SEI and VUI
- JM – Brief description
- Implementation aspects
- Carriage of AVC (MPEG-2 TS / RTP)
- Storage of AVC (File format)
- Scalable Video Coding
- References

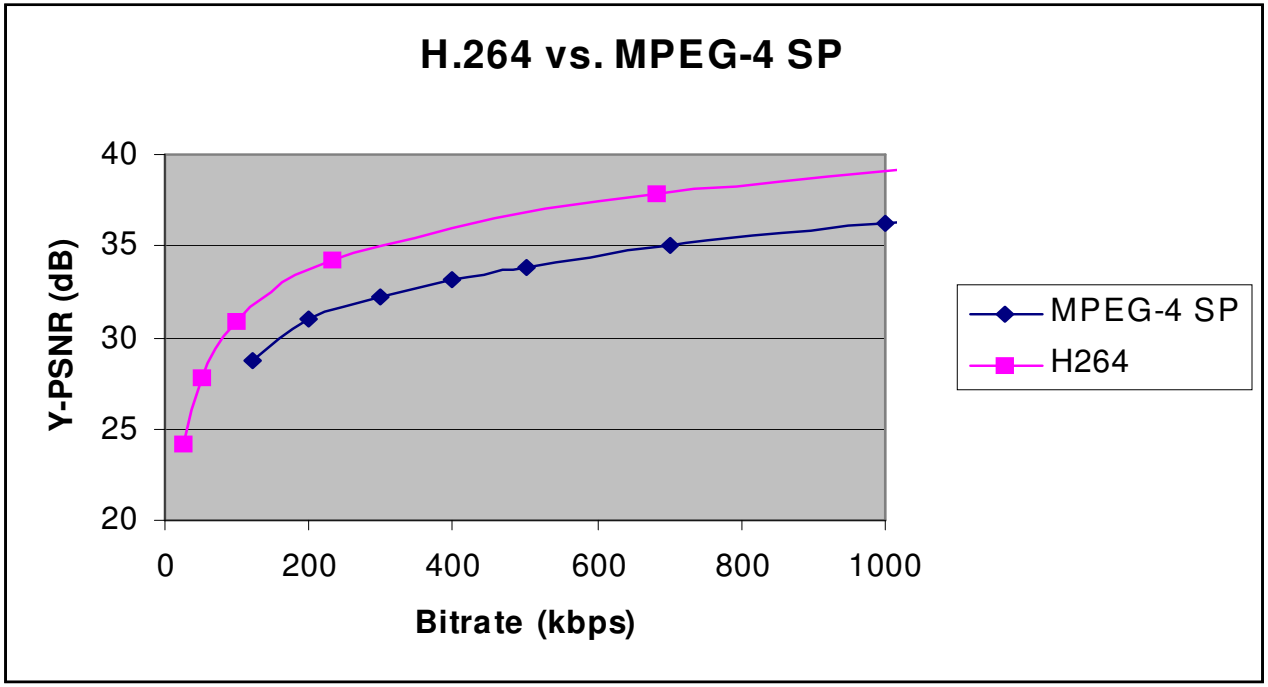
Ittiam

AVC in the market

- 50+ companies that have already announced products
- Span a broad range of product categories
  - DVB Broadcasting (HD/SD)
    - Harmonic, Tandberg
  - Digital Multimedia Broadcast (DMB) or ISDBT
    - Several companies in Korea, Japan
  - IPTV/VoD
    - Skystream, Minerva
  - Portable media player
    - Sony PSP, Apple's video iPod
  - ASICs
    - Broadcom, Conexant, Sigma Designs, ST Micro
  - STBs
    - Pace, Scientific Atlanta, Amino, Sentivision, Ateme, etc.
  - Video conferencing systems
    - Polycom
  - DSP IP
    - Ittiam, Ateme, Ingenient, Sentivision, etc.
  - RTL IP
    - Sciworx, Amphion, etc.
  - PC based
    - QuickTime 7, Main Concept, Elecard
  - Analysis tools
    - Tektronix (Vprov), Interra

Ittiam

H.264 – Compression Advantage



- **H.264 with CABAC, no B-frames, 1 reference frame**
- **Sequence used was foreman CIF, 240 frames**

Ittiam

Video Coding Standards - History

- ITU-T H.261 (1990)
- ISO 11172-2 (MPEG-1, 1993)
- ITU-T H.262 /ISO 13818-2 (MPEG-2, 1995)
- ITU-T H.263 (1996)
- ITU-T H.263+ (1998)
- ISO 14496-2 (MPEG-4, 1999)
- ITU-T H.263++ (2000)
- **ITU-T H.264 / ISO 14496-10 (2003)**
- **ITU-T H.264 FREXT (2004)**
- **Corrigendum and amendments (2005)**

Ittiam

VCEG vs. MPEG

- VCEG (Video Coding Experts Group)
  - ITU-T SG16Q6
  - Focus on video coding for communication
  - H.324, H.323 umbrella standards for systems
- MPEG (Moving Picture Experts Group)
  - ISO/IEC/JTC1/SC29/WG11
  - Focus on video coding for entertainment
  - Specifies carriage of media in Systems specifications
- H.264 is a joint standard
  - ITU-T H.264
  - ISO/IEC 14496-10

Ittiam

Standards vs. Applications

- H.261
  - Video telephony
- MPEG-1
  - CD-ROM
- MPEG-2
  - DTV, DVD, Studio
- H.263, +, ++
  - Video telephony, security, MMS
- MPEG-4
  - Streaming, security, network camera
- H.264
  - Video telephony, streaming, storage (HD-DVD/BD-ROM), Digital Cinema

Ittiam

Coding Tools Progression

- Motion Compensated Transform Coding of a macroblock
- Intra coding
  - DC prediction
  - Improved intra prediction (AC/DC prediction)
- Inter coding
  - Half/Quarter pixel accurate motion compensation
  - Bidirectional motion compensation
  - Variable block size motion compensation
  - Unrestricted motion compensation
- Loop filter
  - Simple in-block smoothing
  - Deblocking filter
- Interlaced coding tools
- 2D/3D VLCs
- Error resilience tools (Resync, Data partitioning, RVLC)
- Scalability (Spatial/Temporal/SNR/FGS)

