Upcoming events at Jigsaw24

Celebrating 25 years of creative design

London

EXPAND

Post-NAB open house

London

EXPAND

Jigsaw24's Tech Breakfasts

London

EXPAND

Tech Breakfast:
Video Compression 101

From macroblocks to GOPs

phil.crawley@jigsaw24.com
@IsItBroke on Twitter
http://www.root6.com/author/phil

Video Compression 101

- Compression - Uncompressed video and codecs
- DCT principles
- Intra / Inter-frame compression
- Acquisition vs. Editing vs. Transmission codecs
- Measuring compression effects
- Codecs vs. File Formats & wrappers
- Compression concatenation
- MPEG4, H.264 & H.265

This is by its nature an introduction – you can’t cover in a few hours what BBC training takes weeks over but it will give you a confidence in the basics to start investigating for yourself.

Compression - Uncompressed video and codecs

- The data rate of uncompressed high definition video is 1.48Gbits^{-1} and 3Gbits^{-1}
- UHD/4k comes in at 6 Gbits^{-1}, 12 Gbits^{-1}, 18 Gbits^{-1} etc.
- These data rates are far too high to easily save to a SAN, record on videotape or send over a network
- Using mathematical techniques the digital data that represents pixels – colour and luminance values – are transformed into a description that allows the pixels to be re-constituted and hence occupies much less space
- Depending on the application video can be compressed to 10% or less of its original size.
- The particular mathematical function used to achieve this is called a codec
- Different codecs have pros and cons depending on application

Early computer-based graphics and video formats

image: Wikipedia

With the exception of MPEG2 most early computer video formats tend to;

- Have square pixels (because computer monitors do)
- Use RGB for their colour representation
- Use varying framerates (from 12fps up)

None of these lend those early computer video systems to television!

The same can be said of computer still image formats – TIFF, Targa, BMP, etc – also;

- They may use CMYK colour space
- Graphics software may work in DPI rather than absolute resolutions
- Varying degrees and quality of anti-aliasing

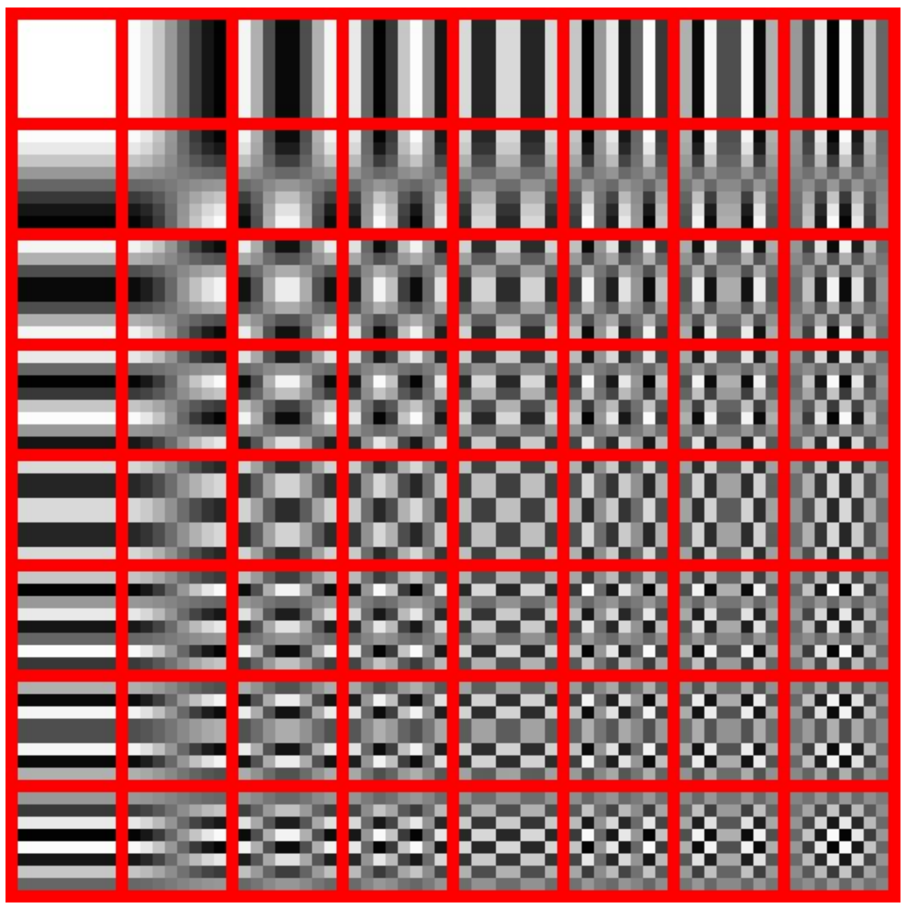
Video compression – how does that work?

