Appendix 3

Video Formats and Compression

Material borrowed from Teppo Räisänen, Klara Nahrstedt, and Roger Cheng

General Information

Originally video material was processed using analog tools

Nowadays it is common, that digital processing is used in

- recording
- editing
- broadcasting

Besides TV-broadcasting and movies video material is widely used within multimedia production

With higher bandwidths the Internet has become a common way of video transfers / broadcasting

General Information

The most recent uses of digital video are e.g.

- DVD/BlueRay
- Digital TV-Broadcasting
- Mobile Devices used for playing video material
- Streaming services

Hardware Environment

Nowadays production environments are mostly based on digital tools

There can, however, be some analog tools integrated into production chain

Editing of video material can be done using either a dedicated device or a computer

Software Environment

For viewing a video file a player is needed, e.g.

- Media Player
- Quicktime Player
- RealPlayer
- VLC player

There are often major differencies between the release versions of players

Software Environment

For editing of video a dedicated piece of software is needed e.g.

- Adobe Premiere
- Avid
- Final Cut Pro

WinXP includes a Movie Maker –tool which is capable of basic video editing

In Linux, KDEnlive and other video editing tools are avaialble

Video as an Element of Multimedia

Video material holds a central position within multimedia production

If the multimedia contents are delivered using CDs of DVDs large file sizes can be used

If the contents are delivered via network file sizes and video quality needs to be compromized

Video File Formats/Containers

Video for Windows

- Microsoft product
- Also referred as AVI files (Audio Video Interleaved)

Quicktime

- Apple product
- .mov files

Matroska

- .mkv files

MP4

- .mp4 files

Video File Formats

MPEG

- A standard covering video & audio
- Most commonly used layers are MPEG-1, MPEG-2 ja MPEG-4
- File extensions .mpeg and .mpg
- VideoCD is based on MPEG-1
- DVD uses MPEG-4 compression

Video File Formats

RealMedia

- Enables streaming broadcasting
- A single .rm file can include multilple versions of the video directed for different connection speeds
- The version can be changed dynamically during video's broadcasting

Video File Formats

Windows Media

- Microsoft product
- A compressed format aimed especisally for Internet delivery of video
- .wmv & .wma files

Properties of Video Files

Properties affecting to the quality of video (not dependent on the file format) are:

- Screen size
- Frames per second
- Compression algorithm (if used)
- Compression (if used) settings

Screen Size

Is declared as a px x px value

There are multiple aspect ratios available, but the screen is always of a rectangular shape

Many screens are of typical ratios like 4:3 (640x480, 800x600) or 16:9 (1280x720, 1920x1080, 3840x2160)

Frames / Second

The human eye views individual pictures as a continuous flow when the frame rate exceeds 20

Lower fps will result to a non-continuous impression

Pal TV uses fps of 50 (interlaced)

Modern high end TV systems use frame rates over 60

Compression Algorithms

Uncompressed video files' sizes can be of terabytes

To enable delivery using portable media the material is compressed (lossy compression)

The common name for compression methods is 'codec' (**Co**mpress/**Dec**ompress)

Typically compression process of video takes much more time than processing audio

Compression Methods

Compression can be based on

- Compression of the contents of individual frames
- Compression of the consequtive frames (interframe compression)

Interframe compression

- Works especially well then the contents of consequtive frames are almost similar
- Uses keyframes
- Of the frames between keyframes, only changing contents are stored within the video file

Compression Algorithms

M-JPEG

- Based on JPEG
- Enables compression ratio of 5:1

MPEG

- Umbrella concept of many substandards
- Compression of both individual fames and interframe technology

Compression Algorithms

Indeo Video Interactive

- A software based codec
- Enables e.g. embedding of hyperlinks to the video material

Cinepak

- A software based codec
- Video for Windows & QuickTime use Cinepak
- Decoding of Cinepak video does not require much computing power

Compression Algorithms

DivX

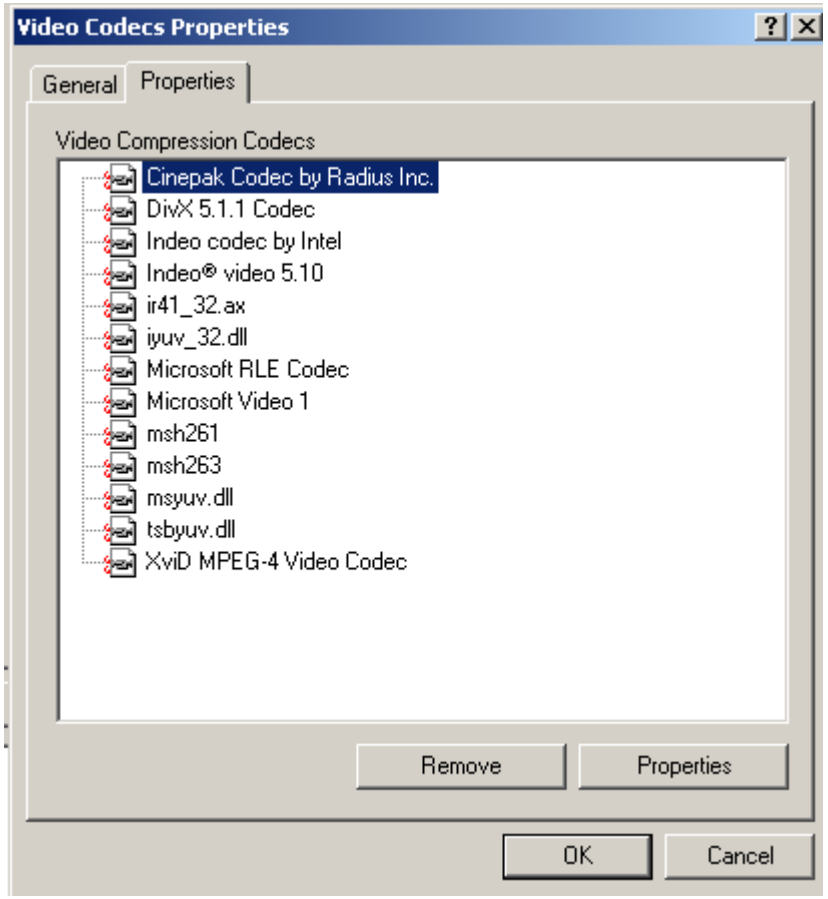
- A Registered Trademarik
- Tools for producing and playing compressed video
- An effective compression method

Besides the aforementioned there are many products available for the compression of video material

