Unit 5
Digital Video

CS 3570

Video

- Video is made up of a series of still images (*frames*) played one after another at high speed
- This fools the eye into believing that it is observing a continuous stream

  - Video (real-world pictures)
  - Animation (Computer generated)

Video and Film

- ***Movie*** : a story told with moving images and sound
- The word ***film*** seems to imply a movie that is shot and/or stored on cellulose film.
- Film and video both rest on the same phenomenon of human perception, called ***persistence of vision*** – the tendency of human vision to continue to "see" something for a short time after it is gone.
- A related physiological phenomenon is ***flicker fusion***— the human visual system's ability to fuse successive images into one fluid moving image.

Standard film aspect ratios

- ***Aspect ratio*** is the ratio of the width to the height of a frame, expressed as *width : height*.
- Two common video graphic aspect ratios are **4:3**(1.33:1), the universal video format of the 20th century, and **16:9** (1.77:1), universal for high-definition television
- The most common aspect ratios used today in the presentation of films in cinemas are **1.85:1** and **2.39:1**

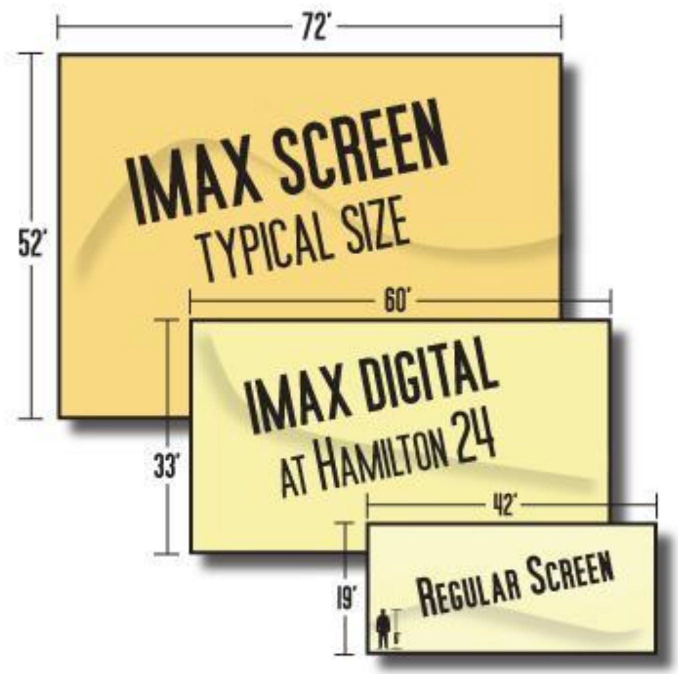
Academy aspect ratio 1.33:1

Widescreen, 1.85:1

Anamorphic widescreen, 2.39:1

Film Development

- Silent movies and early sound movies were shot mostly on 16 mm film, introduced by Eastman Kodak in 1923
- **IMAX** movies are shot on 70 mm film with aspect ratio of 1.43:1.



- **IMAX** movies are on very large screens, e.g. a standard IMAX screen is 22 m × 16.1 m (72 ft × 52 ft).

High-definition television

- In 1981, NHK began broadcasting what came to be known as ***high-definition television, HDTV.***
- The current definition of HDTV is television that has an aspect ratio of 16: 9, surround sound, and one of three resolutions: 1920 $\times$ 1080 using interlaced scanning(**1080i**), 1920 $\times$ 1080 using progressive scanning(**1080p**), or 1280 $\times$ 720 using progressive scanning (**720p**).
- Digital encoding is not part of this definition, and, historically, HDTV was not always digital.

Video and film displays

- Like film, video is created by a sequence of discrete images, called **frames**, shown in quick succession.
- Film is displayed at 24 frames/s. The standard frame rate for NTSC video is about 30 frames/s. The frame rate for PAL and SECAM video is 25 frames/s.
- A film frame is a continuous image. Video frames, in contrast, are divided into lines. Television has to be transmitted as a signal, line-by-line.
- Video is displayed (and recorded) by a process called ***raster scanning.*** The raster refers to a single frame.

Raster scanning

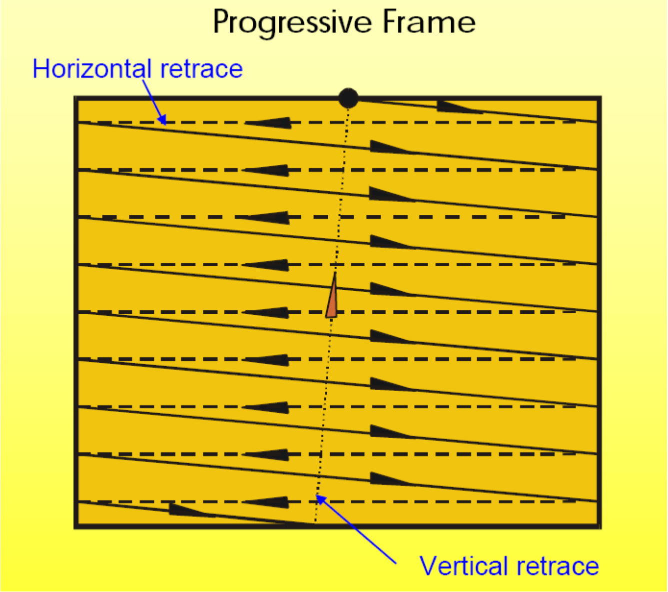
- The scanning process is a movement from left to right and top to bottom.
- When the scanner has finished with one line, it moves back to the left to start another in a motion called ***horizontal retrace***.
- ***Vertical retrace*** takes the scanner from the bottom of the monitor to the top again.
- In the case of video camera, the purpose of the scanning is to record the data that will be saved and/or transmitted as the video signal.

Raster scanning

- For many years, the dominant video display technology was the ***cathode ray tube*** (***CRT***). Most television sets were built from CRTs, as were the computer monitors.
- Scanning can be done by one of two methods: either **interlaced** or **progressive scanning.**
- In ***interlaced scanning***, the lines of a frame are divided into two ***fields***: The odd-numbered lines, called the ***upper field (odd field)***, and the even-numbered lines, called the ***lower field (even field).***
- Video standards are sometimes described in terms of ***field rate*** rather than frame rate.
  - For PAL analog video, 50 fields/s = 25 frames/s

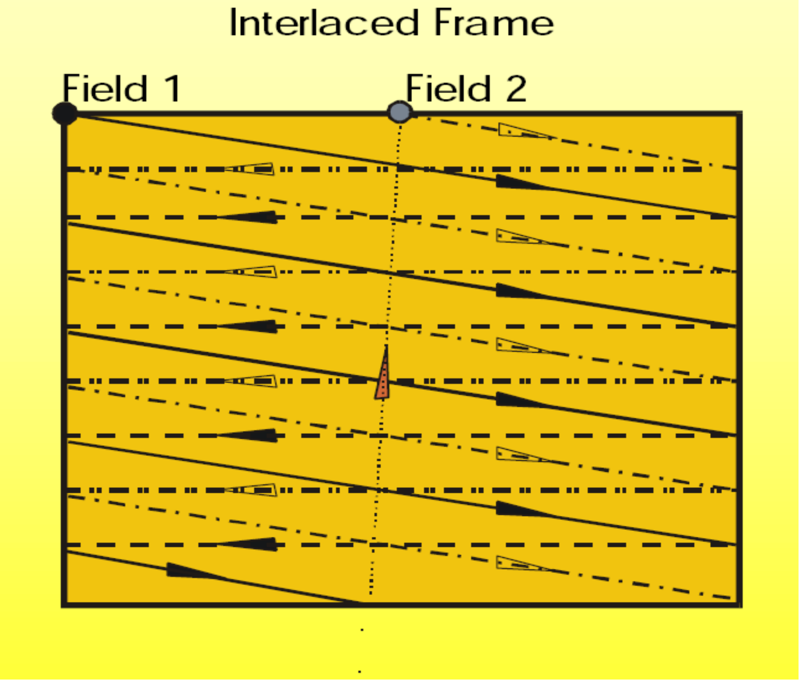
Progressive scanning

- In ***progressive scanning***, each frame is scanned line-by-line from top to bottom.
- For progressive scanning, the frame rate and field rate are the same because a frame has only one field.
- Computer monitors and many digital televisions use progressive scanning.

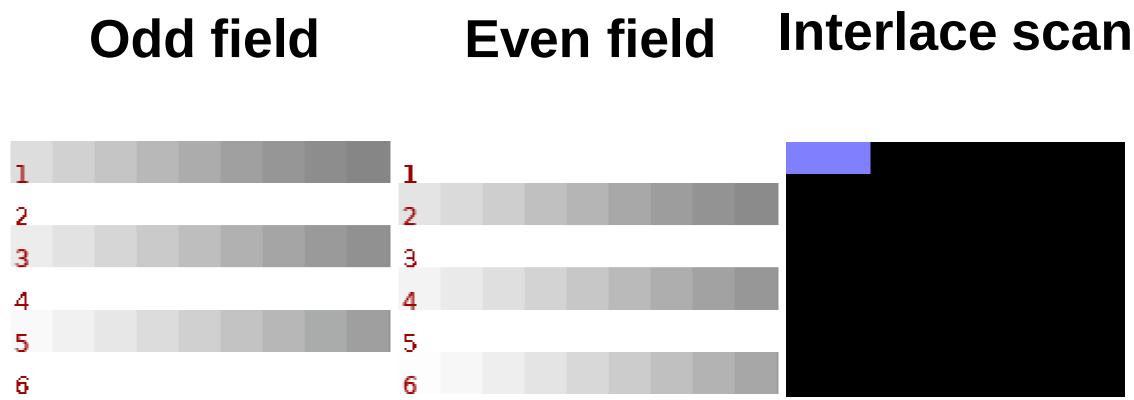


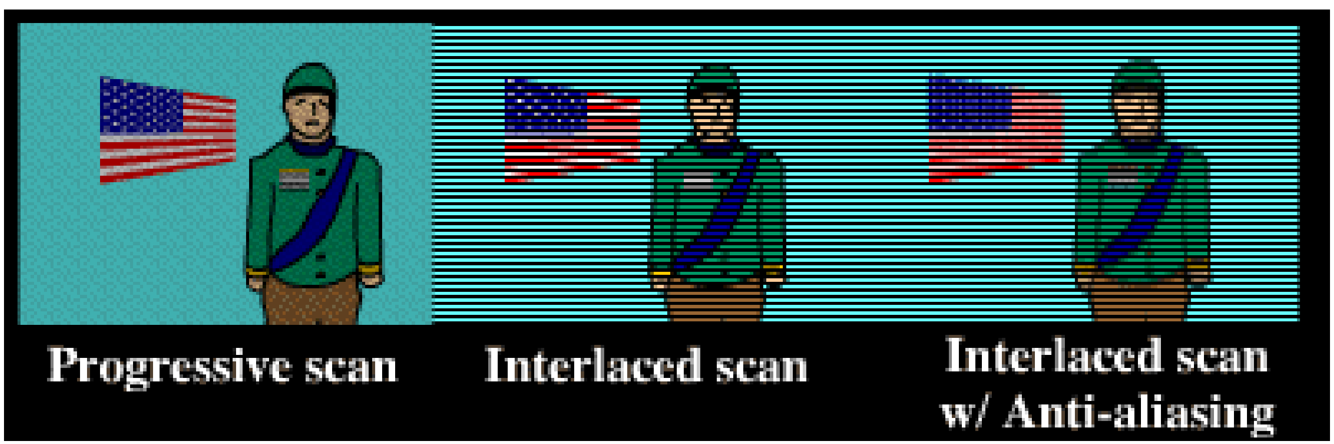
Interlaced Scan

- **In interlaced scan, each frame is scanned in two fields and each field contains half the number of lines in a frame.**