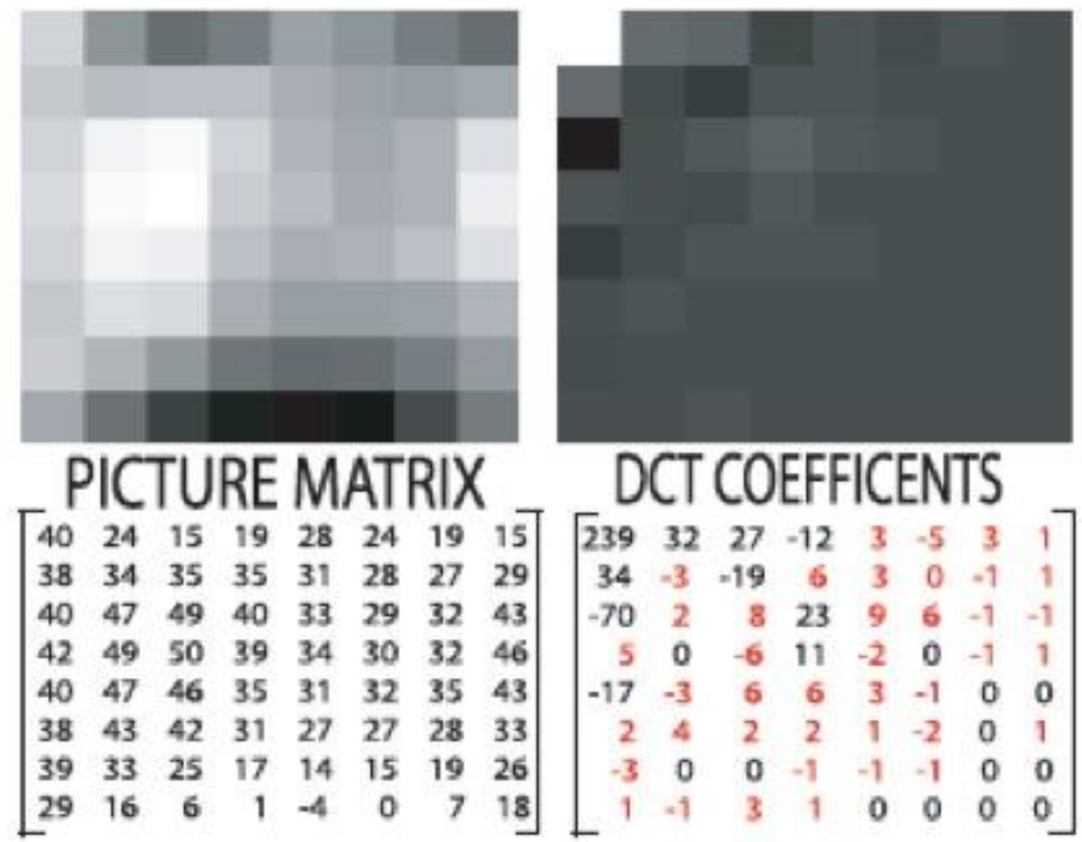
There are two aspects of video that we can exploit to reduce the data load;

- Large areas of similarity within each video frame; ***Intra*** *frame encoding*
- Large areas of similarity between video frames; ***Inter*** *frame encoding*

DCT principles

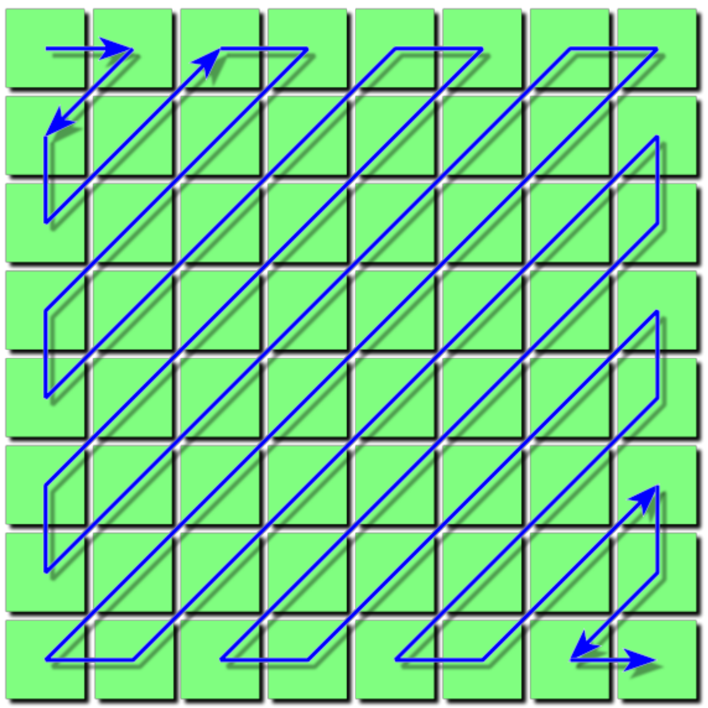
The Discrete Cosine Transformation is an example of a Fourier Transform (but using only real numbers). When applied to pixels (still and moving images) the picture information is transformed from pixel values into frequency values.





