



Τεχνολογία Πολυμέσων

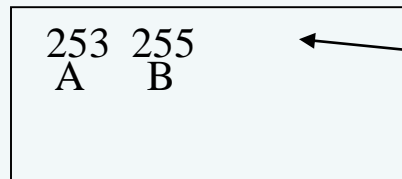


4. Διάλεξη 4: Video Coding

Image/Video Redundancy

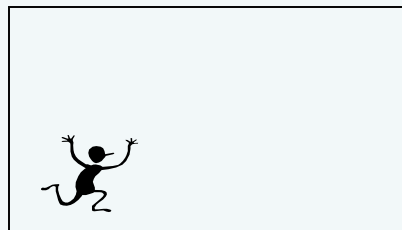
❖ Spatial redundancy **within** frames

- Pixels that are duplicated within a frame.
- It is unusual for intensity values to vary widely in a small area, e.g. a 8x8 pixel block

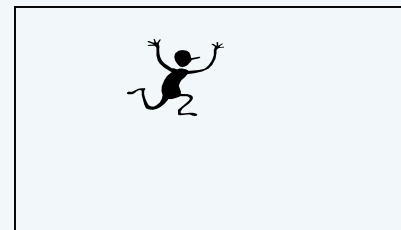


Use A to code B

❖ *Temporal redundancy* **between** frames: only position has changed

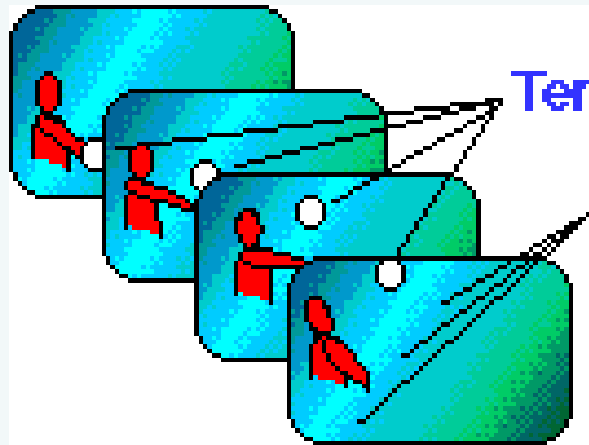


A
Frame N-1



B
Frame N

Image/Video Redundancy



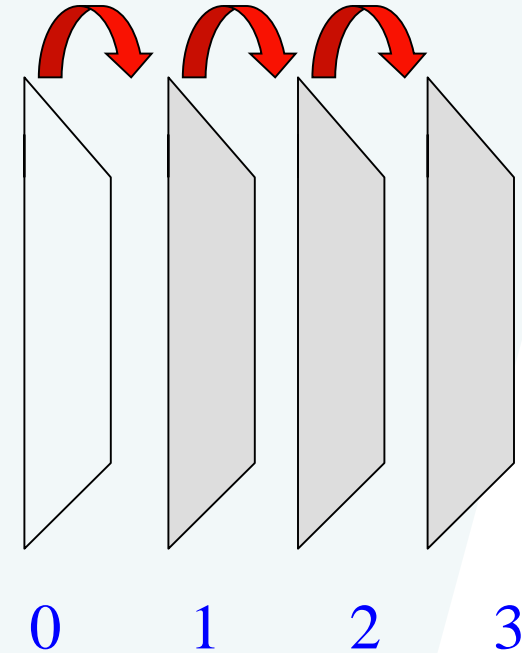
Temporal redundancy (inter-frame)

Spatial redundancy (intra-frame)



Simplest Temporal Coding

- ❖ Frame 0 (still image)
- ❖ Difference frame 1 = Frame 1 – Frame 0
- ❖ Difference frame 2 = Frame 2 – Frame 1
- ❖ If no movement in the scene, all difference frames are 0. Can be greatly compressed!
- ❖ If movement, this can be seen in the difference images





Difference Frames

- ❖ Differences between two frames can be caused by:
 - Camera motion: the outlines of background or stationary objects can be seen in the Diff Image
 - Object motion: the outlines of moving objects can be seen in the Diff Image
 - Illumination changes (sun rising, headlights, etc.)
 - Scene Cuts: Lots of stuff in the Diff Image
 - Noise



Difference Frames

- ❖ If the only difference between two frames is noise (nothing moved), then you won't recognize anything in the Difference Image
- ❖ But, if you can see something in the Diff Image and recognize it, there's still correlation in the difference image
- ❖ Goal: **Estimation motion and reduce correlation**

Difference Frames (example)



Difference Frames

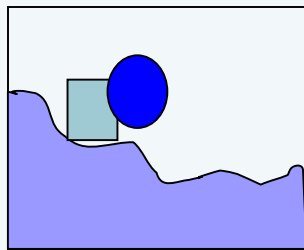


Difference Frames

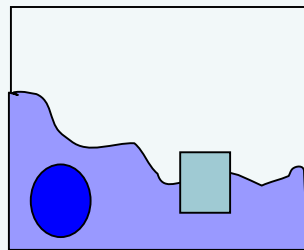


Types of Motion

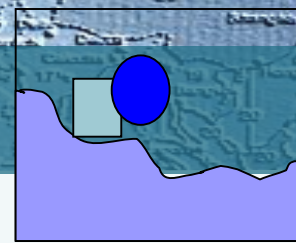
- ❖ Translation: simple movement of typically rigid objects
- ❖ Camera pans vs. movement of objects



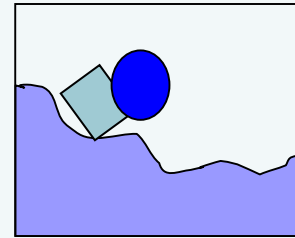
Frame n



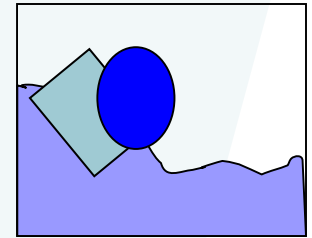
Frame n+1



Frame n



Frame n+1
(Rotation)



Frame n+2
(Zoom)

Rotation: spinning about an axis

- Camera versus object rotation

Zooms –in/out

- Camera zoom vs. object zoom (movement in/out)



Describing Motion

❖ Translational

- Move (object) from (x,y) to $(x+dx,y+dy)$

❖ Rotational

- Rotate (object) by $(r \text{ rads})$ (counter/clockwise)

❖ Zoom

- Move (in/out) from (object) to increase its size by $(t \text{ times})$

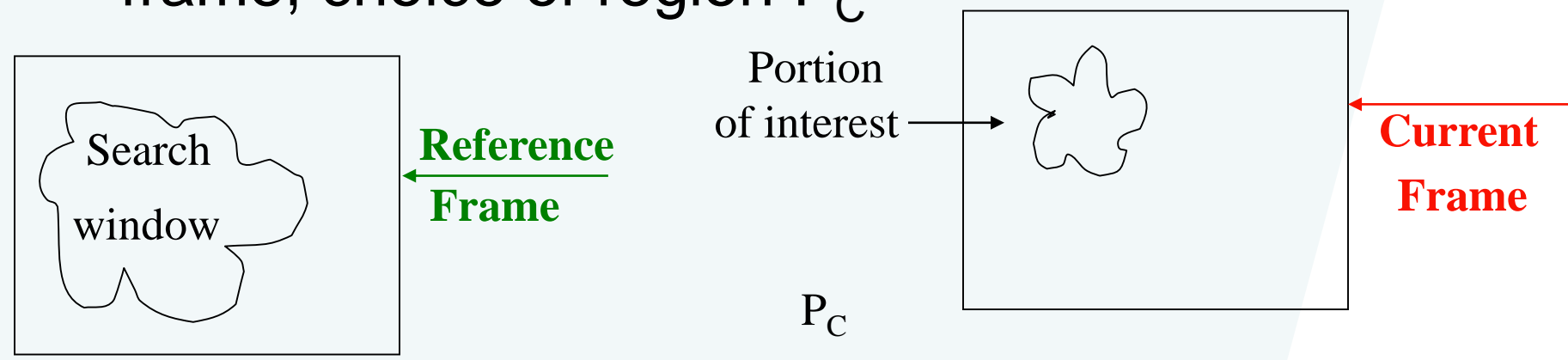


Motion Estimation

- ❖ Determining parameters for the motion descriptions
- ❖ For some portion of the frame, estimate its movement between 2 frames- the *current frame* and the *reference frame*
- ❖ What is some portion?
 - Individual pixels (all of them)?
 - Lines/edges (have to find them first)
 - Objects (must define them)
 - Uniform regions (just chop up the frame)

