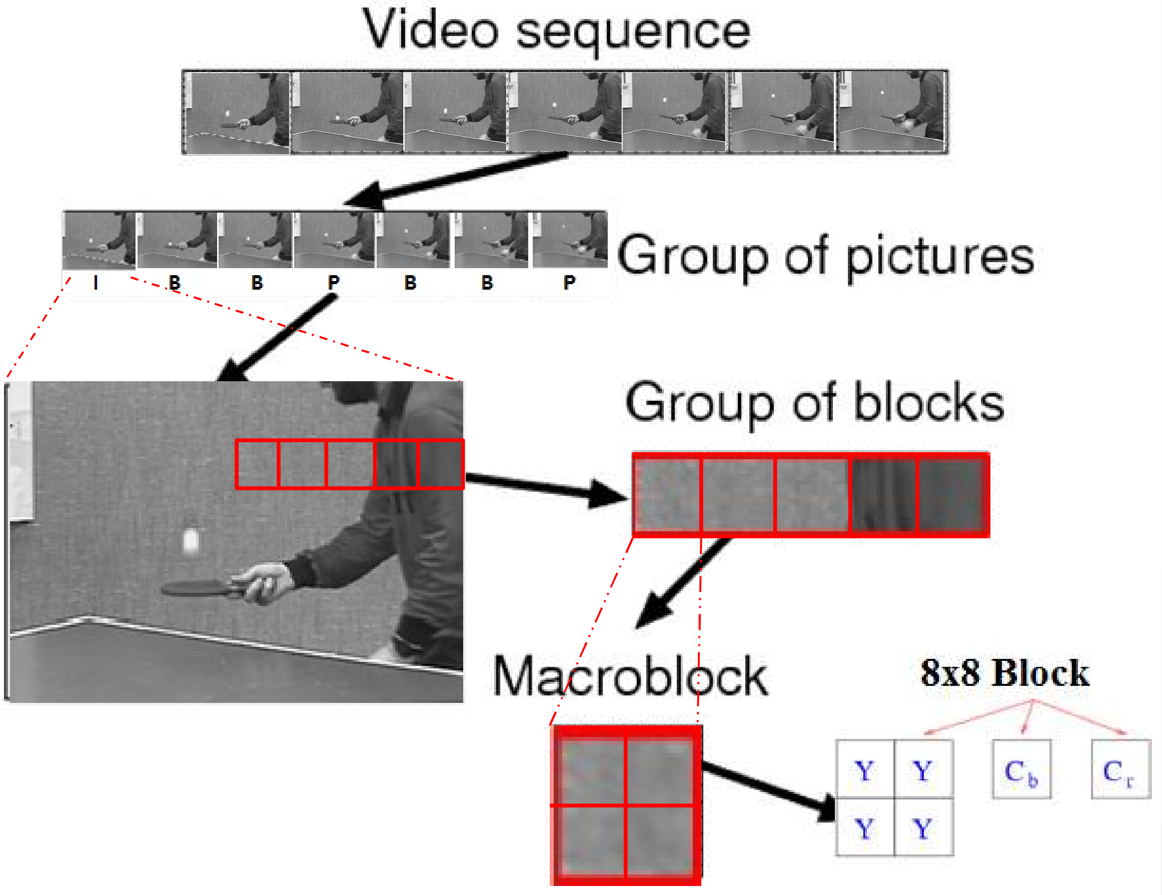
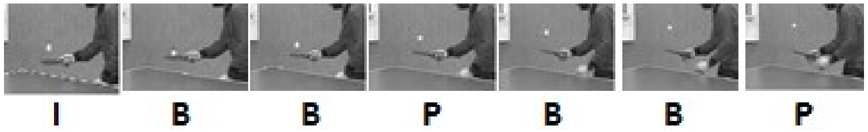
Data Hierarchy

- A hierarchy of data structures in the video stream.
- The hierarchy begins with the video sequence layer and ends with the block layer.
- For the picture layer, MPEG-1 specifically defines three types of pictures: intra, predicted, and bidirectional.



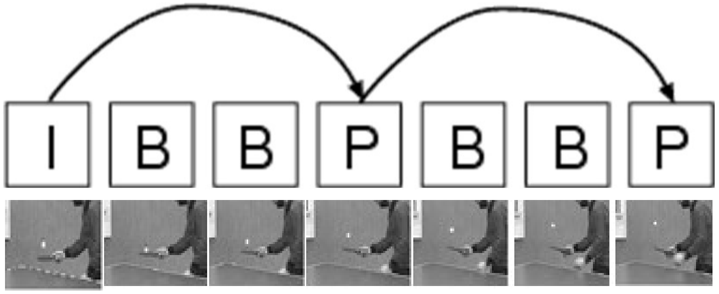
Intra Pictures

- Intra pictures (I-pictures) are coded using only the current frame information and is coded by transform coding technique. They provide random access points into the compressed video data. I-pictures give a moderate compression.



Predicted Pictures

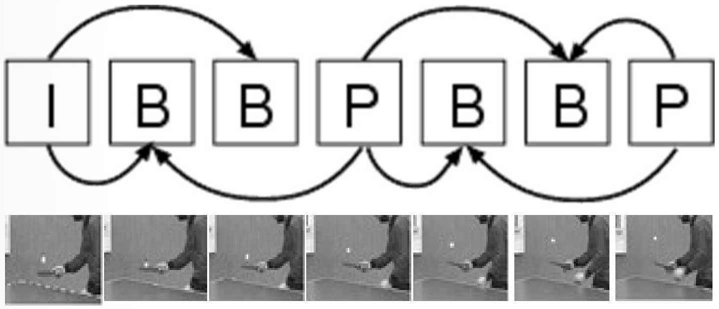
- Predicted pictures (P-pictures) are coded using the nearest previous I or P-picture. This is called forward prediction and is illustrated in figure.

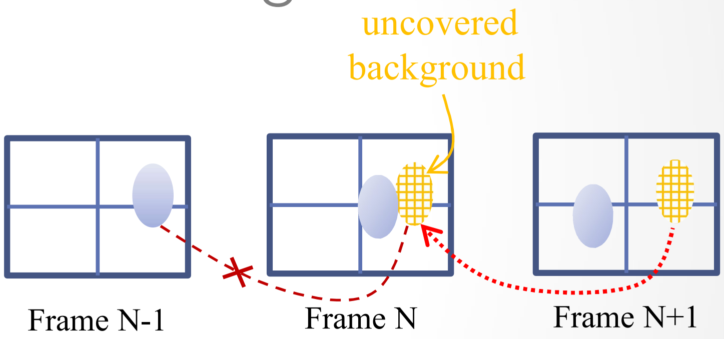


- P-pictures also serve as a prediction reference for bidirectional pictures. P-pictures can always achieve higher compression than I-pictures. Errors can be propagated by P-pictures since P-pictures can be referenced by other predicted or bidirectional pictures.

Bidirectional Pictures

- Bidirectional pictures (B-pictures) are coded using both past and future pictures as reference. This technique is called bidirectional prediction and is illustrated in figure.

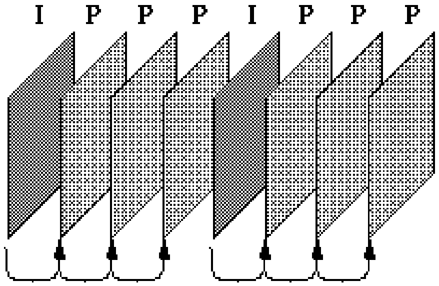




- B-pictures give the most compression and do not propagate errors since they are never used as a reference. They can solve the problem of uncovered background by reference a future picture instead of a previous one. Moreover, they decrease the noise effect by averaging two reference pictures.

H.261

- **1. Overview of H. 261**
- Developed by CCITT (Consultative Committee for International Telephone and Telegraph) in 1988-1990
- Designed for videoconferencing, video-telephone applications over ISDN telephone lines. Bit-rate is *p* x 64 Kb/sec, where *p* ranges from 1 to 30.
- Frame types are CCIR 601 CIF (352 x 288) and QCIF (176 x 144) images with 4:2:0 subsampling.

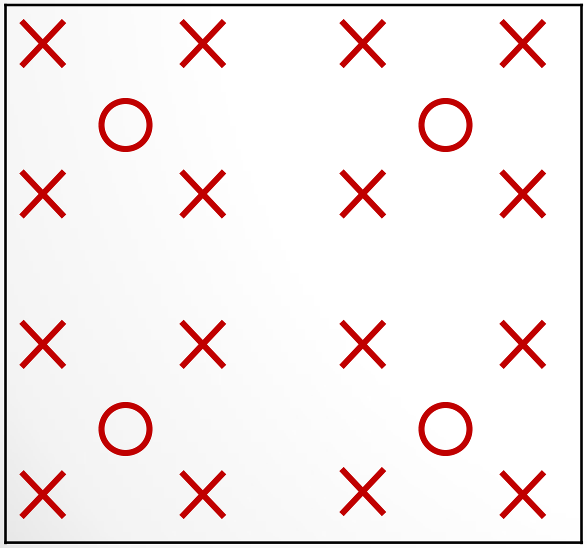


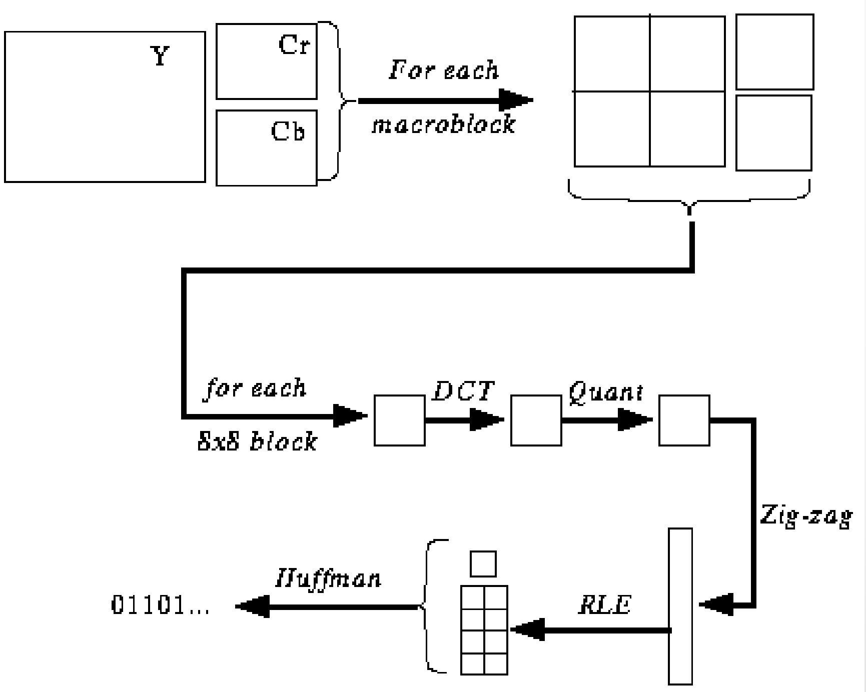
- Two frame types: Intra-frames (*I-frames*) and Inter-frames (*P-frames*): I-frame provides an accessing point, it uses basically JPEG.
- P-frames use "pseudo-differences" from previous frame ("predicted"), so frames depend on each other.

Intra-Frame Coding

- Macroblocks are 16 x 16 pixel areas on Y plane of original image. A macroblock usually consists of 4 Y blocks, 1 Cr block, and 1 Cb block.
- Quantization is by constant value for all DCT coefficients (i.e., no quantization table as in JPEG).

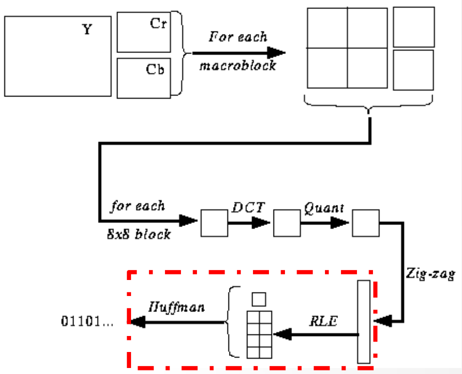
4:2:0 chroma sub-sampling

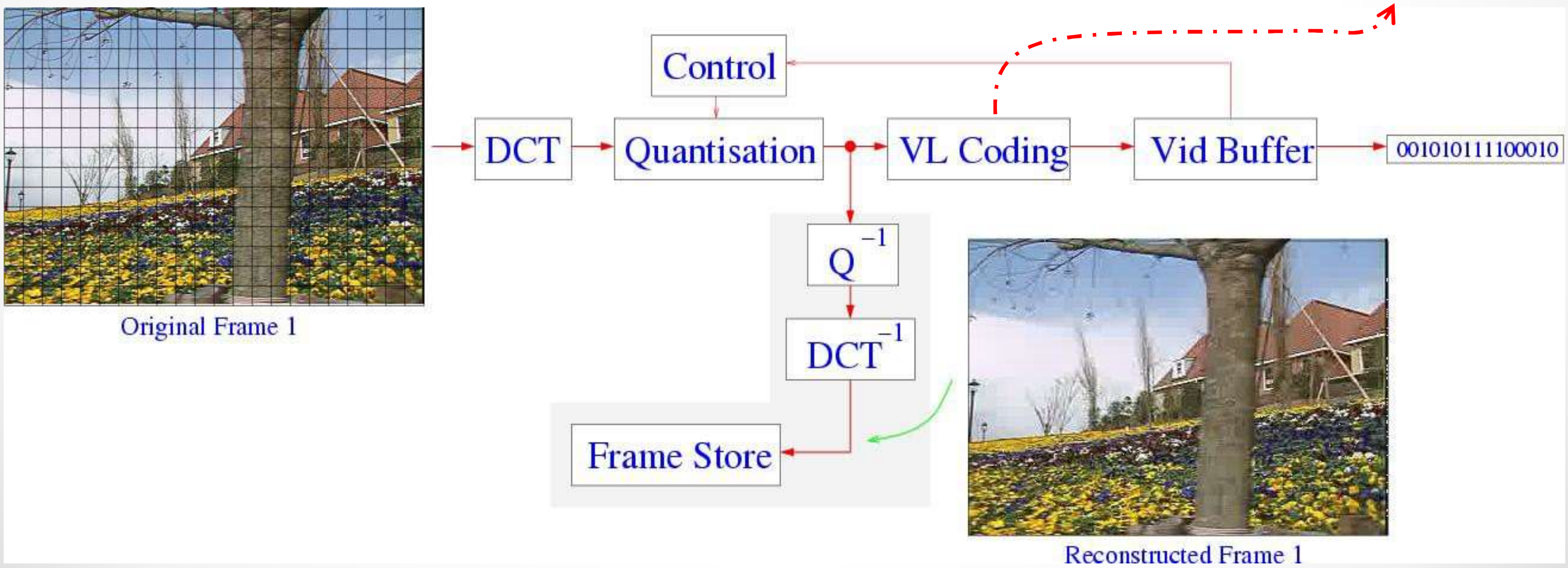




Intra-Frame Coding

- Intra frames (I-frames) are coded using only the current frame information and is coded by transform coding technique. They provide random access points into the compressed video data. I-pictures give a moderate compression.