Compression Algorithms

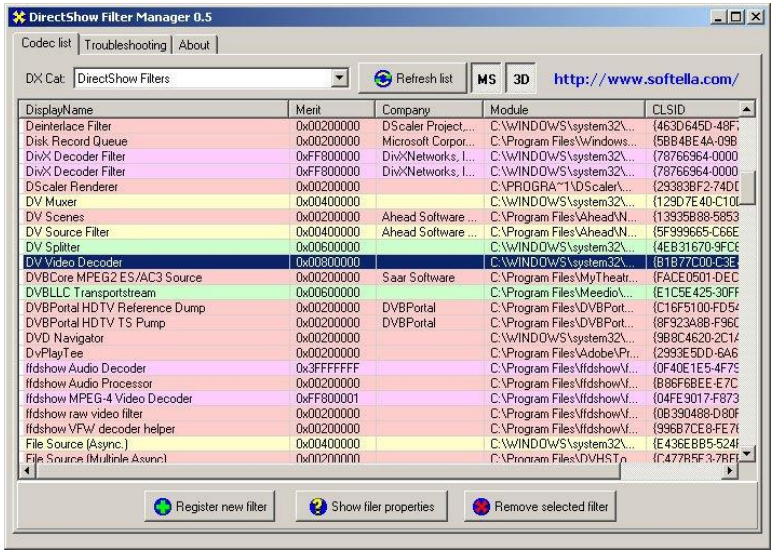
The codecs installed to a specific computer can be viewed via Control Panel (Windows)

In many cases codecs are installed automatically when e.g. installing a media player application

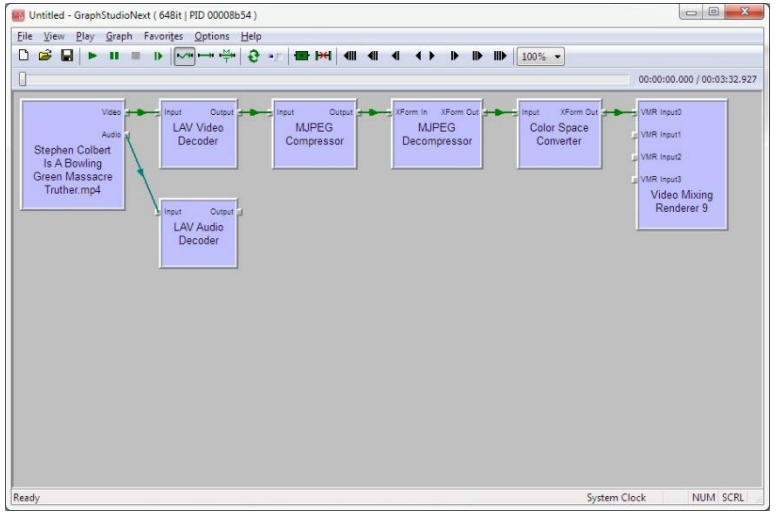


Codecs in Windows

Manipulating the codecs installed on Windows can be done with tools, such as Directshow Filter Manager



GraphStudio shows you the graph Windows uses to combine the DirectShow filters to form the entire pipeline



Images taken from videohelp.com

Compression Algorithms And File Sizes

E.g. transfer/playback of 800x600 px uncompressed video using 24bit colors and 25 fps would require ~36 Mb/s data transfer speed

When editing the video high resolutions are typically used

For different media versions, different levels of quality can be produced

Compression Algorithms And File Sizes

Compression ratio specifies the file size ratio between the original and the compressed file (e.g. 4:1)

The algorithms are based on the human eye's features for e.g. detecting motion and colors

Video Encoding/Compression

Once video is in digital format, it makes sense to compress it

Similarly to image compression, we want to store video data as efficiently as possible

Again, we want to both maximize quality and minimize storage space and processing resources

This time, we can exploit correlation in both space and time domains

TMI! (Too Much Information)

Unlike image encoding, video encoding is rarely done in lossless form

No storage medium has enough capacity to store a practical sized lossless video file

- Lossless DVD video - 221 Mbps
- Compressed DVD video - 4 Mbps
- 50:1 compression ratio!

Definitions

- Bitrate
  - Information stored/transmitted per unit time
  - Usually measured in Mbps (Megabits per second)
  - Ranges from < 1 Mbps to > 40 Mbps
- Resolution
  - Number of pixels per frame
  - Ranges from 160x120 to 1920x1080
- FPS (frames per second)
  - Usually 24, 25, 30, or 60
  - Don’t need more because of limitations of the human eye

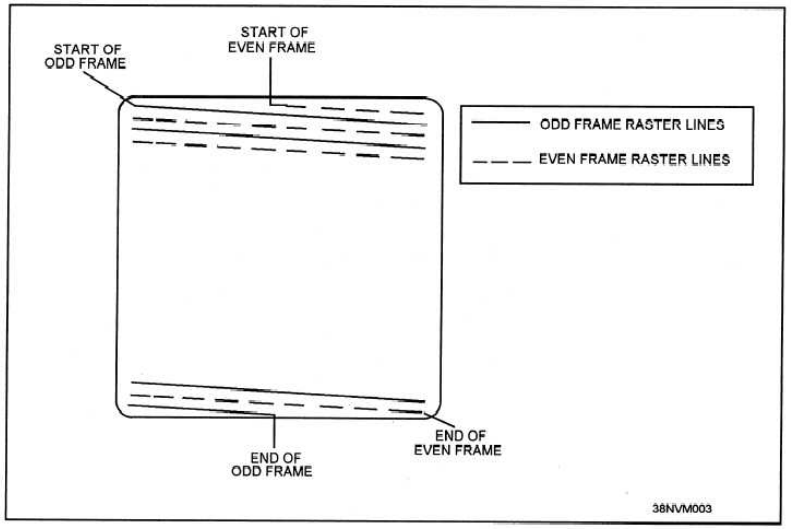
Scan types

Interlaced scan

Odd and even lines displayed on alternate frames

Initially used to save bandwidth on TV transmission

When displaying interlaced video on a progressive scan display, can see “comb effect”



Scan types

Progressive scan

- Display all lines on each frame
- New “fixed-resolution” displays (such as LCD, Plasma) all use progressive scan
- Deinterlacing is not a trivial task

MPEG (Moving Pictures Expert Group)

Committee of experts that develops video encoding standards

Until recently, was the only game in town (still the most popular, by far)