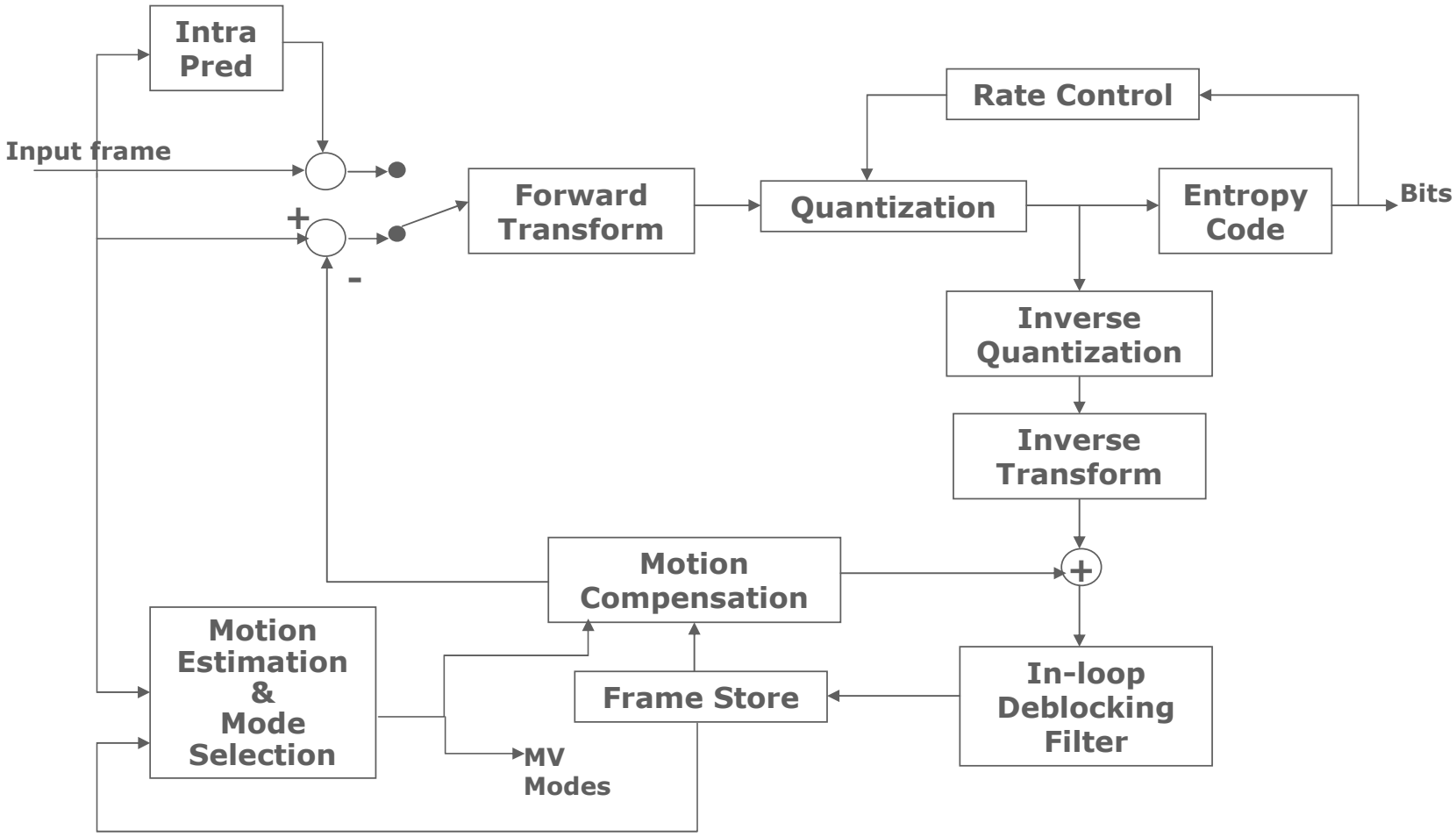
Ittiam

H.264 - Features

- 4x4 block sizes
- Integer transform
  - No mismatch between encoder and decoder
- Low complexity, wider range quantization
- Spatial intra prediction
  - Compression efficiency on par with JPEG-2000
- Enhanced temporal prediction
  - Variable block size motion compensation
  - Multiple reference frames
  - Multi-hypothesis MC
  - Anti-aliased sub-pixel motion compensation
    - Quarter pel accurate luma/One-eight pel accurate chroma MC
  - Weighted prediction
- Efficient entropy coding
  - Content adaptive VLC
  - Context adaptive binary arithmetic coding (CABAC)
- Content and quantizer adaptive in-loop deblocking filter
- Picture adaptive and MB-adaptive Frame/Field coding

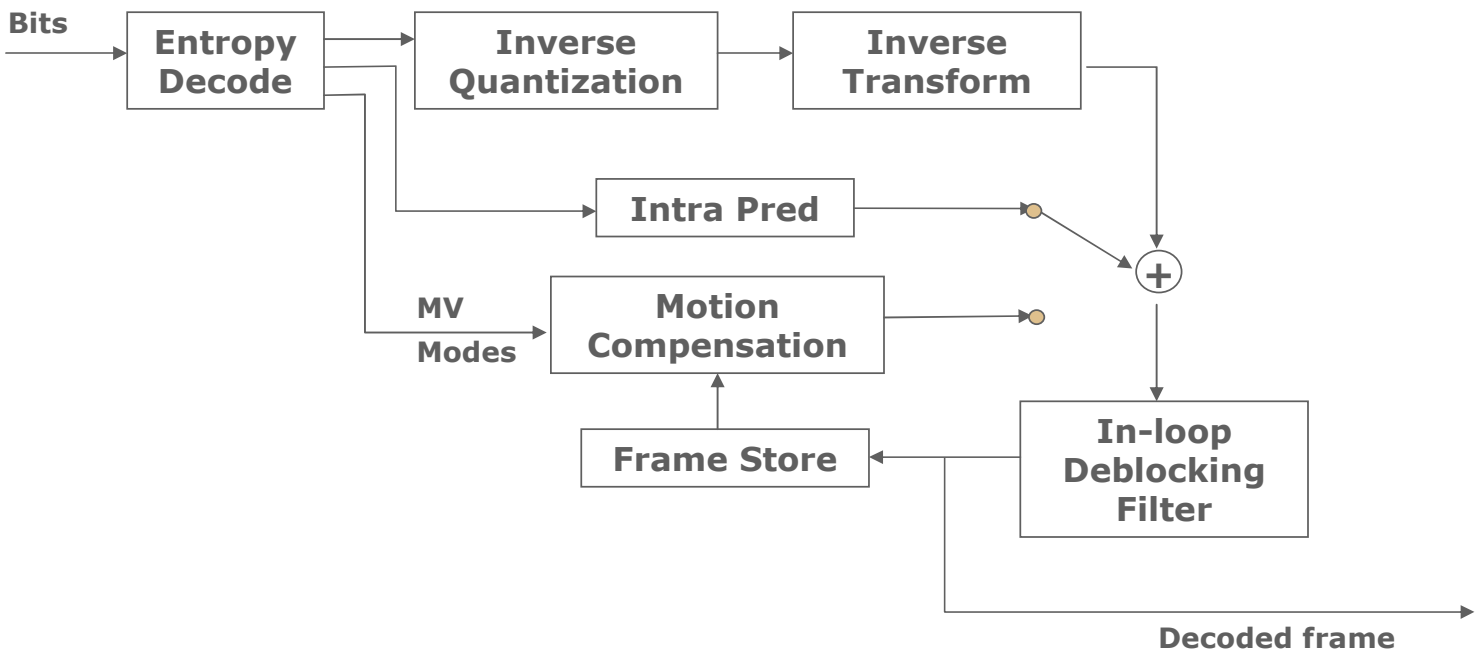
Ittiam

H.264 Encoder



Ittiam

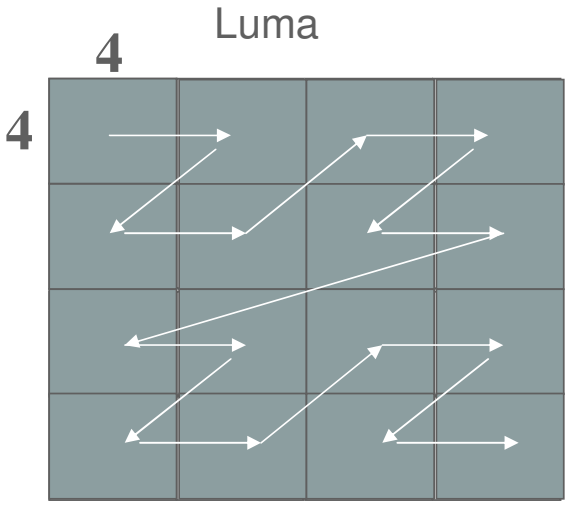
H.264 Decoder

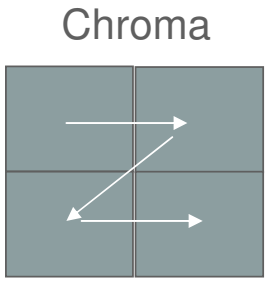


Ittiam

Coding Structure

- Coding mode: 4:2:0 YCbCr
- Sequence
  - Access Unit
  - Coded Picture – loosely defined
    - Complementary field picture pair
- Slices
  - Predictions suppressed across slice boundaries
  - I, P, B slices
    - B-Pic can have I/P/B slices
    - B-Pic can be used as a reference picture
  - Macroblock pairs (only used in MBAFF)
  - Macroblock (16x16 spatial area)
    - 16x16 Y, 8x8 Cb, 8x8 Cr
  - Blocks (8x8)
    - Raster scan within blocks
    - Sub-blocks (4x4)
      - Scan pattern within sub-blocks





Ittiam

SPS and PPS

- Sequence Parameter Set
  - Profile and level
  - Chroma format
  - Width and height
  - Bit-depth
  - Picture order count (POC) type
  - Progressive only coding
    - 8x8 inference
  - Maximum number of reference pictures
  - Cropping parameters
- Picture Parameter Set
  - Entropy coding type
  - Slice group info
  - Num. of active references
  - Weighted prediction ON/OFF
    - Type of weighted prediction
  - Picture quantizer
  - Deblocking filter control
  - Constrained intra-pred use
  - Custom quantization matrices

- Can be sent in bitstream or sent out of band or can be pre-agreed
- IDs for SPS and PPS
  - Which PPS belongs to which SPS; which slice belongs to a given PPS
- Up to 32 SPSs and 256 PPSs are supported at a time

