NUMÉRIQUE ET CONSUMATION ENERGÉTIQUE

CODAGE VIDEO

Mihai MITREA, Département ARTEMIS
mihai.mitrea@telecom-sudparis.eu
Thanks to my PhD student: Mohamed ALLOUCHE

OUTLINE

VIDEO ENCODING PRINCIPLES

WHY VIDEO ENCODING?

How many bits for 1 sec of (uncompressed) 4K video?

1. 1 frame of 4K video:
    - According to DCI: 4096 × 2160 pixels
    - According to SMPT: 3840 × 2160 pixels
2. A pixel:
    - Between 24 and 48
3. A second of video:
    - 30, 60, 120, 144 frames
4. **The total: between 11Gb/s and 61Gb/s**

VIDEO ENCODING PRINCIPLES

WHAT IS VIDEO ENCODING?

Creating a derived representation that reduces the size of the digital representation of the video content without affecting the content to be displayed on the screen

1. Lossless compression:
    - once decoded, the content is ***identical*** with the original one

2. Lossy compression:
    - once decoded, the content is ***visually identical*** with the original one

VIDEO ENCODING PRINCIPLES

WHAT PRINCIPLES?

Lossless video encoding principle: Shannon's information theory

1. Entropy and redundancy for a discrete, ergodic information source
   - Entropy: the average information bought by a symbol
   - Redundancy: the difference between the maximal possible entropy and the the actual entropy of the source
2. Source coding theorem (Shannon's first theorem)
   - The redundancy can be eliminated through coding

     or, equivalently,
   - Any discrete, ergodic information source can be transformed into a zero-redundancy information source by coding its symbols

Disclaimer: these are only principles, the topic itself requires at least 12h of lectures

VIDEO ENCODING PRINCIPLES

WHAT PRINCIPLES?

Lossless video encoding principle: Shannon's information theory

1. Example (1/2): lossless coding for a binary information source $X = \{x_1, x_2\}$, with $p(x_1)=1/3$
   - Example of symbols sampled from this source:
     - $x_2, x_1, x_2, x_2, x_1, x_2, x_2, x_2, \ldots$
   - Naïve code: $\{x_1\} = 0$, $\{x_2\} = 1$ => average length of the code: 1bit/symbol
   - Entropy: $H(X) = - p(x_1)\log_2 p(x_1) - p(x_2)\log_2 p(x_2) = 0.92$ bit/symbol => the naïve code is not optimal
   - Redundancy: max $H(X) - H(X) = \log_2(2) - H(X) = 0.08$

VIDEO ENCODING PRINCIPLES

WHAT PRINCIPLES?

Lossless video encoding principle: Shannon's information theory

1. Example (2/2): lossless coding for a binary information source $X = \{x_1, x_2\}$, with $p(x_1)=1/3$
   - Coding group of 2 symbols generated by the X source
     - $x_2, x_1, x_2, x_2, x_1, x_2, x_2, x_2, \ldots$
     - $\mathbf{x_2\, x_1}, \mathbf{x_2\, x_2}, \mathbf{x_1\, x_2}, \mathbf{x_2\, x_2}, \ldots$ where $p(\boldsymbol{x_1\, x_1})=1/9$, $p(\boldsymbol{x_1\, x_2})=2/9$, $p(\boldsymbol{x_2\, x_1})=2/9$, $p(\boldsymbol{x_2\, x_2})=4/9$
   - Finding a new code for these groups of symbols, where more frequent symbols are associated shorter codelengths
     - Example: $\{x_1\, x_1\}$ **= 110,** $\{x_1\, x_2\}$ **= 111,** $\{x_2\, x_1\}$ **= 10,** $\{x_2\, x_2\}$ **= 0**
     - Average length of the code = (3*1/9+3*2/9+2*2/9+1*4/9)/2=0.94
   - Experimental conclusion: when trying to reach the $H(X)$ = 0.92 limit:
     - Coding individual symbols from X => average length = 1
     - Coding groups of two symbols from X => average length = 0.94

VIDEO ENCODING PRINCIPLES

WHAT PRINCIPLES?

Lossless video encoding principle: Shannon's information theory

1. Conclusion
   - The $H(X)$ theoretical limit can be achieved by coding groups of n symbols, $n \to \infty$:
     - Perfect compression can be achieved assuming infinite memory/computing resources
   - The Shannon's model implicitly assumes that the mathematical model of the X information source exists and it is known:
     - What is ergodicity in practice?
     - Who tells you what are the theoretical probabilities?

- **Practical lossless coding scheme are implicitly sub-optimal and they must reach a trade off between the coding performances for the the memory/storage resources**

VIDEO ENCODING PRINCIPLES

WHAT PRINCIPLES?