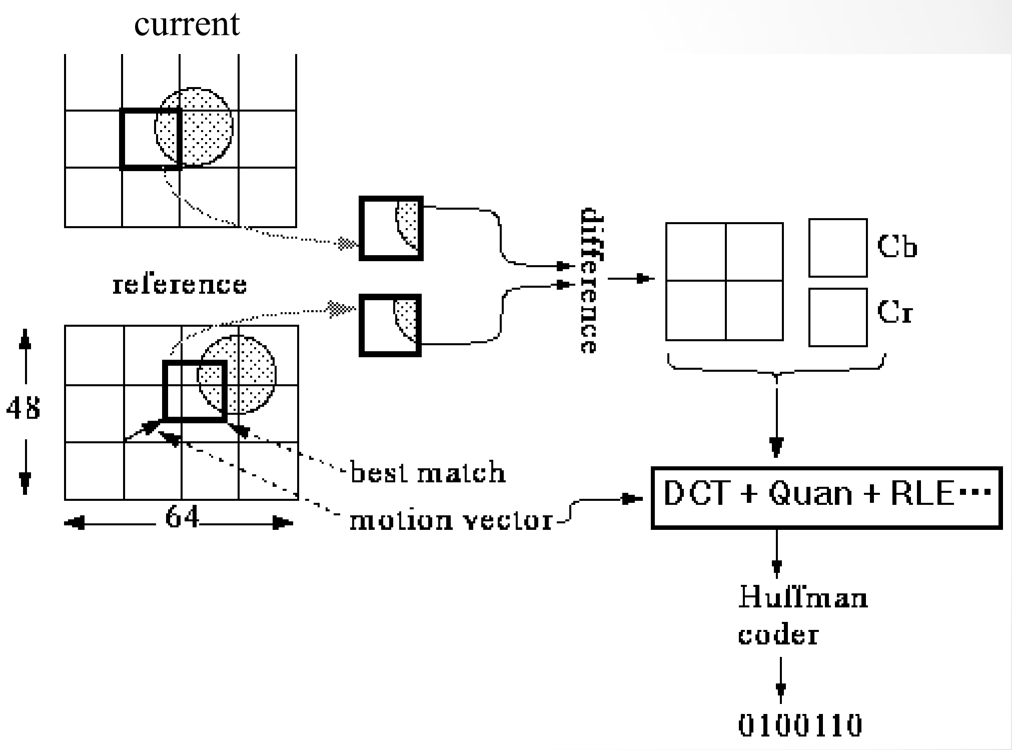




Inter-Frame (P-frame) Coding

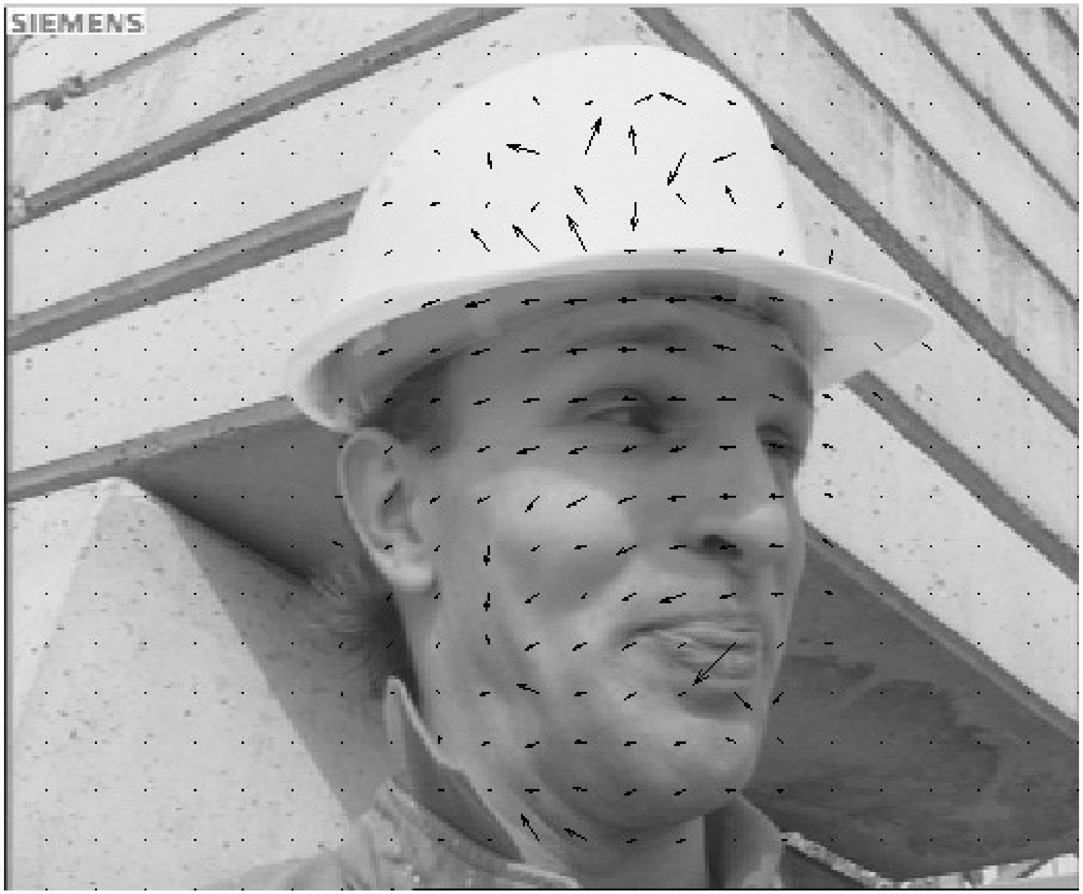
- Previous image is called *reference image*, the image to encode is called *target image*.
- **Points to emphasize:**

1. The difference image (not the target image itself) is encoded.
2. Need to use the decoded image as reference image, *not* the original.



- Motion compensated video coder

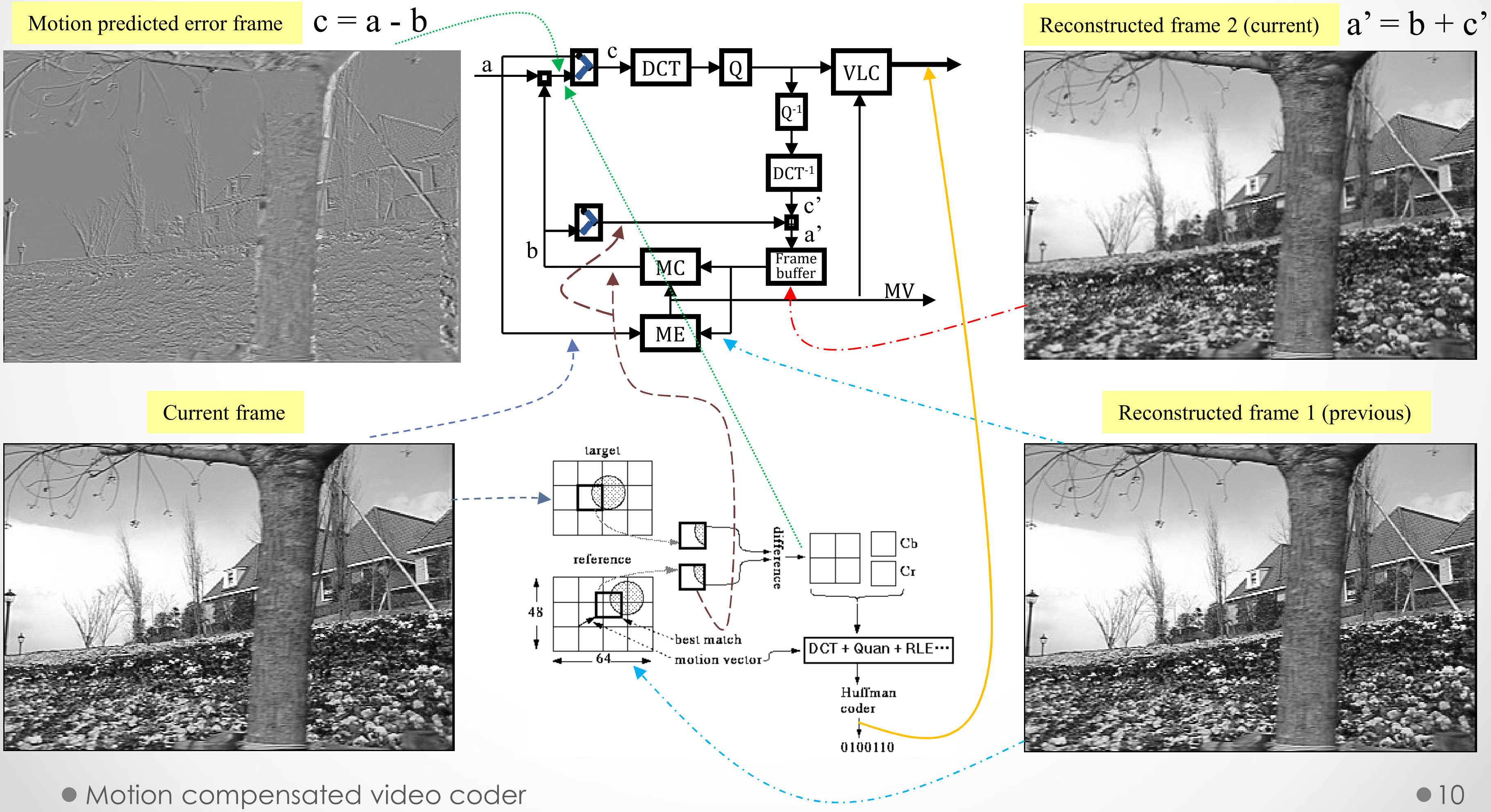
Inter-Frame (P-frame) Motion Vector



- Motion compensated video coder

Inter-Frame Motion Compensation

- Help reduce temporal redundancy of video



Motion predicted error frame $c = a - b$

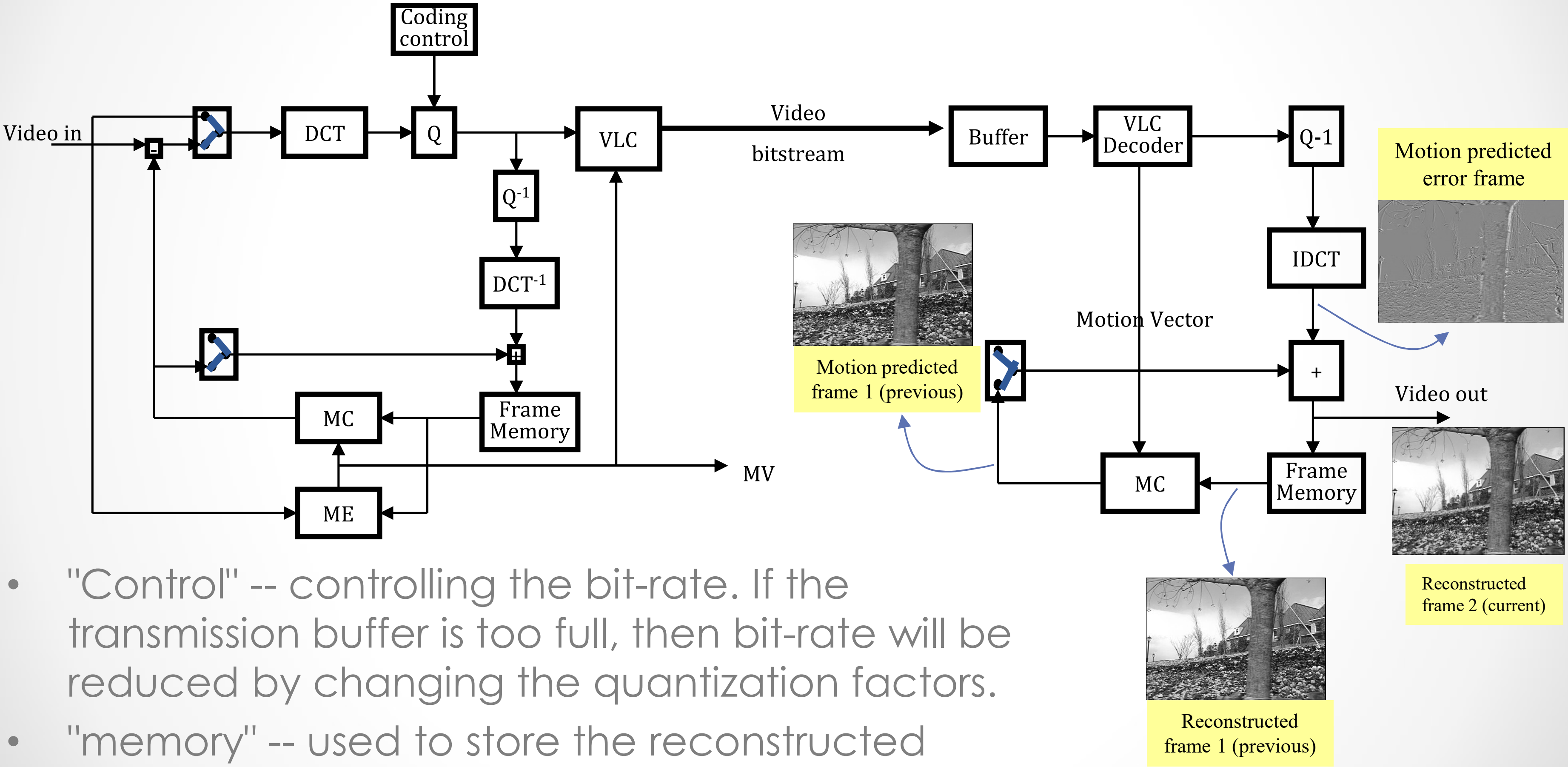
Reconstructed frame 2 (current) $a' = b + c'$