Ittiam

Intra Prediction

- Predicted spatially from reconstructed pixels of causal neighbors
  - DPCM at a block level
- Several modes
  - Four 16x16 block prediction modes for luma
    - H, V, DC, Plane fit (from left, top, and top-left pixels)➔
  - Nine 4x4 sub-block prediction modes for luma
    - Interpolate or extend reconstructed causal neighbor pixels
      - Various orientations to take care off different textures
      - Uses left, top, top-left, and top-right pixels
- Chroma prediction (4 8x8 modes – similar to 16x16 luma modes)
- Increased coding efficiency
  - Mode signalling overhead vs. Texture coding bits
  - Performance of intra coding close to JPEG-2000
- Increased complexity
- Pipeline complications

Ittiam

Transform & Quantization

- Integer approximation to DCT kernel ➡
  - No mismatch
  - 16-bit intermediate precision
- Second level Hadamard transform of DC
  - 4x4 DC block for 16x16 Intra-predicted MBs
  - 2x2 DC block for chroma
- Low complexity, wider range quantization
  - Quant with only multiplications and shifts
  - 52 quantization steps
  - 6 steps lead to doubling of quantization scale factor
    - ⇨ this fact is used in scaling and inverse scaling
    - ⇨ Transform normalization combined with quantization

Ittiam

DCT factorization

$$Y = \left(CXC^T\right) \otimes E =$$

$$\left(\begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & d & -d & -1 \\ 1 & -1 & -1 & 1 \\ d & -1 & 1 & -d \end{bmatrix} \begin{bmatrix} x_{00} & x_{01} & x_{02} & x_{03} \\ x_{10} & x_{11} & x_{12} & x_{13} \\ x_{20} & x_{21} & x_{22} & x_{23} \\ x_{30} & x_{31} & x_{32} & x_{33} \end{bmatrix} \begin{bmatrix} 1 & 1 & 1 & d \\ 1 & d & -1 & -1 \\ 1 & -d & -1 & 1 \\ 1 & -1 & 1 & -d \end{bmatrix} \right) \otimes \begin{bmatrix} a^2 & ab & a^2 & ab \\ ab & b^2 & ab & b^2 \\ a^2 & ab & a^2 & ab \\ ab & b^2 & ab & b^2 \end{bmatrix}$$

$a = 1/2$

$b = \sqrt{1/2} \times \cos(p/8)$

$c = \sqrt{1/2} \times \cos(3p/8)$

d = c/b = sqrt(2) - 1

- By choosing d=1/2 instead of 0.4412, the transform can be done with only shifts/adds.
- To maintain orthogonality, adjust b and c (b=2, c=1).
- Combine E matrix with quantization

Ittiam

Transforms

Forward Transform Matrix

$$\begin{pmatrix} 1 & 1 & 1 & 1 \\ 2 & 1 & -1 & -2 \\ 1 & -1 & -1 & 1 \\ 1 & -2 & 2 & -1 \end{pmatrix}$$

- Orthogonal bases
- Not orthonormal
  - Norms absorbed while quantizing
- Shifts and adds only
- Small values ensure that intermediate values are within 16-bits

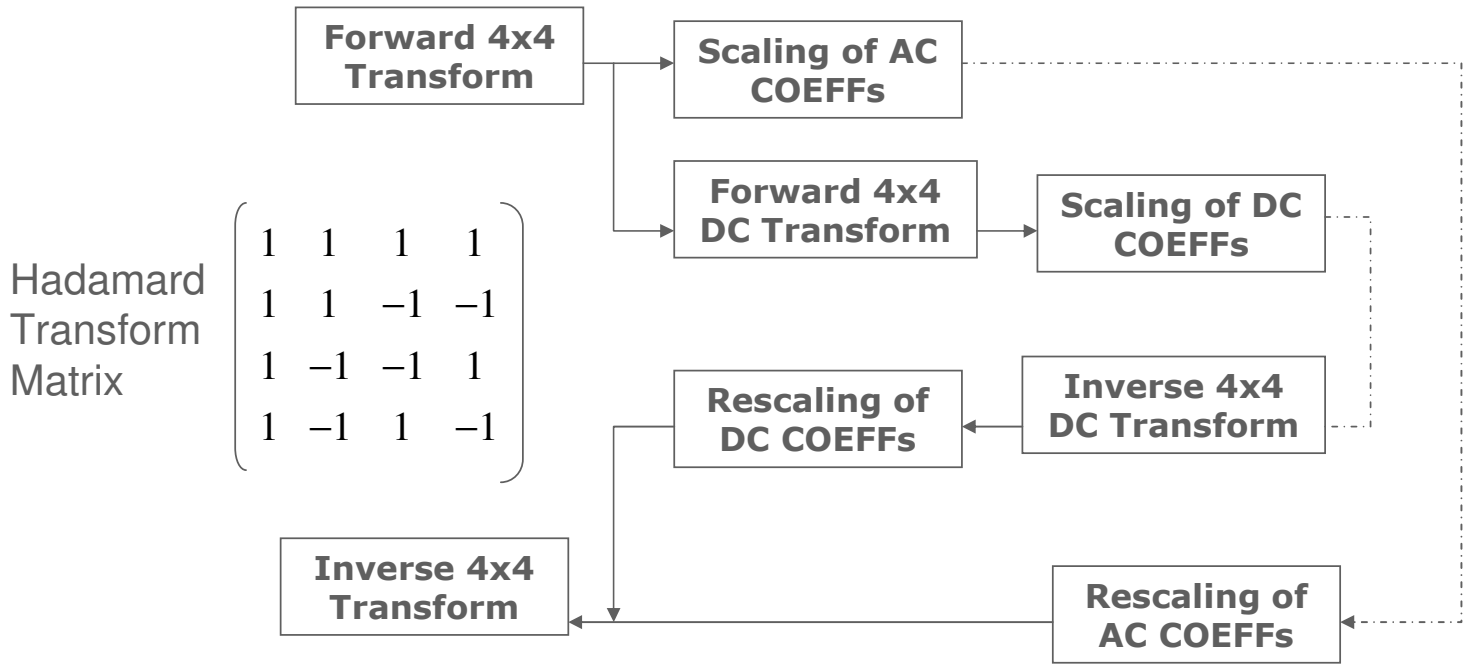
Inverse Transform Matrix

$$\begin{pmatrix} 1 & 1 & 1 & \frac{1}{2} \\ 1 & \frac{1}{2} & -1 & -1 \\ 1 & -\frac{1}{2} & -1 & 1 \\ 1 & -1 & 1 & -\frac{1}{2} \end{pmatrix}$$

- Modified transpose of forward transform to keep dynamic range low
- Sign preserving right shift to compute 0.5
- Norms absorbed while inverse quantizing

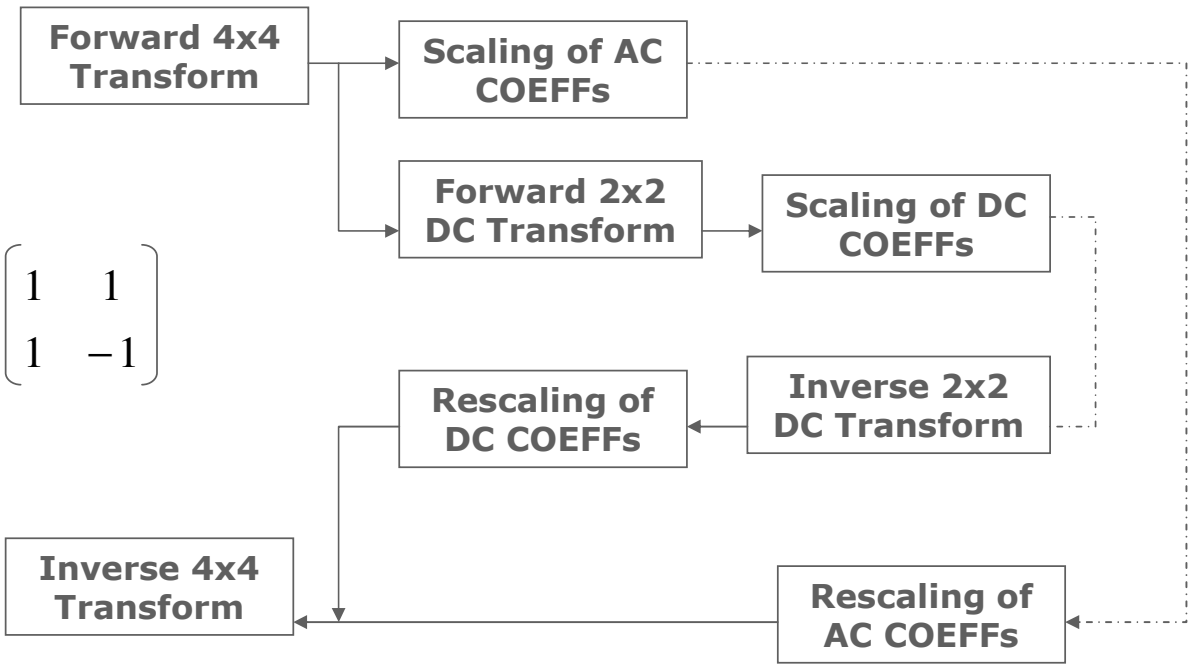
Ittiam

IntraLuma16x16 Coding



Ittiam

Chroma Coding



Hadamard Transform Matrix

$$\begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}$$

Ittiam

Quantization & Inverse Quantization

- Only 6(*3) entries in quant and dequant tables
  - Corresponding to QP%6 and 3 norm values
- Quantization:
  - (F(i,j)*Q(QP%6,i,j) + f)>>(2^(15 + (QP/6)))
- Dequantization:
  - (QF(i,j)*R(QP%6,i,j)) >> (6 – (QP/6))

