Transform Video Compression summary

Prof. Ioannis Pitas
Aristotle University of Thessaloniki
pitas@csd.auth.gr
www.aiia.csd.auth.gr

Version 3.1

Artificial Intelligence & Information Analysis Lab

Transform Video Compression

- **Video compression**
- Intraframe video coding
- Interframe video coding
- Transform Video Coding
- Predictive coding
- MPEG2
- MPEG4

Video compression

Video compression facilitates:

- Handling and storage of high resolution video
- Video transmission over computer networks
- TV broadcasting

Application areas:

- Digital television
- Video conferencing
- Video streaming
- Digital Cinema
- Distance learning

Video compression

Use of inherent ***spatiotemporal video redundancy.***

- If we compress each frame **seperately** (as an image), we only employ spatial redundancy within the frame
- Prediction of current blocks of frames $f(\mathbf{n}, t) = f(n_1, n_2, t)$ from previous (or future) video frame blocks $f(\mathbf{n}, t - l)$ employs temporal redundancy.
- Compression of displaced frame difference (assumed to be small).

Two operation modes:

1. Intraframe coding.
2. Interframe coding.

Transform Video Compression

- Video compression
- **Intraframe video coding**
- Interframe video coding
- Transform Video Coding
- Predictive coding
- MPEG2
- MPEG4

Intraframe video coding

In ***intraframe video coding***, Video frame $f(\mathbf{n},t)$ coding does not take input from video other frames.

- $f(\mathbf{n},t)$ is transformed using ***Discrete Cosine Transform*** (***DCT***).

DCT coefficients are:

- Quantized and
- VLC encoded.

The video frame is:

- compressed and transmitted and received by decoder

The decoder decodes compressed frame and produces $\hat{f}(\mathbf{n},t)$.

Intraframe video coding

An ***I-frame*** is a fully intra-encoded video frame.

- They are used periodically to stop decompression error propagation.
- Very useful for quick video browsing.
- First video frame always encoded as an I-frame.

Transform Video Compression

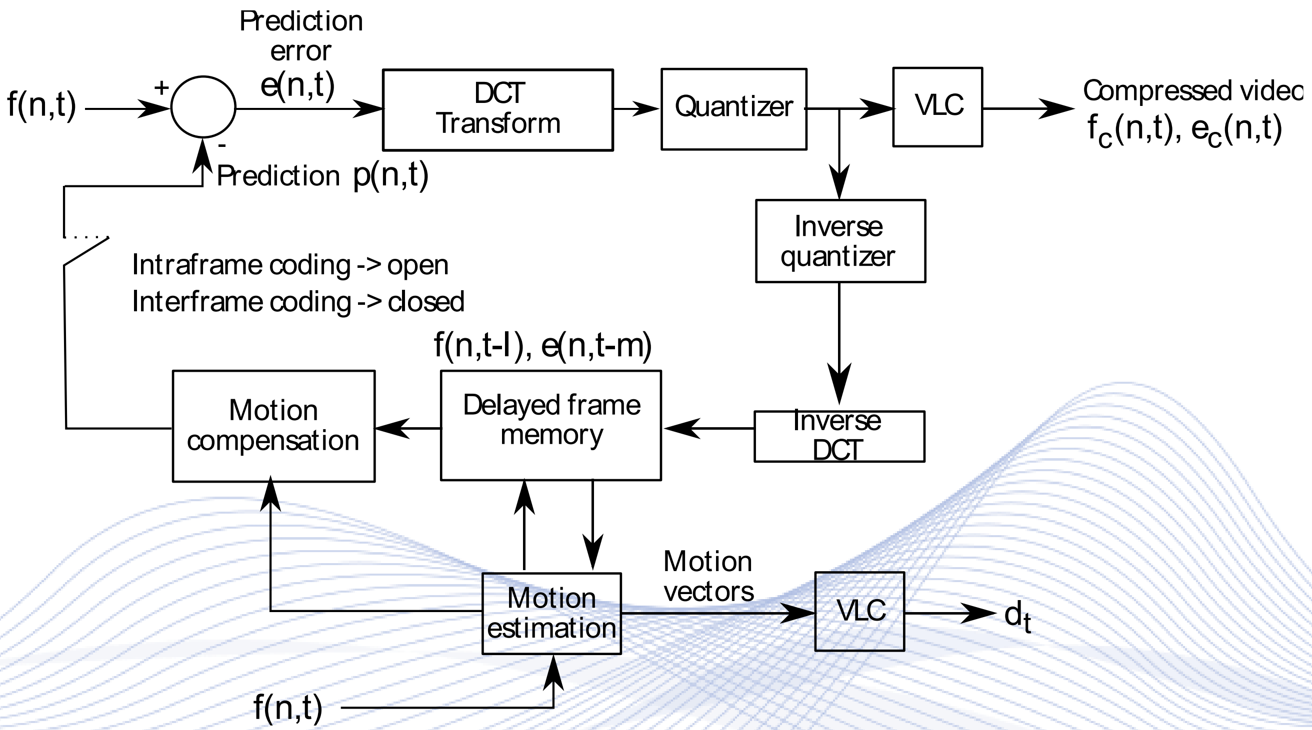
- Video compression
- Intraframe video coding
- **Interframe video coding**
- Transform Video Coding
- Predictive coding
- MPEG2
- MPEG4

Interframe video coding

Predicted video frames:

- ***P-frames*** (forward prediction).
  - Prediction from previous video frame.
  - Reference frame encoded using intraframe transformation or prediction from another frame.
- ***B-frames*** (bidirectional prediction).
  - Bidirectional prediction from previous and subsequent video frames.
- Both encoding methods employ block matching.

