Lecture notes on Image Compression and Video Compression

7. Video Compression

Topics

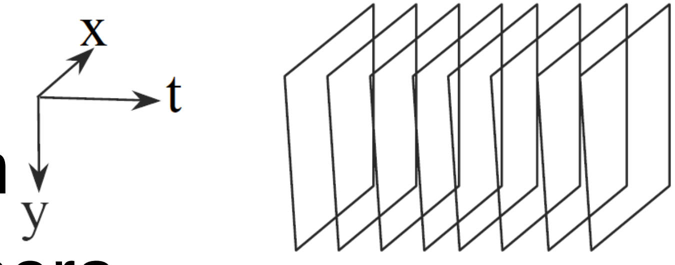
- Introduction to Image Compression
- Transform Coding
- Subband Coding, Filter Banks
- Introduction to Wavelet Transform
- Wavelet Image Compression
- Perceptual Audio Coding
- Video Compression

Contents

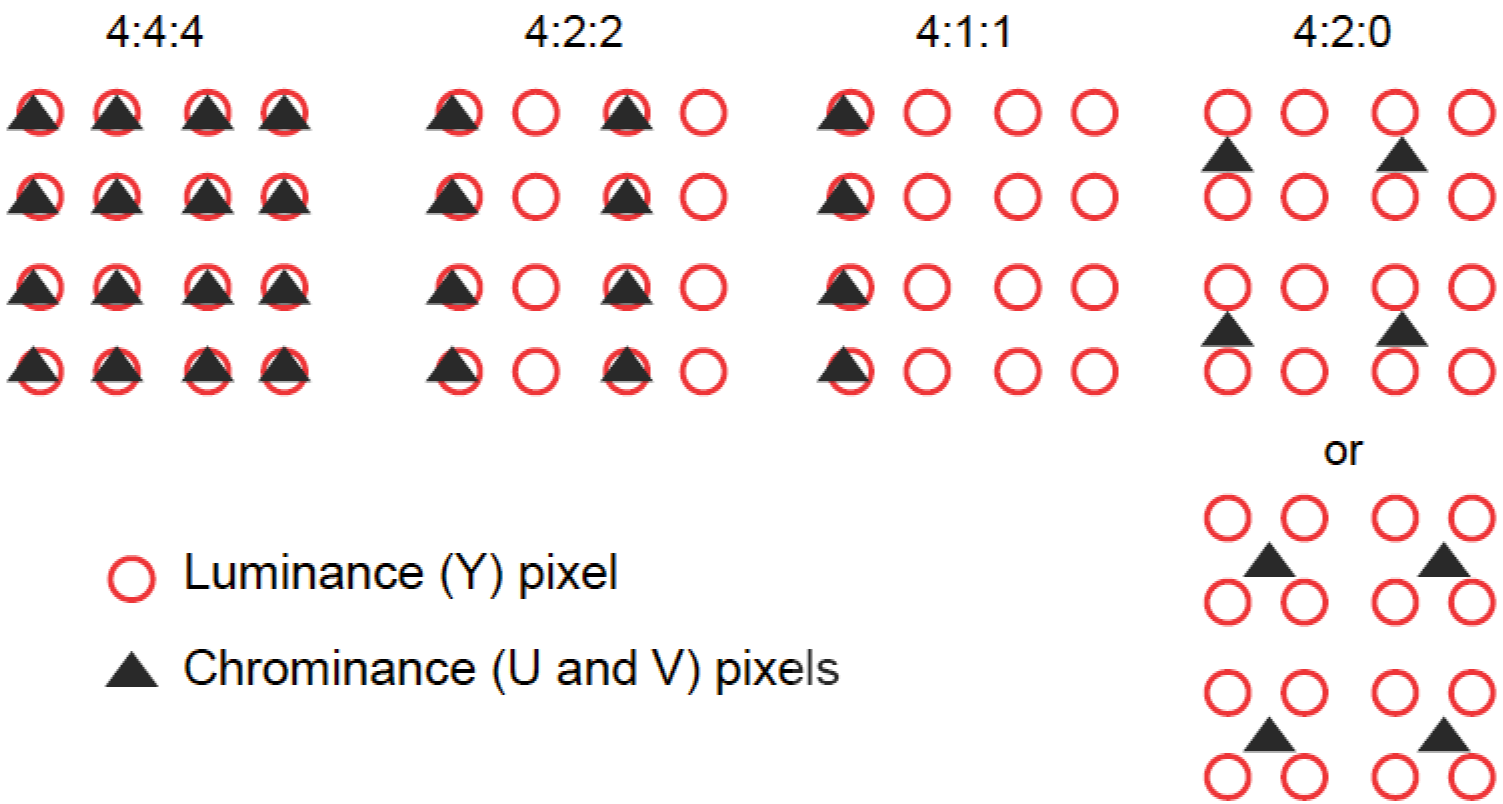
- Introduction to Video Compression
- MPEG 1 Video Compression
- Motion Compensation
- Video Compression Standards

