Introduction to Multimedia Computing

Video Media Type

Topics

- Basics of Digital Videos
- Video Representation and Compression
  - Temporal Redundancy
  - Inter-frame and Intra-frame coding
  - Forward Motion Prediction
  - Bi-directional Motion Prediction
  - Motion Compensated Coding
  - MPEG

Basics of Digital Videos

- Digital videos are a sequence of frames
- At least 16 frames per second are needed for a smooth video (generally 30-35 frames per second are used in high quality videos)
- Each frame is a 2D matrix of pixels

Example Video Size

- One frame of a high quality video can have 3112 rows and 4096 columns
- Each color pixel needs 24 bits.
- 30 frames are stored for 1 second video
- As a result, 1 sec. of video takes more than 1 Gbytes ! (More than the capacity of a CD)

HDTV Example

- HDTV video may have 720 rows and 1280 columns.
- Each pixels uses 24 bits.
- The data rate corresponding to a frame rate of 60 frames/sec is 720 x 1280 x 3 x 60 = 165 Mbytes/sec! (too much for real time transmission over the Internet)

The Need for Compression

- Video data requires much higher rate of compression.
- Storing and transmitting un-compressed video over the Internet (video on demand) is almost impossible
- Beside the redundancy types available in still images, video has temporal redundancy

Approaches to Video Compression

- Intra-frame compression considers each frame of a video as a still image.

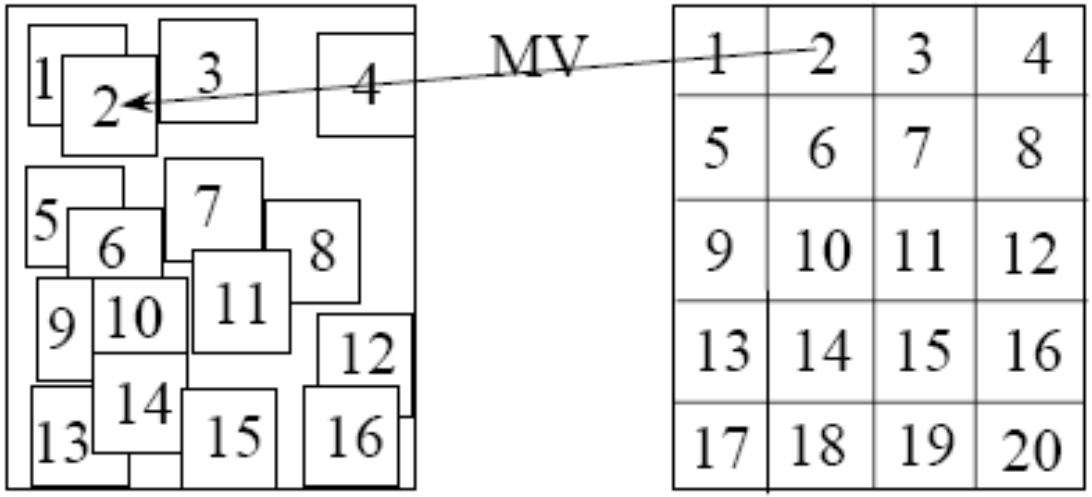
- Inter-frame compression uses temporal redundancy predictions.

Using Temporal Redundancy

- The frames of a video are very similar if the time interval between them is short.