Lossless video encoding principle: practical software tools with performances depending on the solution and on the information source (just for illustration)

1. Gzip: approximative compression ratio for text between 2.5 and 3 for text

2. PNG: approximative compression ratio for images between 5 and 13 => for 1s of 4K video would result in more than 1G compressed content

Conclusion: state of the art lossless data compression cannot deal with video volumetry

VIDEO ENCODING PRINCIPLES

LOSSY COMPRESSION

Reconsidering and extending the Shannon's coding principles (yet, exiting the proper theoretical framework)

1. Working assumption 1: Shannon's redundancy relates to what it can be guess, to what it can be a priori known about the content:
   - there are similitudes among areas in a same frame
   - there are similitudes among successive frames in a video sequence
2. Working assumption 2: By coding groups of symbols, the encoding gains in efficiency
3. Working assumption 3: Visual redundancy – there are several (theoretically unlimited number of) images that are digitally different yet identical for the human eyes

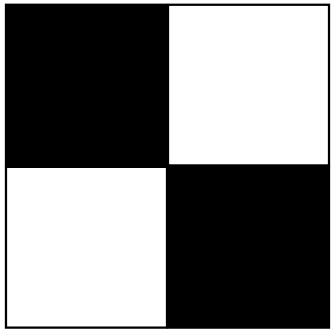
Disclaimer: modeling video sequences as natural information sources is out of the scope of this lecture, yet what an amazing research field!

VIDEO ENCODING PRINCIPLES

LOSSY COMPRESSION

Basic intuitive examples (1/2):

Example 1: *Working assumption 1 and 2*



512x512 pixels, pixels having only 0 value (black) and 255 (white): 512x512x8 bits

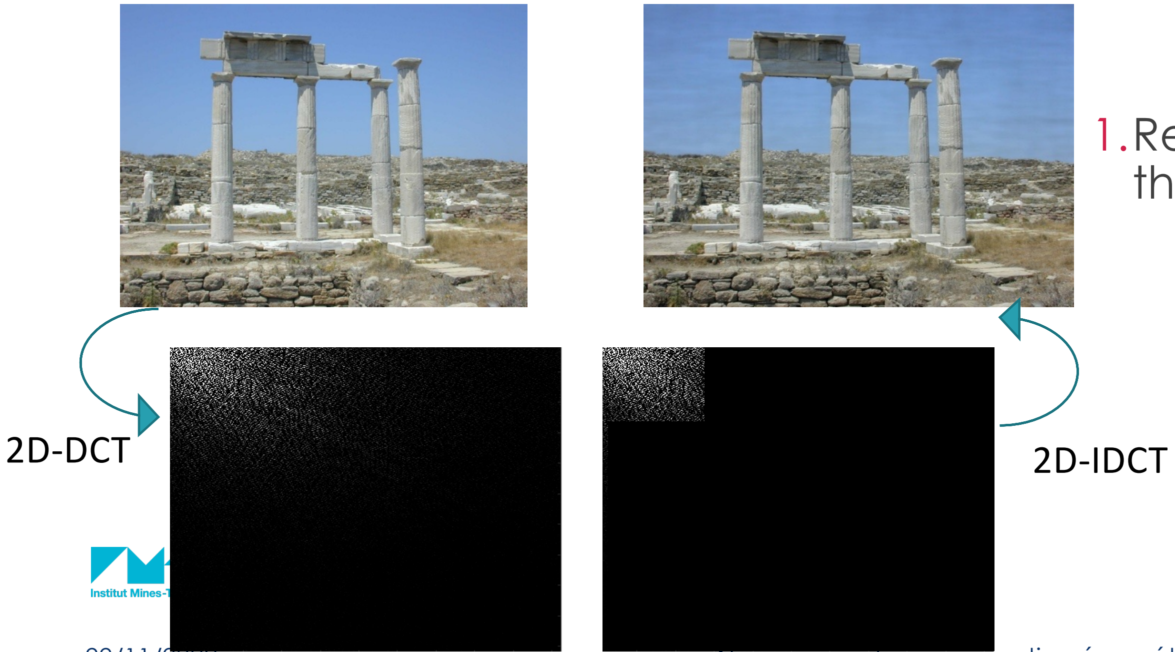
1. Reduce the number of bits for representing the image by:
   - **Identifying blocks with similar values**
   - **Identifying relations between blocks**

VIDEO ENCODING PRINCIPLES

LOSSY COMPRESSION

Basic intuitive examples (2/2):

Example 2: *Working assumption 3*



1. Reduce the number of bits for representing the image by ***Working assumption 3***
   - **Use transformed representations (example: 2D-Discrete Cosine Transform)**
   - **Reduce the dynamic of the values and/or quantization (example: ¾ of the 2D-DCT coefficients are set to 0)**

