MPEG

by Erol Seke

For the course “**Data Compression**”

ESKİŞEHİR OSMANGAZİ UNIVERSITY

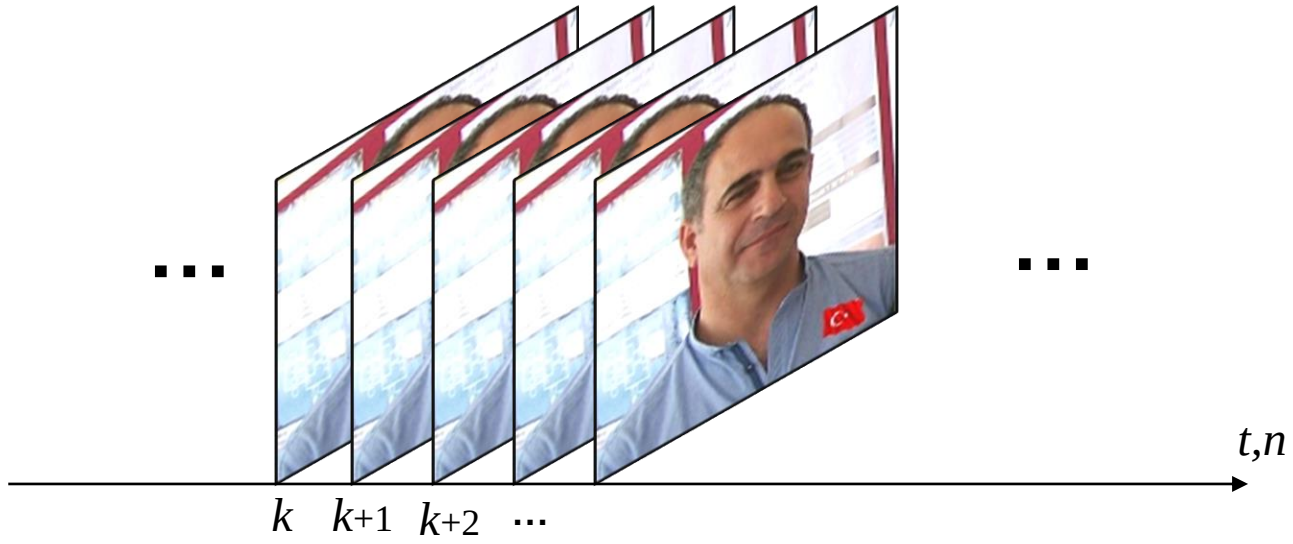
MPEG stands for Motion Picture Experts Group

is a group of Working Groups that aim to develop to set standards for Audio & Video compression

We do aim to discuss neither the group itself nor individual standards developed over the years.

We aim to talk about the underlying approaches used in Video Compression (not the Audio) where together are shortly called MPEG.

Video



Video is generally accepted as a sequence of pictures with slight differences between neighboring ones.
It is assumed that these pictures are captured with small time intervals between them from a moving scene so that when displayed in the correct order and appropriate speed, a human observer will feel like he/she is watching the actual scene.
Human eye/brain is intimidated to fill the gaps between the pictures and object motions/displacements between pictures.
Obviously, video data covering even the shortest time interval will make a lot of data to store/transmit.

Example

Scene duration : 10 seconds

Pictures per second : 30

Picture size : 1024 x 768 pixel

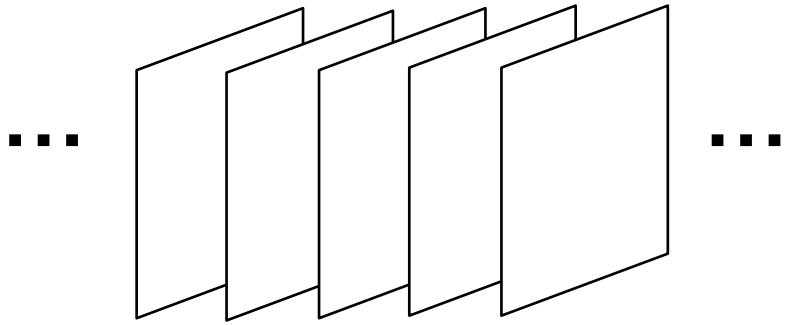
Bits/pixel (assume RGB model, 8 bits per channel) : 24 bits

