Digital Video Processing

Video Coding and Representation

Why Video Compression?

Bandwidth of S-VGA Monitors –(Medium res)

- **(8bits/pixel)*(1024pixels/line)*(768 lines/frame) * (72 frames/sec)**
  = 453 Megabits/sec

$$MSE = \frac{1}{MN}\sum_{i=1}^{M}\sum_{j=1}^{N}\left[I(i,j)-\hat{I}(i,j)\right]^2$$

$$PSNR = 20\log_{10}\left(\frac{2^n}{MSE^{1/2}}\right)$$

Network capability

- **Telephone – 56 kbps**
- **Ethernet - 10-100 Mbps**
- **10G Ethernet – 10 Gbps**

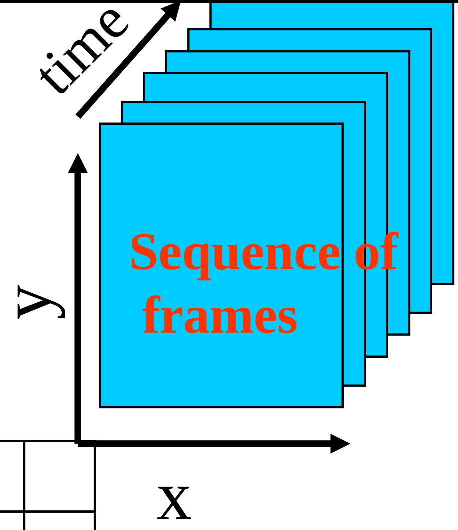
Video Compression standards

- **H.261 / H.263 - Video conferencing**
- **MPEG-1 (1991)- 1.5 Mbps (for CD-ROM and Hard Disk)**
- **MPEG-2 (1994)– 4-9 Mbps (for Digital Television set top boxes and DVDs**
- **MPEG-4 / H.264 (1999)– 4.8 – 6.5 kbps(Object based video coding)**
- **MPEG-7 - Content based description**

These standards do not prescribe an algorithm; Instead syntaxes for Data streams are defined

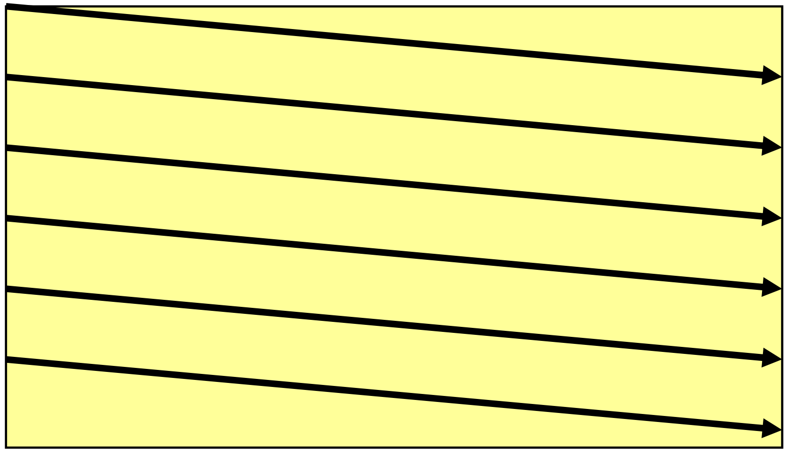
Video representation

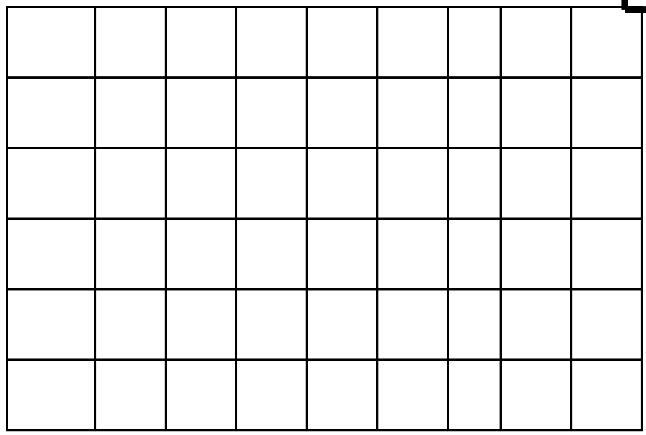
Analog video —sampled→ Digital Video



- **Progressive**

Frame based





Time
k=k0

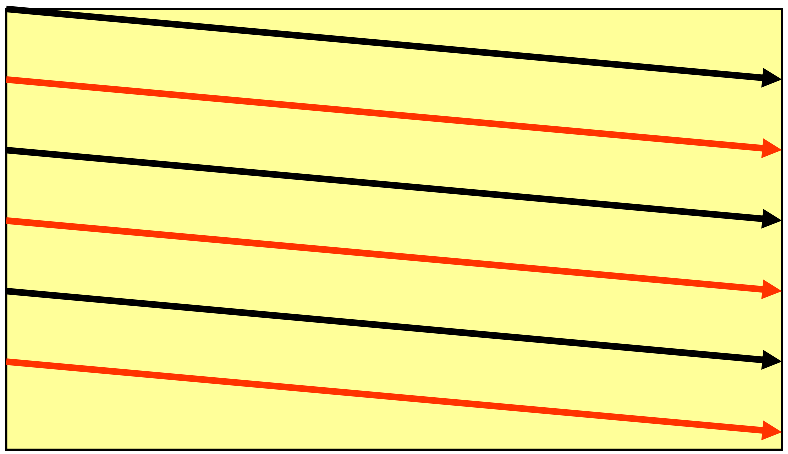
MPEG 1

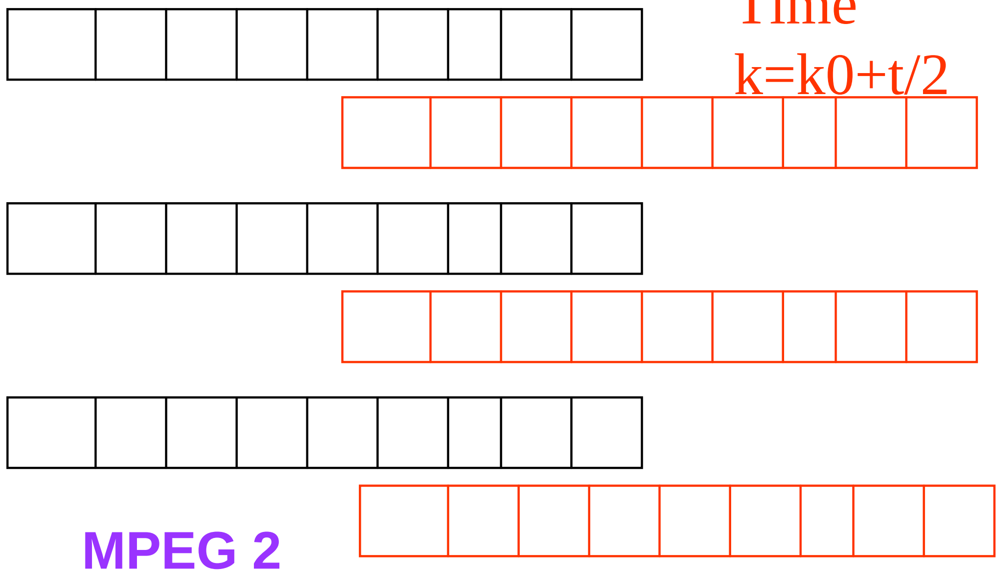
- **Interlaced analog video**

Time k=k0

Field based

Time
k=k0+t/2





MPEG 2

Color Representation

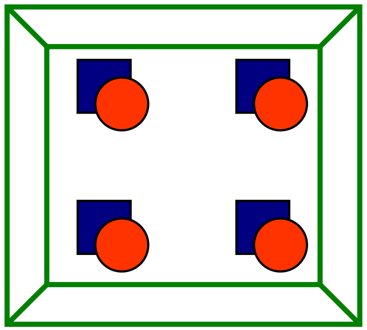
- **Y-Cr-Cb representation** $\longrightarrow$ **Reduced Bandwidth**

Luma Chroma subsampling

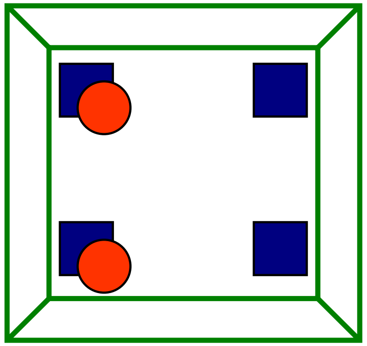
$$Y = k_r R + (1 - k_b - k_r)G + k_b B$$

$$Cb = \frac{0.5}{1-k_b} B - Y$$

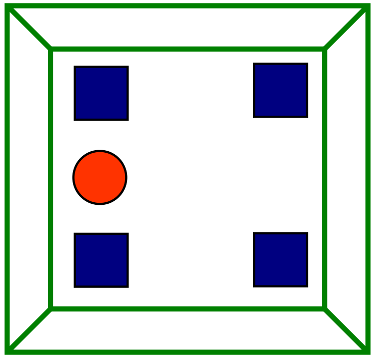
$$Cr = \frac{0.5}{1-k_r} R - Y$$



4:4:4 sampling

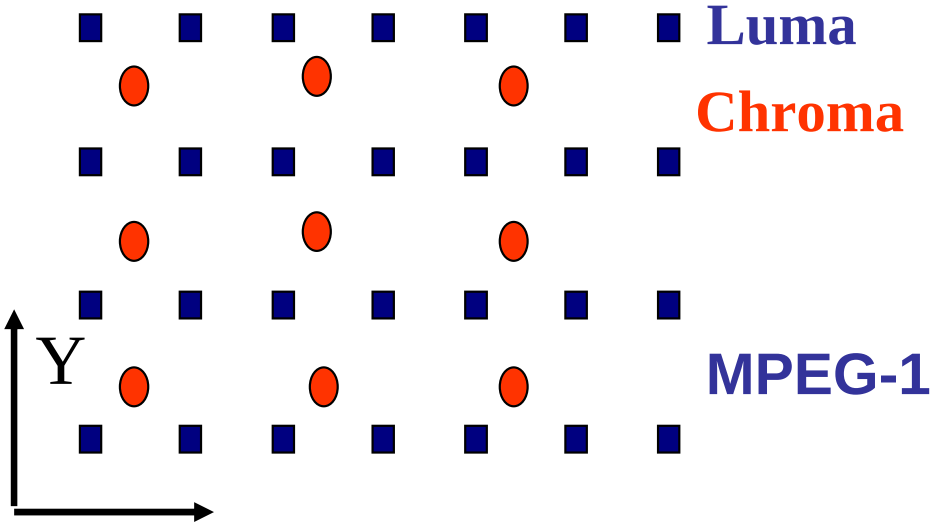


4:2:2 sampling



4:2:0 sampling

Find:
(R, G, B) = $F_{CT}(Y, C_b, C_r)$



4:2:0 sampling

For **4:2:2 sampling (YUV2),** the Chroma components have the *same vertical resolution* as Luma, but *half the horizontal resolution*; used for high quality video.

For **4:2:0 sampling (YV*12*)** both chroma components have the *half the resolution in both directions* (X and Y).

For image resolution - 640*480 (VGA):

4:4:4 (Y, Cb, Cr) Res.: 640*480*8*3 = 900 KB.

4:2:0 (Y, Cb, Cr) Res.: 640*480*8 + 320*240*8*2 = 450 KB ; half as 4:4:4.

4:2:0 sampling (YV*12*) – 12 bits per pixel; WHY ??