Current frame

Reconstructed frame 1 (previous)

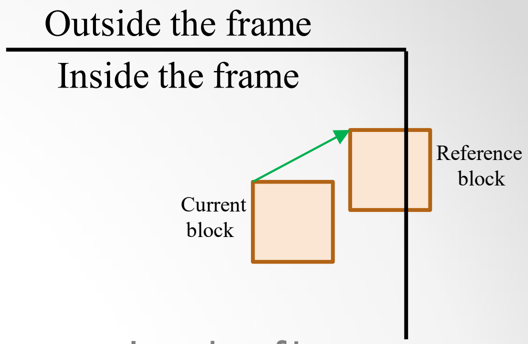
Motion compensated video coder

H. 261 encoder/decoder



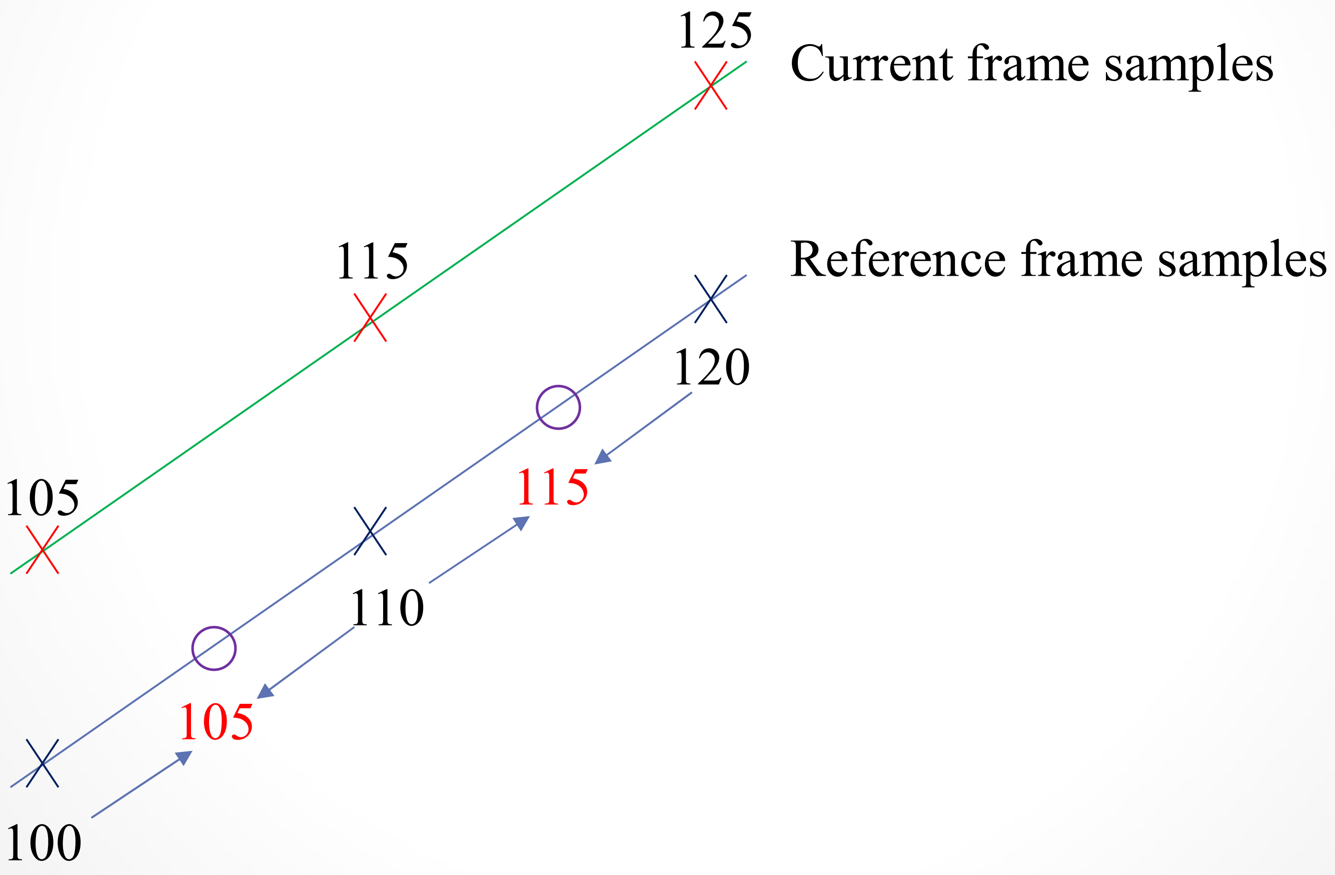
- "Control" -- controlling the bit-rate. If the transmission buffer is too full, then bit-rate will be reduced by changing the quantization factors.
- "memory" -- used to store the reconstructed image (blocks) for the purpose of motion vector search for the next P-frame.
  - Motion compensated video coder

H.263



- It was designed for low bitrate communication, early drafts specified data rates less than 64 Kbits/s.
- The coding algorithm of H.263 is similar to that used by H.261, however with some improvements and changes to improve performance and error recovery.
- The differences between the H.261 and H.263 coding.
  - Half pixel precision is used for motion compensation whereas H.261 used full pixel precision and a loop filter.
  - Unrestricted Motion Vectors,
  - Syntax-based arithmetic coding (SAC): SAC used instead of VLC.
  - Advance prediction, and forward and backward frame prediction similar to MPEG called P-B frames.

Half pixel precision

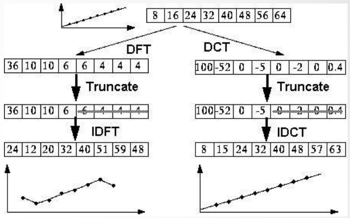


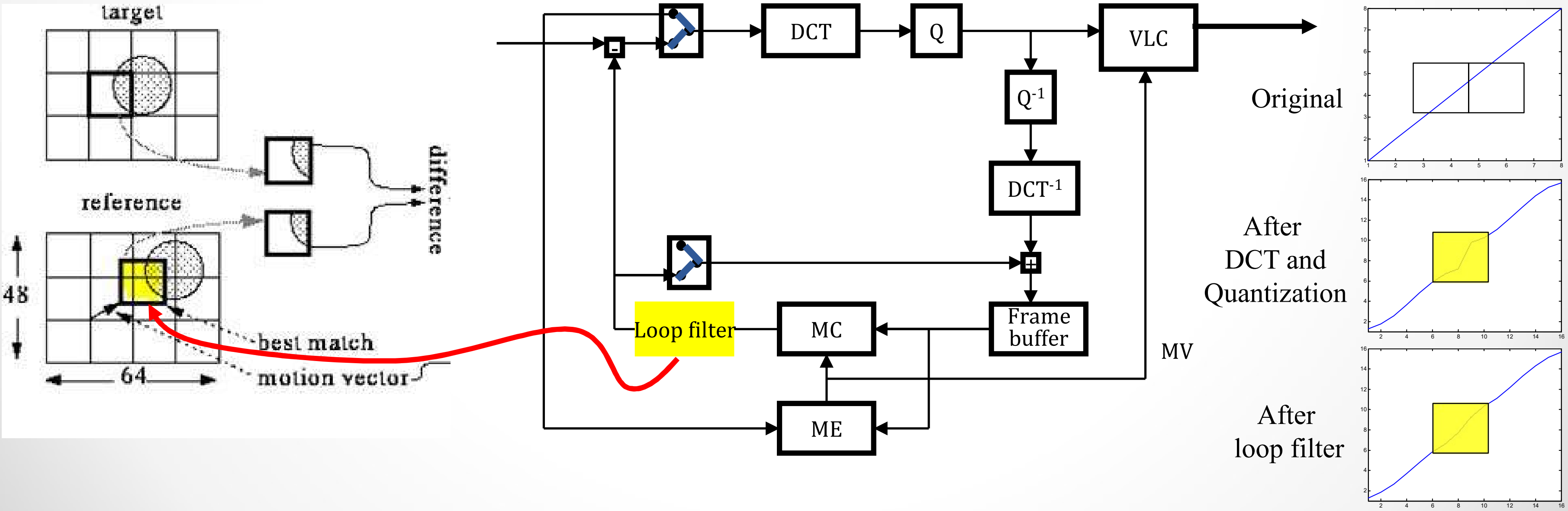
Syntax-based arithmetic coding (SAC)

(Run,Size)	~~Code Word~~	[lower, upper)	(Run,Size)	~~Code Word~~	[lower, upper)
(0,1)	~~00~~	$[0,l_1)$	(0,6)	~~1111000~~	
(0,2)	~~01~~	$[l_1,l_2)$	(1,3)	~~1111001~~	
(0,3)	~~100~~	$[l_2,l_3)$	(5,1)	~~1111010~~	
(EOB)	~~1010~~	:	(6,1)	~~1111011~~	
(0,4)	~~1011~~	:	(0,7)	~~11111000~~	
(1,1)	~~1100~~	:	(2,2)	~~11111001~~	
(0,5)	~~11010~~		(7,1)	~~11111010~~	
(1,2)	~~11011~~		(1,4)	~~111110110~~	
(2,1)	~~11100~~				
(3,1)	~~111010~~		(ZRL)	~~11111111001~~	
(4,1)	~~111011~~				

Loop filter

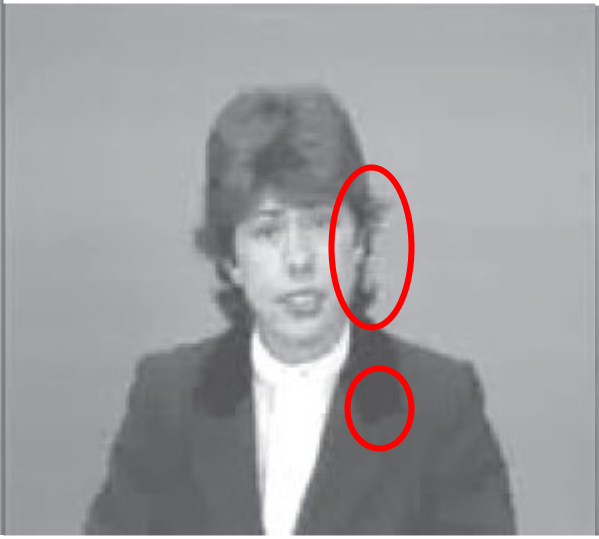
- Motion compensation block obtained from frame buffer may cross a few blocks.
- Blocking effect for integer pixel is obvious.
- Loop filter is separable into one-dimensional horizontal and vertical functions. Both are non-recursive with coefficients of ¼, ½ , ¼.
- Edge blocks have coefficients of 0, 1, 0.



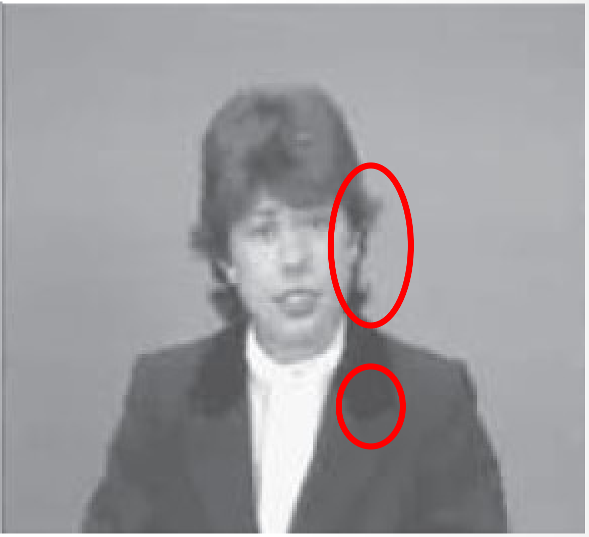


Loop Filter

Original



128kbps

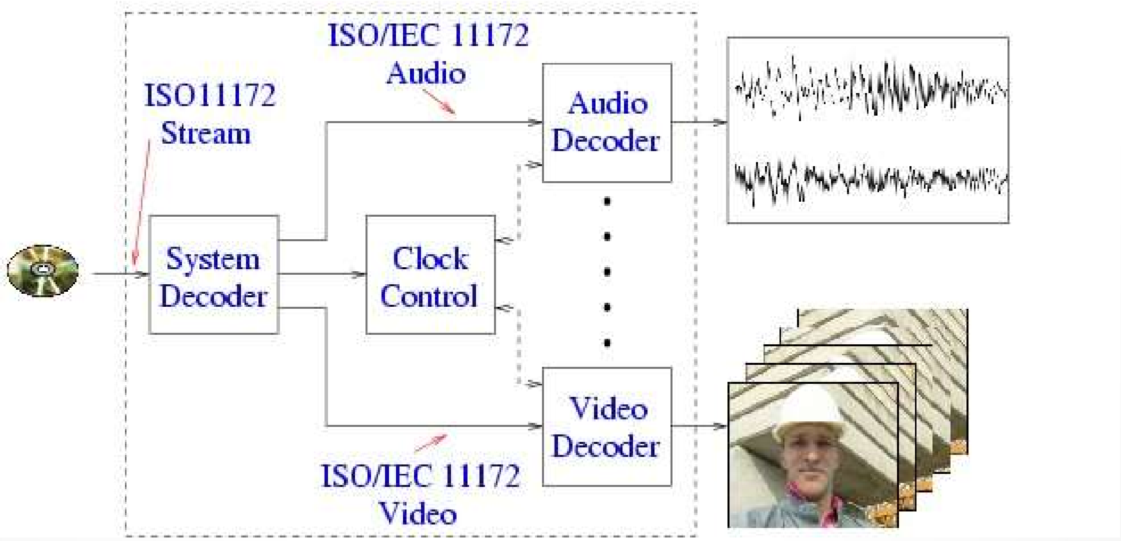


128kbps with loop filter

** Image from: Standard Codecs Image compression to advanced video coding, 3rd Edition, Mohammed Ghanbari

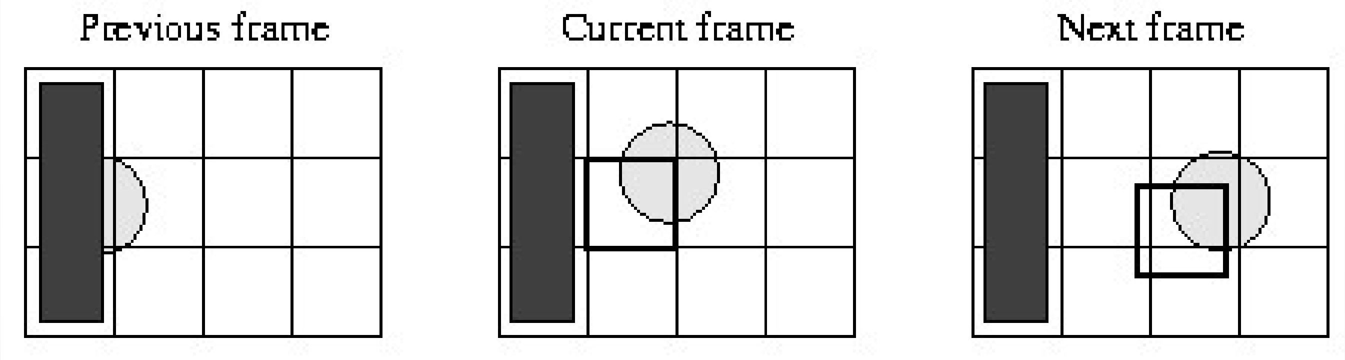
MPEG-1

- MPEG-1 officially referred to as ISO 11172 developed for CD-ROM applications up to about 1.5 Mbps (System part covered in Part 1).