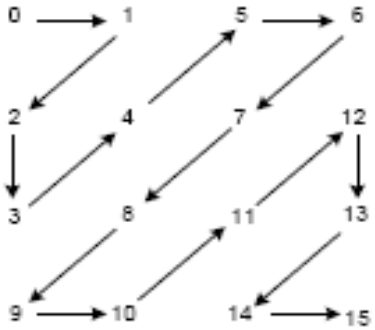
$$Q = \begin{pmatrix} 13107 & 5243 & 8066 \\ 11916 & 4660 & 7490 \\ 10082 & 4194 & 6554 \\ 9362 & 3647 & 5825 \\ 8192 & 3355 & 5243 \\ 7282 & 2893 & 4559 \end{pmatrix}$$

$$R = \begin{pmatrix} 10 & 16 & 13 \\ 11 & 18 & 14 \\ 13 & 20 & 16 \\ 14 & 23 & 18 \\ 16 & 25 & 20 \\ 18 & 29 & 23 \end{pmatrix}$$

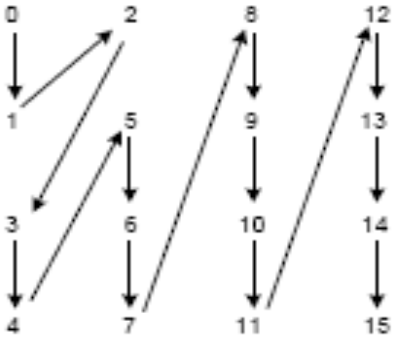
Ittiam

Coefficient Scans

- Zig-zag Scan

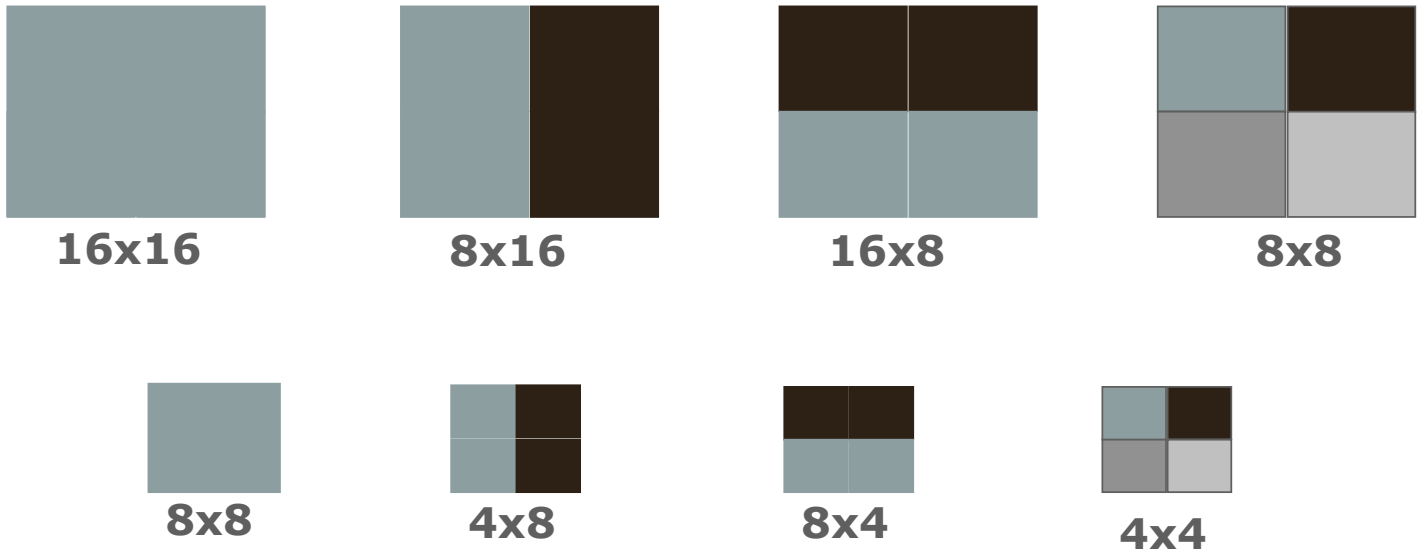


Field Scan



Ittiam

Segmented Motion Compensation



- Total of 1 - 16 motion vectors per MB
- Each 8x8 can in turn be sub-divided into one of the above partition types
- Reduces residual to be coded considerably by adapting to underlying to motion
- In conjunction with unrestricted MC, picture boundary MBs are better compensated

Ittiam

Multiple Reference MC

- Not just from the most recent past/future reference pictures
- Motion induced aliasing
- Short term buffer
  - Good to handle uncovered background/periodic motion
- Long term buffer
  - Good to handle scene changes
- Memory Management Control Operation (MMCO)
  - Insertion, deletion, promoting to long term, resetting, etc.
- Up to 15 reference frames at a given time
- Signaled at a partition level – refIdx
  - Sub 8x8 come from the same refPic
- Can re-order references for each slice
  - Most preferred reference will have the least cost for refIdx
- Error resilience
- B-pictures can be used as references
- P-pictures need not be used as references

Ittiam

Multi-hypothesis MC

- Extension to bi-directional prediction in MPEG-2/MPEG-4
  - Bi-prediction
- Two references can
  - Both be in the past
    - No extra latency
  - Both be in the future
  - One in the past and the other can be in the future
- Non-reference picture
  - Not linked to picture type
  - Bi-pred pictures can be reference pictures

Ittiam

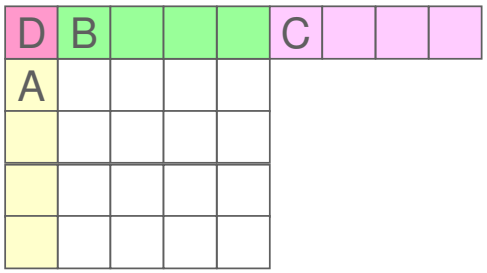
Weighted MC

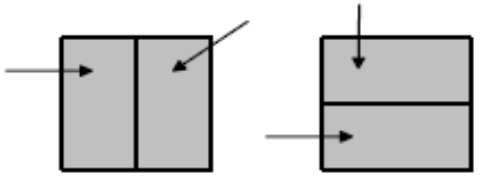
- Pixel values can be weighted during MC prediction
  - Not just 0.5, 0.5 averaging for B-pictures
  - Weights (multiplicative gain) and offsets (additive gain) for Y/Cb/Cr sent in slice header
  - Even P-pictures can have weighting
- Useful in handling fades, dissolves, aperture change in camera, lighting changes etc.
- Default, explicit, and implicit types for B-slices
  - Default – 0.5, 0.5 weighting
  - Explicit – Weights and offsets sent in bitstream
  - Implicit - Weighting based on distances to reference pictures

Ittiam

MV Prediction

- A, B, C, D – Neighboring pixels w.r.t top-left pixel of current partition
  - A – left
  - B – Top
  - C – Partition width apart to the top
  - D – Top-left
- Get MVs of the partitions to which these pixels belong
- If only one of the neighbor's refIdx matches that of current partition, use that MV alone.
- Most other cases – median(A,B,C)
  - D is used only if C is unavailable





Refldx of neighboring partition needs to be equal to the refldx of current partition

Ittiam

Fractional Pel Motion Compensation

- Motion aliasing error highest at half pixel positions
- Anti-aliasing filter to reduce aliasing effect
  - 6-tap filter for half sample interpolation
  - Not restricted to current macroblock
  - Need a 22x22 area for each MB
- Bilinear interpolation between nearest full and half sample values for quarter samples
- Chroma samples interpolated bi-linearly
  - Quarter sample accurate luma MV results in 1/8th sample accurate chroma MVs
- Chroma MV taken on a partition by partition basis
  - No MV averaging (as used in MPEG-4)