Interlaced Scan

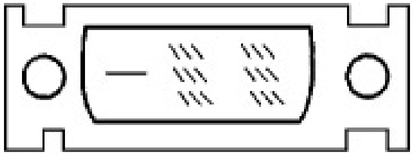




Digital video transmission format

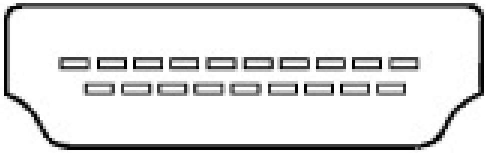
- There are two main types of digital video transmission format: ***DVI (digital video interface),*** and ***HDMI (high definition multimedia interface).***
- *DVI* connects an uncompressed digital video source (e.g., from a video card) to a digital display device.
- There are three basic DVI formats:
  - DVI-D, for a true digital-to-digital connection
  - DVI-A, connects and convert a digital signal to an analog display
  - DVI-I, transmit both digital-to-digital and analog-to-analog

DVI-D interface



Digital video transmission format

- *HDMI* is an audio/video connection for transmitting uncompressed digital data.
- It is backward compatible with DVI and accommodates audio data on the signal.
- HDMI connections can apply ***HDCP (high bandwidth digital content protection)*** to signal.
  - HDCP is a digital copy protection protocol that prevents unauthorized copying.



HDMI interface

High Definition HDV Format

Picture Format		720/25p, 720/30p, 720/50p, 720/60p	1080/50i 1080/60i
Pixel Dimensions		1280×720	1440×1080
Frame Aspect Ration		16:9	
Pixel Aspect Ratio		1.0	1.33
Data Rate	Video data only:	approx. 19 Mbps	approx. 25 Mbps
Color Sampling Method		YUV 4:2:0	
Audio Setting	Sampling rate and bit depth:	48 kHz 16-bit	
	Bit rate after compression:	384 kbps	

HDV Picture Format Notation

1080 / 60 i

"i": interlaced
"p": progressive

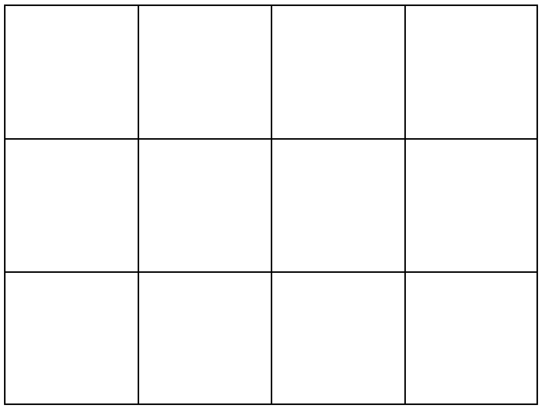
frame/field frequency

frame height

Frame Size Examples

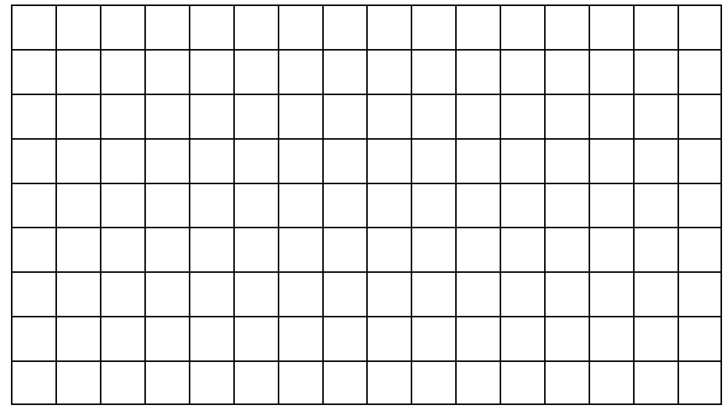
		Frame size
NTSC	standard definition	720 x 480 pixels
	high definition HDV format	1280 x 720 pixels 1440 x 1080 pixels
PAL	standard definition	720 x 576 pixels

Frame Aspect Ratio Examples



4:3

Example:
•Standard definition NTSC standard format



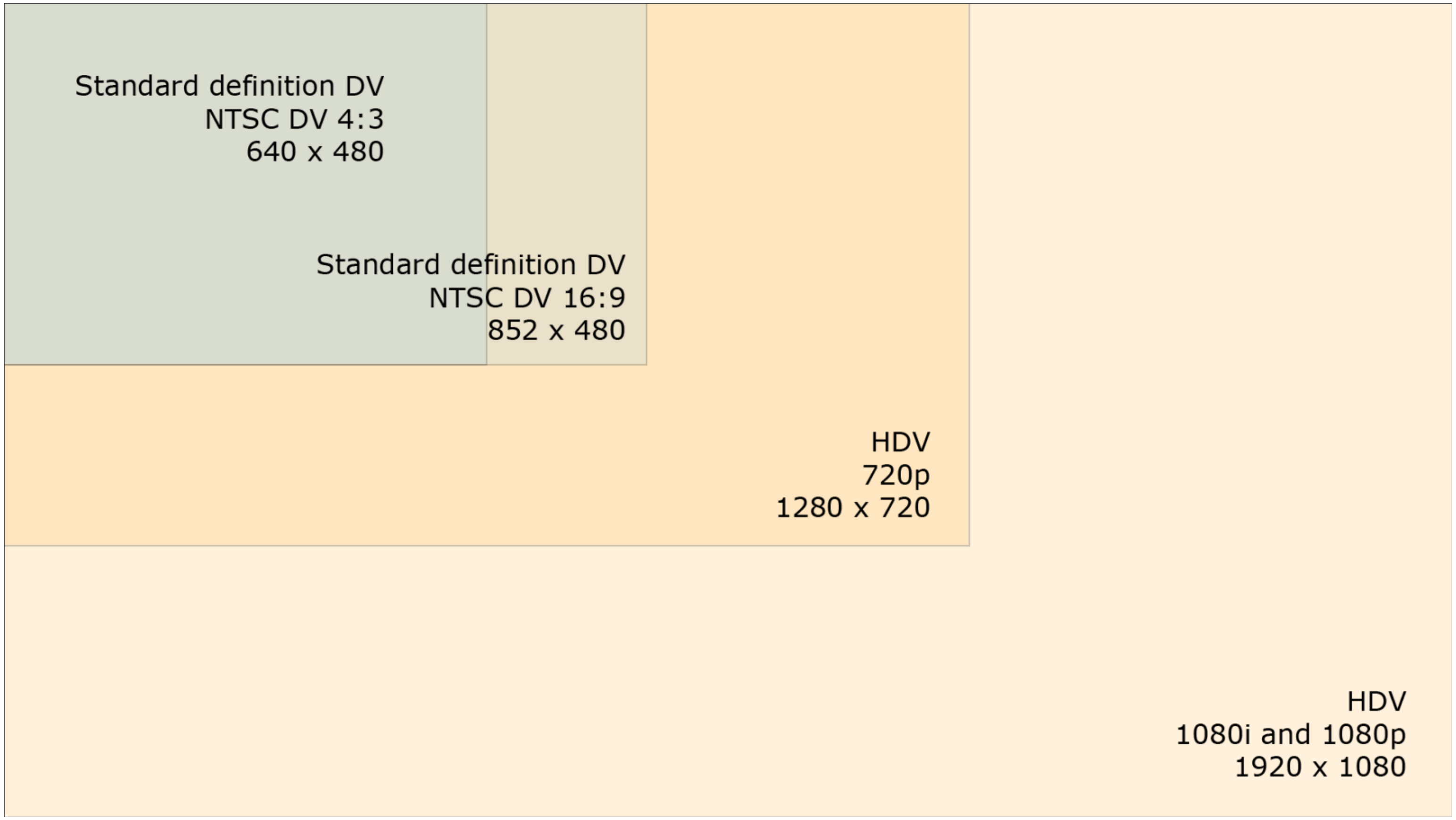
16:9

Examples:
•Standard definition NTSC wide-screen format

•High definition digital video

•High definition TV

Frame Size (Resolution) Comparison between Standard Definition and High Definition



By viewing frame size

Comparison of 4: 3 and 16: 9 image aspect ratio

16:9 Aspect ratio

16:9 Aspect ratio displayed on a 4:3 screen (letter box)

→ letter box

4:3 Aspect ratio

4:3 Aspect ratio displayed on a 16:9 screen (pillar box)

→ Pillar box

Frame Size (Resolution) Comparison between Standard Definition and High Definition

A frame from a 1080i video

Frame Size (Resolution) Comparison between Standard Definition and High Definition

Same frame as 720p

Frame Size (Resolution) Comparison between Standard Definition and High Definition

Same frame as standard definition DV wide-screen (16:9)

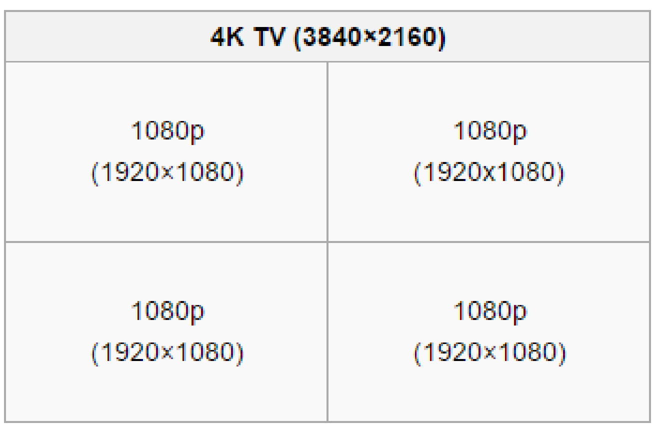
Same frame as standard definition DV standard 4:3

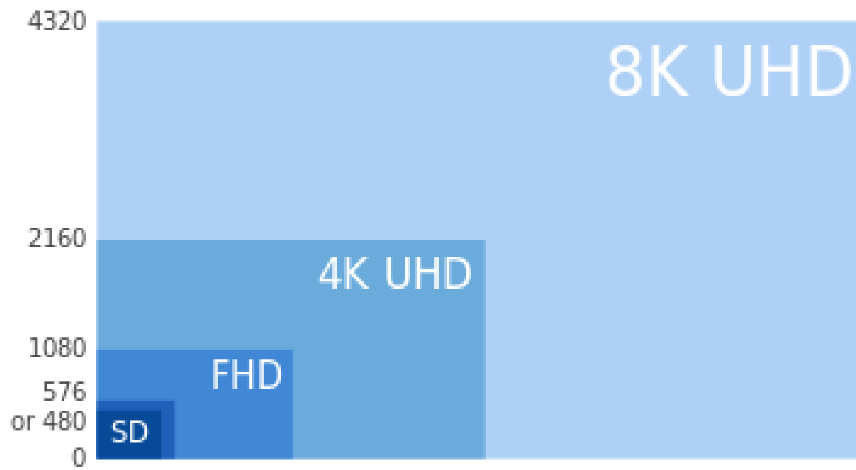
Higher Resolution Nowadays

There are 2 common standards on 4K videos:

- A 4K resolution, as defined by Digital Cinema Initiatives, is 4096 x 2160 pixels(1.9:1 aspect ratio)
- Ultra HD(UHD) 4K resolution is 3840 x 2160 (1.78:1 aspect ratio)

8K resolution or FUHD (Full Ultra HD) is 7680 x 4230.