Digital Video

- A series of frames (i.e. digital images)
- Can be sampled
  - From raster analog scan
  - Directly with a CCD camera
- 3 color components
- Usually, color components down-sampled



Color Down-sampling



The Need for Compression

- Image sequences must be significantly compressed for efficient storage and transmission as well as for efficient data transfer among various components of a video system.
- Examples (Motion Picture)
  - One frame of a Super 35 format motion picture may be digitized (via Telecine equipment) to a 3112 lines by 4096 pixels/color, 10 bits/color image.
  - As a result, 1 sec. of the movie takes ~1 Gbytes !

The Need for Compression - Examples

- HDTV
  - A typical progressive scan (non-interlaced) HDTV sequence may have 720 lines and 1280 pixels with 8 bits per luminance and chroma channels.
  - The data rate corresponding to a frame rate of 60 frames/sec is 720 x 1280 x 3 x 60 = 165 Mbytes/sec!

Applications of Digital Video Compression

- Teleconference or video phone
  - Very low delay (1/10 second is a standard)
- Live Broadcast Video
  - Modest delay is tolerable (seconds is normal)
  - Error tolerance is needed.
- Video-in-a-can (DVD, Video-on-Demand)
  - Random access to compressed data is desired
  - Encoding can take a lot of time
- Decoding must always be at least the frame rate.

Approaches to Video Compression

- Intraframe compression treats each frame of an image sequence as a still image.
  - Intraframe compression, when applied to image sequences, reduces only the spatial redundancies present in an image sequence.
- Interframe compression employs temporal predictions and thus aims to reduce temporal as well as spatial redundancies, increasing the efficiency of data compression.
  - Example: Temporal motion-compensated predictive compression.

Why can Images be Compressed?

- Image compression can be achieved primarily because image data are highly redundant.
- The degree of redundancy determines how much compression can be achieved.
- Four types of redundancy can be identified:
  - Spatial Redundancy
    - Correlation between adjacent data points
  - Temporal Redundancy
    - Correlation between different frames in an image
  - Spectral Redundancy
    - Correlation between different color planes or sensors
    - Limitation of Low-level Human Vision System
  - Psycho-visual Redundancy
    - Limitation of high-level Human Vision System

Human Perception of Video

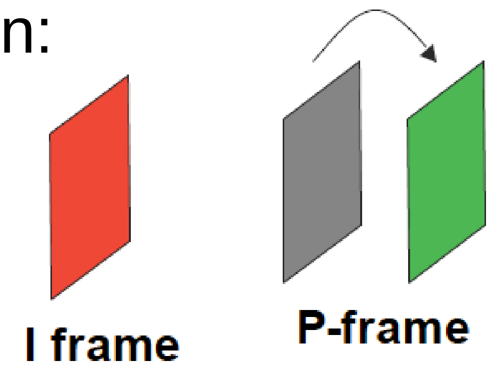
- 30 frames per second seems to allow the visual system to integrate the discrete frames into continuous perception.
- If distorted, nearby frames in the same scene should have <u>only</u> small details wrong.
  - A difference in average intensity is noticeable
- Compression choice when reducing bit rate
  - skipped frames cause stop action
  - lower fidelity frames may be better

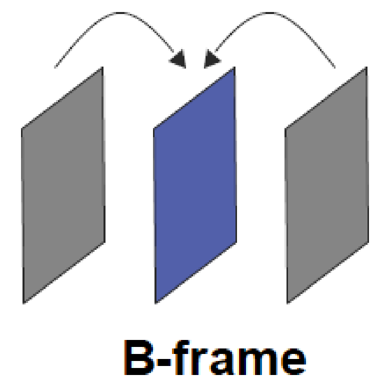
High Compression Ratios Possible

- Nearby frames are highly correlated. Use the previous frame to predict the current one.
- Need to take advantage of the fact that usually objects move very little in 1/30th of a second.
  - Video coders use motion compensation as part of prediction.

