An Introduction to Video Compression

An Introduction to Video Compression

AGENDA

- Video Basics
  - Analog Video
  - Digital Video
  - Scanning Formats
- Video Compression
  - Intra-Frame Coding
  - Inter-Frame Coding
- Video Quality
- Video Coding Standards
- Audio Compression

Video Basics

Typical Camera

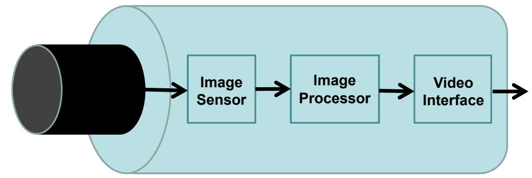


Image Sensors (CCD/CMOS)

- Each cell is a monochrome detector
- A filter array is used to make them sensitive to different color frequencies
- More green cells are used than red and blue cells because human vision is more sensitive to green frequencies
- This is called a Bayer pattern

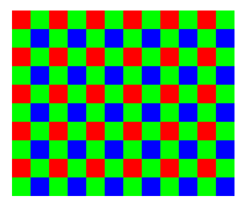


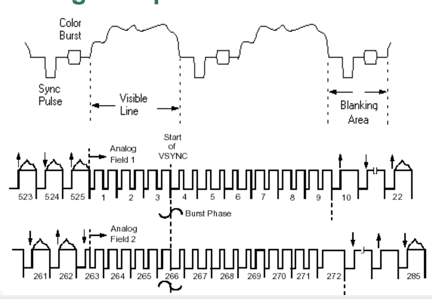
Image Processing

- Image processing internal to the camera converts the Bayer pattern to the desired camera output format
- Demosaicing and Interpolating algorithms are used to generate the output pixels
- Typical outputs include RGB and YCrCb

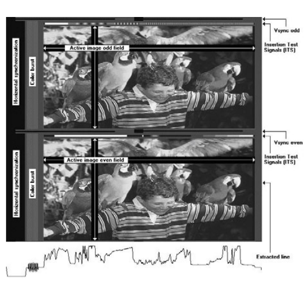
Video Interface

- Converts the camera's pixel array data to an output signal
- Horizontal/Vertical scanning of image array
- Parallel to Serial conversion of pixel data
- Synchronization, timing and status insertion
- Electrical signal generation

Analog Composite Video Waveform



Analog Video Image



Analog Composite Video Formats

- **National Television Standard Committee (NTSC)**
  - Interlaced: 2 fields/frame
  - 525 lines/frame, 262.5 lines/field
  - 29.97 frames/second, 59.94 fields/second
  - 63.5 uSeconds/line
  - Color burst: 3.58 MHz
- **Phase Alternate Line (PAL)**
  - Interlaced: 2 fields/frame
  - 625 lines/frame, 312.5 lines/field
  - 25 frames/second, 50 fields/second
  - 64 uSeconds/line
  - Color burst: 4.43 MHz
- **Séquentiel couleur à mémoire (SECAM)**

Digitized Composite Video Formats

- **NTSC**
  - 720 horizontal x 480 vertical pixels
  - 2:1 Interlaced
  - 29.97 Frames per Second (59.94 Fields per Second)
  - 165.888 M Bits per Second
- **PAL**
  - 720 horizontal x 576 vertical pixels
  - 2:1 Interlaced
  - 25 Frames per Second (50 Fields per Second)
  - 165.888 M Bits per Second

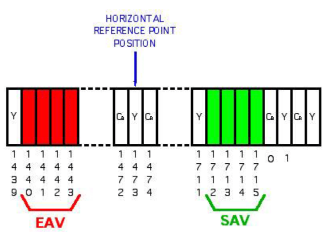
Other Analog Formats

- **S-Video**
  - Y – Luminance Signal
  - C – Chrominance Signal
  - Eliminates need for color separation filtering
  - Improved quality over Composite Video
- **YPrPb Component**
  - Y – Luminance Signal
  - Pr – Red-Green Color Difference Signal
  - PB – Blue-Green Color Difference Signal
  - Supports High Definition Resolutions
- **RGB Component**
  - Red, Green, and Blue Signals
  - Separate Sync signal
  - Sync on Green
  - Primarily a computer monitor format

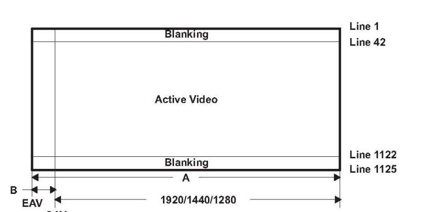
Digital Broadcast Interfaces

- Serial Digital Interface
  - SMPTE 259 SDI (270Mbps)
    - 480i30 (720x480)
    - 576i25 (720x576)
  - SMPTE 292 HD-SDI (1.485Gbps, 1.485/1.001Gbps)
    - 720p60 (1080x720)
    - 1080i30 (1920x1080)
    - 1080p30 (1920x1080)
  - SMPTE 424 3G-SDI (2.970Gbps, 2.970/1.001Gbps)
    - 1080p60 (1920x1080)

Serial Digital Interface Line



Serial Digital Interface Frame



SMPTE274M Progressive Vertical and Horizontal Components

New Digital Broadcast Interfaces

- New Serial Digital Interfaces
  - SMPTE 2081 6G-SDI (5.94Gbps)
    - 4Kp30 (3840x2160)
  - SMPTE 2082 12G-SDI (11.88Gbps)
    - 4Kp60 (3840x2160)
  - SMPTE 2083 24G-SDI (23.76Gbps)
    - 8Kp30 (7680x4320)

Broadcast Interface

- Fixed resolutions and frame rates
- Unidirectional interface
- No control channel
- Fire and forget
- Format indicated within the signal
- Coax cable
- No trigger capability
- No camera power

Machine Vision Interface

- Random resolutions and frame rates
- Bidirectional interface
- Control channel
- Often requires handshake
- Format configured in control channel
- Most use complex cables
- Triggers
- Camera power

Machine Vision Interfaces

- GigE Vision
  - Uncompressed video over UDP on 1-Gigabit and 10-Gigabit Ethernet networks
  - Cat5 cabling, Ethernet infrastructure
  - 125MB/s (1250MB/s) transfer rate
- USB3 Vision
  - Same API as GigE Vision over USB 3.0 connections

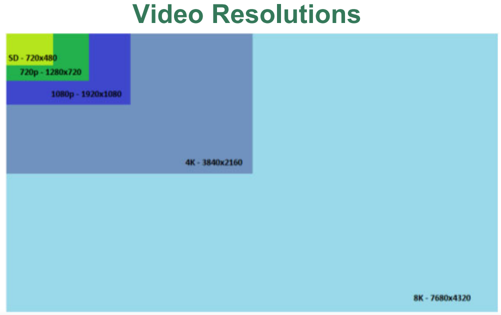
Machine Vision Interfaces (Cont)

- CoaXPress
  - 6.25 Gbps over coax cable
  - Group multiple cables (4 cables = 25 Gbps)
  - Future expansion to 10 and 12.5 Gbps
- CameraLink
  - 26-pin Mini-D ribbon connector
  - LVDS signal pairs, 2.04 Gbps
  - Group multiple cables (2 cables = 4.08 Gbps)
  - Video data, discrete signals, control channel

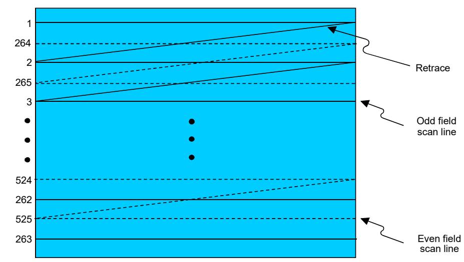
Machine Vision Interfaces (Cont)

- IEEE-1394 FireWire
  - Tree bus topology, bus master negotiations and self identification
  - 400 and 800 Mbps (1600 and 3200 Mbps defined but not widely supported).
  - 4, 6 and 9 conductor connectors
- DVI/HDMI
  - 19 pin connector
  - 48 Gbps data rate
  - I2C Control channel
  - Copy Protection negotiation

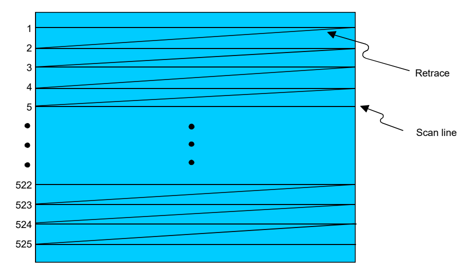
Video Resolutions



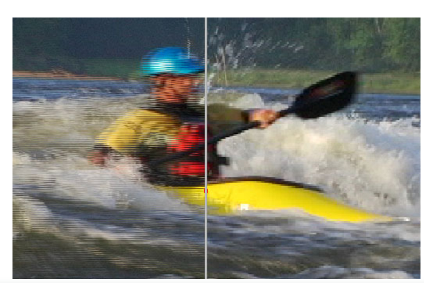
Interlaced Scanning Format



Progressive Scanning Format



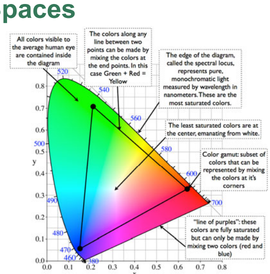
Interlace vs Progressive



Interlace vs Progressive

Color Spaces

- Monochrome
- Red, Green, Blue (RGB)
  - Separate sync, Sync on Green
- YIQ (NTSC, PAL)
- YUV
- YPbPr (Analog), YCbCr (Digital)
  - Pb/Cb = Blue - Luma
  - Pb/Cr = Red - Luma
- HSV, HSB, HSL
- CMYK



Anatomy of a CIE Chromaticity Diagram

Other Characteristics

- Pixel Bit Depth
  - 8, 10, 12, 14, 16 bits per component
- Chroma Sampling
  - 4:4:4, 4:2:2, 4:2:0, 4:0:0

Video Compression

The need for Video Compression

- **Digital Transmission**.
  - Need to match video data rate to digital communications system bandwidth.
- **Digital Storage**.
  - Need to match video data rate to digital storage system bandwidth.
  - Need to reduce storage capacity or increase storage time.
- **Multiplexing**
  - Send more programs or other data over the same channel.