MPEG-1

- MPEG-1 Target: VHS quality on a CD-ROM or Video CD (VCD) (352 x 240 + CD audio @ 1.5 Mbits/sec).
- Problem: some macroblocks need information not in the previous reference frame. Example: The darkened macroblock in Current frame does not have a good match from the Previous frame, but it will find a good match in the Next frame.



- Motion compensated video coder

MPEG-1

- A close GOP is an encoding of a sequence of frames that contain all the information that can be completely decoded within that GOP. For all frames within a GOP that reference other frames (P/B frames), the reference frames (I/P frames) are also included within that same GOP.

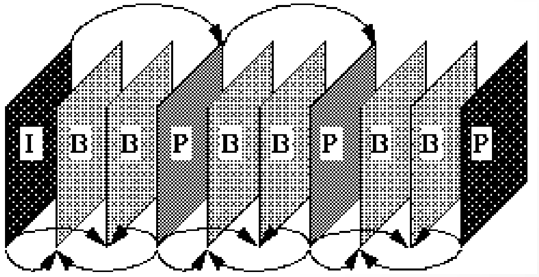
Close GOP   I B B P B B P   I B B P B B P

- An Open GOP will contain some references from the previous GOP.

Open GOP   B B I B B P B B P   B B I B B P B B P

- Actual pattern is up to encoder, and need not be regular.
- Example:

176*144, 4:2:0,   10fps, 8 bits
Average Compression Ratio = 50
Bitrate=176*144*1.5*10*8/50
= 60.8256 kbps
10 min. recording: 60.8256 kbps*10*60
= 4561920Bytes

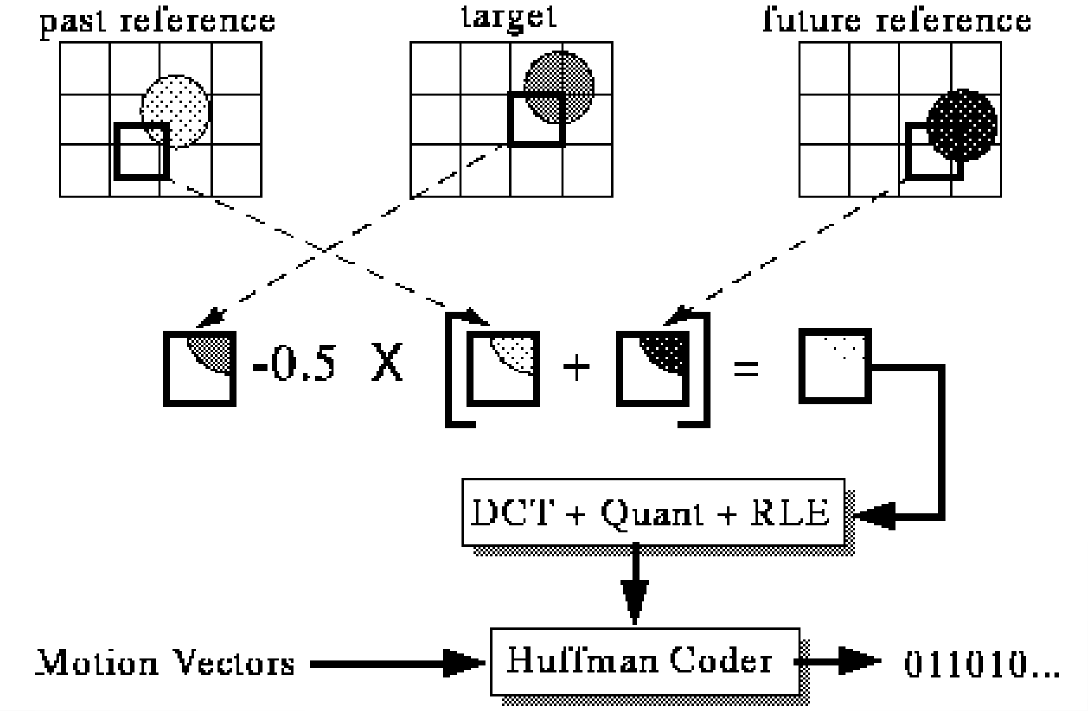


MPEG-1

- Differences from H. 261
  - o Larger gaps between I and P frames as B frames are used, so need to expand motion vector search range.
  - o To get better encoding, allow motion vectors to be specified to fraction of a pixel (1/2 pixel).
  - o Bitstream syntax must allow random access, forward/ backward play, etc.
  - o Added notion of slice for synchronization after loss/corrupt data.

MPEG-1

- B frame macroblocks can specify two motion vectors (one to past and one to future), indicating result is to be averaged.

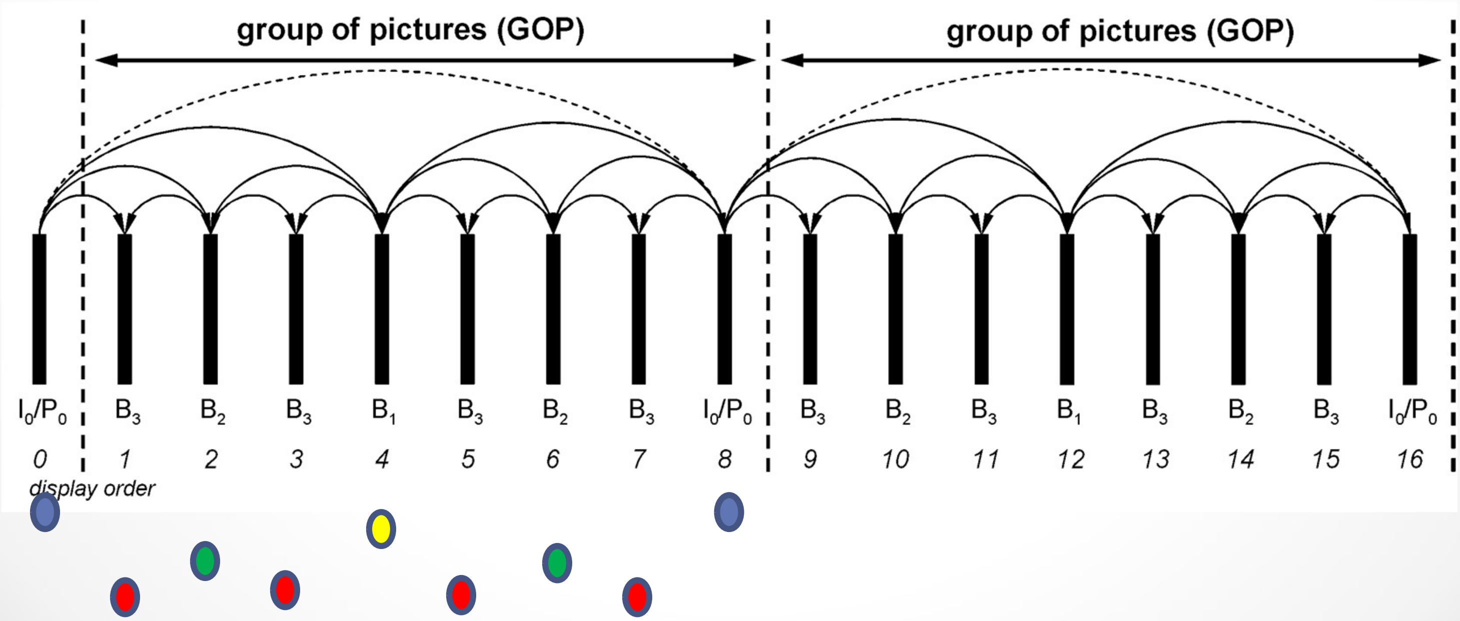


- Motion compensated video coder

Hierarchical B frame

- In new video coding standard (H.264), B frames can also be a reference frame for other frames in order to increase the coding efficiency.

Hierarchical B



MPEG-2

- Unlike MPEG-1 which is basically a standard for storing and playing video on a single computer at low bit-rates, MPEG-2 is a standard for digital TV. It meets the requirements for HDTV and DVD (Digital Video/Versatile Disc).
- Other Differences from MPEG-1:

1. Support both field prediction and frame prediction.
2. Besides 4:2:0, also allow 4:2:2 and 4:4:4 chroma subsampling
3. Scalable Coding Extensions: (so the same set of signals works for both HDTV and standard TV)
   - SNR (quality) Scalability -- similar to JPEG DCT-based Progressive mode, adjusting the quantization steps of the DCT coefficients.
   - Spatial Scalability -- similar to hierarchical JPEG, multiple spatial resolutions.
   - Temporal Scalability -- different frame rates.
4. Frame sizes could be as large as 16383 x 16383
5. Non-linear macroblock quantization factor, default quantization matrix can be overridden by other matrix that may be sent to the decoder in an embedded bitstream.

MPEG-2 field prediction

- The MB to be predicted is split into top field pels and bottom field pels. Each 16×8 field block is predicted separately. (Source: Textbook 1, Fig. 13.18, p.430)

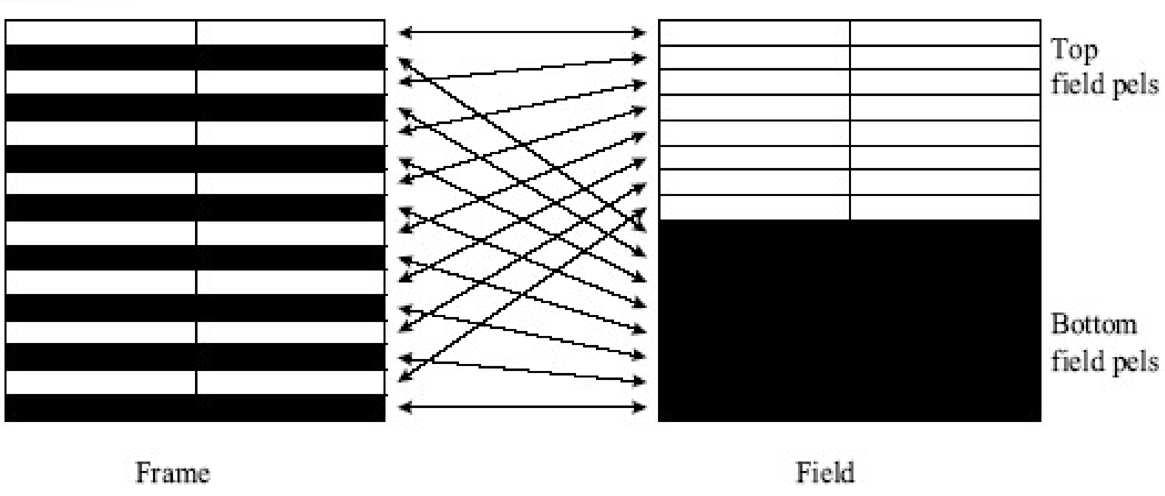
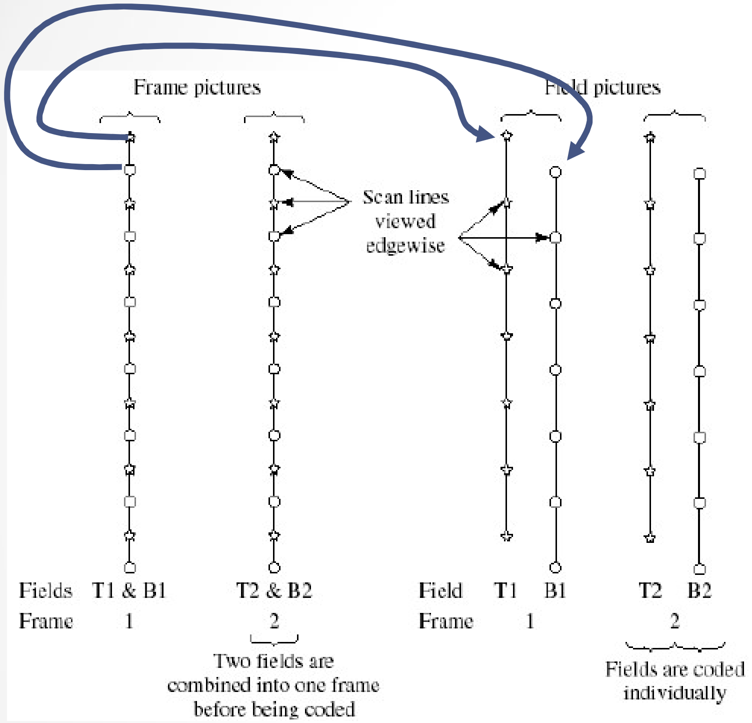