In case of 4:4:4 sampling, 12 samples require **96 (=12*8) bits**.
Average is **96/4 = 24 bits per pixel.**

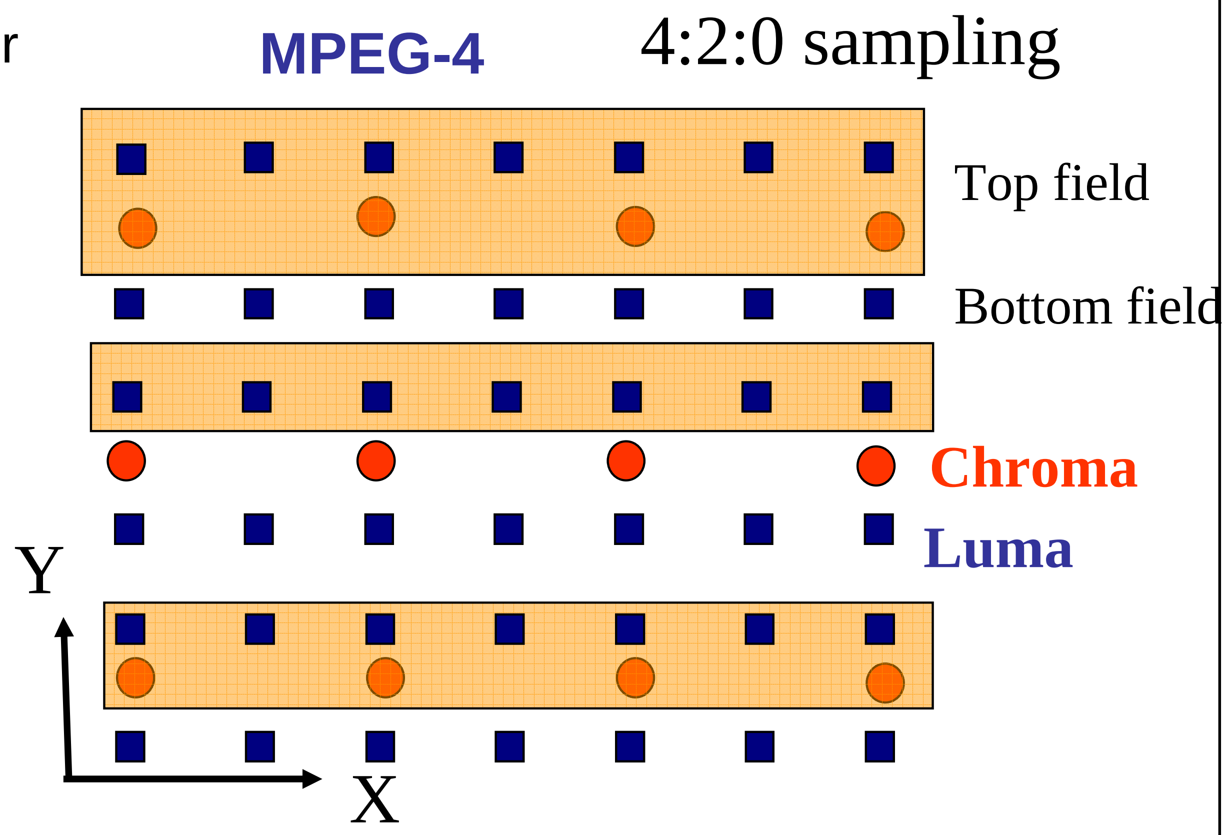
In case of 4:2:0 sampling, 6 samples are required: 4 for Y and one each for Cb and Cr.

Thus we need **6 * 8 = 48 bits**;

Average is **48/4 = 12 bits per pixel**.

$$PSNR_{dB} = 20\log_{10}\left(\frac{2^n - 1}{MSE^{1/2}}\right)$$

N = No. of bits/image sample



Video Format	Y Size	Color Sampling	Frame Rate (Hz)	Raw Data Rate (Mbps)	**Bits/ Frame**
HDTV Over air. cable, satellite, MPEG2 video, 20-45 Mbps					
SMPTE296M	1280x720	4:2:0	24P/30P/60P	265/332/664	
SMPTE295M	1920x1080	4:2:0	24P/30P/60I	597/746/746	
Video production, MPEG2, 15-50 Mbps					
BT.601	720x480/576	4:4:4	60I/50I	249	
BT.601	720x480/576	4:2:2	60I/50I	166	
High quality video distribution (DVD, SDTV), MPEG2, 4-10 Mbps					
BT.601	720x480/576	4:2:0	60I/50I	124	
Intermediate quality video distribution (VCD, WWW), MPEG1, 1.5 Mbps					
SIF	352x240/288	4:2:0	30P/25P	30	
Video conferencing over ISDN/Internet, H.261/H.263, 128-384 Kbps					
CIF	352x288	4:2:0	30P	37	
Video telephony over wired/wireless modem, H.263, 20-64 Kbps					
QCIF	176x144	4:2:0	30P	9.1	
4-CIF	**704x576**	**4:2:0**	**30P**	**138**	**46.6 MB**

CIF – Common Intermediate Format; **SIF – Source Input Format;**

ITU-R Recommendation BT.601; or Rec. 601 (or CCIR 601,) is a standard published by International Telecommunication Union - Radiocommunications sector (formerly CCIR); SMPTE - Society of Motion Picture and Television Engineers (USA).

Video (MPEG2)

- 16 kbit/s – videophone quality (minimum necessary for a consumer-acceptable "talking head" picture)
- 128 – 384 kbit/s – business-oriented videoconferencing system quality
- 1.25 Mbit/s – VCD quality
- 5 Mbit/s – DVD quality
- 15 Mbit/s – HDTV quality
- 36 Mbit/s – HD DVD quality
- 54 Mbit/s – Blu-ray Disc quality

Format	Video Resolution
SQCIF	128 × 96
QCIF	176 × 144
CIF	352 × 288
4CIF	704 × 576
16CIF	1408 × 1152

Standard	Name	Bitrates	Example Video Formats
SMPTE 259M	SD-SDI	270 Mbit/s, 360 Mbit/s, 143 Mbit/s, and 177 Mbit/s	480i, 576i
SMPTE 344M		540 Mbit/s	480p, 576p
SMPTE 292M	HD-SDI	1.485 Gbit/s, and 1.485/1.001 Gbit/s	720p, 1080i
SMPTE 372M	Dual Link HD-SDI	2.970 Gbit/s, and 2.970/1.001 Gbit/s	1080p
SMPTE 424M	3G-SDI	2.970 Gbit/s, and 2.970/1.001 Gbit/s	1080p

Main Points in Video Compression

- **Predict a new frame from a previous frame and only code the prediction error --- Inter prediction**

- **Predict a current block from previously coded blocks in the same frame --- Intra prediction (introduced in the latest standard H.264)**

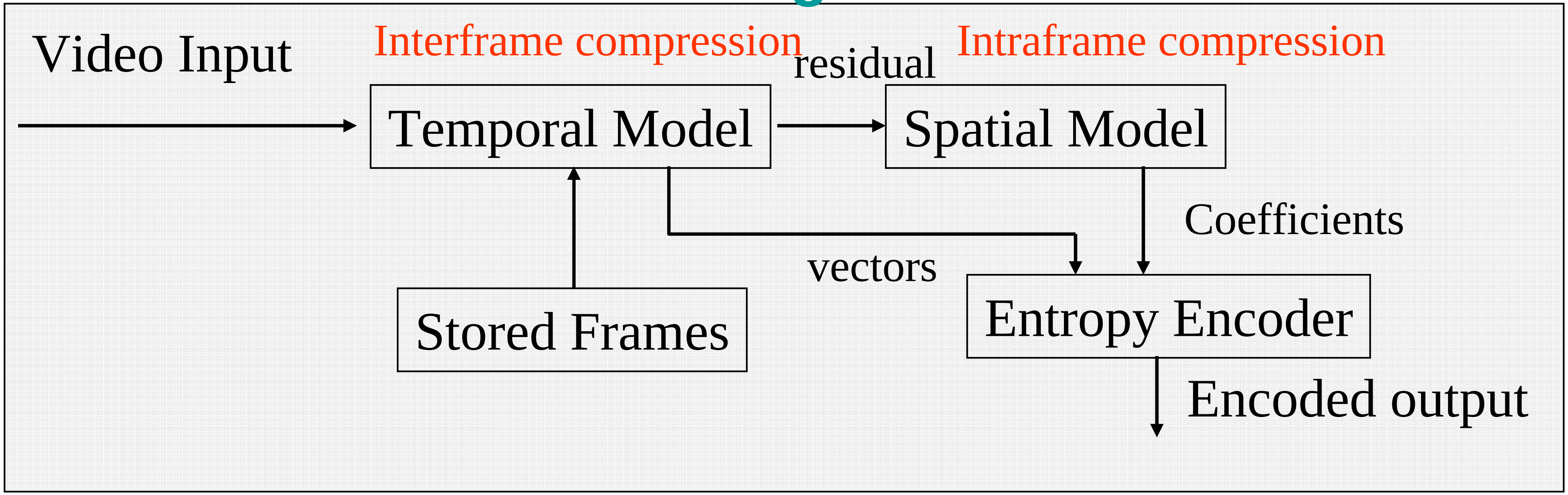
- **Prediction error will be coded using the DCT method**

- **Prediction errors have smaller energy than the original pixel values and can be coded with fewer bits**

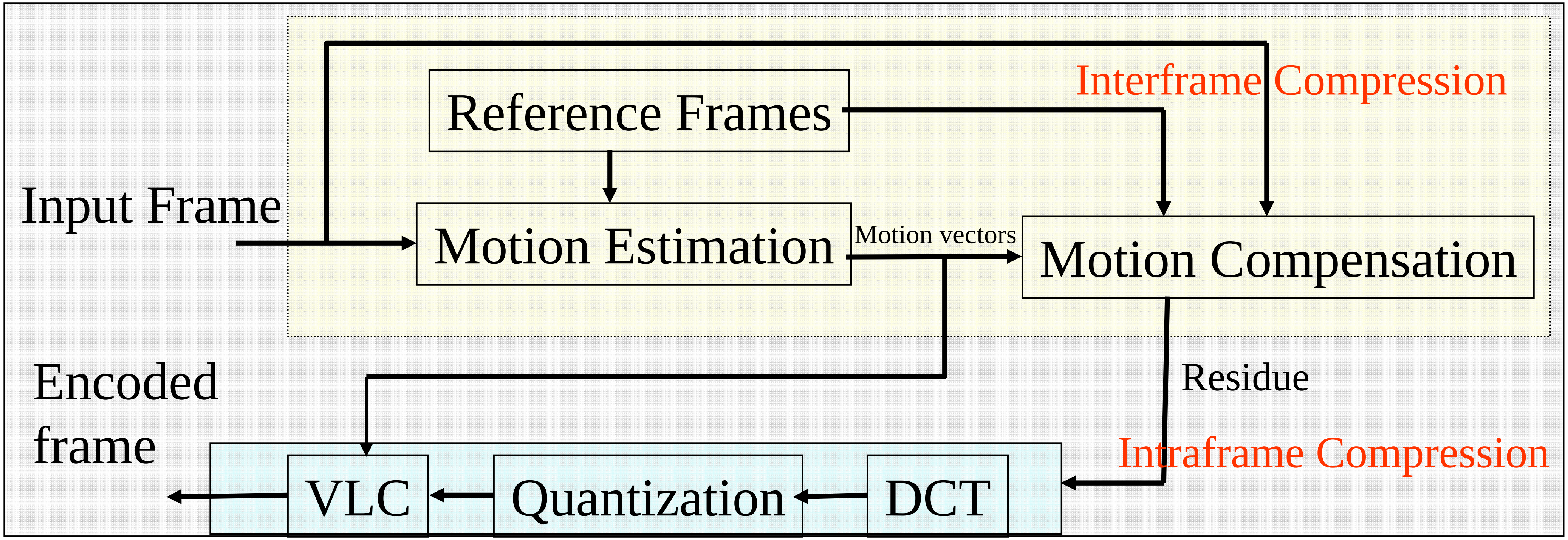
- **Those regions that cannot be predicted well will be coded directly using DCT --- Intra coding without intra-prediction**

- **Work on each macroblock (MB) (16x16 pixels) independently for reduced complexity**
  - **Motion compensation done at the MB level**
  - **DCT coding of error at the block level (8x8 pixels)**