Video Data Rates

- Uncompressed SD Video
  - 720x480, 30 fps, 16 bpp
    - 166 Mbps
- Uncompressed HD Video
  - 1920x1080, 60 fps, 16 bpp
    - 1,990 Mbps
- Uncompressed UHD Video
  - 7680x4320, 120 fps, 16 bpp
    - 63,701 Mbps

Digital Transmission Rates

- Transmission System Data Rates
  - Ethernet: 10/100 Mbps, 1/10 Gbps
  - OC12: 622 Mbps
  - OC3: 155 Mbps
  - DS3: 45 Mbps
  - T1: 1.544 Mbps
  - DSO: 64 Kbps
  - Modem: 33.6 Kbps
  - Cellular: 9600

Digital Storage Capacity

- Uncompressed SD Video: 166 Mbps
  - 1 Min. Clip = 1.24 G Bytes
  - 30 Min. TV Show = 37.3 G Bytes
  - 2 Hour Movie = 149.3 G Bytes
- Uncompressed HD Video: 1,990 Mbps
  - 1 Min. Clip = 14.9 G Bytes
  - 30 Min. TV Show = 447.8 G Bytes
  - 2 Hour Movie = 1.79 T Bytes
- Uncompressed UHD Video: 63,701 Mbps
  - 1 Min. Clip = 477.8 G Bytes
  - 30 Min. TV Show = 14.3 T Bytes
  - 2 Hour Movie = 57.3 T Bytes

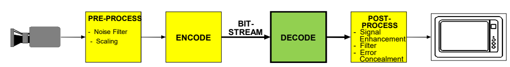
Digital Storage Devices

- CD
  - 185 to 870 M Bytes (various formats)
- DVD
  - 1.46 to 17.08 G Bytes (SS, SD, SL, DL)
- BluRay Disk
  - 25 G Bytes (50 GB for dual layer disks)
- Disk Drives
  - > 1 T Byte

Types of Compression

- **Lossless**
  - Output image is numerically identical to the original image on a pixel-by-pixel basis.
  - Only statistical redundancy is reduced
  - Compression ratio is usually low – 2:1 to 4:1
  - Reversible (infinite compress/decompress cycles)
- **Lossy**
  - Output image is numerically degraded relative to the original.
  - Statistical and perceptual redundancy is reduced
  - High compression due to reduction of perceptual redundancy
  - Can be visually lossless
  - Irreversible (compress/decompress degrades images)

Video Compression Standards



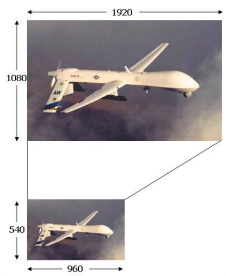
- **Functions Standardized**
  - Bit stream syntax
  - How the decoder interprets the bit stream
- **Functions Not Standardized**
  - Pre-processing
    - Filtering, scaling, noise reduction
  - Encoding strategy
    - Mode Selection
    - Quantizer Selection
    - Block Pattern Selection
    - Motion Estimation
  - Post-filter
    - Scaling, block filtering, error concealment

Preprocessing (1/5)

- Noise Reduction
  - Filter out high frequency information and improves compression efficiency
  - Spatial Noise Reduction
    - Reduces high frequency information within a picture
  - Temporal Noise Reduction
    - Reduces high frequency information between frames

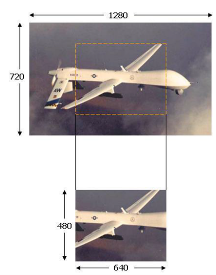
Preprocessing (2/5)

- Resolution Reduction
  - Reduces the amount of spatial information that needs to be compressed
  - Scaling
    - Maintains field of view



Preprocessing (3/5)

- Resolution Reduction
  - Reduces the amount of spatial information that needs to be compressed
  - Cropping
    - Maintains pixel resolution



Preprocessing (4/5)

- Frame Rate Reduction
  - Reduces the amount of temporal information that needs to be compressed

1	2	3	4	5	6

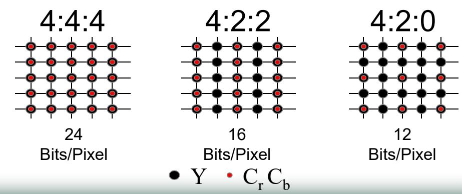
Video Sequence – 30 Frames per Second

1	X	3	X	5	X

Video Sequence – 15 Frames per Second

Preprocessing (5/5)

- Color representation
  - Reduces the amount of spatial information that needs to be compressed



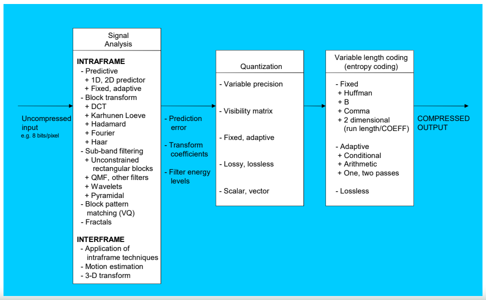
Intra-Frame Coding

- Removes the redundancies within a frame
- Each frame is encoded separately without regard to adjacent frames
- Similar to JPEG

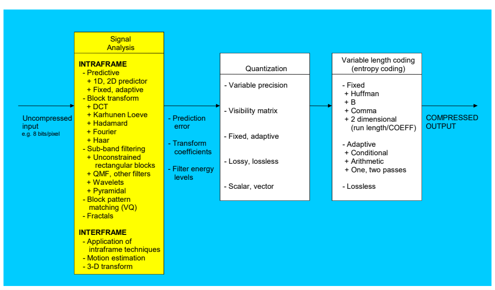
Encoding Stages

- Signal Analysis
  - Transform from spatial domain to another domain that is easier to compress
- Quantization
  - Discard less important information
- Variable Length Coding
  - Last stage of efficient coding

Video Compression System



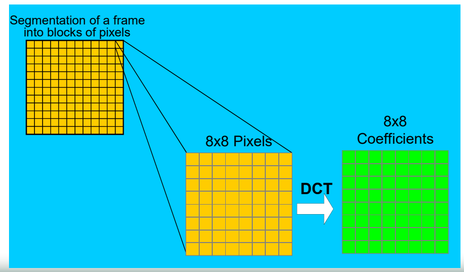
Signal Analysis Phase



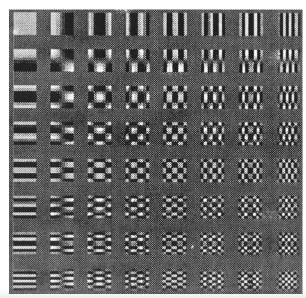
Discrete Cosine Transform

- Converts spatial information to frequency information
- Important information is concentrated in lower frequency bands
- Operations typically performed on 8x8 blocks of pixels
- Lossless and reversible

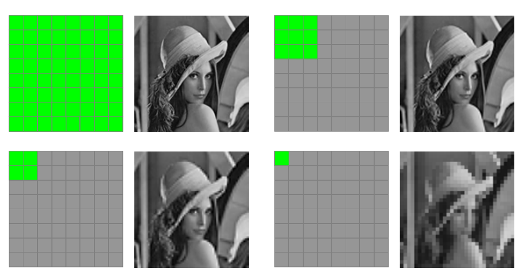
Transform Coding by Discrete Cosine Transform (DCT)



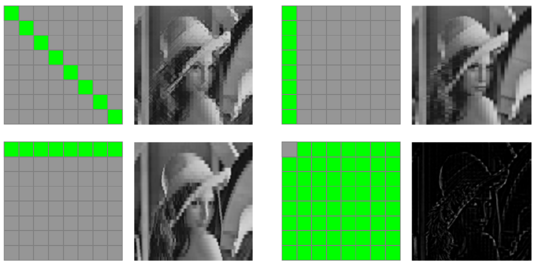
DCT Basis Function



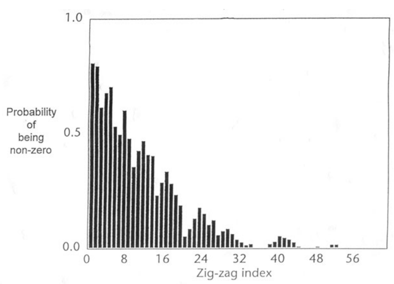
Effect of DCT Coefficients



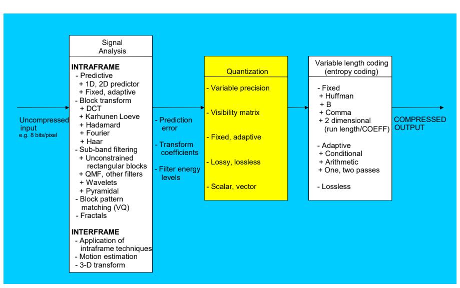
More Effects



Coefficient Levels



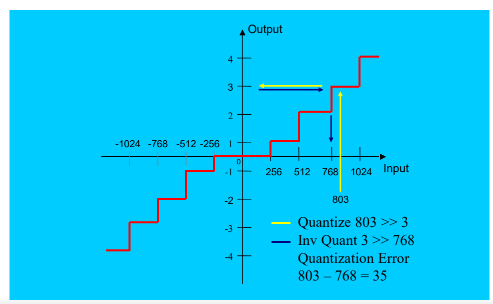
Quantization Phase



Quantization

- Converts 12 bit DCT coefficients to 8 bit values
- Try to force less important information toward zero values
- Takes less bits to code zero
- Most of loss occurs here