#bits = 10 x 30 x 1024 x 768 x 24 = 5.662.310.400 bits

Spectral bandwidth coverage = 5.662.310.400 Hz ≈ 5.6 GHz for a baseband signal

≈ 11.3 GHz for a passband signal (BPSK for example)

Simplest Approach with Compression



Compress each picture (frame from now on) using JPEG

Average JPEG compression ratio is told to be 1:10 (no scientific reference)

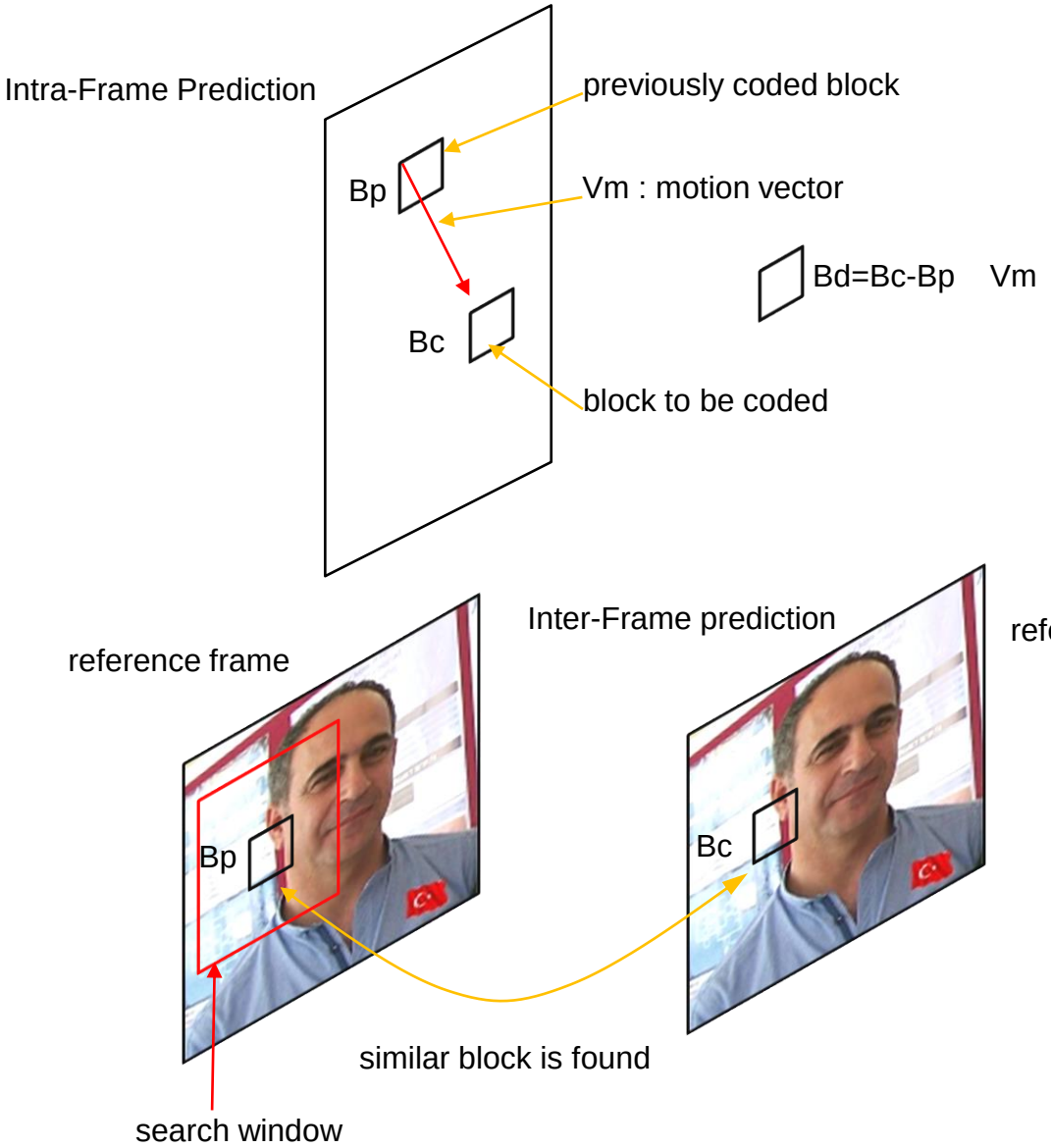
For the previous example

#bits = 566.231.040 bits

This approach do not use similarities/correlations between consecutive frames

It is obvious that,
a lot more compression can be achieved when inter-frame correlations are used.