Video Compression

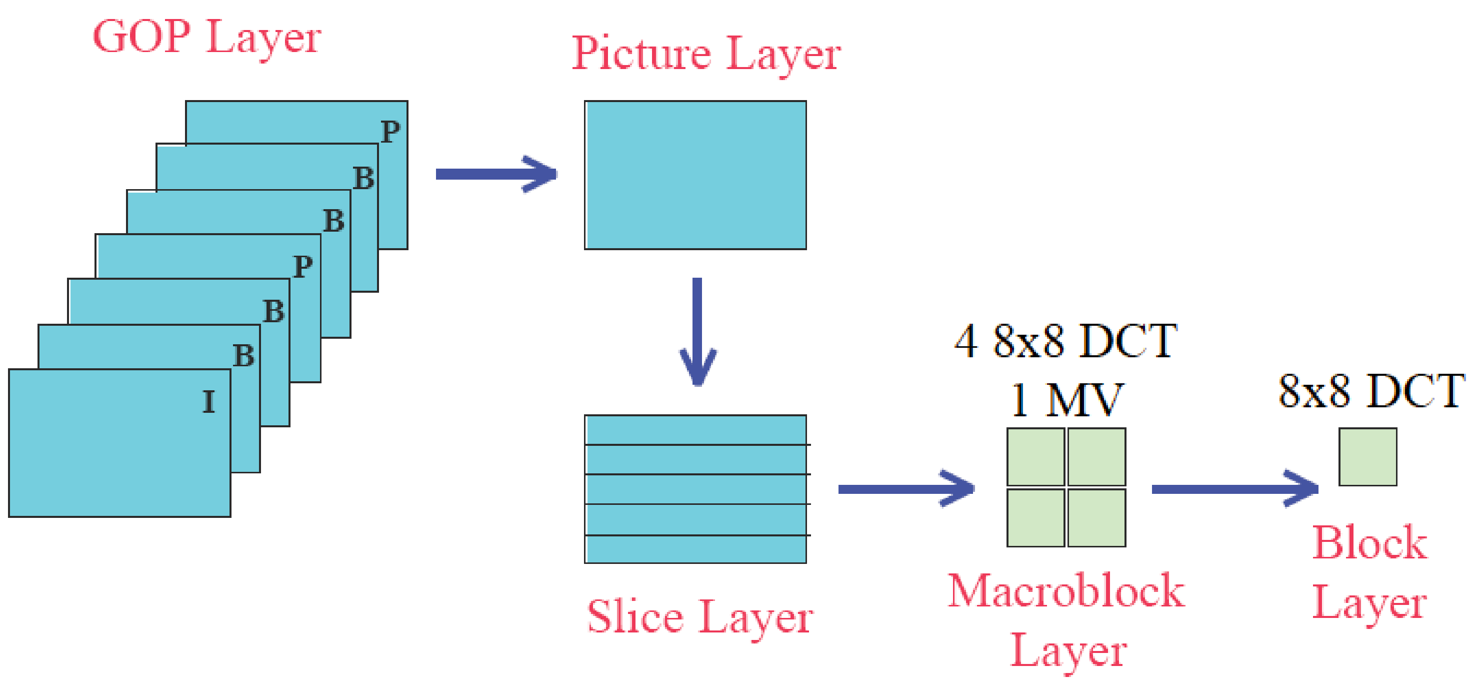
- Main addition over image compression:
  - Exploit the temporal redundancy
- *Predict current frame* based on previously coded frames
- Three types of coded frames:
  - ***I-frame:*** Intra-coded frame, coded independently of all other frames
  - ***P-frame:*** Predictively coded frame, coded based on previously coded frame
  - ***B-frame:*** Bi-directionally predicted frame, coded based on both previous and future coded frames





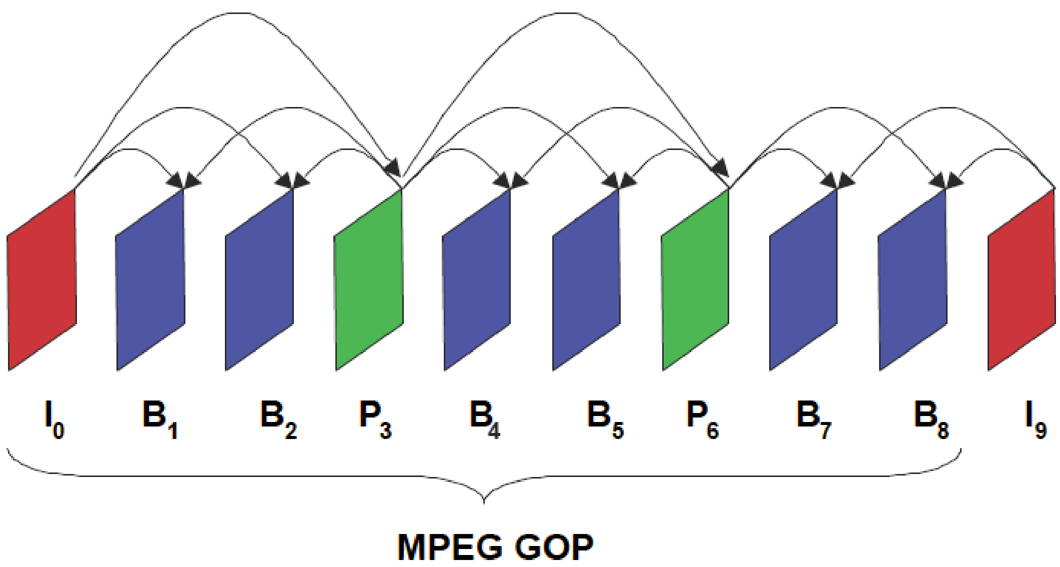
MPEG 1 Structure

- MPEG codes video in a hierarchy of layers.