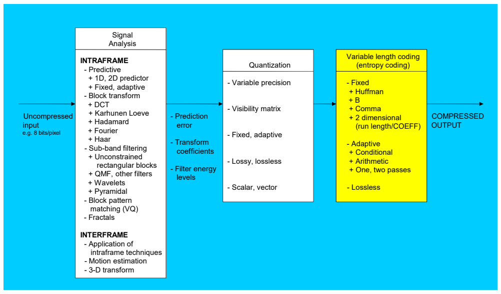
Scalar Quantization



Scanning Order in a Block

1	2	6	7	15	16	28	29
3	5	8	14	17	27	30	43
4	9	13	18	26	31	42	44
10	12	19	25	32	41	45	54
11	20	24	33	40	46	53	55
21	23	34	39	47	52	56	61
22	35	38	48	51	57	60	62
36	37	49	50	58	59	63	64

Entropy Coding Phase



Variable Length Coding

- A technique for assigning shorter codes to high probability symbols, and longer codes to lower probability symbols
- Entropy Coding, Huffman Encoding, Arithmetic Coding
- Modest compression gain (2:1-4:1)
- Lossless
- Reversible

VLC Examples

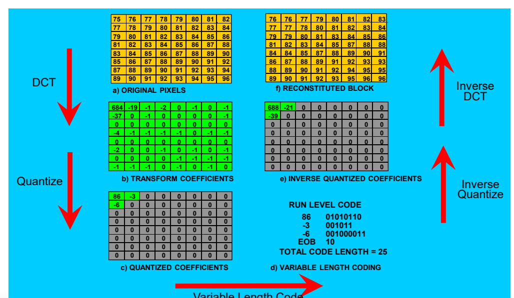
Huffman Variable Length Coding

Symbol	Probability	VLC Code	Fixed Length Code
A	½	0	00
B	¼	10	01
C	1/8	110	10
D	1/8	111	11

For 128 symbols: Fixed Length = 256 bits

Variable Length = 224 bits (12.5% savings)

Sample Intra Block Coding



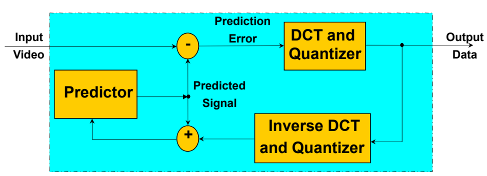
Inter-Frame Coding

- Removes the redundancies between frames
- Uses one or more previous or future frames as a prediction of the current frame
- Produces a difference between the current frame and the prediction
- Only encode the differences

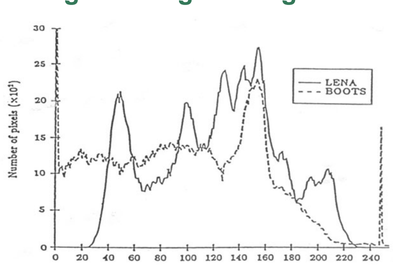
Encoding Stages

- Signal Analysis
  - Transform from spatial domain to another domain that is easier to compress
- Quantization
  - Discard less important information
- Variable Length Coding
  - Last stage of efficient coding
- Prediction
  - Create a prediction for the frame being encoded and only compress the difference

Generic Inter-Frame Coding System



Original Image Histograms

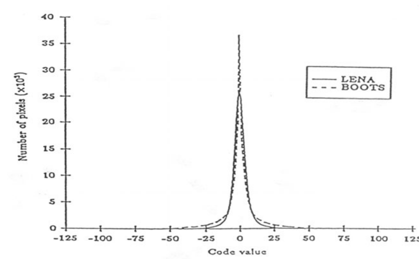


Difference Image

Original Image

Inter-Frame Difference

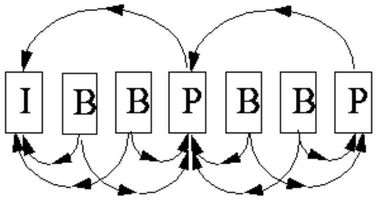
Difference Image Histograms



Pictures Types

- **Intra Frame (I)**
  - Full picture coded independent of other frames
  - Key frame
- **Predicted Frame (P)**
  - Inter coded frame predicted from a previous frame
- **Bidirectional Frame (B)**
  - Inter coded frame predicted from a future frame
- **GOP**
  - # of P/B frames between I frames
  - # of B frames between I/P frames
  - Open or Closed GOP

Group of Pictures (GOP)



Transmission Order: IPBBPBB

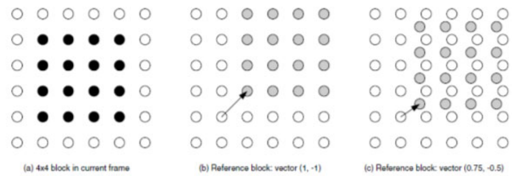
Picture Type Efficiency

- I Frames: 1 bit per pixel
- P Frames: 0.25 bits per pixel
- B Frames: 0.03 bits per pixel
- All I Frames: 10Mbps
- IP15: 6 Mbps
- IPB15: 2.5 Mbps

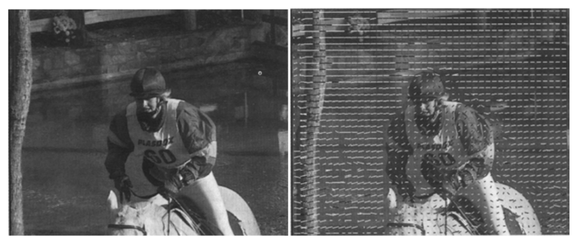
Improving Prediction

- Objects in a scene do not usually change instantaneously
  - Objects move
  - Camera pans and zooms
- Improve prediction by accounting for the motion of blocks
- Extremely computation intensive

Improved Prediction with Motion Vectors

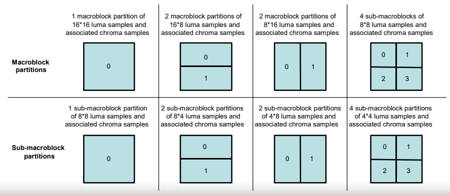


Motion Vector Example

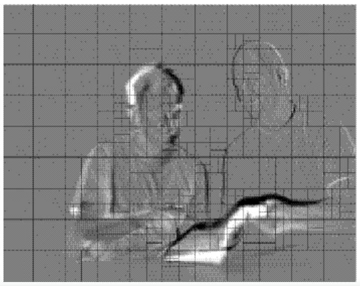


Block Sizes

- H.261/H.263 use 16x16 block for motion estimation
- H.264 uses variable block sizes

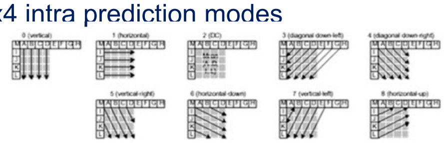


Block Size Example

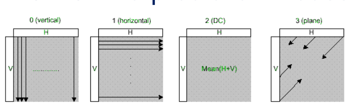


Intra-Frame Prediction

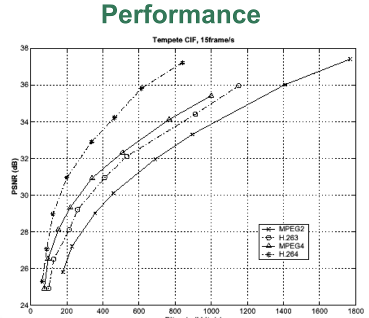
- 4x4 intra prediction modes



- Also 16x16 intra prediction modes



Performance



Video Quality

Rate Control Algorithm (1/2)

- This is the encoder magic sauce
- Determines how to code each macro-block
- Rate Control Modes
  - Constant Quality (CQ) – Attempt to product a consistent video quality regardless of the data rate required
  - Constant Bit Rate (CBR) – Attempt to product the best quality video while holding the data rate constant. May use fill data to maintain the constant data rate. May still be some minor variability in data rate.
  - Variable Bit Rate (VBR) – Attempt to produce the best quality video while not exceeding some maximum data rate, but allowing the data rate to drop when not needed

Rate Control Algorithm (2/2)

- For VBR and CBR, encoder attempts to use the finest quantizer (sharpest quality pictures)
- If the encoder is generating too much data for the selected data rate, the encoder increases the coarseness of the quantizer up to the maximum user selected value (picture quality degrades)
- If the encoder is still generating too much data for the selected data rate, the encoder begins dropping frames (reduced frame rate)

Encoder Trade-offs

- Multidimensional trade-off space
  - Data rate
  - Resolution
  - Frame Rate
  - Quality
  - Latency
  - Error resilience

Compression Ratio

- Old compression algorithms had a fixed bits per pixel and therefore had a defined Compression Ratio.
- New compression algorithms have a variable number of bits per pixel depending on a number of factors, most notable is picture quality.
  - 1000:1 compression is very possible, but picture quality will be unusable
  - 100:1 is reasonable
- Compression efficiency and picture quality are also highly dependent on scene complexity.