VIDEO ENCODING PRINCIPLES

LOSSY COMPRESSION

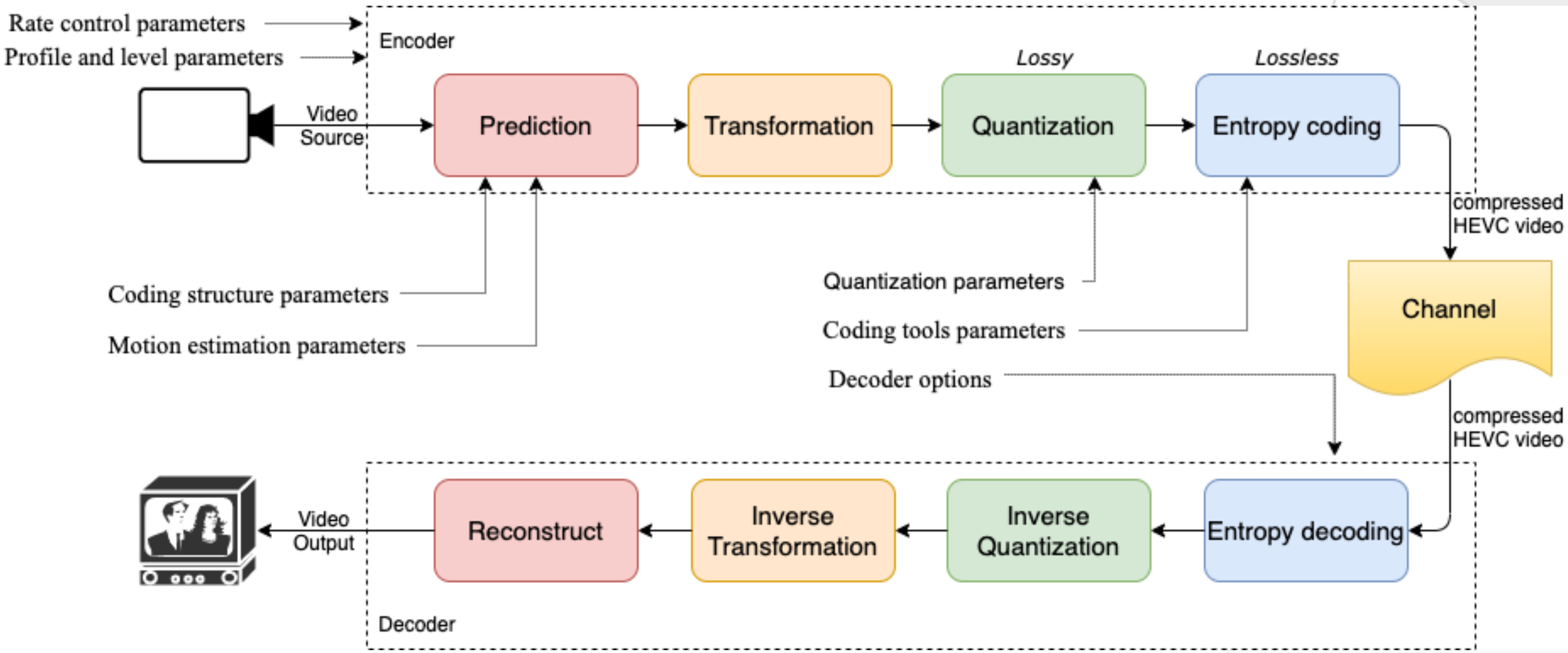
Intuitive conclusions: the performances of image/video encoding can be increased by:

1. Exploiting the three preceding working assumptions
   - The image content can be partitioned into blocks and some blocks can be **predicted** from their neighbors
   - By using orthogonal **transforms**, information relevant for the Human Visual System can be compacted on a smaller number of coefficients
   - The dynamics of the values can be reduced by quantization-like operations
2. Lossless coding can also be considered on the top of this