Motion Estimation

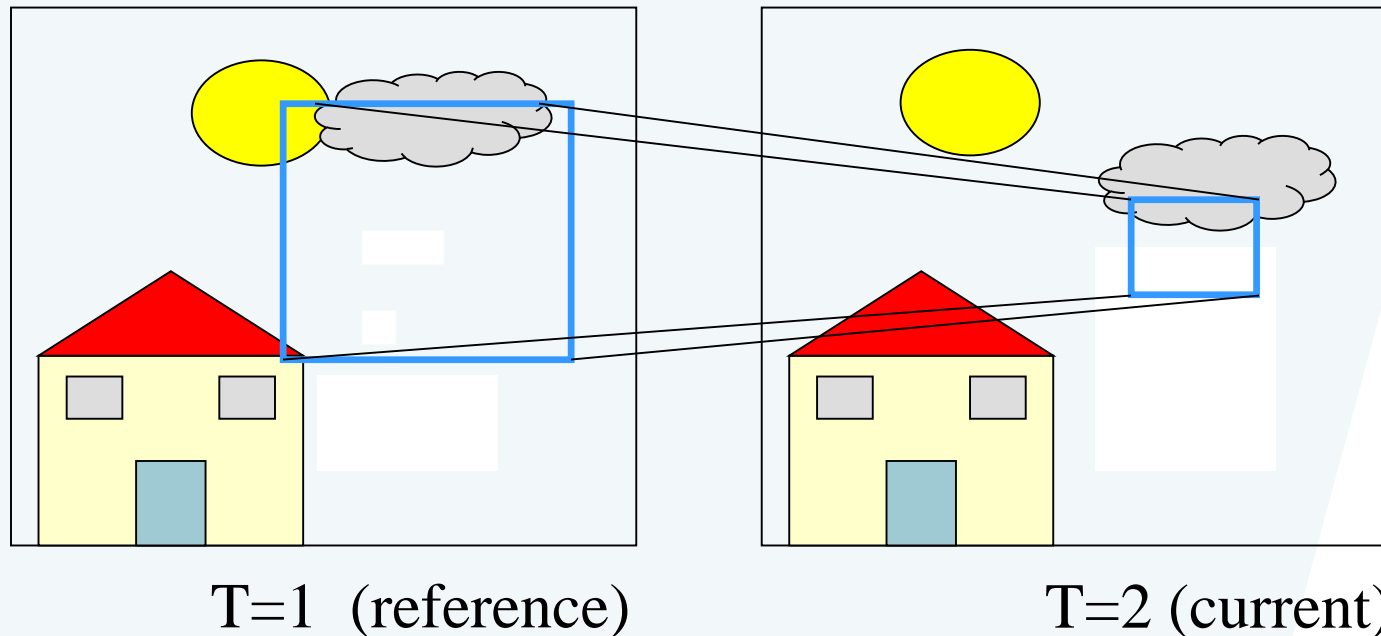
- ❖ For a region P_C in the current frame, find a region P_R in the search window in reference frame so that $\text{Error}(P_R, P_C)$ is minimized
- ❖ Issues: Error measures, search techniques, choice of search window, choice of reference frame, choice of region P_C



Reference frames are frames of a compressed video that are used to define future frames.

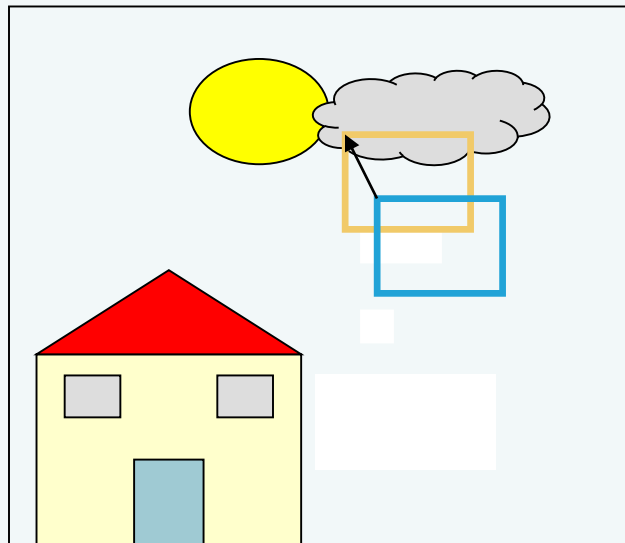
Block-based Motion Estimation

- ❖ P_C is a block of pixels (in the current frame)
- ❖ The search window is a rectangular segment (in the reference frame)

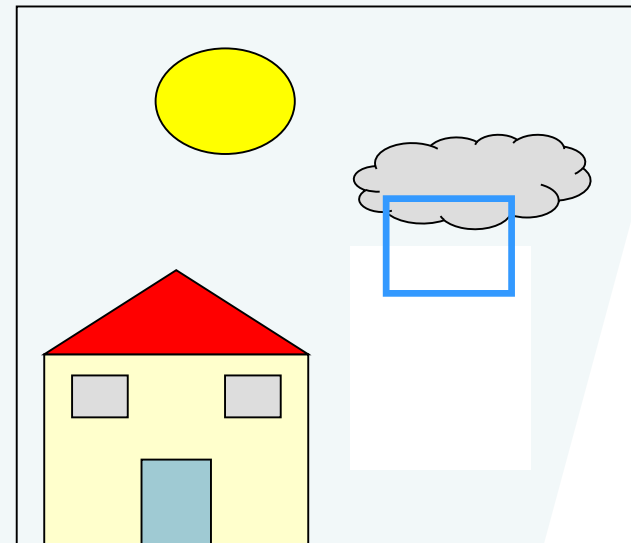


Motion Vectors

- ❖ A motion vector (MV) is the key element in the motion estimation process
- ❖ A *motion vector* (MV) represents a block in a picture being coded, based on the position of this block (or a similar - best-match block) in another picture, called the reference picture.



T=1 (reference)

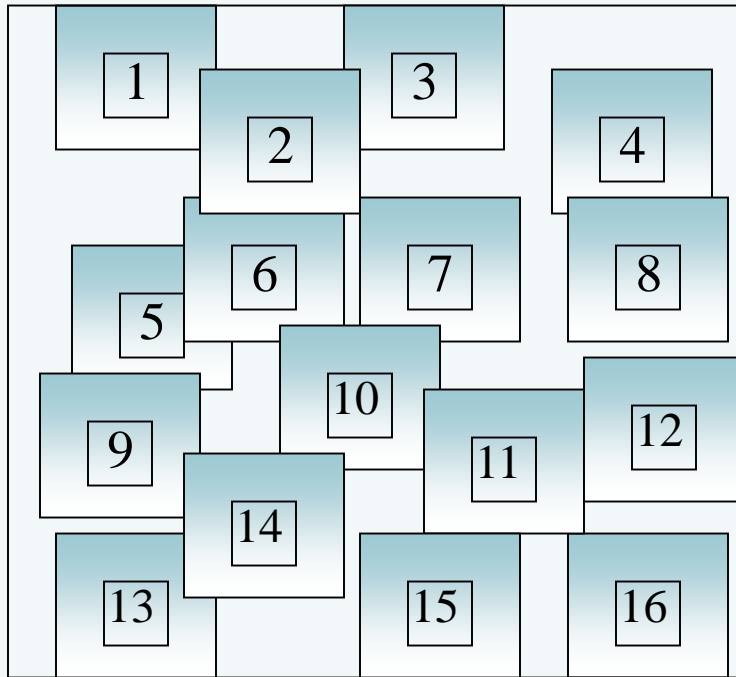


T=2 (current)

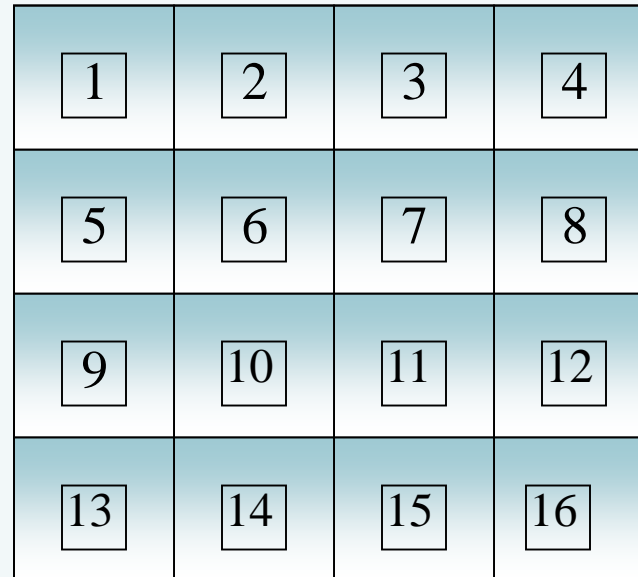
Motion Compensation

- **Motion compensation** is an algorithmic technique used to predict a frame in a video, given the previous and/or future frames by accounting for motion of the camera and/or objects in the video

The blocks are NOT on a grid



The blocks are on a grid



T=1 (reference frame defines the blocks)

T=2 (current)



Motion Vector Search

❖ 1. Mean squared error

- Select a block in the reference frame to minimize

$$\Sigma(b(B_{\text{ref}})-b(B_{\text{curr}}))^2/n$$

❖ 2. Mean absolute error

- Select block to minimize

$$\Sigma|b(B_{\text{ref}})-b(B_{\text{curr}})| /n$$

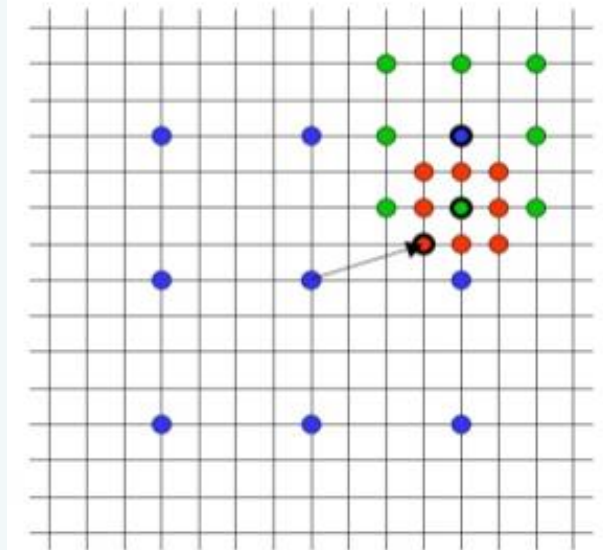
❖ **Given error measure, how to efficiently determine best-match block in search window?**