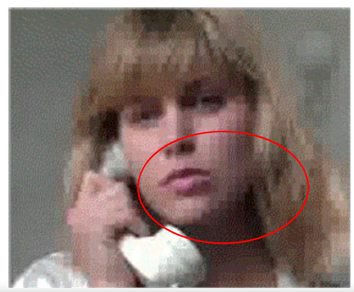
Typical Video Quality Problems

- Data rate too low to support desired quality
  - Communications link bandwidth
  - Codec efficiency
  - Resolution and frame rate
- Too many errors
  - Bit errors
  - Packet loss
  - Burst errors

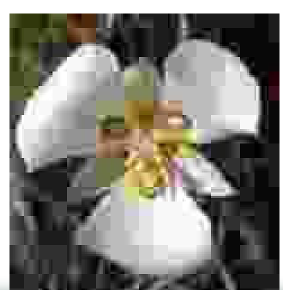
Video Quality Effects

- **Video Frame Rate**
  - Reduced Frame Rate
  - Jerky Video (irregular frame rate)
- **Video Artifacts**
  - Multi-color blocks
  - Black or white stripes
  - Tearing or melting images
  - Ghosts and shadows behind moving objects
  - Block edges
  - Edge noise around sharp edges
- **Decoder Issues**
  - Loss of sync, cannot detect start codes
  - Video freeze
  - Decoder crash

Compression Artifacts - Blockiness



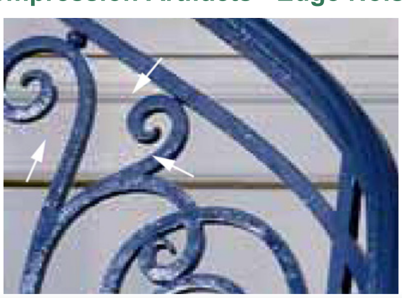
Compression Artifacts - Blockiness



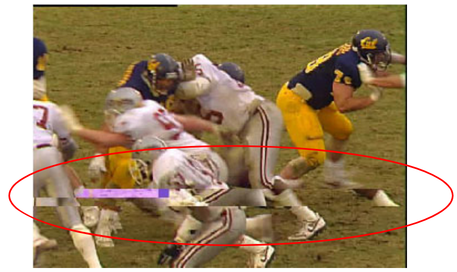
Compression Artifacts - Blurriness

Compression Artifacts - Edge Noise

Compression Artifacts - Edge Noise

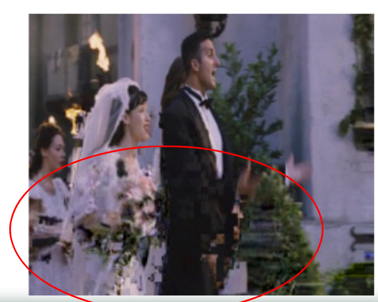


Error Artifacts - I-Frame Data



Error Artifacts – P/B-Frame Data

Error Artifacts – P/B Frame Data



Packet Loss Artifacts

1% Packet Loss Rate

5% Packet Loss Rate

Compressed Data Error Sensitivity

- Sparse Start Codes – usually at the beginning of a picture
- Variable length codes – a bit error not only changes the value, but also the length of the code and the following values
- Inter-frame coding propagates an error over the rest of the GOP

Error Resilience Features

- GOP Structure – distance between I-Frames
- Intra Refresh – Intra-coded macroblocks distributed throughout long GOPs
- Slices – partitioning image into sub images with re-sync points
- Flexible Macroblock Ordering, Arbitrary Slice Ordering, Redundant Slices
- Forward Error Correction Schemes

Error Concealment Features

- Decoder functions to reduce the effect of errors
- Intra-frame concealment – use other pixels within the frame to predict the value of the missing pixels
- Inter-frame concealment – use pixels from previous frames to predict the value of the missing pixels
- Estimation of missing motion vectors

Error Resilience Schemes

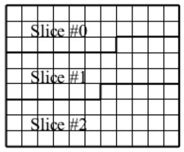
- Flexible Macroblock Ordering (FBO)
- Arbitrary Slice Ordering (ASO)
- Data Partitioning (DP)
- Redundant Slices (RS)
- SP/SI Frames for bit rate switching
- Reference Frame Selection
- Intra-block refreshing

Video Encoder Settings

- Video Resolution
  - Autodetect
  - Scaling/Cropping
- Frame Rate Decimation
- GOP Size/Structure
  - Infinite GOP=0 (only P-Frames)
  - GOP=1(only I-Frames)
  - Normal GOP=2 to N (I and P-frames)
- Quantizer Settings

Video Encoder Settings

- Slice Modes
  - Error Recovery
  - Full Frame
  - Macroblocks
  - Bytes
- Intra-Refresh
  - Force Intra Macroblocks during P-Frames
  - Sequential
  - Random



Video Encoder Setting

- Date Rate
- Rate Control Mode
  - Constant Quality
  - Constant Bit Rate
  - Variable Bit Rate
- Rate Control Parameters
  - Not common

Contact Information

747 Dresher Road
HORSHAM, PA 19044
Ph: 215-657-5270
FAX: 215-657-5273
www.deltadigitalvideo.com

Backup Slides

High Definition Formats

- **720p**
  - 1280 horizontal x 720 vertical pixels
  - Progressive
  - 60 Frames per Second
  - 884.736 M Bits per Second
- **1080i**
  - 1920 horizontal x 1080 vertical pixels
  - 2:1 Interlaced
  - 30 Frames per Second (60 Fields per Second)
  - 995.328 M Bits per Second
- **1080p**
  - 1920 horizontal x 1080 vertical pixels
  - Progressive
  - 60 Frames per Second
  - 1990.656 M Bits per Second

Interlace Artifacts

Video Compression Standards

Standards Organizations

- In the beginning there was COST 211. They created the H.120 standard for interactive video communications.
- The ITU-T VCEG (Video Coding Expert Group) was formed to improve upon H.120.
- MPEG (Moving Pictures Experts Group) was formed to find a way to incorporate codecs for broadcast work.
- To this day, the two committees – MPEG and VCEG, work side-by-side. MPEG specializes in broadcast (television), while the ITU (International Telecommunications Union) focuses on telecommunications (phone, internet).

ITU-T H.120

- Adopted in 1984
- 1544 kbit/s for NTSC and 2048 kbit/s for PAL
- Used differential pulse-code modulation, scalar quantization, variable-length coding
- Showed that a better algorithm was needed to produce usable quality at reasonable data rates

ITU-T H.261 Overview

- Adopted 1988
- Motion Video Algorithm typically used for Video Conferencing over ISDN networks
- Target Data Rate
  - 56 Kbps to 2048 Kbps
- Available Resolutions
  - CIF: 352x288
  - QCIF: 176x144

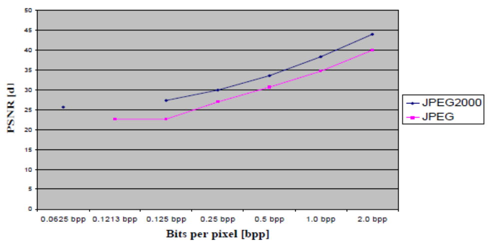
ISO/IEC JPEG Overview

- Adopted in 1993
- ITU-T T.81, ISO/IEC 10918-1
- Still Image Compression Algorithm
- Based on Discrete Cosine Transform (DCT)
- Motion video mode – sequence of still images
- Lossless and lossy compression modes

ISO/IEC JPEG-2000 Overview

- Adopted in 2000
- ITU-T T.800, ISO/IEC 15444-1
- Still Image Compression Algorithm
- Based on Discrete Wavelet Transform (DWT)
- Motion video mode – sequence of still images
- Lossless and lossy compression modes

JPEG vs JPEG2000



JPEG vs JPEG-2000
(at same bit rate)

JPEG

JPEG-2000

ISO/IEC MPEG-1

- Adopted in 1992
- ISO/IEC 11172
- Motion Video Compression primarily for CD-ROM video storage
- Target Data Rate: 1.5 Mbps
- Most common resolution was the Source Input Format (SIF)
  - 352x240 (for NTSC)
  - 352x288 (for PAL)
  - 320x240 (for VHS)

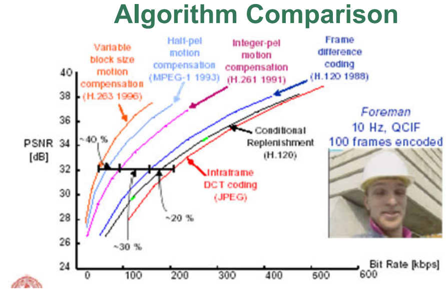
ITU-T H.262 / ISO/IEC MPEG-2

- Adopted in 1996
- ITU-T H.262, ISO/IEC 13818
- Typical Data Rates
  - 3-5 Mbps for broadcast quality standard definition video
  - 15-20 Mbps for high definition video
- Typical Resolutions
  - SIF: 360x240p
  - HHR: 360x480i
  - SD: 720x480i, 720x576i
  - ED: 720x480p
  - HD: 1280x720p, 1920x1080i
- Enabled digital broadcast, cable and satellite television

ITU-T H.263 Overview

- Adopted in 1999
- ITU-T Recommendation H.263
- Target Data Rate: 56 kbps
- Target Resolution: Sub QCIF 128x96
- Designed for video phone over analog telephone lines
- Significant improvements in compression efficiency

Algorithm Comparison



ISO/IEC MPEG-4 Overview

- Adopted in 1999
- ISO/IEC 14496-2 MPEG-4 Part 2
- Based on H.263 as baseline
- Optimized for internet streaming applications
- Same picture sizes as MPEG-2 but also smaller sub-QCIF sizes

ITU-T H.264

- Adopted in 2003
- ITU-T H.264
- Also called ISO/IEC 14496-10 – MPEG-4 Part 10, Advanced Video Coding
- H.264 is the most widely used codec on earth, even surpassing broadcast MPEG-2, simply due to the power of the internet. It is used by Youtube, and every other video provider of note.
- It is also the codec that drives Blu-ray. Most cable, satellite and television channels have moved to H.264.

