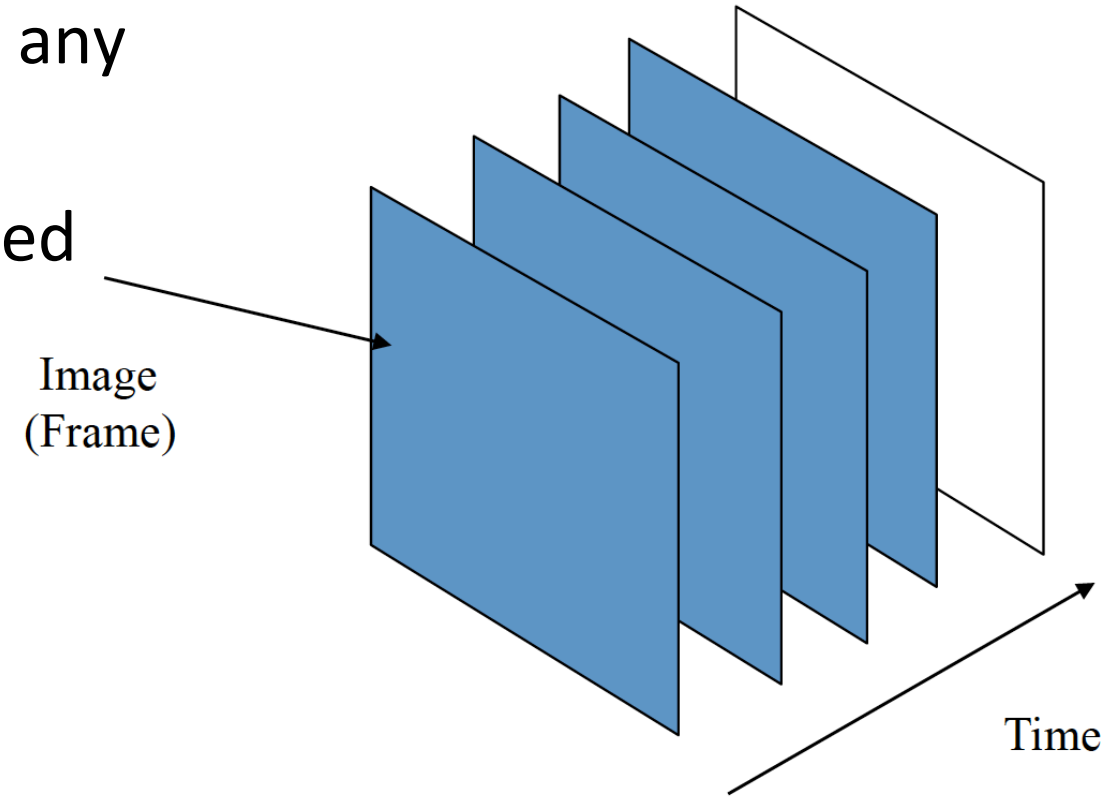
Multimedia Processing and Applications

Video Compression

Digital Video

- A sequence of images
  - Can be edited
  - Can be stored on any digital medium
  - Can be compressed

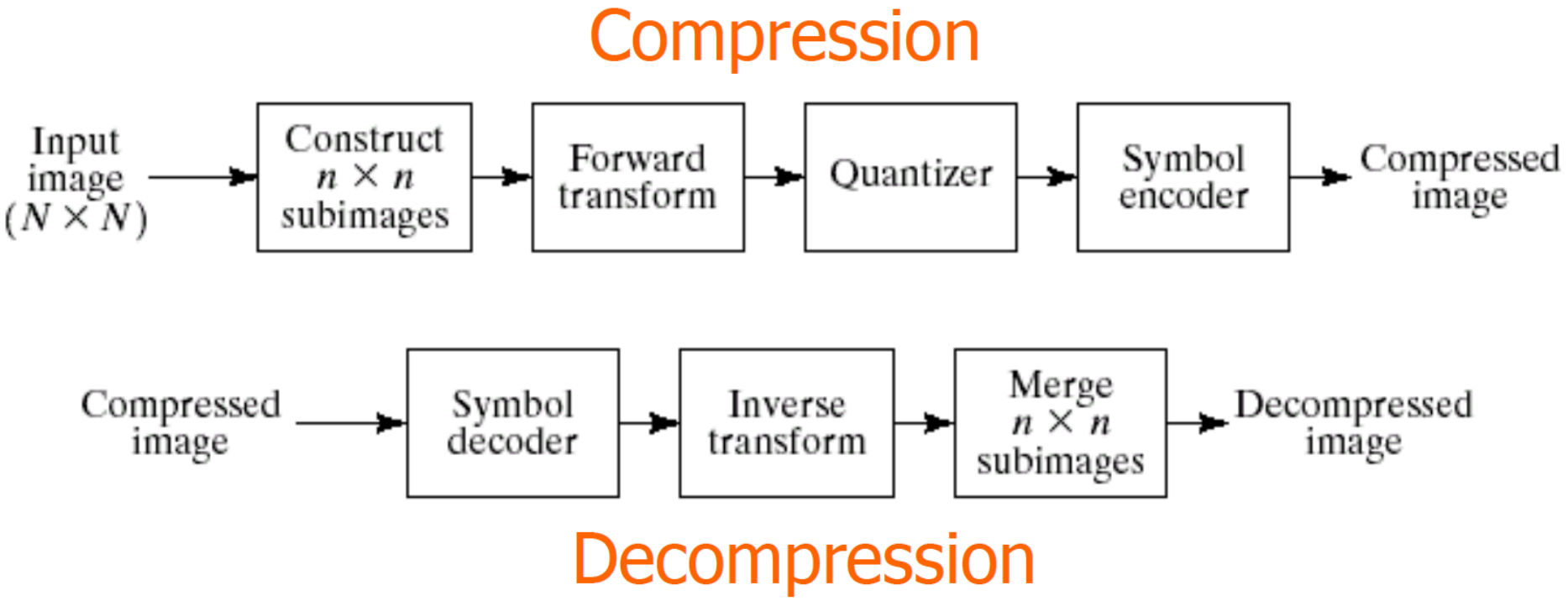


Video Compression: Need

- Huge data
- Example: High-Definition Television (HDTV)
- 1920x1080
- 30 frames per second (full motion)
- 8 bits for each three primary colors → 1.5Gb/sec
- Channel bandwidth 19.2 Mb/sec
- Reduced to 18 Mb/sec w/audio + control …
- Compression rate must be 83:1 !!

Image Compression: Transform Coding-> JPEG

- Pipeline

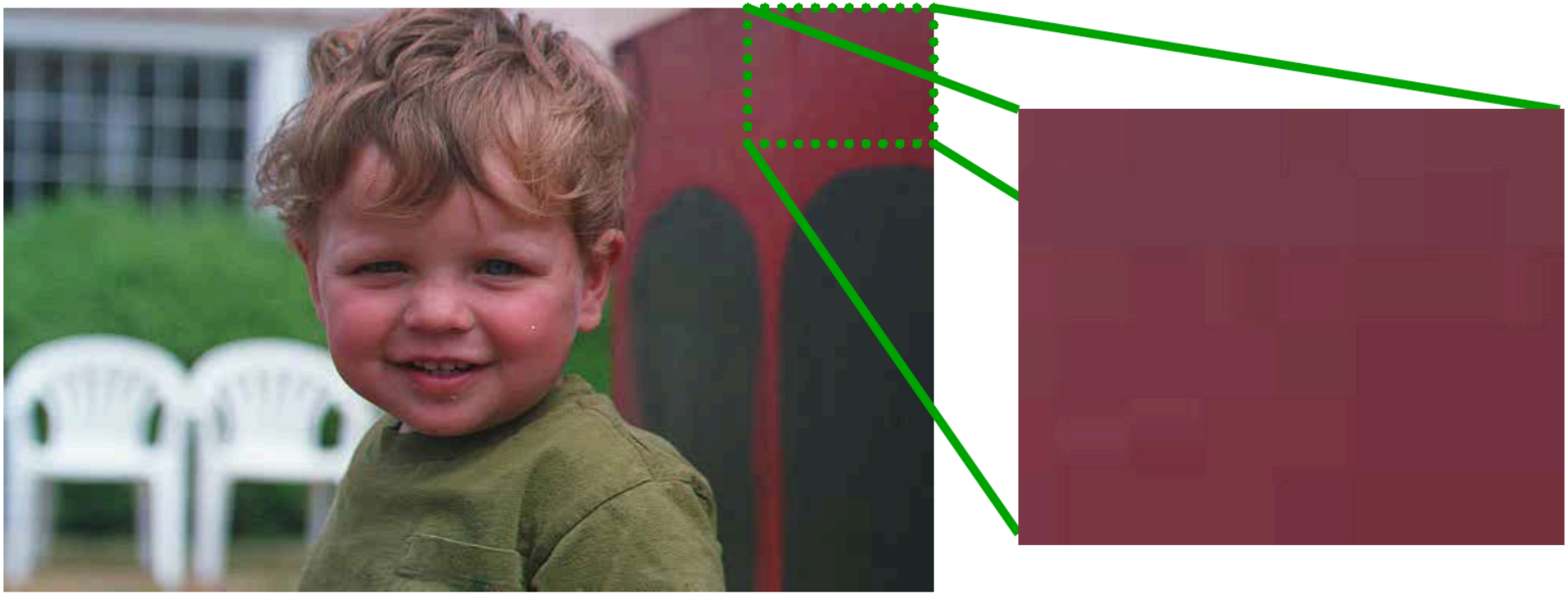


MJPEG (Motion JPEG)

- Each frame can be compressed as a single image
- Compression is achieved only due to spatial redundancy in the frame
- Takes care of intra-frame redundancy

Spatial Redundancy

- Takes advantage of similarity among most neighboring pixels



Temporal Redundancy

- Video: Sequence of images in time (that are related!)
- Take advantage of similarity between successive frames

950

951

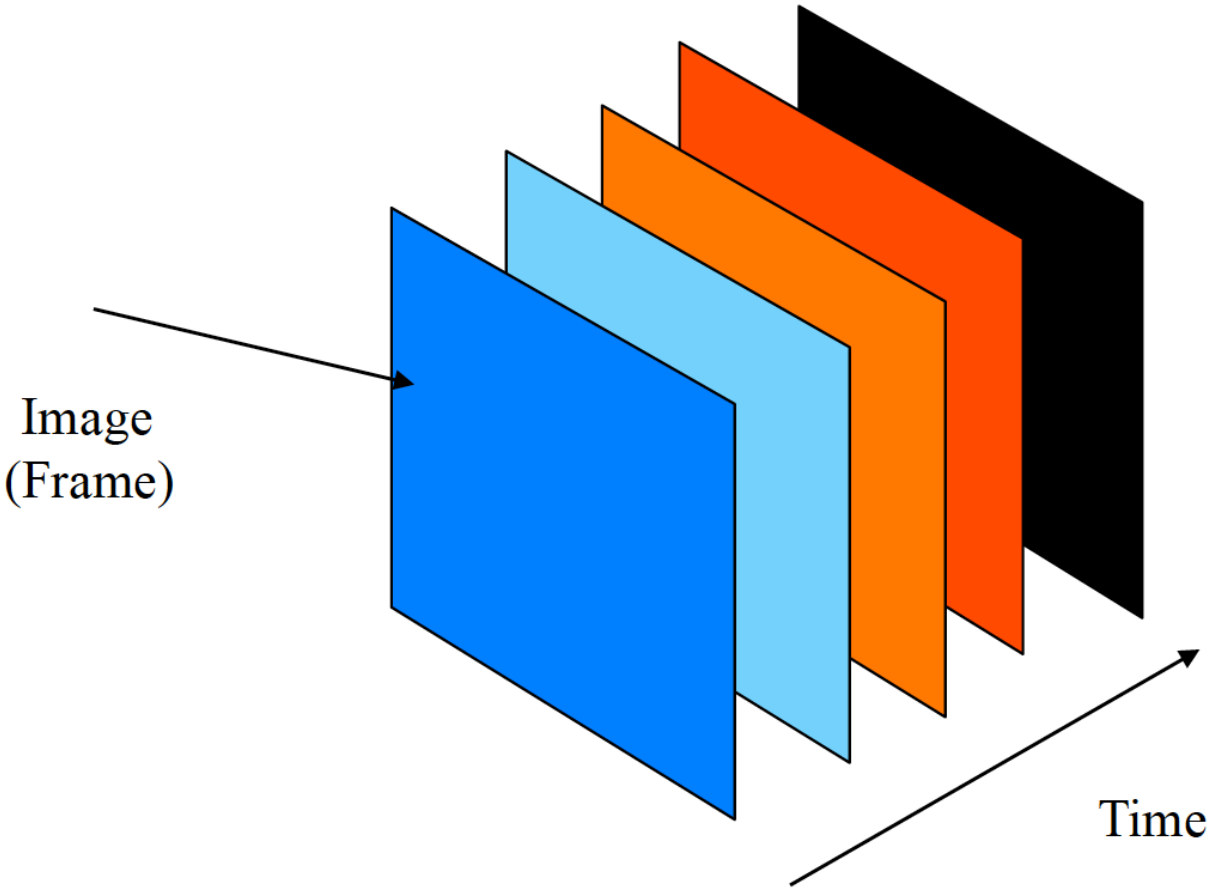
952

Intuitive Methods

- Subsampling
  - Drop frames
- Differencing
  - Differential coding of pixels
- Block Differencing
  - Differential coding of blocks (big pixels)
- Motion Compensation
  - Figure out the motion vector and compensate for it