Digital Video Distribution Media

- Overview of types of CDs and DVDs

TABLE 6.7 Digital Video Distribution Media

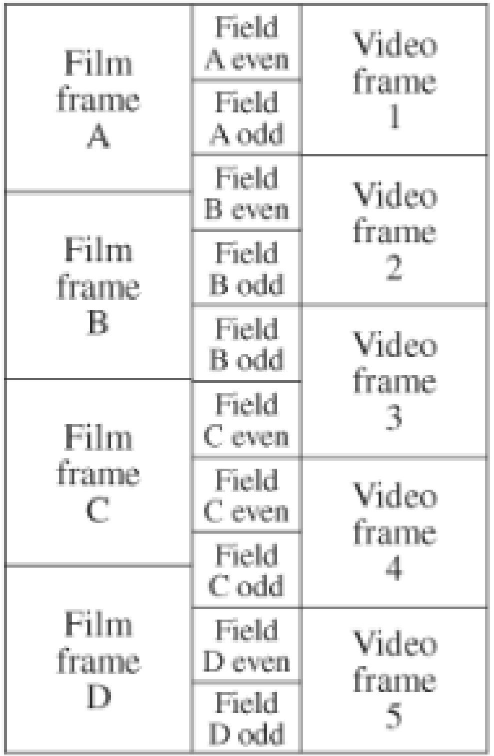
Format	VCD	SVCD	DVD	HD-DVD	Blu-ray
NTSC Resolution	352 × 240	480 × 480	720 × 480	1920 × 1080	1920 × 1080
Video Compression	MPEG-1	MPEG-2	MPEG-2	MPEG-2, MPEG-4 AVC, SMPTE-VC1	MPEG-2, MPEG-4 AVC, SMPTE-VC1
Audio Compression	MP1	MP1	PCM, DD, DTS Surround	PCM, DD, DD$^+$, DD, TrueHD, DTS, DTS-HD	PCM, DD, DD$^+$, DD, TrueHD, DTS, DTS-HD
Video Bit Rate	~1.2 Mb/s	~2 Mb/s	~10 Mb/s	~28 Mb/s	~40 Mb/s
Length (in time)	74 min. on CD	35–60 min. on CD	1–4 hours or more of SD	2 hours or more of HD, depending on the number of layers	2 hours or more of HD, depending on the number of layers

Telecine and pulldown

- The word ***telecine*** refers to both the process that transfers film to video and the machine that performs the process.
- The major difficulty is that film and television have different frame rates. A better way, one that creates smoother video and truer audio, is called ***pulldown.***
- ***Pulldown*** is a method for using interlaced fields more than once, across frames, to make up for a discrepancy in frame rates as film is translated to video.
- For NTSC video, it uses ***3:2 pulldown.*** The first step is to slow down the film by 0.1% so that we can get to an integer-based ratio of frame rates.

Telecine and pulldown

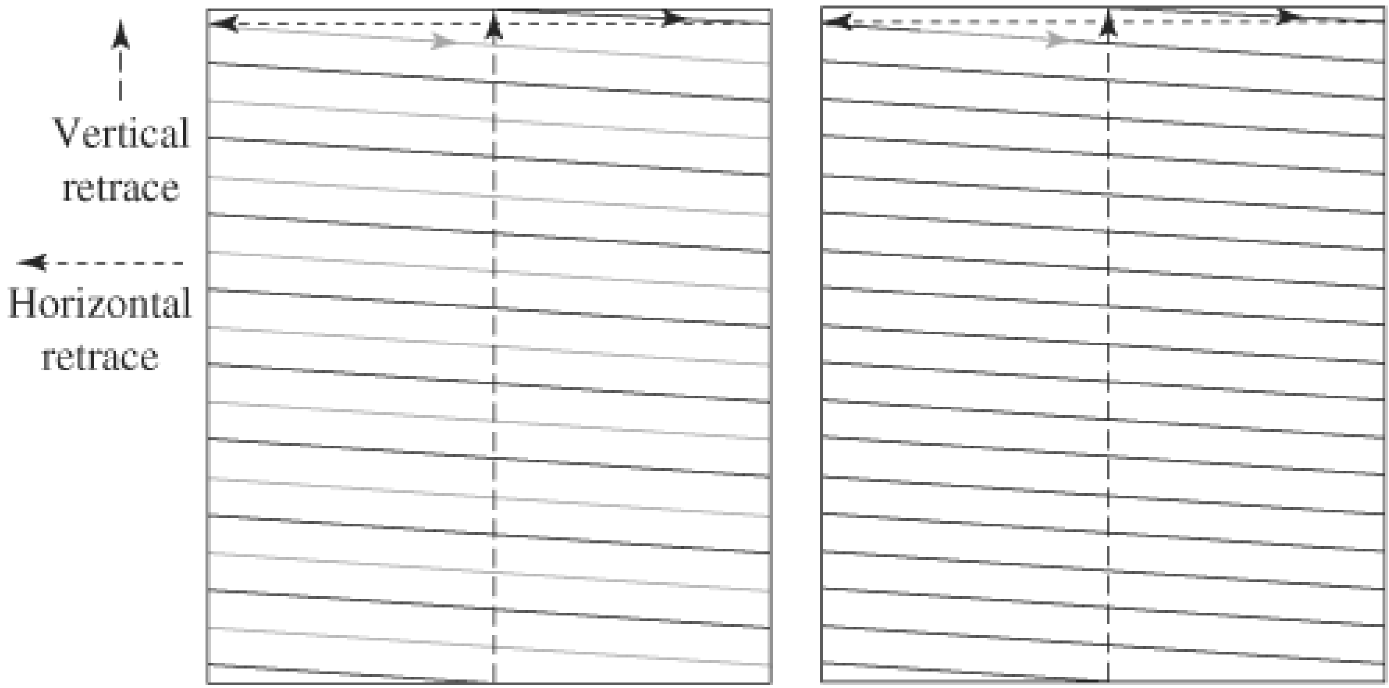
- If you multiply 24 * 0.999, you get 23.976. This gives you a ratio of 23.976/29.97, which is 4/5.
- Now these numbers are something we can deal with. For each four frames of film, we need to create five frames of video.
- Figure 6.21 illustrates 3: 2 pulldown.
- The name reflects the pattern of how many fields are used from a frame .



Interlaced vs. Progressive Scan

- Interlaced
  - Developed for CRT (Cathode Ray Tube) technology
  - Divides scans into odd and even lines
  - Alternately refreshes odd lines, then even lines
  - Slight delay between refreshes causes "jaggedness" or interlace artifacts
  - Deinterlacing can compensate somewhat
- Progressive
  - Computer monitors
  - Scans entire picture line by line
  - Eliminate flicker seen in interlaced

Interlaced and progressive scanning



Interlaced scanning:
Lower field (shown in gray) displayed first, one line at a time from top to bottom; then upper field (shown in black) displayed

Progressive scanning:
Lines displayed in order from top to bottom

Interlaced and progressive scanning

Interlaced

Deinterlaced Result

Deinterlacing

- ***Deinterlacing*** is the process of putting fields together in a way that creates a coherent frame that can be shown by a progressive-scanning display.
- It is required in two situations:
  - When material has been transferred from film to video and is being transferred back to film.
  - When video material in interlaced format is being shown on a display device that uses progressive scanning.
  - The second situation arises when HDTV is transmitted in 1080i format and received by a television that is 1080p.

Deinterlacing

- In the case of the film-to-video-to-film translation
  - If film has been transferred to video by a telecine or film-scanning process, then the pattern of field usage is known, as described above for 3:2 pulldown.
  - If nothing changed the frames after they were scanned to video, this pattern could simply be reversed.
  - Once the video is edited before being returned to film, and the pattern is disrupted.
  - Image analysis techniques have to be employed, to find segments where the pulldown patterns may still be intact.
  - Where pulldown patterns are not intact, interpolation methods have to be applied.

Deinterlacing

- In the case of translating interlaced video into progressive video
  - With video that was created from film, if two fields are created from one frame, you know that they belong to a picture that was captured at a single moment in time.
  - If the inverse telecine process can find the correct two fields, they will go together perfectly, with **no combed edges**.
  - Video frames that were created *as video* are different. With video frames, the even and odd fields from one frame are captured at different moments in time.
  - Thus, if there is motion in the scene, objects aren't in the same place in the second field as they were in the first.

Deinterlacing

- This is alright when the video is displayed in an interlaced manner, since the two fields are not displayed at exactly the same time.
- However, if you put the two fields into one progressively scanned frame and show that frame all at once, you get a **combed-edge effect.**

Combed effect from combining even and odd interlaced fields

Deinterlacing by interpolation

Deinterlacing

- An easy way to accomplish interlacing is called ***doubling :*** choosing either the even or the odd field and using the chosen field twice to create a frame.
- An alternative is to average the even and odd fields and use the average for both.
- Both doubling and averaging reduce the resolution of the frame.

Properties of codecs

- Digital video files are very large. With no compression or subsampling, NTSC standard video would have a data rate of over 240 Mb/s; HD would have a data rate of about 1 Gb/s.
- Remove redundancies and extraneous information within one frame is called ***intraframe compression***. It also can be referred to as ***spatial compression.***
- There are two commonly used methods for accomplishing spatial compression: transform encoding and vector quantization.
- ***Temporal compression*** is a matter of eliminating redundant or unnecessary information by considering how images change over time. it is also called ***interframe compression***.

Properties of codecs

- The basic method for compressing between frames is to detect how objects move from one frame to another, represent this as a vector.
- Determining the motion vector is done by a method called ***motion estimation.***
- Some codecs allow you to select either ***constant*** or ***variable bit rate encoding*** (***CBR*** and ***VBR***, respectively). Variable bit rate varies the bit rate according to how much motion is in a scene.
- Codecs are mostly ***asymmetrical***. This means that the time needed for compression is not the same as the time needed for decompression.

Different kinds of codecs

- **Vector quantization**
  - Create a palette for a frame. The palette represents the frame's dominant colors and color patterns and serves as a code table.
  - Divide the frame into areas.
  - Encode the area by an index into the code table.
- **Motion JPEG compression (MJPEG)**
  - Apply JPEG compression frame-by-frame
- **DV compression**
  - Standard ***DV compression*** produces resolutions of 720×480 for NTSC and 720×576 for PAL
  - DV cameras take an RGB color signal, convert it to YCbCr, downsample to 4: 1: 1 (NTSC) or 4: 2: 0 (PAL)

Compression Standards

- **JPEG** issued by ISO in 1989 (but adopted by ITU as ITU T.81)
- **MPEG 1** released by ISO in 1991,
- **H.261** released by ITU in 1993,
- **H.262** (better known as MPEG 2) released in 1994.
- **H.263** released in 1996 extended as H.263+, H.263++.
- **MPEG 4** released in 1998.
- **H.264** releases in 2002 to lower the bit rates with comparable quality video and support wide range of bit rates, and is now part of MPEG 4 (Part 10, or AVC - Advanced Video Coding).
- **H.265 (HEVC)** releases in 2013.