MPEG-2 vs H.264

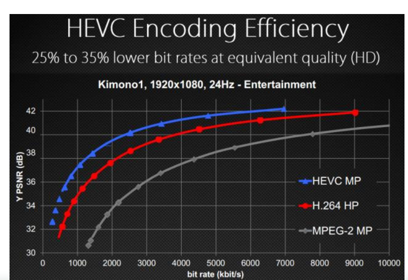
Algorithm Comparison

MPEG-2 MPEG-4 H.264

ITU-T H.265

- Initially adopted 2013 and revised several times.
- ISO/IEC 23008–2 MPEG-H Part 2, High Efficiency Video Coding (HEVC).
- Up to 8K UHDTV (8192×4320 maximum)
- Up to 12-bit color bit depth 4:4:4 and 4:2:2 chroma sub-sampling
- Supports up to 300 fps
- Data rates up to several GB/s
- Special features to support high efficiency encoding of very high resolution video

Algorithm Comparison



ITU/ISO/IEC New Work

- Formed Joint Video Exploration Team (JVET) to investigate next generation video compression standard
- Target release in 2020

Other Codecs

- Microsoft
  - WMV7, WMV8, WMV9
- Google
  - VC6, VP7, VP8, VP9
- DIRAC
- Alliance of Open Media (AOMedia)
  - AV1

Contact Information

747 Dresher Road
HORSHAM, PA 19044
Ph: 215-657-5270
FAX: 215-657-5273
www.deltadigitalvideo.com

Backup Slides

Overview of IRIG-210

The Horace Standard

IRIG 210

- Developed in late 1980's under SBIR program
- Standardized by IRIG in 1993
- Differential Pulse Code Modulation Algorithm (DPCM)
- Monochrome, RS-170 video
- Full, Fixed or variable frame rate
  - Field or frame skip modes
- Typical Resolution
  - 640, 512, 450, 320, 256, 225, 160 or 128 pixels per line
  - 240 lines per field or 480 lines per frame

Differential PCM (DPCM)

- Video is encoded line by line
  - Line Synchronization code
  - Format Information
  - Line pixel data
- PCM – 7 bit value for each pixel
- DPCM – difference between the current pixel and the previous pixel
  - Three DPCM modes: Normal, High, 2-Bit
  - Adaptive on a line by line basis
- Entropy Coding
  - Normal and High modes only
  - 1 to 8 bit codes based on previous and current DPCM code

DPCM Modes

- Normal
  - 3 bits per pixel (8 difference values)
    - 0, +3, -3, +8, -8, +20, -20, +/-40
- High Level (course)
  - 3 bits per pixel (8 difference values)
    - 0, +4, -4, +10, -10, +25, -25, +/-50
- 2-Bit Mode
  - 2 bits per pixel (4 difference values)
    - +4, -4, +26, -26

Entropy Coding

TABLE 2-2. ENTROPY CODE TABLE, KERNEL 000

	To Next L-Code							
From L-Code↓	1	2	3	4	5	6	7	8
1	1	001	01	00001	0001	0000001	00001	00000001
2	1	01	001	0001	00001	000001	0000001	00000001
3	1	0001	01	00001	001	0000001	000001	00000001
4	001	01	00001	1	000001	0001	0000001	00000001
5	001	00001	01	000001	1	0000001	0001	00000001
6	0001	001	00001	01	000001	1	0000001	00000001
7	001	00001	0001	0000001	01	000001	1	00000001
8	1	001	01	00001	0001	0000001	000001	00000001

DPCM Example 1

Input Pixel	50	45	48	43	40	44	47	51
Difference	+50	-5	+3	-5	-3	+4	+3	+4
DPCM Code	8	2	2	3	3	2	2	2
DPCM Jump	+40	+3	+3	-3	-3	+3	+3	+3
Output Pixel	40	43	46	43	40	43	46	49
Error	10	2	2	0	0	1	1	2
Output Bits	0000 0001	001	01	001	01	000 1	01	01

26 Bits for 8 Pixels = 3.25 Bits/Pixel

DPCM Example 2

Input Pixel	2	4	8	16	32	64	127	127
Difference	+2	+2	+4	+8	+16	+32	+64	+0
DPCM Code	2	2	2	4	6	6	8	6
DPCM Jump	+3	+3	+3	+8	+20	+20	+40	+20
Output Pixel	3	6	9	17	37	57	97	117
Error	1	2	1	1	5	7	30	10
Output Bits	001	01	01	0001	0001	1	0000 0001	0000 001

31 Bits for 8 Pixels = 3.875 Bits/Pixel

DPCM Example 3

Input Pixel	0	0	0	0	127	127	127	127
Difference	+0	+0	+0	+0	+127	+0	+0	+0
DPCM Code	1	1	1	1	8	8	8	4
DPCM Jump	+0	+0	+0	+0	+40	+40	+40	+8
Output Pixel	0	0	0	0	40	80	120	128
Error	0	0	0	0	87	47	7	1
Output Bits	000 1	1	1	1	0000 0001	0000 0001	0000 0001	0000 1

36 Bits for 8 Pixels = 4.5 Bits/Pixel

IRIG 210 Performance

- Best possible performance
  - Black screen
  - 1-bit per pixel
  - 512x480 @ 30 fps: 7.7Mbps
  - 256x480 @ 30 fps: 4.0 Mbps
- Typical performance
  - About 2.2-bits per pixel
  - 512x480 @ 30 fps: 16.5 Mbps
  - 256x480 @ 30 fps: 8.4 Mbps

Overview of H.261 and H.263

The original video conferencing codec

ITU-T H.261 Format

- Picture Structure
  - Picture [CIF: 2x6 GOBs, QCIF: 1x3 GOBs]
    - Group of Blocks (GOB) [11x3 macro blocks]
      - MacroBlock [4 Y, 1 Cr, 1 Cb blocks]
        - » Block [8x8 pixels]
- Forward Error Correction
  - BCH (511,493)
  - 512 bit frame with sync bit and fill indicator

ITU-T H.261 Intra-frame Coding

- Full Intra-frame coded on startup
- After that all frames Inter-frame coded
- Intra Refresh
  - Intra code individual blocks
  - X macroblocks per frame
  - To correct transmission and roundoff errors
- Full Intra-frame coded or on command
  - Usually requested by decoder

ITU-T H.261 Inter-frame Coding

- Prediction based on previous frame with motion compensation
- Motion vector applies to macroblock
- Full pixel motion vector resolution
- +/- 15 pixel motion vector range

JPEG and JPEG2000

- JPEG (Joint Photographic Experts Group) became a popular codec of choice for still image compression
- Intra Frame compression algorithms – compress each frame independently without regard to prior or later frames

ISO/IEC JPEG-2000 Features

- Superior compression performance compares to JPEG.
- Multiple resolution representation.
- Progressive transmission by quality and resolution accuracy.
- Random code-stream access and processing, also referred as Region Of Interest (ROI).
- Error resilience.
- No blocking artifacts
- Supports very large format images and tiling

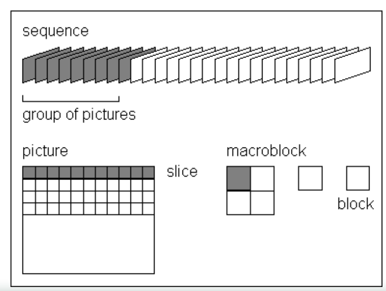
Overview of MPEG-1, 2, 4

Entertainment quality video

MPEG-1 Overview

- ISO/IEC 11172 MPEG-1 Standard (1992)
- Target bit-rate about 1.5 Mbit/s
- Typical image format SIF (352x240), no interlace
- Frame rate 24 ... 30 fps
- Main application: video storage for multimedia (e.g., on CD-ROM)

MPEG Components



Discrete Cosine Transform

- 4 x 4 Block size
  - DCT requires O(n^2) operations
  - 8x8: 4096 operations
  - 4x4: 256 operations (x4=1024)
- 2 x 2 block size for Chrominance
- 4 x 4 block size for DC coefficients of 16x16 group
- Integer implementation
  - 16 bit integer math
  - No floating point math
- No multiplication or division
  - Only add, subtract, shift
- Integrated with quantizer
  - Scaling integrated with quantizer so it is only done once.