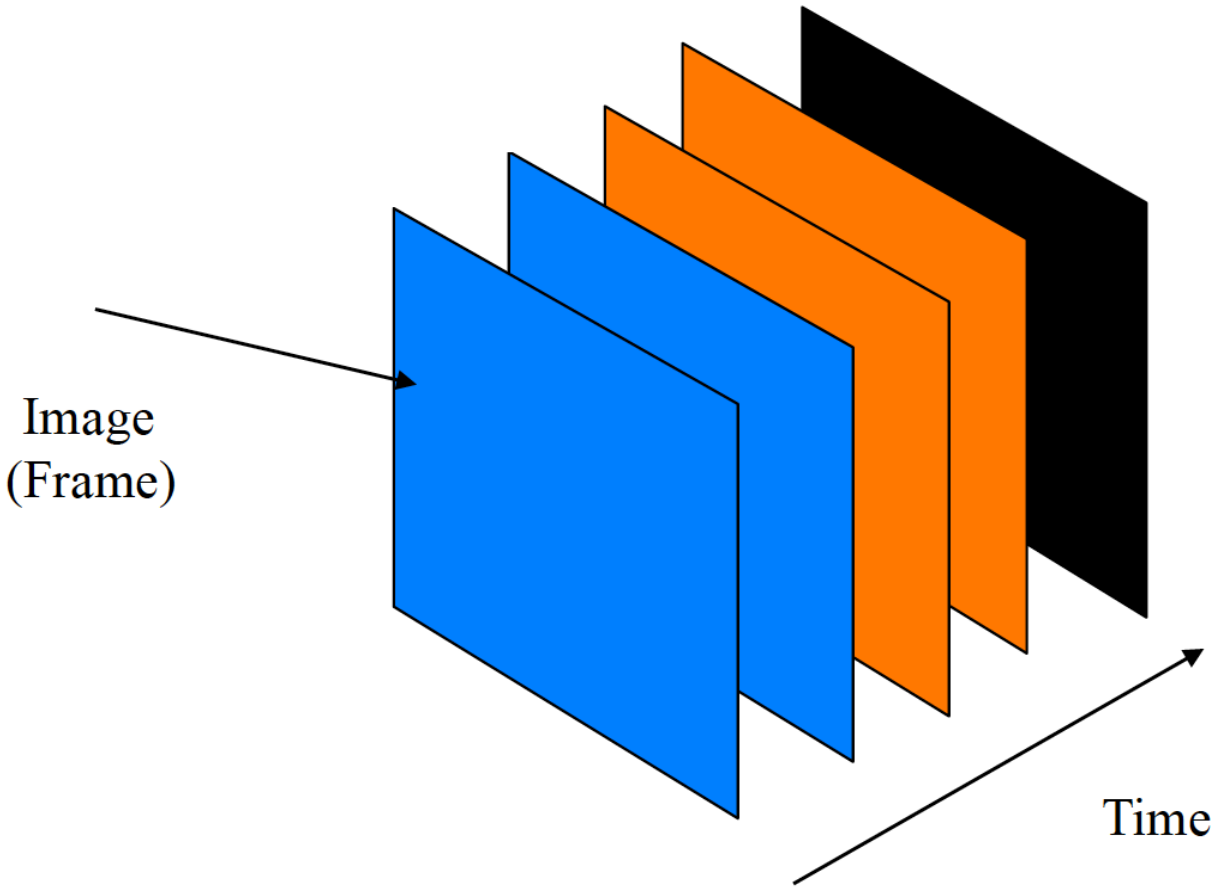
Subsampling

- Drop frames (repeat frames)

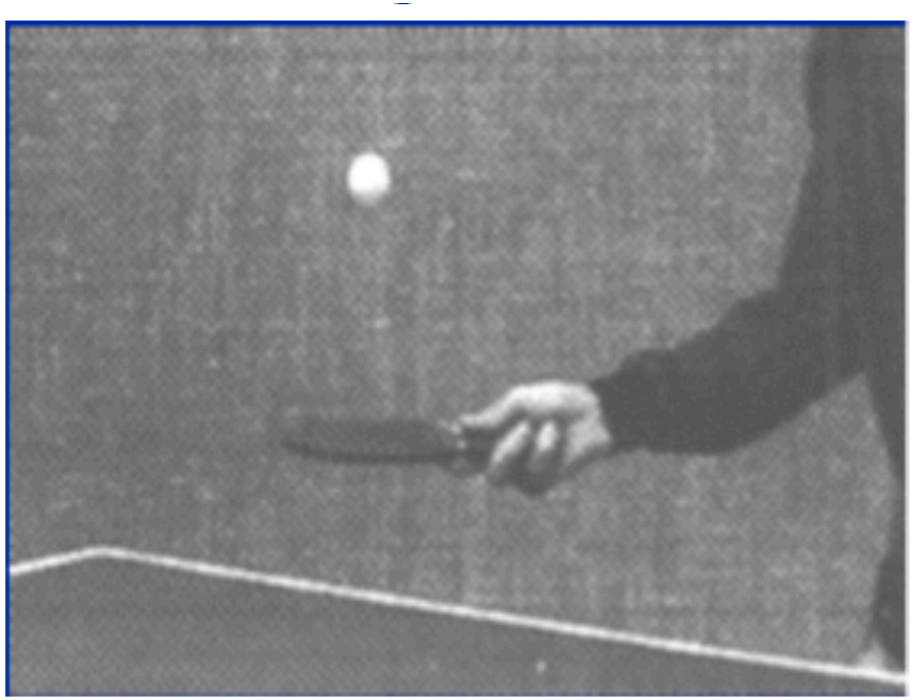


Subsampling

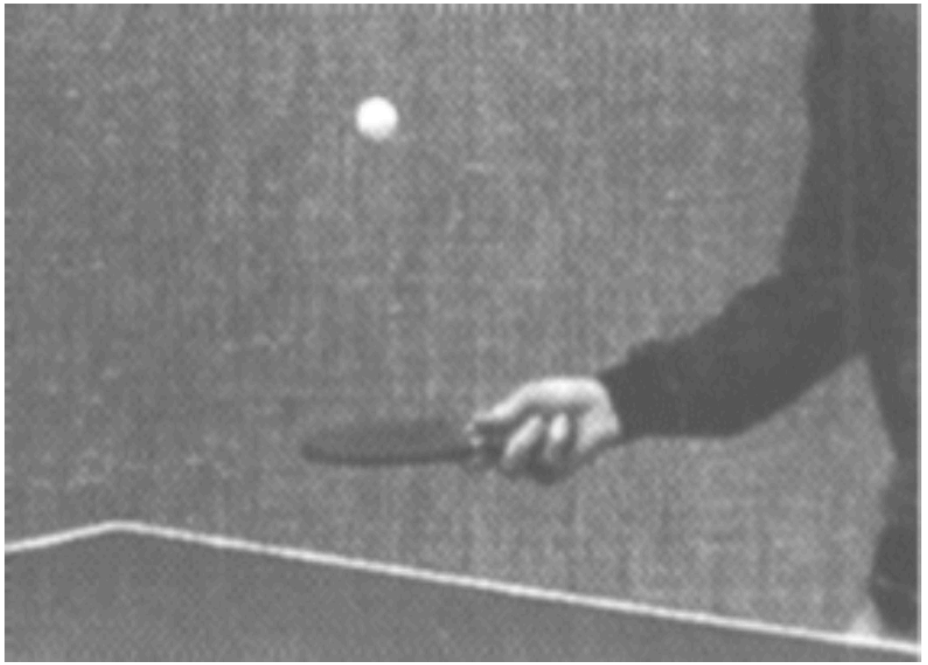
- Drop frames (repeat frames)