OUTLINE

CONVENTIONAL VIDEO ENCODERS

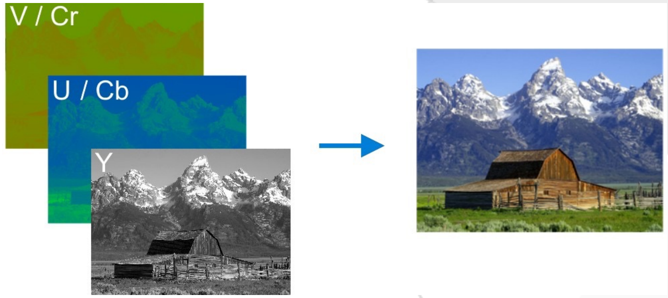
OVERVIEW



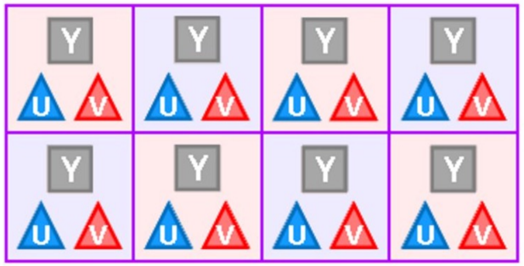
CONVENTIONAL VIDEO ENCODERS

PIXEL REPRESNETATION - YC_BC_R

1. Y = Luma (after nonlinear Gamma correction)
2. C_b = blue
3. C_r = red

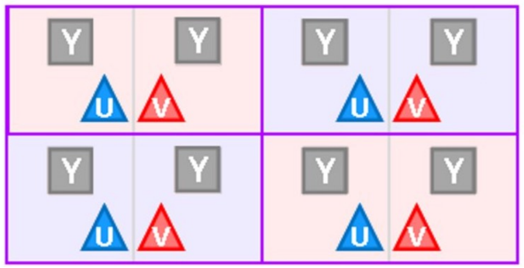


4:4:4



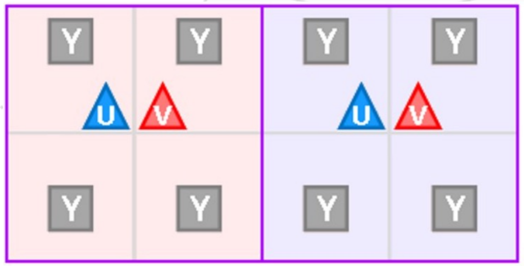
- Raw format (no compression)
- No gains compared to RGB

4:2:2



- 1 C_b color will be used to render 2 pixels (same for C_r)
- 33% gains

4:2:0



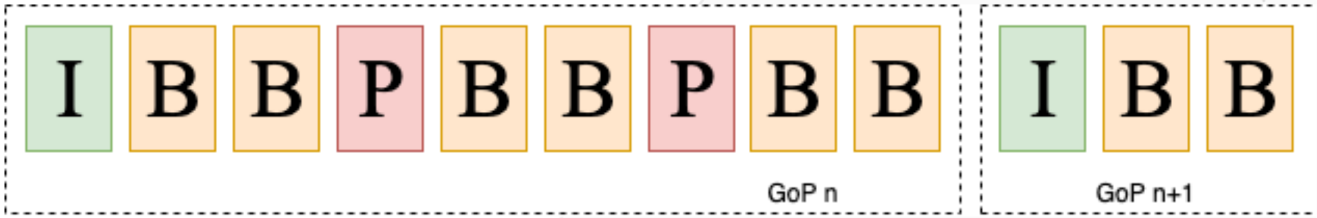
- The mainstream subsampling (TV, Internet, Blu-ray)
- 50% gains

CONVENTIONAL VIDEO ENCODERS

GOP - GROUPS OF PICTURES

The GOP is the coding unity:

1. **I** frames
   - Intra prediction frames
   - always the first frame in a GOP
2. **P** and **B** frames
   - Inter (frame) prediction

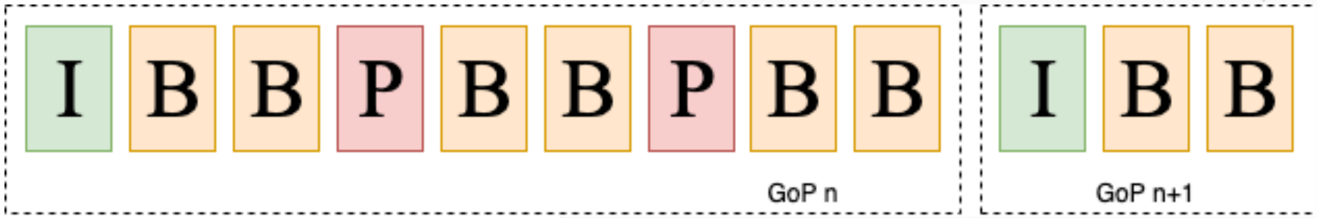


CONVENTIONAL VIDEO ENCODERS

GOP - GROUPS OF PICTURES

I frames:

1. Encoded independently from others
   - All encoding information is extracted at the frame level
   - Virtually acts as an image encoding algor
2. Ensures error robustness
3. Lower complexity but also lower compression efficiency