- Full search: best results, most computation
- Logarithmic search – heuristic, faster
- Hierarchical motion estimation

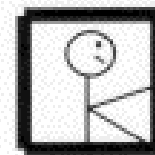
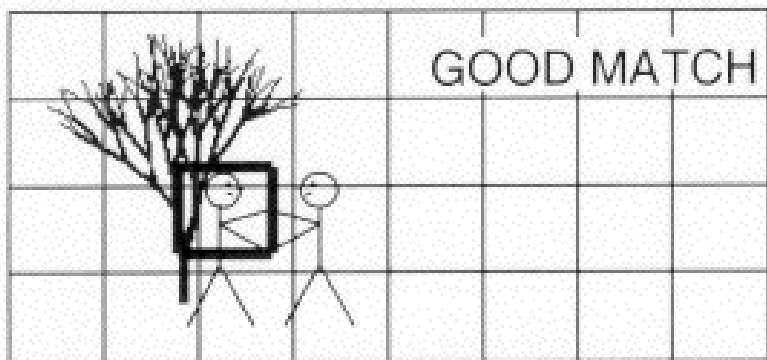
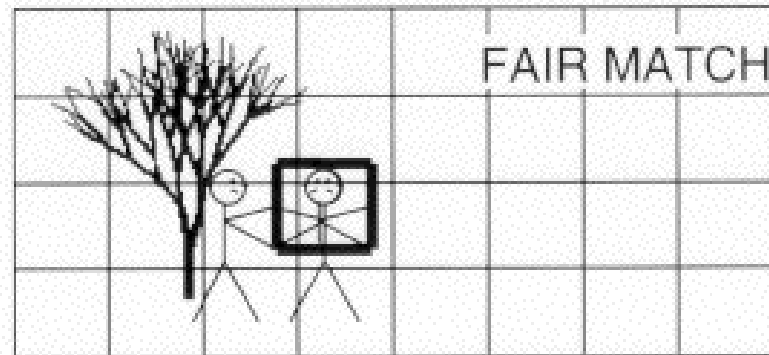
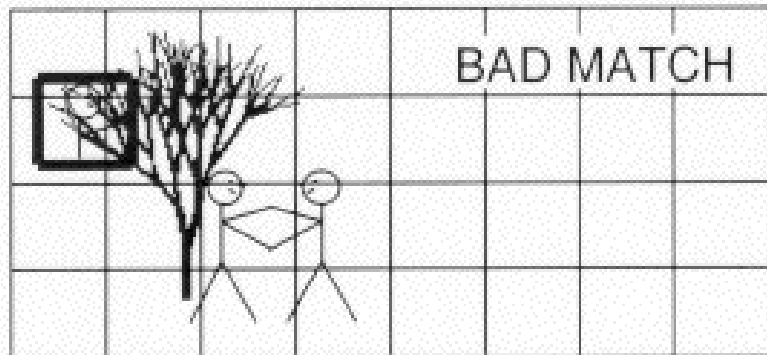
Motion Vector Search

❖ 2D Logarithmic search $\rightarrow O(\log n)$

- Reduces number of search points
- In each step
 - Best match is found
 - Compare 9 blocks evenly spaced within distance $p/2$ and find best matching block
 - Search is centered around the best match of the prior step
- Speedup increases for larger search areas



Block-based Motion Estimation



Macroblock to be coded

Motion Compensation

- ❖ Encoder forms motion compensated diff frame:

$$\text{MCD}(n) = F(n) - \text{MC}(n)$$

- ❖ Decoder can then do

$$F(n) = \text{MCD}(n) + \text{MC}(n)$$

- ❖ With **no** motion compensation encoder could do frame diff:

$$\text{FD}(n) = F(n) - F(n-1)$$

- ❖ Decoder can then do

$$F(n) = \text{FD}(n) + F(n-1)$$

- ❖ This difference frame formed using MC should have less correlation between pixels than the difference frame formed without using MC
- ❖ If successive frames are very similar:
 - **fewer bits to send Motion Vectors + MCD(n) instead of FD(n)**
 - **fewer bits to send FD(n) instead of F(n)**

Motion Compensation



Full original frame, as shown on screen.

Differences between the original frame and the next frame.



Differences between the original frame and the next frame, shifted right by 2 pixels. Shifting the frame *compensates* for the panning of the camera, thus there is greater overlap between the two frames.

Motion Compensation

Frame 1 $s[x,y,t-1]$ (previous)



Frame 2 $s[x,y,t]$ (current)

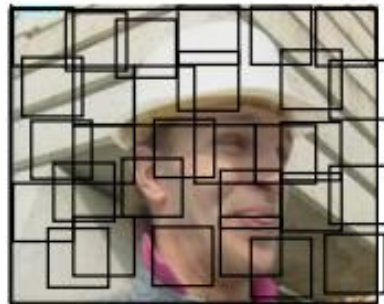


Partition of frame 2 into blocks (schematic)



Size of Blocks

Accuracy of Motion Vectors



Referenced blocks in frame 1



Frame 2 with displacement vectors



Difference between motion-compensated prediction and current frame $u[x,y,t]$

Τι είναι το MPEG;

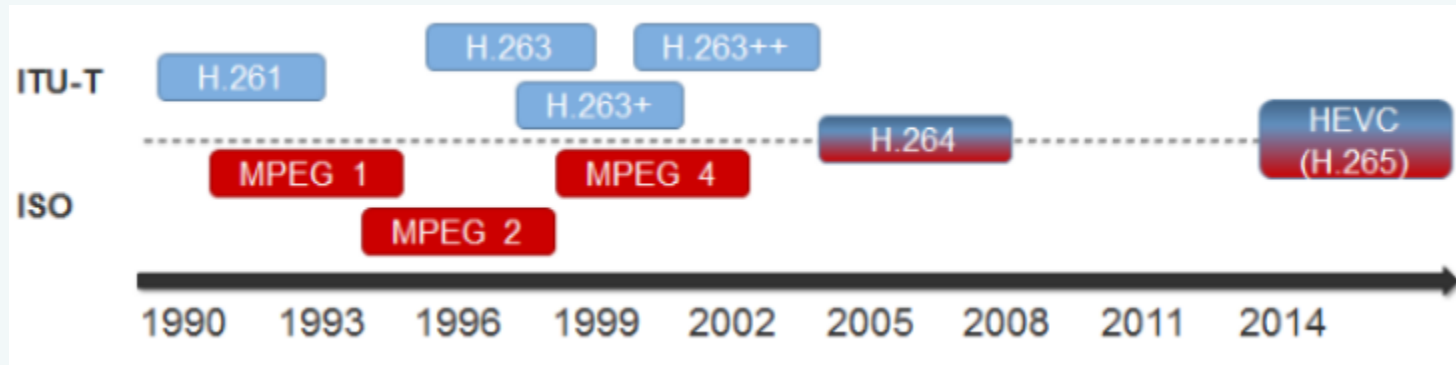
- ❖ MPEG (Motion Pictures Experts Group)
 - Παρόμοιο όνομα με το JPEG
 - Βίντεο και ήχος υψηλής ποιότητας
 - Διανομή ή μετάδοση μέσω
 - MPEG-1: CD-ROM, T1 (~1,5 Mbps)
 - MPEG-2: DVD, DTV
 - MPEG-4: HDTV, κινητά, αλληλεπίδραση



Τι είναι το MPEG;

- ❖ Αξιοποίηση άλλων προτύπων
 - JPEG: ενδοπλαισιακή (intra-frame) κωδικοποίηση
 - H.261: διαπλαισιακή (inter-frame) κωδικοποίηση
- ❖ Σύστημα MPEG
 - Κωδικοποίηση ήχου
 - Κωδικοποίηση βίντεο
 - Πολύπλεξη ροών δεδομένων

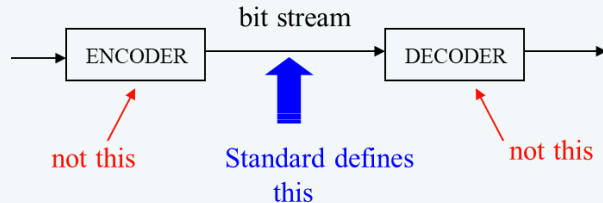
Video Coding Standards



Standards	MPEG2 (MP)	MPEG4 (ASP)	H.264 (HP)
Picture coding type	I, P, B	I, P, B	I, P, B
Entropy Coding	VLC	VLC	UVLC, CAVLC, CABAC
MV resolution	½ pel	¼ pel	¼ pel
Transform	8x8 DCT	8x8 DCT	4x4 & 8x8 Integer
Vector block size	16x16, 16x8	16x16, 8x8	16x16, 16x8, 8x16, 8x8, 8x4, 4x8, 4x4
Special Intra prediction	No	No	Yes
Loop filter	Post	Post	In-loop