MPEG compression

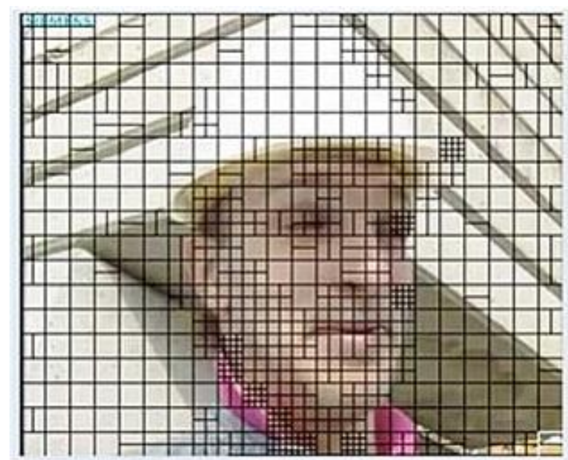
- MPEG compression was developed in two lines.
  - The first was the work of ITU-T and their subcommittee, the **Video Coding Experts Group**. We know this line of codecs as the H.26* series
  - The second line emerged from the **Motion Picture Experts Grou**p, from which we get the name MPEG
- The revolutionary advance in MPEG-4 compression is the use of object-based coding.
- ***MPEG-4 AVC*** (Advanced Video Coding) and equivalent to ***H.264***, is an improved MPEG-4 version introduced in 2003 that quickly achieved wide adoption for DVD; videoconferencing; videophone…

MPEG General Information

- Goal: data compression 1.5 Mbps
- MPEG defines video, audio coding and system data streams with synchronization.
- MPEG information
  - Aspect ratios: 1:1 (CRT), 4:3 (NTSC), 16:9 (HDTV)
  - Refresh frequencies: 23.975, 24, 25, 29.97, 50, 59.94, 60 Hz

MPEG Image Preparation - Blocks

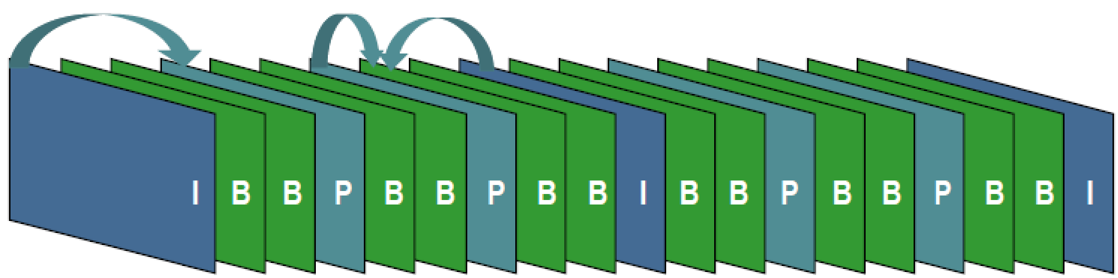
- Each image is divided into **macro-blocks.**
- Macro-block : 16x16 pixels for luminance; 8x8 for each chrominance component.
- Macro-blocks are useful for Motion Estimation.
- No MCUs which implies sequential non-interleaving order of pixels values.



H.264

Main steps of MPEG compression

- **Step 1.** Divide the sequence of frames into GOPs, identifying I, P, and B frames.
  - A ***GOP*** is a ***group of pictures***, that is, a group of *n* sequential video frames
  - ***I frames***, or ***intraframes***, are compressed independently, as if they were isolated still images, using JPEG compression.
  - I frames serve as reference points for the P frames (***interframes***, also called ***forward prediction frames***) and B frames (***bidirectional frames***), which are compressed both spatially and temporally.



Main steps of MPEG compression

- **Step 2.** Divide each frame into macroblocks.
  - A ***macroblock*** is a 16 × 16 pixel area.
  - A 16 × 16 macroblock can be divided into 8 × 8 blocks. The way that macroblocks are divided depends on the particular compression standard, which can apply different types of chrominance subsampling.

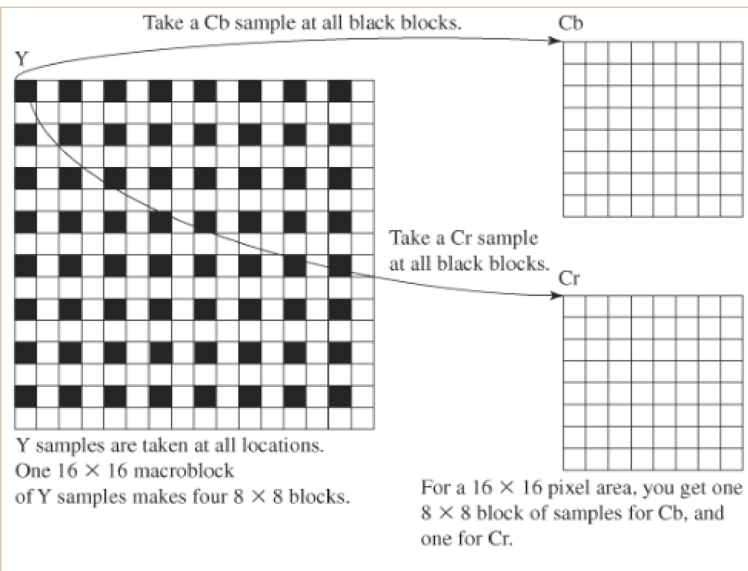


Figure 7.22 Macroblock for 4: 2: 0 subsampling

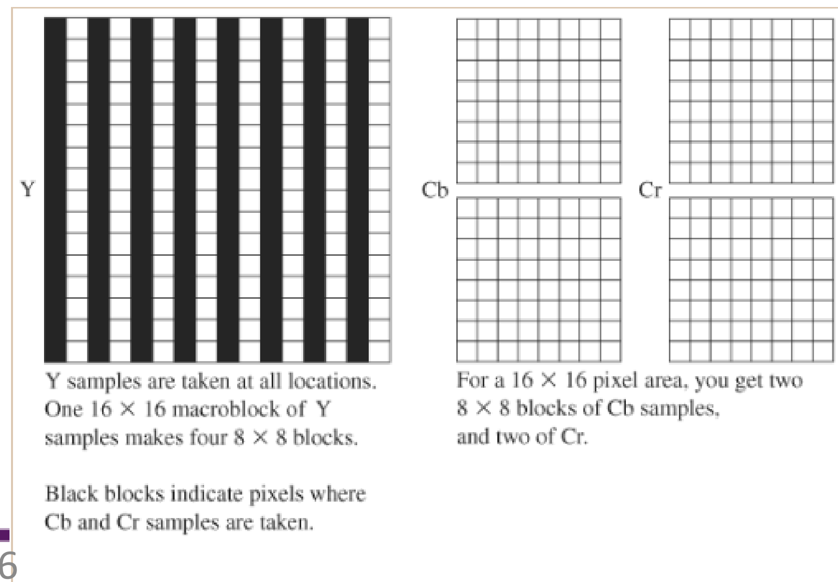
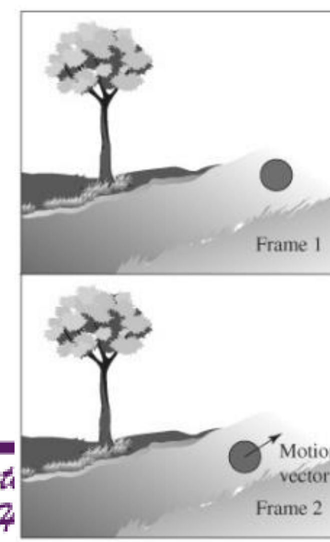


Figure 7.23 Macroblock for 4: 2: 2 subsampling

Main steps of MPEG compression

- **Steps 3 and 4.** For each P and B frame, compare the frame to the related I frame to determine a motion vector. Record differential values for P and B frames.
  - This step is called ***motion estimation.***
  - It's more economical to convey the difference between one frame and the next, a method called ***differential encoding.***
  - Motion estimation determines how much a frame has "moved" since the previous frame.
  - The difference between the macroblock in frame 2 and the matching macroblock in frame 1 is called the ***prediction error***.