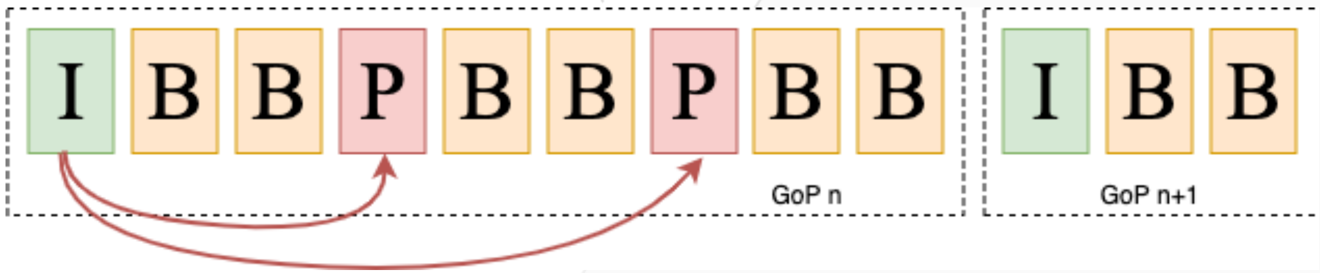


CONVENTIONAL VIDEO ENCODERS

GOP - GROUPS OF PICTURES

P frames:

1. Encoded from previous I and P frames (in the same GoP)
2. Higher complexity but also higher compression efficiency

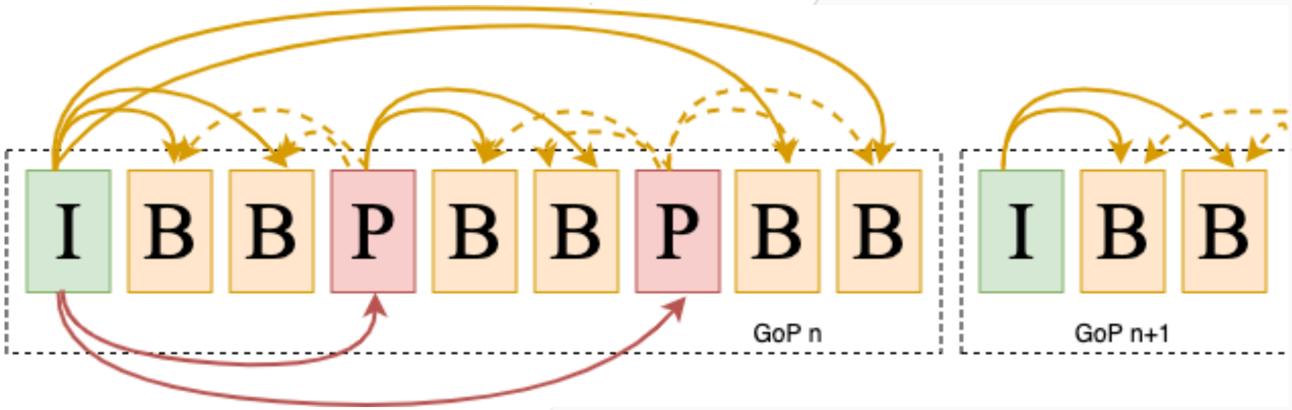


CONVENTIONAL VIDEO ENCODERS

GOP - GROUPS OF PICTURES

B frames:

1. Encoded from:
   - previous I frames
   - previous and succeding P frames (in the same GoP)
2. Highest complexity but also highest compression efficiency



CONVENTIONAL VIDEO ENCODERS

MACROBLOCKS

Principle (MPEG-4 AVC - a.k.a. H.264):

1. Define areas in the image that are expected to have homogeneous behaviour:
   - 4x4 pixels
   - 16 x 16 pixels

CONVENTIONAL VIDEO ENCODERS

MACROBLOCKS & CTU (CODING TREE UNITS)

Principle (HEVC - a.k.a. H.265):

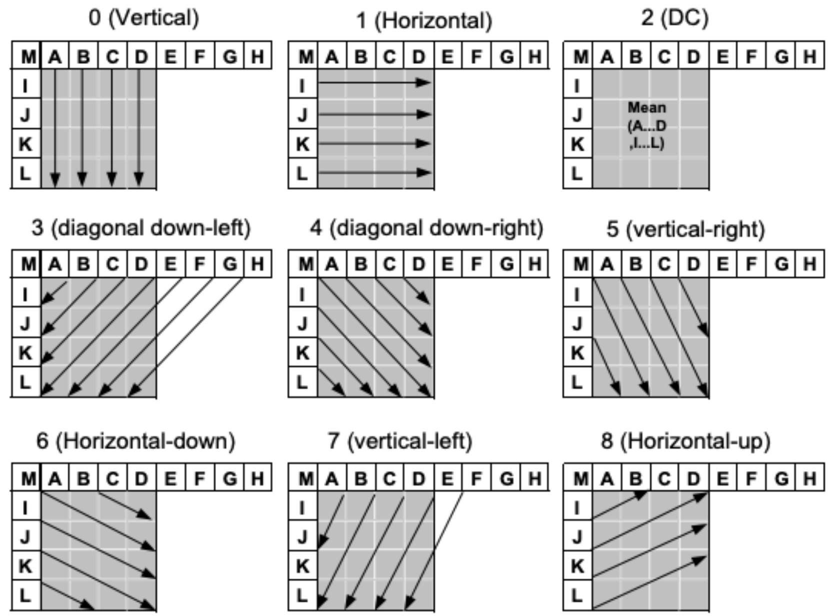
1. Define areas in the image that are expected to have homogeneous behaviour:
    - The CTU is the basic unit of coding and can be up to 64x64 pixels
    - CTU can be between 16×16 pixels and 64×64 pixels
    - A Coding Tree Unit can be subdivided by dichotomy regions known as Coding Units (CUs) using a quadtree structure