More Motion Estimation

- Forward and Backward Reference Frames (Blocks)
- Multiple Reference Frames (Blocks)
  – Average of blocks in multiple different frames
  – Weighted factor for each reference frame
- 1/4 pixel accuracy/resolution

Improved De-blocking Filter

- In-loop filtering, not post filtering
- Improves the quality of the predicted image as well as the resultant image
- On edge level, filtering strength is dependent on the code mode, motion vector, and values of residuals
- On sample level, quantizer dependent thresholds can turn off filtering for any individual sample

Entropy Coding

- Exponential Golomb code
  – Header and other information
- Context-Adaptive Variable Length Coding (CAVLC)
  – Quantized transform coefficients
  – Run-length coding
  – Runs of zeros, runs of ones
- Context-Adaptive Binary Arithmetic Coding (CABAC)
  – Quantized transform coefficients
  – 5-15% improvement over CAVLC, but more complex

Exponential Golomb Codes

Bit string form	Range of Symbols (#)
1	0 (1)
0 1 x_0	1-2 (2)
0 0 1 x_1 x_0	3-6 (4)
0 0 0 1 x_2 x_1 x_0	7-14 (8)
0 0 0 0 1 x_3 x_2 x_1 x_0	15-30 (16)
0 0 0 0 0 1 x_4 x_3 x_2 x_1 x_0	31-62 (32)
…	…

Slices

- Groups of one or more Macro Blocks (MB)
  – Horizontal, vertical, rectangular, dispersive
  – Independently decodable
  – Limits error propagation within a picture
- I Slice
  – Only intra coded MB
- P Slice
  – Forward predicted and intra coded MB
- B Slice
  – Backward, forward, multiple reference and intra coded MBs
- SI, SP Slice
  – For bitstream switching
- Redundant Slices
  – For error resilience
- Arbitrary Slice Order

Parts and Layers

- MPEG standards have sub-divisions, called Parts.
- Traditionally, Part 1 is always for the ‘System’ (file or stream format). Part 2 is for video, and Part 3 is for audio.
- Parts are further sub-divided into Layers.
- Audio has three or more layers, called Layer I, Layer II and Layer III and so on.
- MPEG-1 Part 3 Layer III is called MP3.

Profiles and Levels

- Beginning with MPEG-2, is the concept of Profiles and Levels – in addition to Parts and Layers.
- Profiles are sets of capabilities within the specification to be used for specific applications.
- Levels specify a range of parameters used within a profile.
- The algorithms are so complex and address so many different applications that it does not make sense for an encoder or software to comply with all of it.

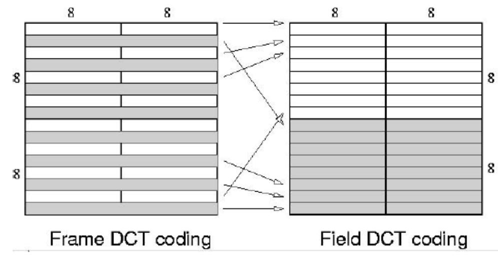
ISO/IEC MPEG-2 Overview

- Part 1, Systems defined two major classifications: Program Stream and Transport Stream
- Part 2, Video is also called H.262. It had additional support for interlacing and 4:2:2
- Part 3, Audio incorporated 5.1 channels

ISO/IEC MPEG-2 Features

- More input flexibility
  - Interlaced and progressive pictures
  - Field and frame coding
  - Wide range of picture resolutions
- Motion estimation
  - Half pixel motion vectors
  - 16x16 or 16x8 motion blocks
  - Estimation across fields (Dual prime)
- Non-linear quantization
- Improved VLC
- Various scalability tools

Frame/Field Coding



ISO/IEC MPEG-2 Profiles

- Simple
  - 4:2:0 sampling, I/P pictures only, no scalable coding
- *Main*
  - *As above, plus B pictures*
- SNR
  - As above, plus SNR scalability
- Spatial
  - As above, plus spatial scalablity
- High
  - As above, plus 4:2:2 sampling, 11 bit precision
- 422
  - As above, no scalability, higher data rate

ISO/IEC MPEG-2 Levels

- Low
  - 352 x 288 luminance samples, 30 Hz
- *Main*
  - *720 x 576 luminance samples, 30 Hz*
- High-1440
  - 1440 x 1152 luminance samples, 60 Hz
- High
  - 1920 x 1152 luminance samples, 60 Hz

ISO/IEC MPEG-4 Features

- Adds object based coding
- Natural and synthetic objects
- Shape coding
  - Identifies objects and background
  - Codes them separately
- Motion Coding
  - Accounts for object motion
  - 1/4 Pixel motion vectors, 8x8 motion blocks
  - Global motion vectors
- Texture Coding
  - Accounts for object color and texture

ISO/IEC MPEG-4 Profiles and Levels

- Profiles
  - Version 1: Simple, Simple Scalable, Core, Main, N-bit, Scaleable Texture, Simple Face Animation, Basic Animated Texture, and Hybrid.
  - Version 2: Core Scalable, Advanced Core, Advanced Coding Efficiency, Advanced Real Time Simple, Advanced Scaleable Texture, and Simple FBA.
  - 1st Extension: Simple Studio and Core Studio.
  - 2nd Extension: Advanced Simple and Fine Granularity Scalability.
- Levels
  - Various levels constraining profile parameters

Network Abstraction Layer (NAL)

- Formats the compressed data and provides header information in a manner appropriate for conveyance on a variety of communication channels or storage media
- NAL Unit Stream
  - Groups data by slice
  - Used for packet transmission or storage applications
- Byte Stream
  - NAL Unit plus Start Code Prefix and fill Suffix
  - Used for transmission applications

H.264 Levels

- Restricts range of parameter values
- Bit rate
- Frame size
- Motion vector range
- Slice size

H.264 Summary

- 50% coding gain over MPEG-2 and H.263 Baseline profile
- 25% coding gain over H.263 CHC profile
- Encoder complexity 3x that of older algorithms
- Decoder complexity 2x that of older algorithms

ITU-T H.264 Features

- 40-50% the bit rate reduction at the same visual quality compared to MPEG-2
- Hybrid spatial-temporal prediction model
- Flexible partition of Macro Block (MB) , sub MB for motion estimation
- Intra Prediction (extrapolate already decoded neighboring pixels for prediction) with 9 directional modes
- Entropy coding is CABAC and CAVLC
- Support Up to 4K (4,096×2,304)
- Supports up to 59.94 fps
- 21 profiles ; 17 levels

Why is it so good?

- Intra-frame prediction
- Integer-based DCT algorithm
- Improved de-blocking filter
- Flexible block sizes from 4x4 to 16x16 for improved motion estimation
- Improved entropy coding (CABAC)
- Multiple Reference Frames

Overview of H.264

(a.k.a. MPEG-4 Part 10)

(a.k.a. AVC)

Key H.264 Profiles

- Baseline Profile
  - All inter and intra coding tools (I and P slices)
  - CAVLC
  - Error concealment tools
- Main Profile
  - Same as Baseline except no ASO, FMO, RS
  - CABAC
  - Multiple Reference frames, weighted prediction (B slices)
  - Macro Block Adaptive Frame/Field (MBAFF) coding
  - Adaptive Block-size Transform (ABT)
- Constrained Baseline Profile
- High Profile (HiP)
  - The primary profile for broadcast and disc storage applications, particularly for high-definition television applications (for example, this is the profile adopted by the Blu-ray Disc storage format and the DVB HDTV broadcast service).

More H.264 Profiles

- Extended Profile
  - Data Partitioning
  - Switching I and P slices
- High 10 Profile (Hi10P)
  - Going beyond typical mainstream consumer product capabilities, this profile builds on top of the High Profile, adding support for up to 10 bits per sample of decoded picture precision.
- High 4:2:2 Profile (Hi422P)
  - Primarily targeting professional applications that use interlaced video, this profile builds on top of the High 10 Profile, adding support for the 4:2:2 chroma subsampling format while using up to 10 bits per sample of decoded picture precision.
- High 4:4:4 Predictive Profile (Hi444PP)
  - This profile builds on top of the High 4:2:2 Profile, supporting up to 4:4:4 chroma sampling, up to 14 bits per sample, and additionally supporting efficient lossless region coding and the coding of each picture as three separate color planes.

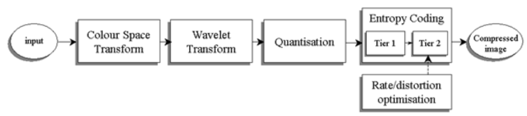
Even More H.264 Profiles

- High 10 Intra Profile
  - The High 10 Profile constrained to all-Intra use.
- High 4:2:2 Intra Profile
  - The High 4:2:2 Profile constrained to all-Intra use.
- High 4:4:4 Intra Profile
  - The High 4:4:4 Profile constrained to all-Intra use.
- CAVLC 4:4:4 Intra Profile
  - The High 4:4:4 Profile constrained to all-Intra use and to CAVLC entropy coding (i.e., not supporting CABAC).
- Scalable Baseline Profile
  - Primarily targeting video conferencing, mobile, and surveillance applications, this profile builds on top of a constrained version of the H.264/AVC Baseline profile to which the base layer (a subset of the bitstream) must conform. For the scalability tools, a subset of the available tools is enabled.
- Scalable High Profile
  - Primarily targeting broadcast and streaming applications, this profile builds on top of the H.264/AVC High Profile to which the base layer must conform.
- Scalable High Intra Profile
  - Primarily targeting production applications, this profile is the Scalable High Profile constrained to all-Intra use.
- Stereo High Profile
  - This profile targets two-view stereoscopic 3D video and combines the tools of the High profile with the inter-view prediction capabilities of the MVC extension.
- Multiview High Profile
  - This profile supports two or more views using both inter-picture (temporal) and MVC inter-view prediction, but does not support field pictures and macroblock-adaptive frame-field coding.