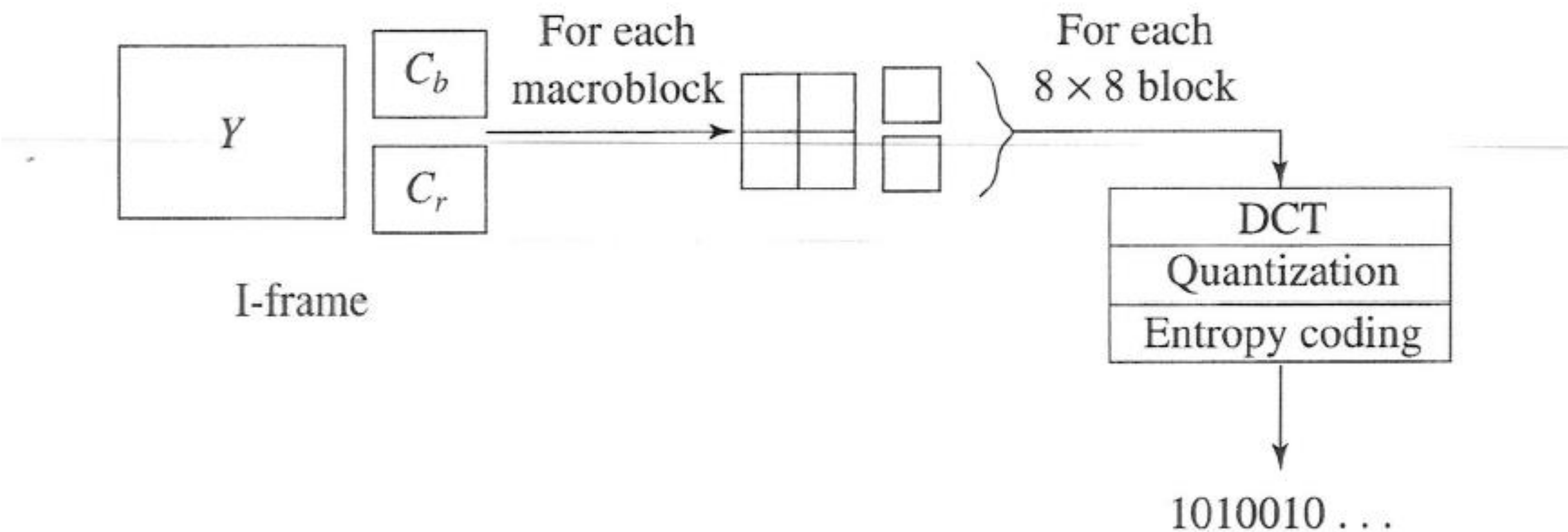


Main steps of MPEG compression

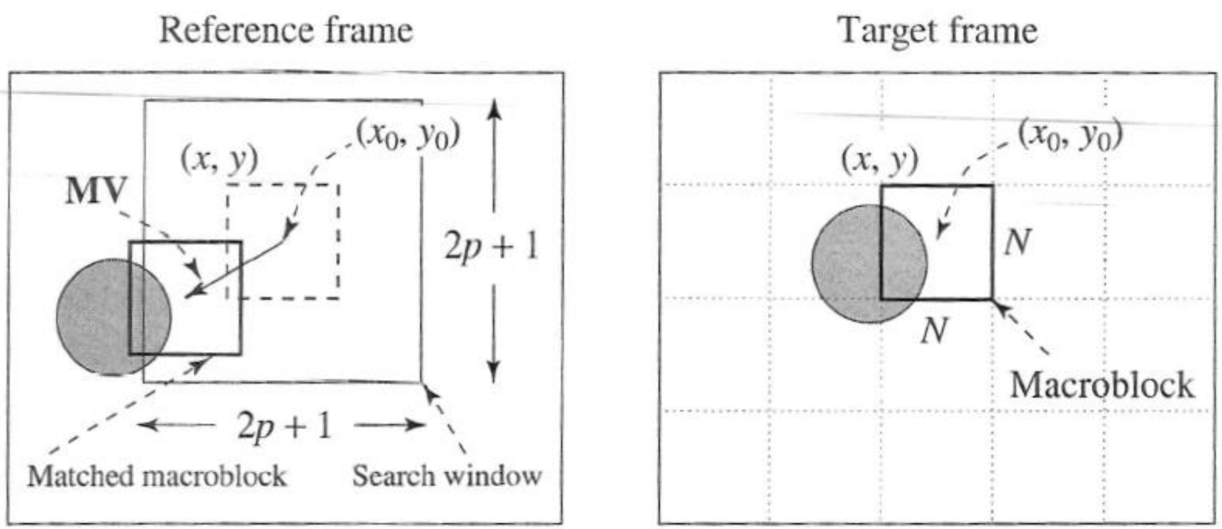
- The P or B frame being compressed is called the ***target frame***.
- The reference frame to which a P frame is compared is called its ***forward prediction frame***. The reference frame to which a B frame is compared is called its ***backward prediction frame.***
- Assume we have a macroblock in the target frame T. We will search for a matching macroblock in reference frame R. We want to look in the vicinity of $\boldsymbol{R}_{x,y}$ for the macroblock that most closely matches $\boldsymbol{T}_{x,y}$.

Intra-frame coding

- Various lossless and lossy compression techniques use – like JPEG

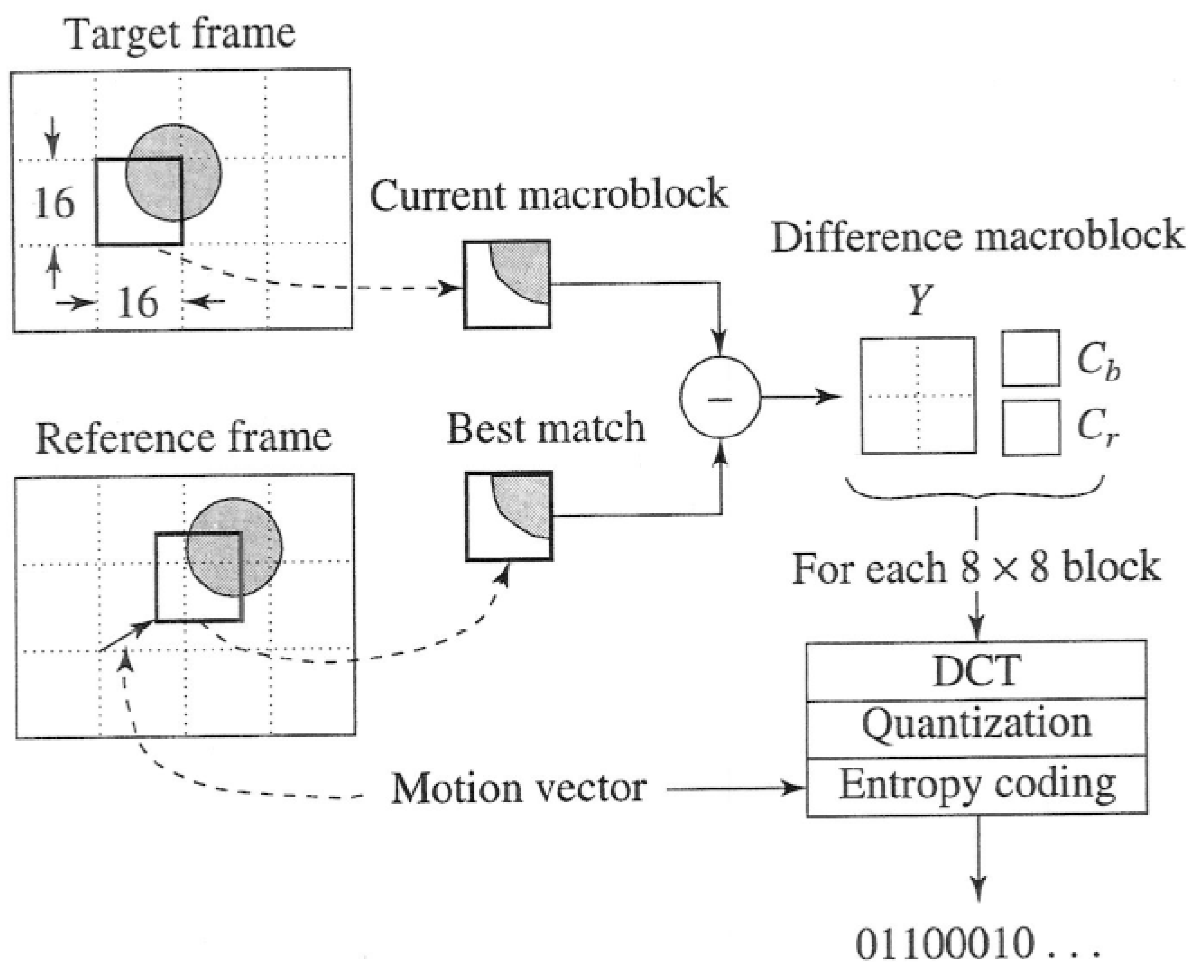


Motion Compensation



In H.261, motion vectors are in the range [-15,15]x[-15,15], e.g, p = 15.

MPEG Video for P-Frames



Estimate motion vectors

So how do we find the motion?

Basic Ideas is to search for Macroblock (MB)

- Within a $\pm n$ x m pixel search window
- Work out for each window
  Sum of Absolute Difference (SAD)
  (or Mean Absolute Error (MAE))
- Choose window where SAD/MAE is a **minimum**.

If the encoder decides that no acceptable match exists then it has the option of

- Coding that particular macroblock as an intra macroblock,
- Even though it may be in a P frame!
- In this manner, high quality video is maintained at a slight cost to coding efficiency.

Sum of Absolute Difference (SAD)

SAD is computed by:

$$SAD(i, j) = \sum_{k=0}^{N-1} \sum_{l=0}^{N-1} |C(x + k, y + l) - R(x + k + i, y + l + j)|$$

- $N =$ size of macroblock window typically (16 or 32 pixels),
- (x, y) the position of the original macroblock, C, and
- R is the **reference** region to compute the SAD.
- $C(x + k, y + l)$ — pixels in the macro block with upper left corner (x, y) in the **target**.
- $R(x + k + i, y + l + j)$ — pixels in the macro block with upper left corner $(x + i, y + j)$ in the **reference**.

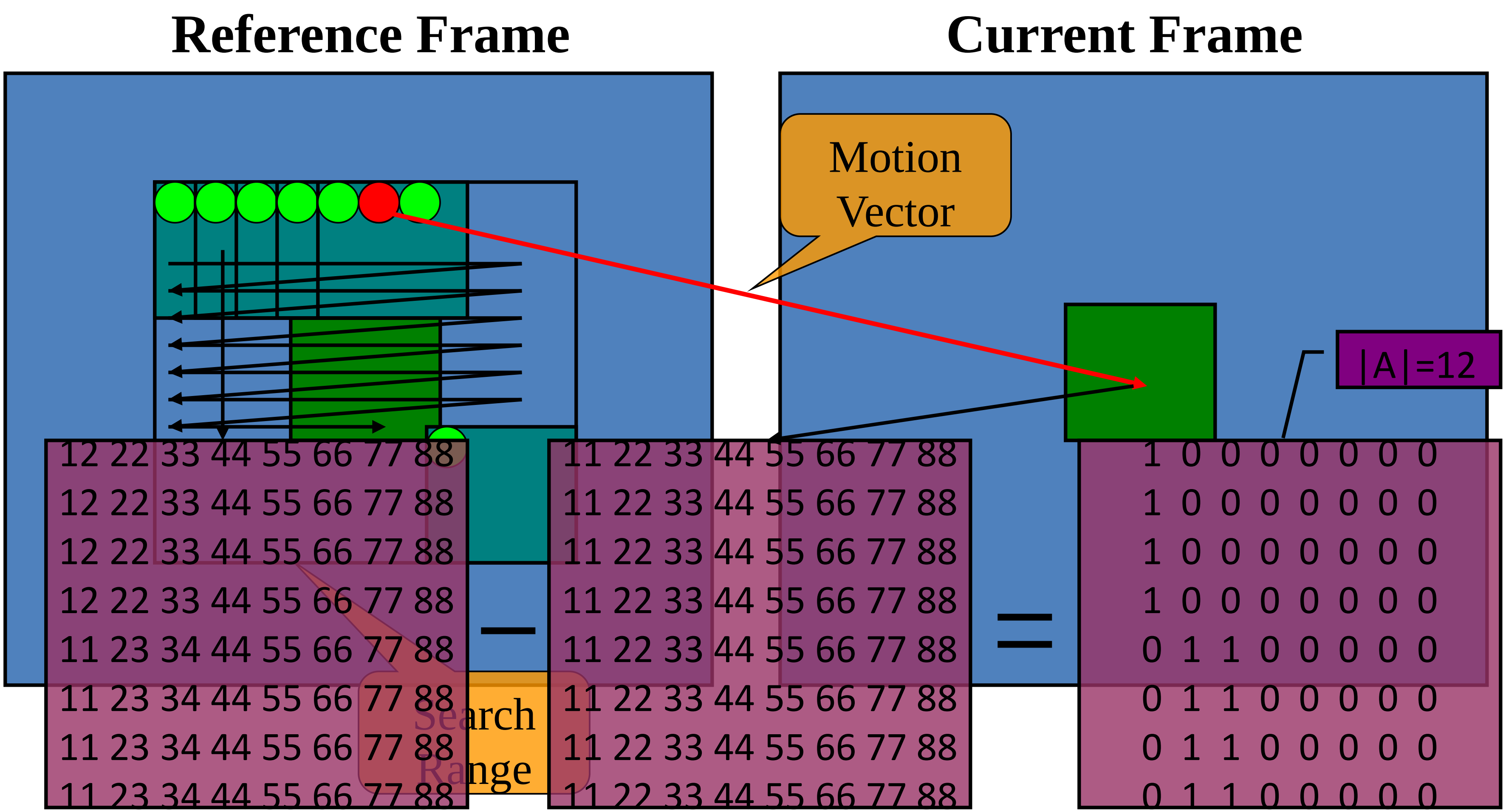
Sum Square Differences (SSD)

- **Alternatively:** sum of squared differences

$$SSD(i,j) = \sum_{k=0}^{N-1}\sum_{l=0}^{N-1}(C(x+k,y+l)-R(x+k+i,y+l+j))^2$$

- Goal is to find a vector (i,j) such that SAD/SSD (i,j) is minimum.