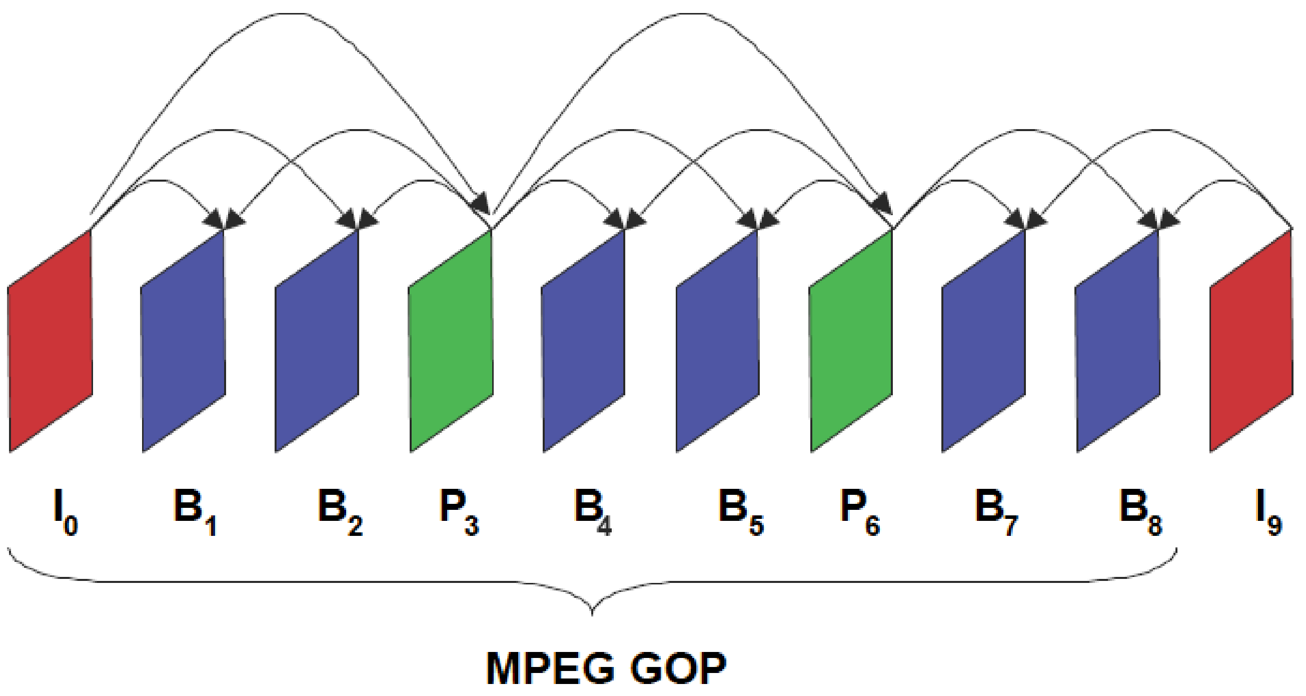
MPEG 1 Group of Pictures (GOP) Structure

- Composed of I, P, and B frames
- Periodic I-frames enable random access into the coded bit stream.
- Parameters: (1) Spacing between I frames (2) number of B frames between I and P frames



Example Use of I-,P-,B-frames: MPEG 1 Group of Pictures (GOP)

- Arrows show prediction dependencies between frames



Definition of P Frames

- P Pictures are composed of macroblocks that are either
  - forward predictive (non-intra) coded . Or
  - intra coded (using the same quantization and VLC as macroblocks of the I pictures).
- The encoder is allowed to make an Intra/Non-Intra decision depending on the accuracy of the prediction.
- This decision can be made in many different ways and the choice is up to the encoder.

P Frames

- A possible simple decision mechanism compares the variance of the luminance component of the original macroblock with that of the prediction error macroblock.
- If the variance of the prediction error macroblock is higher, then the macroblock is intra coded.