MPEG

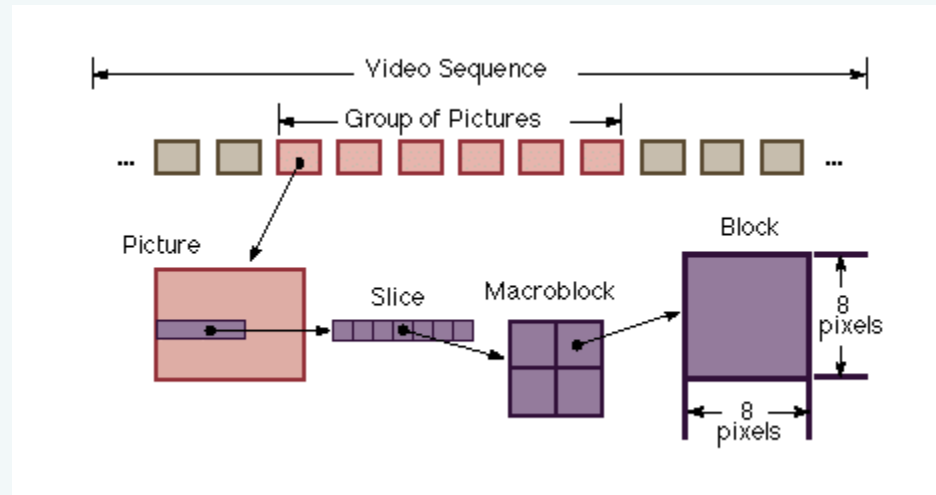
- ❖ The video compression standards define **syntax and semantics for the bit stream between encoder and decoder**



- ❖ Encoder is not specified by MPEG except that it produces a compliant bit stream
 - The standards do not specify how the encoder will find the motion vectors (MVs)
 - The encoder can use exhaustive/fast search, MSE /MAE/other error metric, etc.
- ❖ **The standard DOES specify**
 - **The allowable syntax for specifying the MVs**
 - **What the decoder will do with them**
- ❖ Compliant decoder must interpret all legal/eligible MPEG bit streams
- ❖ *This allows future encoders of better performance to remain compatible with existing decoders.*
- ❖ *Also allows for commercially secret encoders to be compatible with standard decoders*

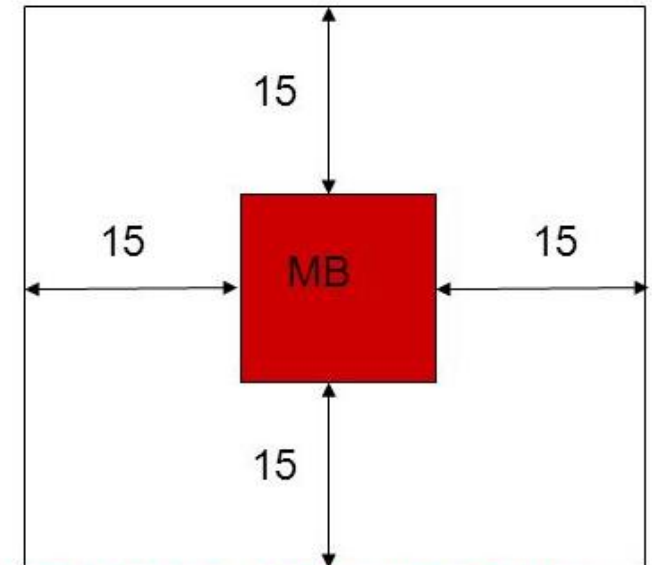
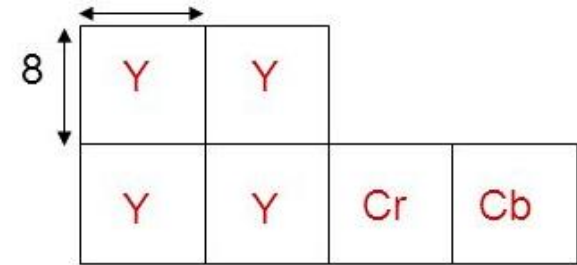
MPEG Frame Structure Terminology

- ❖ A **block** contains 8x8 pixels
- ❖ A **macroblock** (MB) contains 4 blocks from the luminance, plus the corresponding chrominance blocks
 - Macro-blocks are useful for motion estimation
- ❖ A **slice** is a collection of macroblocks
- ❖ A **picture** is a frame; the primary coding unit
- ❖ A **Group of Pictures** (GOP) contains ≥ 1 frame.



MPEG Motion Estimation

- ❖ Macro-block:
 - Luminance: 16x16, four 8x8 blocks
 - Chrominance: two 8x8 blocks
 - Motion Estimation only performed for luminance component
- ❖ Motion vector range
 - e.g. [-15, 15]

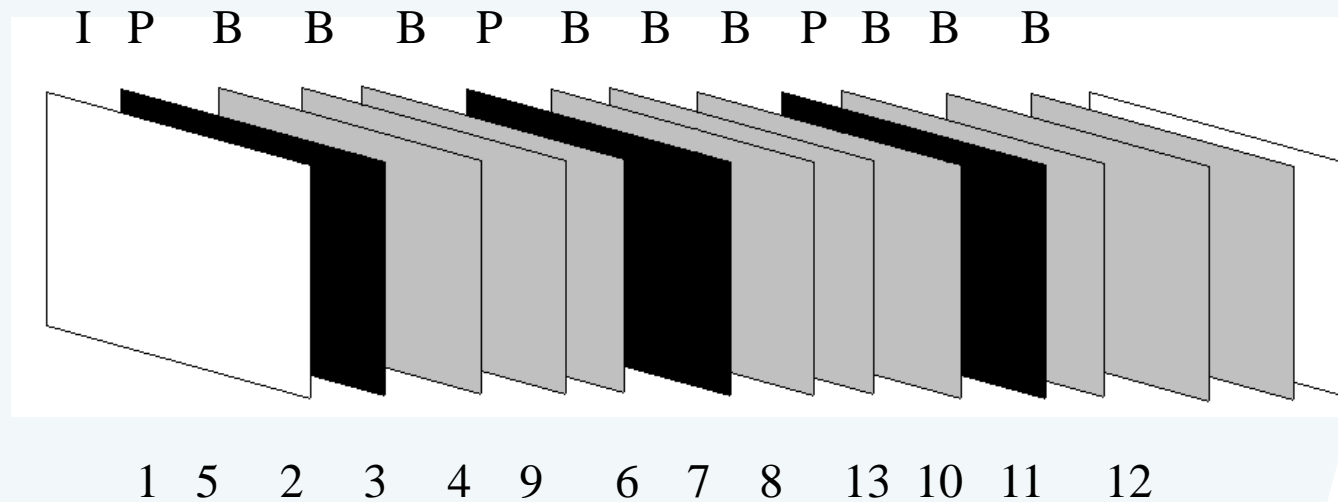


Search Area in Reference Frame

MPEG

- ❖ Consider a sequence of n frames of video.
- ❖ It consists of:
 - I-frames
 - P-frames
 - B-frames
- ❖ A sequence of one I-frame followed by P- and B-frames is known as a Group of Pictures – GOP, e.g. IBBPBBPBBPBBP
 - A GOP can be decoded without any other reference frames
 - Usually 12 or 15 frames
 - Random access to middle of stream – Start with I frame

Frame Types in MPEG

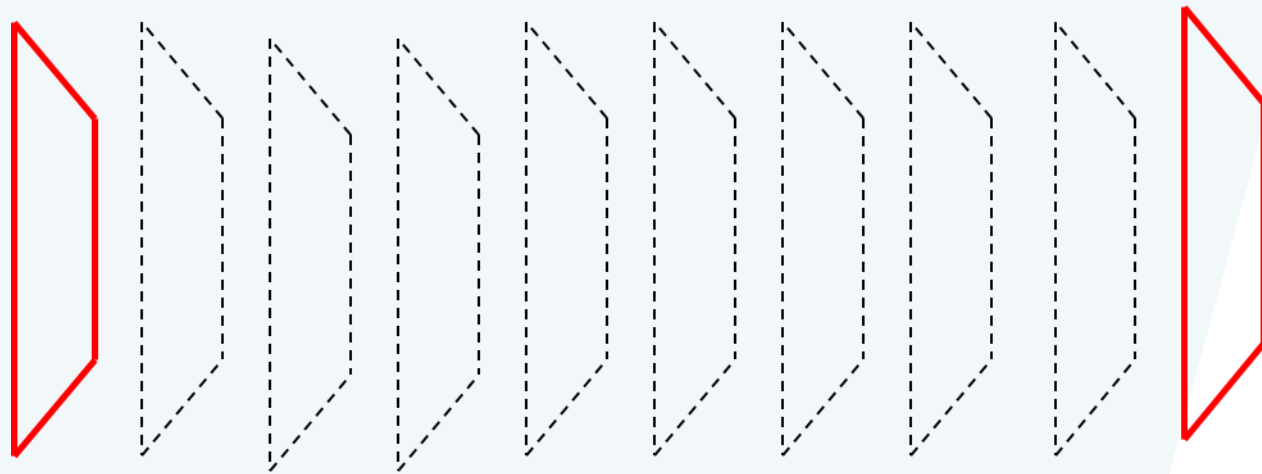


Transmission order and display order may differ
Reference frames must be transmitted first

MPEG Frame Types

❖ Intra frames (I-frames)

- Points of random access in a MPEG stream
- A I-frame is encoded using only information from within that frame (intra coded), as still (JPEG) images
- No temporal compression (inter coded)
- No motion vectors