Inter/Intra-Frame Prediction

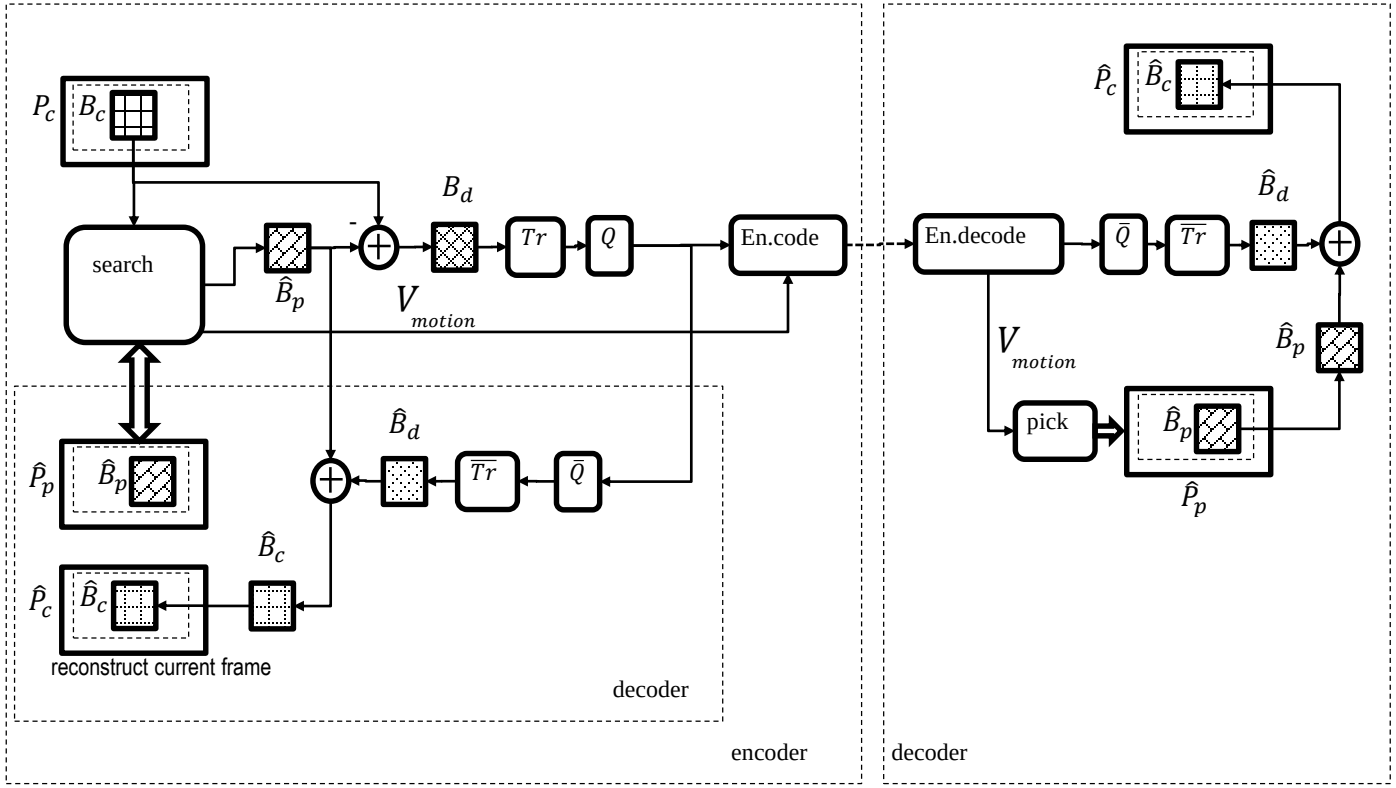


The block Bc is searched within the search window and the difference between the most similar block is lossy coded along with the motion vector