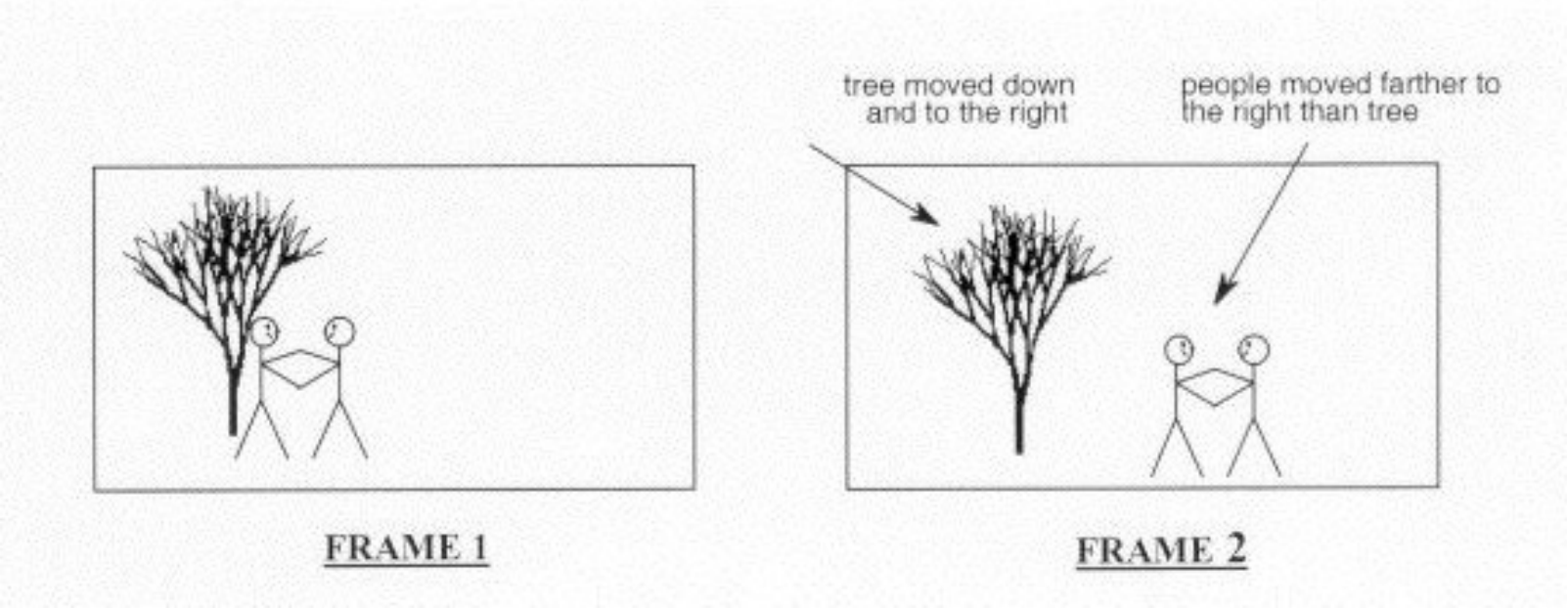
Using Temporal Redundancy

- The objects in the first frame are slightly moved to a new place.