The Exhaustive Block-Matching Algorithm



12	22	33	44	55	66	77	88
12	22	33	44	55	66	77	88
12	22	33	44	55	66	77	88
12	22	33	44	55	66	77	88
11	23	34	44	55	66	77	88
11	23	34	44	55	66	77	88
11	23	34	44	55	66	77	88
11	23	34	44	55	66	77	88

−

11	22	33	44	55	66	77	88
11	22	33	44	55	66	77	88
11	22	33	44	55	66	77	88
11	22	33	44	55	66	77	88
11	22	33	44	55	66	77	88
11	22	33	44	55	66	77	88
11	22	33	44	55	66	77	88
11	22	33	44	55	66	77	88

=

1	0	0	0	0	0	0	0
1	0	0	0	0	0	0	0
1	0	0	0	0	0	0	0
1	0	0	0	0	0	0	0
0	1	1	0	0	0	0	0
0	1	1	0	0	0	0	0
0	1	1	0	0	0	0	0
0	1	1	0	0	0	0	0

Fast Block-Matching Algorithms

- **The characteristics of fast algorithm**
  - Not as accurate as exhaustive search algorithm
  - Save large amount of computation
- **fast motion estimation algorithm**
  - 2-D logarithm search method

2-D logarithm Search Method

- The search is accomplished by successively reducing the area of search.
- Each step consists of searching five locations which contain the center of the area, and the midpoints between the center and the four boundaries of the area along the axes passing through the center.
- In the final step all the nine locations are searched and the location corresponding to the minimum is the *direction of minimum distortion*.

J.R. Jain and A.K. Jain, "Displacement measurement and its application in interframe image coding," *IEEE Trans. Commun.*, Vol. COM-29, pp. 1799–1808, Dec. 1981

The algorithm is given below.

For any integer $m > 0$, we define

$$N(m) = \{(i, j); \quad -m \leqslant i, j \leqslant m\}$$

$$M(m) = \{(0, 0), (m, 0), (0, m), (-m, 0), (0, -m)\}.$$

A 2-D Logarithmic Search Procedure for DMD:

Step 1: (initialization)

Error function

$$D(i, j) = \infty \qquad (i, j) \notin N(p)$$

P is search range

$$n' = \lfloor \log_2 p \rfloor$$

$$n = \max \cdot \{2, 2^{n'-1}\}$$

$$q = l = 0 \text{ (or an initial guess for DMD)}$$

where $\lfloor \cdot \rfloor$ is a lower integer truncation function.

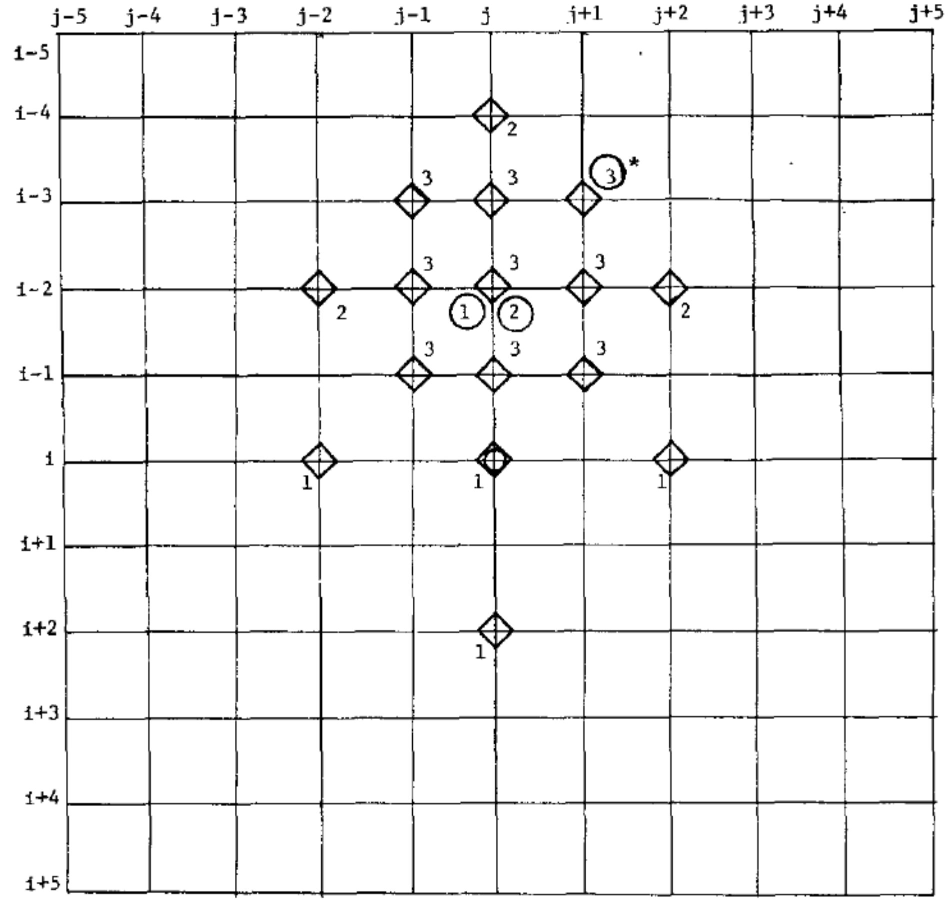
Step 2: $M'(n) \leftarrow M(n)$.

Step 3: Find $(i, j) \in M'(n)$ such that $D(i + q, j + l)$ is minimum. If $i = 0$ and $j = 0$, go to Step 5; otherwise go to Step 4.

Step 4: $q \leftarrow q + i, l \leftarrow l + j$; $M'(n) \leftarrow M'(n) - (-i, -j)$; go to Step 3.

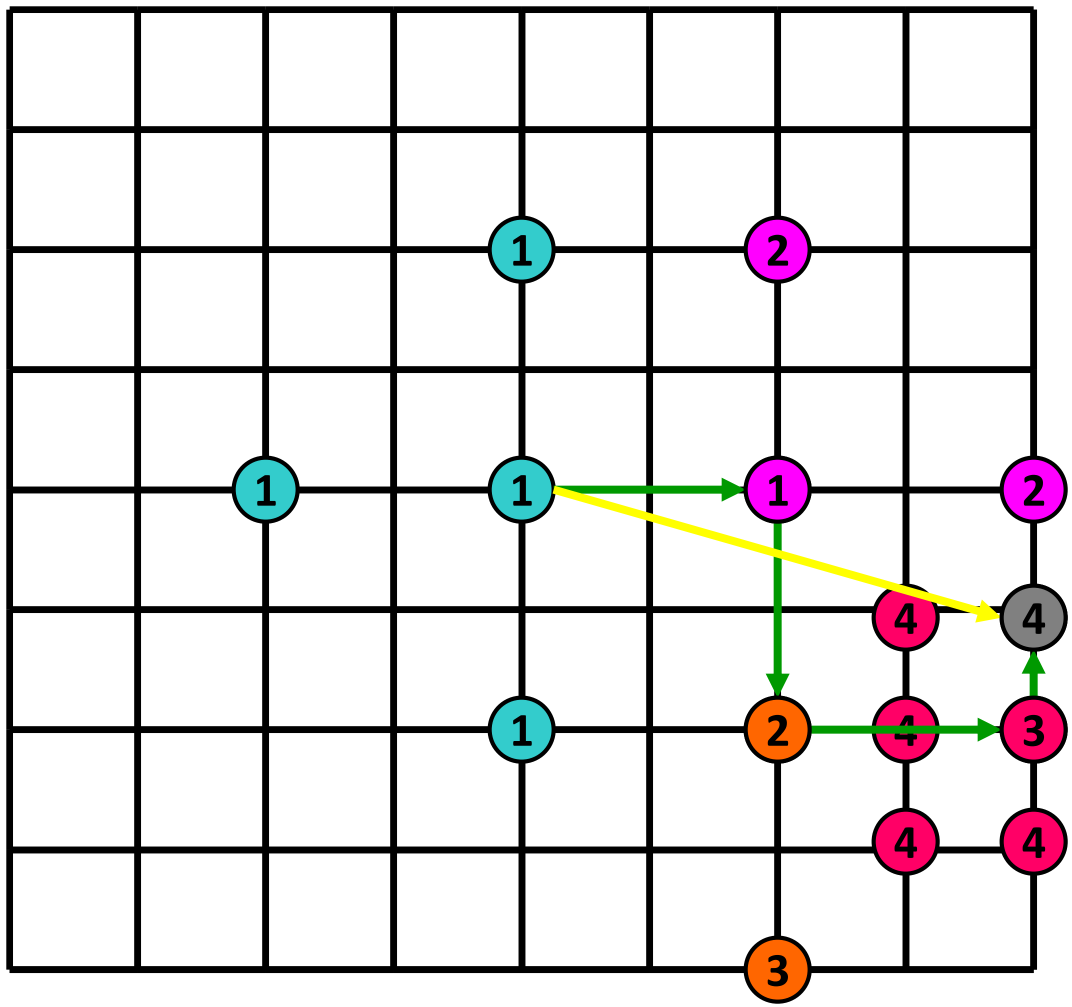
Step 5: $n \leftarrow n/2$. If $n = 1$, go to Step 6; otherwise, go to Step 2.

Step 6: Find $(i, j) \in N(1)$ such that $D(i + q, j + l)$ is minimum: $q \leftarrow i + q, l \leftarrow l + j$. (q, l) then gives the DMD.



P=5

2-D logarithm Search Method



The main steps of MPEG compression

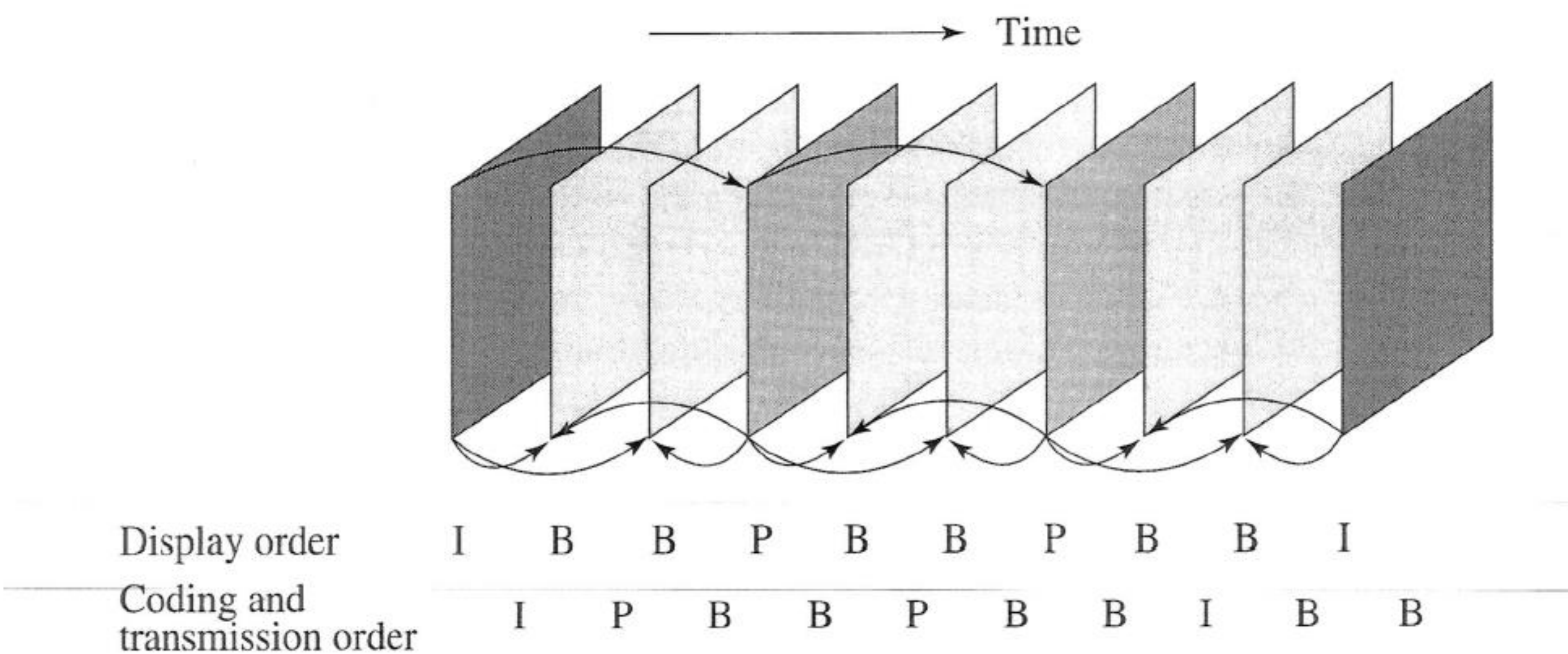
- **Step 5:** For all frames, compress with JPEG compression.
  - Compressing a frame of video is just like compressing a still image, and thus JPEG compression can be applied
  - **I frames** undergo intraframe compression only, without reference to any other image.
  - **P and B frames** first undergo motion prediction. Then the difference between the expected value of a pixel and its actual value is encoded.