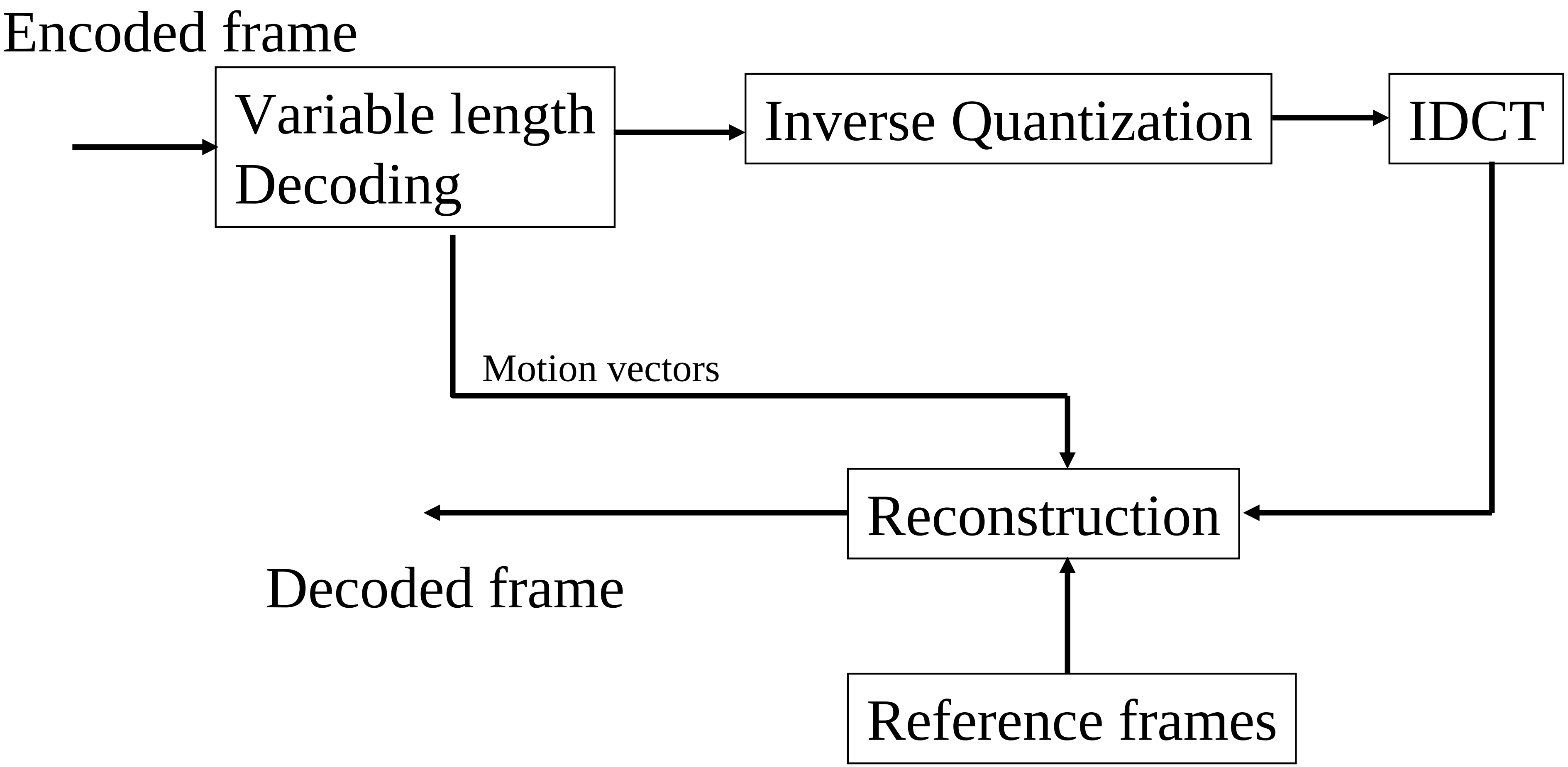
Video Encoder block diagram

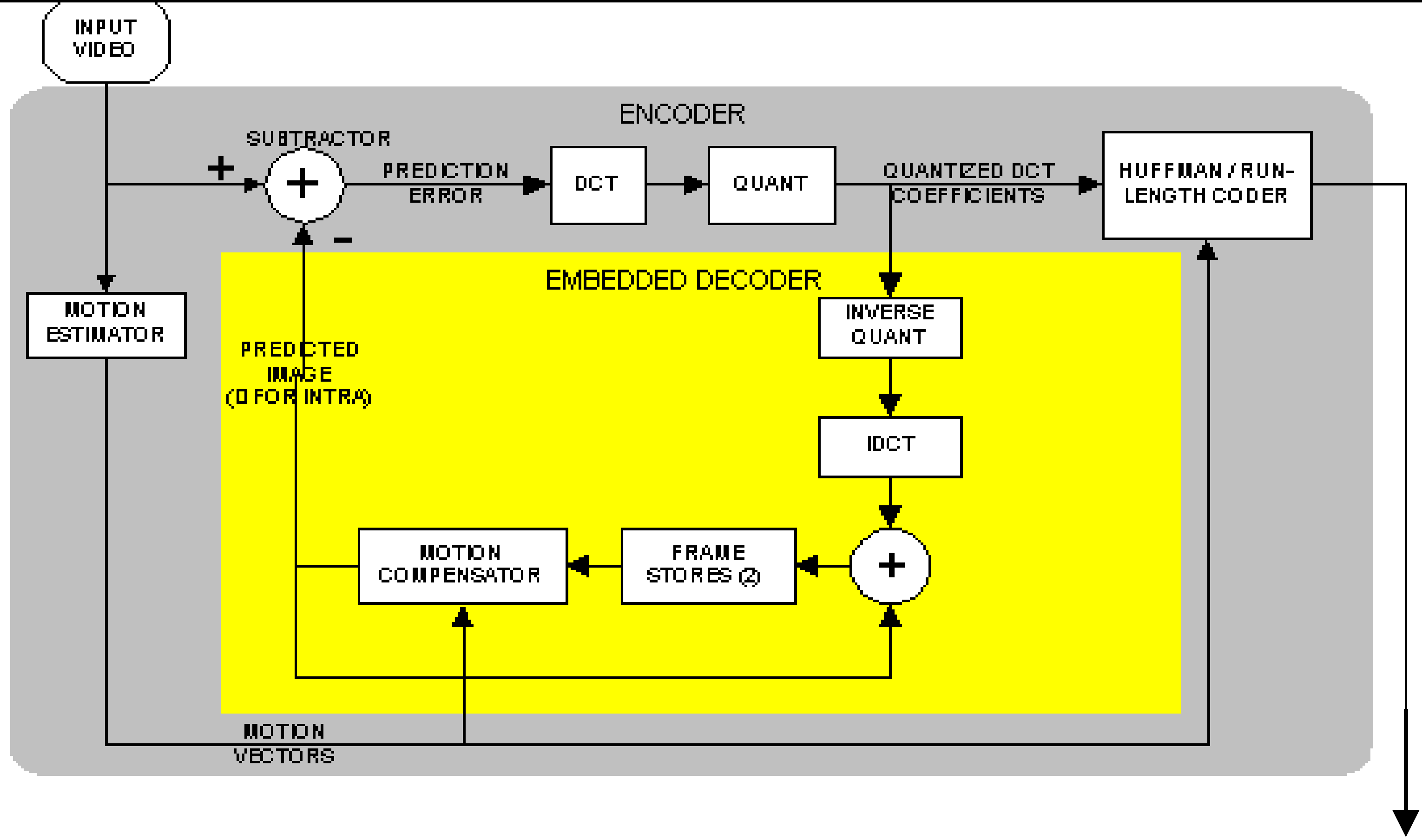


MPEG-2 Coder



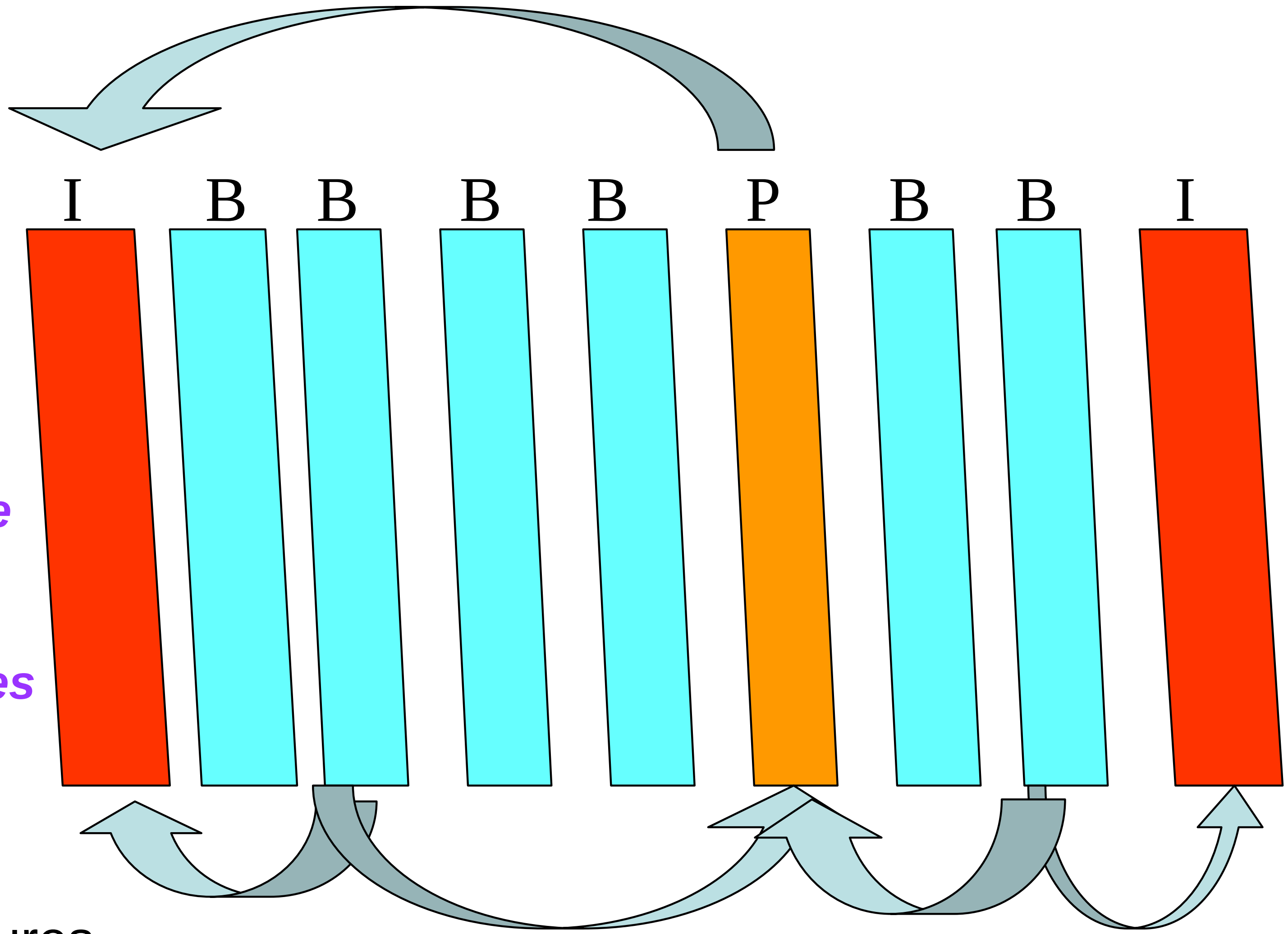
MPEG-2 Decoder





INPUT VIDEO
ENCODER
SUBTRACTOR
+
−
PREDICTION ERROR
DCT
QUANT
QUANTIZED DCT COEFFICIENTS
HUFFMAN / RUN-LENGTH CODER
MOTION ESTIMATOR
EMBEDDED DECODER
INVERSE QUANT
IDCT
PREDICTED IMAGE (0 FOR INTRA)
MOTION COMPENSATOR
FRAME STORES (2)
+
MOTION VECTORS

MPEG Group of Pictures

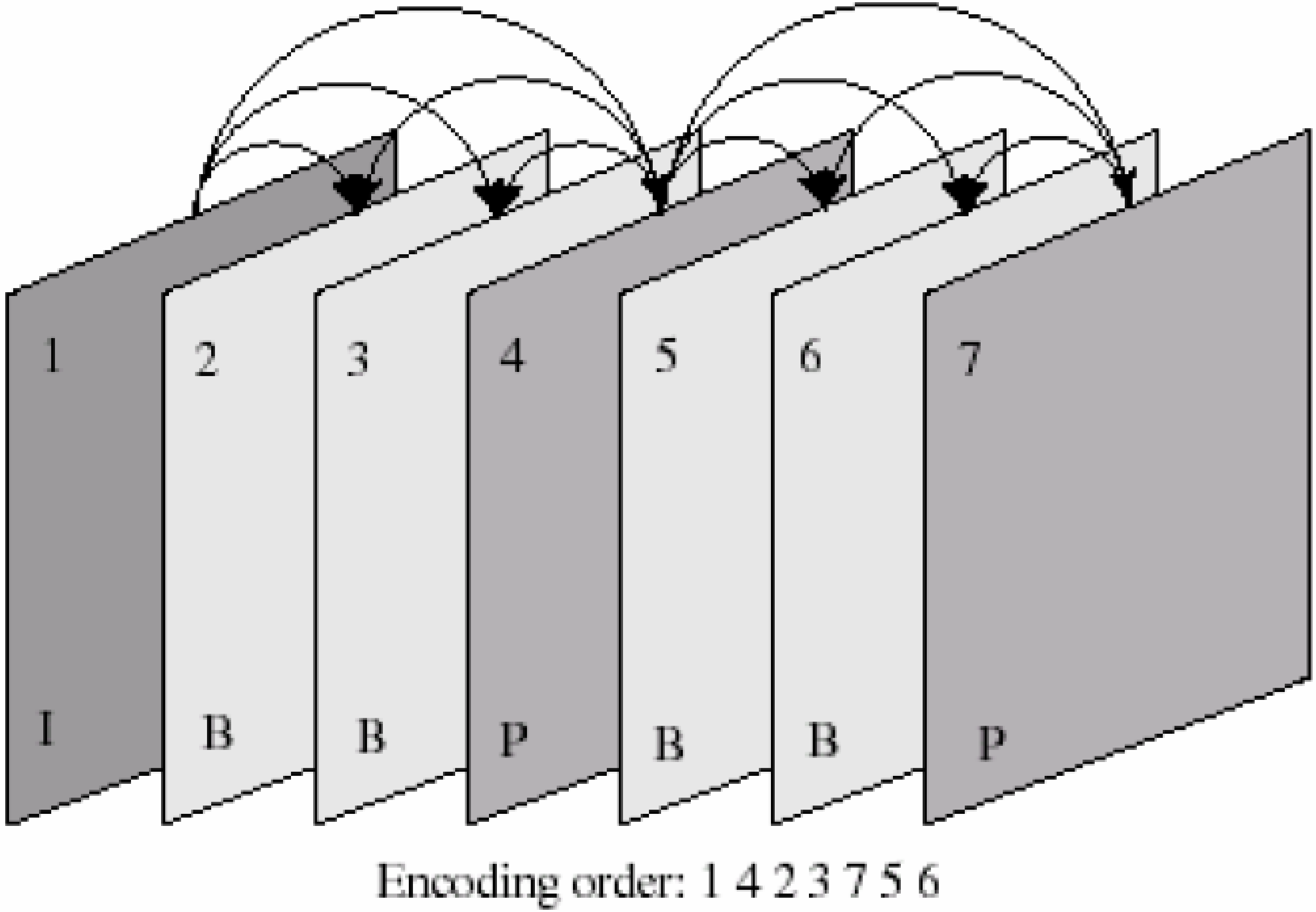


***I**ntra coded Picture*

***P**redictive coded picture*

***B**idirectionally predictive coded pictures*

- Advantage of B-pictures
  - Object occlusion problem avoided
  - Noise reduction
- Disadvantage
  - Requires previous and future frames