Interframe video coding



General system for Transform Video Compression.

Interframe video coding

Forward prediction predicts video ***P-frame*** pixel values based on corresponding pixel in a previous frame.

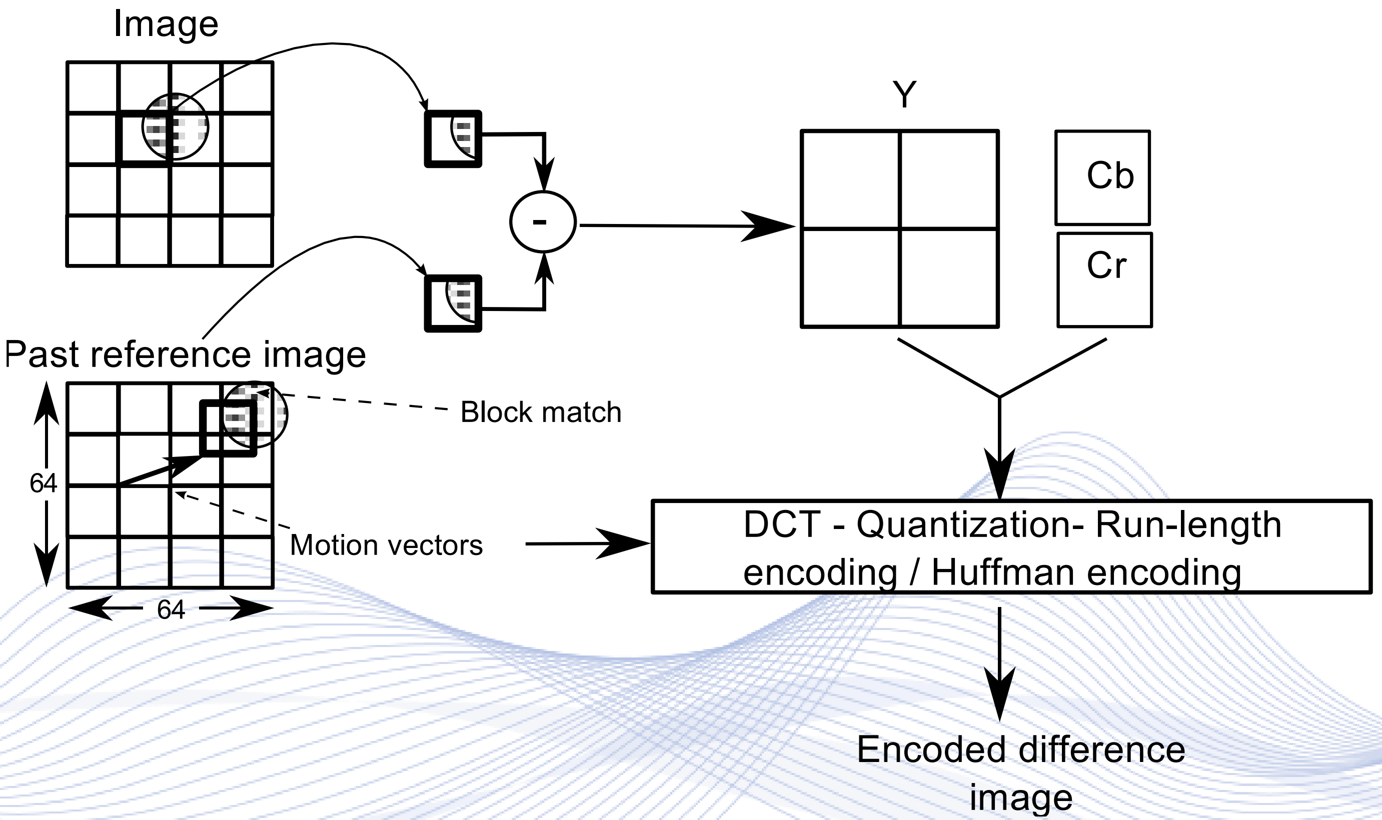
- Object from a video moves between frames - Pixel values change

Motion compensated prediction:

$$p(\mathbf{n}, t) = f(\mathbf{n} + \mathbf{d}_t, t - 1).$$

- $f(\mathbf{n}, t)$: pixel luminance in location $\mathbf{n} = (n_1, n_2)$
- $\mathbf{d}_t = [dx, dy]^T$: motion vector of a pixel from frame $t-1$ (***reference frame***) to frame t (***predicted frame***).

Interframe video coding



P-frame prediction.