MPEG-1 Video

- MPEG-1 was approved by ISO and IEC in 1991 for “Coding of Moving Pictures and Associated Audio for Digital Storage Media at up to about 1.5Mbps”.
- MPEG-1 standard is composed of
  - System
  - Video
  - Audio
  - Conformance
  - And Software
- MPEG-1’s video format is called SIF(Source Input Format)
  - 352x240 for NTSC at 30f/s
  - 352x288 for PAL at 25f/s

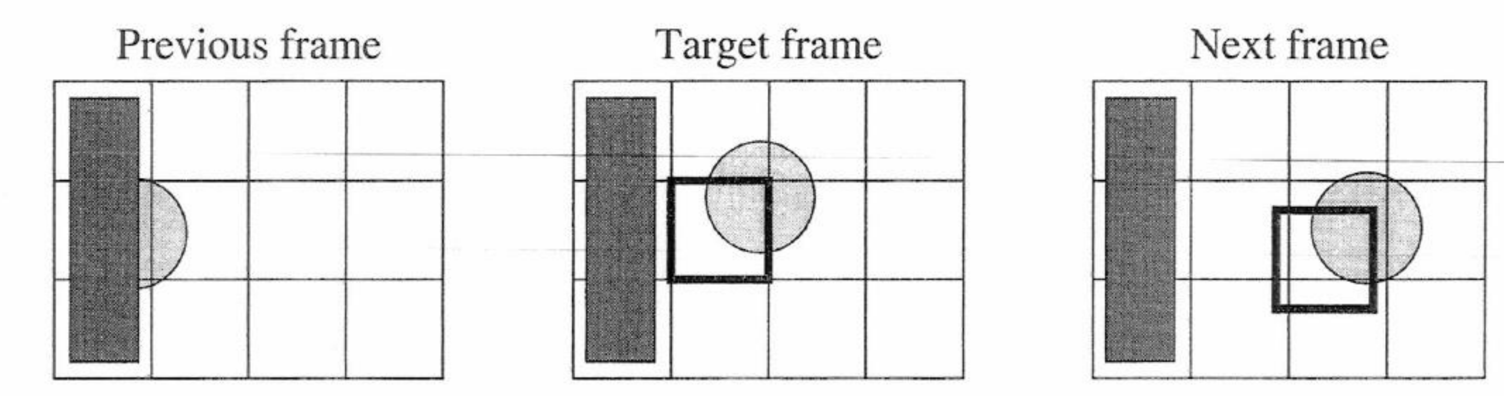
MPEG-1 Motion Compensation

- MPEG-1 introduces a new type of compressed



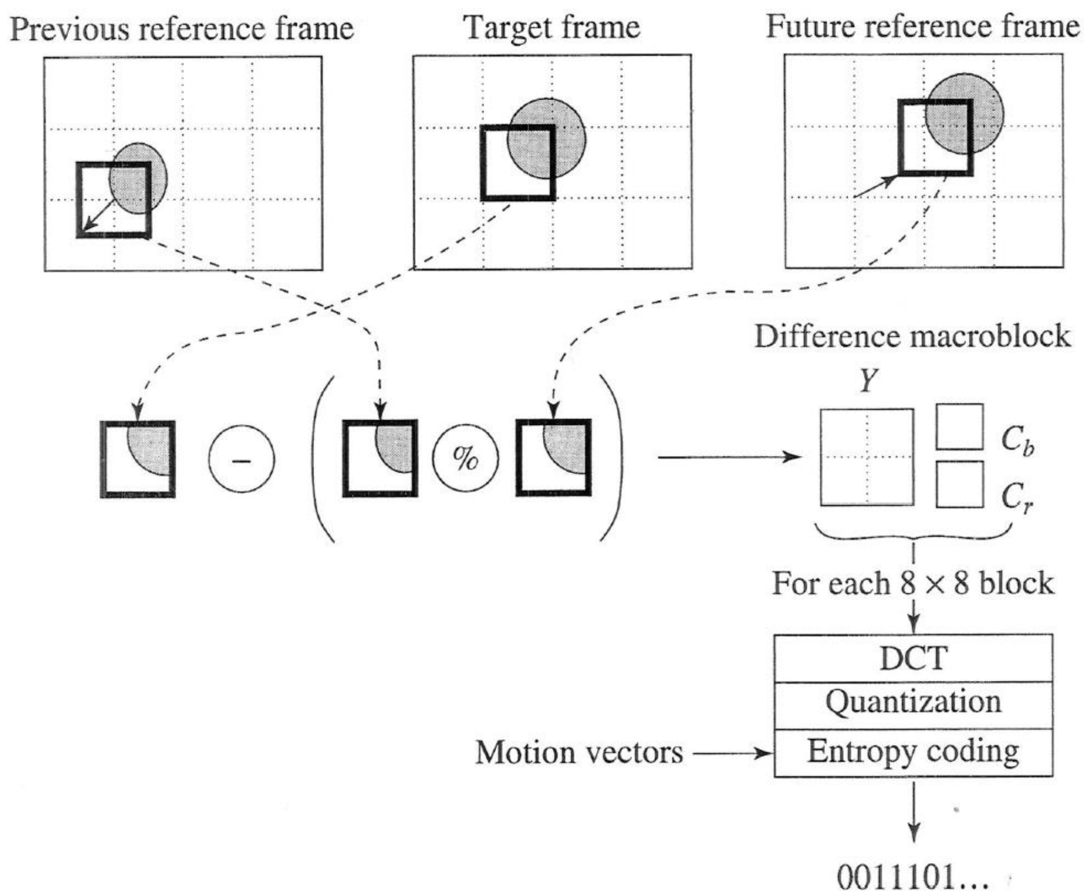
Why do we need B-frames?

- Bi-directional prediction works better than only using previous frames when occlusion occurs.

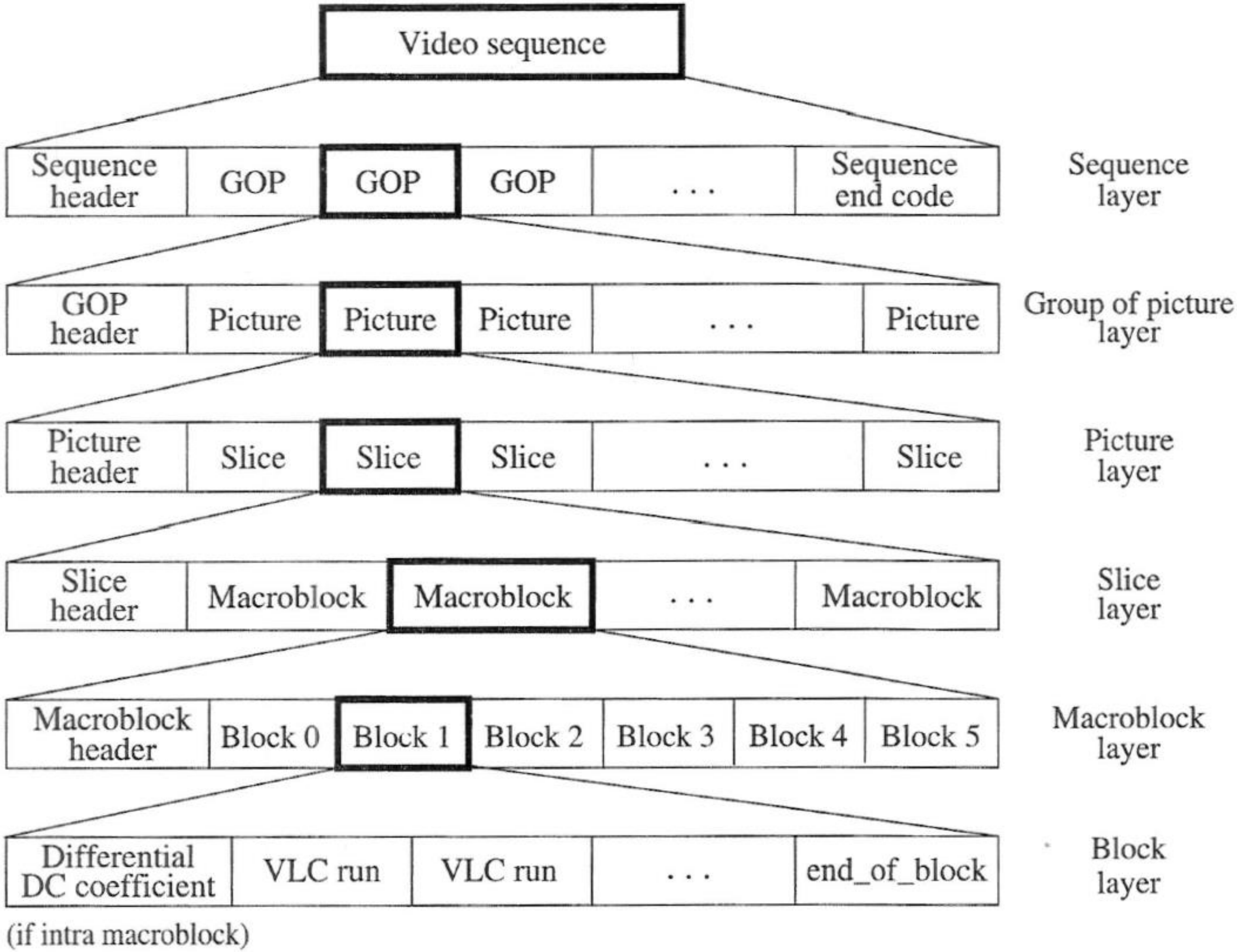


For this example, the prediction from next frame is used and the prediction from previous frame is not considered.