CONVENTIONAL VIDEO ENCODERS

PREDICTION MODES – MPEG-4 AVC

Principle: information in a macroblock can be computed from information in preceding blocks and from a *predictiuon error*

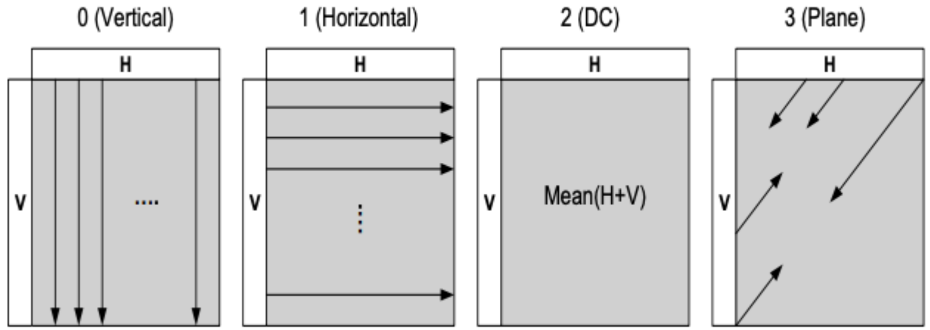
INTRA 4×4



Intra prediction modes

1. For intra prediction includes two different block sizes:
   - Intra4×4 performed for each 4x4 block and containing 9 directional prediction modes
   - Intra16×16 applied for each 16x16 MB and including 4 directional prediction modes

INTRA 16X16

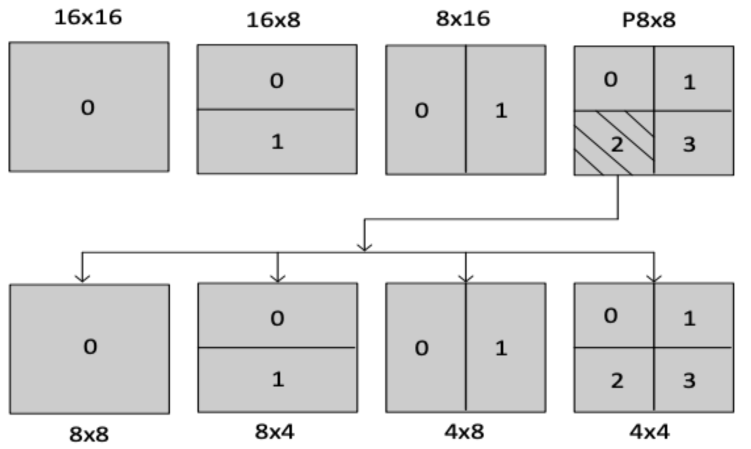


CONVENTIONAL VIDEO ENCODERS

PREDICTION MODES – MPEG-4 AVC

Inter prediction modes

- For inter prediction, there are seven different block sizes: 16×16, 8×16, 16×8, P8x8 {8×8, 8×4, 4×8 and 4×4}

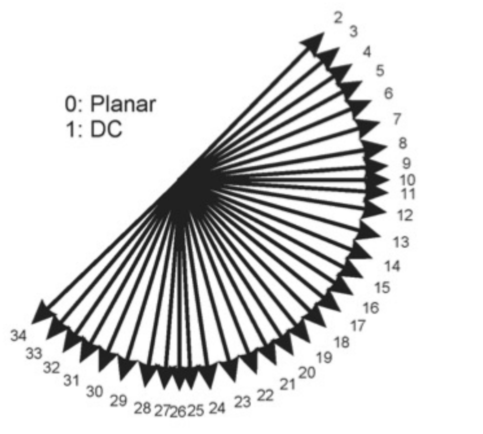


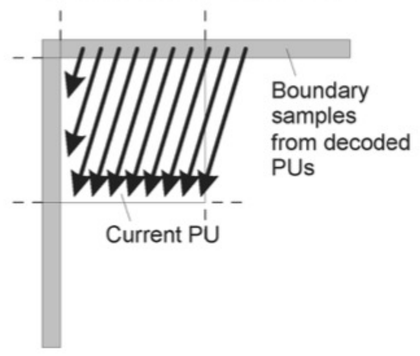
CONVENTIONAL VIDEO ENCODERS

PREDICTION MODES – HEVC

Intra prediction modes

1. Each CU is predicted from neighbouring data in the same frame
    1. Using planar prediction (fitting a plane surface to the CU)
    2. Using DC prediction (an average value for the CU)
    3. Using directional prediction (extrapolating from neighbouring data)