Image from https://en.wikipedia.org/wiki/Interlaced_video

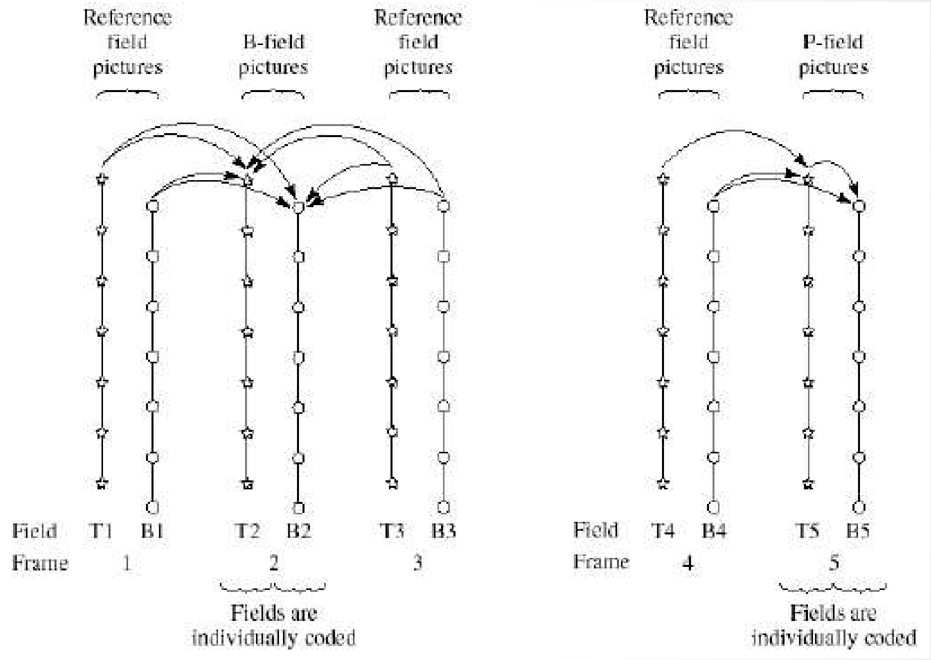
Field prediction for frame image: the MB to be predicted is split into top field and bottom field. Each 16x8 field block is predicted separately with its own motion vector for P frame and two motion vectors for B frame.

Source: textbook 1, fig. 13.18.

MPEG-2 field prediction



Frame and field image structure
(T: top field, B: bottom field)

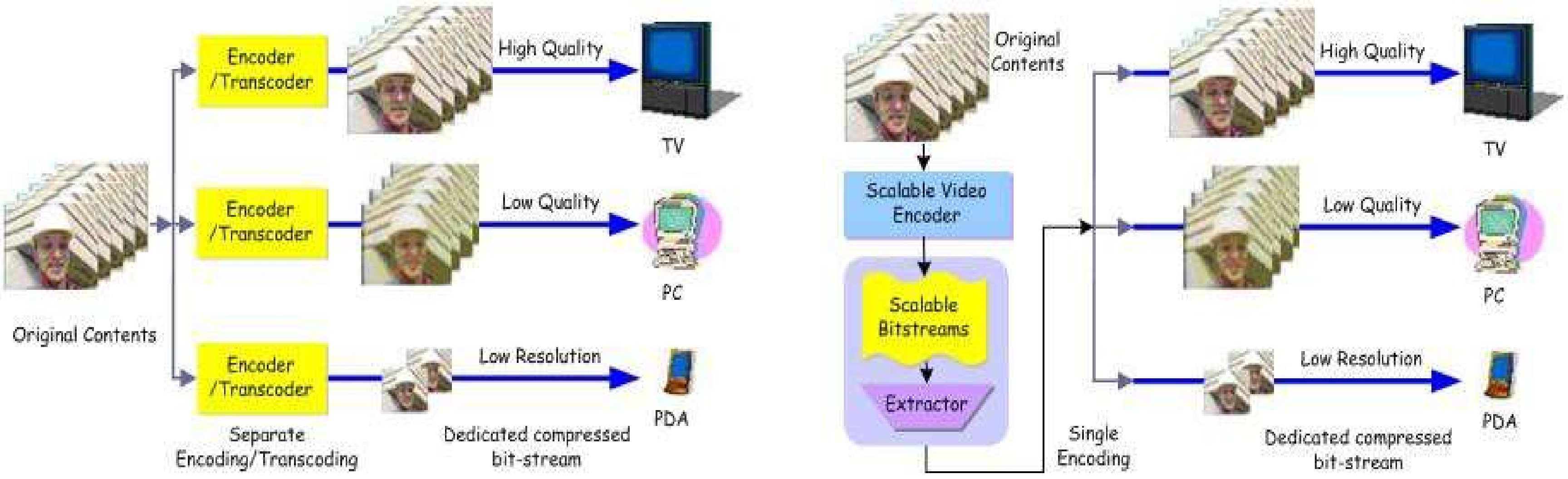


Every MB relevant for field prediction is located within one field of reference image

Source: textbook 1, fig. 13.16 and fig. 13.17.

Scalability

Non-Scalable v.s. Scalable

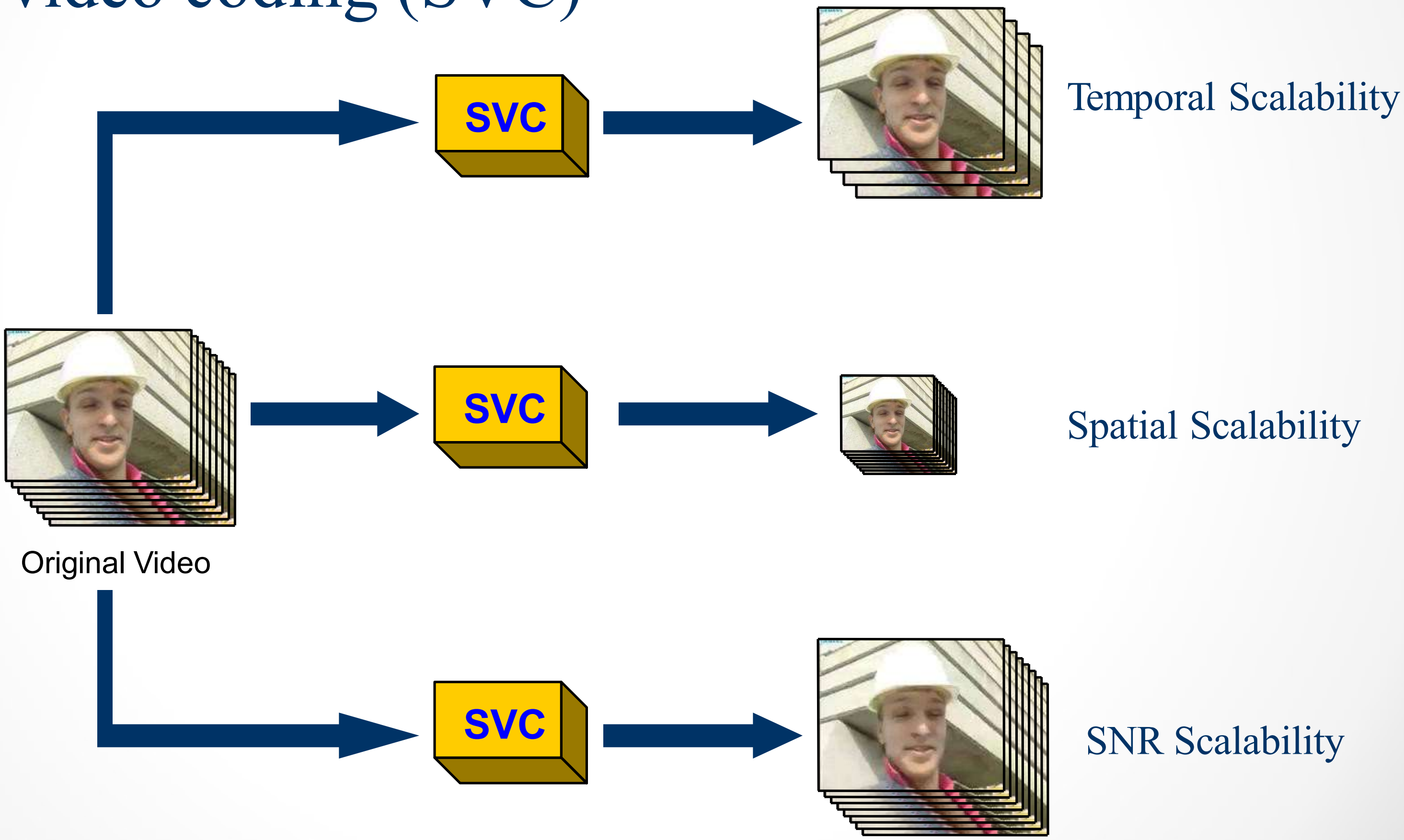


Non-Scalable

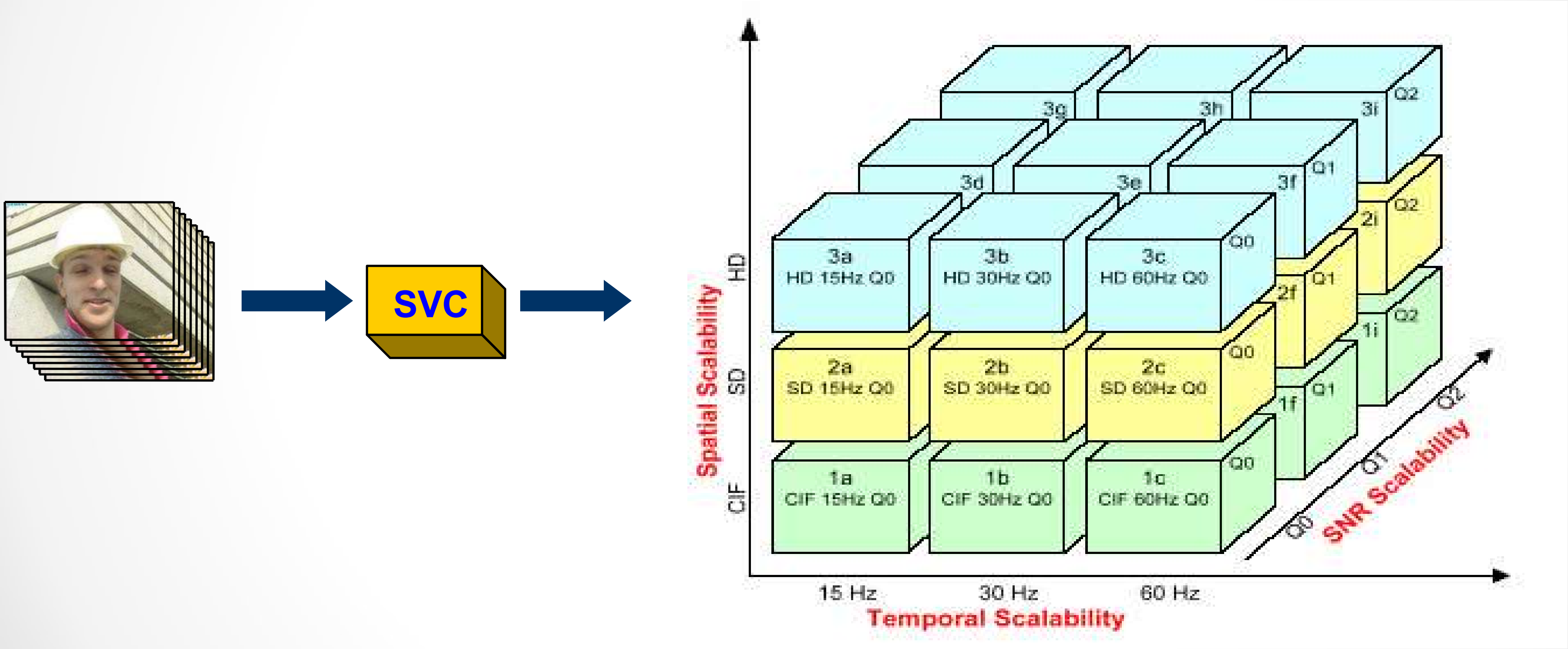
Scalable

Scalability

Scalable video coding (SVC)