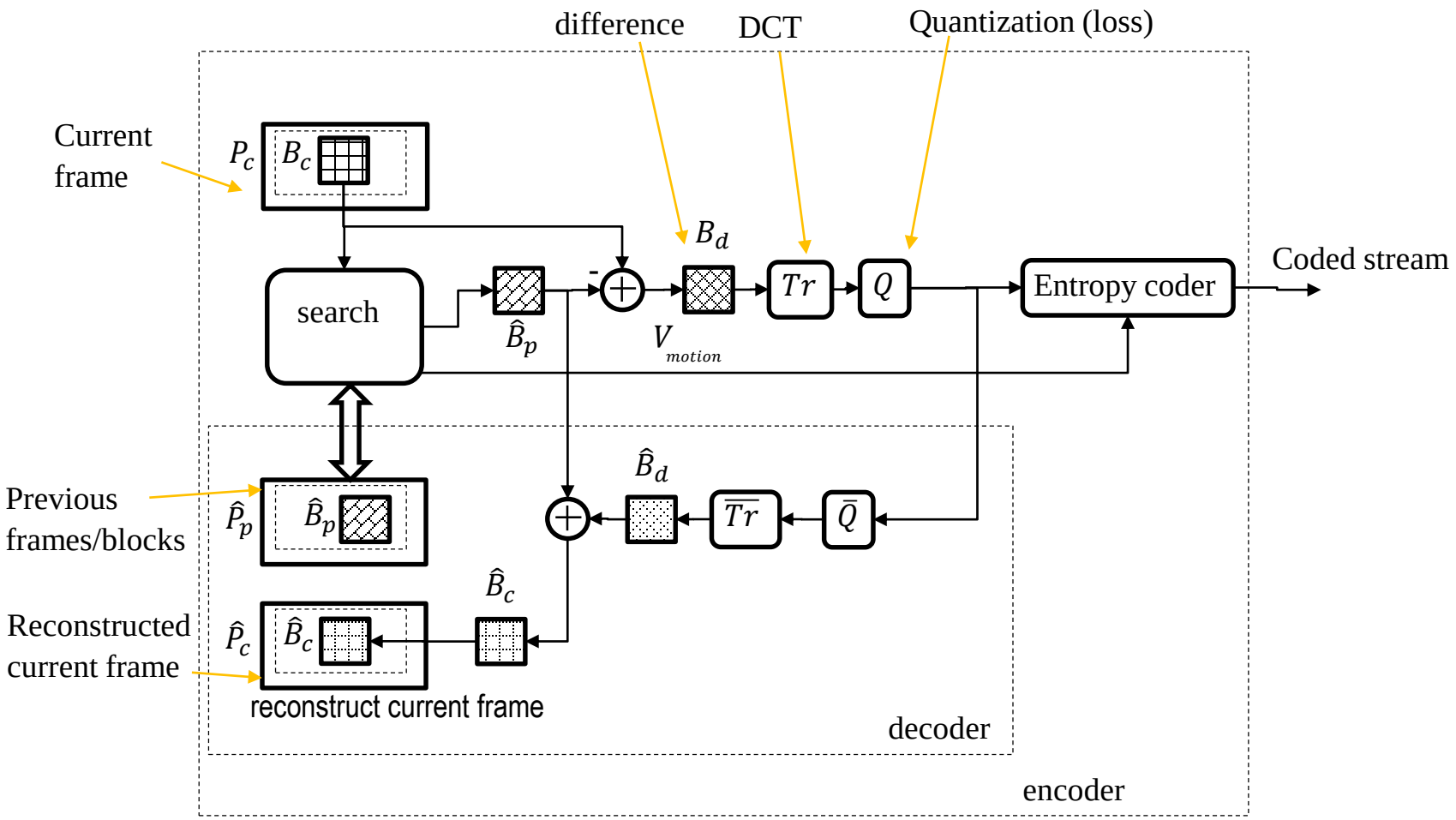
see predictive coding and jpeg blocks

Similar to