1
2
3
4
5
6
7
I
B
B
P
B
B
P
Encoding order: 1 4 2 3 7 5 6

Temporal Model in MPEG

- Changes due to motion – Optical flow

- Better prediction by compensating for motion between two frames

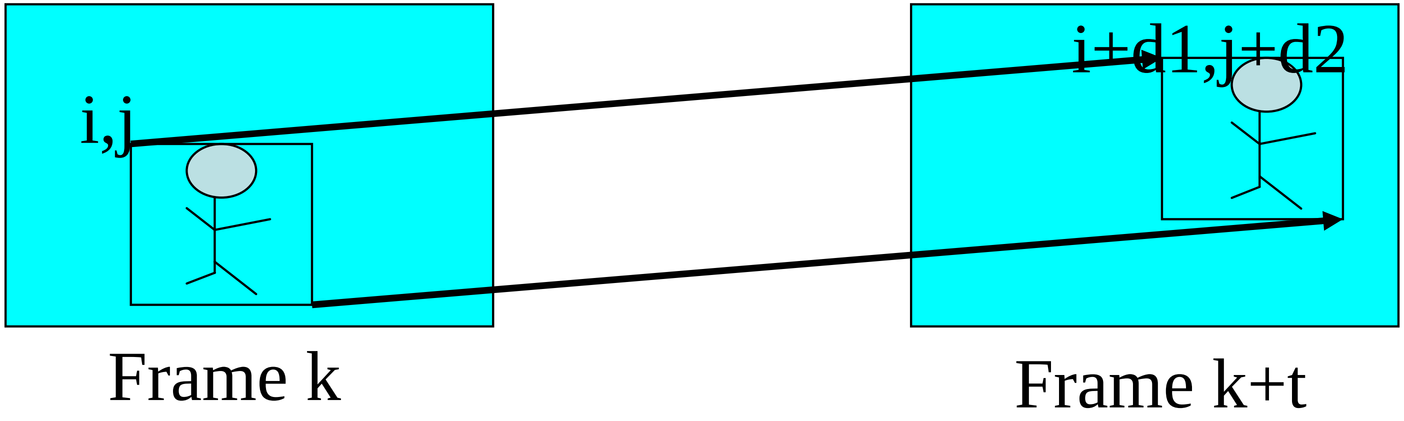
Block based Motion Estimation and Compensation

Macroblock – 16*16 pixel region

- basic unit for motion compensated prediction

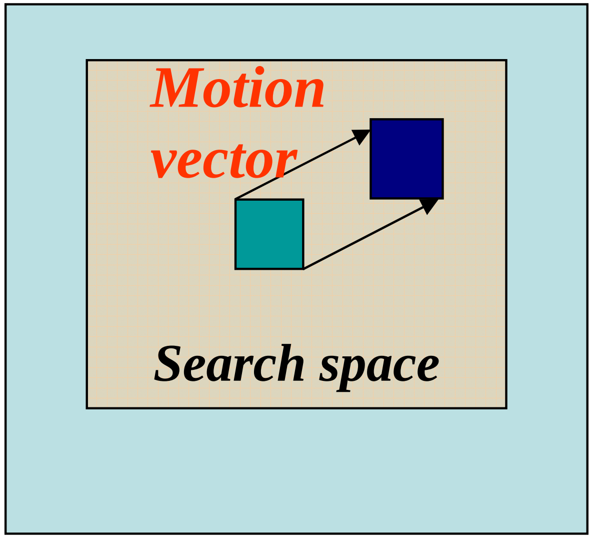
Block motion estimation

- Assumes video consists of moving blocks
- Assumes simple translational motion

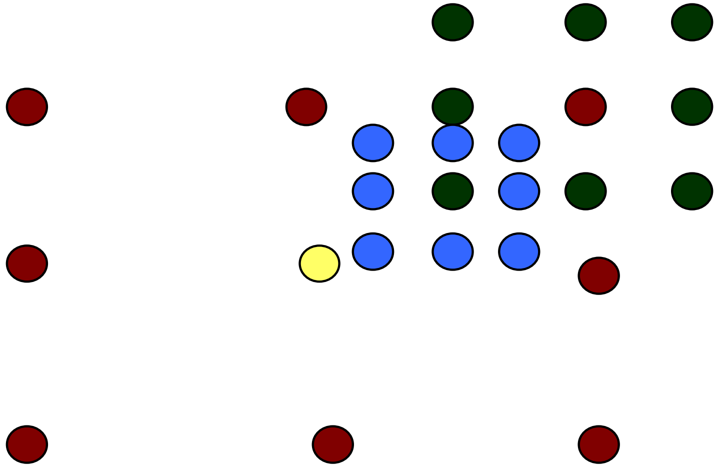


Only one motion vector needed per block

Exhaustive search



Three Step Search



Bi-directional Motion Compensation

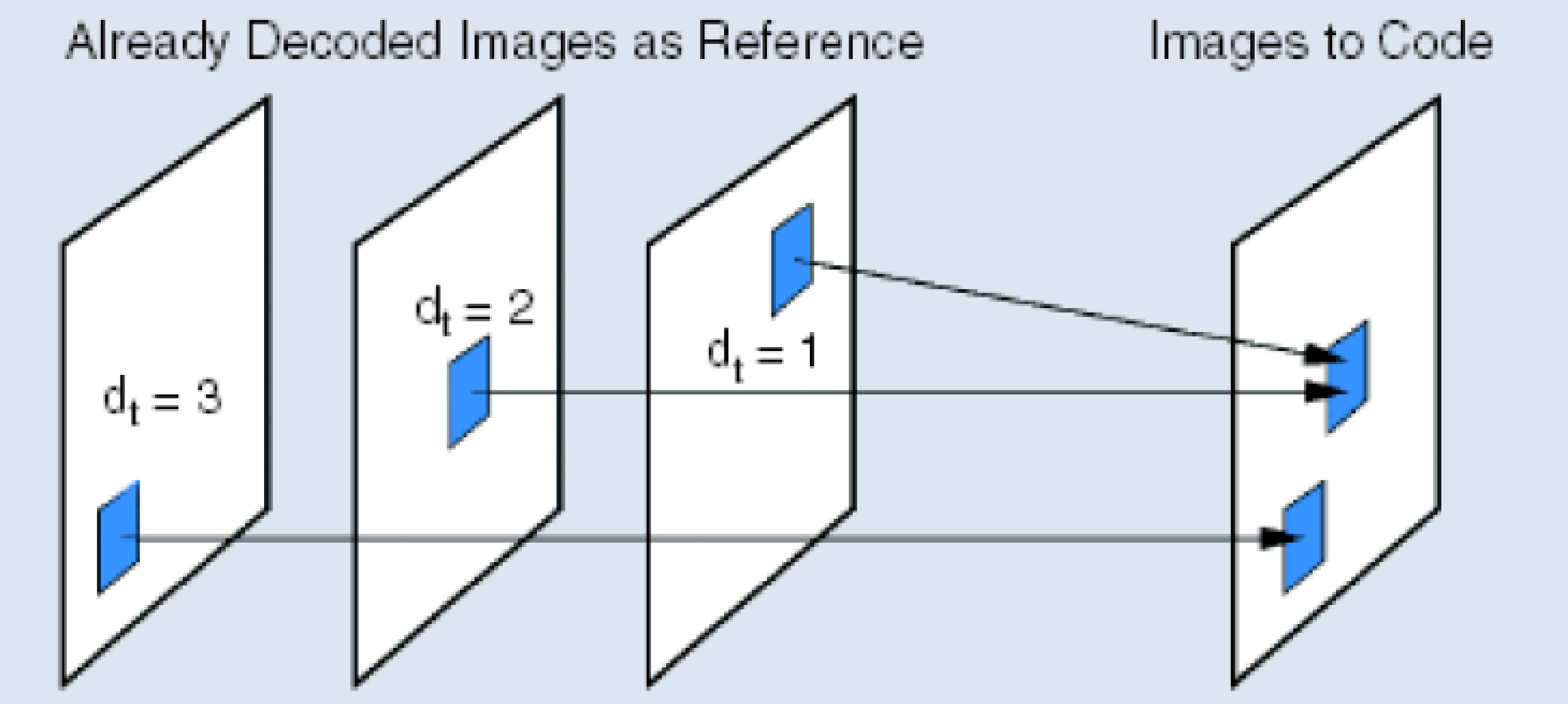
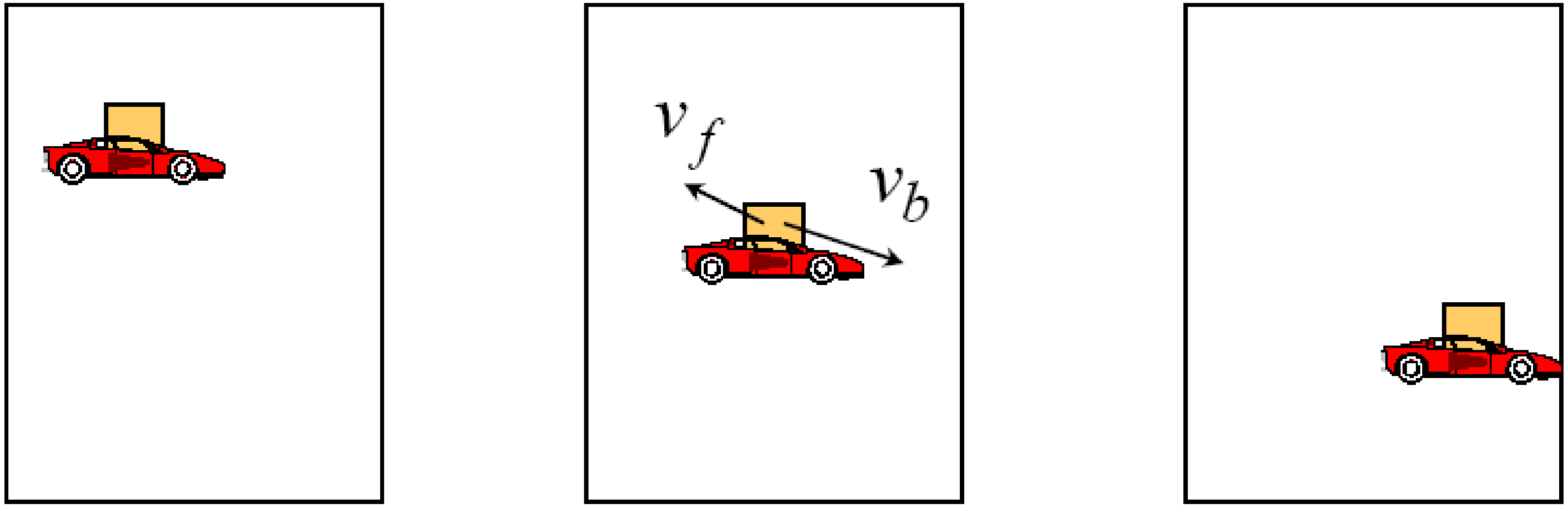


Figure 9. Motion-compensated prediction with multiple reference images. In addition to the motion vector, also an image reference parameter d_t is transmitted.

$$\hat{f}(t,m,n) = w_b f(t-1, m-d_{b,x}, n-d_{b,y}) + w_f f(t+1, m-d_{f,x}, n-d_{f,y})$$