Interframe video coding

Bidirectional Temporal Prediction:

- Current frame is predicted both from previous and subsequent frames.
- The delayed frame memory (video frame buffer) is employed.
- Predicted video frame t:

$$p(\mathbf{n}, t) = a_1 f(\mathbf{n} + \mathbf{d}_t^-, t-1) + a_2 f(\mathbf{n} + \mathbf{d}_t^+, t+1).$$

- $\mathbf{d}_t^- = [dx^-, dy^-]^T$: motion vector from frame $t-l$ to frame t.
- $\mathbf{d}_t^+ = [dx^+, dy^+]^T$: motion vector from frame $t-m$ to frame t.

Interframe video coding

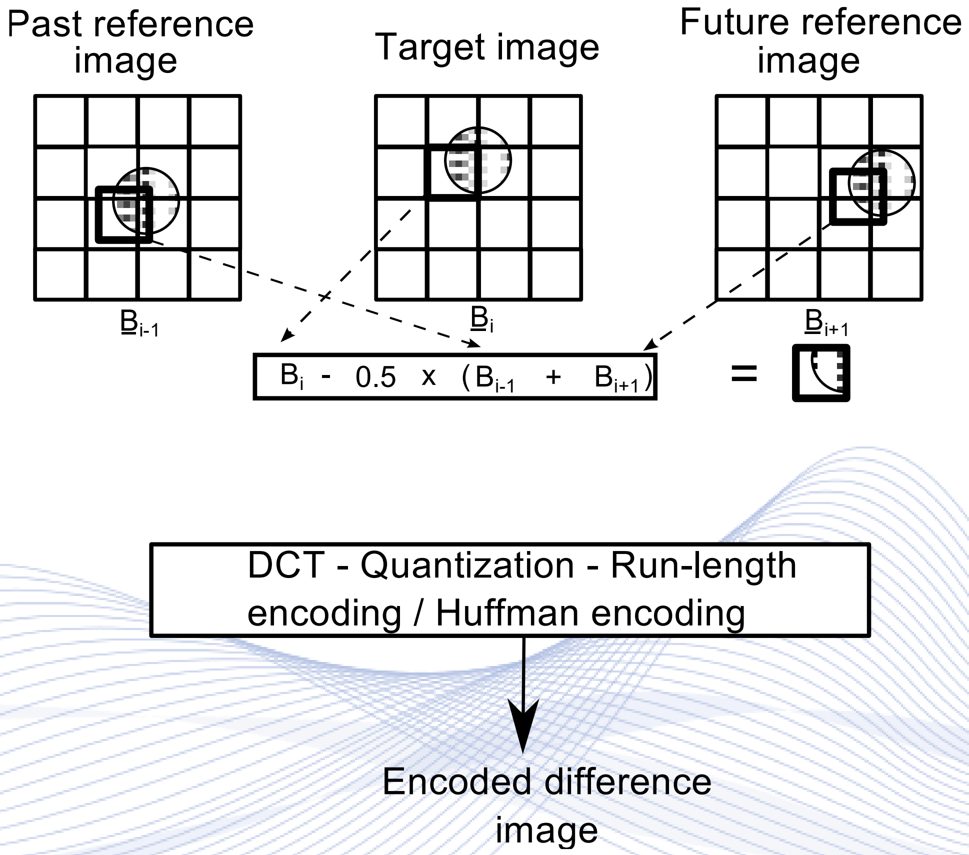
Bidirectional temporal prediction coefficients:

- $a_1 + a_2 = 1$, if mean pixel luminance does not change a lot.
- $a_1 = a_2 = 0.5$ is a good option.

Bidirectional temporal prediction is beneficial if:

- some regions in current frame do not appear in previous frames.
- For example: new objects enter the camera view field.

Interframe video coding



B-frame prediction.

Transform Video Compression

- Video compression
- Intraframe video coding
- Interframe video coding
- **Transform Video Coding**
- Predictive coding
- MPEG2
- MPEG4

Transform video coding

Block-based video coding:

- Video frame content changes in various image regions.
  - Therefore, entire video frame coding is suboptimal.
- In block-based video coding, each video frame is divided into image blocks, e.g., of 8×8 or 16×16 pixels.
- Each block is:
  - processed independently;
  - encoded using temporal prediction and transform coding.
- Block matching can be used for motion estimation.

Transform video coding

There are many 2D linear image transformations:

- DFT, ***DCT***, DST, Haar transform, Hadamard transform, Slant transform.

- They utilize the high special correlation of neighboring pixels.
- They carry image energy to few transform coefficients.
- As image content is not spatially stationary, they are applied to small frame blocks (e.g., of 8×8 or 16×16 pixels).

Most common transform for image coding: ***Discrete Cosine Transform*** (***DCT***).

Transform video coding

Quantization is applied to transform coefficients:

- Lossy compression.
- Significant reduction of bit number.
- Allocated bit number depends on ***Human Visual System*** (**HVS**) characteristics:
  - HVS is more sensitive to low and middle frequencies.
  - Low frequency coefficients: more allocated bits.
  - High frequency coefficients: less allocated bits.
- Use of ***Variable Length Coder*** (***VLC***) on quantizer output.
- Minimization of ***source entropy***.