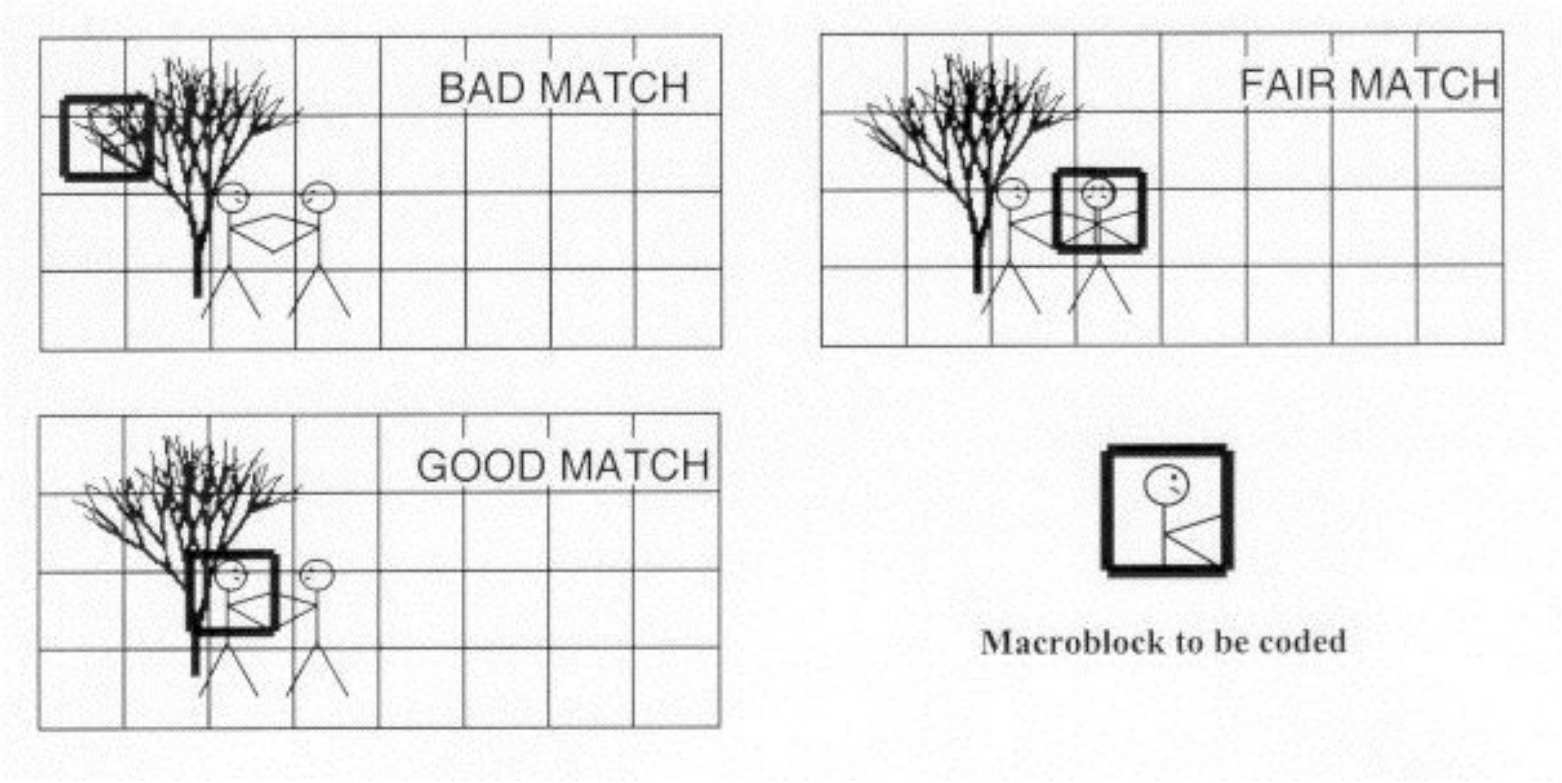


Using Temporal Redundancy

- Solution to the object displacement is estimating the motion before comparing.



Motion Estimation by Finding Similar Region for Each Block



Similarity Measurement

- Sum of Absolute Difference (SAD) is generally used for similarity measurement.

- SAD is defined as:

$$SAD = \sum_i \sum_j | B_i - R_j |$$

where B is the block and R is the reference image

Compaction Property of DCT

- If the values of a block are small, its DCT coefficients will be mostly zero
- Conclusion: Instead of encoding the block data, we should encode its difference with the block from a previous frame.
- To get smaller difference values, compare each block with its most similar region in a previous frame.

Motion Compensation

- If two blocks in two frames contain very similar data, we can encode the difference instead of the data itself.

122	131
144	130

Block from frame 1

121	131
143	126

Block from frame 2

122	131
144	130

Block from frame 1

1	0
1	4

Difference between the blocks

Compaction Property of DCT

– Example: An 8 x 8 block of the Lena image.