Definition of B Pictures

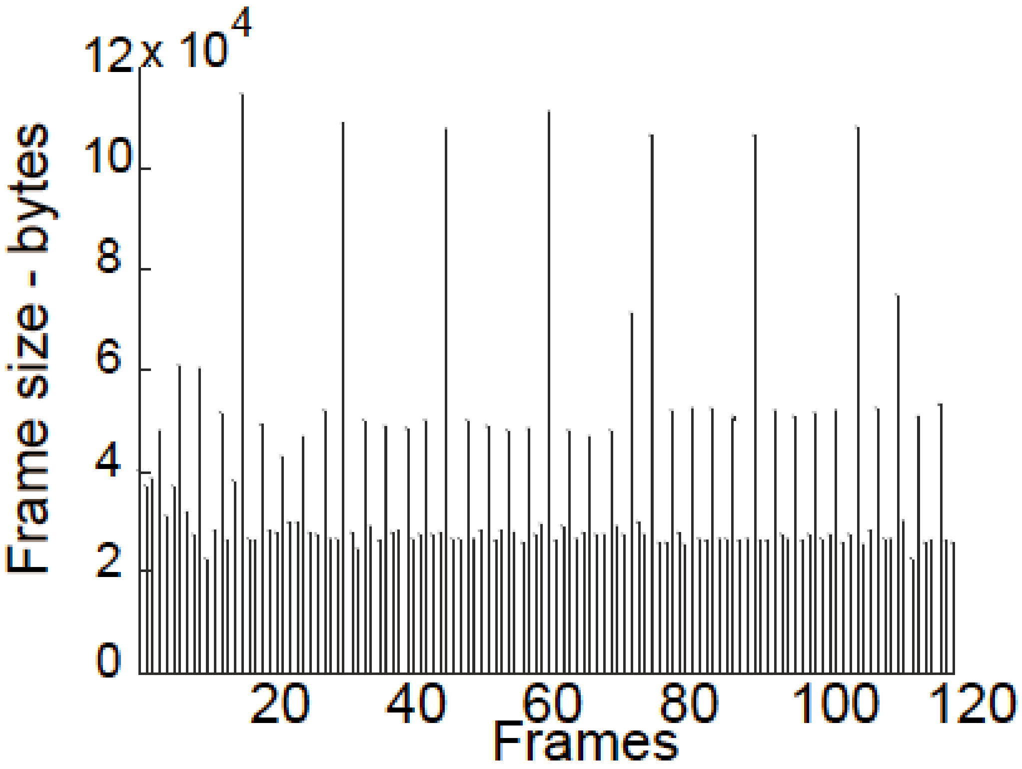
- B Pictures are composed of macroblocks that are
  - bi-directional predictive coded, or
  - backward predictive coded, or
  - forward predictive coded, or
  - intra coded
- A possible decision mechanism is picking the mode that results in the least macroblock (luminance component) variance.
- The macroblocks in the B pictures are not used as references.

MPEG 1

- Relative number of (I), (P), and (B) pictures can be arbitrary.
  - It depends on the nature of the application.
  - For instance it depends on fast access and compression ratio requirements; relatively smaller amount of compression is expected to be achieved at (I) pictures compared to (P) and (B) pictures.
- The (B) pictures are expected to provide relatively the largest amount of compression under favorable predictability conditions.

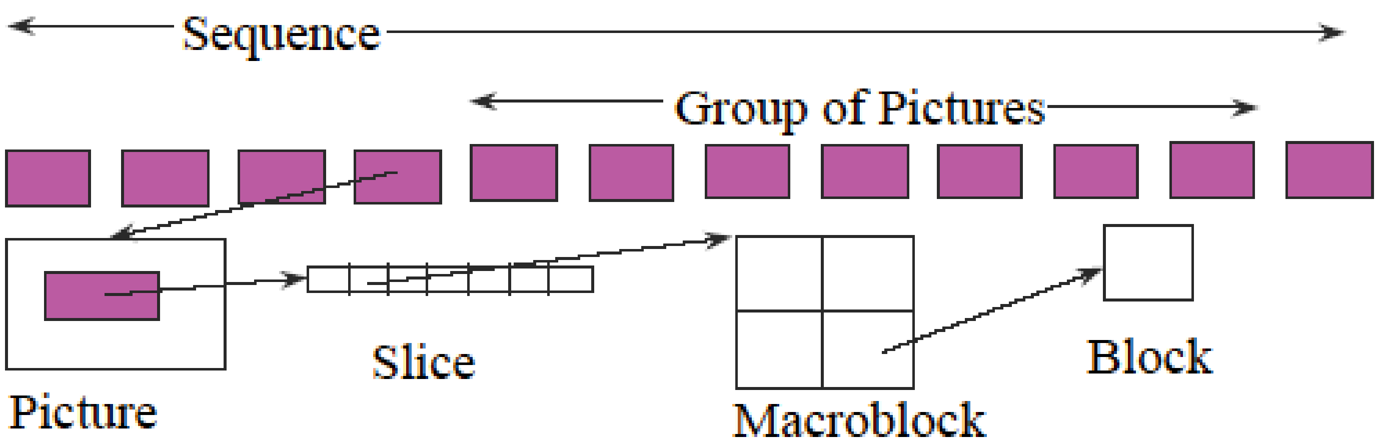
MPEG 1 Frame Size

- Size of I, P, B frames



MPEG-1- Bit Stream Hierarchy

- MPEG-1- Bit Stream Hierarchy



The Bit Stream

Sequence 1 Level	GOP Level	Picture Level	Slice Level	MB Level	Block Level	Sequence 2 Level	