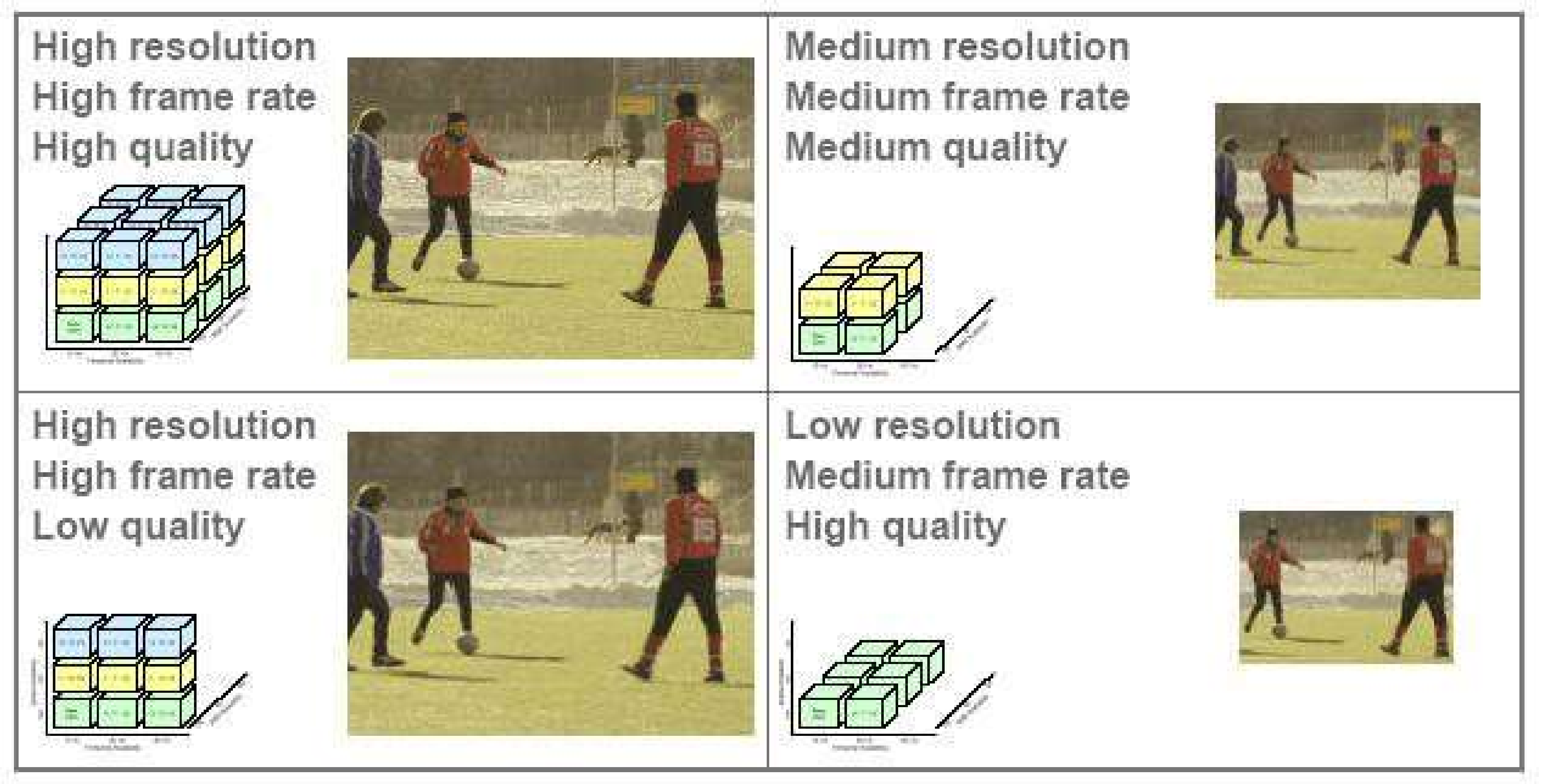


Scalability: One encoding



One embedded bitstream
(Layered coding)

Scalability: Multiple Decoding



High resolution
High frame rate
High quality

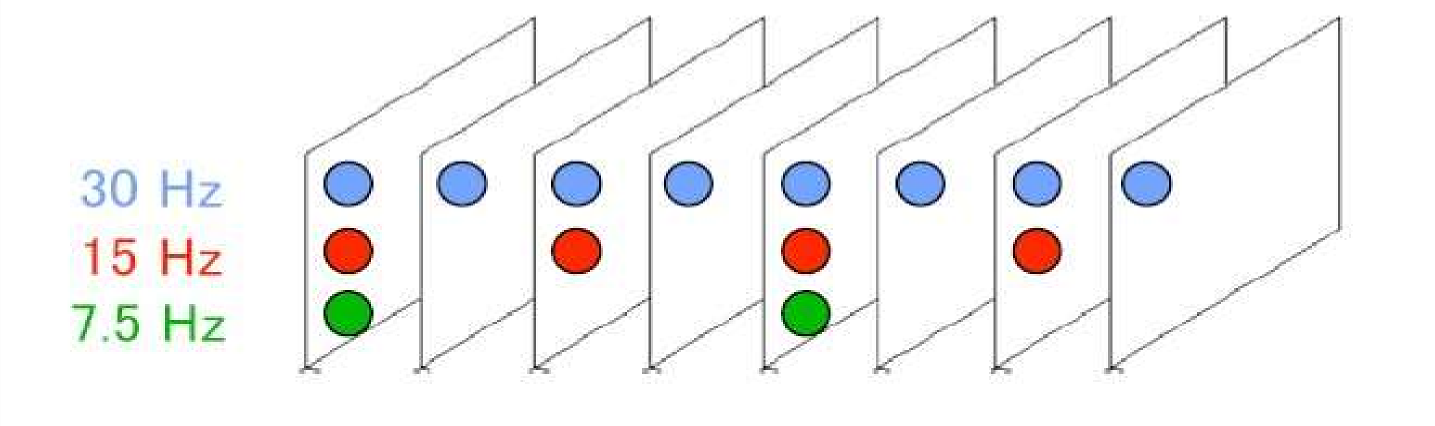
Medium resolution
Medium frame rate
Medium quality

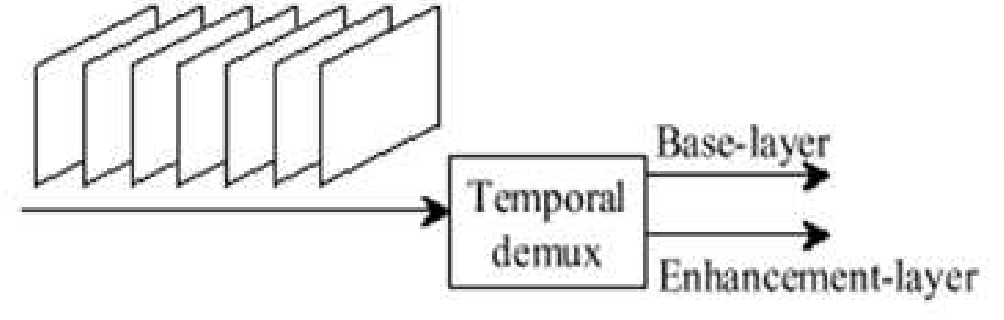
High resolution
High frame rate
Low quality

Low resolution
Medium frame rate
High quality

MPEG-2 Temporal Scalability

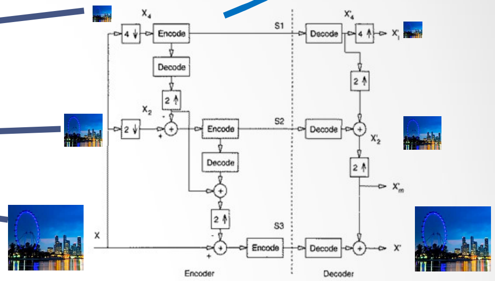
- Change of frame rate





MPEG-2 Spatial Scalability

- Change of frame size



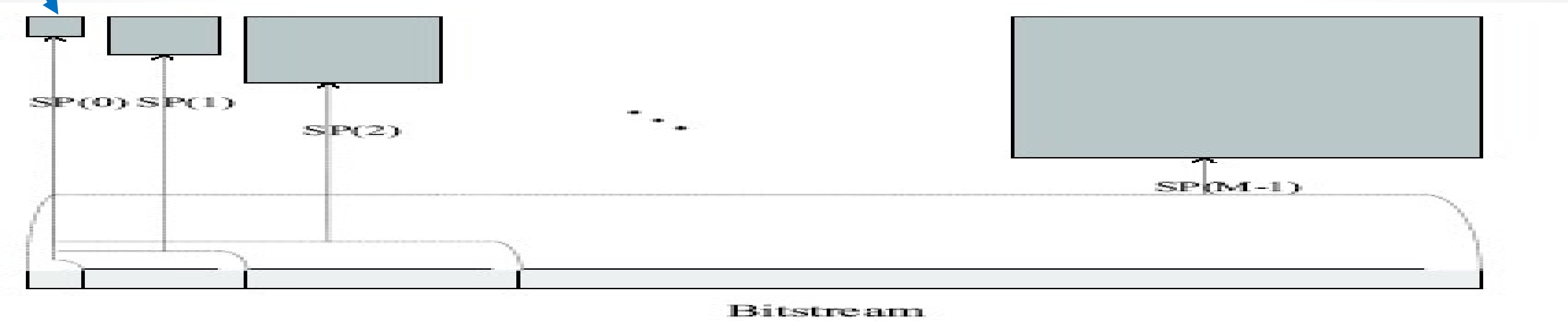
…

…

Mobile device

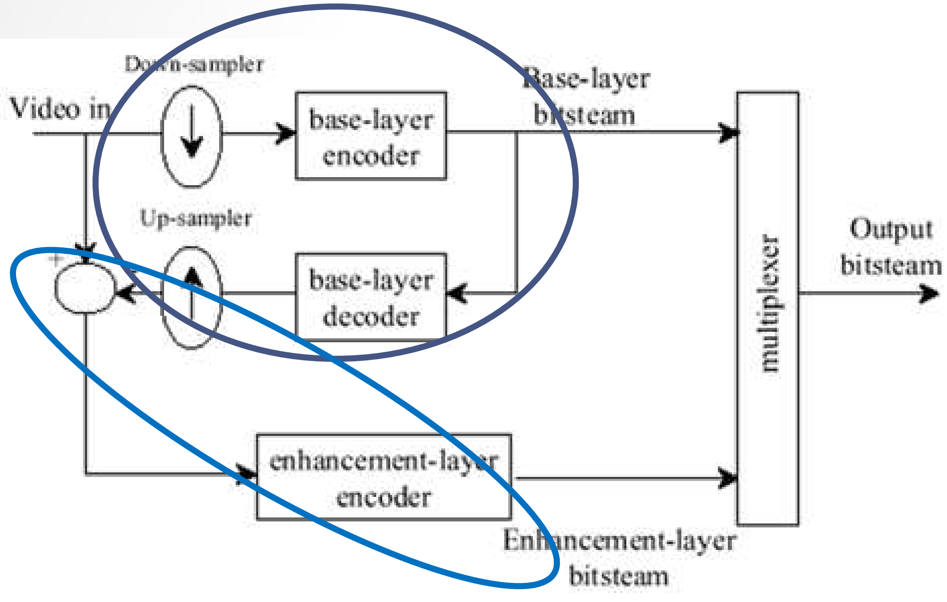
Computer monitor

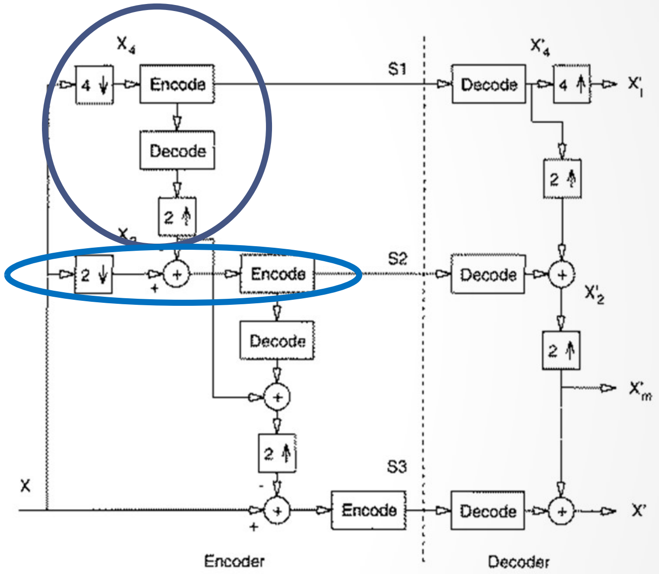
60” TV



Bitstream with M layers of spatial scalability (Source: Textbook 1, Fig. 11.5, p.354)

MPEG-2 Spatial Scalability

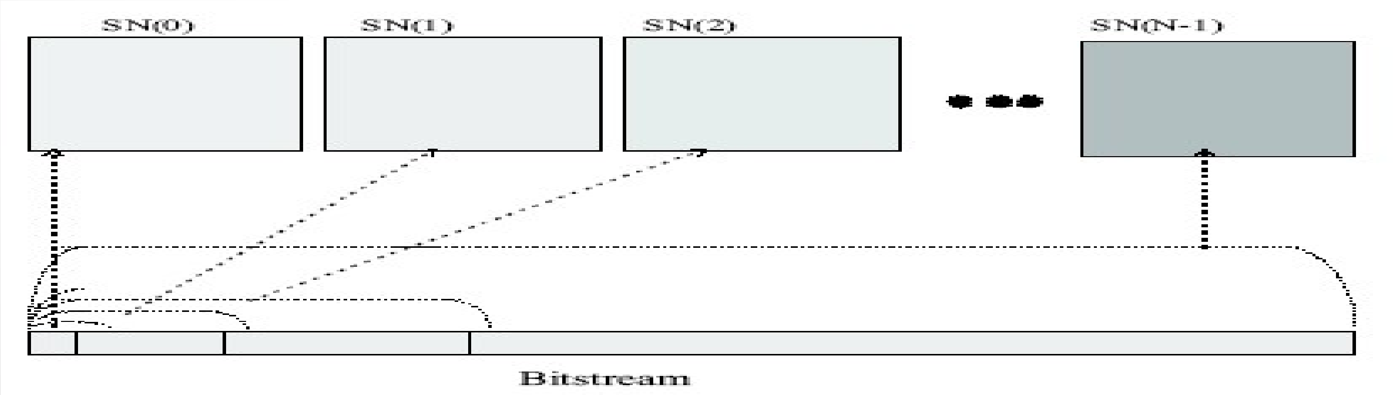




- Motion compensated video coder

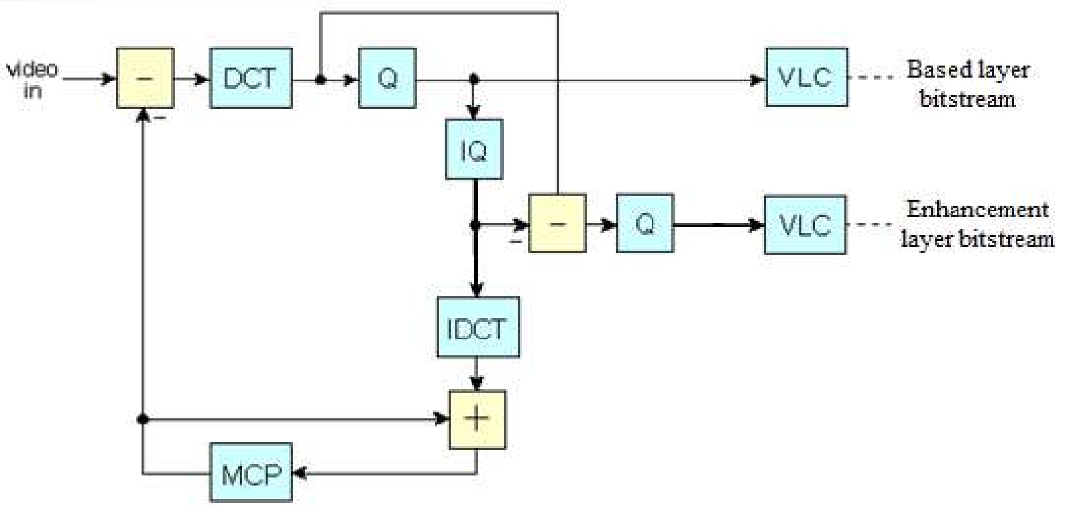
MPEG-2 SNR (Quality) Scalability

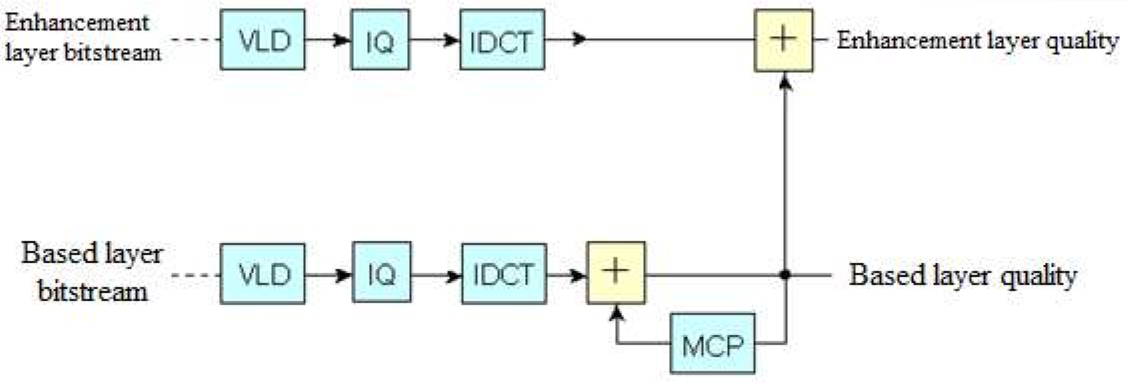
- Change of quality



Bitstream with N layers of quality scalability. (Source: Textbook 1, Fig. 11.2, p.352)