Suitable for wide range of videos

- Low resolution to high resolution
- Slow movement to fast action

Can be implemented either in software or hardware

Evolution of MPEG

MPEG-1

- Initial audio/video compression standard
- Used by VCD's
- MP3 = MPEG-1 audio layer 3
- Target of 1.5 Mb/s bitrate at 352x240 resolution
- Only supports progressive pictures

Evolution of MPEG

MPEG-2

- Current de facto standard, widely used in DVD and Digital TV
- Ubiquity in hardware implies that it will be here for a long time
  - Transition to HDTV has taken over 10 years and is not finished yet
- Different profiles and levels allow for quality control

Evolution of MPEG

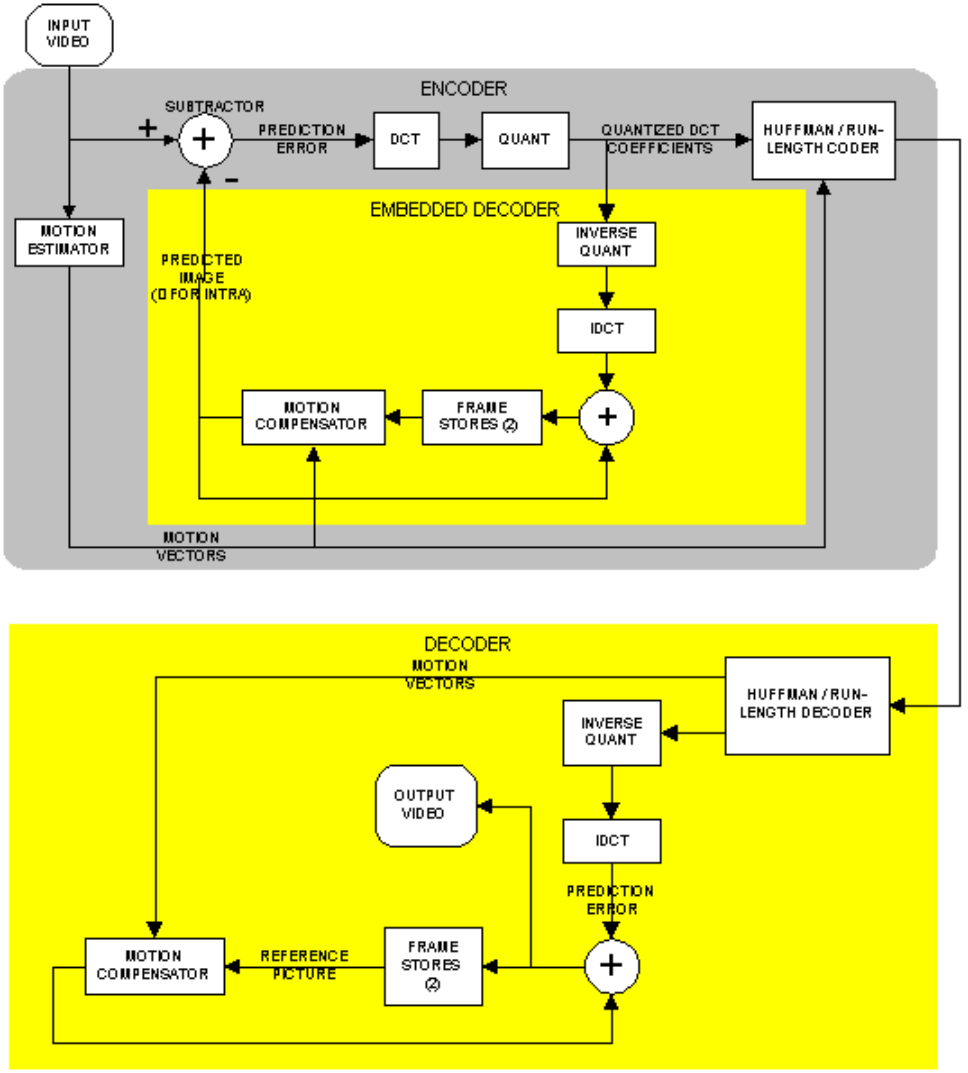
MPEG-3

- Originally developed for HDTV, but abandoned when MPEG-2 was determined to be sufficient

MPEG-4

- Includes support for AV “objects”, 3D content, low bitrate encoding, and DRM
- In practice, provides equal quality to MPEG-2 at a lower bitrate, but often fails to deliver outright better quality
- MPEG-4 Part 10 is H.264, which is used in HD-DVD and Blu-Ray

MPEG Block Diagram



MPEG technical specification

Part 1 - Systems - describes synchronization and multiplexing of video and audio.

Part 2 - Video - compression codec for interlaced and non-interlaced video signals.

Part 3 - Audio - compression codec for perceptual coding of audio signals. A multichannel-enabled extension of MPEG-1 audio.

Part 4 - Describes procedures for testing compliance.

Part 5 - Describes systems for Software simulation.

Part 6 - Describes extensions for DSM-CC (Digital Storage Media Command and Control.)

Part 7 - Advanced Audio Coding (AAC)

Part 8 - Deleted

Part 9 - Extension for real time interfaces.

Part 10 - Conformance extensions for DSM-CC.