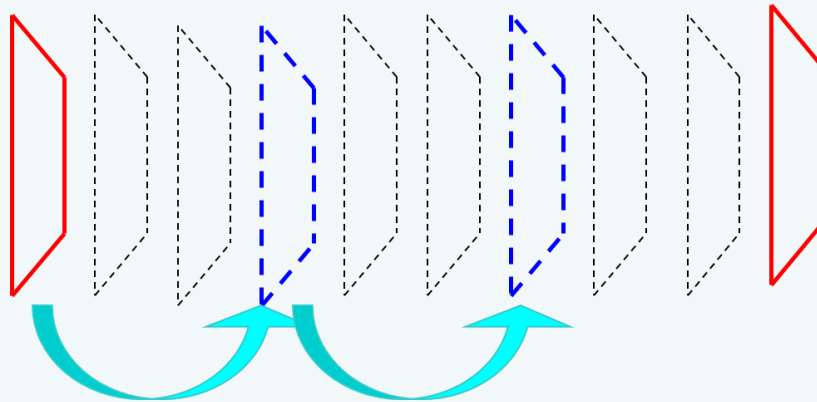


MPEG Frame Types

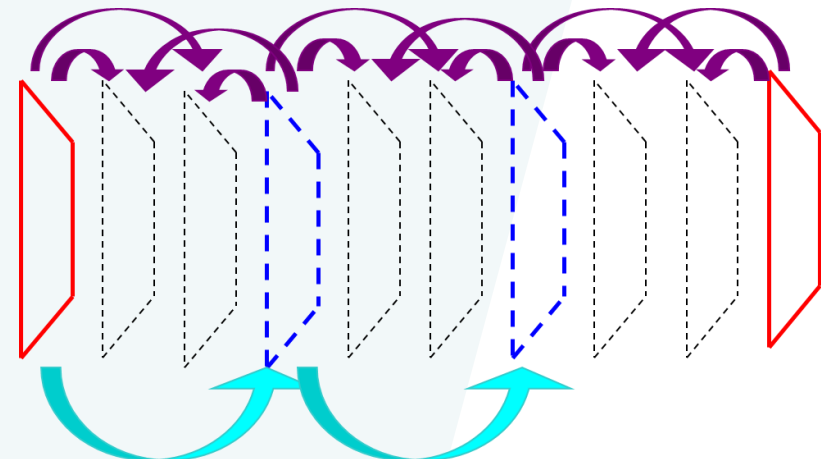
❖ Non-intra frames (P-frames and B-frames)

- Motion compensated information will be used for coding
- P frame (predicted frame) use preceding I and P frames for motion compensation
- B frame (bidirectional frame) use both preceding and followings I and P frames for motion compensation

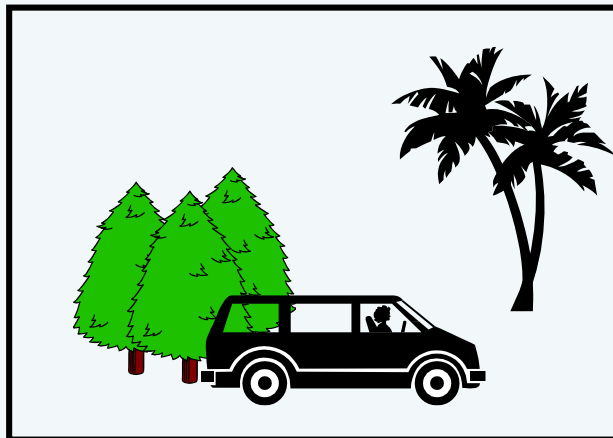
P frames



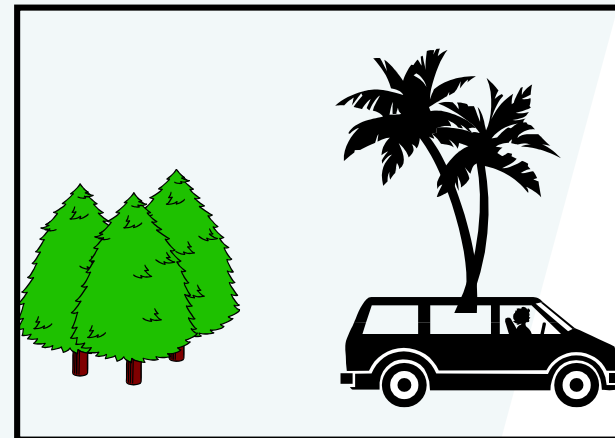
B frames



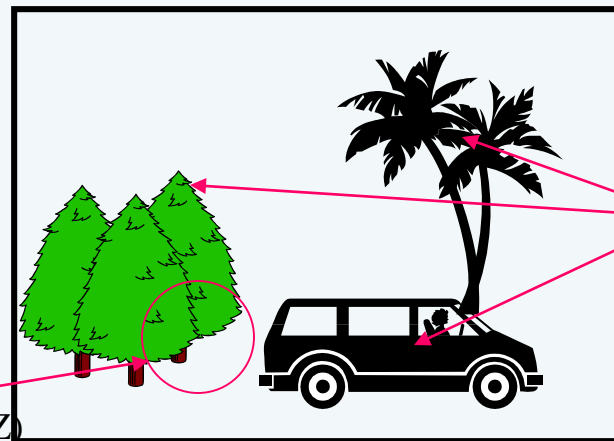
MPEG Frame Types



X



Z



Available from later frame (Z)

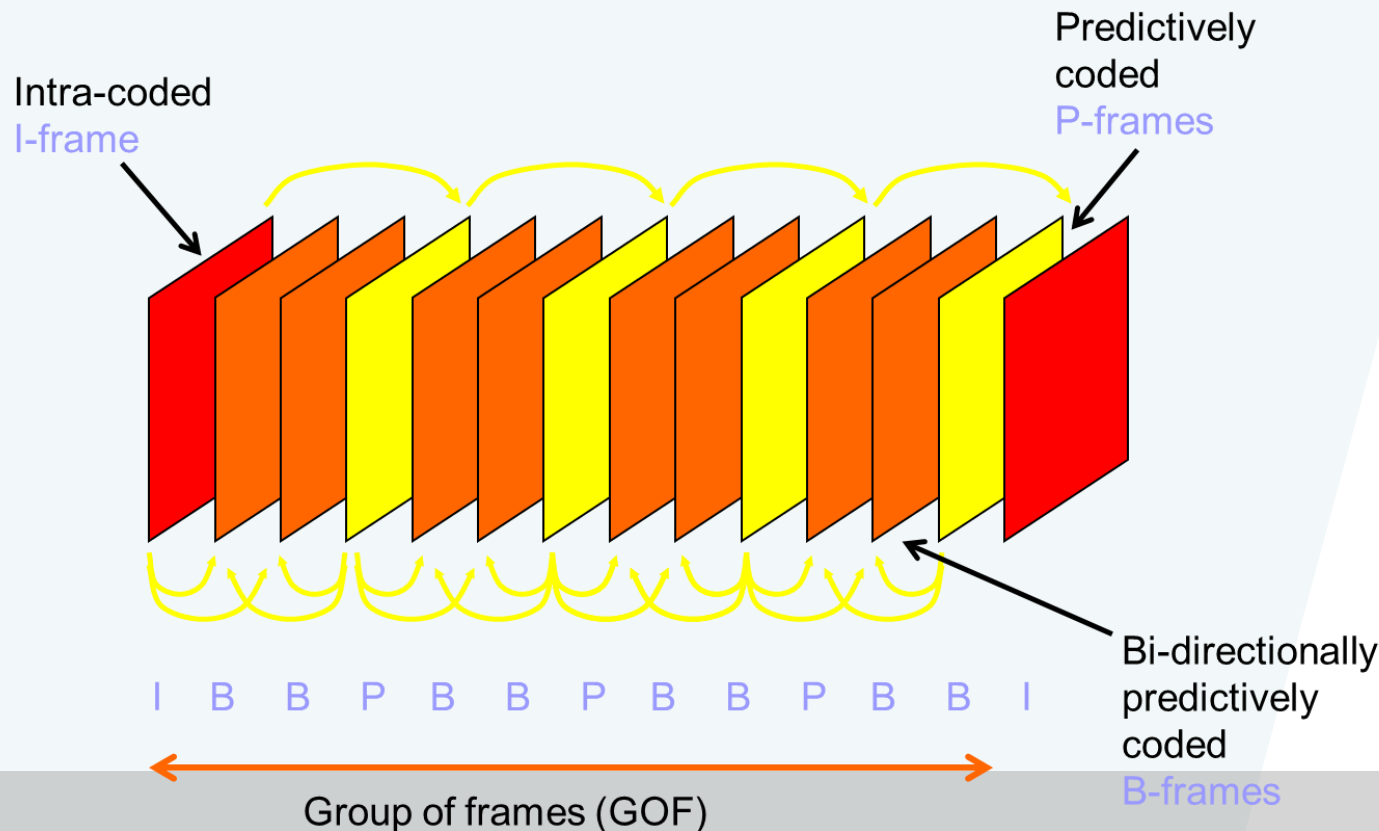
Available from earlier frame (X)

Y

MPEG Frame Types

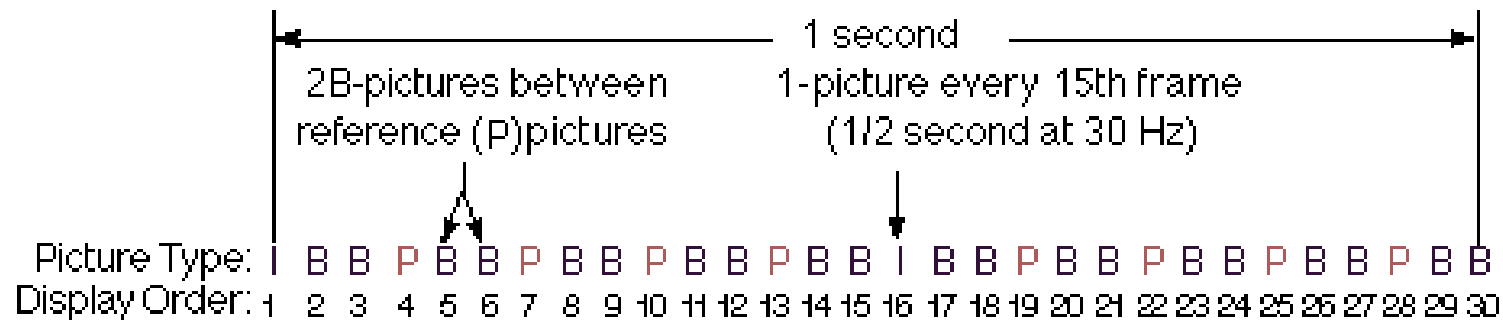
❖ Selecting I, P, or B Frames

- change of scenes should generate I frame
- limit B and P frames between I frames
- B frames are computationally intense