Differencing

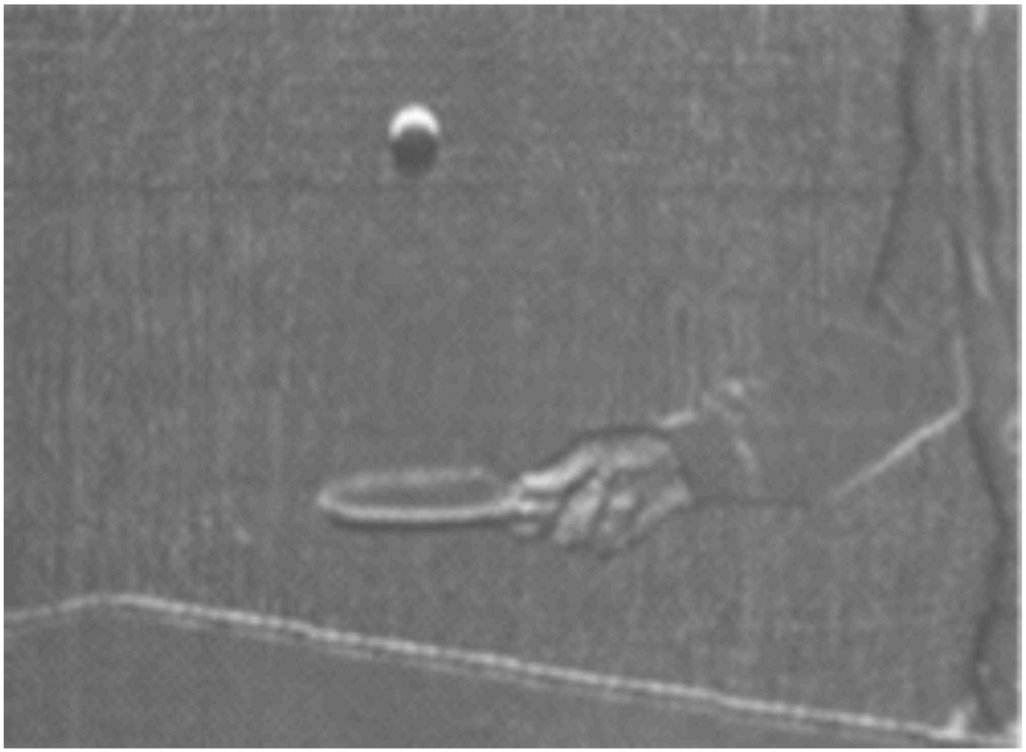


Frame N



Frame N+1

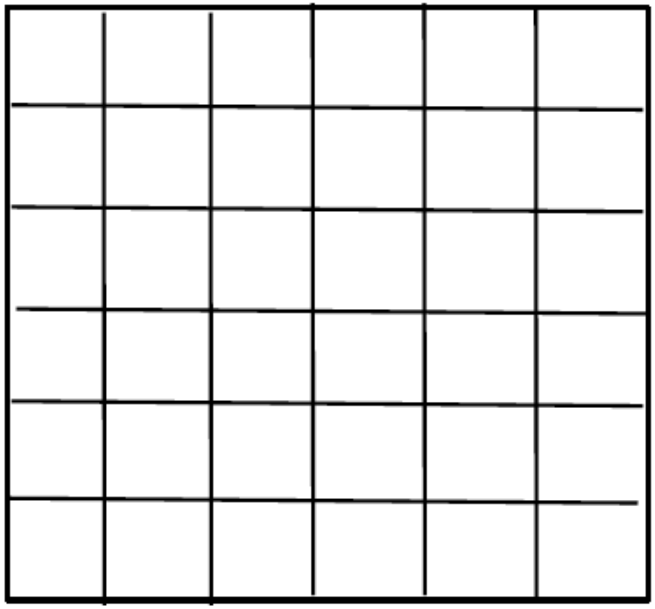
Differencing



Difference frame

Block Differencing

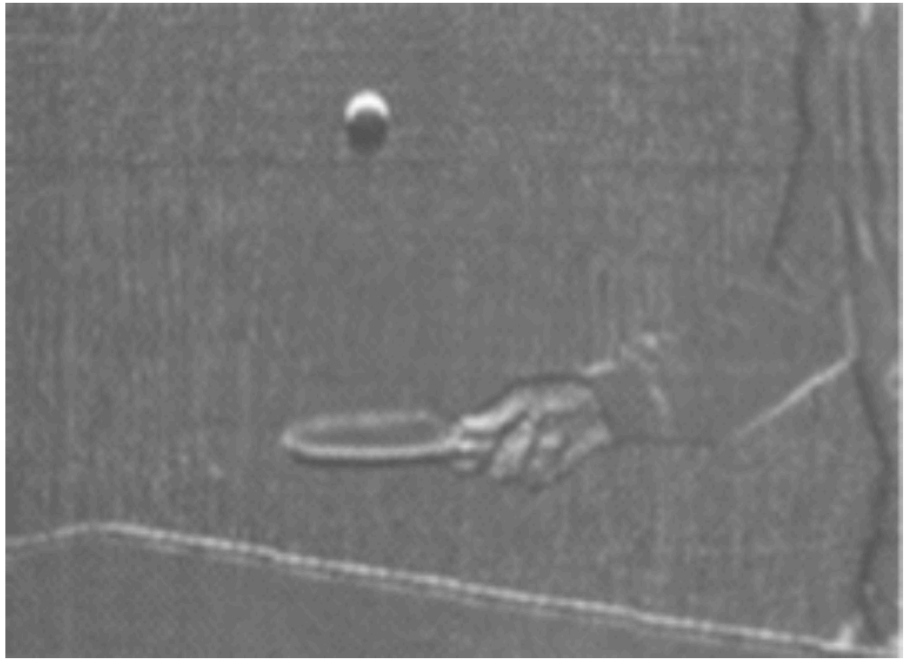
- Frame is divided into non-overlapping block
- Block level comparison rather than pixel level to decide which blocks for the difference is to be coded
- May work when the motion is relatively small of foreground objects
- If the motion is large and not limited to portion of image then it may not be effective



Motion Compensation

- Simple frame difference will fail if there is a significant motion
- Should account for the motion
  - Motion-compensated (MC) prediction
- How can we estimate motion?

Motion Compensation



Difference frame without motion prediction

Difference frame with motion prediction

Intra-Coded and Inter-Coded Frames

- Motion estimation and compensation requires current and previous frame
- Intra-coded frames: No previous frame for the first frame of video
  - Only spatial or intra-frame redundancy is exploited
  - Periodically introduced to prevent accumulation of prediction error over frames
- Inter-coded frames: Second frame onwards, both temporal and spatial redundancy can be exploited

Motion Estimation

Approach 1

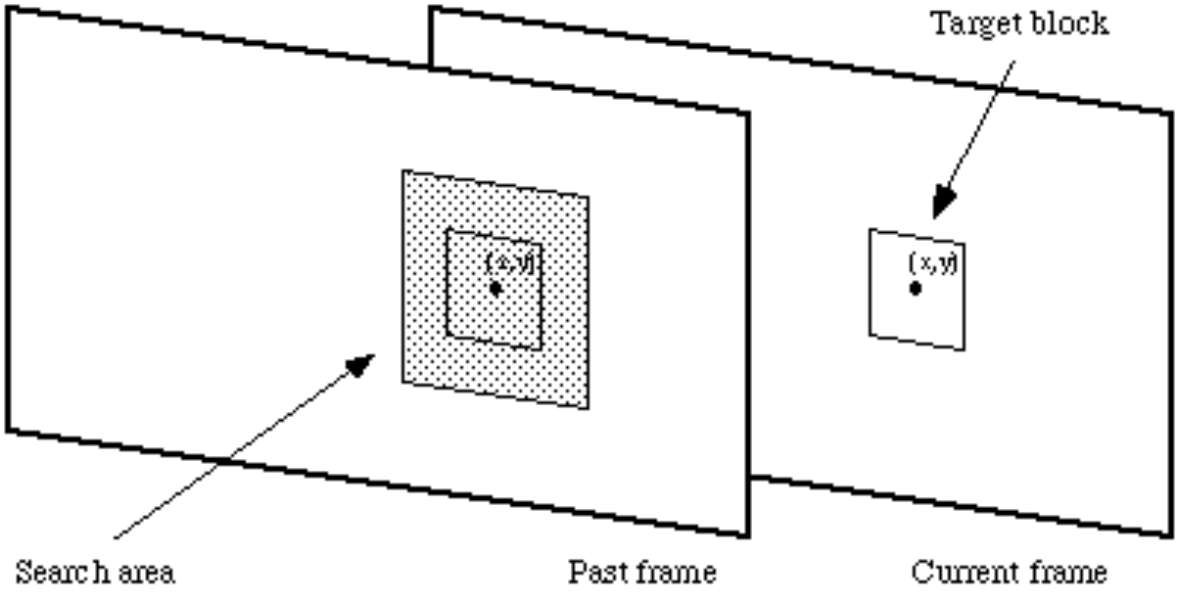
- Segment video into moving objects
- Describe (model) object motion
  - May be some what difficult

Approach 2 (Practical)

- Block matching motion segmentation
- No object segmentation and identification required
- Good performance

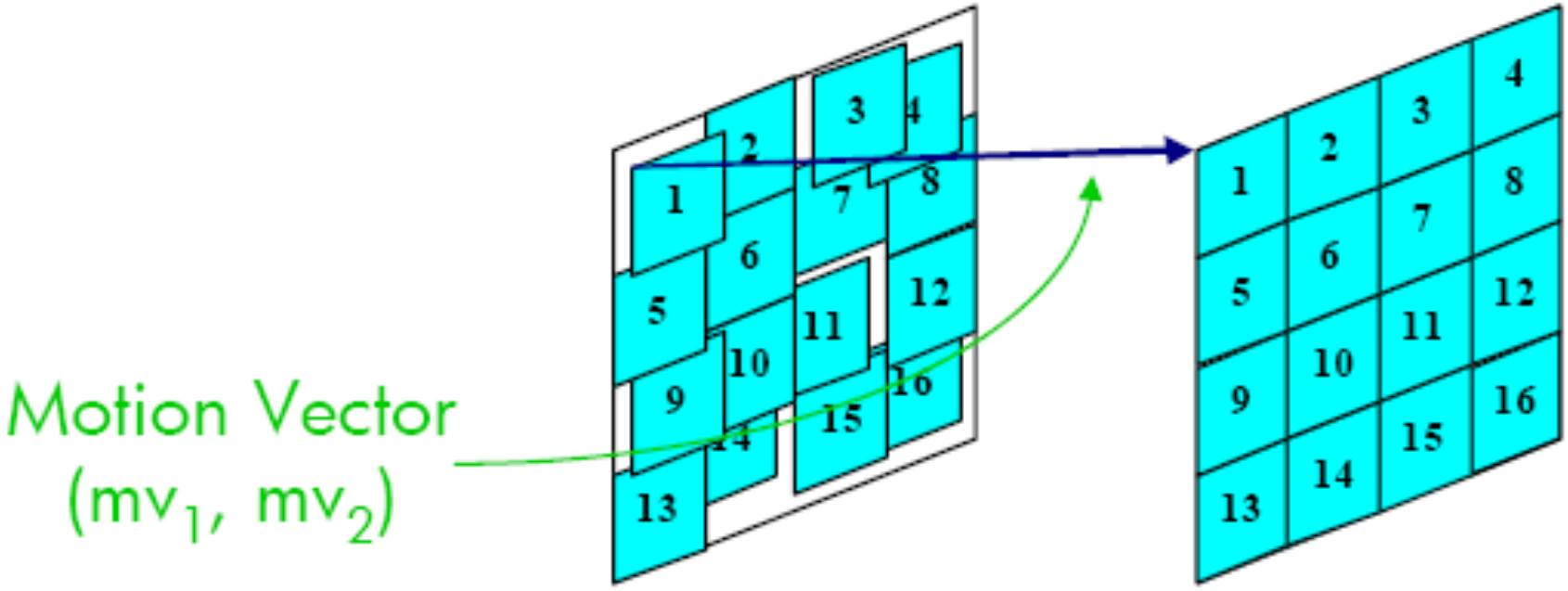
Motion Estimation

- Computes the displacement between the current frame and a stored past (reference) frame
- For a pixel/block in current frame, search area in previous frame is usually limited to a region close to the target pixel/block



Motion Estimation

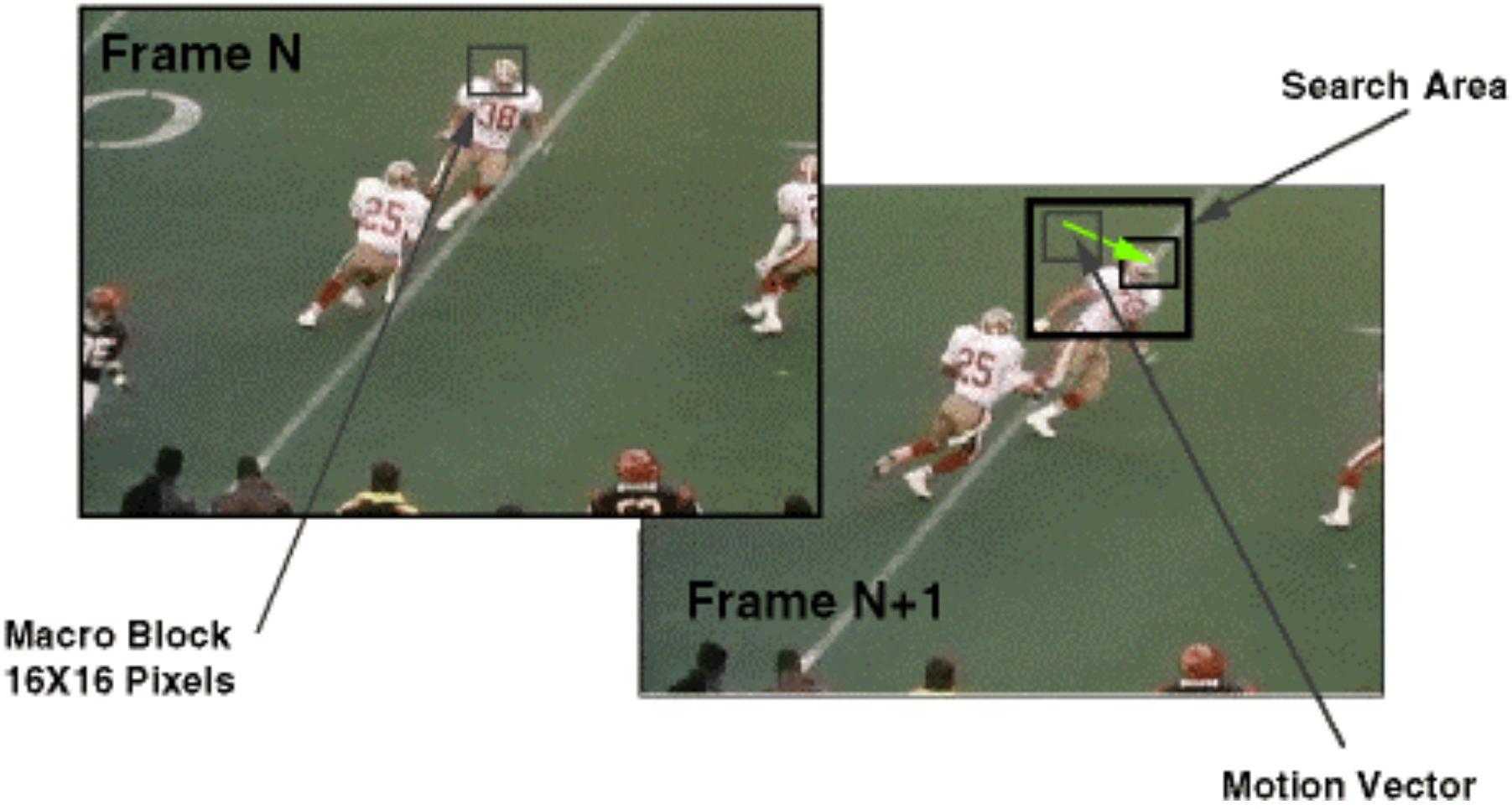
- Motion Vector: The difference in position between the candidate and its match in the reference frame is defined as displacement vector or motion vector