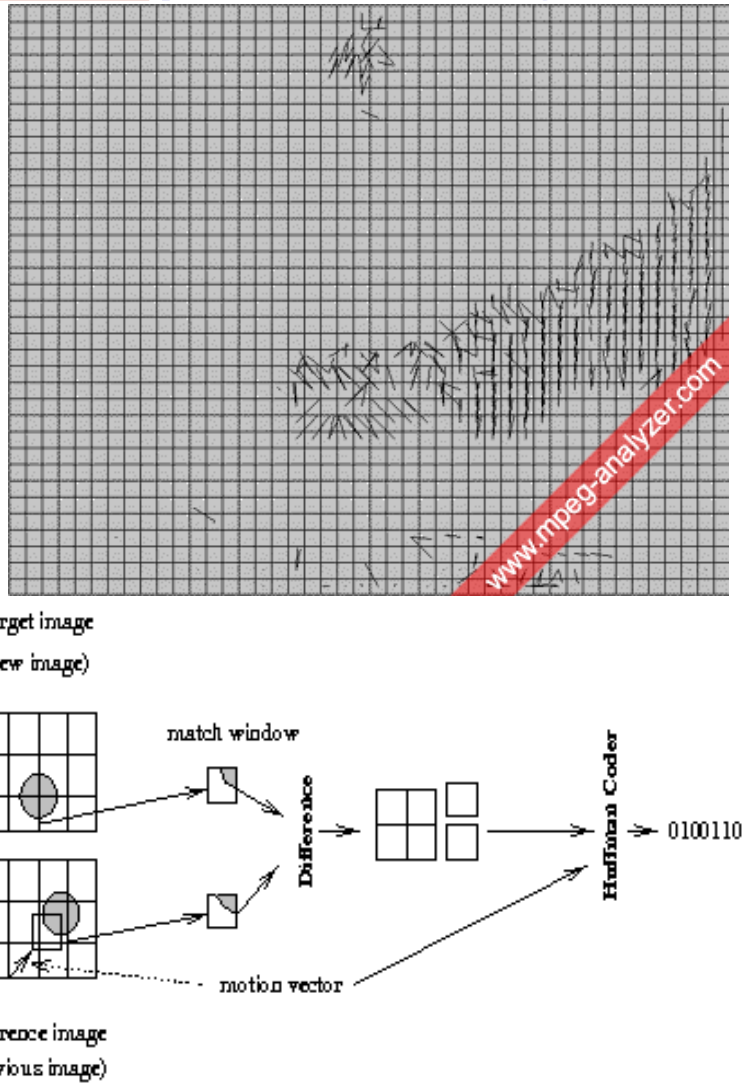
MPEG video spatial domain processing

Spatial domain handled very similarly to JPEG

- Convert RGB values to YUV colorspace
- Split frame into 8x8 blocks
- 2-D DCT on each block
- Quantization of DCT coefficients
- Run length and entropy coding

MPEG video time domain processing

- Totally new ballgame (this concept doesn't exist in JPEG)
- General idea – Use motion vectors to specify how a 16x16 macroblock translates between reference frames and current frame, then code difference between reference and actual blocl



Types of frames

I frame (intra-coded)

- Coded without reference to other frames

P frame (predictive-coded)

- Coded with reference to a previous reference frame (either I or P)
- Size is usually about 1/3rd of an I frame

B frame (bi-directional predictive-coded)

- Coded with reference to both previous and future reference frames (either I or P)
- Size is usually about 1/6th of an I frame

GOP (Group of Pictures)

GOP is a set of consecutive frames that can be decoded without any other reference frames

Usually 12 or 15 frames

Transmitted sequence is not the same as displayed sequence

Random access to middle of stream – Start with I frame

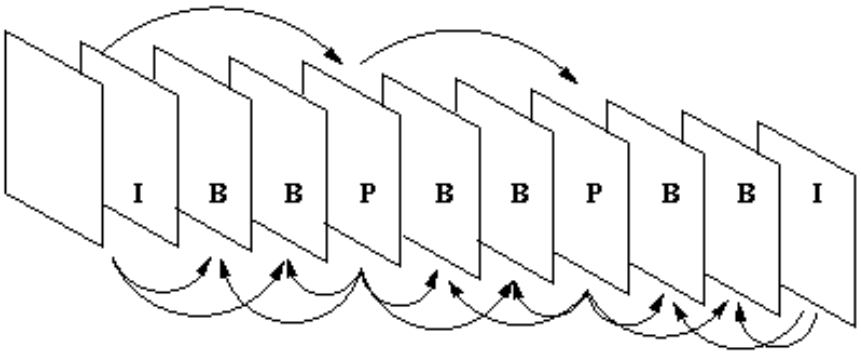


Figure 1: Prediction between MPEG-2 Frames

Things about prediction

- Only use motion vector if a “close” match can be found
  - Evaluate “closeness” with MSE or other metric
  - Can’t search all possible blocks, so need a smart algorithm
  - If no suitable match found, just code the macroblock as an I-block
  - If a scene change is detected, start fresh
- Don’t want too many P or B frames in a row
  - Predictive error will keep propagating until next I frame
  - Delay in decoding

Bitrate allocation

CBR – Constant BitRate

- Streaming media uses this
- Easier to implement

VBR – Variable BitRate

- DVD's use this
- Usually requires 2-pass coding
- Allocate more bits for complex scenes
- This is worth it, because you assume that you encode once, decode many times

MPEG audio

MPEG-1 – 3 layers of increasing quality, layer 3 being the most common (MP3)

- 16 bits
- Sampling rate - 32, 44.1, or 48 kHz
- Bitrate – 32 to 320 kbps
- De facto - 44.1 kHz sample rate, 192 kbps bitrate

MPEG-2 – Supports > 2 channels, lower sampling frequencies, low bitrate improvement

AAC (Advanced Audio Coding)

- More sample frequencies (8 kHz to 96 kHz)
- Higher coding efficiency and simpler filterbank
- 96 kbps AAC sounds better than 128 kbps MP3

Usually CBR, but can do VBR

MPEG Container Format

Container format is a file format that can contain data compressed by standard codecs

2 types for MPEG

- Program Stream (PS) – Designed for reasonably reliable media, such as disks
- Transport Stream (TS) – Designed for lossy links, such as networks or broadcast antennas

AV Synchronization

Want audio and video streams to be played back in sync with each other

Video stream contains “presentation timestamps”

MPEG-2 clock runs at 90 kHz

- Good for both 25 and 30 fps

PCR (Program Clock Reference) timestamps are sent with data by sender

Receiver uses PLL (Phase Lock Loop) to synchronize clocks

Audio/Video Streams

Audio and Video are encoded separately in different containers, such as MPEG, mkv, or mp4

Streams can be extracted from the container:

demux – demultiplex

Different streams can be added/included in the container:

mux – multiplex

This allows one to include different audio streams to support various languages or different quality video streams