MPEG-2 SNR (quality) Scalability





- Motion compensated video coder

MPEG-2 Data Partitioning

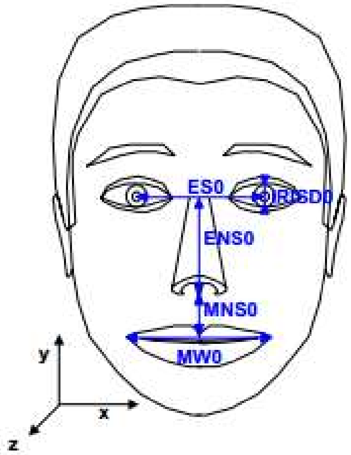
- Data partitioning splits the video bit stream into two or more layers, e.g., including different frequency components in each layer, with the base layer containing low frequency components, and other layers containing increasingly higher frequency components.
- Syntactic elements are placed into the high priority base layers and low priority enhancement layers.
- The decoder can decode the base layer only if the decoder implements a bitstream loss concealer for the enhancement layers.

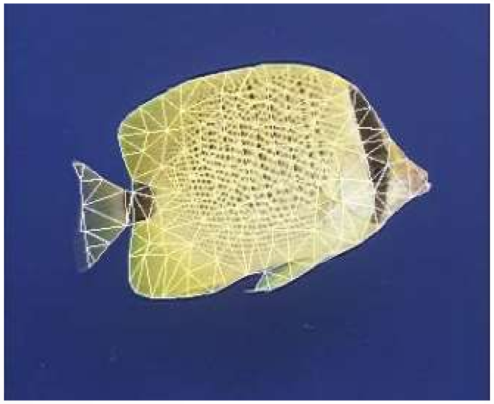
MPEG-4

- Originally targeted at very low bit-rate communication from 4.8kbps to 64 kbps, it now aims at 5kbps – 10Mbps.
- It emphasizes the concept of Visual Objects --> Video Object Plane (VOP)
  - objects can be of arbitrary shape, VOPs can be non-overlapped or overlapped
  - supports object-based interactivity
  - individual audio channels can be associated with objects
- Good for video composition, segmentation, and compression.
- Standards being developed for shape coding, motion coding, texture coding, etc.

MPEG-4 Video Main Features

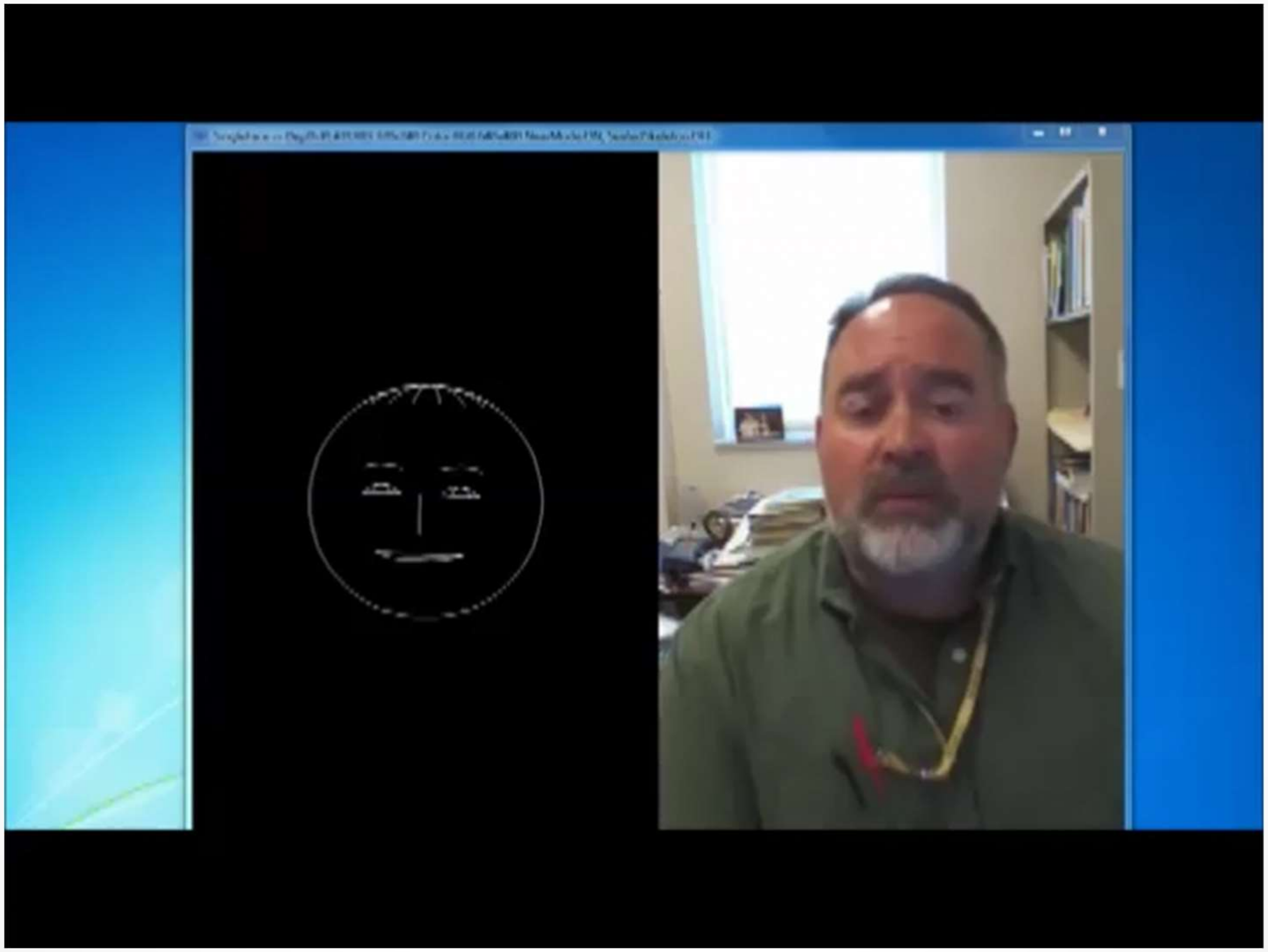
- Object based coding
  - Video object, face object, mesh object
  - Arbitrary shape, rectangular shape
  - Object based coding techniques
  - I-VOPs, P-VOPs, B-VOPs
  - Object based scalability
- Error resilience
  - Resynchronization, data recovery





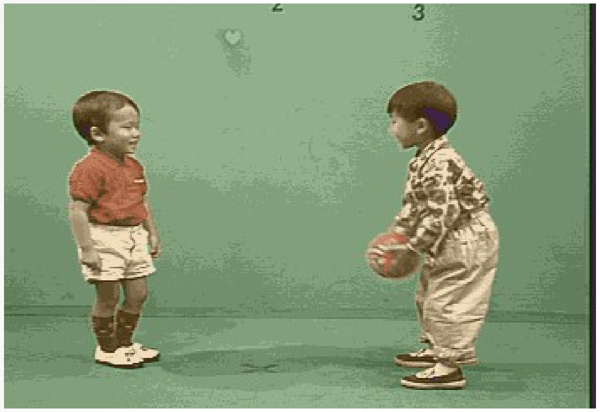
* Pictures from “Feature point based mesh deformation applied to MPEG-4 facial animation”, Sumedha Kshirsagar, Stephane Garchery, and Nadia Magnenat-Thalmann and from “Face and 2-D Mesh Animation in MPEG-4”, A. Murat Tekalp and Jörn Ostermann