DCT principles

The Discrete Cosine Transformation doesn't actually reduce the data load – there are still 64 bytes of data required for an 8x8 macroblock, but looking at the DCT data shows that by 'zig-zagging' across the matrix allows us to use run-length encoding.



It's the run-length encoding that allows us to reduce the data rate. By the time the encoder gets to the bottom right of the matrix of the block there are long runs of similar values and these can be recorded as a value and a number of repetitions.

- If the block was all of one colour the 64-bytes might be reduced to two bytes.
- Even if this part of the picture is very complicated (lots of detail) quite high compression is possible.
- At some point though throwing away too much of the data results in pictures that don't come back nicely!

3, 15, 25, 50, 150:1

Compression ratios

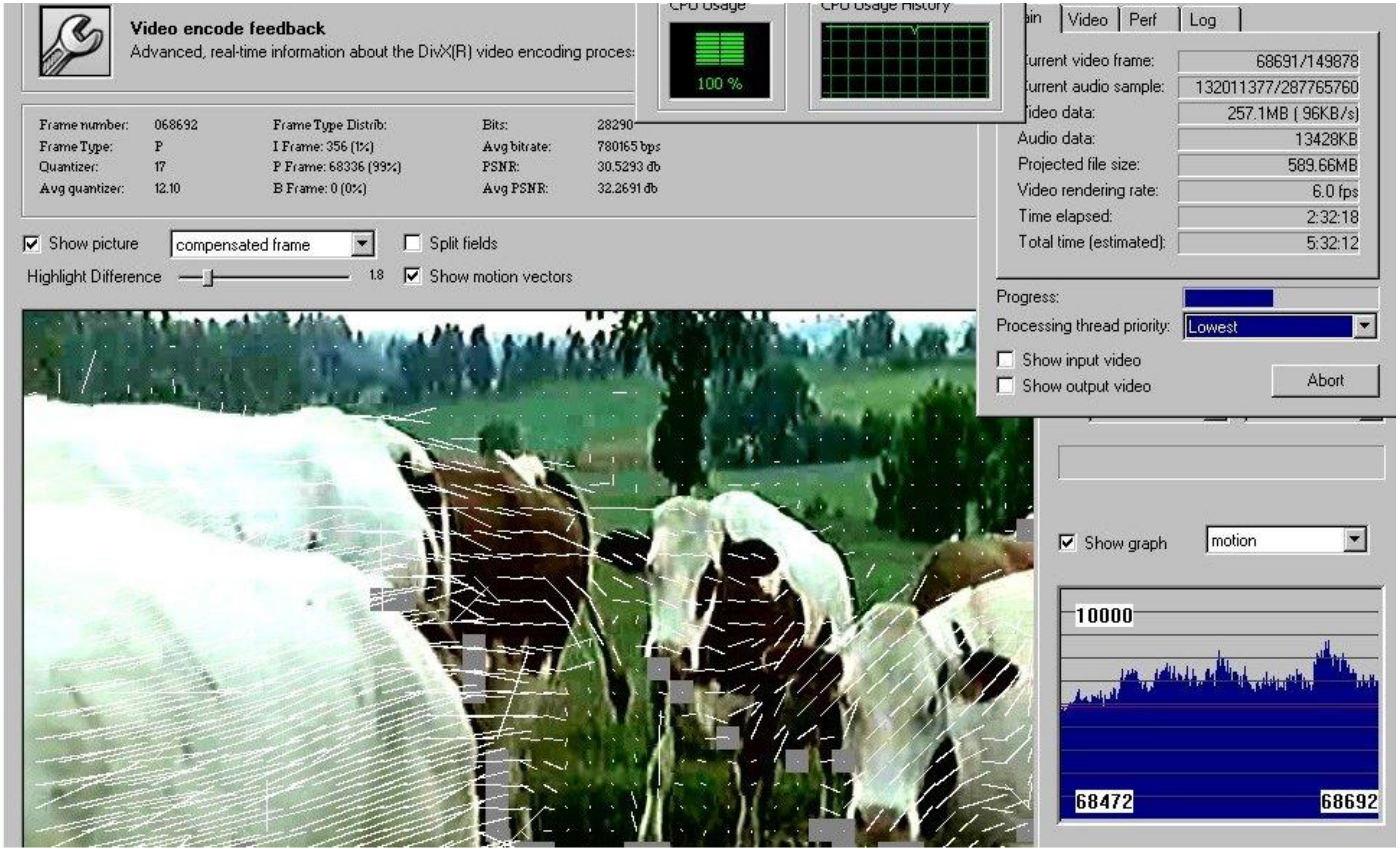
Inter-frame compression

- Applied to a sequence of video frames, rather than a single image.
- In general, relatively little changes from one video frame to the next.
- Inter-frame compression exploits the similarities between successive frames, known as temporal redundancy, to reduce the volume of data required to describe the sequence.