Real time video encoding

Motion estimation will be worse, so need higher bitrate to compensate

Very hard to do in software, need dedicated hardware or hardware assistance

Tivo, ReplayTV do this

Streaming media

Common types include Flash, RealVideo, Quicktime, WebM

Usually have low bandwidth available, need to optimize as such

Want dedicated network protocols for this purpose

- TCP will wait indefinitely for retransmission, so is often not suitable

MPEG data stream

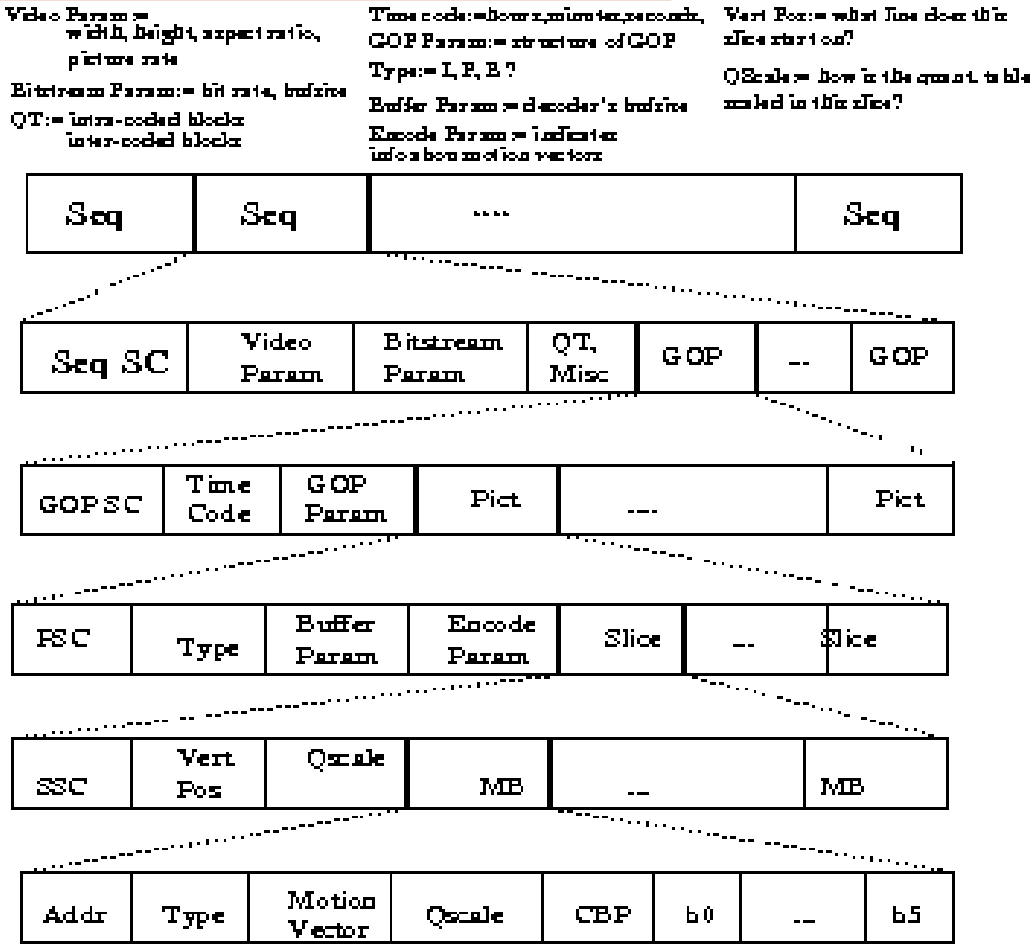


Figure 15: Video Data Stream

H.261 – Video Coding for Video Conferencing

H.261 – CCITT Recommendation of ITU-T Standard

- Developed for interactive conferencing applications
- Symmetric coder - real-time encoding and decoding
- Rates of p x 64 Kbps for ISDN networks
- **Only I and P frames**

H.261 Design

ITU-T Video Coding Experts Group (VCEG) Standard – 1988

- Bit rates between 40 kbps-2 Mbps
- Video frame sizes
  - CIF (352x288 luma, 176x144 chroma)
  - QCIF (176x144 luma, 88x72 croma) using 4:2:0 sampling scheme

H.261 Design

Basic processing unit – macroblock

Macroblock consists of

- **16x16 luma samples**
- two corresponding **8x8 chroma** samples,
- 4:2:0 sampling and **YCbCr color space**

DCT transform coding is used to reduce spatial redundancy

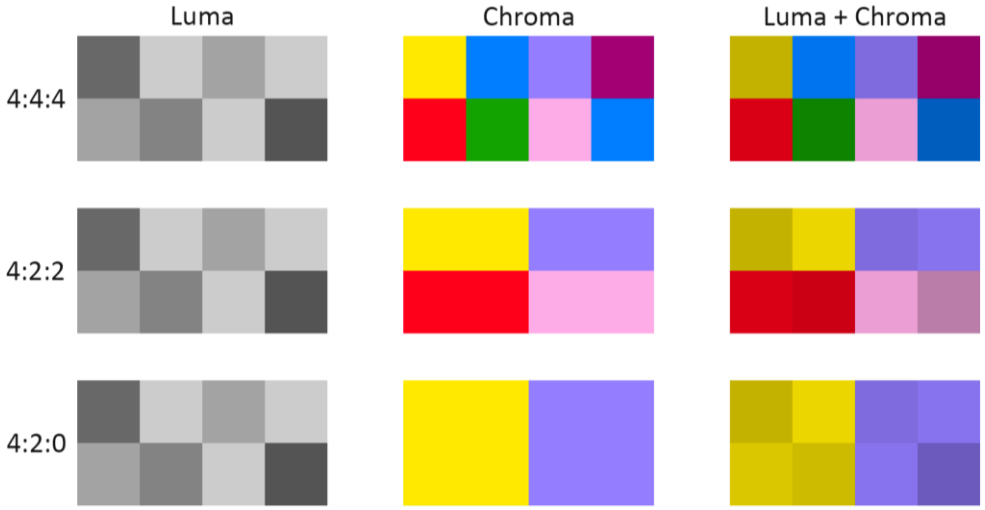
Scalar quantization and Zig-zag scanning

Entropy coding with RLE

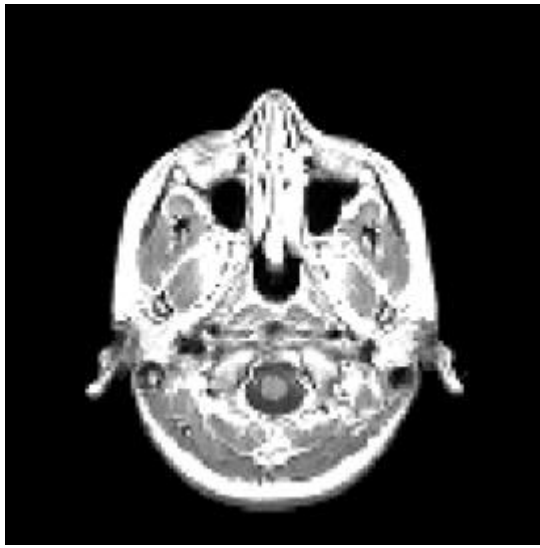
H.261 Design

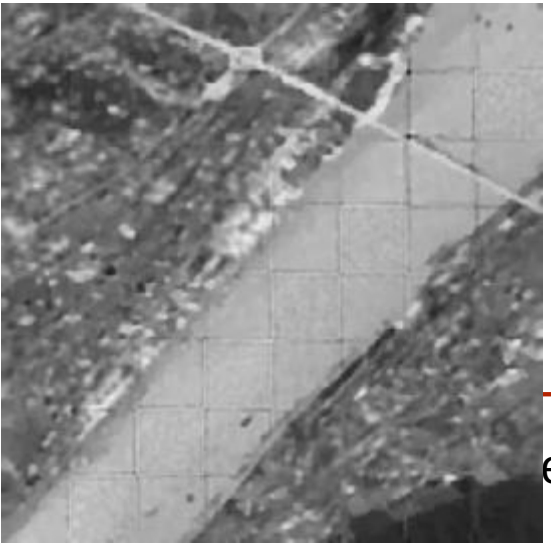
YCbCr: Y is the luma component and CB and CR are the blue-difference and red-difference chroma components