Discrete Cosine Transform

Two-dimensional DCT:

- DCT expresses a digital signal as of a sum of cosine functions at different frequencies.

- 2D DCT is a separable transformation:

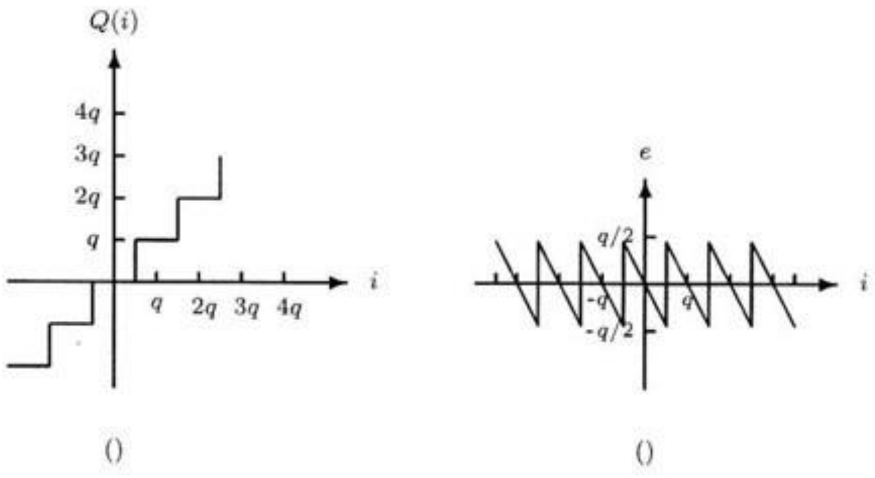
$$C(k_1, k_2) = \sum_{n_1=0}^{N_1-1} \sum_{n_2=0}^{N_2-1} 4x(n_1, n_2) \cos \frac{(2n_1+1)k_1\pi}{2N_1} \cos \frac{(2n_2+1)k_2\pi}{2N_2},$$

$$0 \leq k_1 \leq N_1 - 1, 0 \leq k_2 \leq N_2 - 1.$$

DCT Coefficient Quantization

- DCT coefficient quantization results in lossy image/video compression.
- Minimization of mean square error (MSE) between original and quantized coefficients.

- Two ***types*** of quantization:
  - Uniform quantization.
  - Non-uniform quantization.



a) Input-output relation of a uniform quantizer;
b) b) Quantization error.

DCT Coefficient Quantization

- Resulting bit number:

$$B_k = B + \frac{1}{2}\left[\log_2 \epsilon_k^2 \sigma_k^2 - \frac{1}{N}\log_2\left(\prod_{k=1}^{N} \epsilon_k^2 \sigma_k^2\right)\right]$$

- Corresponding Mean Square Error:

$$E = \left(\prod_{k=1}^{N} \epsilon_k^2 \sigma_k^2\right)^{1/\mathrm{N}} 2^{-2\mathrm{B}_k}.$$

8	7	6	5	3	3	2	2	2	1	1	1	1	1	0	0
7	6	5	4	3	3	2	2	1	1	1	1	1	0	0	0
6	5	4	3	3	2	2	2	1	1	1	1	1	0	0	0
5	4	3	3	3	2	2	2	1	1	1	1	1	0	0	0
3	3	3	3	2	2	2	1	1	1	1	1	0	0	0	0
3	3	2	2	2	2	2	1	1	1	1	1	0	0	0	0
2	2	2	2	2	2	1	1	1	1	1	0	0	0	0	0
2	2	2	2	1	1	1	1	1	1	1	0	0	0	0	0
2	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0
1	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0
1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0
1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0
1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0
1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

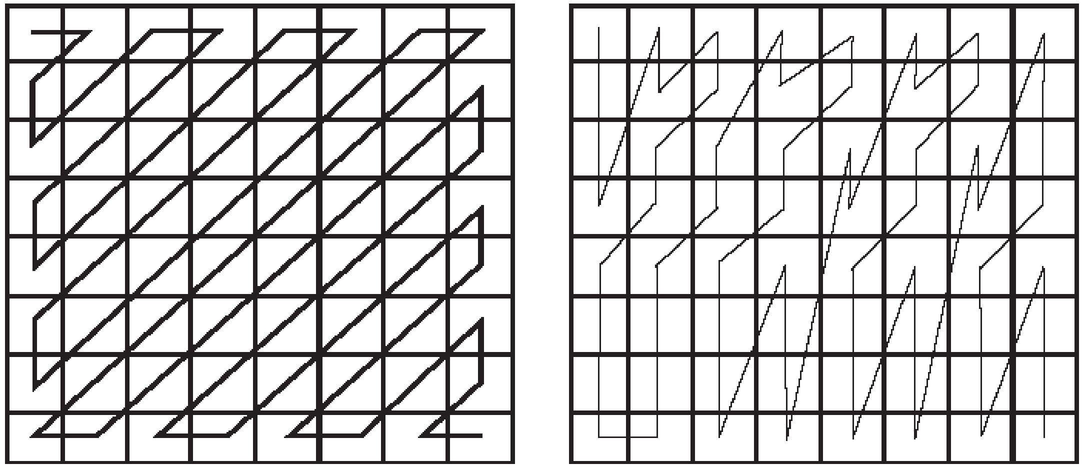
Bit allocation to 16×16 DCT coefficients.