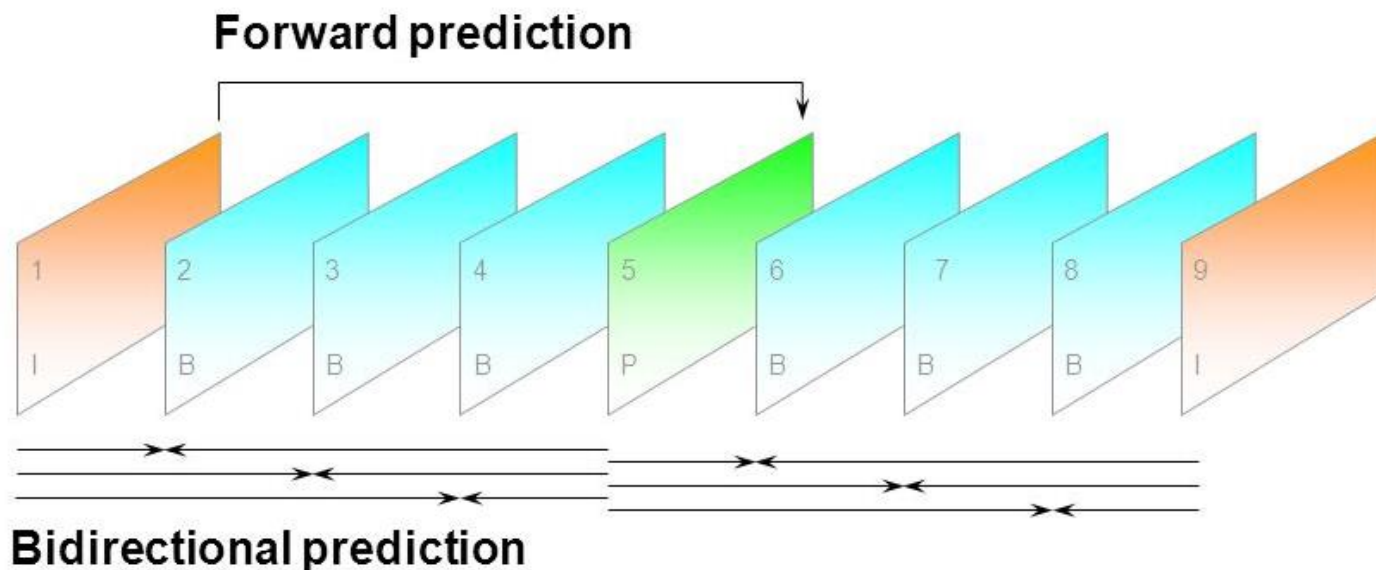
Video Stream Composition

- ❖ Encoder chose frequency and location of I-Pictures
- ❖ Typical display order



Video Stream vs Display order

- ❖ Transmission order and display order may differ
 - Reference frames must be transmitted first



Transmission Order : 1 5 2 3 4 9 6 7 8
I P B B B I B B B



MPEG

❖ I-frames

- Intraframe coded
 - No motion compensation

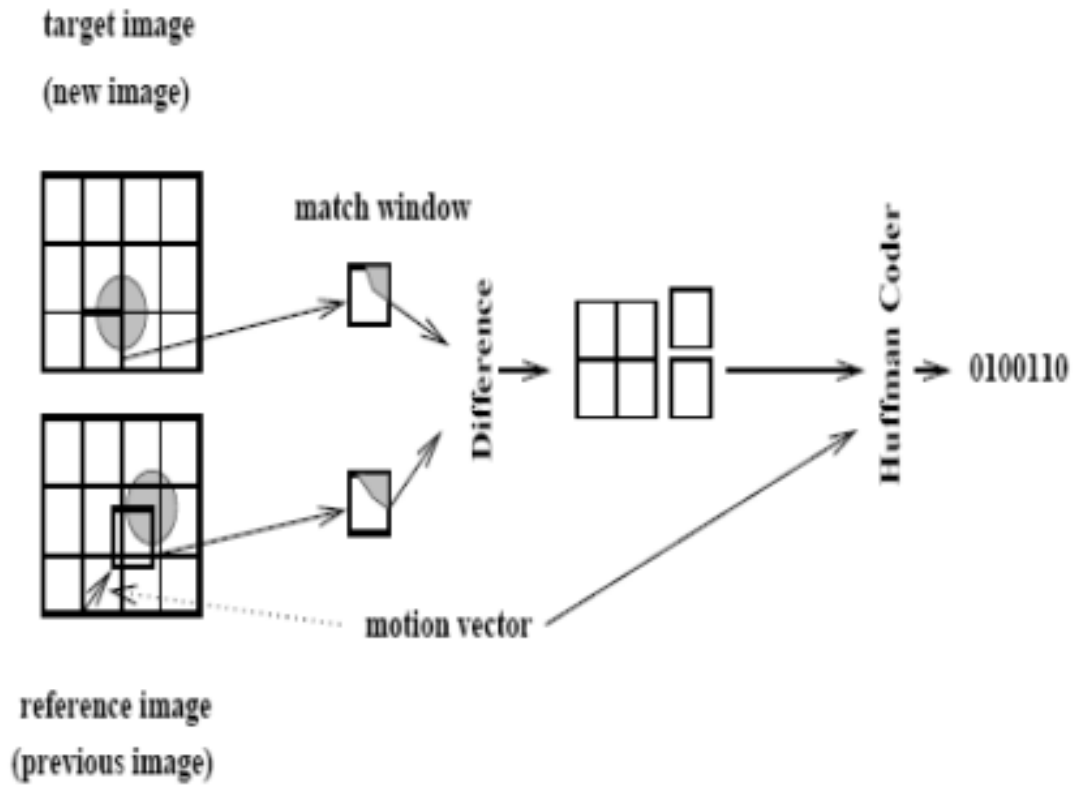
❖ P-frames

- Interframe coded
 - Motion compensation
 - Based on past frames only

❖ B-frames

- Interframe coded
 - Motion compensation
 - Based on past and future frames

MPEG-coding of P-frames



Motion compensated prediction

- ❖ Divide current frame, i , into disjoint 16×16 macroblocks
- ❖ Look for match window within a given search window in previous frame, $i-1$
 - Match window – macro-block
 - Search window – arbitrary window size depending how far away are we willing to look
- ❖ Displacement of two match windows is expressed by motion vector
- ❖ Calculate the difference
- ❖ **Transmit motion vector + entropy coded difference**

Encoder forms motion compensated diff frame:

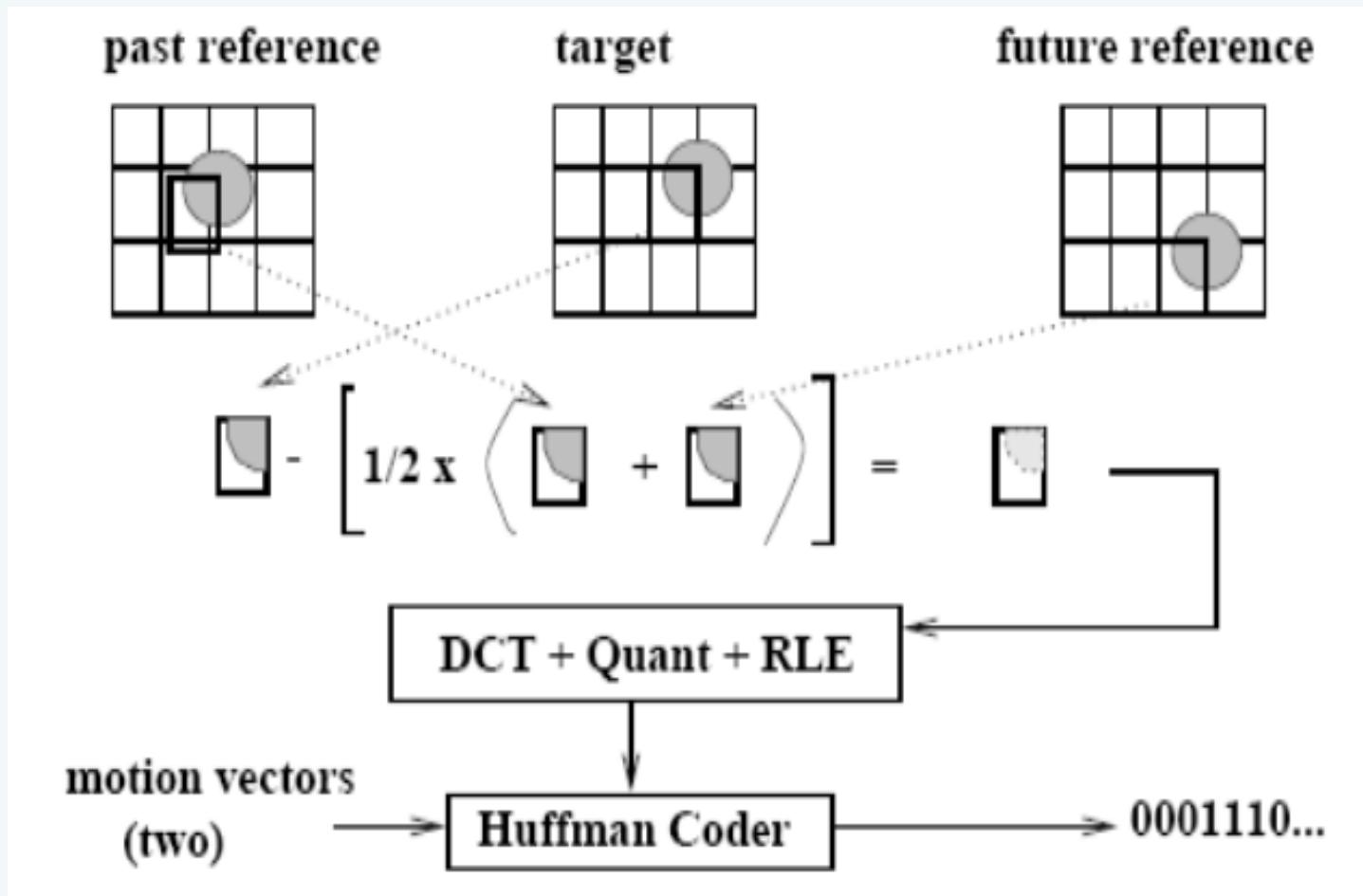
$$\mathbf{MCD}(n) = \mathbf{F}(n) - \mathbf{MC}(n)$$