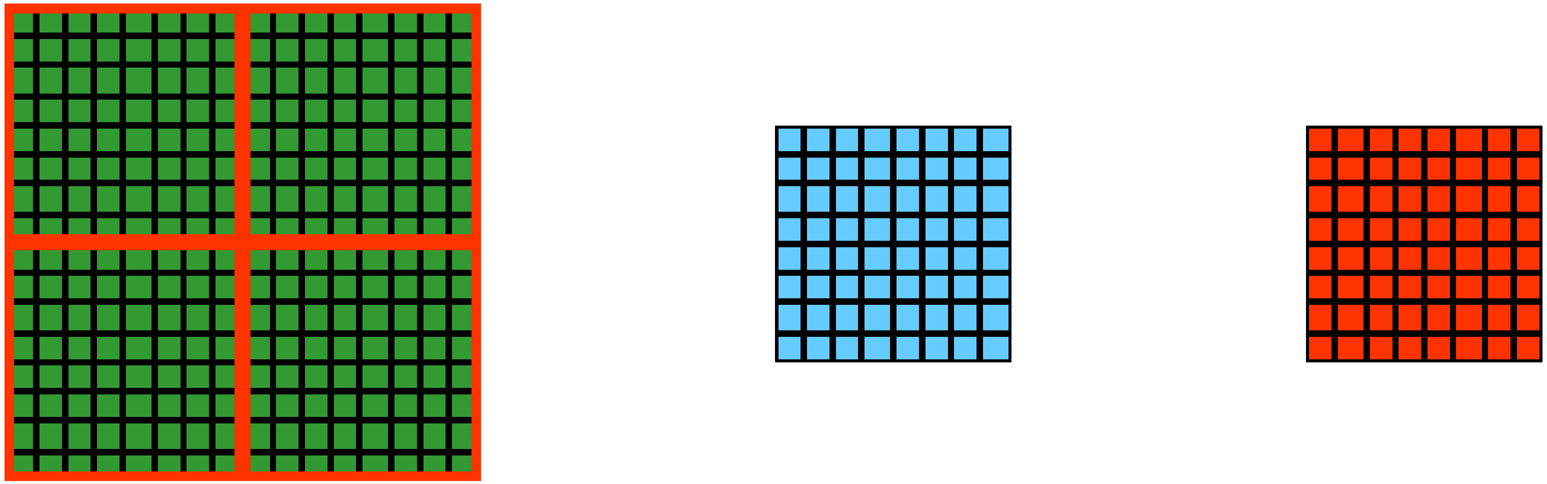


- **Conventional block motion compensation**
  - **One best matching block is found from a reference frame**
  - **The current block is replaced by the best matching block**

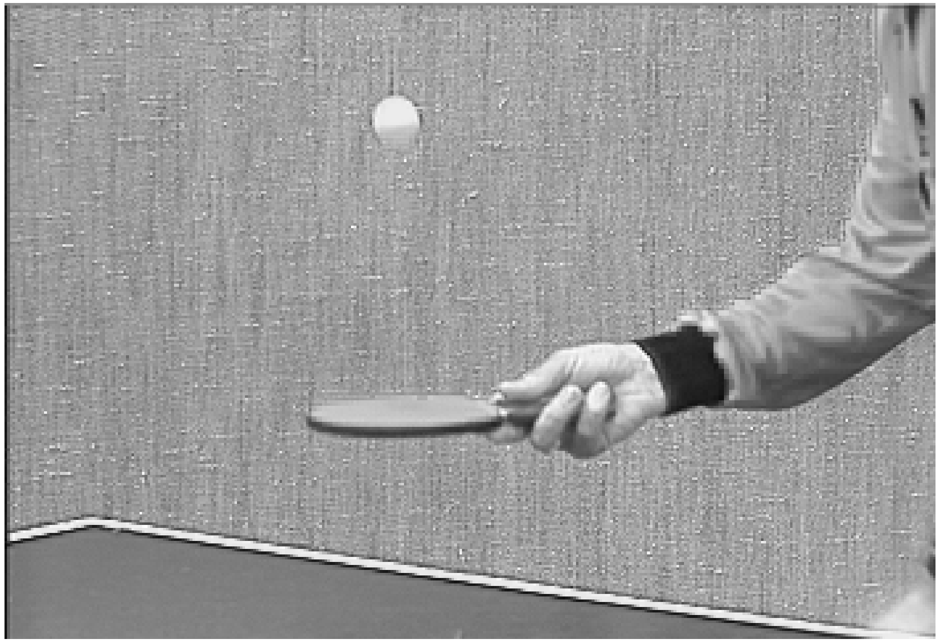
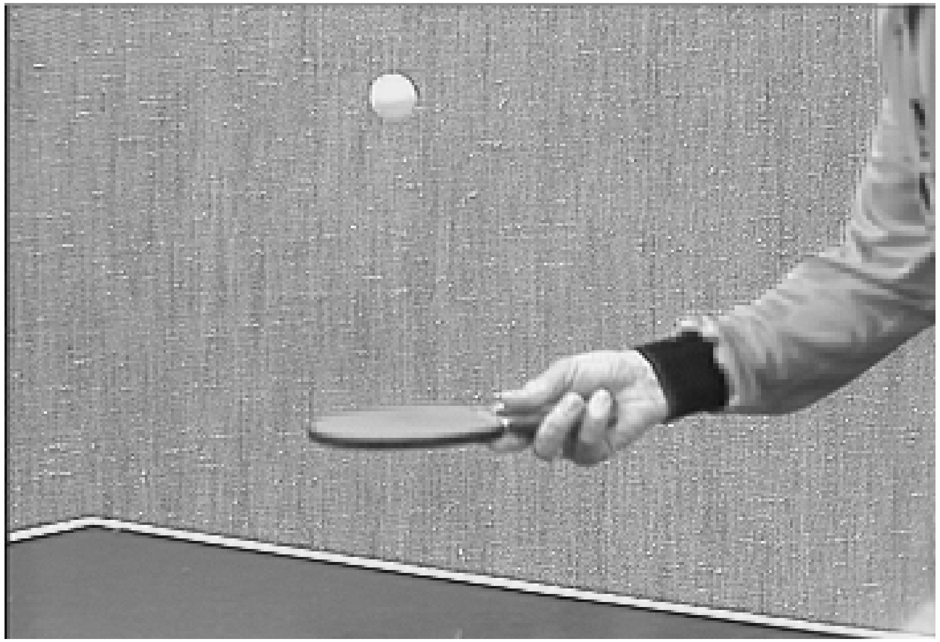
- **OBMC** - Overlapped Block Motion Compensation

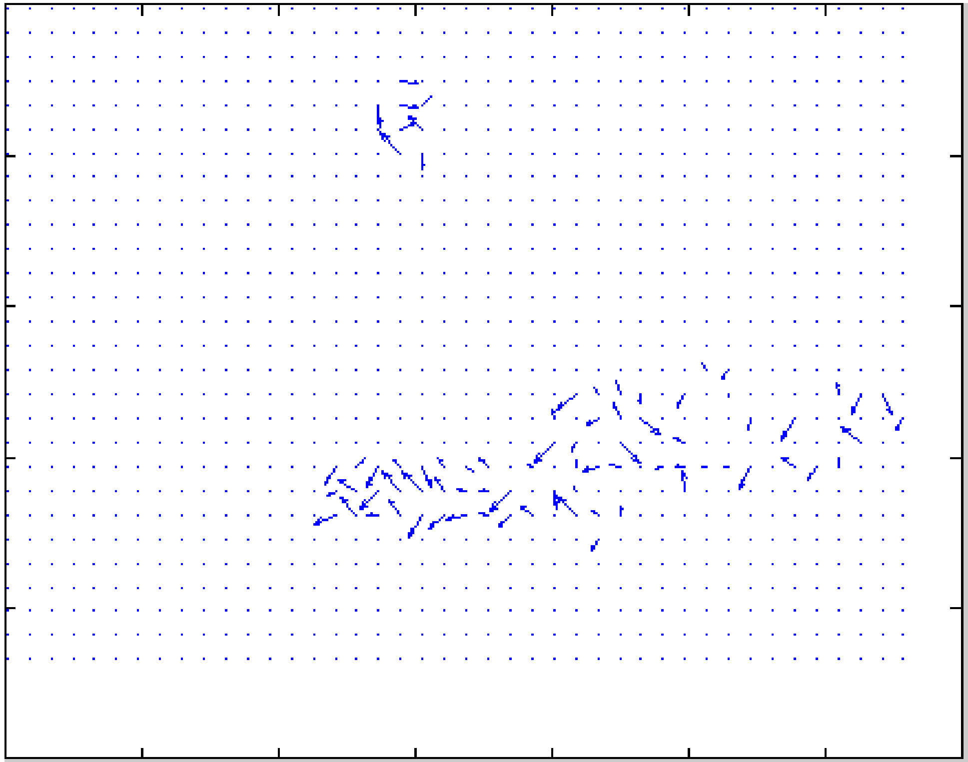
  - **Each pixel in the current block is predicted by a weighted average of several corresponding pixels in the reference frame**
  - **The corresponding pixels are determined by the MVs of the current as well as adjacent MBs**
  - **The weights for each corresponding pixel depends on the expected accuracy of the associated MV**

MB Structure in 4:2:0 Color Format



4 8x8 Y blocks   1 8x8 Cb blocks   1 8x8 Cr blocks

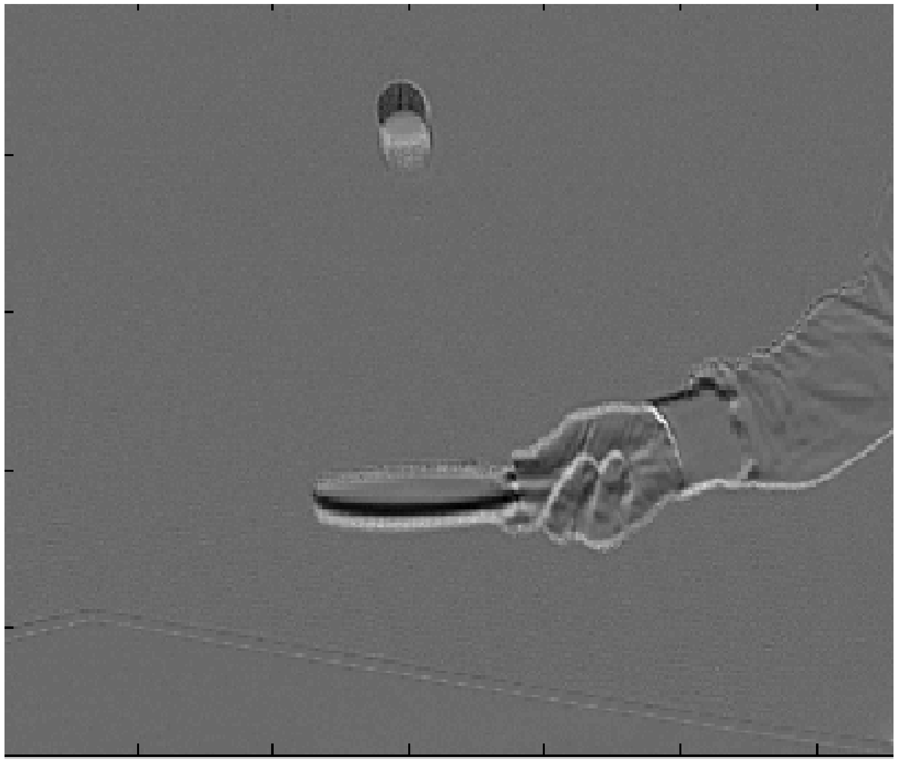
 



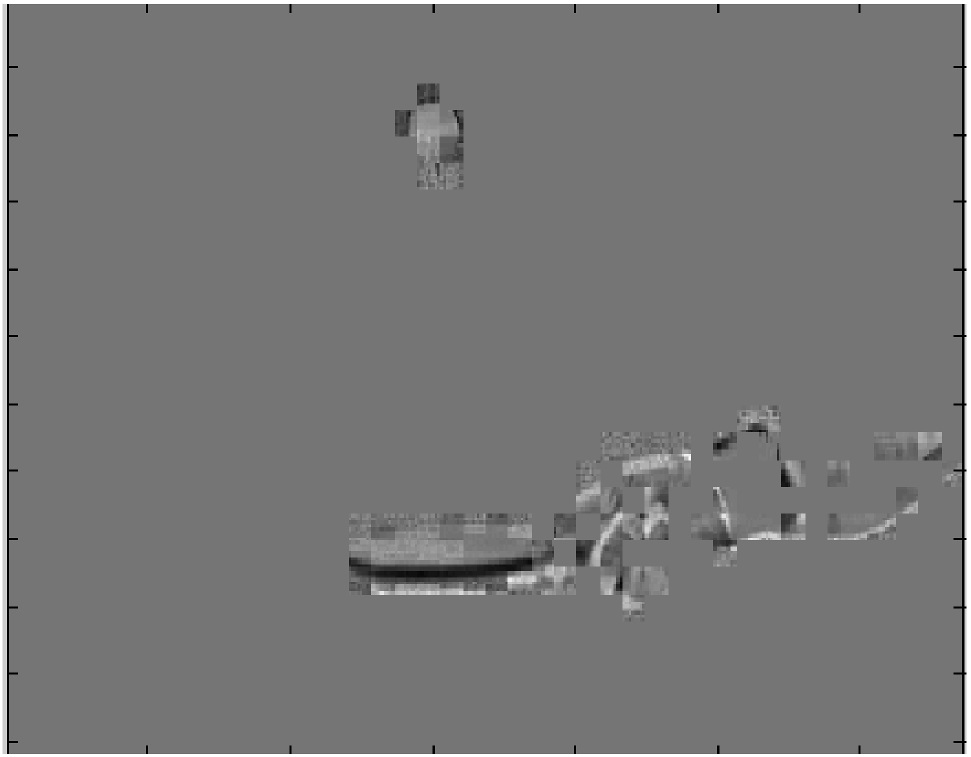
Motion vector

Motion Compensation

- Best matching region subtracted from current macroblock to produce residual macroblock
- Motion vector and residual are encoded and transmitted