Intra-Frame, but between different frames.

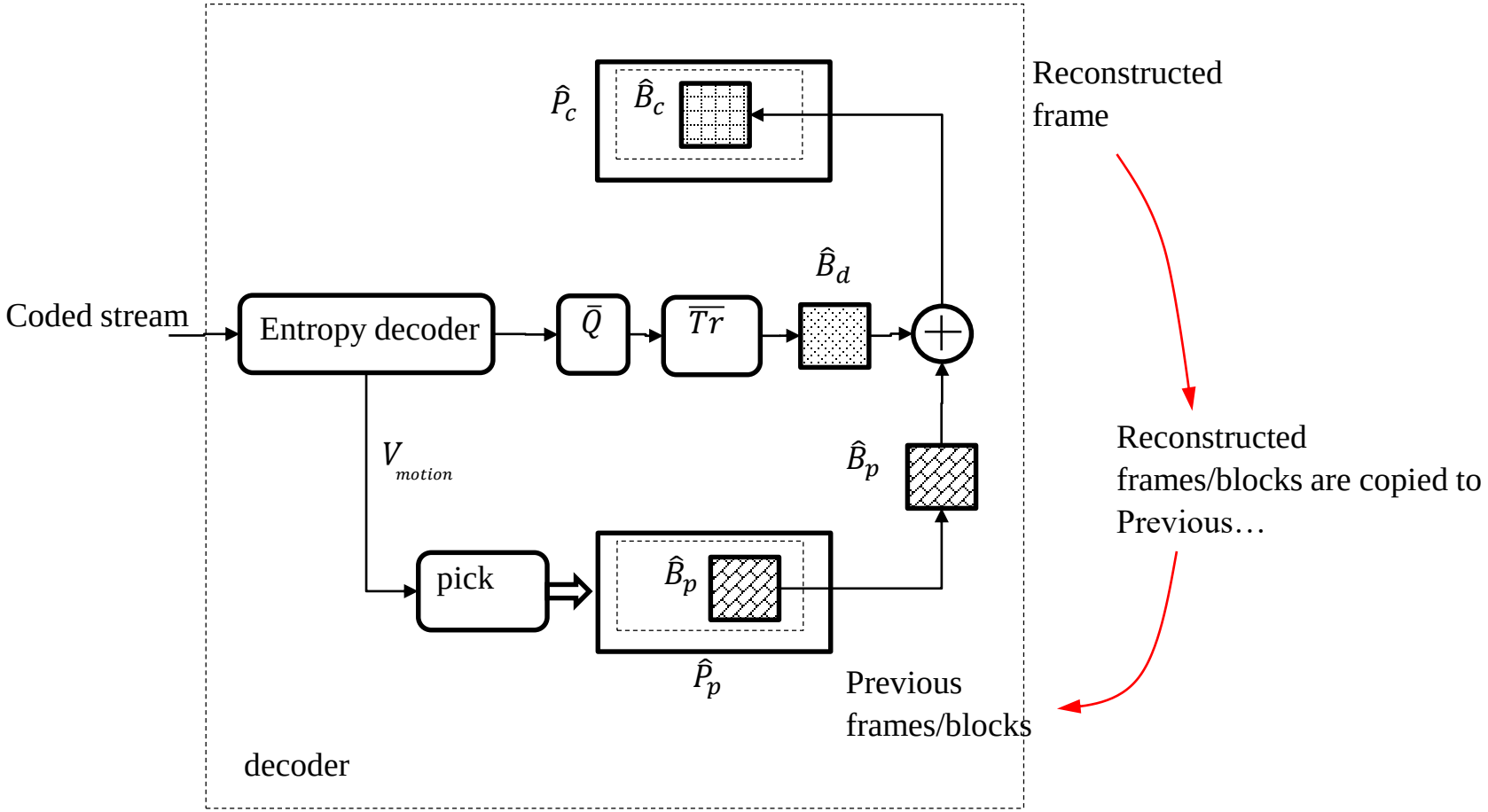
MPEG4 (H264)