There are several inter-frame compression techniques, of various degrees of complexity, most of which attempt to more efficiently describe the sequence by reusing parts of frames the receiver already has, in order to construct new frames.

- The location of the similar or matching block in the past frame might be different from the location of the target block in the current frame.
- The relative difference in locations is known as the motion vector.
- If the target block and matching block are found at the same location in their respective frames then the motion vector that describes their difference is known as a zero vector.

Inter-frame compression



Inter-frame compression

When coding each block of the predicted frame;

- The motion vector detailing the position (in the past frame) of the target block's match is encoded in place of the target block itself
- Because fewer bits are required to code a motion vector than to code actual blocks, compression is achieved.

During decompression;

- The decoder uses the motion vectors to find the matching blocks in the past frame (which it has already received)
- It copies the matching blocks from the past frame into the appropriate positions in the approximation of the current frame, thus reconstructing the image.
- In general a perfect replica is not possible with block based motion compensation and thus the technique is lossy.

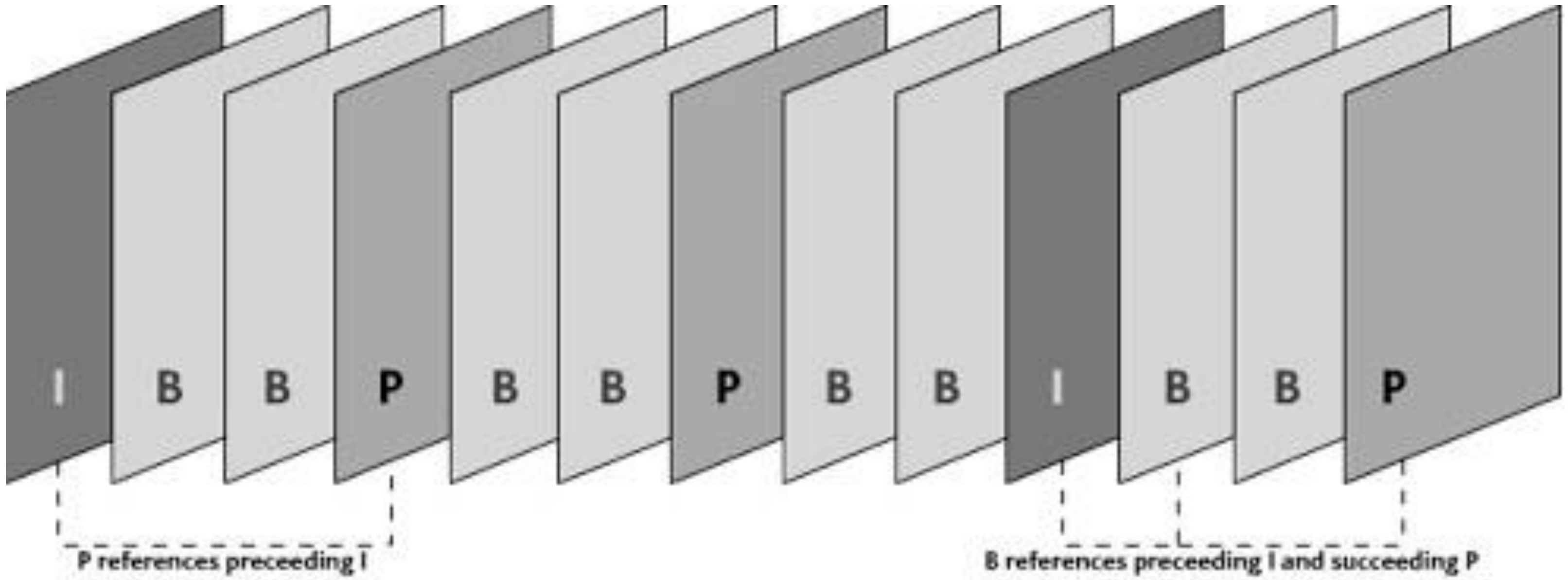
MPEG – Motion Picture Expert Group

- **MPEG-1**, 1993 – The basis for CDi and VideoCD (early DVD predecessors) – only 1.5Mbit/s and a quarter-screen resolution of 350x288 pixels. No interlaced video, 4x3 and stereo audio only. Long GOP
- **MPEG-2**, 1995 – Full standard def video, basis of DVD and DVB-T, C & S. Variable data rate and full 601 resolution using Y, Cr, Cb colour sampling. 16x9 or 4x3 video, up to 5.1 audio via AC3. 23.976, 25, and 29.97 FPS. Long GOP or i-frame (TX vs editing).
- **MPEG-4**, 1998 –The basis of most modern video encoding for acquisition, post and delivery. Multi-resolution, multi-framerate, multi-audio standards, editing or TX variants.
- **MPEG-4 part 10, AKA "AVC", AKA "H.264",** 2003 – further development which better optimises for HD video
- **MPEG-H part 2, AKA "HEVC", AKA "H.265"**, 2012 – further optimised for higher resolutions; up to 8k/UHD2-TV

Each of these formats comes with various "profiles" and "levels" which seek to standardise raster, GOP-length as well as data rate for specific applications.