Motion Compensation

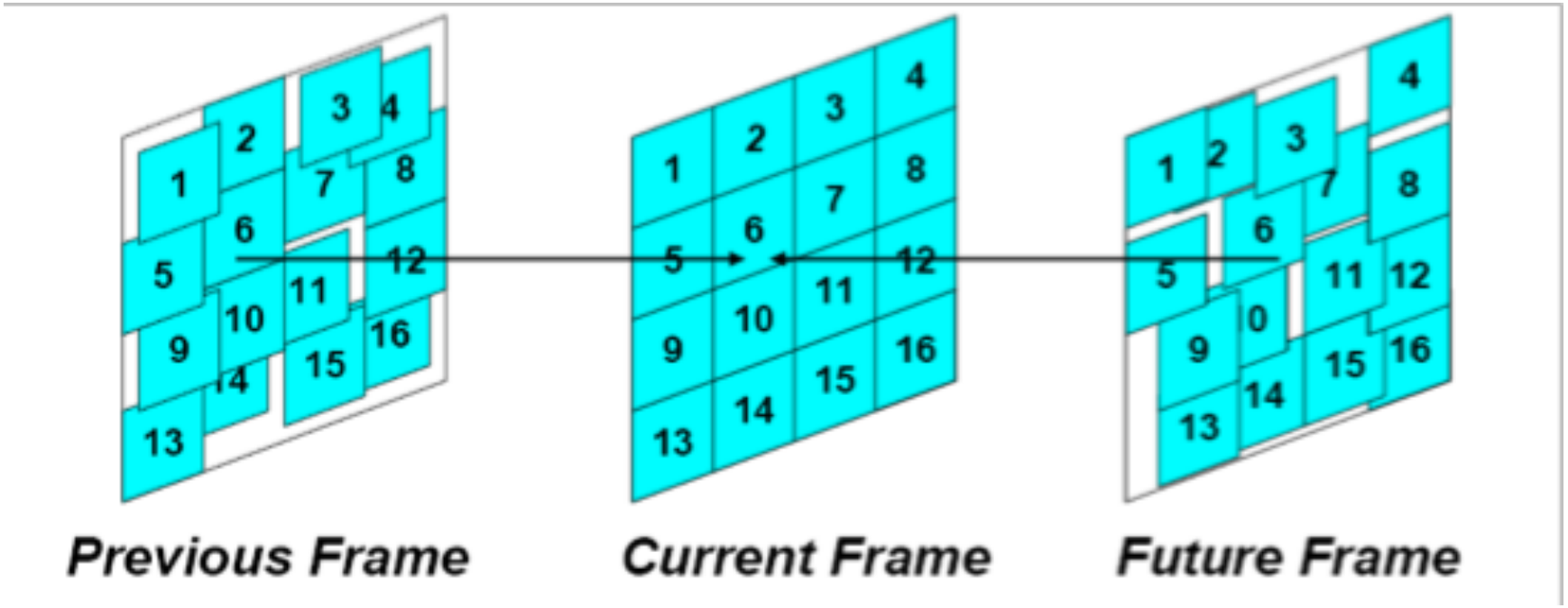
- Composes how the current frame would have looked if corresponding displacements were applied at different regions of the reference frame.

Motion Compensation



Motion Estimation

- Forward, backward, bidirectional

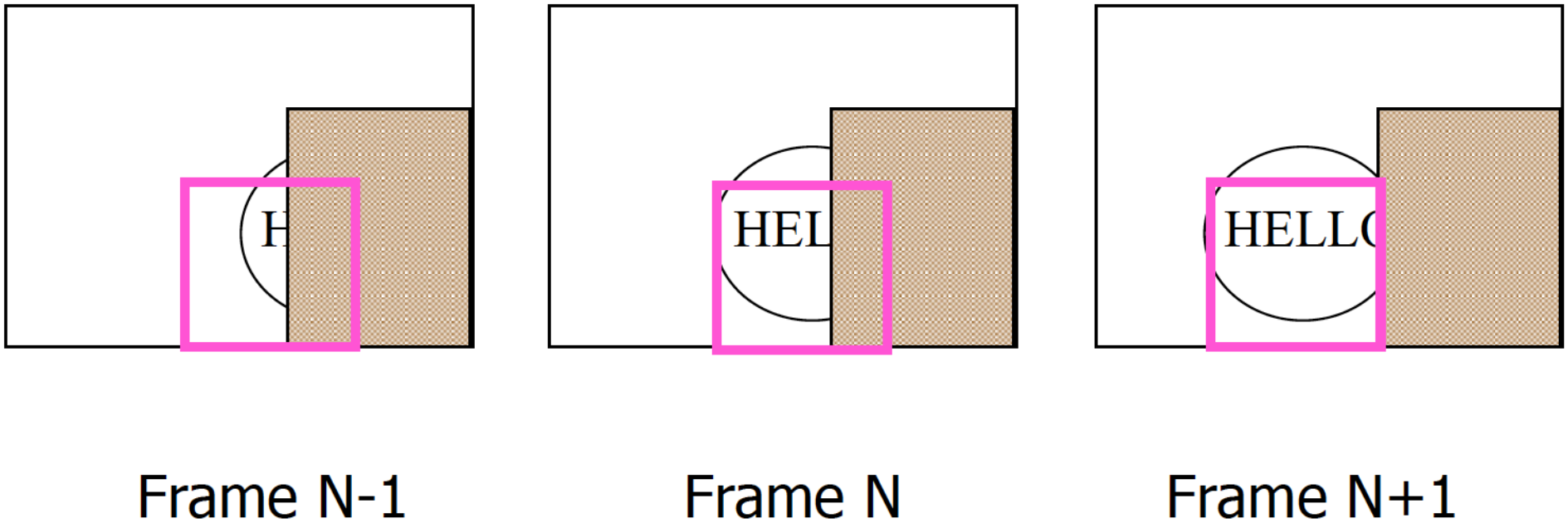


Motion Estimation

- As a general scheme a block in current frame can be estimated from a block in
  - Previous frame: Backward motion estimation
  - Future frame: Forward motion estimation
  - Average of block from the previous frame and a block from the future frame: Bidirectional

Motion Estimation

- Example



Frame N-1 Frame N Frame N+1

- Due to unexpected movements and occlusions in real scenes, the target macroblock may not have a good matching entity in the previous frame

Basic Approaches to Motion Estimation

- Pixel-based: determine motion vectors for every pixel
  - Brightness constancy
  - Additional constraints are required
  - Computationally not effective
- Block-based: Candidate frames are divided into non-overlapping blocks
  - Motion vector is computed for each block
  - More popular and computationally effective

Matching Criteria for Block-Based Motion Estimation

- Mean of squared error (MSE)
- Mean of absolute difference (MAD)
- Matching Pixel Count (MPC)

Motion Estimation

- Issues:
  - Block size?
  - Search range?
  - Motion vector accuracy
  - Complex motion

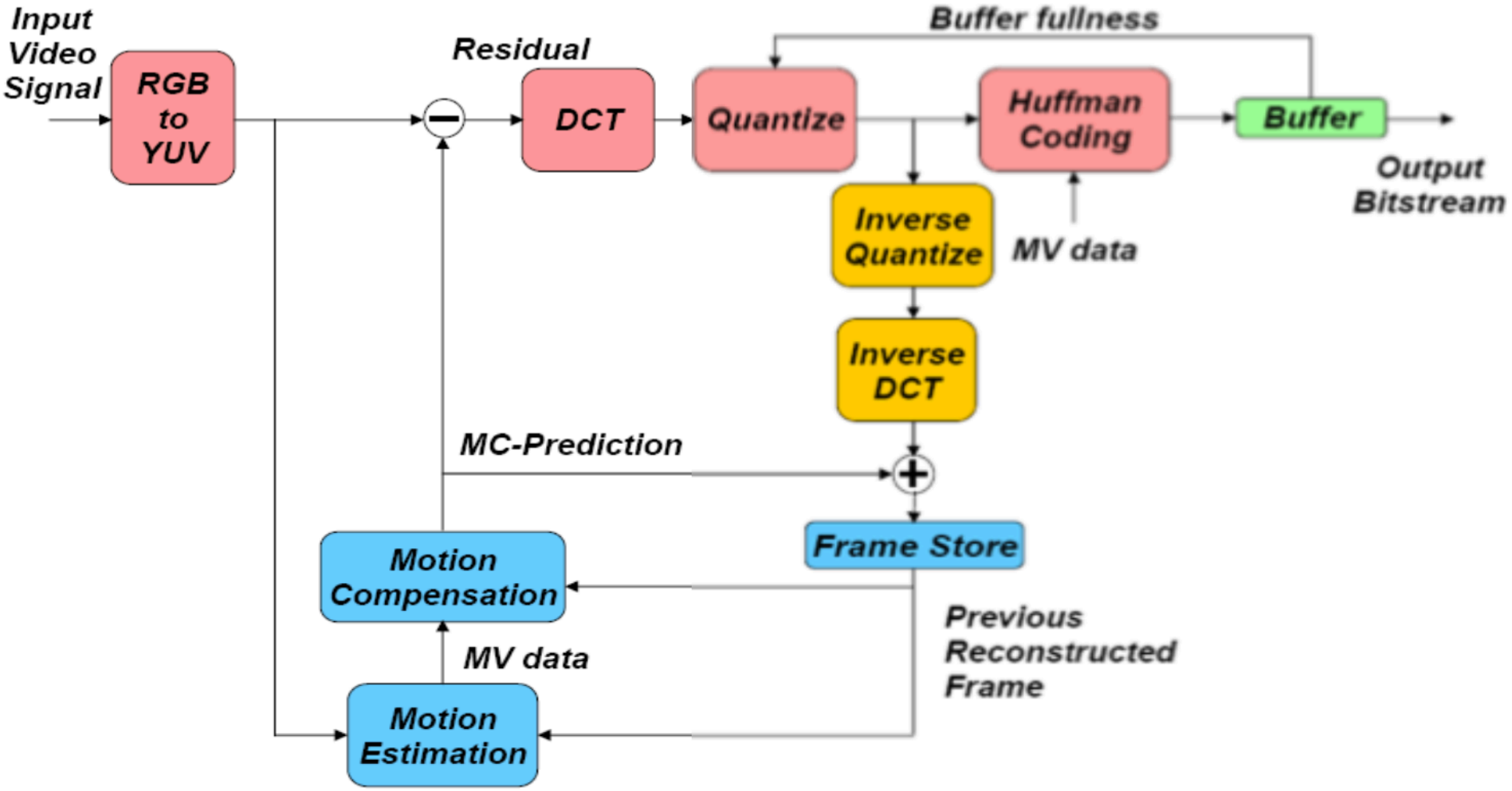
Motion Estimation

- Translation motion model
- All pixels within the block have the same motion
- Motion is estimated using only luminance
- The motion vector is encoded in place of the target block itself
- Fewer bits are required to code a motion vector

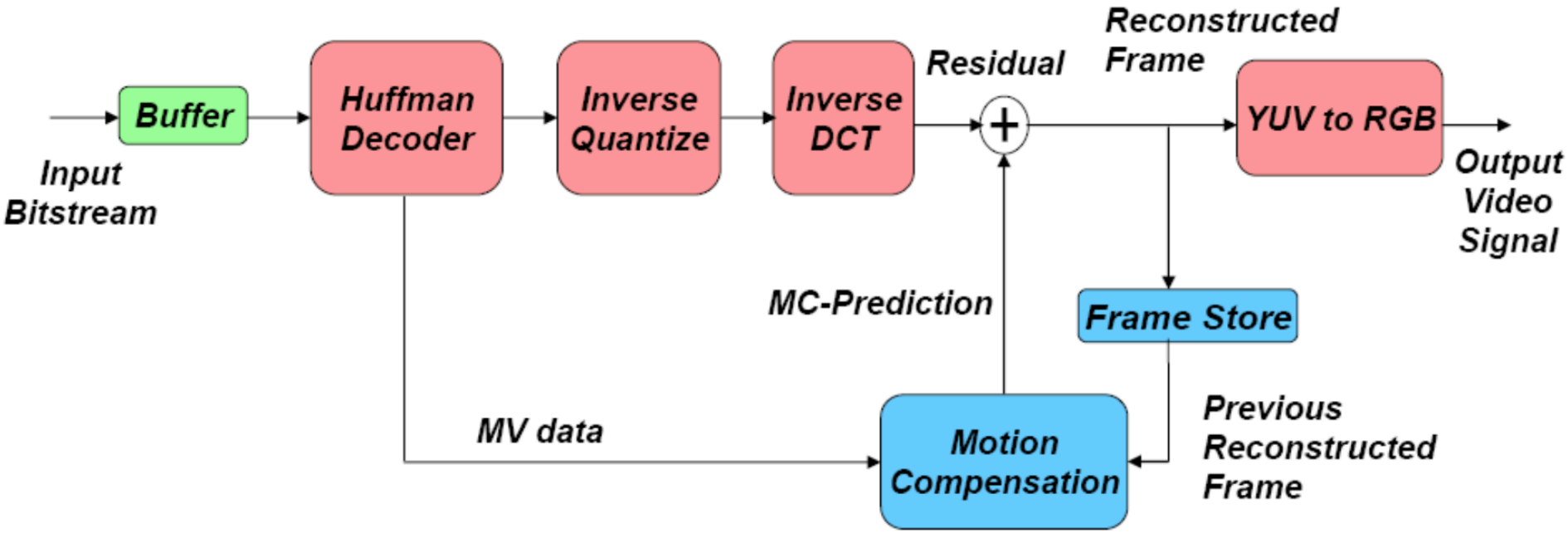
MC-Based Video Encoding

- In motion estimation, each macroblock of the target frame is assigned a best matching macroblock from previously encoded frame. This is called prediction.
- The difference between macroblock and matching block is the prediction error.
- Error is sent to DCT and subsequent encoding steps.