Decoder can then do:

$$\mathbf{F}(n) = \mathbf{MCD}(n) + \mathbf{MC}(n)$$

MPEG-coding of B-frames

Bi-directionally Predictive-coded frames



A background map of the United States, showing state boundaries and major cities. The map is partially obscured by a blue gradient overlay at the top and bottom of the slide.

MPEG Video Quantization

- ❖ If data rate increases over threshold, then quantization enlarges step size (increase quantization factor Q)
- ❖ If data rate decreases below threshold, then quantization decreases Q

Στόχοι του MPEG-1 (1 από 2)

- ❖ Ρυθμός δεδομένων έως 1,2-1,5 Mbps
 - CD-ROM μονής ταχύτητας ή γραμμή T-1
- ❖ Μορφότυπα εικόνας
 - Συνιστώσες YUV, αναλογία (4:1:1 ή 4:2:0)
 - 8 bits ανά εικονοστοιχείο ανά συνιστώσα
- ❖ 14 λόγοι διαστάσεων εικονοστοιχείων
 - 1:1 για τετράγωνα εικονοστοιχεία
 - Διαφορετικοί λόγοι για 4:3 και 16:9

Στόχοι του MPEG-1 (2 από 2)

- ❖ Τυποποιημένο μορφότυπο ανταλλαγής (SIF)
 - NTSC
 - 352 x 240 (φωτεινότητα)
 - 176 x 120 (χρωμικότητα)
 - PAL/SECAM
 - 352 x 288 (φωτεινότητα)
 - 176 x 144 (χρωμικότητα)
 - Ρυθμοί πλαισίου από 23,97 Hz έως 60 Hz

Δομή εικόνας MPEG-1



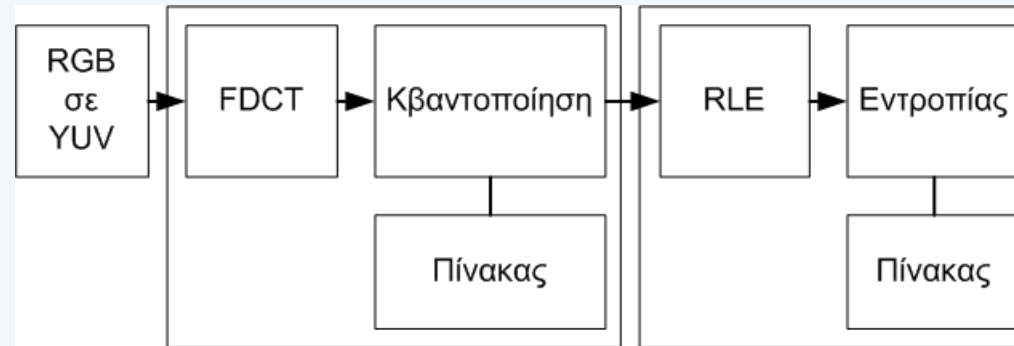
❖ Μακρομπλόκ

- 16 x 16 φωτεινότητα, 8 x 8 χρωμικότητα

❖ Οριζόντια τμήματα (slices)

- 352 x 240 (NTSC): 15 τμήματα x 22 μακρομπλόκ
- 352 x 288 (PAL/SECAM): 18 τμήματα x 22 μακρομπλόκ

Καρέ-I (1 από 2)



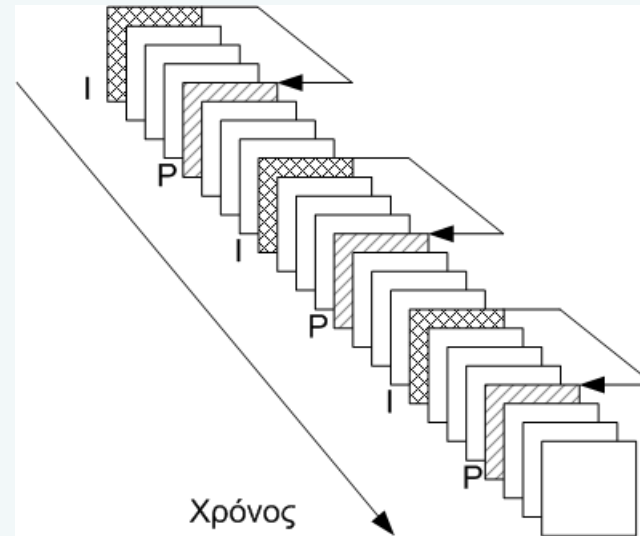
❖ Καρέ-I (I-frames)

- Κωδικοποίηση χωρίς αναφορές σε άλλα καρέ
- Κατάλληλα για τυχαία προσπέλαση
- JPEG με απωλεστικό ακολουθιακό ρυθμό
 - Μικρές αποκλίσεις για απλότητα

Καρέ-Ι (2 από 2)

- ❖ Είσοδος: Μπλοκ (8x8)
 - Μετασχηματισμός DCT
 - Κβαντοποίηση συντελεστών
 - DPCM για συντελεστές DC
 - Διάταξη ζιγκ-ζαγκ για συντελεστές AC
 - Κωδικοποίηση μήκους σειρών (RLE)
 - Κωδικοποίηση εντροπίας παρόμοια με Huffman

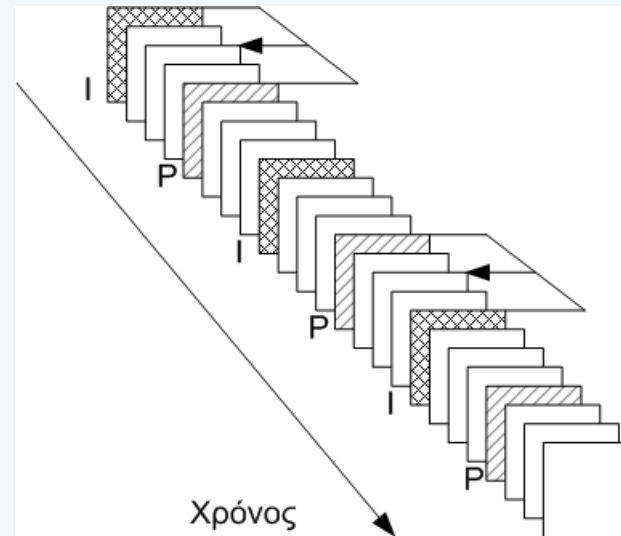
Καρέ-P (1 από 2)



❖ Καρέ-P (P-frames)

- Επανόρθωση κίνησης με καρέ αναφοράς
- Αναφορά: αμέσως προηγούμενο καρέ-I ή -P
- Εντοπισμός πλέον παρόμοιου μακρομπλόκ
- Επιτρέπονται και μακρομπλόκ χωρίς πρόβλεψη

Καρέ-B



❖ Καρέ-B (B-frames)

- Πρόβλεψη με δύο καρέ αναφοράς
- Ίδια μέθοδος κωδικοποίησης με τα καρέ-P
- Περιοδική χρήση καρέ-I και καρέ-P
- Κωδικοποίηση ενδιάμεσων καρέ-B

Καρέ-D

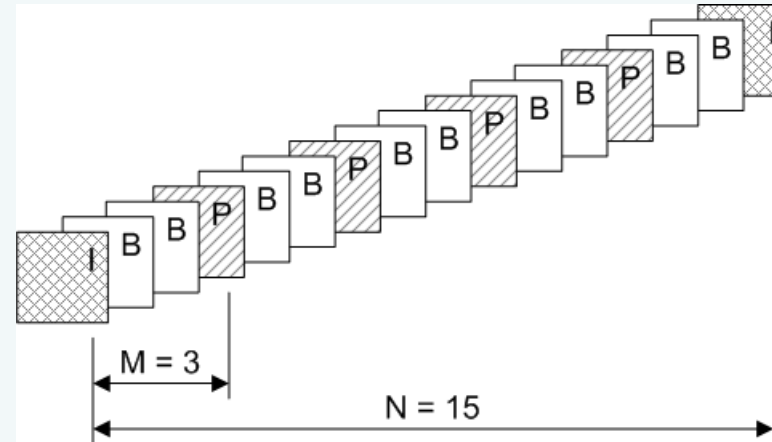
❖ Καρέ-D (D-frames)