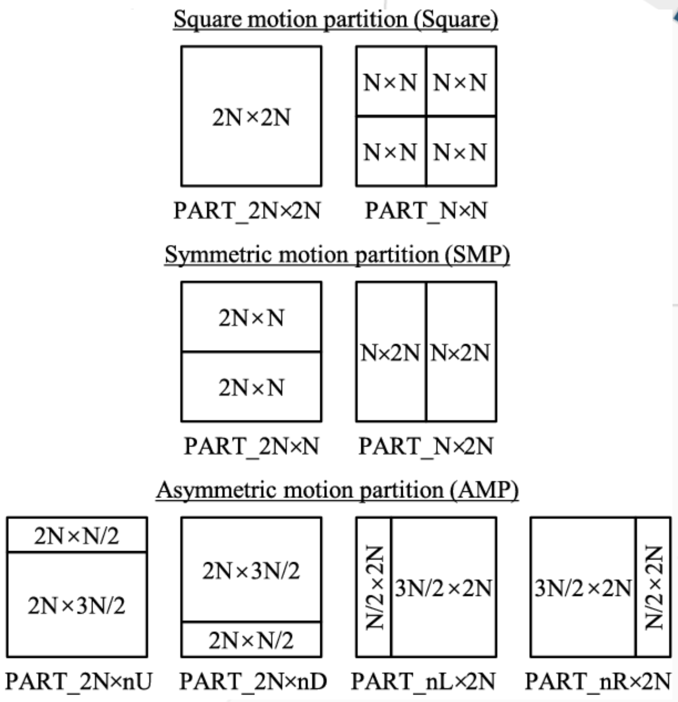
CONVENTIONAL VIDEO ENCODERS

PREDICTION MODES – HEVC

Inter prediction modes

1. Each CU is predicted from image data in one or two reference pictures

2. Motion compensated prediction is used



CONVENTIONAL VIDEO ENCODERS

TRANSFORMATIONS – MPEG-4 AVC

Principle: representing the prediction errors on an orthonormal basis

1. Integer-coefficient DCT approximation:
   - Encoder and decoder are always synchronized (no precision errors)
   - Coefficients: 0, ±1, ±2 → implemented by bit-shift and sum or subtraction

2. Two transform levels:
   - 4x4 transform over the 16 DC coefficients of an Intra16 block
   - 2x2 transform over the 4 DC coefficients of Chroma blocks

CONVENTIONAL VIDEO ENCODERS

TRANSFORMATIONS – HEVC

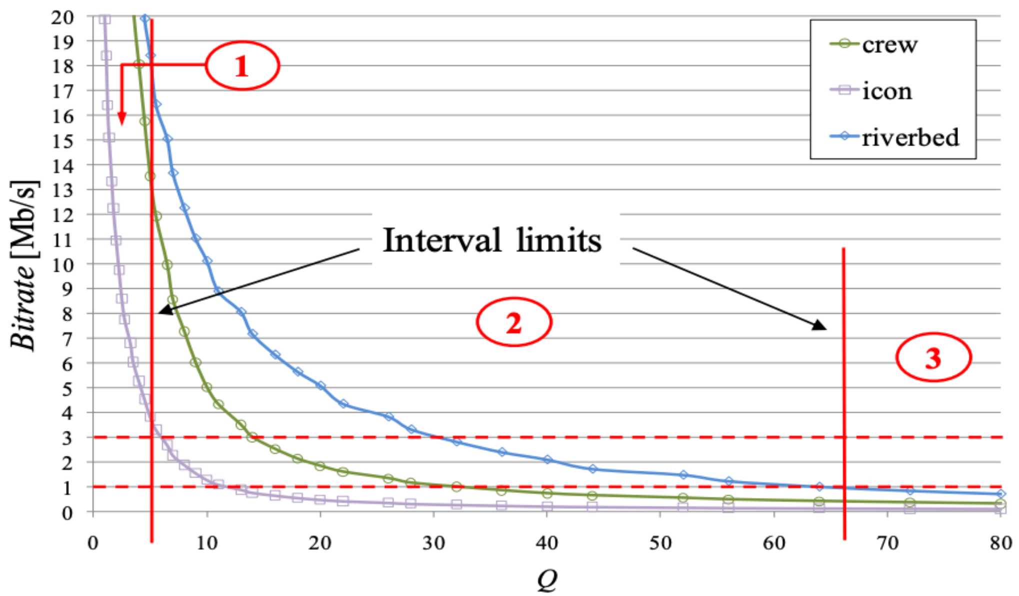
Principle: representing the prediction errors on an orthonormal basis