Ittiam

P-Skip

- mbType is INTER16x16
- refIdx is L0[0]
- MV is (0,0)
  - If A or B is not available (or)
  - If at least one neighbor out of A, B has a (0,0)MV
- Else, MV is set as the predicted MV
- Cbp is set to zero
  - No coded residuals

Ittiam

B-Direct & B-Skip

- Spatial Direct
  - Inherits partitions from spatially co-located L1[0] MB
  - Inherits (0,0) MV for partitions that have (0,0) MV in co-located
  - Other partitions take the spatially predicted MV for Inter16x16
  - Prediction utilization – inherited from neighbors A, B, C
  - RefIdx in L0 and/or L1 obtained as MinPositive(refIdx(A,B,C))
- Temporal Direct
  - Inherits partitions from spatially co-located L1[0] MB
  - MV of each partition w.r.t its reference is scaled by distance to B-pic
  - Always Bi-pred (between co-located partition's reference and L1[0])
- Direct_8x8 inference
  - Flagged (or) true always for interlaced
- Direct16x16 - Direct mode for the entire MB
- Direct8x8 – Direct mode for only an 8x8 block
- B-Skip – Direct16x16 and Cbp=0

Ittiam

Variable length Coding

- Single VLC for most syntax elements
  - Exponential Golomb codes
    - Easy to parse
    - Optimal for one/two-sided geometric pdfs
    - No need for tables
  - Signed values by mapping 0,1,-1,2,-2,… to above
  - Cbp coded by mapping the codes in a table according to empirical ordering of Cbp

1
010
011
00100
00101
00110
00111

- Syntax elements covered
  - mbType, sub_mbType, MVD, refIdx, cbp, dquant, etc.

Ittiam

Context-adaptive VLC for residual coding

- Uses causal neighbor's number of non-zero coefficients as context for current number of non-zero coefficients
  - Selection among multiple VLC tables based on context
  - Average symbol length is reduced
- Coeff_token (total_coeff, trailing_ones)
- Sign of trailing ones
- Coeffs coded as (prefix, suffix) from the last nzCoeff in scan order
  - Prefix is a pure prefix code (from 0 to 15)
  - Suffix is FLC
  - Two levels of escape code (that have longer levelSuffixSize)
  - SuffixLength adapted to previously decoded coeff's level (7 levels)
- TotalZeros coded first using a TotalCoeff context based table
- Runs before a nzCoeff coded using zerosLeft context based table

Ittiam

Arithmetic Coding - Basics

- Huffman codes cannot reach the entropy of the source when probability of a symbol is greater than 0.5
  - Other work-arounds exist (such as joint coding of symbols)
    - Not always easy to do this
- Arithmetic coding
  - Repeatedly partition an interval based on the symbol probability
  - From only the knowledge of the interval
    - (and the bitstream syntax and an identical probability update mechanism between the enc and dec)
    - the symbols can be decoded
  - L – low end of the interval
  - R – Range of the interval
  - Fixed point implementation – use b bits to represent L and R
  - Initially, L=0 and $R=2^{b-1}$

Ittiam

Basic Encoding Process

- A symbol occurs *(h – l)* times out of *t*
  - Allot probability range [*l/t*, *h/t*) to this symbol
  - t is the sum of frequency counts of all symbols
    - Represented in (b-2) bits
- When this symbol needs to be encoded,
  - L = L + R**l/t* Move low-end to start of new interval
  - R = R**(h – l)/t* Update interval by probability of the symbol
  - Update occurrence counts
  - Renormalize to maintain 2^{b-2} < R <= 2^{b-1}
    - While R<= 2^{b-2}
      - If L+R <= 2^{b-1}, output 0 (followed by bits_outstanding 1s)
      - If L >= 2^{b-1}, subtract 2^{b-1} – output 1 (followed by bits_outstanding 0s)
      - Else, subtract 2^{b-2} from L, increment bits_outstanding
      - Expand L and R by 2

Ittiam

Basic Decoding Process

- At the decoder,
  - Read into V, b-bits from the bitstream
  - From V and t, compute T = ((V – L + 1)*t – 1)/R
  - Divide up the current interval according to the symbol probabilities
  - See as to in which of the new intervals does T fall
    - The symbol corresponding to this interval is the decoded symbol
  - Update L and R to this interval
  - Renormalize L and R (just like the encoder); in addition
    - While R<= 2^{b-2}
      - If L >= 2^{b-1}, subtract 2^{b-1} from V
      - Else if L+R is in the mid segment, subtract 2^{b-2} from V
      - Expand L, R and V by 2
      - Shift in a new bit into V

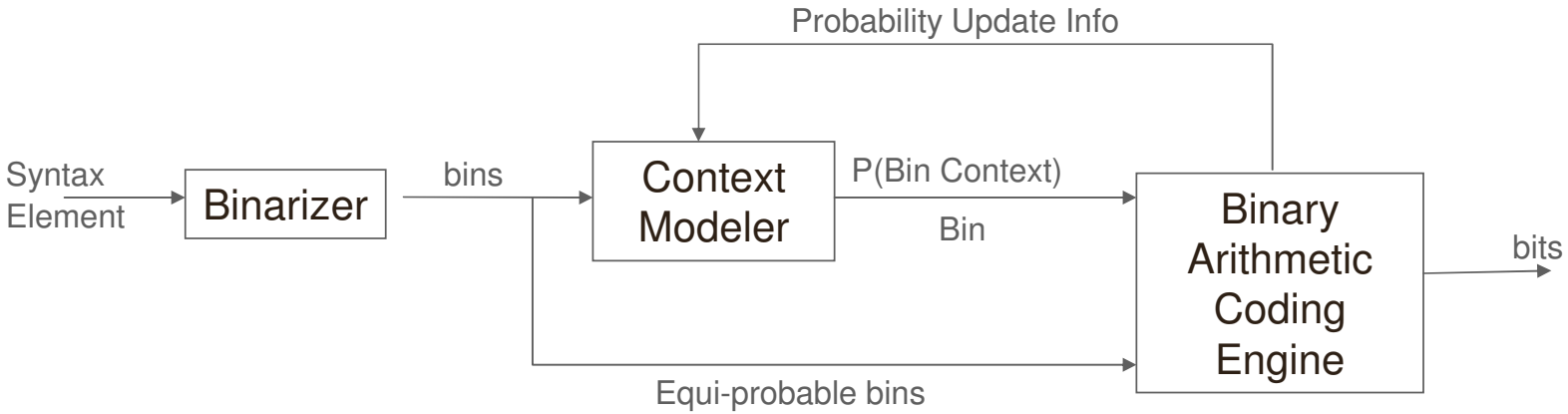
Ittiam

Binary Arithmetic Coding

- Symbols encoded are always binary
  - p_{LPS} and $p_{MPS} = (1 - p_{LPS})$
- Interval [L, L+R) is subdivided into
  - [L, L+R-R* p_{LPS}) and [L+R-R* p_{LPS},L+R)
- The search complexity to find which symbol corresponds to the interval goes down
  - If V > (R-R*p_{LPS}), then LPS; else MPS.

Ittiam

CABAC Encoding – Conceptual Blocks



- Bypass mode for equi-probable bins
  - E.g.: Sign of MVD (or) Coeffs
  - No probability model or update
  - L is updated
  - Renormalization performed as necessary

Ittiam

Binarization

- Why?
  - Reduces complexity of the arithmetic coder compared
  - Context can be modeled at a sub-symbol level
  - Higher order modeling can be employed for the higher density regions of the pdf
- How?
  - Best might have been Huffman coding; but results in higher complexity
  - 4 types of common binarization
    - Unary – String of 1s followed by a 0 — E.g.: refIdx
    - Truncated Unary - Trailing 0 is suppressed — E.g: Chroma IntraPredModes
    - EGk – Unary (up to k-bits) + exp-Golomb of rest
    - FLC
    - Combinations of above:
      - UEGk – Trunc unary up to Cutoff + EGk — E.g. MVD , absCoeff
  - A few custom binarizing schemes
    - E.g. mb_type, sub_mb_type

Ittiam

Bin Context Models

- Based on past symbols encountered, choose a context modeling function to get conditional prob. of each symbol given this function over a neighborhood
- Adapt spatially
- Types
  - Based on left and top neighbor's corresponding bins
  - Based on the previously coded bin
    - Used for coding mb_type, sub_mb_type
  - Based on position in the scan order
  - Based on number of coeffs of a particular level encountered
- 398 contexts in the standard
  - Multiple context values for a particular syntax element's bin
    - Context Index table
    - Context Index Offset – starting index for the context of an SE's bin
    - Context Index Inc – a particular value of the context for the SE's bin

Ittiam

Examples

- Mb_skip
  - Context function: !mb_skip(left) + !mb_skip(top)
    - Possible values: 0, 1, 2
    - Correspond to ctxIdx 11-13 for P-slices and 24-26 for B-slices
- MVD
  - First bin contexts are based on
    - f = |mvdx(left)| + |mvdx(top)|
    - f < 3, 3 <= f <= 32, f > 32
      - 3 contexts each for horiz. and vert.