Quantized DCT Coefficient Compression

Quantized DCT coefficients are scanned to form an one-dimensional vector.

- ***Zig-zag scanning***:
  - It is used for video frames where low-frequency DCT coefficients have equal importance along horizontal and vertical direction.
- ***Alternate scanning***:
  - For interlaced video.
  - Higher frequency content in the vertical direction.

Quantized DCT Coefficient Compression



DCT coefficient scanning: a) Zig-zag scanning; b) Alternate scanning.

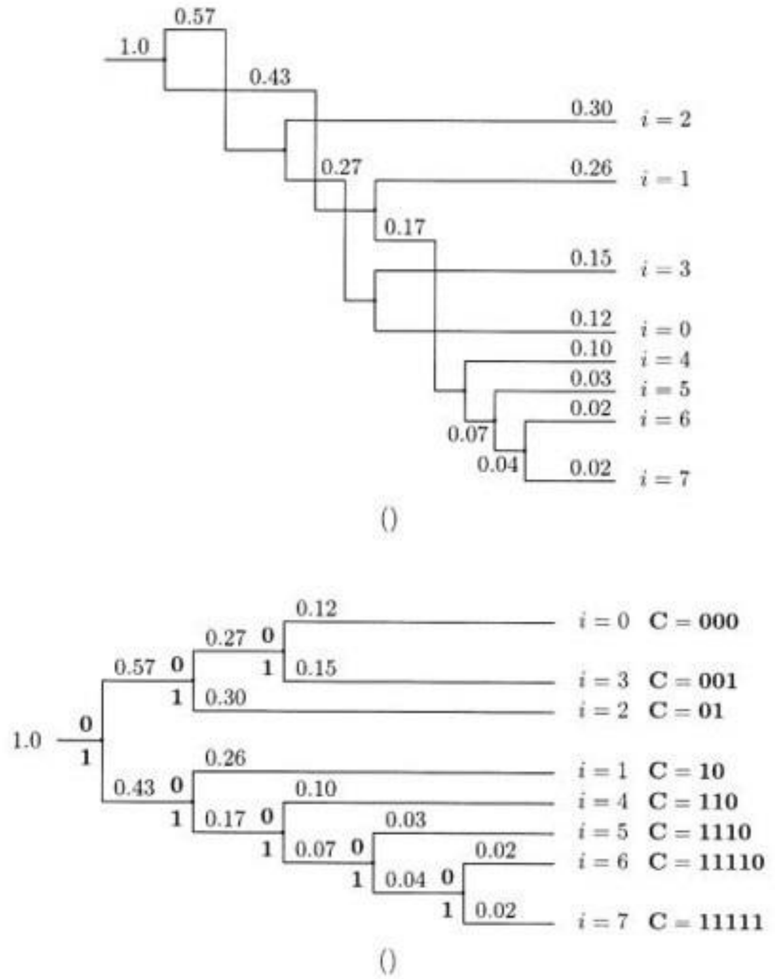
Quantized DCT Coefficient Compression

Huffman encoding process:

1. Pick the two child nodes with the smallest probabilities.
2. Form a parent node, whose probability is the sum of its children node probabilities.
3. Repeat process until all luminance levels (symbols) are used.
4. Root node probability should be 1.
5. Rearrange tree branches to disentangle them.
6. Assign 0/1, when traversing tree from root to leaves (upwards/downwords).
7. The codeword of each luminance value consists of ones and zeros in the path from the tree root to the corresponding leaf.

Quantized DCT Coefficient Compression

- Image with 8 luminance levels.
- 3 bits/pixel required for PCM encoding ($B = 3$).
- $p(i),\ i = 0, \dots, 2^{B-1}$ known probabilities.



a) Huffman tree; b) Tree re-arrangement.

Transform Video Compression

- Video compression
- Intraframe video coding
- Interframe video coding
- Transform Video Coding
- **Predictive coding**
- MPEG2
- MPEG4

Predictive Coding

Predictive coding can be used to remove information redundancy from:

- DCT coefficients.
- Motion vector of a video frame block.

Value $f(n)$ prediction from its prediction window $\mathcal{A}$:

$$\hat{f}(n) = L[f(n-i), i \in \mathcal{A}, i \neq 0].$$

- $f(n)$, $n = 1,2,\ldots$: DCT coefficient value or one component of motion vector $\mathbf{d}_t = [dx, dy]^T$.
- $f(n-i)$, $i \in \mathcal{A}$: values in prediction window $\mathcal{A}$.
- operator L is usually a linear function.