Encoding


Input Video Signal
RGB to YUV
Residual
DCT
Quantize
Buffer fullness
Huffman Coding
Buffer
Output Bitstream
Inverse Quantize
MV data
Inverse DCT
MC-Prediction
Frame Store
Motion Compensation
Previous Reconstructed Frame
MV data
Motion Estimation

Decoding


Buffer
Input Bitstream
Huffman Decoder
Inverse Quantize
Inverse DCT
Residual
Reconstructed Frame
YUV to RGB
Output Video Signal
MC-Prediction
Frame Store
MV data
Motion Compensation
Previous Reconstructed Frame

Video Coding Standards

- Exploit temporal redundancies
  - Predict video frames using motion estimation
- Error in temporal prediction is encoded by transform domain techniques (e.g., DCT)
- Spatial redundancy
  - Quantization and entropy coding

Video Coding Standards

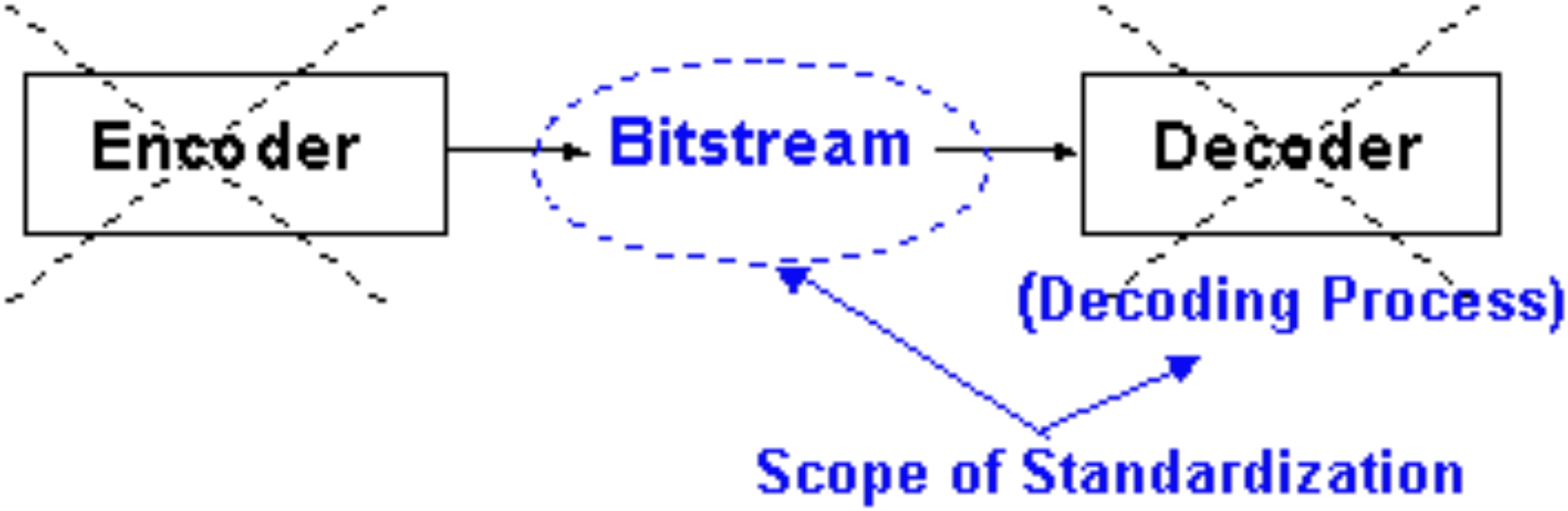
- Video codecs follow hybrid encoding scheme
- All standards are generic. They do not specify the operations of encoder but the syntax and semantics of coded bitstream and the method of decoding

Video Compression Standards

- A video compression system consists of the following:
  - An encoder
  - Compressed bit-streams
  - A decoder

Video Compression Standards

- Compressions standards specify
  - Not the encoder, not the decoder



Major Video Coding Standards

- ITU Video Coding Experts Group (VCEG)
  - H.261 for ISDN video conferencing
  - H.263 for very low bit-rate and Plain Old Telephone Systems (POTS) video conferencing
  - H.264 for video telephony and video streaming in wireless applications

Major Video Coding Standards

- ISO/IEC Moving Pictures Experts Group (MPEG)
  - MPEG-1 for storing videos on CD-ROM, bit-rate up to 1.5Mbits/sec
  - MPEG-2 for storing videos on DVDs and in HDTV applications

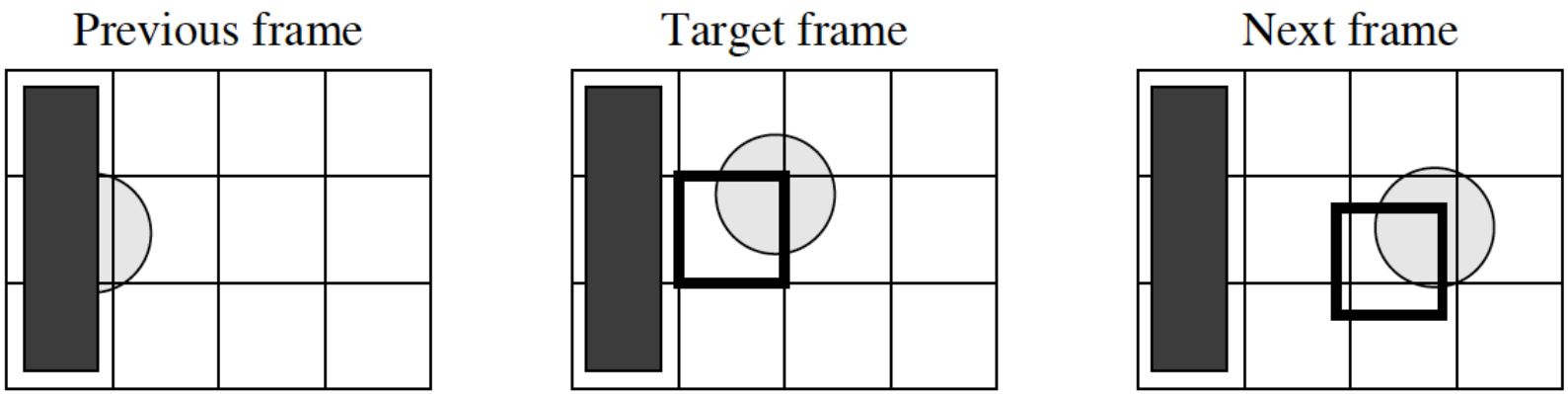
MPEG-1 Standard

- ISO/IEC standard for "*Coding of Moving Pictures and Associated Audio for Digital Media at up to about 1.5Mbits/s*"
- Supports only non-interlaced video
- Uses 4:2:0 chroma-subsampling

Motion Compensation in MPEG-1

- Motion Compensation (MC) based video encoding works as follows:
  - In Motion Estimation (ME), each macroblock (MB) of the Target P-frame is assigned a best matching MB from the previously coded I or P frame - prediction.
  - prediction error: The difference between the MB and its matching MB, sent to DCT and its subsequent encoding steps.

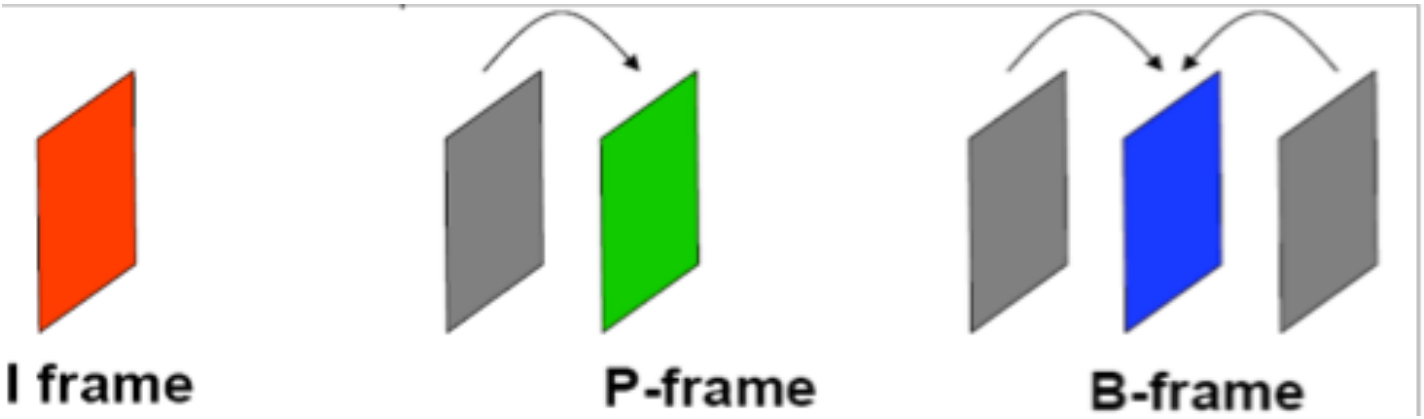
The Need for Bidirectional Search.



- The MB containing part of a ball in the Target frame cannot find a good matching MB in the previous frame because half of the ball was occluded by another object. A match however can readily be obtained from the next frame.