P_c
B_c
search
$\hat{B}_p$
B_d
Tr
Q
En.code
V_{motion}
$\hat{P}_p$
$\hat{B}_p$
$\hat{B}_d$
$\overline{Tr}$
$\overline{Q}$
$\hat{B}_c$
$\hat{P}_c$
$\hat{B}_c$
reconstruct current frame
decoder
encoder
En.decode
pick
decoder

Encoder

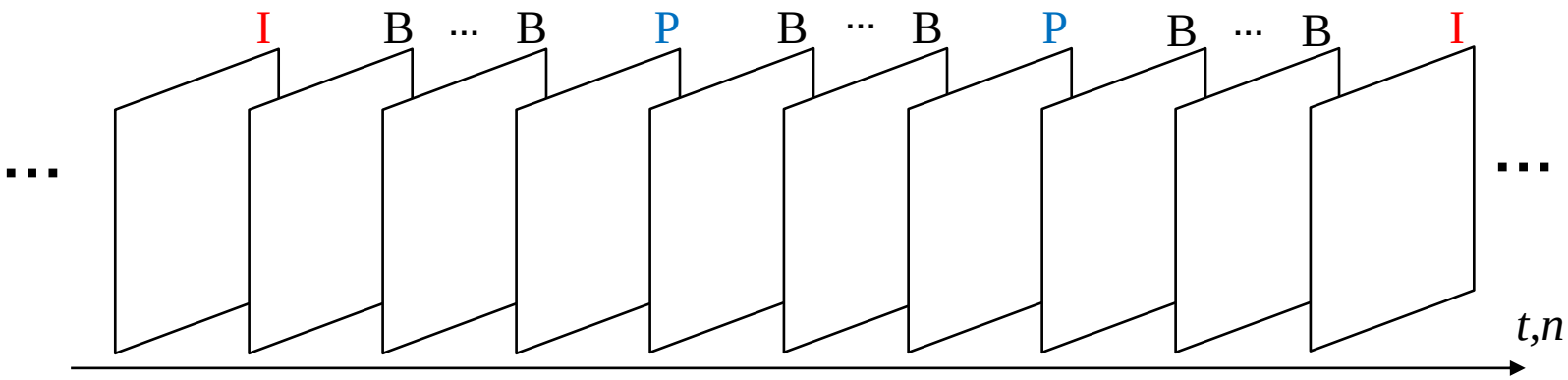

difference
DCT
Quantization (loss)
Current frame
P_c
B_c
B_d
search
$\hat{B}_p$
V_{motion}
Tr
Q
Entropy coder
Coded stream
Previous frames/blocks
$\hat{P}_p$
$\hat{B}_p$
$\hat{B}_d$
$\overline{Tr}$
$\bar{Q}$
$\hat{B}_c$
Reconstructed current frame
$\hat{P}_c$
$\hat{B}_c$
reconstruct current frame
decoder
encoder