Ittiam

Encoding Process

- Range is quantized to 4 values
  - 2 significant bits of the range
- 64 probabilities for LPS are supported from 0.5 down to 3/160
  - Next_prob ~= 0.95 * prev_prob
- $R*p_{LPS}$ can be computed using a 256 entry table
- For each ctxIdx
  - 6-bit stateIdx corresponding to probability(LPS)
  - 1-bit to represent value of LPS
- If MPS is encountered, stateIdx is incremented
- If LPS is encountered
  - a state transition empirical look-up table (64-entry) is used
- If already is at state 0, LPS and MPS are interchanged
- Based on actual bin value, update L and R. Renormalize as needed.

Ittiam

Initialization & Termination

- Initialization
  - Based on slice level QP
    - Empirically derived parameters m and n
    - stateIdx of each ctxIdx as a function of QP, m, and n
    - Set MPS or LPS for each ctxIdx
    - Allows quicker adaptation
  - Cabac_init_idc – 0/1/2
    - Another way to select better initializations at a slice level
    - Signaled in the bitstream
- Termination
  - End of slice flag assigned a non-adapting state index
  - Always produces the same pattern at the end of a slice
  - Flush happens when end_of_slice flag is true or I_PCM is encountered

Ittiam

In-loop Deblocking filter

- Aimed at reducing blocking at 4x4 sub-block boundaries
  - Introduced due to block-based transform followed by quantization
- In loop filter improves motion compensation performance at low bit-rates
- Strength of filtering is adapted based on
  - Quantization step size
  - Coding mode (intra/inter)
  - Inter-MB edge or intra-MB edge
  - Number of nonzero coeffs
  - Motion vectors and reference indices
  - Mixed Frame/Field mode edge
- Signal adaptive filtering with quantizer adaptive thresholds
  - Block boundary difference threshold (alpha)
    - To not accidentally smooth true edges
  - Differences at pixels on either side of block boundary pixels threshold (beta)
    - To not smooth texture
  - Limit corrections applied by a quantizer dependent threshold

Ittiam

Deblocking filter details

- BS = 4
  - Strongest filtering for I-MB edges
  - 4 or 5-tap filter with heavy smoothing
- BS = 1, 2, 3
  - Similar filtering
    - Uses 4 pixels
    - Clipping amount depends on the BS
- Chroma filtering alters fewer pixels
- Vertical edges are deblocked first in an MB
  - Followed by horizontal edges
- Several conditions to determine BS
- Several conditions within filtering to produce minimal signal distortion
- At low quantizers, introduce no correction
- Level of filtering can be controlled through QP offsets at PPS level

Ittiam

Interlaced coding

- Field picture coding support
  - Mix of Frame and field pictures is allowed
  - Unpaired field picture
  - Complementary field pair
- Macroblock adaptive frame field coding (MB-AFF)
  - Mix of Field pictures and MB-AFF pictures is allowed
  - MB-pair – 32x16 area that share a field/frame MB-pair coding decision
  - Each MB pair can be coded as either 2 frame MBs or 2 field MBs
    - Frame MB – Frame MC + Frame transform
    - Field MB – Field MC + Field transform
- No change in the encoding loop
  - Except for a field scan (instead of zig-zag)
- All other coding tools are extended to support PAFF and MB-AFF
  - Neighbor calculations for different predictions
  - Co-located neighbor calculation for direct mode
  - Deblocking filter
  - Context modeling for CABAC
  - DPB management
  - Reference list creation

Ittiam

Error Resilience

- NAL – Network Adaptation Layer
  - Standardized to communicate
    - Whether a particular picture is a reference picture
    - Provide a relative importance that can be used by the network layer for unequal protection or for dropping
    - A forbidden bit that should be zero normally, but can be set by network to indicate errors
- Slices
  - Independently decodable – no prediction across
  - Deblocking can be conditionally turned off across slice boundaries
  - Intra prediction can be constrained to not use inter MB data
  - Slices are not restricted to any geometry
- Slice groups – ability to code a set of macroblocks in a flexible manner
  - Can be used for concealment
- Arbitrary slice order – used to reduce delay due to out-of-order packet sequencing
- Redundant slices – primary picture and redundant pictures
  - If primary slice is not received, lower quality redundant slices can be used

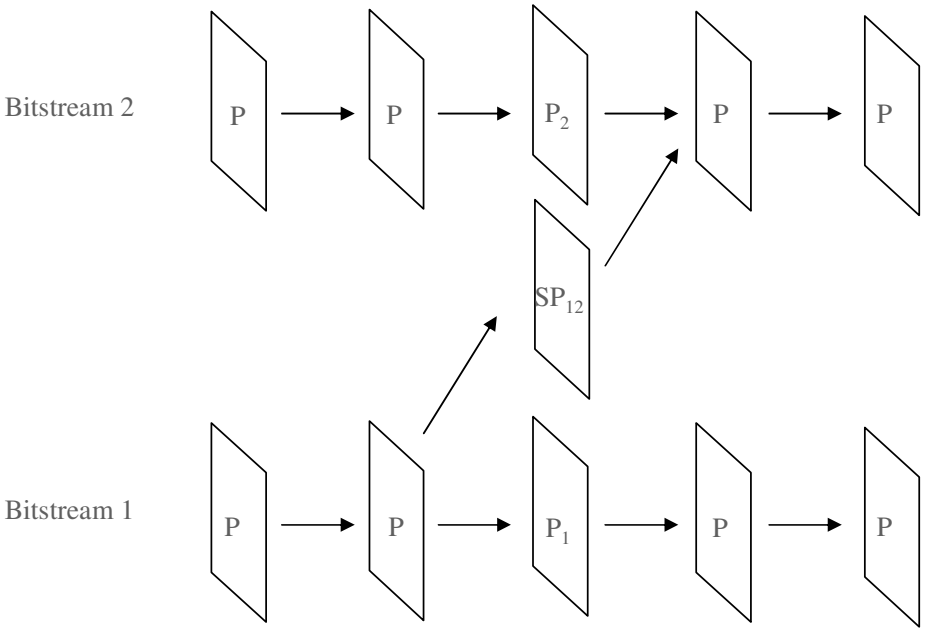
Ittiam

Data Partitioning

- Partition A
  - Slice headers
  - All MB level headers including intraPred modes, MVs and refIdx
- Partition B
  - I or SI macroblock residuals
- Partition C
  - P and B macroblock residuals
- Slice_id to find the corresponding partitions
- A is required for B and C to be decoded
- If only A is received, most predictions can be done
- Unequal protection of A, B, C

Ittiam

Switching Pictures



Ittiam

Switching Pictures

- SI and SP pictures
  - designed to suit multiple bit-rate streaming
  - Supports Fast forward or fast-rewind
  - Can be used to improve error resilience
  - SI is the intra version that matches SP
- Principle behind
  - Match reconstructions of P2 and SP12
  - Transform the reconstructed P2 block
  - Perform MC for SP picture from stream1
  - Perform Forward Transform and quantization
  - Subtract the transformed reconstruction and transmit the difference

Ittiam

Fidelity Range Extension

- Features
  - Beyond 8 bits per pixel-component
    - All dynamic ranges suitably modified
    - Quantization steps extended
  - Beyond YCbCr 4:2:0 format -> 4:2:2, 4:4:4
    - Higher number of chroma blocks
    - Chroma intra prediction modes extended
    - Chroma DC transform modified to support 2x4 and 4x4
  - Auxiliary picture support
    - Alpha layer, disparity map, etc. with luma-only coding support
  - High fidelity coding, including lossless
    - Transform bypass and no quantization loss
    - Prediction (intra/inter) and entropy coding are still used
  - Ability to use RGB color space
    - Reduce color transform errors
    - Only perform residual coding in a modified YCgCo color space
      - No rounding errors
      - No need to increase pixel precision
      - But can be done only with 4:4:4
  - Backward compatibility