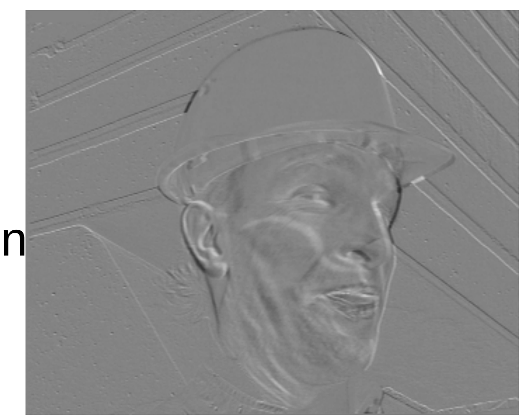
Temporal Redundancy

- Adjacent frames are highly correlated.

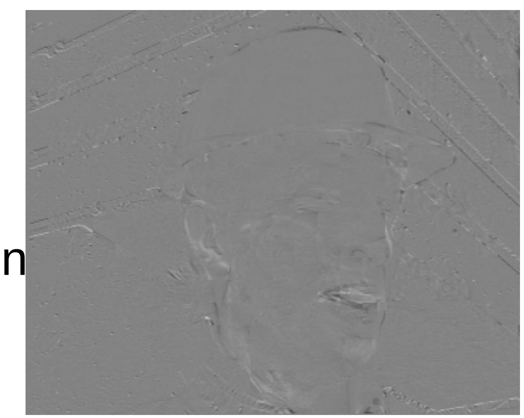
Frame 7

Frame 8

No motion compensation



With motion compensation



Motion Compensation

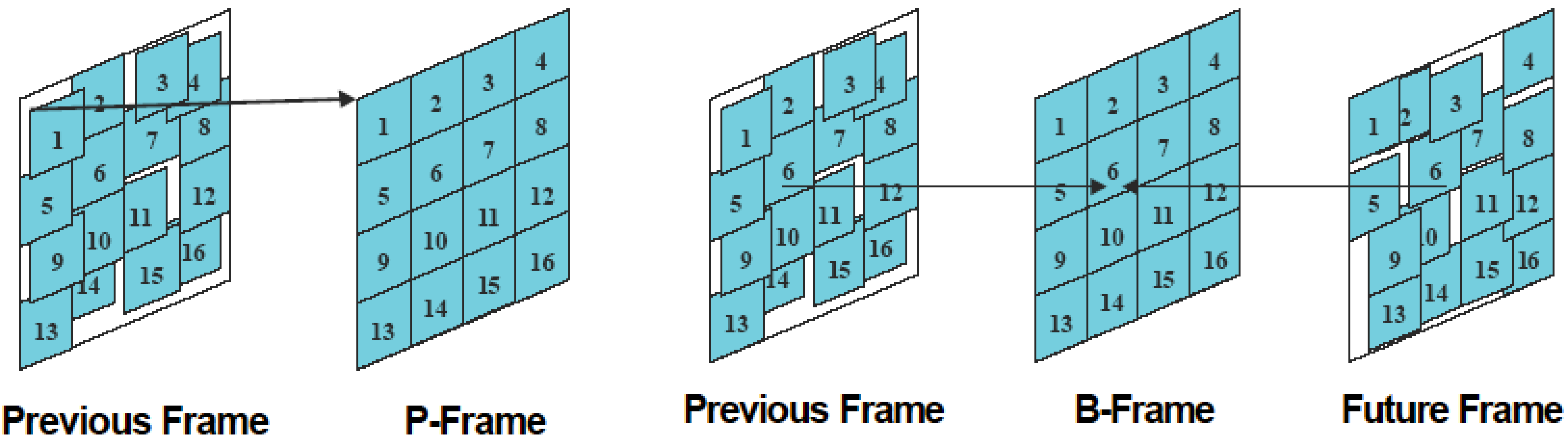
- In general, we speak of motion of objects in 3-D real world.
  - Here, we are concerned with the "projected motion" of 3-D objects onto the 2-D plane of an imaging sensor.
  - By motion estimation, we mean the estimation of the displacement (or velocity) of image structures from one frame to another in a time sequence of 2-D images.
- In the literature, this projected motion is referred to as "apparent motion", "2-D image motion", or "optical flow".
- The detail algorithms on "optical flow“ is beyond the content of this course.