Editing vs Transmission codecs

MPEG2 and all MPEG4 variants are based on a mathematical model of video called the **Discrete Cosine Transform** (DCT). Once applied to the video data this function then allows the codec to reduce the data rate of the video stream – you can transmit it and store more of it on disk.



Once the data in the video frames has had the DCT function applied the codec can also define different types of video frames that go to make up the **Group of Pictures** (GOP).

clip - MUXtrouble.mp4

Editing vs Transmission codecs cont.

- **I-frame:** An intra-frame, or I-frame, is a video frame which has been encoded without any reference to any other frame. A video file will always start with an I-frame and will have subsequent I-frames added at regular intervals. I-frames are also known as key-frames and are important for random access of video files such as rewind, fast-forward and seek operations. The downside to an I-frame is that they are the largest in terms of size as the whole video frame is encoded every time.
- **P-frame:** A predictive inter-frame, or P-frame uses previous I or P-frames as a reference when encoding. This means a P-frame will analyze a previous I or P-frame for any static elements which do not change between frames. Any areas which do not change are not encoded therefore a P-frame only stores video which registers movement making them much smaller than I-frames. The downside to P-frames is that they are sensitive to transmission errors because of their dependency on earlier frames.
- **B-frame:** A bi-predictive inter frame, or B-frame makes reference to both a preceding reference frame as well as a future reference frame. Using B-frames improves the prediction and ultimately the quality of decoded video but it also increases the processing requirements and latency.

Picture faults – GOP faults

In this frame the previous I-Frame has been corrupted and the decoder is unable to make a complete picture, only those elements that have arrived in new B & P-frames.

This problem is worse in MPEG4 as blocks may refer back to pictures from many seconds ago, unlike MPEG2 where there is always a new GOP started every 12-frames (15-frames in NTSC-land).

clip – mpeg4-fin.mp4

Measuring compression effects

“When you can measure what you are speaking about, and express it in numbers, you can know something about it; but when you cannot measure it, when you cannot express it in numbers, your knowledge is of a meagre and unsatisfactory kind: it may be the beginning of knowledge but you have scarcely, in your thoughts, advanced to the stage of science.”

William Thompson, Lord Kelvin.

The compression patch in the SRI-Visualizer allows you to derive a pair of numbers which give you a measure.

- Lines per picture height (resolution)
- Bit depth (dynamic range)

The Visualiser is good for many picture-based measurements, not just compression; we quantify picture quality using the PSNR scale measured in dBs PSNR.

Acquisition vs. Editing vs. Transmission codecs

The requirements of editing and transmission differ somewhat;

- Editing requires immediate access to each video frame and should not have to build a complete frame by looking at the frames that surround it, so an **I-Frame** only (or ‘short-GOP’ codec) is used.
- Transmission would rather leverage the additional image quality available to a long-GOP system and so a 12-frame (typical) GOP is used.

Rule of ten;

- Uncompressed high definition video ~ 1500Mbits/sec
- I-Frame editing codec, ProRes ~ 150Mbits/sec
- Long GOP transmission codec, MPEG4 ~ 15Mbits/sec
- Statistical Multiplexed DVB stream to the home ~ 9Mbits/sec

Codecs vs. File Formats & wrappers

- Avid DNX HD – I-frame only (i.e. editing) **codec**.
- Quicktime – Apple's wrapper format, especially **ProRes codec**.
- MXF –a 'universal' wrapper format that can encapsulate different codecs – also DPP.
- MPEG2 – the **codec** used in HDV, XDCam
- DV – I-Frame only **codec**, initially domestic cameras but extends up to HD (100Mbits^{-1})