Predictive Coding

- If prediction window $\mathcal{A}$ scans the frame blocks row-wise, the prediction $\hat{f}(n)$ is causal and can be based on already reconstructed past values $f_r(n-i)$:

$$\hat{f}(n) = L[f_r(n-i), i \in \mathcal{A}].$$

- Prediction error:

$$e(n) = f(n) - \hat{f}(n).$$

Predictive Coding

- Signal $f(n)$ is reconstructed as follows:

$$f_r(n) = L[f_r(n-i), i \in A] + e_q(n).$$

- $e_q(n)$: quantized prediction error.

- Good prediction produces small error and results in better compression.

Transform Video Compression

- Video compression
- Intraframe video coding
- Interframe video coding
- Transform Video Coding
- Predictive coding
- **MPEG2**
- MPEG4

MPEG-2 Standard

The MPEG-2 standard describes a combination of lossy video compression and lossy audio data compression methods:

- It is used as a digital television signal format for terrestrial (over-the-air), cable or satellite TV broadcasting.
- It is also used as a digital cinema and video streaming/storage format.
- It is compatible with the MPEG-1 standard.
- It serves many applications of various video rates (2-20 Mbps) and resolutions.

MPEG-2 Standard

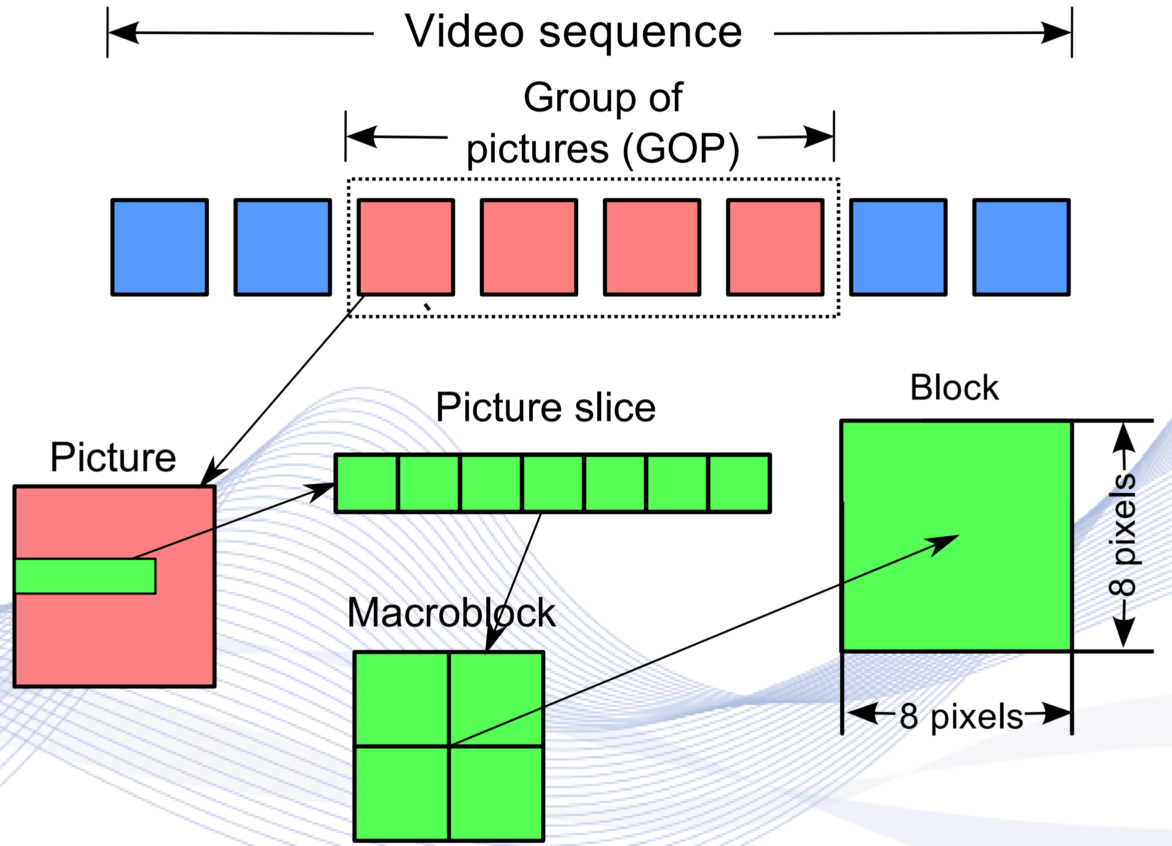
Table 13.9.1: The profiles MPEG-2

Abbreviation	Profile Name	Compression mode	Chrominance Subsampling	Picture aspect ratio	Scalability
SP	Simple	, R	4:2:0	4:3, 16:9	None
MP	Main	, R,	4:2:0	4:3, 16:9	None
SNR	SNR scalable	, R,	4:2:0	4:3, 16:9	SNR
Spatial	Spatially scalable	, R,	4:2:0	4:3, 16:9	SNR or spatial
HP	High	, R,	4:2:2, 4:2:0	4:3, 16:9	SNR spatial
422	4:2:2	I, P, B	4:2:2, 4:2:0	4:3, 16:9	SNR or spatial
MVP	Multiview	I, P, B	4:2:0	4:3, 16:9	temporal