1. DCT (Discrete Cosine Transform) and DST (Discrete Sine Transform) transforms can be used
    - Various transform sizes are possible : 32x32, 16x16, 8x8, and 4x4
    - Square transforms only
        - Intra: DCT + DST
        - Inter: DCT only

CONVENTIONAL VIDEO ENCODERS

QUANTIZING

Principle: adapt the bitrate as a function of a quantizer over the transform results



CONVENTIONAL VIDEO ENCODERS

ENTROPY CODING

Principle: adding a final binary encoding

1. Two techniques are supported
   - A low-complexity method, based on variable length coding, with a dictionary depending on the context (CAVLC)
     - Gain :2–7% rate reduction
     - Available only for MPEG-4 AVC
   - A high-complexity method, using context adaptive arithmetic coding (CABAC)
     - Gain :7–20% rate reduction
     - Available for both MPEG-4 AVC and HEVC

OUTLINE

CONCLUSION

Video encoding at work: between MPEG-4 AVC and HEVC

	AVC (H.264)	HEVC (H.265)
Approved date	2003	2013
Improvements	- 40-50% bit rate reduction compared with MPEG-2 - Available to deliver HD sources for Broadcast and Online	- 40-50% bit rate reduction compared with H.264 at the same visual quality - implement Ultra HD, 2K, 4K for Broadcast and Online
Maximal support	Up to 4K	Up to 8K
Partition sizes	Macroblocks 16x16	Coding Units 8x8 to 64x64
Partitioning	Sub-block down to 4x4	Prediction Unit Quadtree down to 4x4 square, symmetric and asymmetric (only square for intra)
Intra prediction modes	13 modes	35 modes
Entropy coding	CABAC, CAVLC	CABAC

CONCLUSION

Video encoding is a complex task

1. Theoretical solution exists but requires “infinite” complexity
2. Practical solutions exist but requires a combinatory complexity among:
   - GOP partitioning
   - computing Intra/Inter prediction modes
   - computing DCT/DST transform at the block level
   - dynamically adjusting the quantizing step
   - binary encoding

CONCLUSION

Illustrating video encoding performances

1. Example 1: The role of parameters (MPEG-4 AVC)
   - Comparison original video *vs*. automatic encoded video *vs*. I frames only video
   - Comparison on I frame only with variable number of frames per sec (30 *vs*. 10 *vs*. 5)

CONCLUSION

Illustrating video encoding performances

1. Example 2: The role of encoding technology
   - Comparison original video *vs*. MPEG-4 AVC *vs*. HEVC

CONCLUSION

Illustrating video encoding performances

1. Example 3: The role of original content
   - Comparison original video *vs*. MPEG-4 AVC *vs*. HEVC for a new type of content

CONCLUSION

How to turn green the video encoding?

1. No theoretical (general) solution, as the actual solution depend on the targeted application:
   - reduce computation during encoding can lead to more bandwidth consumption (during transmission / storage)
   - reducing the bandwidth would a priori increase the computation during encoding (sometimes even during decoding)
   - there is a hidden energy consumption during the displaying porcess
2. Come tomorrow for the VIDMIZER industrial conference

REFERENCES

- https://techjury.net/blog/video-consumption-statistics/#gref
- https://webtribunal.net/blog/video-consumption-statistics/#gref
- https://thesocialshepherd.com/blog/video-marketing-statistics
- https://2014.jsconf.eu/speakers/raul-fraile-how-gzip-compression-works.html

- N. Bahri, et al. 'Fast Intra Mode Decision Algorithm for H264/AVC HD Baseline Profile Encoder'. International Journal of Computer Applications in Technology, Inderscience, 2012, 37 (6), pp.8 - 13. ⟨10.5120/4610-6601⟩. ⟨hal-01797174⟩
- G. J. Sullivan, J. -R. Ohm, W. -J. Han and T. Wiegand, "Overview of the High Efficiency Video Coding (HEVC) Standard, » in IEEE Transactions on Circuits and Systems for Video Technology pp. 1649-1668,. 2012, doi: 10.1109/TCSVT.2012.2221191.
- T. Grajek, M. Domański, "Single Frame Rate-Quantization Model for MPEG-4 AVC/H.264 Video Encoders", September 2010, Computer Vision and Graphics - International Conference, ICCVG 2010, Warsaw, Poland, September 20-22, 2010, Proceedings, Part I