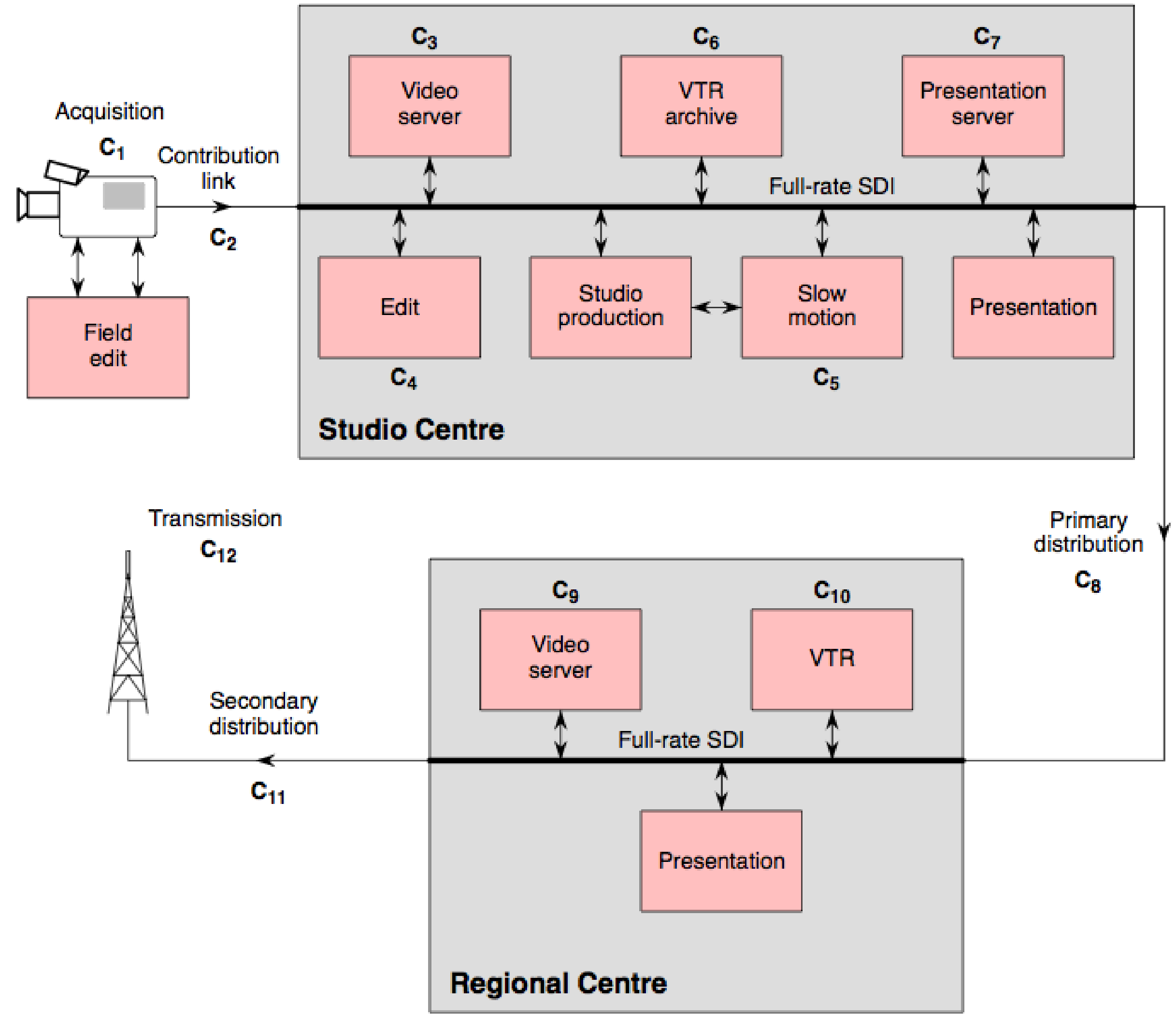
You have to distinguish between ***codecs*** (that mathematical function that changes raw pixels into a file-description of how to re-create the pixels; compression) and ***container*** formats (AKA wrapper).

All of these codecs "talk" to each other – best to avoid codec-concatenation.

Compression concatenation

The ideal video compressor will (given enough bandwidth) will be able to reduce the file or bitstream size to make it manageable but without introducing annoying changes to the picture. However, even the best encoders do leave some (hopefully invisible!) artefacts in the moving images.

The problem now is that different manufacturers and different codecs are all introducing picture and sound changes, some of which will interfere with each other producing truly terrible effects!



MPEG4 & H.264

- MPEG4 improves the performance of MPEG2 by allowing macroblocks to the referenced across I-Frame boundaries.
- MPEG4-part 10 (aka 'H.264' or AVC) further improves performance by using variable sized macroblocks and blocks that can 'look around' for similarity.
- It's hard to see how much further DCT-based codecs can go!