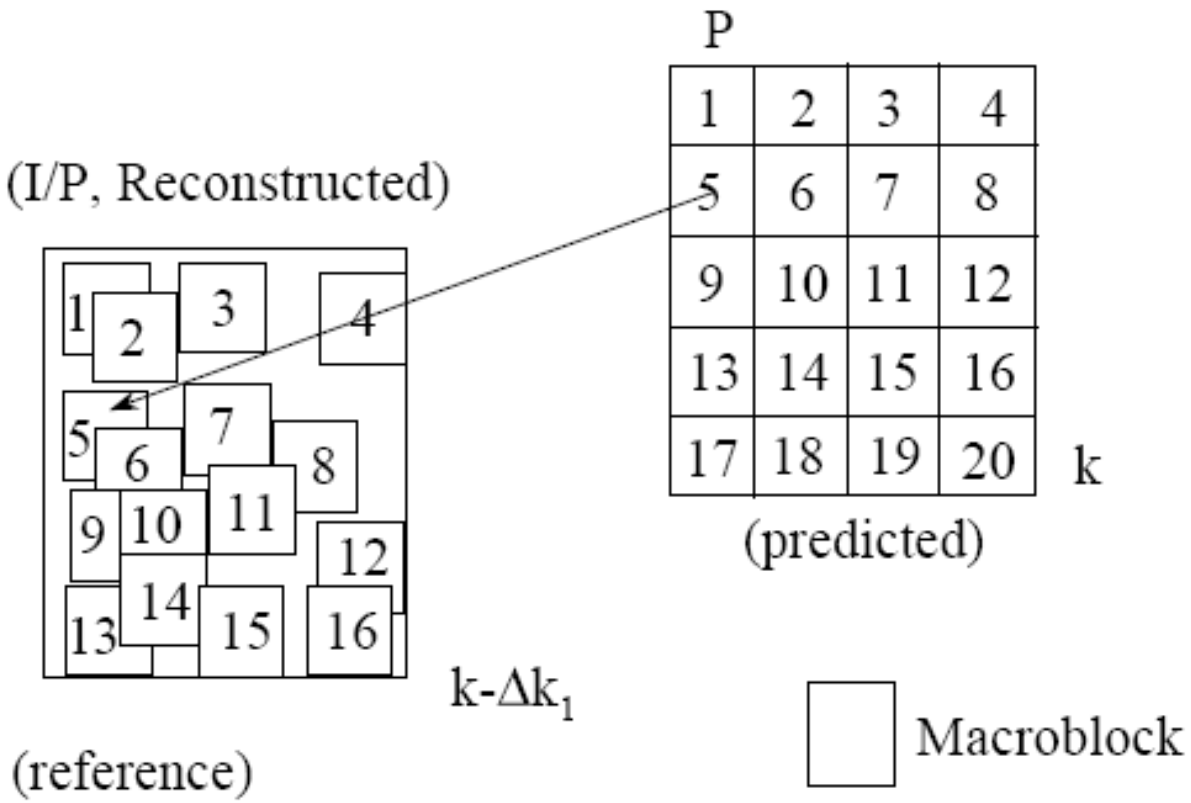
Image Block

139	144	149	153	155	155	155	155
144	151	153	156	159	156	156	156
150	155	160	163	158	156	156	156
159	161	162	160	159	159	158	159
159	160	161	162	162	155	155	155
161	161	161	161	160	157	157	157
162	162	161	163	162	157	157	157
162	162	161	161	163	158	158	158