- Ενδοπλαισιακή κωδικοποίηση
- Χαμηλότερη ποιότητα από τα καρέ-I
 - Κωδικοποιούνται μόνο οι συνιστώσες DC
- Κωδικοποιούνται επιπλέον των καρέ-I
- Δεν χρησιμοποιούνται για πρόβλεψη
- Γρήγορη κίνηση εμπρός και πίσω
- Δεν χρησιμοποιούνται στις επόμενες εκδόσεις

Ρυθμός δεδομένων

- ❖ Επίτευξη σταθερού ρυθμού δεδομένων
 - Παράγοντας κλιμάκωσης κβαντοποίησης
 - Παρόμοια λογική με H.261
 - Ακέραιος αριθμός ανάμεσα στο 1 και το 31
 - Πολλαπλασιαστής συντελεστών
 - Στο H.261 έχουμε έναν μόνο συντελεστή
 - Παρακολουθεί τη χωρητικότητα του ενταμιευτή
 - Μεταδίδεται σε κάθε αλλαγή τιμής

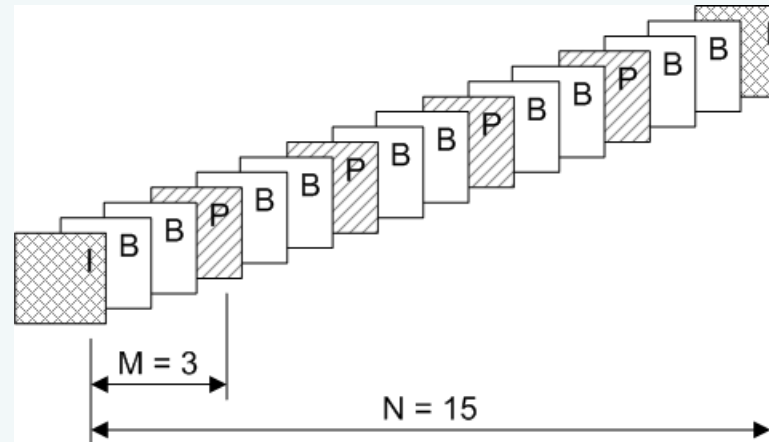
Ομάδα εικόνων



❖ Ομάδα εικόνων (GOP)

- Ξεκινάει με καρέ-I
- Τελειώνει πριν το επόμενο καρέ-I
- Περιλαμβάνει όλα τα ενδιάμεσα καρέ-P και B
- N: ελάχιστη μονάδα τυχαίας προσπέλασης

Ομάδα καρτέ



❖ Ομάδα καρτέ

- Ξεκινάει με καρτέ-I ή P
- Τελειώνει πριν το επόμενο καρτέ-P ή I
- Διάκριση αποκωδικοποίησης και εμφάνισης
- M: ελάχιστη μονάδα αποκωδικοποίησης



Ενταμίευση πλαισίων

- ❖ Πόσο πρέπει να κρατάω κάθε πλαίσιο;
 - Καρέ-I: μέχρι το επόμενο καρέ-P
 - Καρέ-P: μέχρι το επόμενο καρέ-P ή καρέ-I
 - Καρέ-B: καθόλου
 - Δεν χρησιμοποιούνται ποτέ ως καρέ αναφοράς
 - Χρειάζονται όμως το επόμενο P/I για αναφορά

Ακολουθίες καρέ

❖ Συνηθισμένες ακολουθίες καρέ

■ PAL

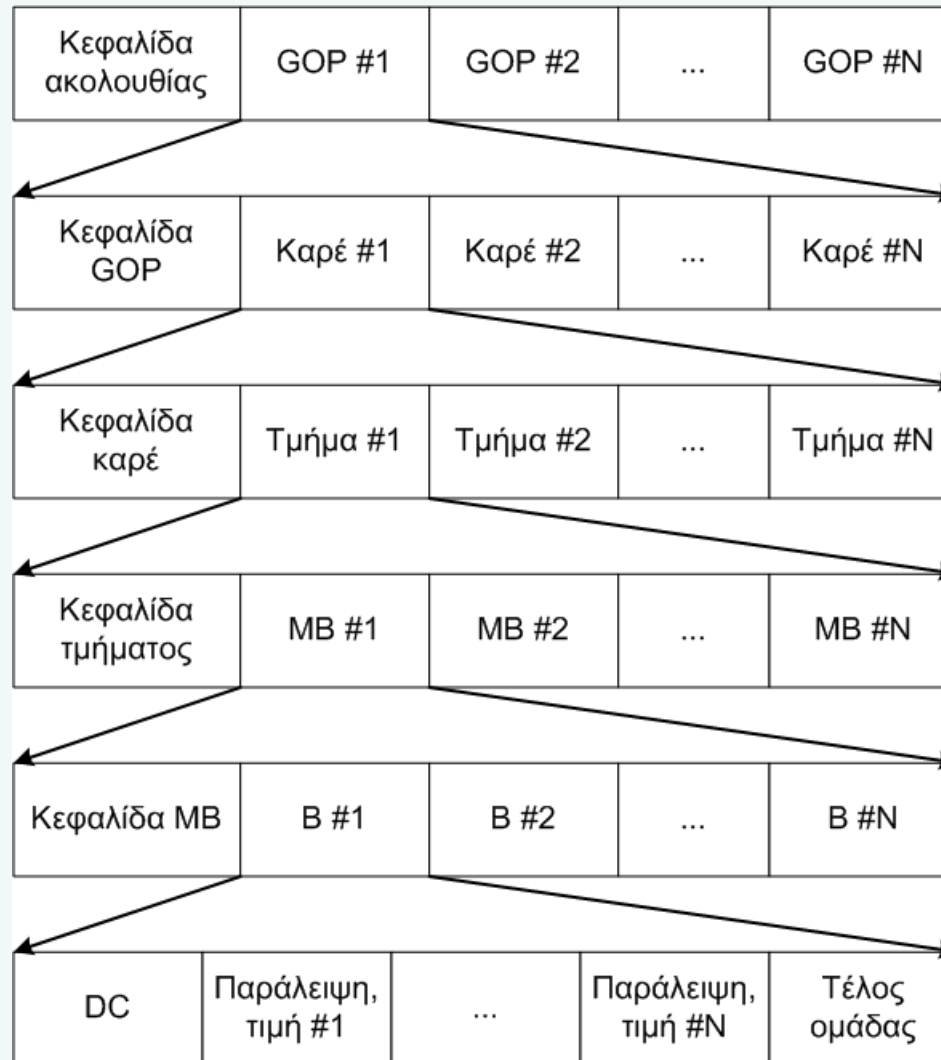
- Εμφάνιση “IBBRBBRBBIBBRBBRBB...”
- Αποκωδικοποίηση “IPBBRBBIBBRBBRBB...”

■ NTSC

- Εμφάνιση “IBBRBBRBBRBBIBBRBBRBBRBB ...”
- Αποκωδικοποίηση “IPBBRBBRBBIBBRBBRBBRBB ...”

■ Μετάδοση με τη σειρά αποκωδικοποίησης

Οργάνωση δεδομένων



Επίπεδα δεδομένων (1 από 2)

- ❖ Ρεύμα δεδομένων βίντεο (6 επίπεδα)
- ❖ Επίπεδο ακολουθίας: ενταμίευση
 - Ρυθμός δεδομένων και χώρος ενταμίευσης
 - Εκτίμηση καθυστέρησης και χώρου
- ❖ Επίπεδο ομάδων εικόνων: προσπέλαση
 - Αποτελείται από ένα τουλάχιστον καρέ-I
- ❖ Δεν υπάρχουν στο H.261



Επίπεδα δεδομένων (2 από 2)

- ❖ Επίπεδο εικόνας: καρτέ
 - Αριθμοί σειράς για συσχέτιση πλαισίων
- ❖ Επίπεδο τμημάτων: τμήματα πλαισίων
- ❖ Επίπεδο μακρομπλόκ: χαρακτηριστικά μακρομπλόκ
- ❖ Επίπεδο μπλοκ: χαρακτηριστικά μπλοκ
- ❖ Παρόμοια με H.261

Πολύπλεξη (1 από 2)

- ❖ Πολύπλεξη βίντεο, ήχου και άλλων μέσων
 - Συνδυασμός ρευμάτων επιμέρους μέσων
 - Διαίρεση ρεύματος κάθε μέσου σε πακέτα
 - Τυπικό μέγεθος 2048 byte
 - Ίσο με έναν τομέα CD-ROM
 - Χρονοσφραγίδες για συγχρονισμό (90 kHz)
 - Στοιχειώδες ρεύμα πακέτων (PES)
 - Προκύπτει από στοιχειώδες ρεύμα (ES) κωδικοποιητή

Πολύπλεξη (2 από 2)

- ❖ Πολύπλεξη ρευμάτων δεδομένων
 - Πολύπλεξη των PES σε πακέτα (racks)
 - Χρονοσφραγίδα πακέτου
 - Συγχρονισμός ρολογιού αποκωδικοποιητή
 - Επικεφαλίδα στο πρώτο πακέτο
 - Μέγιστος ρυθμός δεδομένων, τύποι PES
 - Δυσκολία πρόσβασης για δικτυακά ρεύματα
 - Πρέπει να τα έχουμε διαβάσει από την αρχή