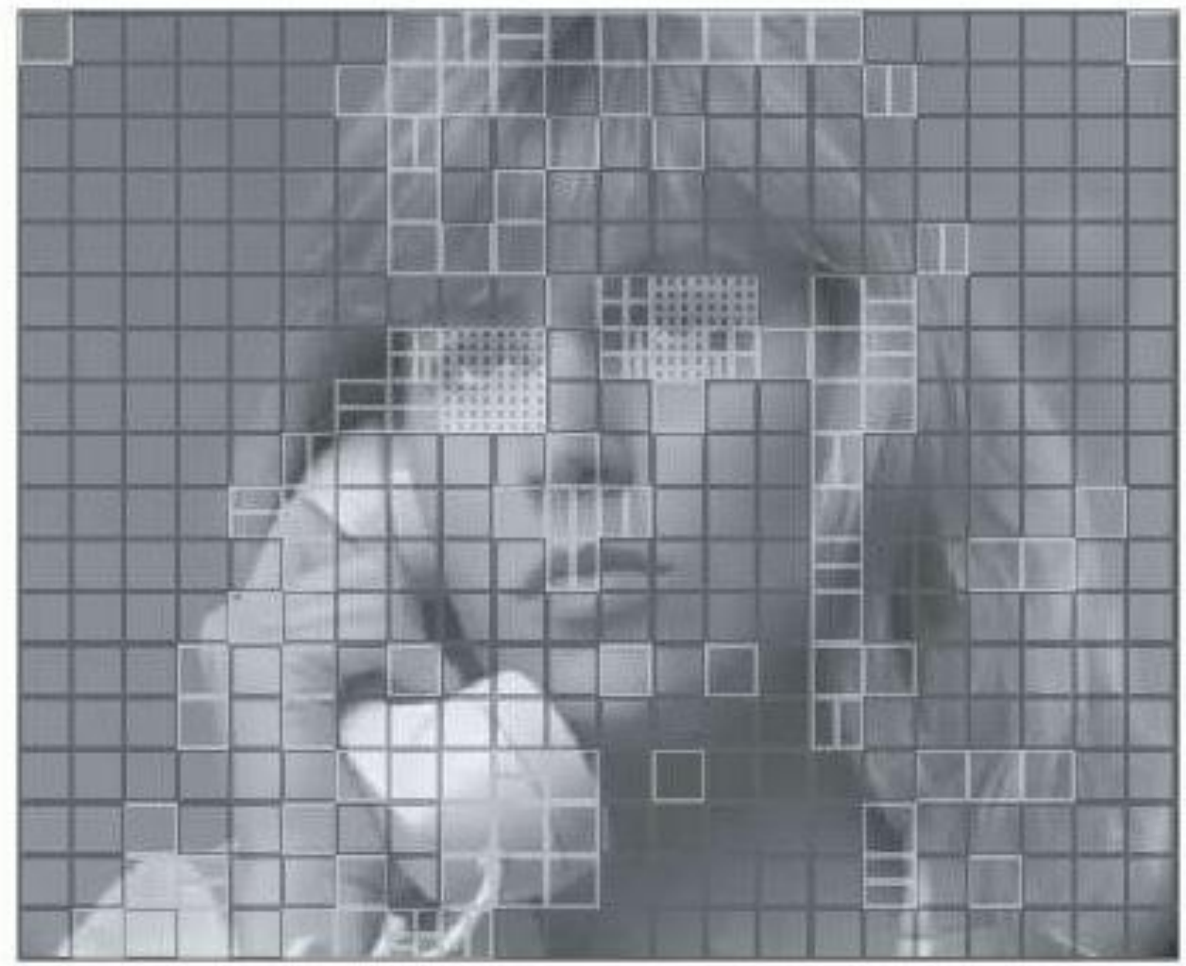
Complexity of H.264; two examples

- Intra-prediction - In the other video standards, in an 'Intra' frame, all the 16x16 Macroblocks are self-contained: they are not generated based upon Macroblocks in preceding or following frames, nor from other Macro-blocks in the same frame. However, H.264/AVC takes advantage of the fact that within a frame, Macroblocks which are located close to each other often have similar data – so that even within an 'Intra' frame it is possible to use prediction techniques where the value of one Macroblock is derived from the values of the video data in one of the surrounding Macroblocks.

- Smaller blocks for enhanced efficiency - H.264/AVC allows the 16x16 Macroblocks to be sub-divided down to blocks as small as 4x4, to enhance the compression efficiency. In MPEG-2 and MPEG-4, there is just one type of 'Intra' Macroblock, containing compressed video which does not refer to any other Macroblock.