Rounded DCT coefficients of the block

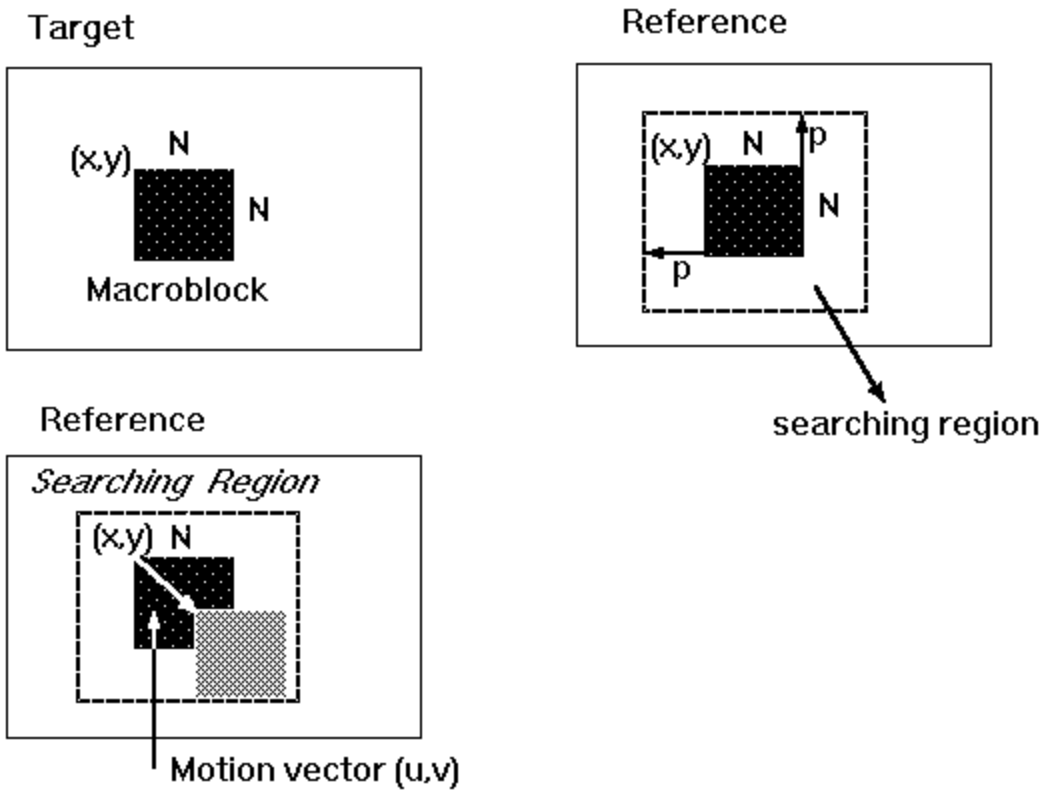
1260	-1	-12	-5	2	-2	-3	1
-23	-17	-6	-3	-3	0	0	-1
-11	9	-2	2	0	-1	-1	0
-7	-2	0	1	1	0	0	0
-1	-1	1	2	0	-1	1	1
2	0	2	0	-1	1	1	-1
-1	0	0	-1	0	2	1	-1
-3	2	-4	-2	2	1	-1	0

Encoding with Motion Prediction

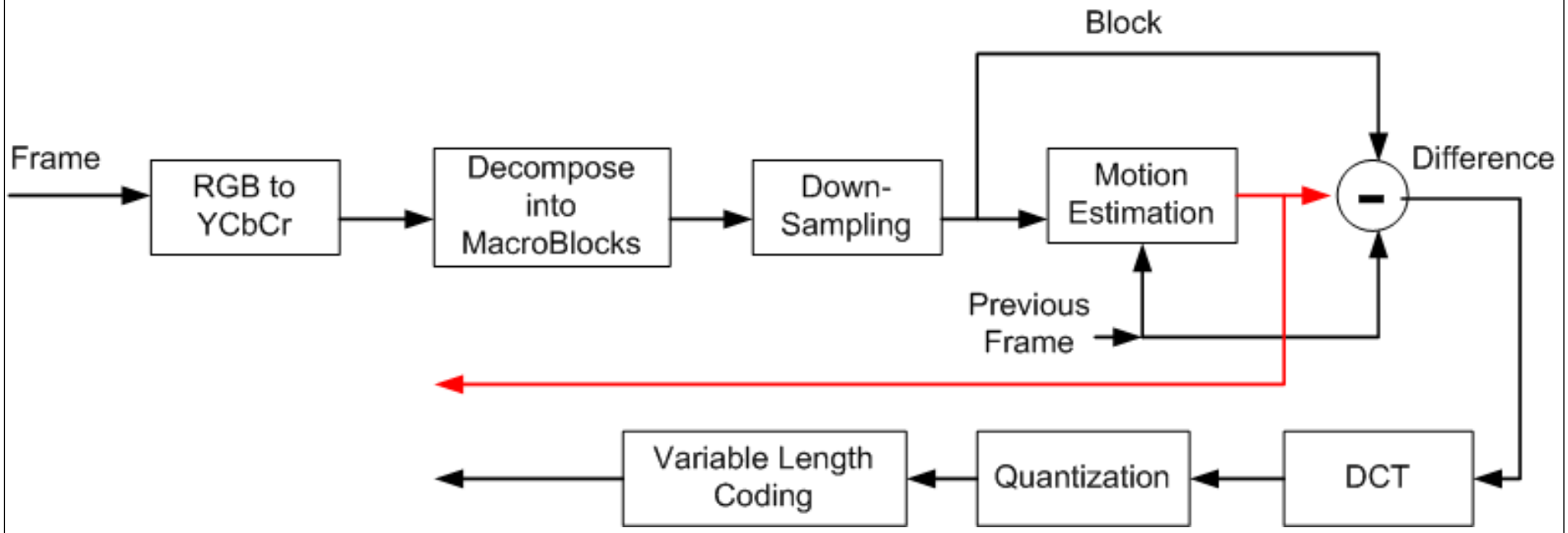


Motion Vectors

- For each block, a vector called motion vector, shows the displacement needed to reach the most similar region in the reference frame