No compensation



Motion compensated

Original Video - Foreman

MPEG compressed videos

MPEG4 standard

- Move away from ‘traditional’ view of video sequence as collection of rectangular frames

- Treats video sequence as collection of one or more video objects - Object based coding

- Targeted for low bit rate video transmission

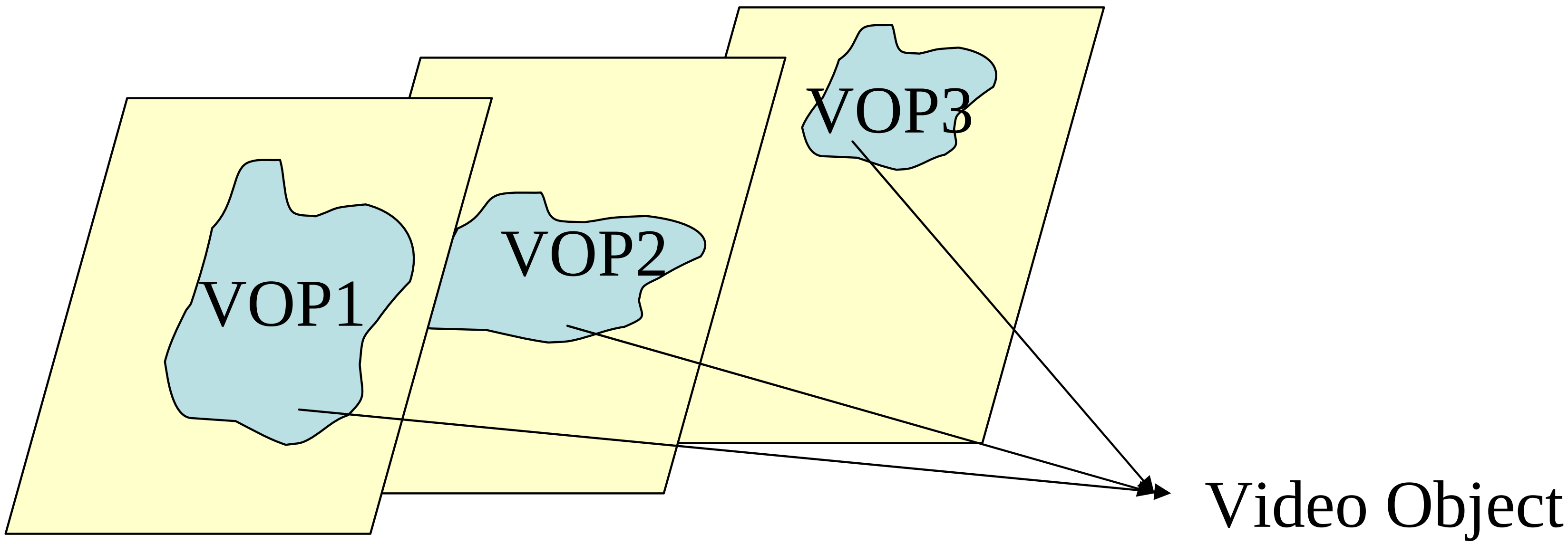
- Supports access, manipulation, and separate coding of objects within scenes

- Challenge- Segmenting objects from background

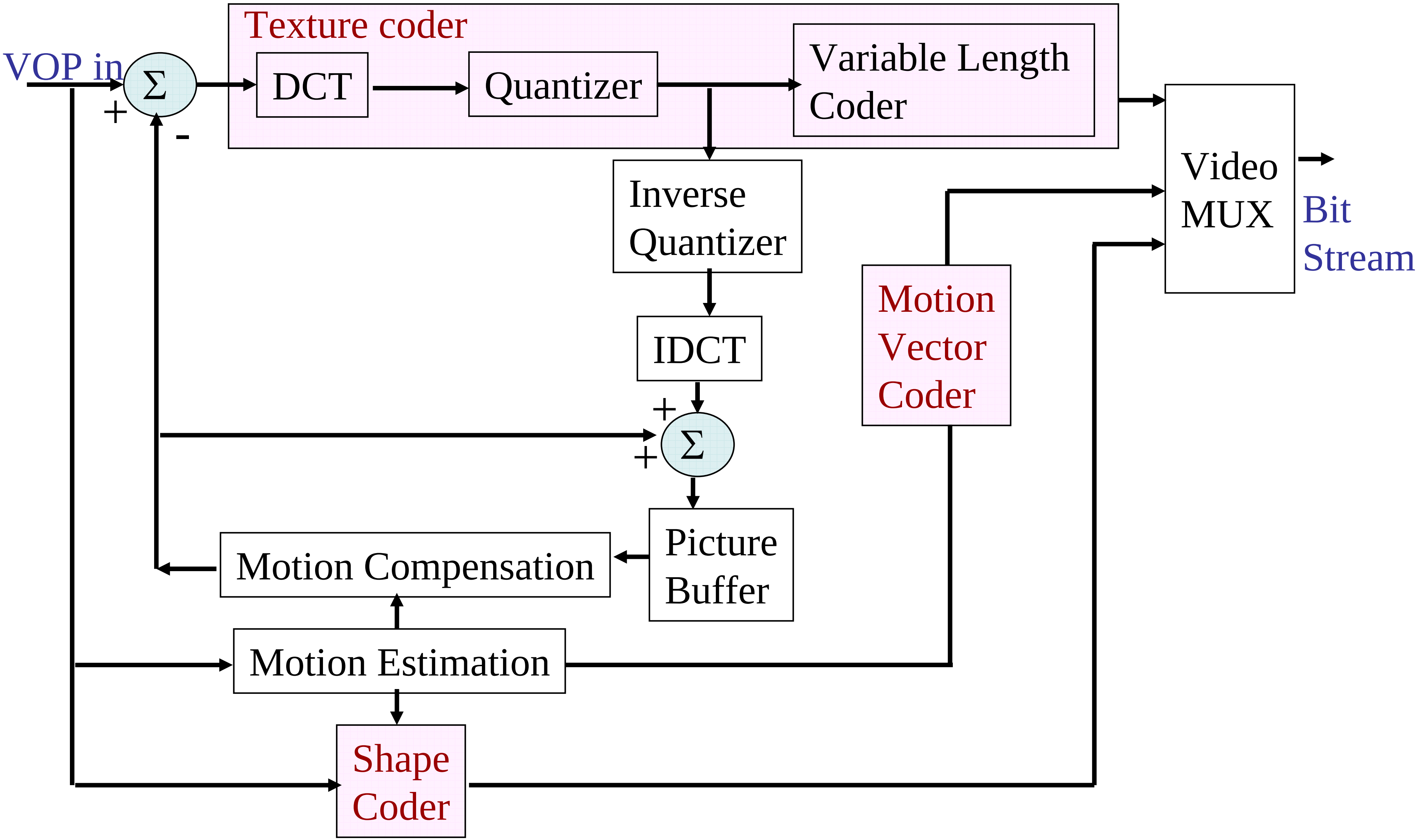
Video Object (VO)

- Area of video scene that may occupy an arbitrarily shaped region and may exist for an arbitrary length of time

- An instance of a VO at a particular point in time is a Video Object Plane(VOP)

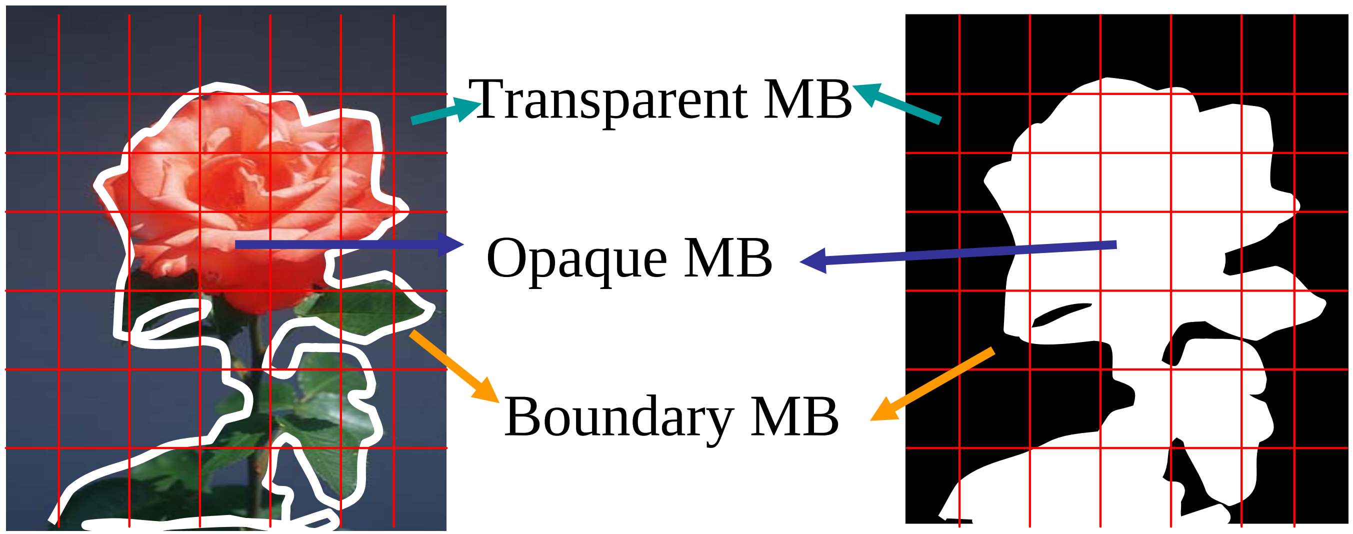


MPEG-4 Encoder *(Object based)*



Coding Arbitrary Shaped Regions

- Shape Coding
  - Shape of a video object is defined by Alpha blocks (16*16 pixel area of video scene)
  - Each alpha block may be entirely external to VO, internal to VO, or cross boundary of VO.
  - BAB (Binary alpha block) is coded using arithmetic coding

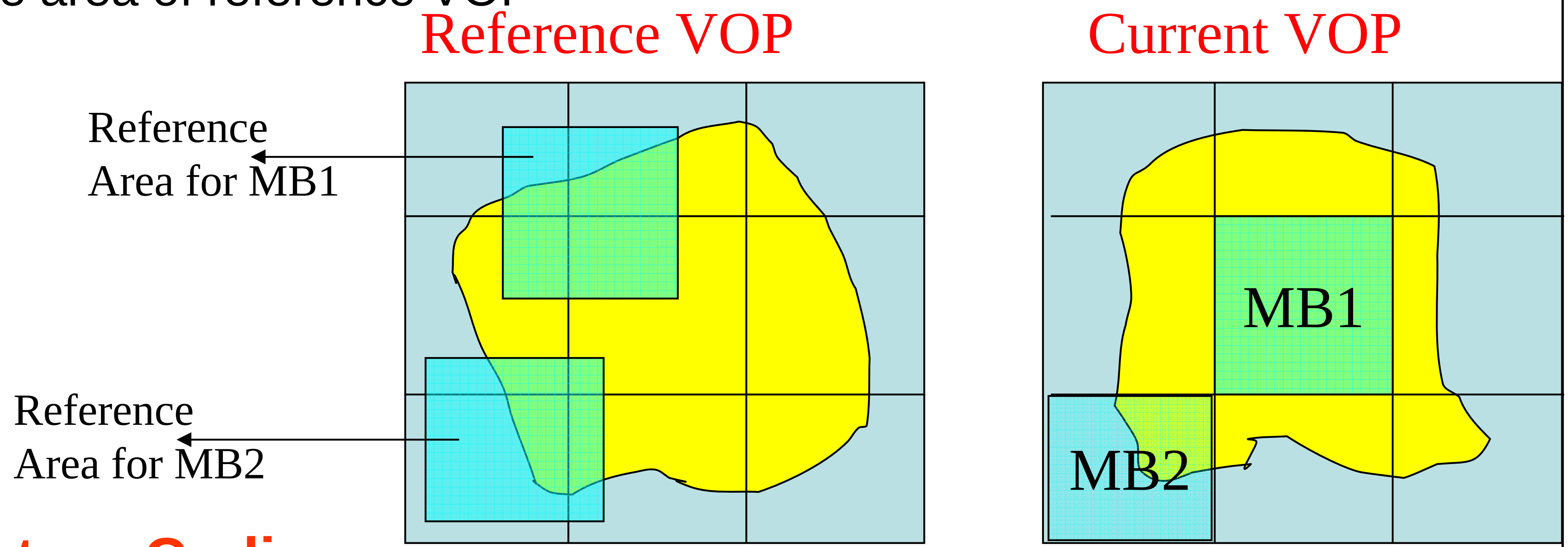


Complete VOP   Binary Alpha Mask

MB → Macroblock

• Motion Compensated coding of VOP

- Each VOP is encoded as I-VOP or P-VOP or B-VOP
- P-VOP, B-VOP is predicted from reference I or P VOP using Motion compensation
- Non-transparent pixels in boundary macroblock are motion compensated from reference VOP where boundary pixels are padded to edges of motion estimation search area
- Motion Vector may point to reference region that extends outside of opaque area of reference VOP