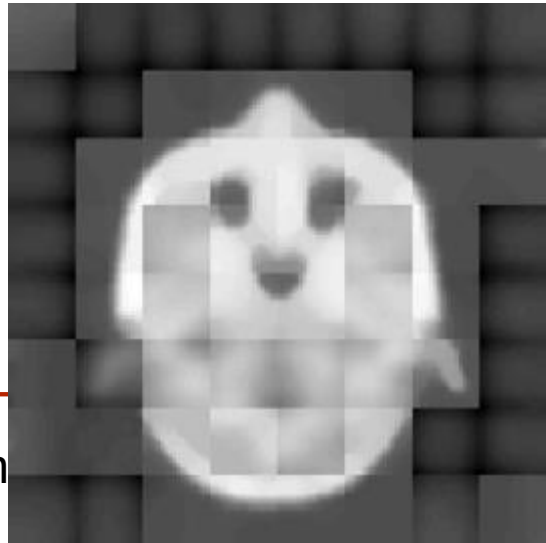
A signal with chroma 4:4:4 has no compression (so it is not subsampled) and transports both luminance and color data entirely. In a four by two array of pixels, 4:2:2 has half the chroma of 4:4:4, and 4:2:0 has a quarter of the color information available. The 4:2:2 signal will have half the sampling rate horizontally, but will maintain full sampling vertically. 4:2:0, on the other hand, will only sample colors out of half the pixels on the first row and ignores the second row of the sample completely.



Blocking Problem in Compressed Images







How Blocking Effect Happens

At low bit rates, the quantization step size is large

Larger step sizes can force many DCT coefficients to zero

If only DC and few AC coefficients remain, reconstructed picture **appears blocky**

H.261 Design

Uses post-processing technique called

- Deblocking filtering (loop filter)
- Key element of H.261 (started here)

Deblocking filtering

- Reduces appearance of block-shaped artifacts caused by block-based motion compensation and spatial transform parts of design

Deblocking Filter

Applied for **low bit rate video** 64kbps and 128 kbps

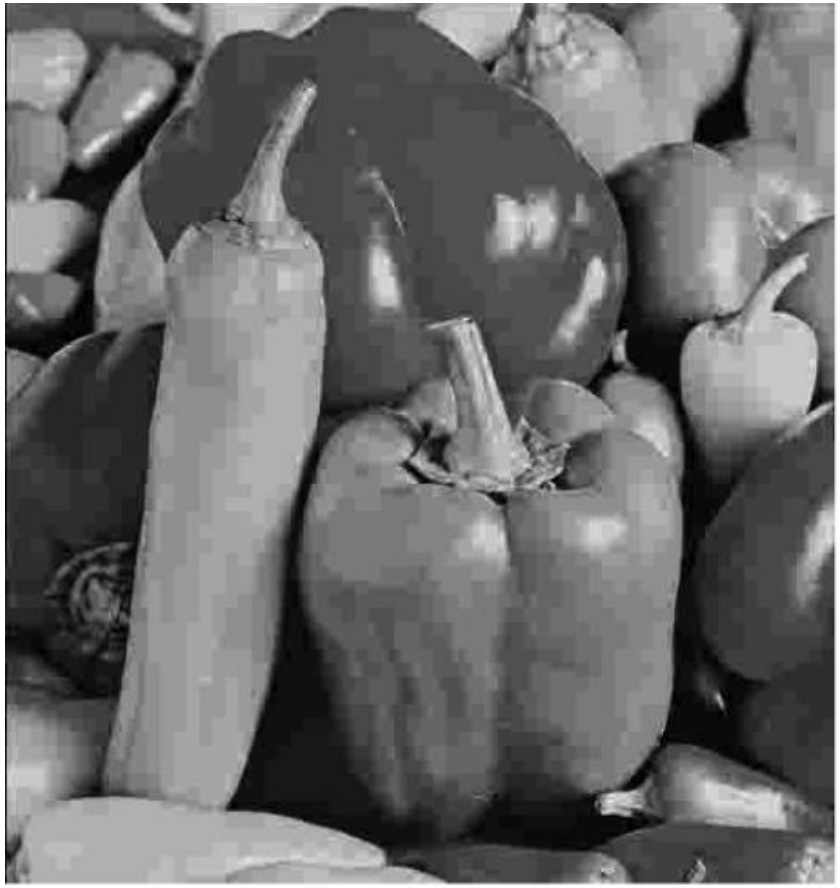
Blockiness degradations appear as staircase noise

- Mosquito noise

Artifacts are reduced by using deblocking filter

- low pass filter removing high frequency and block boundary distortions

Deblocking filter



Without filter

With filter

H.263 – video coding for low bit rate

H.263 – established 1996

- Used for **low bit rate transmission**
- Improvements of error correction and performance
- Takes in PB-frames mode
- Temporal, Spatial and SNR scalability