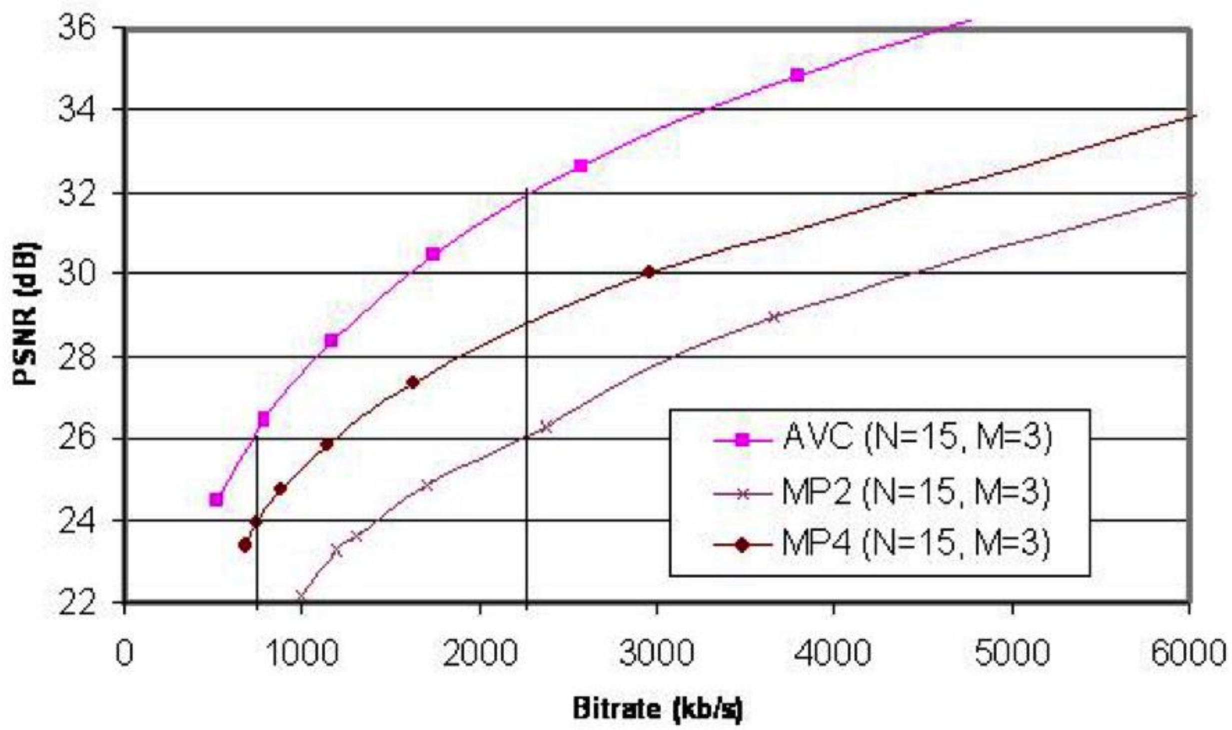
H.264 improvements

However, with the complexities above (and others), in H.264/AVC there are 26 types of Intra Macroblock. There are many other complex elements in H.264/AVC, such as the ‘CABAC’ Entropy coding, where bits are shared such that fractions of 1 bit effectively represent a specific syntax element.



Qualitative comparison of MPEG2, MPEG4 & H.264

Video reams at the same resolution, encoded three times.



HEVC – aka H.265, the new kid on the block?

High Efficiency Video Coding (HEVC) is a video compression standard, a successor to H.264/MPEG-4 AVC (Advanced Video Coding), that was jointly developed by the ISO/IEC Moving Picture Experts Group (MPEG) and ITU-T Video Coding Experts Group (VCEG) - the specification was formally ratified as a standard on April 13, 2013.

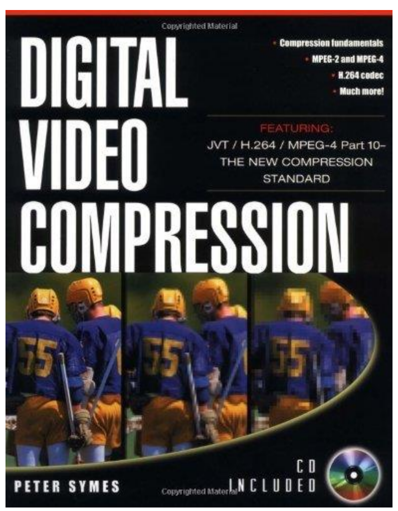
- Double the data compression ratio compared to H.264/MPEG-4 AVC at same level of video quality.
- Used to provide substantially improved video quality at the same bit rate.
- It can support 8K UHD and resolutions up to 8192x4320

Increasing number of products supporting HEVC as well as OTT services;

- Online streamed video – the only way we'll be able to watch 4K Netflix!
- 4K and 8K UHD-TV production; HEVC also supports optimisations for HLG for high dynamic range
- Ultra BlueRay disks

Further afield and further reading

- DCI packages are encoded using JPEG-2000; a short-GOP, Wavelet-based codec; https://en.wikipedia.org/wiki/Wavelet
- Current experimentation is around object-based codecs; primitives such as objects, shaders and vectors describe how objects move in each frame – this gives rise to frameless & pixel-less codecs which can be rendered at arbitrary rasters and framerates without loss of quality; http://opus.bath.ac.uk/43943/



"Digital Video Compression" – Peter Symes.