Video Coding Block Diagram



MPEG Standard

- The Moving Picture Experts Group (MPEG) is a working group that was formed by ISO and IEC to set standards for audio and video compression and transmission in 1988.
- The MPEG standards has different *Parts*. Each *part* has different properties.
- MPEG-1 (1993)
- MPEG-2 (1995)
- MPEG-3 : Merged with MPEG-2
- MPEG-4 (1998): Most recent version

Frame Types in MPEG

- MPEG is the most commonly used standard for video coding.

- MPEG defines three frame types
  - P frames
  - B frames
  - I frames

P Frames

- The frame is split into macro-blocks
- The pixel colors are converted into YCbCr
- Macro blocks are down-sampled
- For each block after down-sampling, the most similar region in the previous frame is found.
- The difference of the block with the most similar region is computed and DCT transformed.
- The DCT coefficients are quantized, and entropy encoded

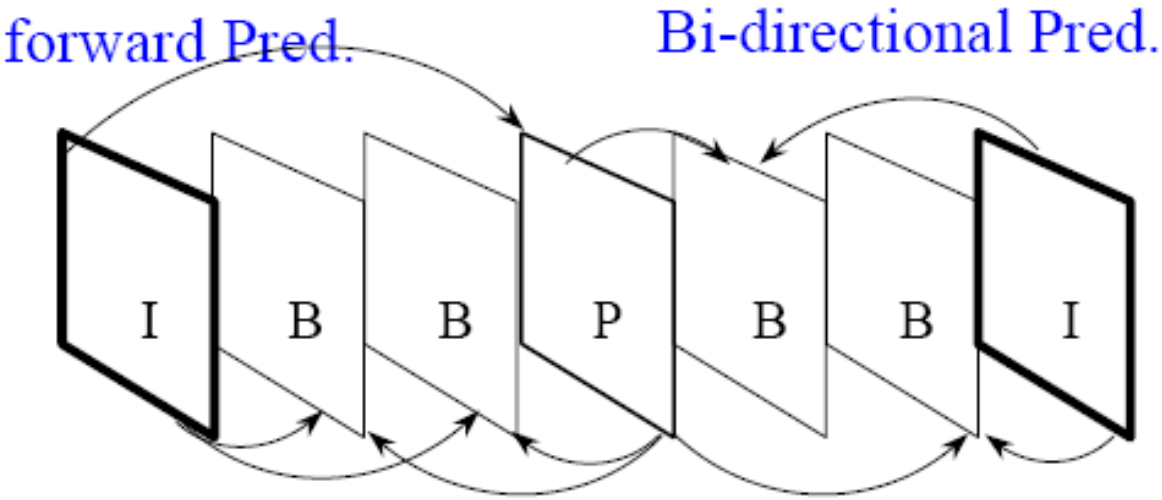
B Frames

- In B frames, each block is compared to its previous and next frames.
- After finding the most similar regions, the differences are found.
- Differences from the previous frame and the next frames are averaged [(Diff1 + Diff2)/2]
- B frames give smaller coefficients than P frames.

I Frames and Group of Pictures (GOP)

- I frames are encoded independently (without comparing to any other frame)
- Frames of a video are grouped as one I frame followed by some P or B frames. This is called a Group Of Pictures (GOP)
- I frame is the reference frame for the first P frame.
- A P frame can be a reference frame for another P frame but B frames are not used as reference frames (MPEG 1 and 2)

Group of Pictures



Summary

- Videos are sequences of two dimensional frames. Each frame is a matrix of pixels.
- To compress videos we make use of the temporal redundancies in the videos.
- MPEG is a standard to store videos.

Questions?