• Texture Coding

Motion compensated residual samples are coded using DCT, quantization, variable length coding

Scalable Video Coding

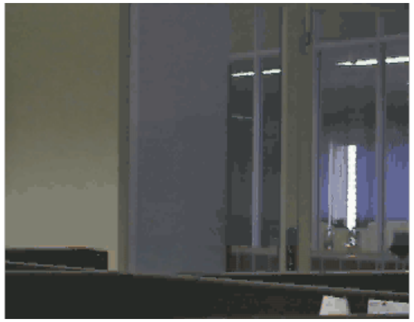
Spatial Scalability

Enables decoder to decode only part of coded bitstream

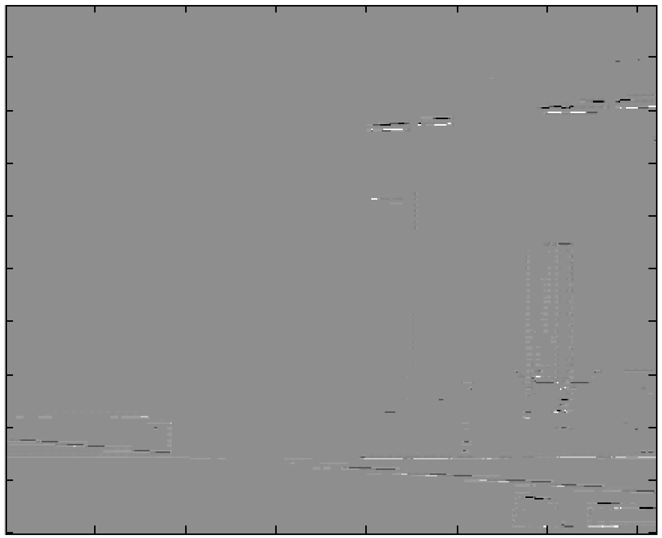
Coded stream has Base Layer, Enhancement layers

Original Frame

Sub sampled frame (base layer)



Upsampled frame (decoded)

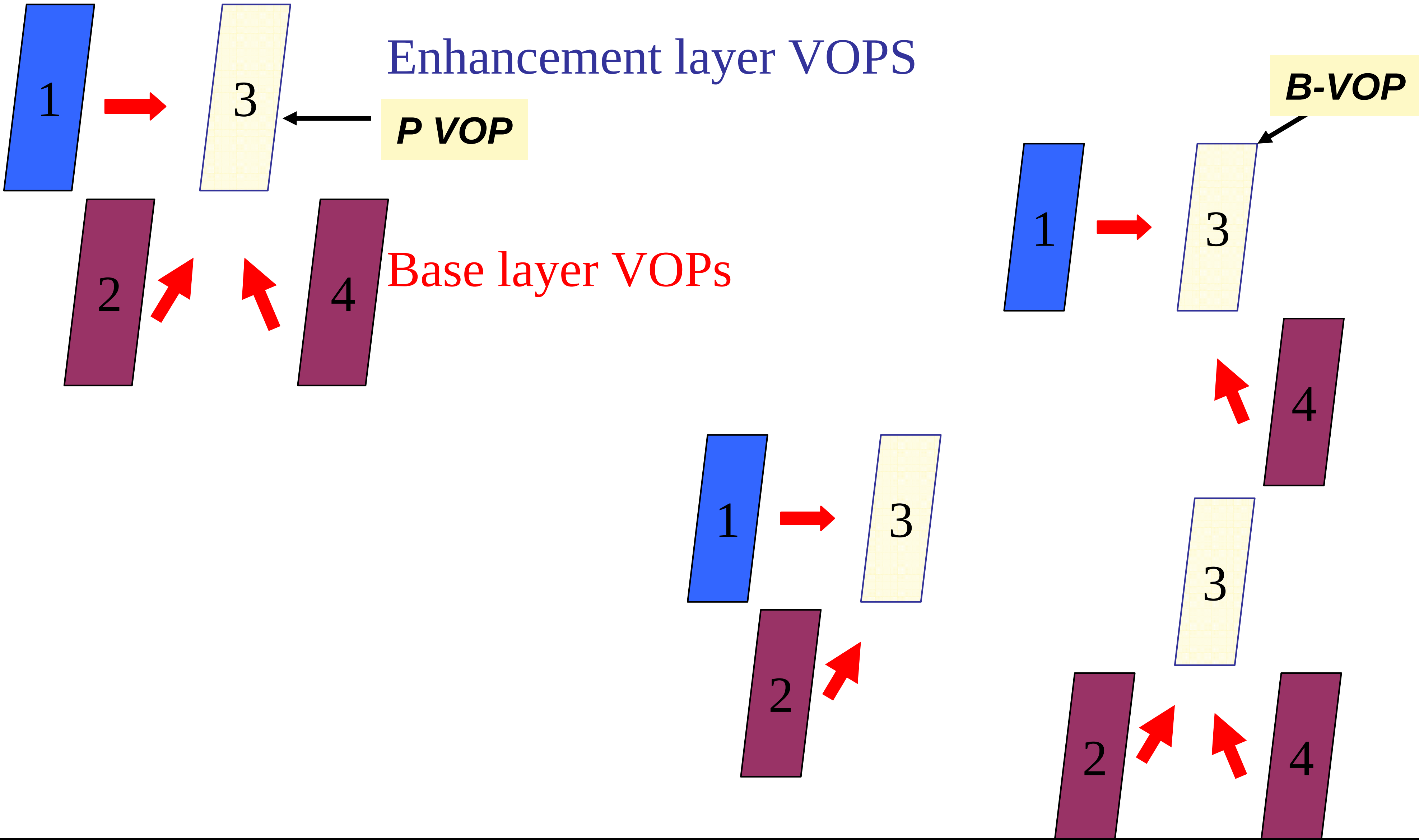


Residual (Enhancement layer)

Temporal Scalability

Base layer is encoded at low video frame rate

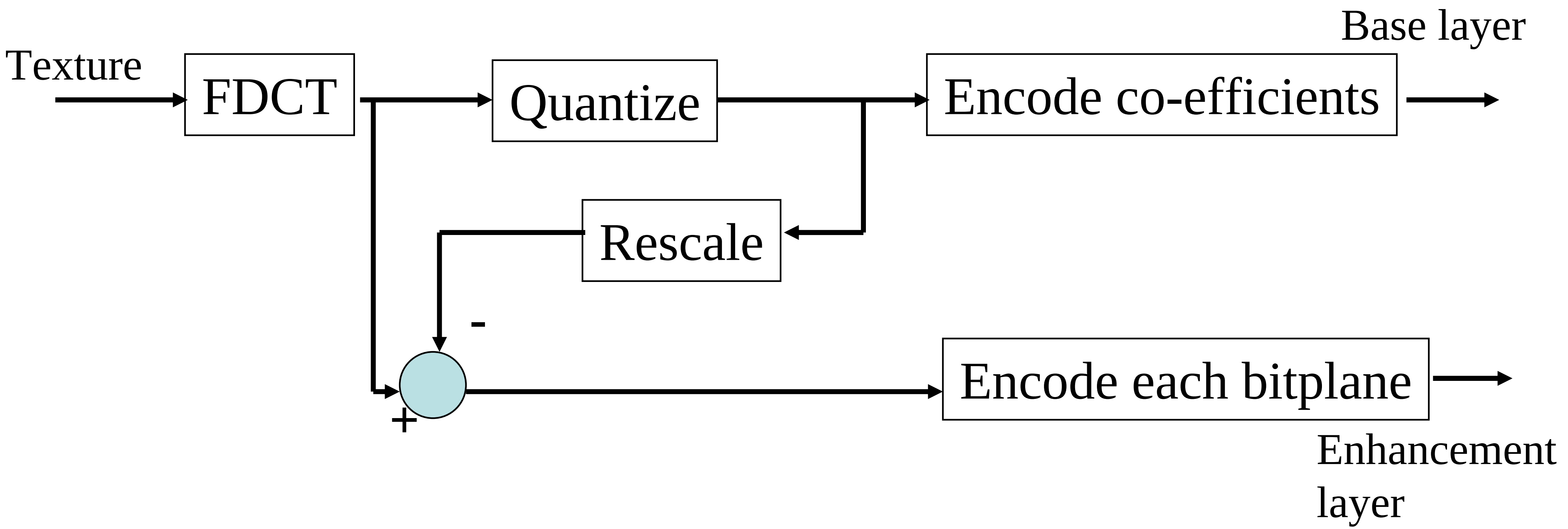
Temporal Enhancement layer provide increased video frame rate



Fine Granular Scalabilty

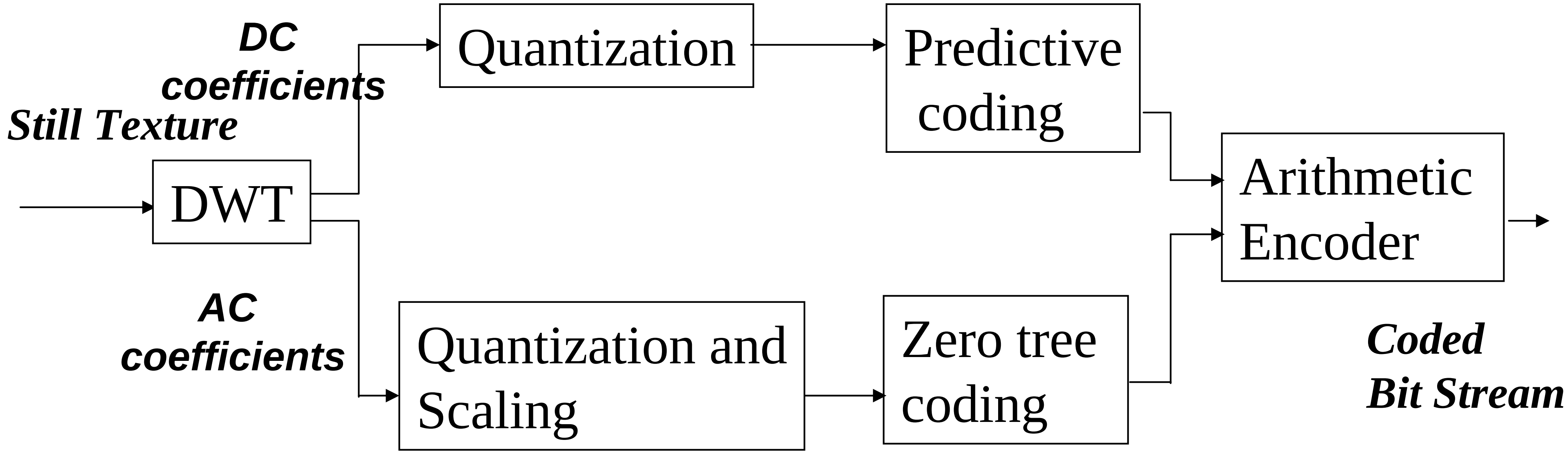
Enhancement layer can be truncated during or after encoding (reducing bit rate and quality)