MPEG-2 Standard

Macroblock:

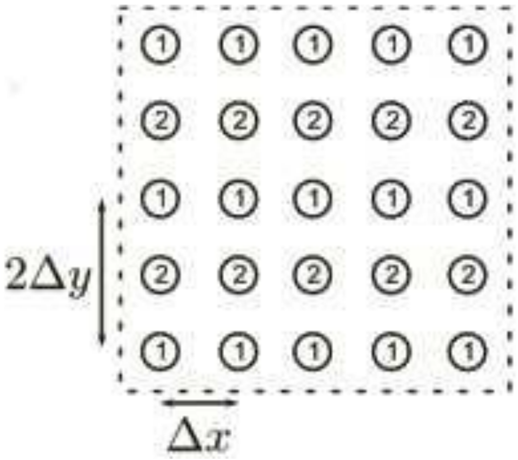
- Four 8×8 luminance blocks.
- Motion estimation performed at Macroblock level.
- Resulting motion vector used in its 4 constituent image blocks.



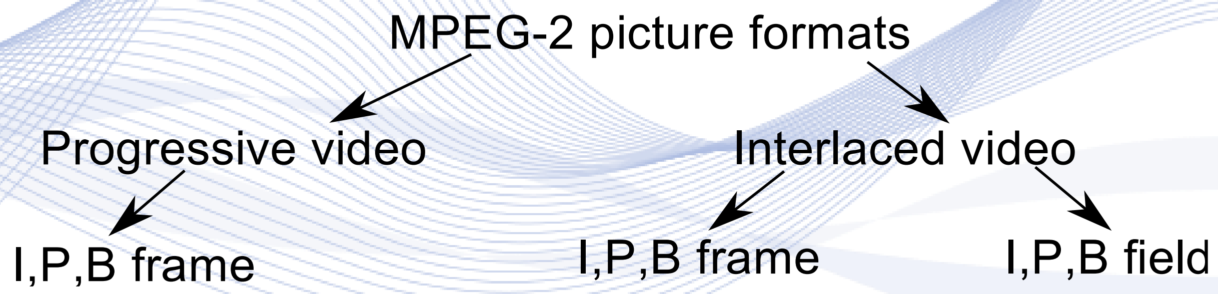
MPEG2 stream structure.

MPEG-2 Standard

- Two interlaced video picture types:
- ***Frames pictures***: obtained by deinterlacing even and odd-numbered fields (I-, P- or B-type).
- ***Field pictures***: even and odd-numbered fields as seperate images (I-, P- or B- type).
- Support of both interlaced and progressive video.



Interlaced luminance video.

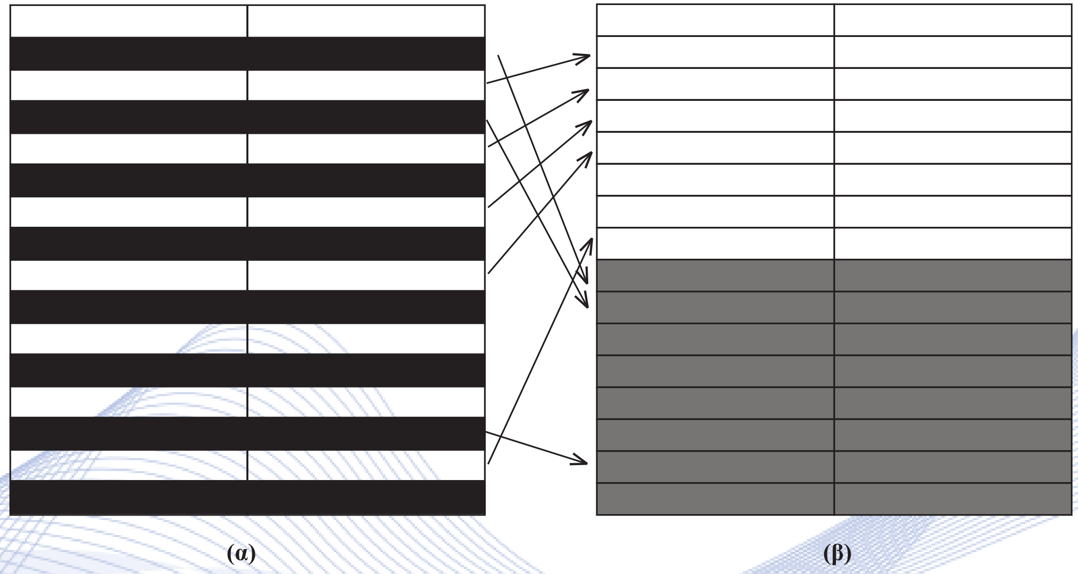


MPEG-2 picture formats.

MPEG-2 Standard

Two encoding options:

- ***Field encoding***: Each field block is encoded independently, if significant motion is present.
- ***Frame encoding***: Two fields encoded together as frame picture (better for static video content).



a) Frame DCT; b) Field DCT.