Ittiam

Fidelity Range Extension

- Adaptive block sized transforms
  - 8x8 or 4x4 size – selectable on an MB by MB basis
  - 8x8 low complexity, multiplication free integer transform introduced
  - Scans for 8x8 size introduced
  - Again norms are different
    - Adjusted with quantization
      - 6x6 matrix
- Intra prediction on 8x8 blocks
  - 9 modes like in Intra4x4
  - An intermediate reference sample smoothing step is used
    - [1 2 1] or [3 1] filters
- Quantization matrix support
  - For 4x4 and 8x8
  - Default or downloadable at sequence level
  - Default or downloadable at picture level
  - Separate for intra, inter, luma and chroma
- Separate quantizers for Cb and Cr

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

6	13	20	28
13	20	28	32
20	28	32	37
28	32	37	42

Default Intra4x4 Scaling matrix

6	10	13	16	18	23	25	27
10	11	16	18	23	25	27	29
13	16	18	23	25	27	29	31
16	18	23	25	27	29	31	33
18	23	25	27	29	31	33	36
23	25	27	29	31	33	36	38
25	27	29	31	33	36	38	40
27	29	31	33	36	38	40	42

Default Intra8x8 Scaling matrix

10	14	20	24
14	20	24	27
20	24	27	30
24	27	30	34

Default Inter4x4 Scaling matrix

9	13	15	17	19	21	22	24
13	13	17	19	21	22	24	25
15	17	19	21	22	24	25	27
17	19	21	22	24	25	27	28
19	21	22	24	25	27	28	30
21	22	24	25	27	28	30	32
22	24	25	27	28	30	32	33
24	25	27	28	30	32	33	35

Default Inter8x8 Scaling matrix

Ittiam

Coding Tools in Profiles

Tools\Profiles	Baseline	Main	Extended	High	High 10	High 4:2:2	High 4:4:4
Intra Prediction	x	x	x	x	x	x	x
4x4 Transform	x	x	x	x	x	x	x
Segmented MC	x	x	x	x	x	x	x
Multiple reference MC	x	x	x	x	x	x	x
Unrestricted MC	x	x	x	x	x	x	x
Luma Quarter pel Chroma 1/8th pel MC	x	x	x	x	x	x	x
CAVLC	x	x	x	x	x	x	x
In-loop Deblocking	x	x	x	x	x	x	x
Slice Groups (FMO/ASO)	x		x				
Redundant Slices	x		x				
Multi-hypothesis MC		x	x	x	x	x	x
CABAC		x		x	x	x	x
MB-AFF		x	x	x	x	x	x
PAFF		x	x	x	x	x	x
Weighted MC		x	x	x	x	x	x
Data partitioning			x				
Spare pictures			x				

Ittiam

Coding Tools in FREXT Profiles

Tools\Profiles	High	High 10	High 4:2:2	High 4:4:4
8x8/4x4 adaptive transform	x	x	x	x
Intra 8x8 Prediction in luma	x	x	x	x
Quantization matrix	x	x	x	x
Cb and Cr QP – Independent	x	x	x	x
Auxiliary picture (monochrome)	x	x	x	x
Up to 10-bit samples		x	x	x
4:2:2 support			x	x
4:4:4 support				x
Up to 12-bit samples				x
Residual color transform				x
Lossless coding				x

Ittiam

Profiles vs. Applications

- Baseline
  - IP videophone
  - Mobile video phones
  - Simple streaming
  - DMB, DVB-H
- Main
  - Broadcast
  - VOD
  - PVR
- Extended
  - Streaming
- High-8
  - BD-ROM
  - HD-DVD
  - Broadcast
- High-10
  - Prosumer
- High422
  - Editing
  - Studio
- High444
  - Digital Cinema
  - Archival

Ittiam

Levels

- To limit the resources needed to decode according to the desired maximum decoding resolution
  - Resources
    - Bits
      - Bit-rate
    - Memory
      - Decoded Picture buffer, Coded picture buffer
    - Processing
      - Maximum number of MBs per second
      - Maximum frame rate per second
      - MBs in picture to number of slices in picture ratio limitation
      - Maximum number of motion vectors per MB (or pair of MBs)
      - Minimum compression ratio
      - Direct 8x8 inference
      - Minimum block size for bi-pred
- QCIF to beyond HD (levels 1 to 5, sub-levels as well)

Ittiam

SEI

- Supplementary Enhancement Information
  - To convey information to improve display, buffering, provide additional information synchronized with the pictures
    - Initial buffering period
    - Picture timing message
      - CPB removal delay
      - DPB output delay
      - Picture structure
        - Frame or field
        - Field or frame repeat with which field first
    - Recovery point
      - To support random access
      - Indicate independence from prior decoded pictures
      - Broken link to indicate splicing
    - Film grain synthesis
      - Different model to synthesize and add back film grain at the decoder
    - Stereo video information
      - Which field/frame is intended for which eye
    - Registered or unregistered user data
    - Freezing display, releasing frozen display
    - Subsequence information (layered coding) – average frame-rate/bit-rate, etc.

Video Usage Information

- Used to specify
  - video format properties
    - Type of video (NTSC/PAL, etc.)
    - Chroma sample position
    - Timing info and time ticks information (or) fixed frame-rate operation
    - Aspect ratio information
    - Restrictions on bitstream
      - Max mvx, mvy
      - Unrestricted MC
      - Max number of re-ordered frames
      - Max DPB
  - HRD parameters
    - Bit-rate
    - Buffer size (CPB)
    - CBR operation

Ittiam

Joint Model

- Reference implementation standardized in WG11
  - Decoder implements almost all features
  - Encoder
    - Exercises most of the important coding tools
    - Provides an elaborate list of control parameters
      - to enable or disable each tool at a fine granularity
    - Offers a rate-distortion optimized implementation
      - Exhaustively considers all possible modes
      - Codes with each mode to get output distortion and number of bits at a given quantizer
      - Picks the best mode using rate constrained minimization of distortion
    - Offers several fast computation options
      - Motion estimation
      - Intra mode selection
    - Provides a CBR rate control implementation
    - Serves as a reference for what is the best quality possible using H.264
    - Good description of the reference algorithms exists (Doc:)
- Currently at v10.1
- Can be downloaded from http://bs.hhi.de/~suehring/tml/download/

Ittiam

Decoder Implementation Aspects

- 8x8 and 4x4 intra prediction modes require reconstruction of current block before moving to the next block
  - Parallel processing on multiple intra blocks is not possible
- Shift and Add only based transform and quantization will help on low-end processors, but not on high end DSPs that already have multiple single cycle MACs
- Multiple reference pictures, segmented MC, and sub-pixel MC using pixels outside the block result in a very high memory bandwidth requirement and short burst sizes
  - 2 directions x 16 partitions/MB x (9x9 area per partition) -> 10 frames
- Deblocking costs a good percentage of total decoder MCPS
  - Boundary strength calculation for B-slices is quite expensive
  - Across luma, chroma, vertical, horizontal and BS, 8 different deblocking filters need to be implemented
  - In Baseline profile, deblocking has to be done at the end of the picture
- CABAC processing is quite serial in nature and it is very hard to use much instruction level parallelism also
  - Cycles increase linearly with increasing bit-rate
  - Worst case compression ratio and number of bins per picture can be very high
- Neighbor availability checks and neighbor calculations for the various predictions add a lot of cycles
  - MB-AFF results in a lot of conditions to be checked to get spatial and co-located neighbors
- The code size, overall, will be quite high with all the conditional processing and requires careful tuning on cache based processors

Ittiam

Carriage of AVC content over RTP

- RFC 3984
  - Defines
    - Marker bit setting guidelines
    - Fragmentation guidelines
      - Start bit, End bit, and fragmentation type
    - Aggregation guidelines
      - Single Time and Multiple Time Aggregation packets to reduce packet level overheads
      - NAL unit size (16-bits) is used to parse out the individual NAL units
    - SEI Pic timing vs. RTP timestamp usage
    - Interleaving facility through Decoding Order Number (DON)
    - Leverages the 8-bit NAL unit header byte
      - NAL_ref_idc is used to set priority (0/1/2/3)
      - Some reserved values are used to signal fragmentation and aggregation
    - Defines MIME settings and SDP configuration parameters
  - FEC as per RFC 2733 can be performed using the fragmented units