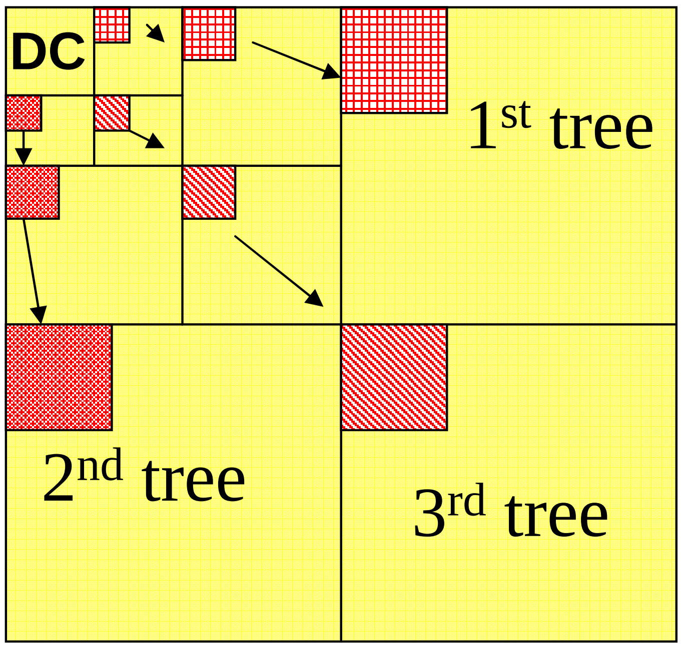
Useful in Video streaming Applications

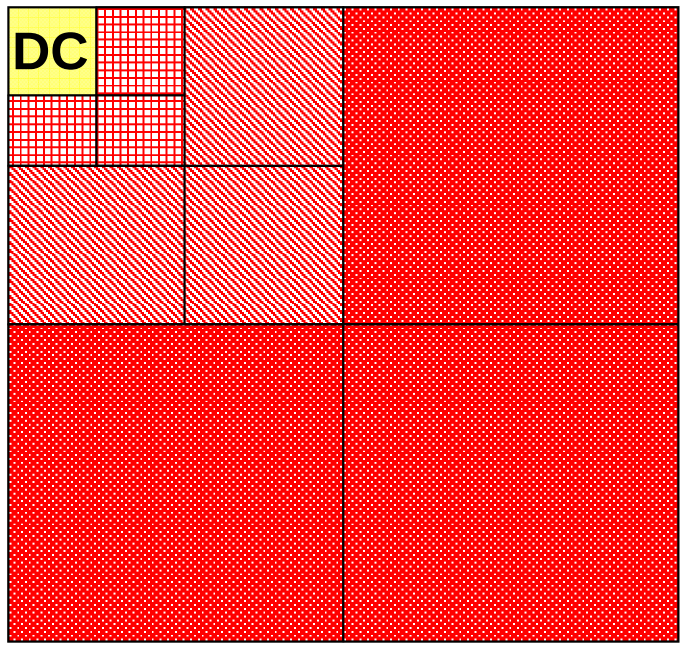


Texture Coding

Discrete Wavelet transform for coding still images







Tree order Scanning **Band by Band Scanning**

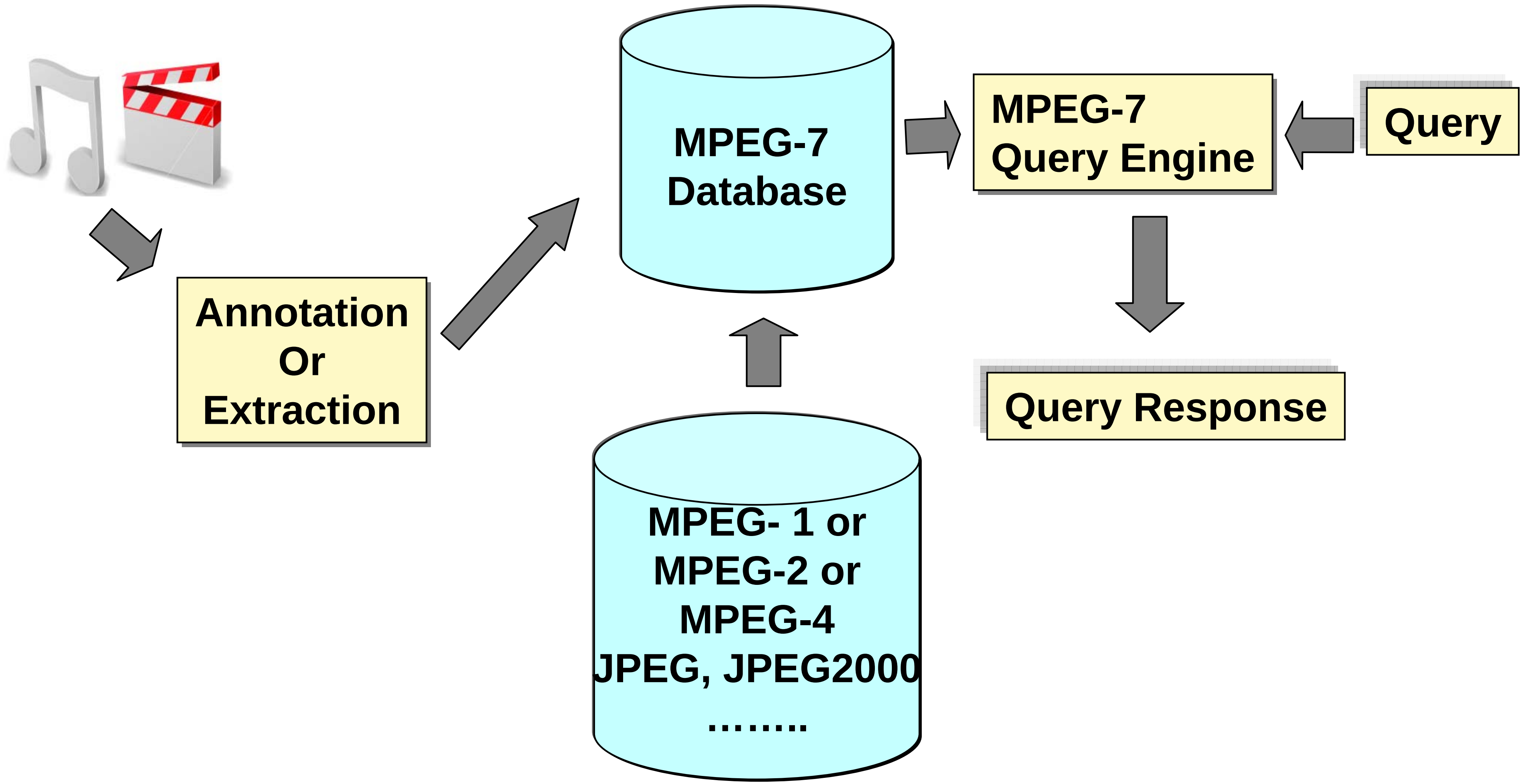

Spatial scalability
6.5 kbps
133.9 kbps
SIEMENS
SIEMENS
21.6 kbps
436.3 kbps
Quality (SNR) scalability

Coding Synthetic Visual Scenes

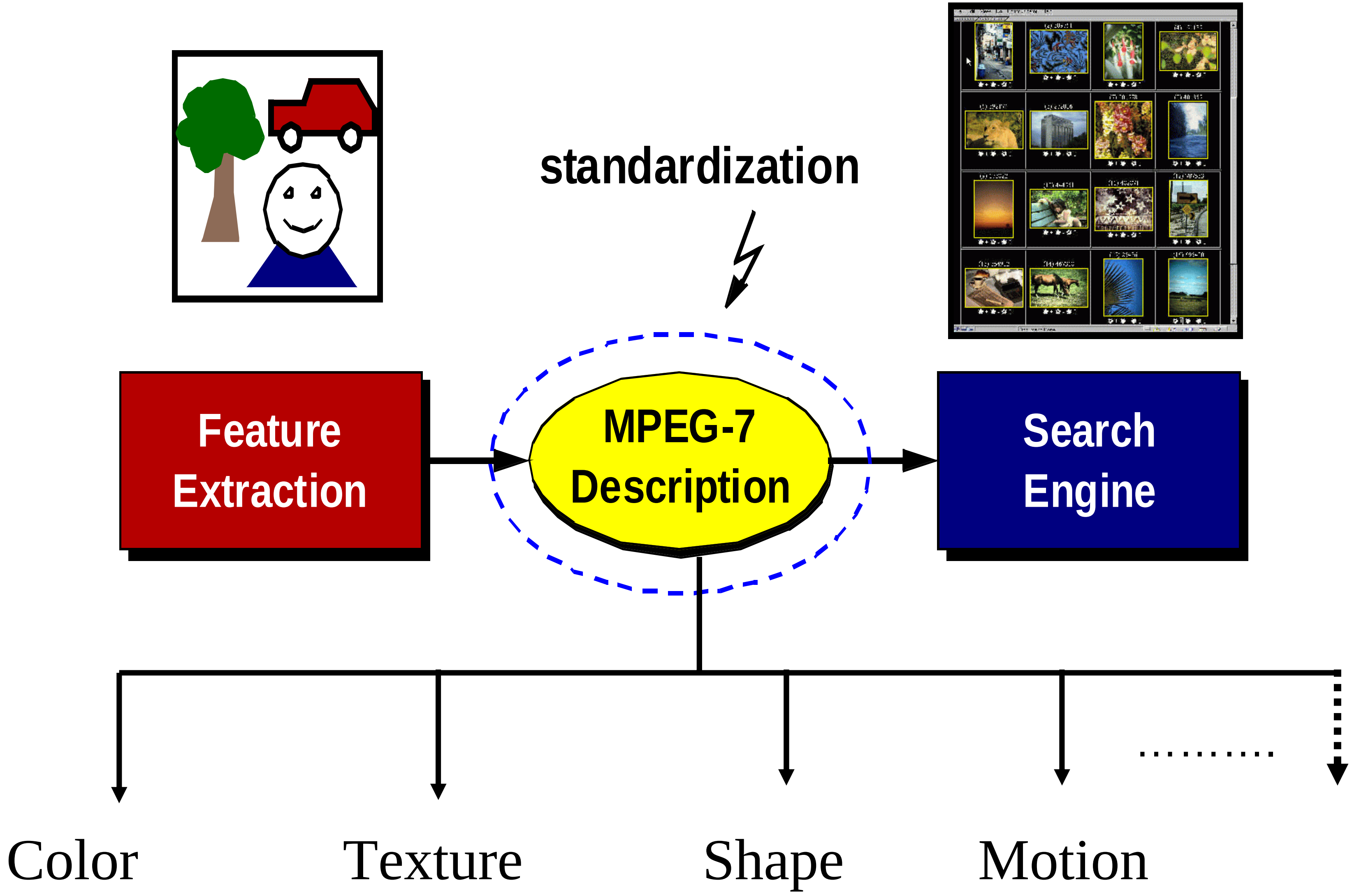
- MPEG4 introduced concept of hybrid synthetic and natural video objects
- Surface texture of 2D and 3D mesh may be compressed as static Texture image (using DWT)
- Mesh parameters are transmitted to generate intermediate frames
- Face model described by Facial Definition parameters and Animated using Facial Animation parameters

MPEG-7 *(Content based Description)*

MPEG-7 is a complementary to MPEG-4 and not an enhancement



MPEG-7


standardization
Feature Extraction
MPEG-7 Description
Search Engine
Color
Texture
Shape
Motion
....

The *MPEG standards* consist of different Parts. Each part covers a certain aspect of the whole specification. The standards also specifies Profiles and Levels. Profiles are intended to define a set of tools that are available, and Levels define the range of appropriate values for the properties associated with them. MPEG has standardized the following compression formats and ancillary standards:

MPEG-1: is the first compression standard for audio and video. It was basically designed to allow moving pictures and sound to be encoded into the bitrate of a Compact Disc. To meet the low bit requirement, MPEG-1 downsamples the images, as well as using picture rates of only 24-30 Hz, resulting in a moderate quality. It includes the popular Layer 3 (MP3) audio compression format.

MPEG-2: Transport, video and audio standards for broadcast-quality television. MPEG-2 standard was considerably broader in scope and of wider appeal--supporting interlacing and high definition. MPEG-2 is considered important because it has been chosen as the compression scheme for over-the-air digital television ATSC, DVB and ISDB, digital satellite TV services like Dish Network, digital cable television signals, SVCD, and DVD.

MPEG-3: Developments in standardizing scalable and multi-resolution compression which would have become MPEG-3 were ready by the time MPEG-2 was to be standardized; hence, these were incorporated into MPEG-2 and as a result there is no MPEG-3 standard. MPEG-3 is not to be confused with MP3, which is MPEG-1 Audio Layer 3.

MPEG-4: MPEG-4 uses further coding tools with additional complexity to achieve higher compression factors than MPEG-2. In addition to more efficient coding of video, MPEG-4 moves closer to computer graphics applications. In more complex profiles, the MPEG-4 decoder effectively becomes a rendering processor and the compressed bitstream describes three-dimensional shapes and surface texture. MPEG-4 also provides Intellectual Property Management and Protection (IPMP) which provides the facility to use proprietary technologies to manage and protect content like digital rights management.

Several new higher-efficiency video standards (newer than MPEG-2 Video) are included (an alternative to MPEG-2 Video), notably:

MPEG-4 Part 2 (or Advanced Simple Profile) and
MPEG-4 Part 10 (or Advanced Video Coding or H.264). MPEG-4 Part 10 may be used on HD DVD and Blu-ray discs, along with VC-1 and MPEG-2.

In addition, the following standards, while not sequential advances to the video encoding standard as with MPEG-1 through MPEG-4, are referred to by similar notation:

MPEG-7: A multimedia content description standard.

MPEG-21: MPEG describes this standard as a multimedia framework.

Read about:

- **Loop Filtering**
- **Rate Control**
- **Scalable Coding – bitstream scalability**
- **Fine granularity**
- **Drift problem**
- **DWT and DCT**
- **Optical Flow equations**