Motion Compensation

- Predict the current frame based on reference frames while compensating for the motion.
- For each motion compensation block
  - Find the block in the reference decoded frame that gives the least distortion.
  - If the distortion is too high then code the block independently. (intra block)
  - Otherwise code the difference (inter block)

MC-Prediction and Bi-Directional MC-Prediction (P- and B-frames)

- Examples of block-based motion-compensated prediction (*P-frame*) and bi-directional prediction (*B-frame*)

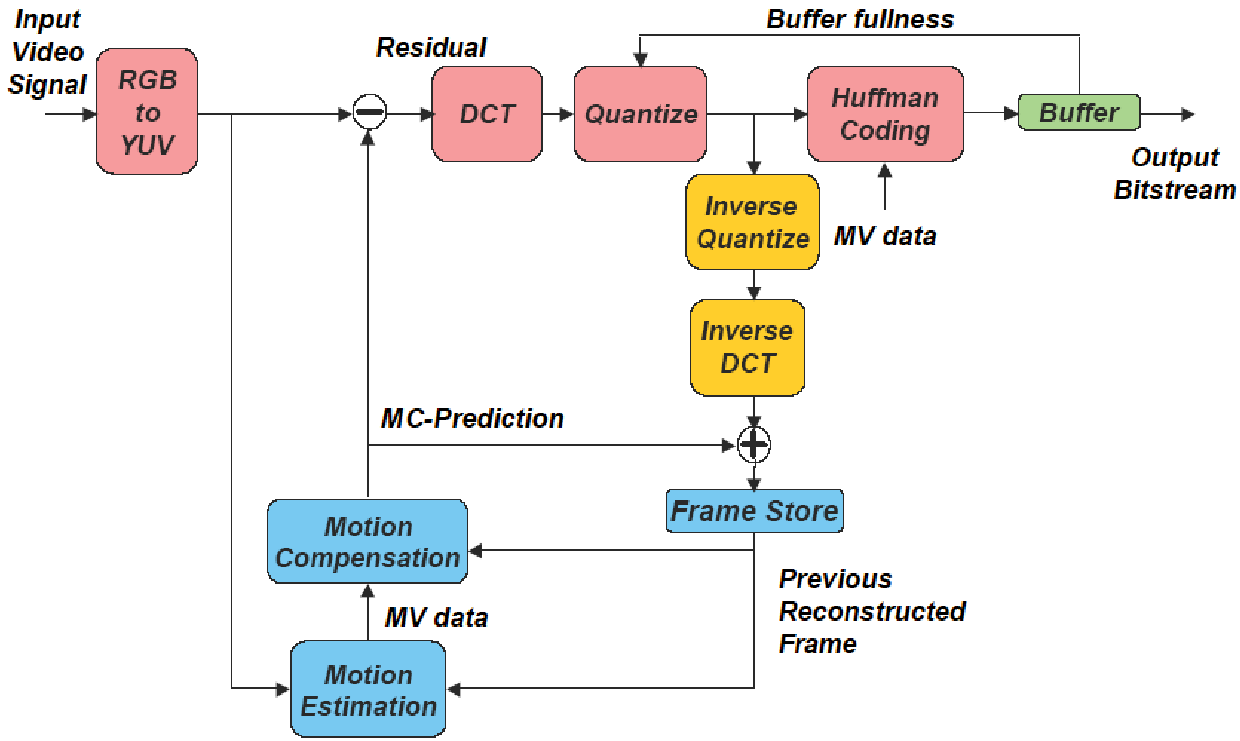


Summary of Temporal Processing

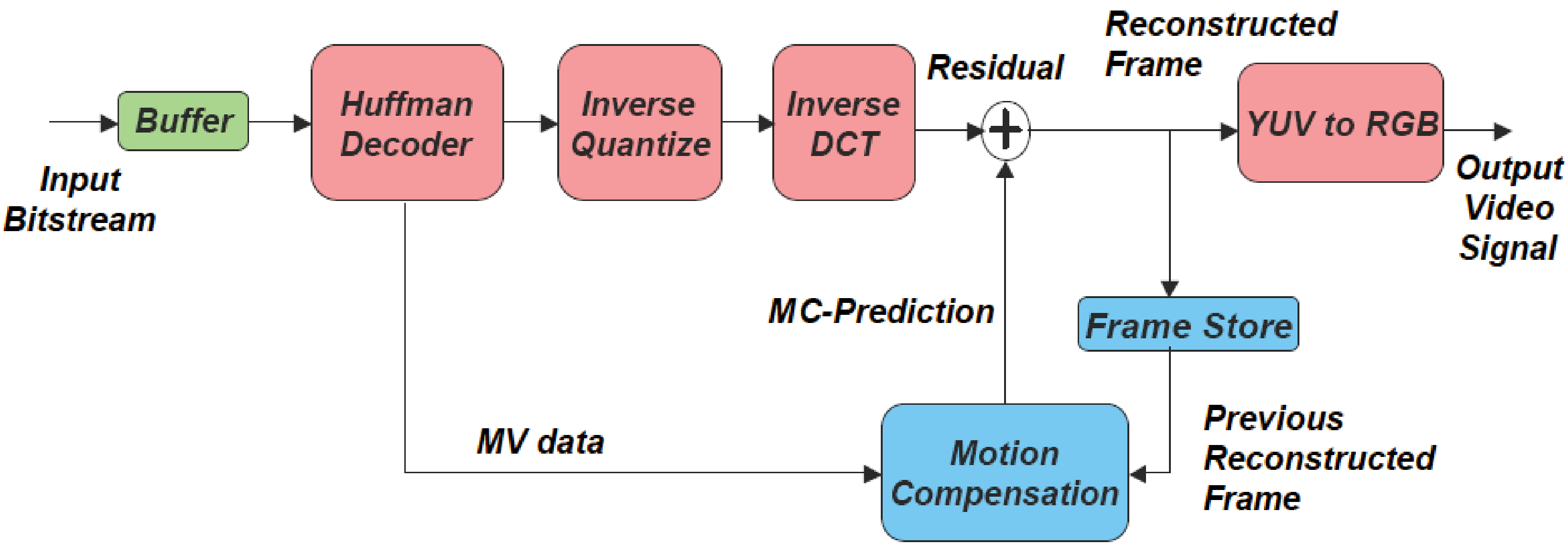
- Use MC-prediction (P and B frames) to reduce temporal redundancy.
- MC-prediction usually performs well; In compression have a second chance to recover when it performs badly.
- MC-prediction yields:
  - Motion vectors
  - MC-prediction error or residual
    -> Code error with conventional image coder.
- Sometimes MC-prediction may perform badly
  - Examples: Complex motion, new imagery (occlusions)
  - Approach:
    - 1. Identify frames or individual blocks where prediction fails
    - 2. Code them without prediction

Example Video Encoder

- Theoretically, MPEG = JPEG + ME.



Example Video Decoder



Video Standard

- ITU-T Standards for Audio-Visual Communications
  - H.261
  - H.263
  - H.263+, H.263++
- ISO Standards for
  - MPEG-1
  - MPEG-2
  - MPEG-4
  - MPEG-7

Multimedia Communications Standards and Applications