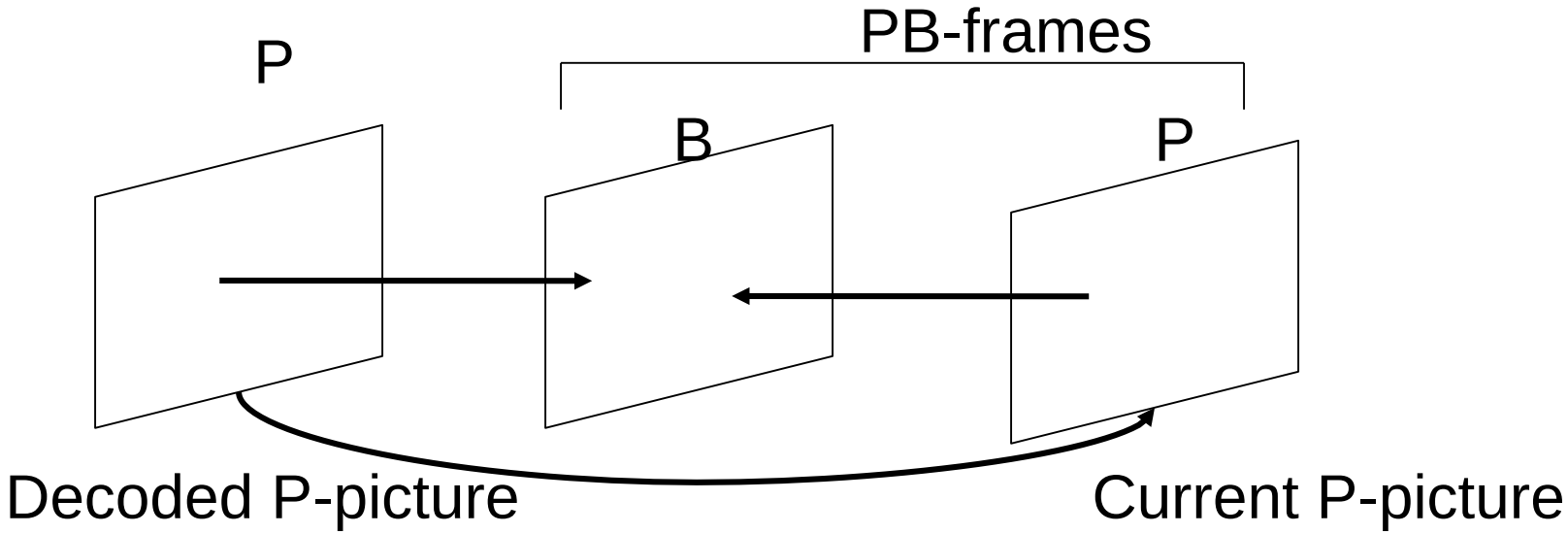
H.263 – PB-Frames Mode

A PB-frames consist of two pictures encoded as one unit.

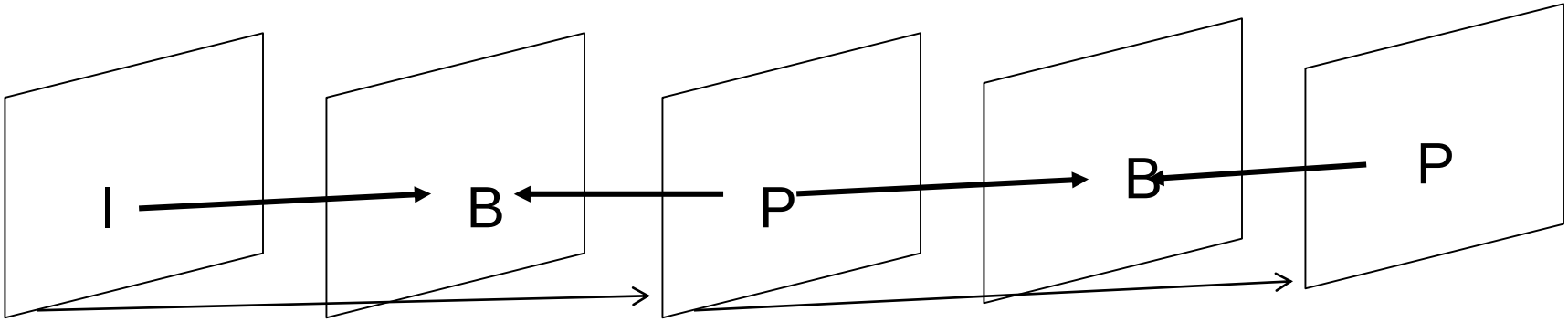
PB-frame consists of

One P-picture which is predicted from last decoded P-picture

One B-picture which is predicted from last decoded P-picture and the P-picture currently being decoded.



Comment on Temporal Scalability



Temporal scalability is achieved using B-pictures

These B pictures differ from B-picture in PB-frames

- they are not syntactically intermixed with subsequent P-picture

H.263 is used for low frame rate apps (e.g., mobile), hence in base layer there is one B-picture between I and P pictures.

H.264/MPEG-4 AVC Part 10

Joint effort between

- ITU- Video Coding Experts Group (VCEG) and
- ISO/IEC Moving Picture Experts Group (MPEG)
- 2003 completed

H.264 – codec

- Standard for Blu-ray Discs
- Streaming internet standard for videos on YouTube and iTunes Store
- web software Adobe Flash Player and Microsoft Silverlight support H.264
- Broadcast services – direct broadcast satellite television services; cable television services

H.264 Characteristics

Sampling structure

YCbCr 4:2:2 and YCbCr 4:4:4

Scalable Video Coding (SVC) allows

Construction of bit-streams that contain sub-bit-streams that also conform to standard

Temporal bit-stream scalability, spatial and quality bit-stream scalability

Completed in 2007

Scalable Video Coding

Encoding of high-quality video stream that contains one or more subset of bitstreams

- Allows for sending video over lower bandwidth networks
  - Reduced bandwidth requires
    - Temporal scalability - Lower spatial resolution (smaller screen)
    - Spatial scalability - Lower temporal resolution (lower frame rate)
    - SNR/Quality/Fidelity scalability - Lower quality video signal
- Subset bitstream can be derived by dropping packets from larger video

H.264 Characteristics

Multi-view Video Coding (MVC)

- Construction of bit-streams that represent more than one video of a video scene
  - Example: stereoscopic (two-view) video
  - Example: free viewpoint television
  - Example: multi-view 3D television
- Two profiles in MVC:
  - Multi-view High Profile (arbitrary number of views);
  - Stereo High Profile (two-view stereoscopic video);
- Completed in 2009

MVC

Contains large amount of inter-view statistical dependencies

- Cameras capture same scene from different viewpoints

Combined temporal and inter-view prediction

- Key for efficient MVC encoding
- Frame from certain camera can be predicted not only from temporally related frames from same camera, but also from neighboring cameras

H.264 Characteristics

Multi-picture inter-picture prediction

- Use previously-encoded pictures as references in more flexible way than in past standards
- Allow up to 16 reference frames to be used in some cases
  - Contrast to H.263 where typically one or in some cases conventional “B-pictures”, two.
- Use variable block size from 16x16 to 4x4
- Use multiple motion vectors per macro-block (one or two per partition where partition can be a block of 4x4)