Overview of JPEG-2000

Wavelet based image compression

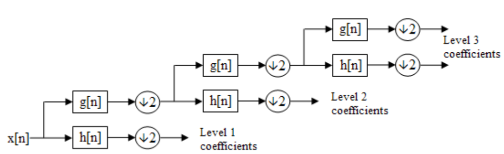
JPEG-2000 System



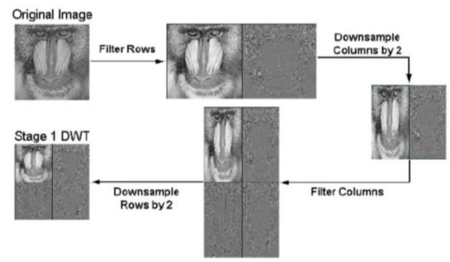
Discrete Wavelet Transform

- Captures both frequency *and* location information
- Effectively a series of filters and down sampling stages
- Reversible Transform for lossless compression
- Irreversible Transform for lossy compression

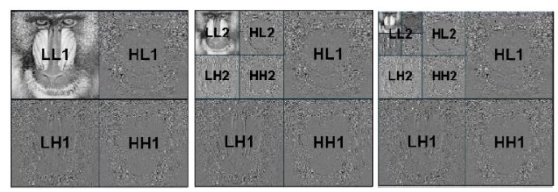
Discrete Wavelet Transform



2D Discrete Wavelet Transform

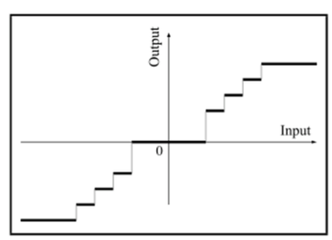


3-Stage DWT



Stage 1 Stage 2 Stage 3

Quantization



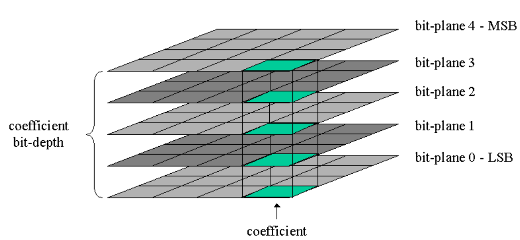
Wavelet Encoding Methods

- Embedded Zero-Tree Wavelet (EZW)
  - The bitstream is embedded and the coefficients are ordered in significance and precision, so that it can be truncated according to the bit-rate requirements of the channel.
  - It efficiently utilizes the self-similarity between the subbands of similar orientation and achieves significant data reduction.
  - However, the EZW algorithm is not exactly optimal.
- Set Partitioning in Hierarchical Trees (SPIHT)
  - Improved performance over EZW, because of its ability to exploit the grouping of insignificant coefficients.
- Embedded Block Coding with Optimized Truncation of bit-stream (EBCOT)
  - Offers both resolution scalability and SNR scalability.
  - Because of its advantages, the EBCOT algorithm has been accepted incorporated within the most recent still image compression standard JPEG-2000.

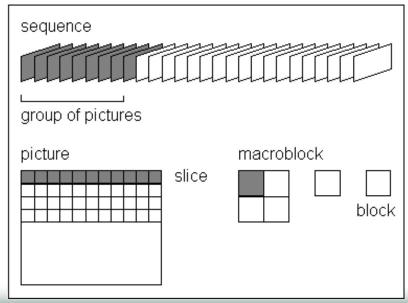
Embedded Block Coding

- Embedded Block Coding with Optimized Truncation (EBCOT)
- Divided into two layers called Tiers.
  - Tier 1 is responsible for source modeling and entropy coding
  - Tier 2 generates the output stream
- Wavelet sub-bands are partitioned into small code-blocks (to not be confused with tiles), typically 64x64
- Code blocks are in turn are separated into bit-planes. (i.e. coefficient bits of the same order are coded together)
- Bit-planes are further partitioned into three coding passes. Each coding pass constitutes an atomic code unit, called chunk.
- Bit-planes are coded using Arithmetic coding (MQ-coder).

Embedded Block Coding



MPEG Components



Levels of Video Compression

- Uncompressed
  - 166 mbps
- True Lossless Compression
  - 45 mbps (4)
- Visually Lossless
  - 3-10 mbps (16-55)
- High Quality Videoconferencing
  - 384 Kbps to 1.5 mbps (110-432)
- Acceptable Quality Videoconferencing
  - 112 Kbps to 256 Kbps (648-1482)
- Videophone
  - 20 Kbps to 33.6 Kbps (4940-8300)

Comparison of Video Coding Standards

	H.261	H.263	MPEG1	MPEG2
General Standard Structure	Narrow Profile	Narrow Profile	Generic Tool Kit	
Picture Format	176 x 144 (mandatory) 352 x 28 (optional)	SQCIF, QCIF, CIF, 4CIF, 16CIF	360 x 240	Field or Frame
Quantization Precision of Motion Vectors	One Pixel	One Half Pixel	One Half or One Pixel	
B-Frames/PB Frame	None	PB Frames	B Frames An Available Tool	
Intraframe Coding	Usually Distributed	Flexible	Full Frame is Mandatory	
Color Coding	4:2:0	4:2:0	4:4:4, 4:2:2, 4:2:0	
Picture Structure	Group of Blocks	GOB	Slice	
Dual Prime (a special motion compensation mode)	No	No	No	Yes
Nominal Bitrate	56 Kbps to 1,936 Kbps	Low Bit Rate	1.5 mbps	4 – 20 mbps
Applications	Interactive Audiovisual Services -Videophone -- Videoteleconferencing	VTC/Videophone via PSTN/Mobile Network	- VCR	- Broadcast TV - + Contribution + Distribution - DBS - HDTV

Video Coding Standards

ITU		ISO	
H.261	S 1990 S CIF 352 x 288 pixels S VHS quality S Videoconferencing/ Videophone (VC/VPO) via N-ISDN; 56 Kbps - 1.920 mbps	MPEG1	S 1993 S SIF 352 x 240 pixels S VHS quality S Star application - CDROM - 1.5 bps
H.262	S 1995 S VC/VP via B-ISDN; 2 - 20 mbps	MPEG2	S 1993 S SIF 352 x 240 pixels S VHS quality S Star application - CDROM - 1.5 mbps
H.263	S March 1996 S SQCIF, QCIF, CIF, 4CIF, 16CIF S H.324, H.323, H.324M S Higher quality than H.261	MPEG4 - Video	MPEG-4 will be interoperable with baseline H.263
H.263+	S January 1998 S 12 optional H.263 Extension S New Functionalities S Improved quality, error robustness		Version 1 - 12/98 Version 2 - 12/99 S Object based coding
H.263++	SFuture options frame-based		
H.264	S March 2003 S Twice the compression of MPEG-2 S More efficient, better tools	MPEG4 Part 10	-Integer DCT - 1/16 pixel motion estimation -Variable motion block shape

Native Video Resolutions

TYPE	NAME	ACTIVE PIXELS	STRUCTURE	FRAME RATES
SD	NTSC – 480i	720 x 480	Interlaced	30 Fps
	PAL – 576i	720 x 576	Interlaced	25 Fps
ED	NTSC – 480p	720 x 480	Progressive	30 Fps, 60 Fps
	PAL – 576p	720 x 576	Progressive	25 Fps, 50 Fps
HD	720p	1280 x 720	Progressive	24, 25, 30, 48, 50, 60 FPS
	1080i	1920 x 1080	Interlaced	24, 25, 30 FPS
	1080p	1920 x 1080	Progressive	24, 25, 30, 48, 50, 60 FPS

Common Scaled Video Resolutions

SubQCIF	128 x 96	single field
QCIF	176 x 144	single field
CIF	352 x 288	single field
2xCIF	704 x 288	single field
4xCIF	704 x 576	pseudo progressive
16xCIF	1408 x 1152	pseudo progressive
SIF	360 x 240	single field
2xSIF	720 x 240	single field
½ D1	360 x 480	interlaced
D1	720 x 480	interlaced

ITU-T H.265 Features

- 40-50% of the bit rate at the same visual quality compared to H.264
- Enhanced Hybrid spatial-temporal prediction model
- Flexible partitioning, introduces Coding Tree Units (Coding, Prediction and Transform Units -CU, PU, TU) and larger block structure (64x64) with more variable sub partition structures
- 35 directional modes for intra prediction
- Superior parallel processing architecture, enhancements in multi-view coding extension
- Up to 8K UHDTV (8192×4320)
- 3 approved profiles, draft for additional 5 ; 13 levels

H.265

- Replaces 16x16 Macroblocks with 64x64 Coding Tree Units (CTU)
- Increases Intra Prediction modes from 9 to 35
- Adds 16x16 and 32x32 to the existing 8x8 and 4x4 Integer DCT Transform
- Decode independent Tiles for increased parallelism

An Introduction to Audio Compression

AGENDA

- Audio Codecs
- Types of Audio Codecs
- Waveform Coders
- Source Coders
- Silence Compression
- Echo Cancellation

Classes of Audio Codecs

- Speech Codecs
  - Optimized for human speech
  - Single channel
  - Low sample rates
  - Limited audio bandwidth
- Audio Codecs
  - Optimized for full fidelity audio (music, sound effects, etc)
  - Multiple channels
  - High sample rates
  - High audio bandwidth

Speech Codecs

Types of Speech Codecs

- Waveform Coding
  - Speech waveform is digitized, coded, and transmitted.
- Source Coding
  - Speech source is modeled by an excitation and a filter. Characteristics of the excitation and filter are coded and transmitted.