Transform Video Compression

- Video compression
- Intraframe video coding
- Interframe video coding
- Transform Video Coding
- Predictive coding
- MPEG2
- **MPEG4**

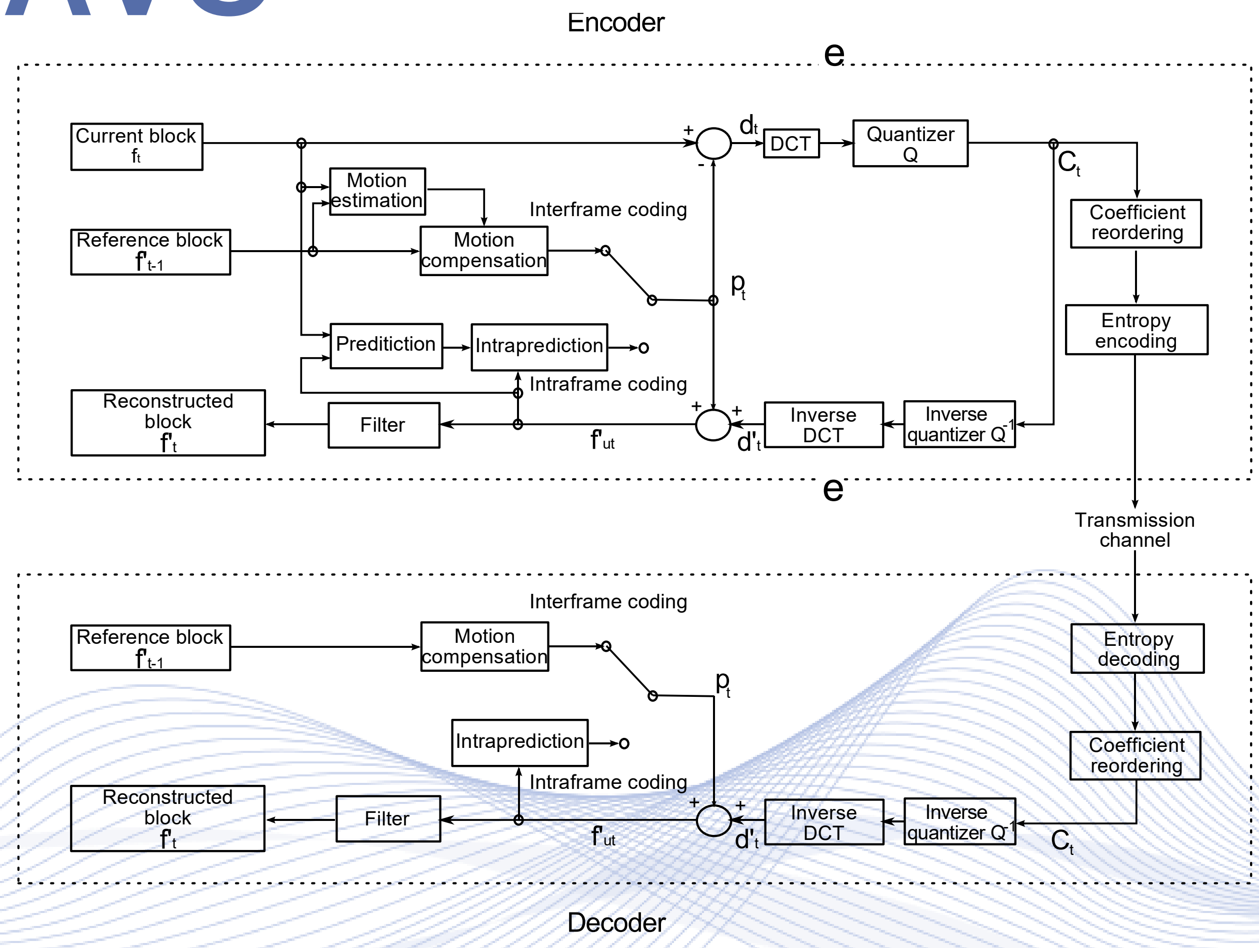
MPEG-4 AVC

MPEG-4 Advanced Video Coding (***AVC***):

also called H.264 or MPEG-4 Part 10, it is a video compression standard based on block-oriented, motion-compensated and DCT coding.

- It was developed by ITU-T Video Coding Experts Group (VCEG) with the ISO/IEC Moving Picture Experts Group (MPEG).
- It has different philosophy from MPEG-2 motion compensation sections.
- Good quality at lower transmission rates.
- Design to avoid increased implementation complexity/cost.

MPEG-4 AVC



MPEG-4 AVC structure.

Bibliography

[PIT2017] I. Pitas, “Digital video processing and analysis”, China Machine Press, 2017 (in Chinese).

[PIT2013] I. Pitas, “Digital Video and Television”, Createspace/Amazon, 2013.

[WAN2002] Wang Y, Ostermann J, Zhang YQ. Video processing and communications, Prentice Hall, 2002.

[PIT2000] I. Pitas, Digital Image Processing Algorithms and Applications, J. Wiley, 2000.

Q & A

Thank you very much for your attention!

More material in
http://icarus.csd.auth.gr/cvml-web-lecture-series/

Contact: Prof. I. Pitas
pitas@csd.auth.gr