Decoder



decoder is identical to “decoder” section in the encoder

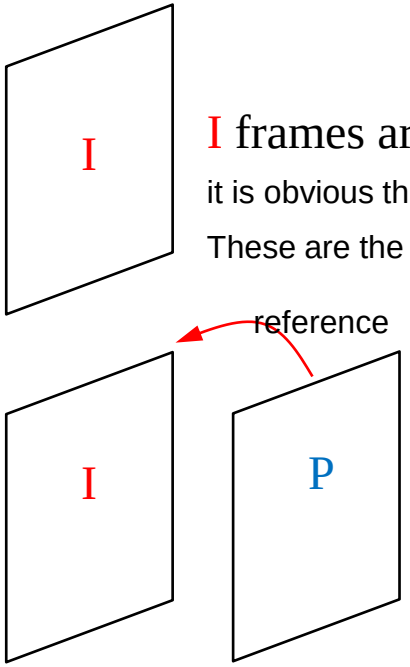
GOP : Group of Pictures



I : intra frames are coded only using blocks within the same frame (no reference to other frames)

P : predicted frames are coded using blocks within the previous frames or current frame

B : bidirectional frames are coded using blocks within all frames in the vicinity



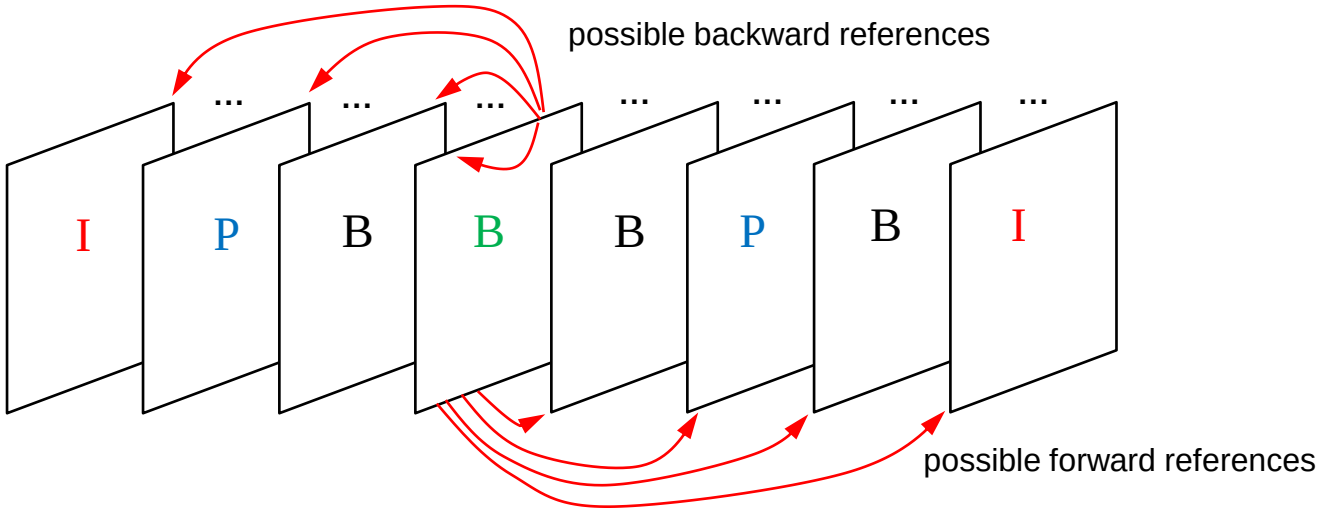
I frames are coded only by itself (no reference to other frames)

it is obvious that, within the current frame, only the previous macroblocks can be used for prediction

These are the key-frames that can be used to jump to the desired point within a large video file

Blocks in P frames are;

- predicted using blocks in previous I frame
- predicted using blocks within current frame
- not predicted (coded alone)



Blocks in B frames can be;

- predicted from previous frames within GOP
- predicted using future frames within GOP
- predicted using frames in both direction (interpolation)
- not predicted, coded by itself.

If GOP includes future referencing B frames, reference frames must be kept in memory for correct order of display.
All referenced frames must be kept in memory until referencing frame is coded.
Previous frames can be deleted from memory when all referencing frames are coded.
On decoder side, previous frames can be deleted from memory when it is displayed and all referencing frames are decoded.

END