Ittiam

Carriage of AVC over MPEG-2 TS or MPEG-4 Systems

- Stream type support has been added in MPEG-2
- ObjectType support added in MPEG-4
- Elementary stream can be encapsulated using PES header
  - Used mostly in Broadcast encoders
  - Annex B byte stream format is used to facilitate NAL parsing
    - Specifies start codes for NALs
- MPEG-4 Sync Layer can also be used as the first level of encapsulation
  - Used in DMB

Ittiam

AVC File Format

- ISO/IEC 14496-15
  - Extends 14496-12 (ISO base media file format) and 14496-14 (MP4)
    - Based on the QuickTime file format
    - MOVIE atom and MOVIE data atom
      - Data is stored as is in MOVIE data atom
      - All metadata to manipulate the data is stored in MOVIE atom
  - avc1 FOURCC
  - Parameter set elementary stream
  - Video elementary stream
  - SI/SP act as shadow sync
  - Alternate tracks are used to keep multiple bit-rate streams
  - Subsequence description (from SEI) added
  - Hint tracks – altered slightly to denote only non-reference pictures as discardable (which used to be B-pictures in MP4)

Ittiam

Scalable Video Coding Extension

- Ongoing work in JVT
  - Currently on Working Draft 3
    - WD4 will be published soon
  - JSVM – Joint Scalable Video Model
    - Software reference implementation
  - Types of scalability targeted
    - Temporal
      - E.g.: 10fps to 30fps
        - Non-reference pictures offer temporal scalability
        - Hierarchical B-pictures – use of reference B-pictures
        - Motion compensated temporal filtering (MCTF) based
    - Spatial
      - Scale from lower to higher spatial resolution
      - Can use lower spatial resolution layer picture for prediction
        - Dyadic or non-dyadic
    - SNR
      - Coarse grain – successive quantization and encoding of residuals
      - Fine grain – bit-plane arithmetic coding
    - Combinations of all the three

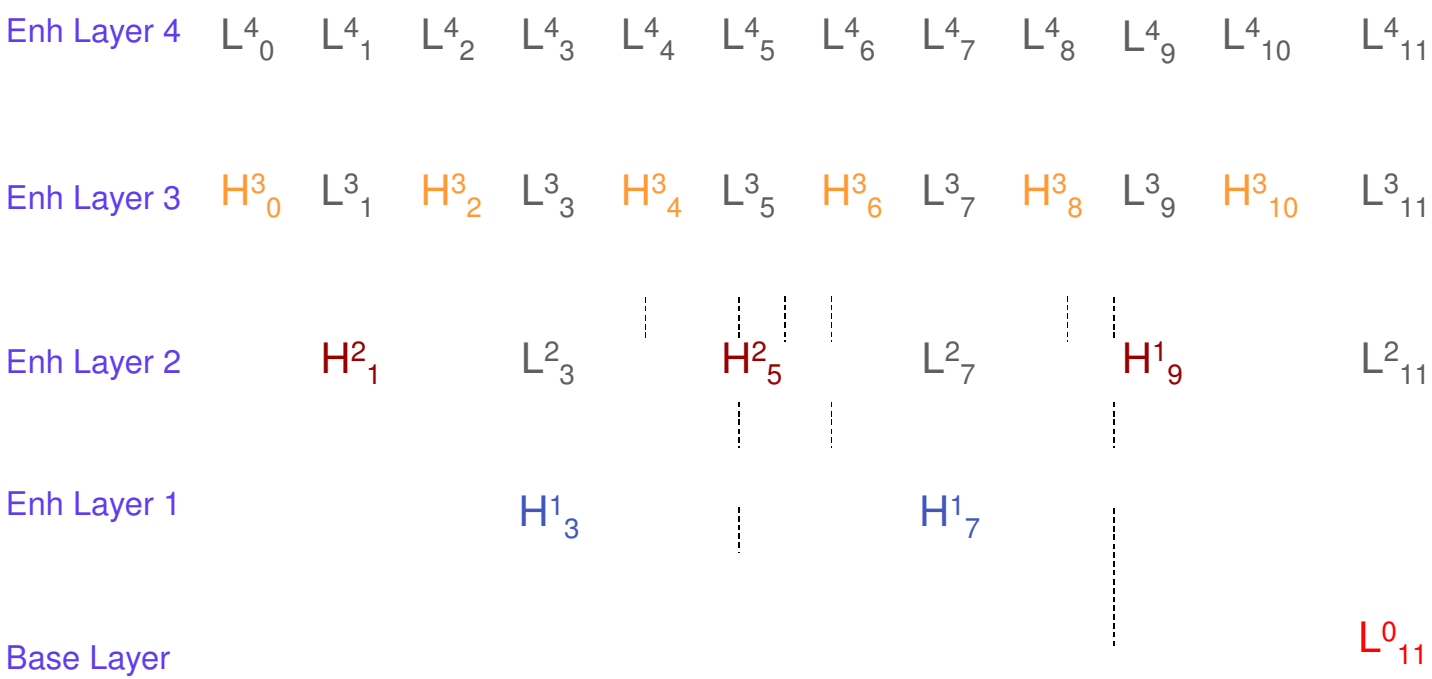
Ittiam

MCTF

- Lifting based implementation
  - Fast Wavelet Decomposition
  - Prediction and update steps
  - Perfect reconstruction on synthesis is guaranteed
    - Even with a non-linear prediction/update step
  - Use motion compensation as the operation
  - Prediction uses
    - Uni-pred MC to align reference to current
      - Equivalent to a Haar filter
    - bi-pred MC to align reference to current
      - equivalent to a 3/5 bi-orthogonal filter)
  - Update step aligns the residual error to the reference co-ordinate
    - Through backward mapping of motion vectors
    - Find all MVs that point into a 4x4 block in current from the references
  - Low-pass version
    - Reference altered with smoothing from current on the motion trajectory
  - High-pass version
    - prediction residuals

Ittiam

Iterative MCTF based Temporal Scalability



- L^0_{11} coded as a H.264 I or P picture
- H pictures are coded as H.264 B pictures
- No additional information is sent for the update step
- Normalization in decomposition absorbed in the quantization step

Ittiam

Spatial Scalability

- Base Layer is H.264 conformant
- Pyramidal representation
  - QCIF -> CIF -> 4CIF
- Each spatial scalable layer can have its own MCTF decomposition
- From a lower spatial resolution layer
  - Re-use MB partitioning and MVs with suitable upscaling
  - Up-sample residuals and use for prediction
  - Up-sample Intra MBs and use for prediction

Ittiam

SNR Scalability

- Coarse SNR Scalability
  - Residual error is re-quantized and transmitted
  - New slice type introduced to code only residuals
  - By coding the enhancement layer at QP-6 of base layer, effect of cascaded quantization is minimal
- FGS
  - Apply transform once
  - Reduce quantizer by one step size progressively and code
  - Uses CAVLC or CABAC
    - First code all coeffs that have been zero so far from base layer to the last enhancement layer encountered
    - Then code coeffs that have been non-zero earlier

Ittiam

Pointers & References

- Good overview of most coding tools
  - IEEE Transactions on Circuits and Systems for Video Technology, Special Issue on H.264, July 2003
  - Gary J. Sullivan, et. al, "The H.264/AVC Advanced Video Coding Standard: Overview and Introduction to the Fidelity Range Extensions", SPIE Conference on Applications of Digital Image Processing XXVII, August 2004.
  - White papers from http://www.vcodex.com/h264.html
- H.264 Standard (Pre-published version)
  - http://www.itu.int/rec/recommendation.asp?type=folders&lang=e&parent=T-REC-H.264
- JVT Contribution Documents – from the beginning of JVT
  - http://ftp3.itu.ch/av-arch/jvt-site/
  - Contains Working drafts of SVC
- Email reflector
  - http://mailman.rwth-aachen.de/mailman/listinfo/jvt-experts
- Reference software
  - http://iphome.hhi.de/suehring/tml/download/
- Books
  - H.264 and MPEG-4 Video Compression, Iain E G Richardson, John Wiley & Sons, September 2003, ISBN 0-470-84837-5.

Ittiam

Closing Remarks

- AVC provides significant coding gain over previous standards
- Implementation complexity is significantly high
  - Decoders – 2-3x as complex as MPEG-4 decoders
  - Encoders – Rich set of picture level and MB level modes
- Carriage of AVC and storage of AVC has been standardized
- Significant adoption level in the market
- Licensing issue is still not very clear
  - VIA and MPEG-LA do not cover all patents
- Faces competition from
  - SMPTE VC-1 (or WMV9) and AVS
- Scalable extensions are already underway

Ittiam