H.264 Characteristics

New Transform design features

Similar to DCT, but simplified and made to provide exactly-specified decoding

Quantization

Frequency-customized quantization scaling matrices

selected by encoder based on perception optimization

Entropy Encoding

Context-adaptive variable-length coding

Context-adaptive binary arithmetic coding (CABAC)

H.265/HEVC/MPEG-H Part 2

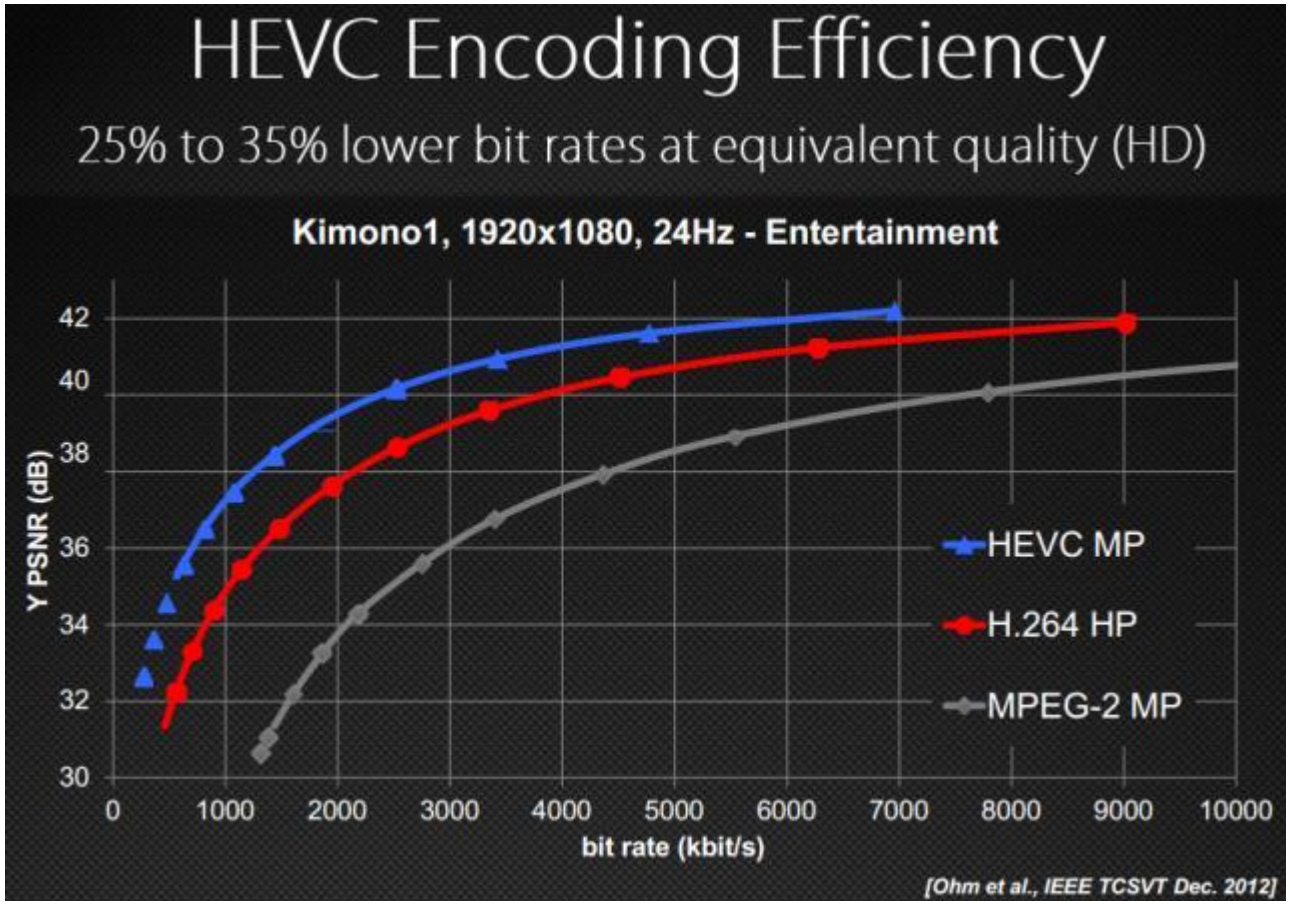
Main drivers

- Get **Low bitrate target – target 2:1** over H.264
- Cheat your eyes – how much can you cut bits and still see the same quality
- Improve **resolutions (8K by 4K and 4K by 2K) and frame rates**
- Launch **1080p50/60 services** to compete against BluRay
- Expect <10x more computational complexity and 2x-3x (decode)

H.265 – High Efficience Video Coding (HEVC)

ISO MPEG and ITU-T

- Proposed in January 2013
- Double the compression rate to H.264



H.266 - Versatile Video Coding (VVC)

Finalized on 6 July 2020, by the Joint Video Experts Team (JVET):

- joint video expert team of the VCEG working group of ITU-T Study Group 16 and the MPEG working group of ISO/IEC JTC 1
- supports resolutions from 4K to 16K as well as 360° videos
- supports YCbCr 4:4:4, 4:2:2 and 4:2:0 with 10 to 16 bits per component
- BT.2100 wide color gamut and high dynamic range (HDR)
- variable and fractional frame rates from 0 to 120 Hz

Conclusion

H.264 – major leap forward towards scalable coding and multi-view capabilities

- Some controversy on patent licensing
  - Qualcomm owns patent on adaptive block size image compression and system
  - Qualcomm owns patent on interframe video encoding and decoding system
- Controversies around H.264 stem primarily from its use within HTML5 Internet standard and its use of video and audio.
  - Fight between Theora and H.264 as the Internet video format

Conclusion

Theora – free lossy video compression format

- Developed by Xiph.Org Foundation
- Distributed without licensing fees
- Goes with Vorbis audio format and the Ogg container
- Comparable in design and bitrate to MPEG-4 Part 2 (early version of Microsoft Media Video and RealVideo)

H.265

Derived from H.264

- More modes, tools and more interdependencies
- More efficient search algorithms
- More complex intra-prediction
- Macroblocks vs Partitions

H.265

AVC

16x16 macro-blocks

8x8 and 4x4 transform sizes

HEVC

Coding unit size 64x64 to 8x8

32x32, 16x16, 8x8 and 4x4 transformsizes