Standards	Application	Video Format	Raw Data Rate	Compressed Data Rate
H.320 (H.261)	Video conferencing over ISDN CIF	CIF/ QCIF	37 Mbps 9.1 Mbps	>=384 Kbps >=64 Kbps
H.323 (H.263)	Video conferencing over Internet	4CIF/ CIF/ QCIF		>=64 Kbps
H.324 (H.263)	Video over phone lines/ wireless	QCIF	9.1 Mbps	>=18 Kbps
MPEG-1	Video distribution on CD/ WWW	CIF	30 Mbps	1.5 Mbps
MPEG-2	Video distribution on DVD / digital TV	CCIR601 4:2:0	128 Mbps	3-10 Mbps
MPEG-4	Multimedia distribution over Inter/Intra net	CIF/ QCIF		28-1024 Kbps
GA-HDTV	HDTV broadcasting	SMPTE 296/295	<=700 Mbps	18-45 Mbps
MPEG-7	Multimedia databases (content description and retrieval)			

Comparing Current Video Compression Standards

- Based on the same fundamental building blocks
  - Motion-compensated prediction (I, P, and B frames)
  - 2-D Discrete Cosine Transform (DCT)
  - Color space conversion
  - Scalar quantization, runlengths, Huffman coding
- Additional tools added for different applications:
  - Progressive or interlaced video
  - Improved compression, error resilience, scalability, etc.
- MPEG-1/2/4, H.261/3/L: Frame-based coding
- MPEG-4: Object-based coding and Synthetic video

Newest Trends

- H.264
  - Just out in 2003, many new features
  - Quarter pixel motion compensation
  - Variable size motion blocks
- 3-D Wavelet Coding
  - Third dimension is time
  - 3-D SPIHT has been implemented
  - Delay is large because GOP is large
- GTV
  - Group testing for video
  - Bits per frame can be controlled enabling off-line rate control to succeed.