Picture Types in MPEG-1

- I-frame: Intra coded frame, coded independently
- P-frame: Predictive coded frame, coded based on previously coded frame
- B-frame: Bi-directionally predicted frame, coded based on previous and future coded frames



Intra-coded Frame/I-Frame

- Coded without reference to other frames
  - Does not use any motion estimation or motion compensation
- The first frame of every video sequence must be an I-frame
- I-frames are encoded at regular intervals to enforce updating with the current content
  - Beginning of every GOP (Group of Pictures)
- Very poor compression

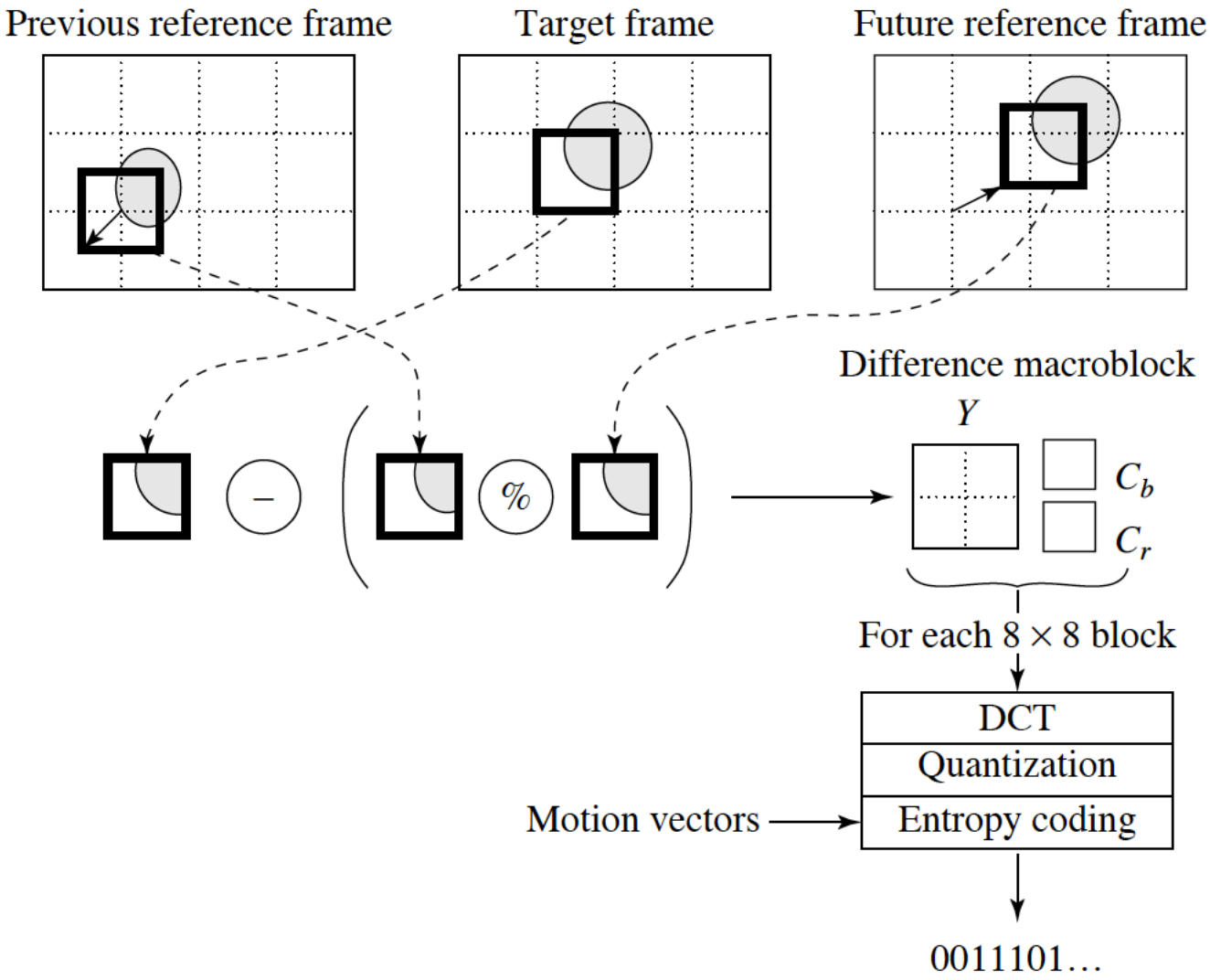
Inter-frame/Predicted Frame (P-Frame)

- Coded with reference to the nearest coded I-Frame or P-Frame using MC for prediction
- Better compression

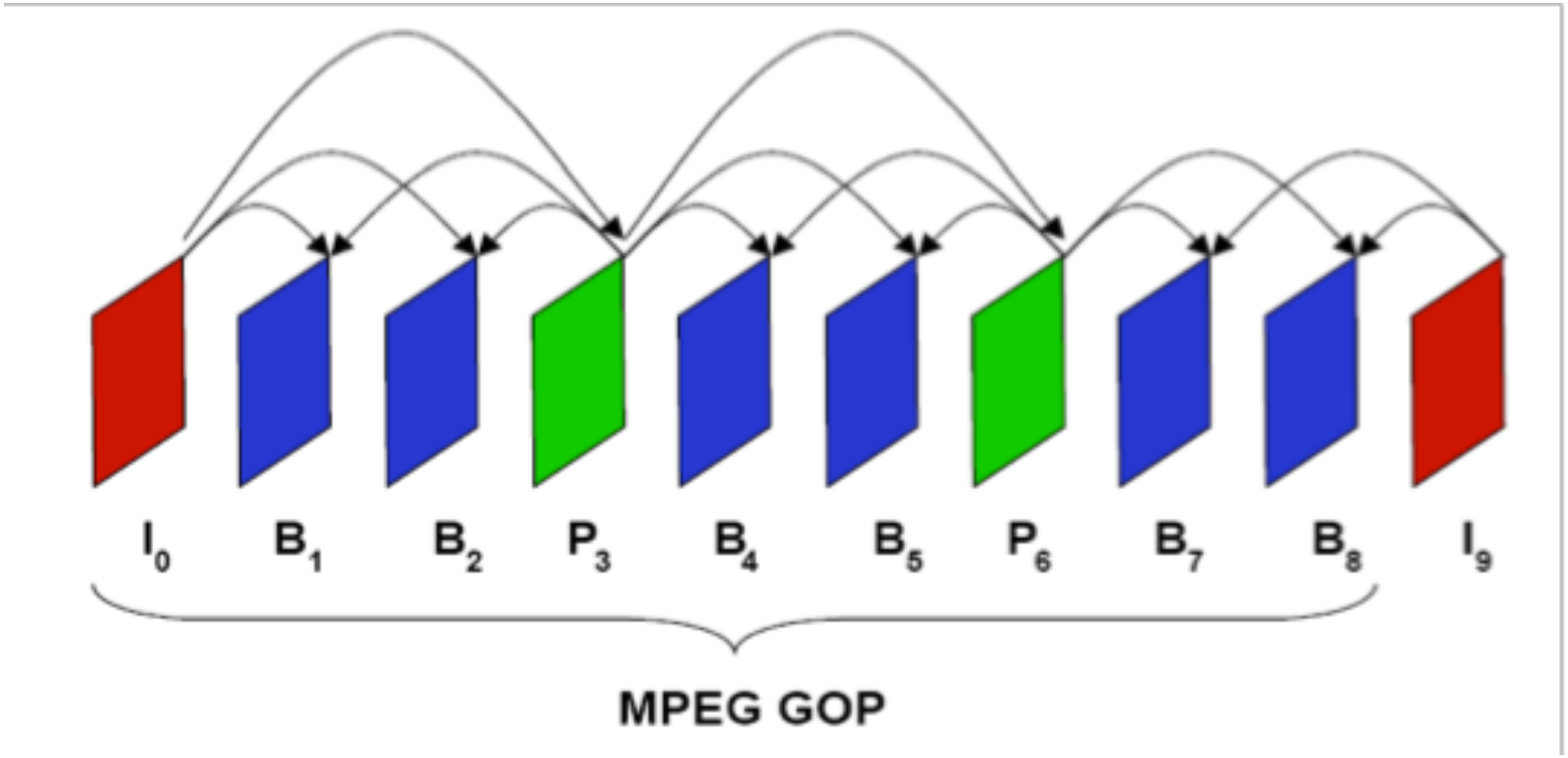
Bi-directionally Predicted Frames

- Each macroblock from a B-frame will specify up to two motion vectors, one from the forward and one from the backward prediction.
- Corresponding matching blocks are averaged before matching with the target macroblock for generating prediction error
- Best compression performance

B-frame Coding Based on Bidirectional Motion Compensation



Coded Frames Sequence

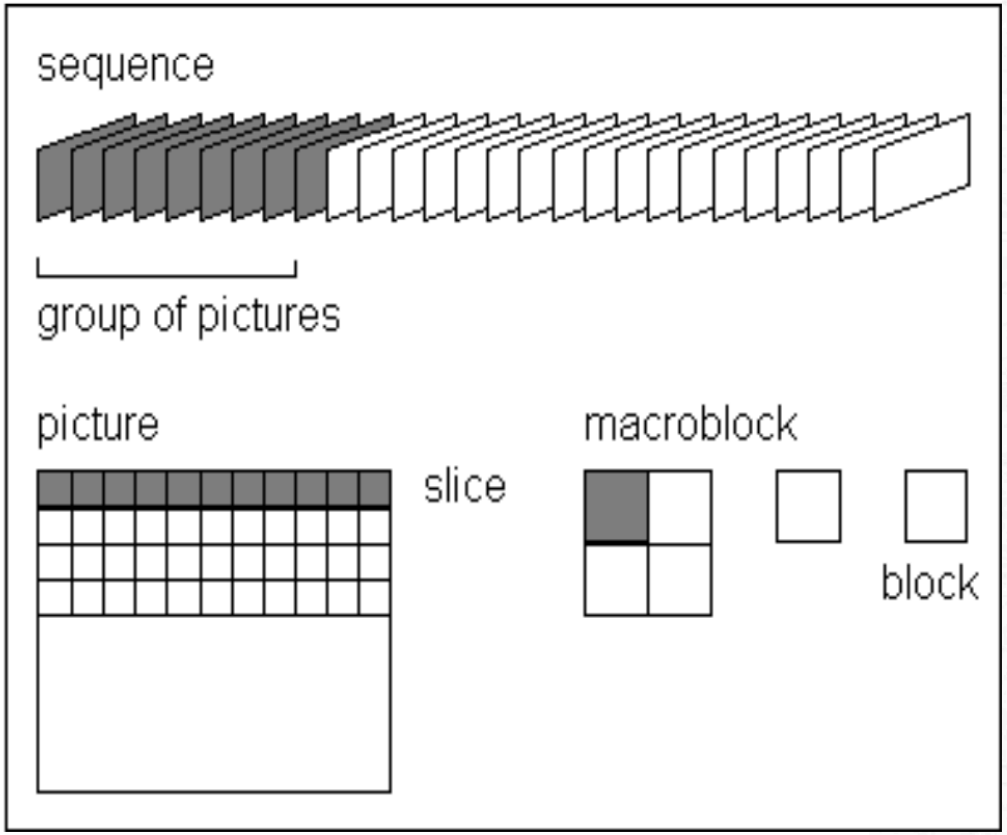


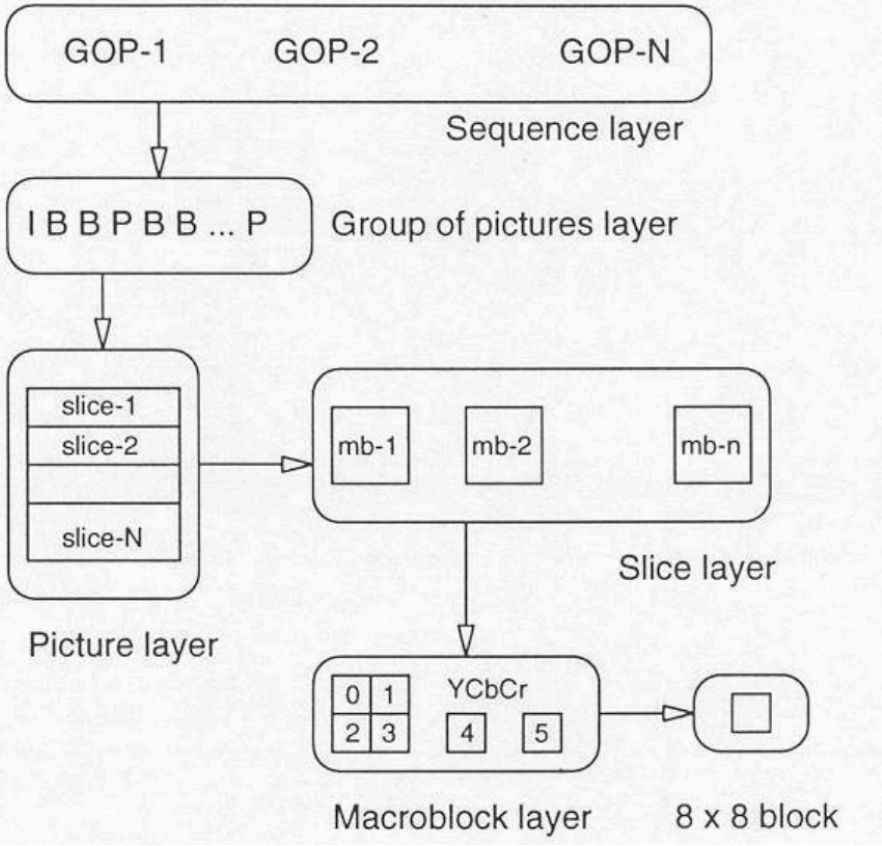
- Display order: I_0, B_1, B_2, P_3, B_4, B_5, P_6, B_7, B_8, I_9
- Transmission order: I_0, P_3, B_1, B_2, P_6, B_4, B_5, I_9, B_7, B_8