Face object

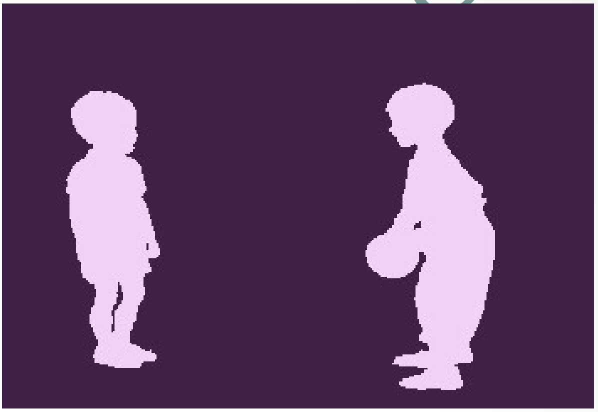


https://www.youtube.com/watch?v=ruSHueJhp3I

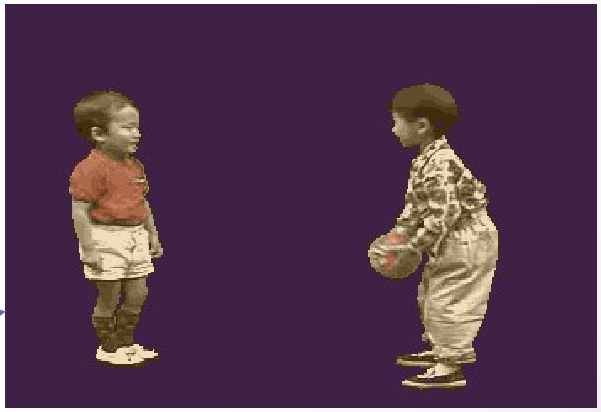
Object based Coding



Scene



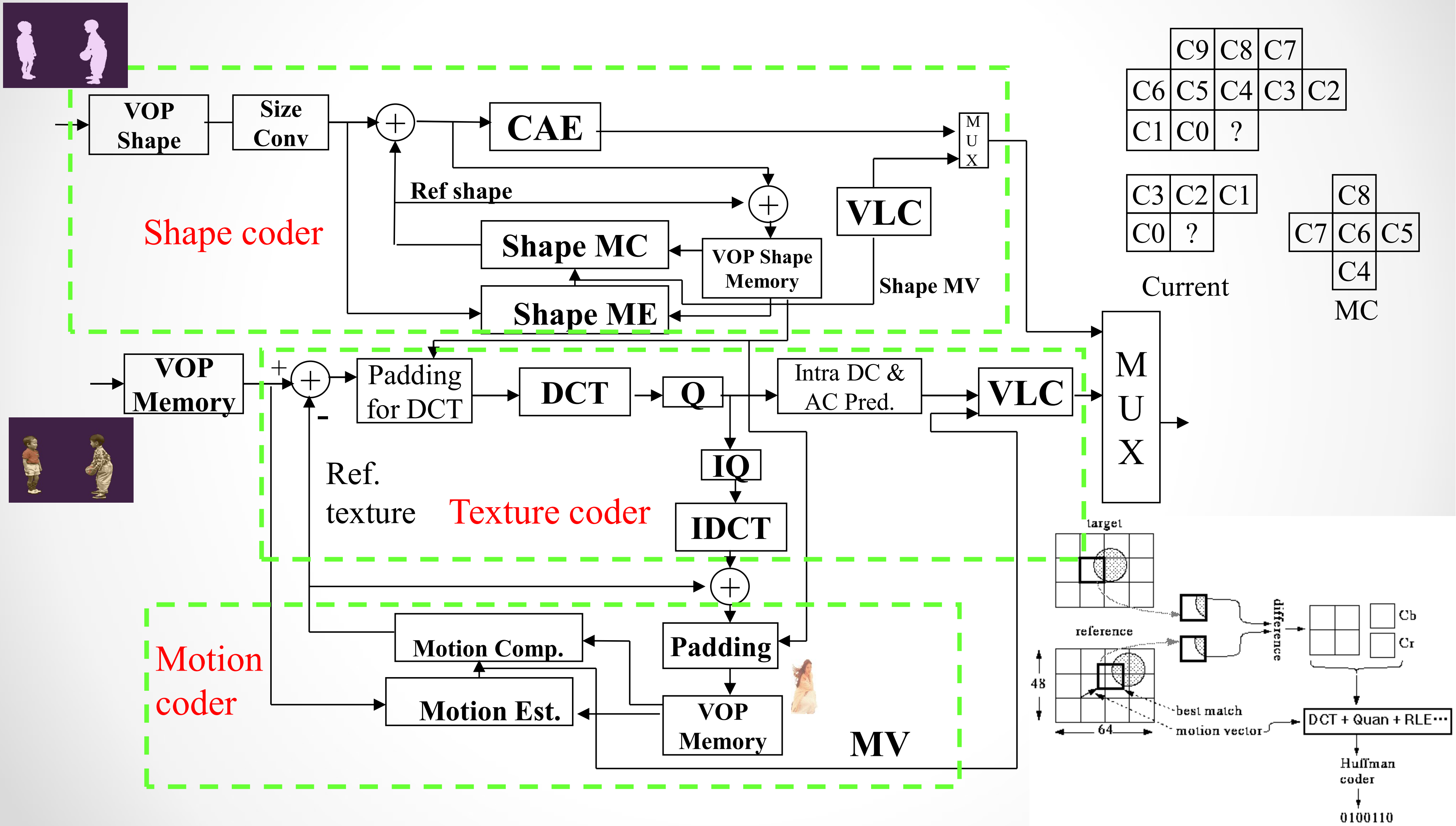
Shape



VOP

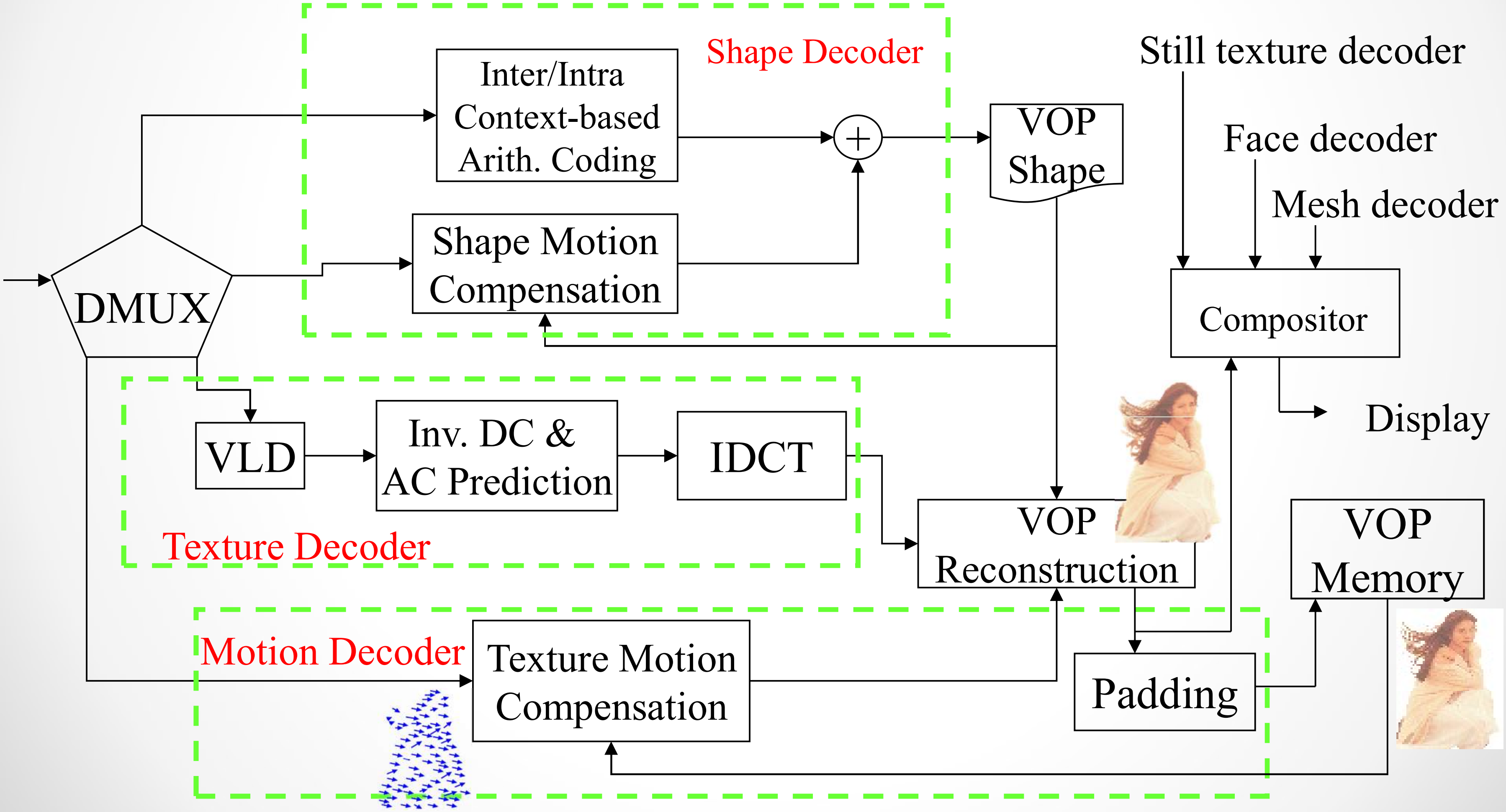
- Shape coding
- Motion coding
- Texture coding

MPEG-4 VOP Encoder (Simplified)



- Motion compensated video coder

MPEG-4 VOP Decoder (Simplified)



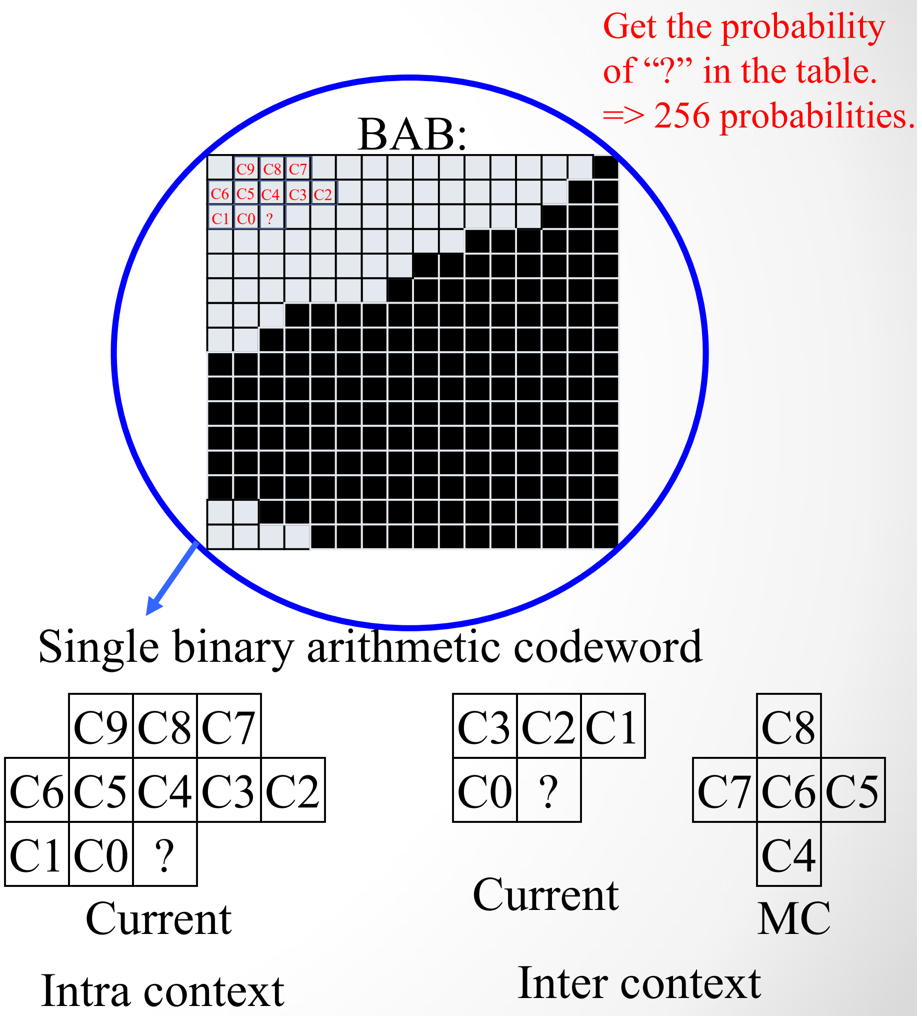
- Motion compensated video coder

Shape Coding

- The shape information is referred to as alpha planes. The techniques to be adopted by the standard will provide lossless coding of alpha planes and lossy coding of shapes and transparency information since the shape size scalability, thus allowing a tradeoffs between bit rate and accuracy of shape representation.
- Furthermore, intra and inter shape coding functionalities employing motion compensated shape prediction is envisioned so as to allow both efficient random access operations as well as efficient compression of shape and transparency information.

Shape Coding

- Binary alpha planes
  - (each pixel is either 0 or 255)
- Binary alpha blocks (BAB) (16 x 16)
- Motion estimation & compensation for shape
- Rate control through size conversion
- BAB coding (intra, inter) by context-based arithmetic encoding (CAE)
- Context-based means that we use C9-C0 as a context to determine the probability of “?”.
- When building contexts, any pixels outside the bounding rectangle of the current VOP to the left and above are assumed to be zero.



Get the probability of “?” in the table. => 256 probabilities.

CAE

- Encode 4 pixels and the value is 1111.
- CAE: assume the probability is 0.1 for the pixel “?” predicted to be “0”

0	0.1	1.0
0.1	0.19	1.0
0.19	0.271	1.0
0.271	0.3439	1.0

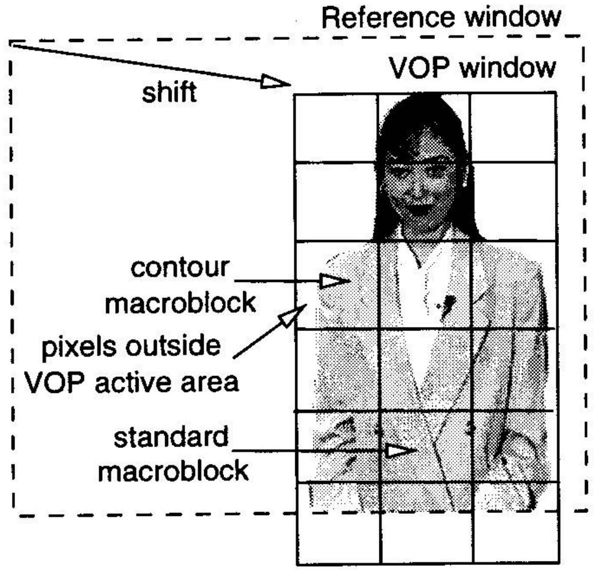
- How many bits to encode this range?
- Any number in this range [0.3439, 1.0) => 0.5
- $0.5_{(10)}$ is located in this range. => $0.1_{(2)}$ => 1 bit!

Motion Coding

- Temporal redundancies between video content in separate VOPs are exploited using block-based motion estimation and compensation.
- In general, these techniques can be viewed as extensions of the standard block-matching techniques in other MPEG standard to image sequences of arbitrary shape.

Motion Coding

- To perform block based motion estimation and compensation for VOPs, an arbitrary shape macroblock approach in the figure is used.
- A shift parameter is coded to indicate the location of VOP with respect to the boarders of the reference window.
- A VOP windows surrounds the foreground object.
- The VOP window contains an integer number of macroblocks horizontally and vertically. Size of each macroblock is 16x16 pixels. Macroblock can be either standard or contour/boundary macroblocks.



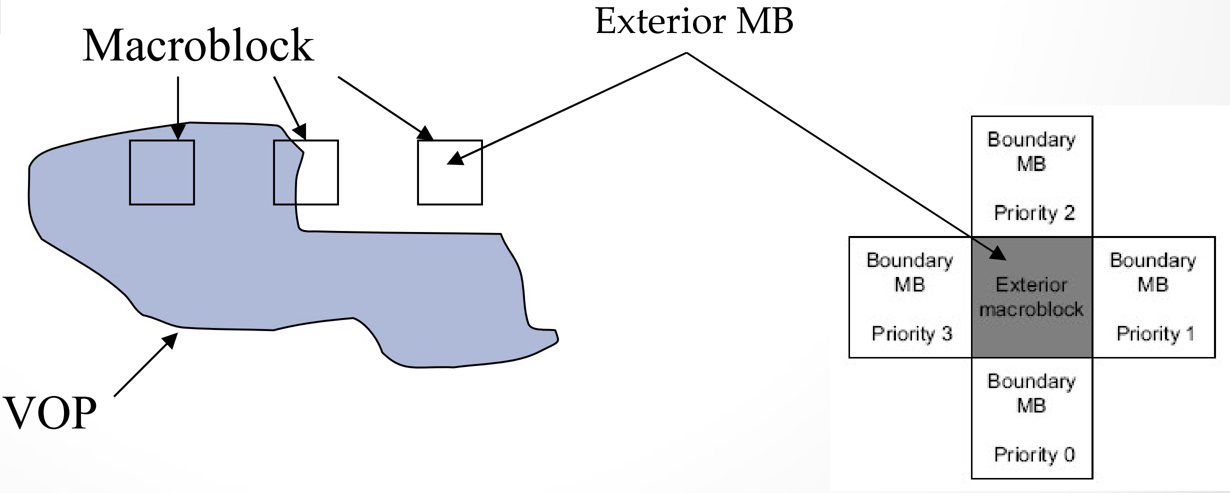
- Motion compensated video coder

Motion Coding

- For a standard macroblock, where all of its pixel are inside the active VOP area, any motion estimation and compensation techniques can be used.
- For contour/boundary macroblock, padding technique is used to fill the pixels outside the active VOP area, before motion estimation & compensation.
- The padding method extrapolates pixels outside the VOP based on pixels inside the VOP.
- After padding, the motion estimation & compensation process is the same as standard macroblock.

Motion Estimation & Compensation

- Similar to MPEG-2 + provisions for arbitrary shape objects
- Pixels outside VOP but inside bounding rectangle of that VOP - padding required for reference VOP
- Modes
    - Motion estimation & compensation
    - unrestricted
    - advanced



Texture Coding

- The intra VOPs as well as residual errors after motion compensated prediction are coded using DCT on 8x8 blocks, in a manner similar to that employed in MPEG-1&2, h.26x.
- For each macroblock, there could be four 8x8 luminance block and two 8x8 chrominance blocks. As in the motion estimation step, 8x8 blocks well within the VOP active area can be coded in a straightforward manner.
- For the coding of boundary macroblock, pixel outside the active area use padding technique.
- After computing the DCT, zig-zag scanning, quantization and run-length coding of DC and AC coefficients are used as in previous MPEG standards.