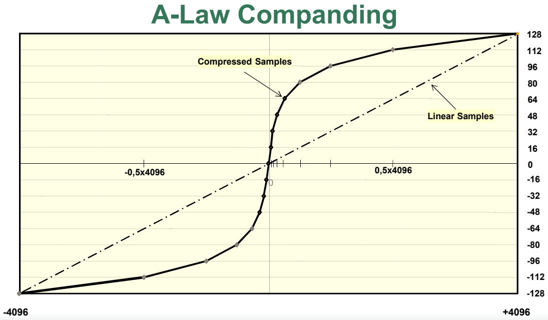
Waveform Codecs

- G.711 A-Law and u-Law PCM
- G.722 ADPCM

G.711 PCM

- A-Law or u-Law Companding
- 8 Ksps = 3.5 KHz bandwidth
- Converts 16-bit sample to 13-bits by truncation.
- Converts 13-bit sample to 8-bits by logarithmic translation.
- Preserves dynamic range
- Compressed data rate: 64kbps

A-Law Companding



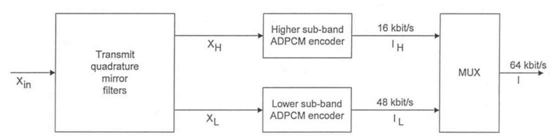
G.722 SubBand ADPCM

- Adaptive Differential PCM
- 16 Ksps = 7 KHz Bandwidth
- Filters signal into a high band and a low band.
- Applies ADPCM to each band.
- Higher bitrate allocated to low band
- Compressed data rate: 64 Kbps

ADPCM

- Previous samples form a prediction of the next samples.
- Take the difference between the prediction and the actual speech.
- Code the difference.
- Adapt the prediction as the speech changes.

Block Diagram of the G.722 Encoder

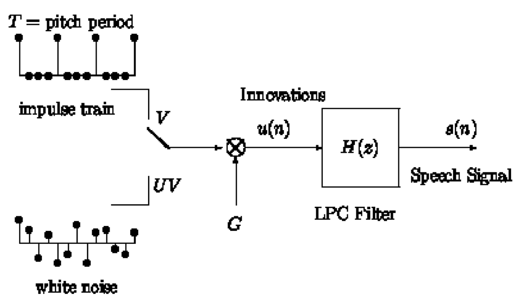


Source Codecs

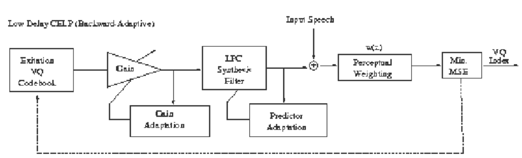
- G.728 LD-CELP
- G.729 ACELP
- G.723.1 MPE/MLQ and ACELP

Linear Prediction Coding

- Model human speech by using an excitation signal passed through a filter.
- Match the model to the input speech using analysis by synthesis.
- Only transmit info about the excitation signal and characteristics of the filter.



Linear Prediction Diagram



G.728 LD-CELP

- Low Delay – Code Excited Linear Prediction
- 8 Ksps = 3.5 KHz Bandwidth
- Find codebook excitation and filter that best matches signal.
- Code and transmit codebook index, gain, and filter coefficients.
- Compressed data rate: 16 Kbps

G.729 CS-ACELP

- Conjugate-Structure Algebraic Code Excited Linear Predictor
- 8 Ksps = 3.5 KHz bandwidth
- Uses Conjugate Structured codebook which simplifies the signal match search
- Compressed data rate: 8 Kbps

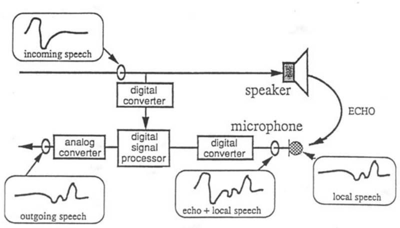
G.723.1 MP-MLQ/ACELP

- Multipulse – Maximum Likelihood Quantization
- Algebraic Code Excited Linear Prediction
- 8 Ksps = 3.5 KHz bandwidth
- MP-MLQ = 6.3 Kbps
- ACELP = 5.3 Kbps

Silence Compression

- Used in G.723.1 and G.729
- Reduces bitrate during silence periods
- Voice Activity Detection (VAD)
  - Spectral analysis: presence or absence of speech
  - Operates on 30 ms speech frames which are encoded or filled with comfort noise
- Comfort Noise Generator (CNG)
  - Generates artificial background noise that matches the actual background noise

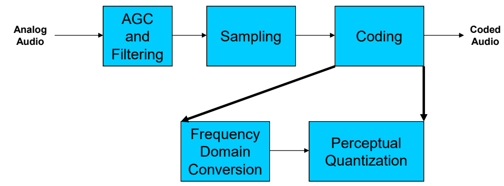
Acoustic Echo Cancellation



Types of Audio Codecs

- Sub-band Coding
  - Audio signal is applied to a series of filter banks which are then entropy encoded.
- Modified Discrete Cosine Transform (MDCT)
  - DCT performed on overlapping blocks data where the last half of the previous block is the first half of the next block.
- Sub-band MDCT
  - Audio signal is prefiltered with a 32-band polyphase quadrature filter (PQF) bank before the MDCT is applied.

Audio Codec



Audio Channels

- Monaural
  - Single channel
- Stereo
  - Two channels
- Surround Sound
  - 5.1 (5 channels plus 1 Low Frequency Enhancement)

Perceptual Quantization

- Removes information that the human auditory system will not be able to easily perceive.
- To choose which information to remove, the audio signal is analyzed according to a psychoacoustic model, which takes into account the parameters of the human auditory system.
- Research into psychoacoustics has shown that if there is a strong signal on a certain frequency, then weaker signals at frequencies close to the strong signal's frequency cannot be perceived by the human auditory system. This is called frequency masking.
- By ignoring information at frequencies that are deemed to be imperceptible, more data can be allocated to the reproduction of perceptible frequencies.

MPEG-1/Layer 2 (MP2)

- Primarily used for Broadcast Audio
- Sub-band
  - Splits the input audio signal into 32 sub-bands using sub-band filter banks, and if the audio in a sub-band is deemed to be imperceptible then that sub-band is not transmitted.
- MPEG-1 Audio Layer II
  - Sampling rates: 32, 44.1 and 48 kHz
  - Bit rates: 32, 48, 56, 64, 80, 96, 112, 128, 160, 192, 224, 256, 320 and 384 kbit/s
  - Up to two audio channels
- MPEG-2 Audio Layer II
  - Sampling rates: 16, 22.05 and 24 kHz
  - Bit rates: 8, 16, 24, 40 and 144 kbit/s
  - Multichannel support - up to 5 full range audio channels and an LFE-channel (Low Frequency Enhancement channel)

MPEG-1/Layer 3 (MP3)

- Primarily used for PC and Internet audio
- MDCT/Hybrid Sub-band
  - Transforms the input audio signal into 576 frequency components in the frequency domain using the MDCT allowing finer application of the perceptual filtering
- MPEG-1 Audio Layer III
  - Bitrate: 32, 40, 48, 56, 64, 80, 96, 112, 128, 160, 192, 224, 256 and 320 kbit/s
  - Sample Rate:32, 44.1 and 48 kHz.
- MPEG-2 Audio Layer III
  - Bitrate: 8, 16, 24, 32, 40, 48, 56, 64, 80, 96, 112, 128, 144, 160 kbit/s
  - Sample Rate: 16, 22.05 and 24 kHz
- MPEG-2.5 Audio Layer III
  - Bitrate: 8, 16, 24, 32, 40, 48, 56 and 64 kbit/s
  - Sample Rate: 8, 11.025, and 12 kHz

Advanced Audio Coding (AAC)

- Primarily used for YouTube, iPhone, iTunes, Nintendo
- Improvement over MP3
- MDCT/Hybrid Sub-band
  - The signal is converted from time-domain to frequency-domain using forward modified discrete cosine transform (MDCT).
  - The frequency domain signal is quantized based on a psychoacoustic model and encoded.
- Sample rates: from 8 to 96 kHz
- Bit Rate: Arbitrary (2kbps to 320kbps)
- Up to 48 channels

Contact Information

747 Dresher Road
HORSHAM, PA 19044
Ph: 215-657-5270
FAX: 215-657-5273
www.deltadigitalvideo.com

Backup Slides

Typical Audio Quality Problems

- Too many errors
  - Packet loss
- Too much delay
  - Congestion
  - Too many router hops
  - QOS parameters

Audio Quality

- Packet Delay
  - Time for voice packets to transit from the sender to the receiver
  - If too large, makes conversation difficult
- Packet Jitter
  - Variation in packet delay
  - If too large, requires larger buffering, increasing delay or causing buffer underflow or overflow
- Packet Loss
  - Packets are dropped due to congestion, poor cabling, equipment malfunction, duplex mismatch, bit errors, CRC errors
  - Since there is no retransmission, results in lost information
  - Clicks, chirps, hum in audio

Audio Quality

- One Way Delay
  - Good (0ms-150ms)
  - Acceptable (150ms-300ms)
  - Unacceptable (> 300ms)
- Jitter
  - Good (0-20ms)
  - Acceptable (20ms-50ms)
  - Unacceptable (>50ms)
- Packet Loss
  - Good(0%-0.5%)
  - Acceptable (0.5%-1.5%)
  - Unacceptable (>1.5%)