Hierarchical Structure for MPEG-1 Bitstream

- Block: A basic unit for DCT operation (8x8)
- Macroblock: A basic unit for motion compensation operation. Consists of 4 Y blocks, one Cb and one Cr block.
- Slice: String of macroblocks
- Picture/Frame: I/B/P
- Group-of-pictures (GOP): Collection of pictures (10 or so)
- Sequence: Contains one or more GOPs. It always starts with sequence header (picture size, pixel aspect ratio, frame rate, quantization matrix and so on)

MPEG-1 Video Bitstream





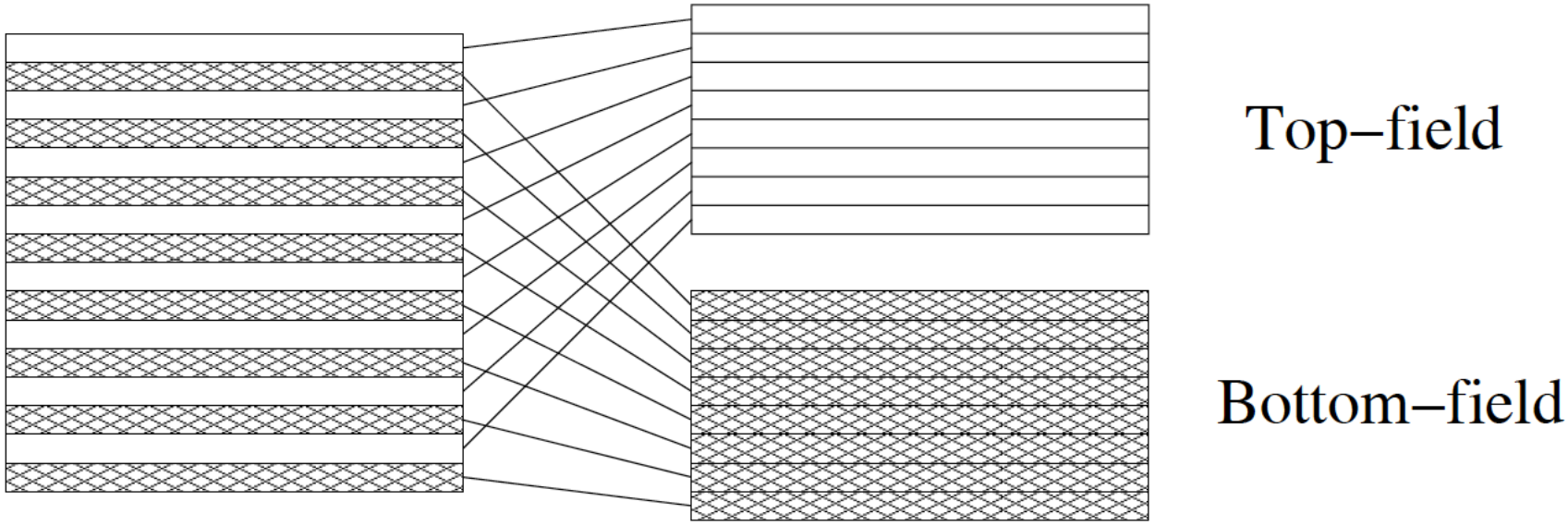
MPEG-2 Standard

- For higher quality video at a bit-rate of more than 4 Mbps
- Supports interlaced video
- Scalable bit streams
- Backward compatible with MPEG-1
- Besides 4:2:0, supports 4:2:2 and 4:4:4 chroma subsampling

Interlaced Video: Frame Picture and Field Picture

- In interlaced video each frame consists of two fields, referred to as the top-field and the bottom-field.
- Field Picture
  - Every field is treated as separate picture and is coded separately
  - Every field is separated into non-overlapping macroblock and DCT is applied on a field basis
- Frame Picture
  - Two fields are coded together as a frame

Frame Picture and Field Picture



(a) Frame–picture vs. Field–pictures

Frame Picture and Field Picture

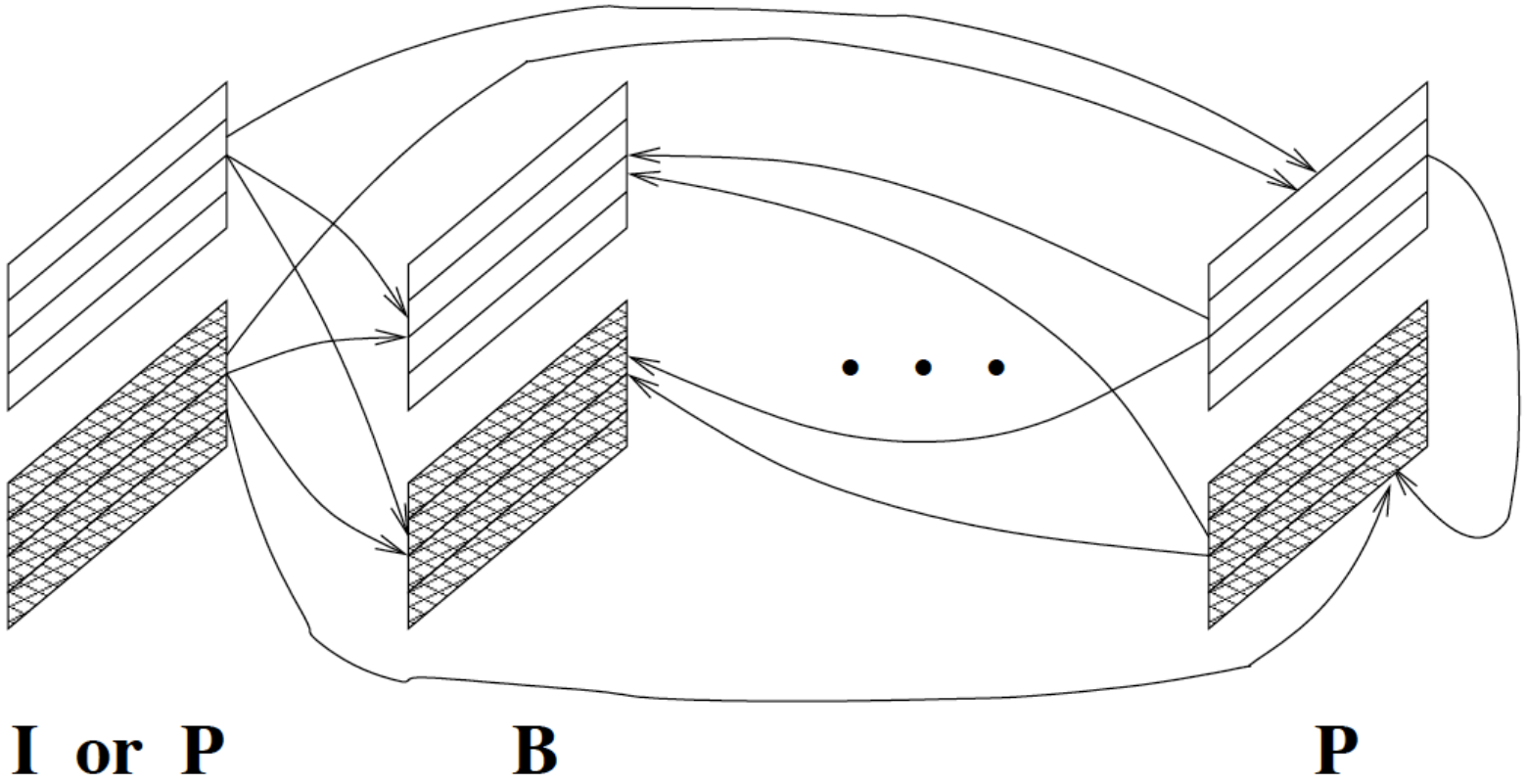
- Frame pictures are preferred for relatively still images and field pictures give better results in presence of significant motion.
- It is possible to switch between the frame picture and the field picture on a frame-by-frame basis.
- Each frame picture or a field picture may be I-type, P-type or B-type.

Field and Frame Prediction

- Frame prediction for frame-pictures: Identical to MPEG-1 MC-based prediction methods in both P-frames and B-frames.

Field and Frame Prediction

- Field Prediction for Field-pictures: A macroblock size of 16 x 16 from Field-pictures is used.



Field and Frame Prediction

- Field Prediction for Frame-pictures:
  - This mode treats the top-field and bottom field of a frame-picture separately
  - Each 16x16 macroblock from the target frame-picture is split into two 16x8 parts, each coming from one field
  - Field prediction is carried out for these 16x8 parts in similar manner

Field and Frame Prediction

- Dual Prime Prediction: Two independent predictions are made - one for the 8 odd (top) field lines and another for the 8 even (bottom) field lines

Scalability Support of MPEG-2

- Bit stream is organized into layers
  - A base layer and one or more enhancement layers
  - The base layer can be independently encoded, transmitted and decoded to obtain basic video quality
  - The encoding and decoding of the enhancement layer is dependent on the base layer or the previous enhancement layer

Scalable Coding Schemes

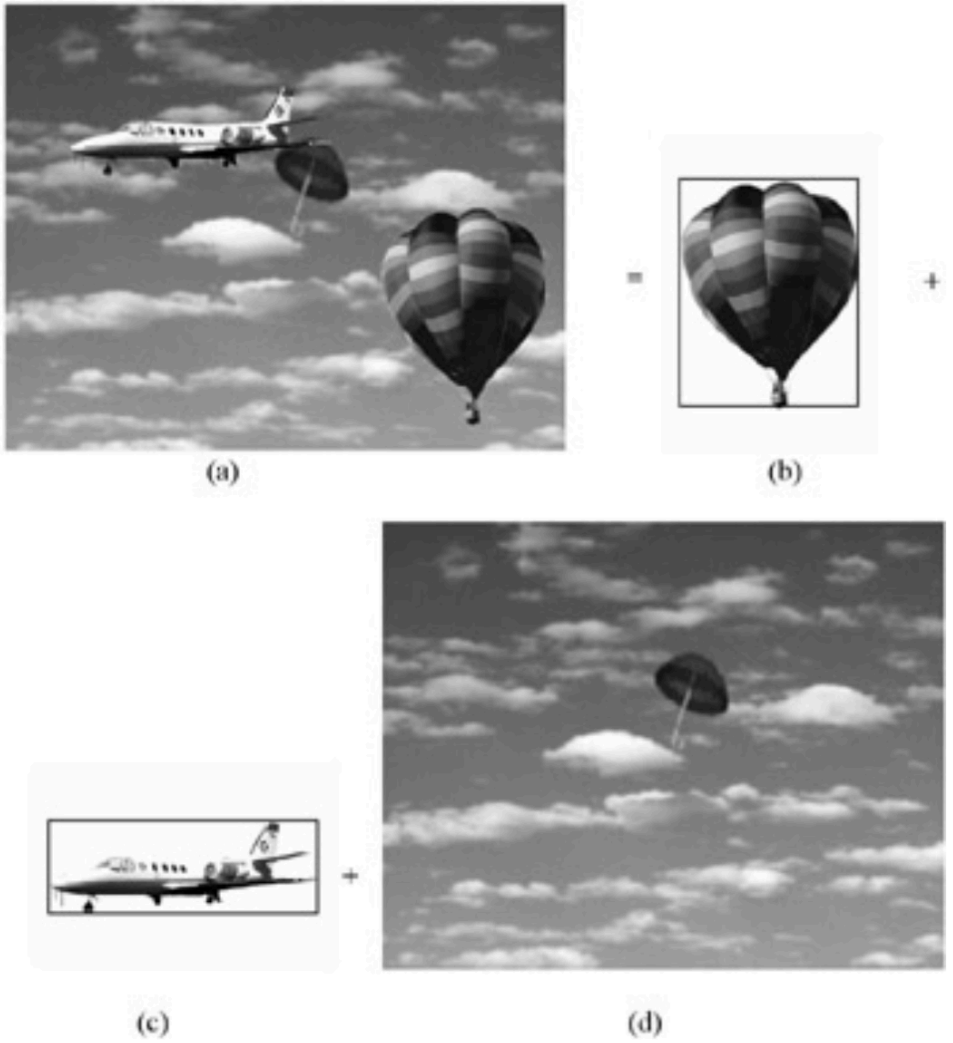
- SNR scalability: The enhancement layer provides higher SNR.
- Spatial scalability: The enhancement layer provides higher spatial resolution.
- Temporal scalability: The enhancement layer facilitates higher frame rate.
- Hybrid scalability: This combines any two of the above three scalabilities.
- Data partitioning: Quantized DCT coefficients are split into partitions.

MPEG-4

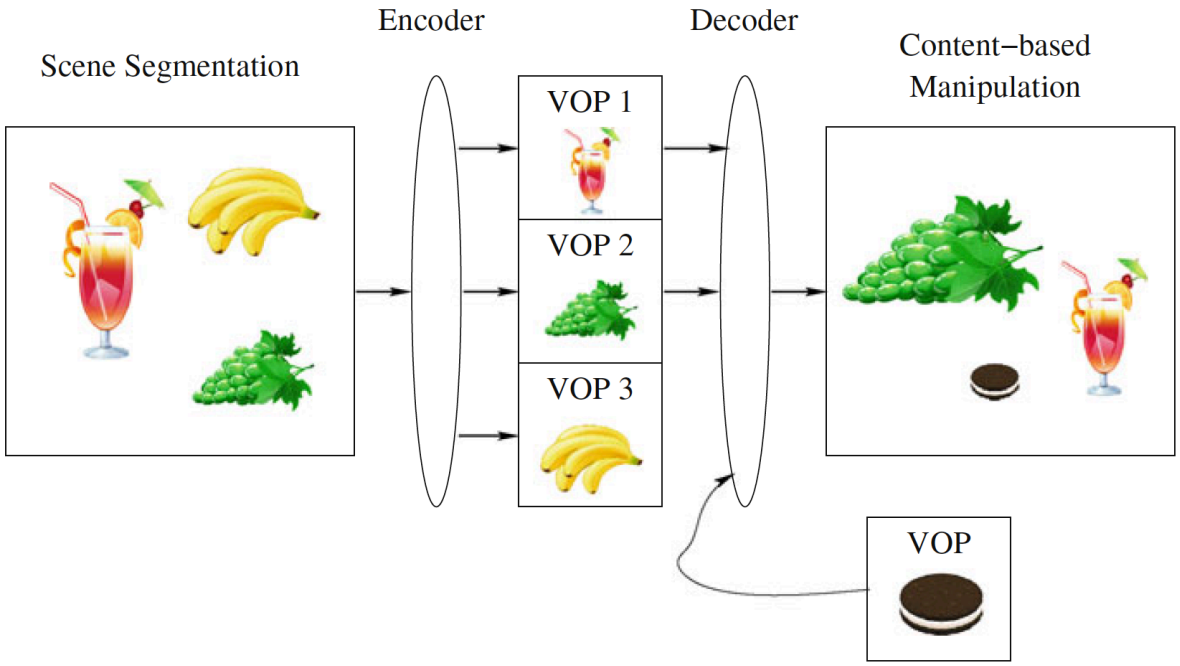
- Supports content-based manipulation and bit-stream editing
- Ability to combine synthetic and natural scene or objects
- Scalability: Automated selection of decoded quality of objects in the scene
- Content-based storage and retrieval

Content-Based Video Coding

- Original scene composed of Video Object Planes (VOPs)
  - Balloon
  - Airplane
  - Clouds/background



Composition and manipulation of MPEG-4 videos



- Each of the contents (objects) are encoded and decoded separately

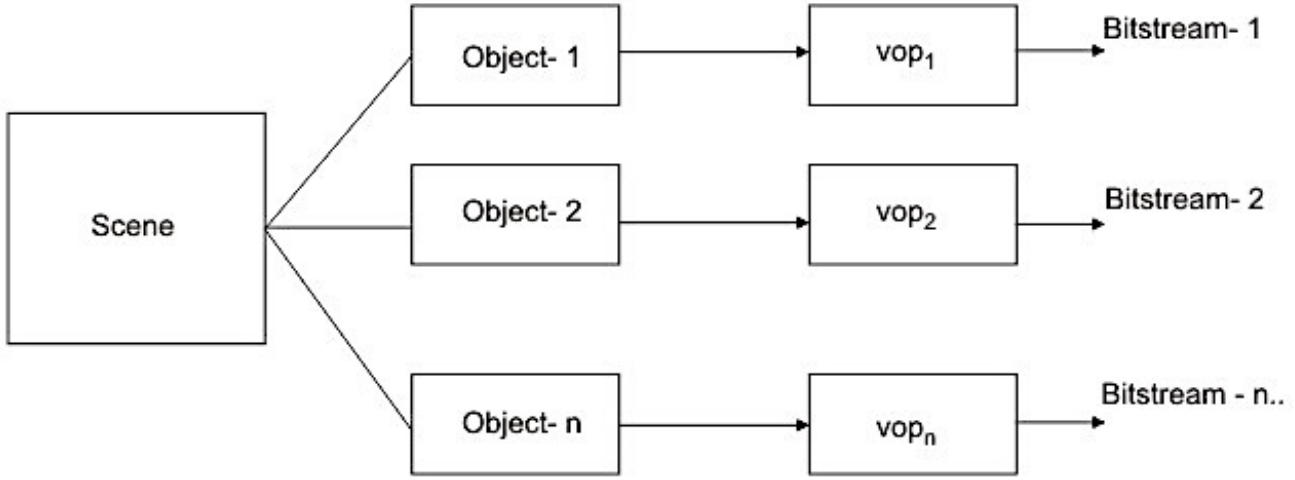
MPEG-Standards On Video Object Representation

- AVO and VO: A sequence is composed of one or more audio visual objects (AVO)
  - Audio objects from speech, music, sound effects, etc.
  - Video object representing a specific object
  - VO may be present over multiple frames
- VOP (Video Object Plane): A snapshot of VO in one frame
  - Most Elementary form of content representation

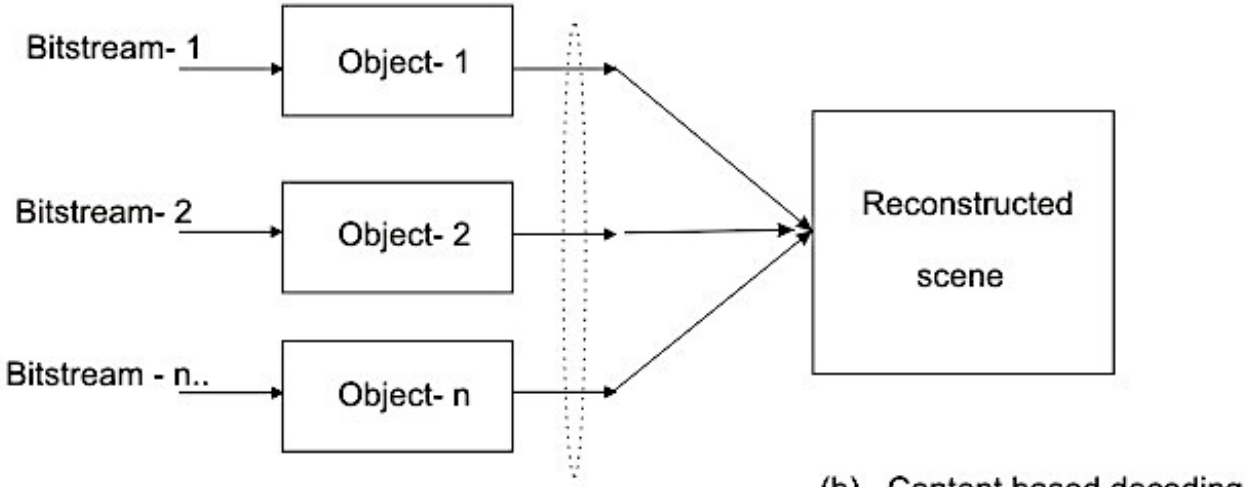
MPEG-Standards On Video Object Representation…

- (VOL) Video Object Layer: Shape, motion and texture information of the VOPs belonging to same VO is encoded and transmitted into the VOL
  - Bit-stream include information on how to combine the different VOLs to reconstruct video
- Each VOL has three components
  - Shape (contour) coding
  - Motion estimation and compensation
  - Texture coding

MPEG-4 Encoding/Decoding



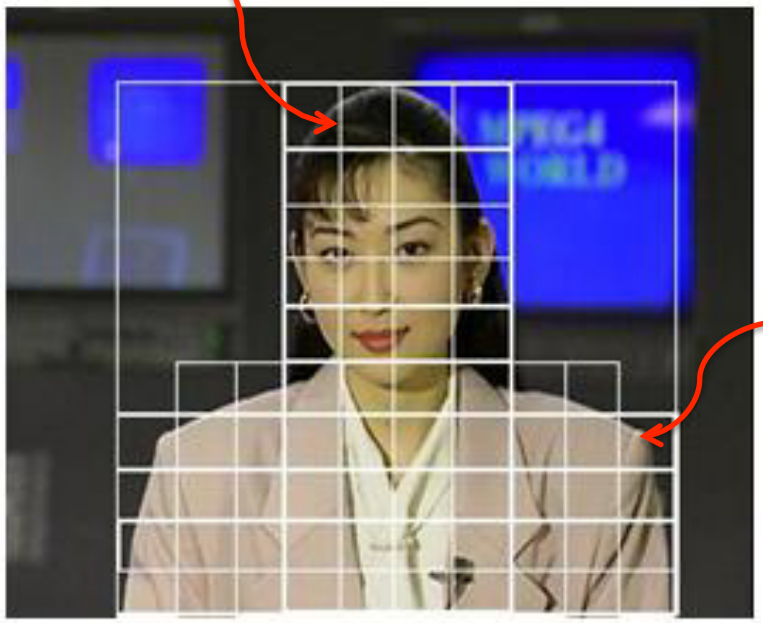
(a) content - based encoding



(b) Content based decoding

Shape Coding

Video Object Plane



Binary Alpha block

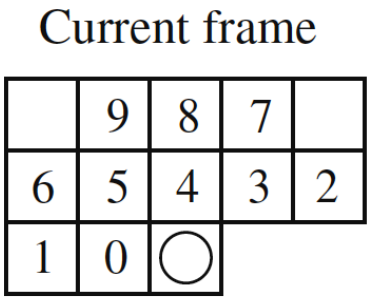
VOP Window and VOP macroblocks

Binary alpha-plane

- Bounding box approach

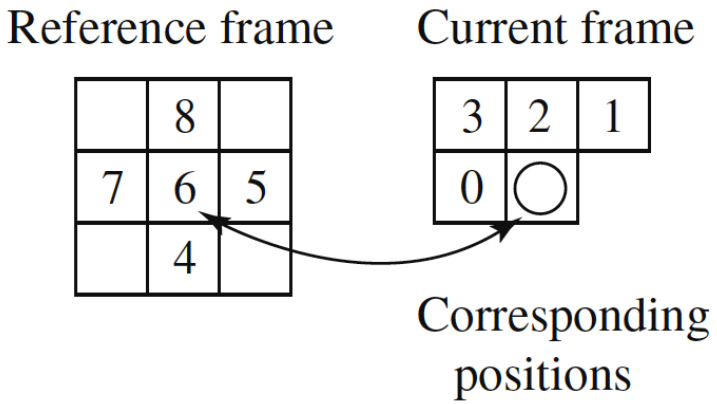
Shape Coding

- Context-Based Arithmetic Encoding (CAE)
  - Context for pixel in BAB
- Intra-CAE
  - Ten neighboring pixels form (2^{10} = 1024) context
  - Probability table is built to indicate probability of occurrence for each 1024 context
  - CAE eventually provide one floating point number to each BAB



Shape Coding

- Inter-CAE
  - Involves both reference and target frame
  - Motion estimation and compensation is invoked to locate matching macroblock in the reference frame
  - Context include 4 neighboring pixels from target alpha map and 5 from reference alpha map
  - Each pixel in BAB is assigned one of the $2^9 = 512$ probabilities
  - CAE algorithm is applied



Texture Coding

- Bounding box and grids are used
- Without motion compensation in case of I-VOPs
- On residual errors after motion compensation in P-VOPs and B-VOPs

Sprite Coding

- Sprite: A graphic image that can freely move around within a larger graphic image or set of images
- Sprite panorama: A still image that describes the static background over a sequence of video frames
- A sprite panorama image is encoded and sent to decoder only once
- Decoder receives separately coded foreground objects and camera parameters to reconstruct scene

Object Coding and Animation

- 3-D wire mesh model of a human head and shoulders
- A sprite image of a person is mapped on to the 3-D surface to represent the texture details of the person
- Later, only few parameters, that represent the motion are transmitted