Compression of B-frames



MPEG-1 Video Stream



MPEG-2

- **MPEG-2** (aka H.222/H.262 as defined by the ITU) is a standard for "the generic coding of moving pictures and associated audio information“
- Backwards compatibility with existing hardware and software means it is still widely used, for example in the DVD-Video standard
- MPEG-2 evolved out of the shortcomings of MPEG-1

MPEG-2

- MPEG-2 profiles and levels:

Level	Simple profile	Main profile	SNR scalable profile	Spatially scalable profile	High profile	4:2:2 profile	Multiview profile
High		*			*		
High 1440		*		*	*		
Main	*	*	*		*	*	*
Low		*	*				

Four levels in the main profile of MPEG-2.

Level	Maximum resolution	Maximum fps	Maximum pixels/sec	Maximum coded data rate (Mbps)	Application
High	$1,920 \times 1,152$	60	62.7×10^6	80	Film production
High 1440	$1,440 \times 1,152$	60	47.0×10^6	60	Consumer HDTV
Main	720×576	30	10.4×10^6	15	Studio TV
Low	352×288	30	3.0×10^6	4	Consumer tape equivalent

Scalability

- SNR scalability
  - Base layer uses rough quantization, while enhancement layers encode the residue errors.
- Spatial scalability
  - Base layer encodes a small resolution video; enhancement layers encode the difference of bigger resolution video with the “un-sampled” lower resolution one.
- Temporal scalability
  - Base layer down-samples the video in time; enhancement layers include the rest of the frames.
- Hybrid scalability
- Data partitioning

MPEG-4

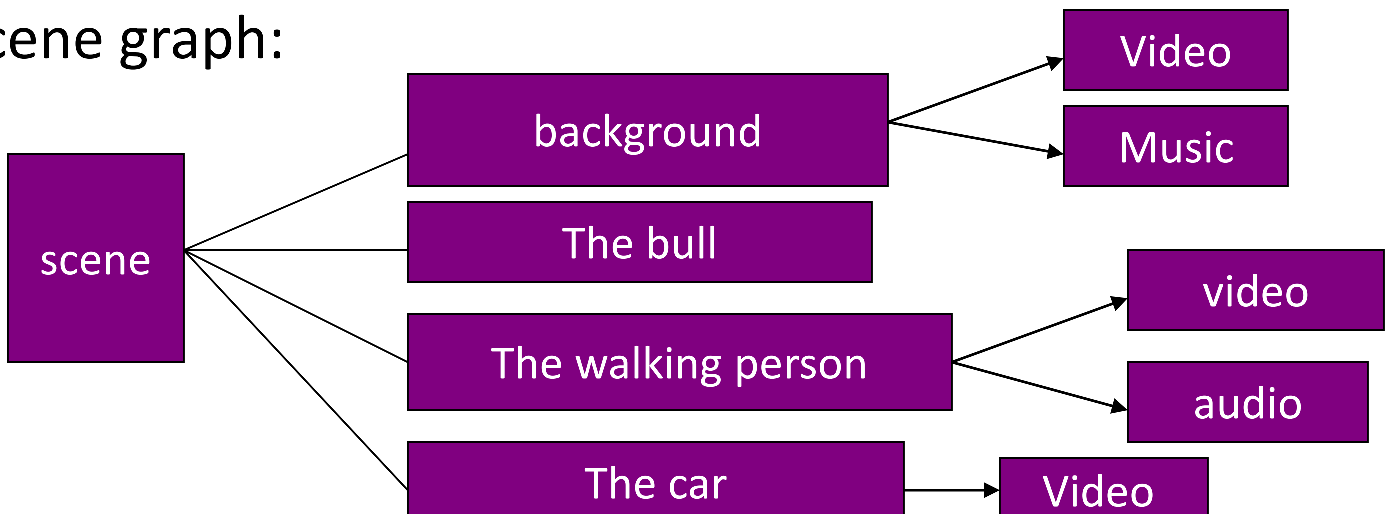
- Officially up to 10 Mbits/sec.
- Improved encoding efficiency.
- Content-based interactivity.
- Content-based and temporal random access.
- Integration of both natural and synthetic objects.
- Temporal, spatial, quality and object-based scalability.
- Improved error resilience.
- Support object-based features for content
- Enable dynamic rendering of content
  - defer composition until decoding
- Support convergence among digital video, synthetic environments, and the Internet

Audio-Video Object

- MPEG4 is based on the concept of media objects.

Audio Video Objects

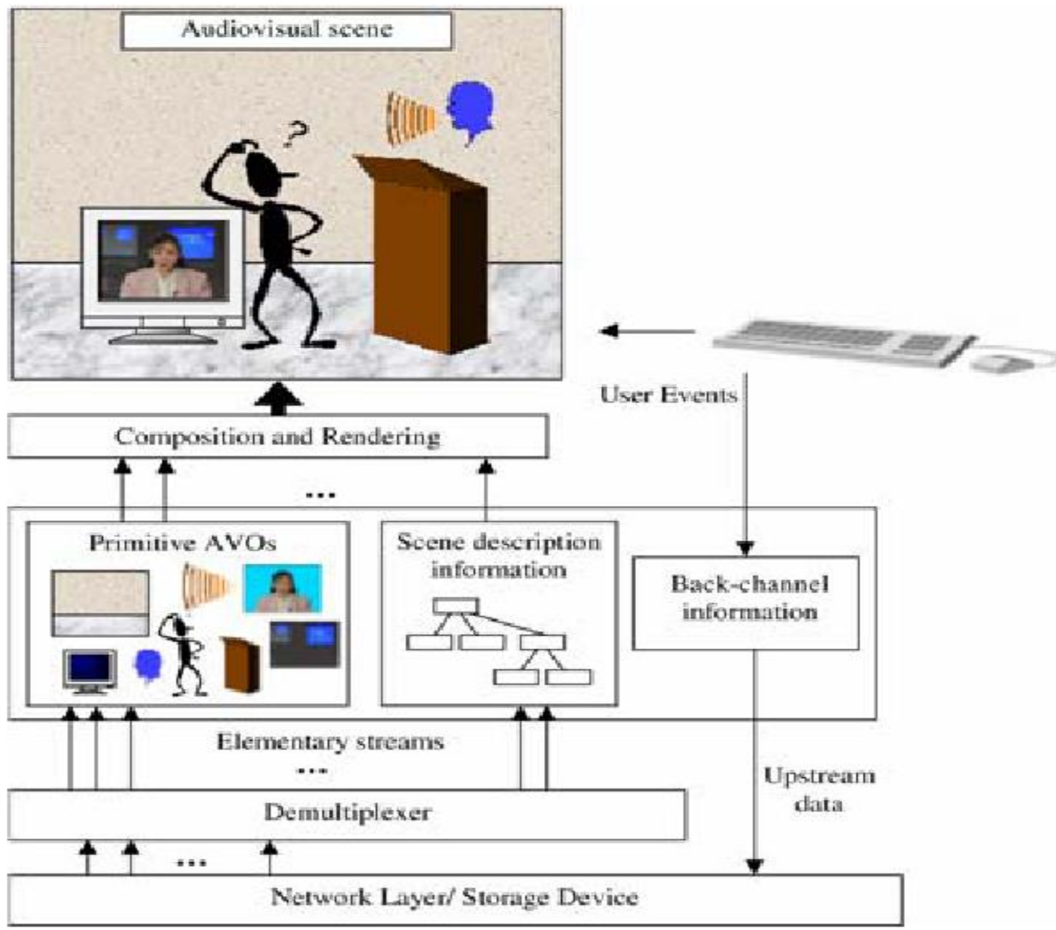
- A media object in MPEG4 could be
  - A video of an object with “shape”.
  - The speech of a person.
  - A piece of music.
  - A static picture.
  - A synthetic 3D cartoon figure.
- In MPEG4, a scene is composed of media objects based on a scene graph:



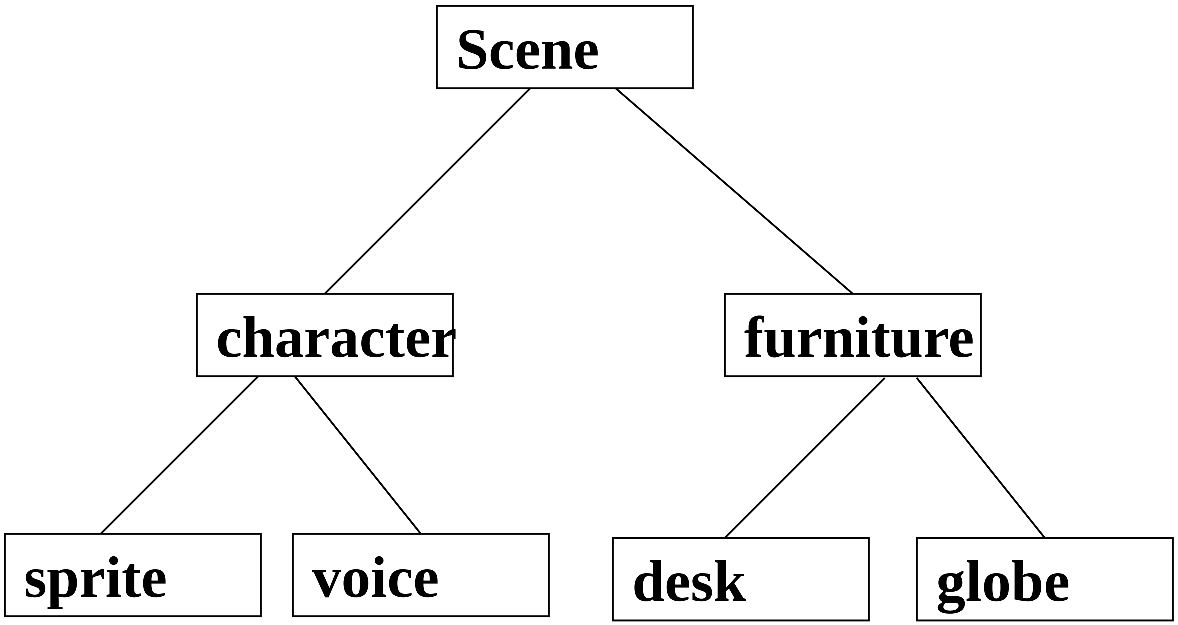
MPEG-4 Standard

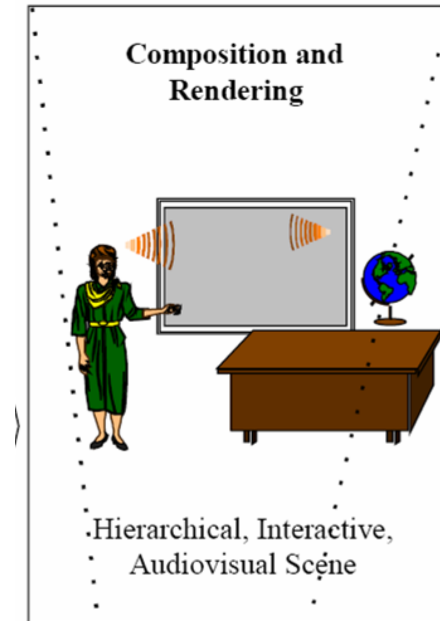
- Defines the scheme of encoding audio and video objects
  - Encoding of shaped video objects.
  - Sprite encoding.
  - Encoding of synthesized 2D and 3D objects.
- Defines the scheme of decoding media objects.
- Defines the composition and synchronization scheme.
- Defines how media objects interact with users.

Composition and Interaction



Composition



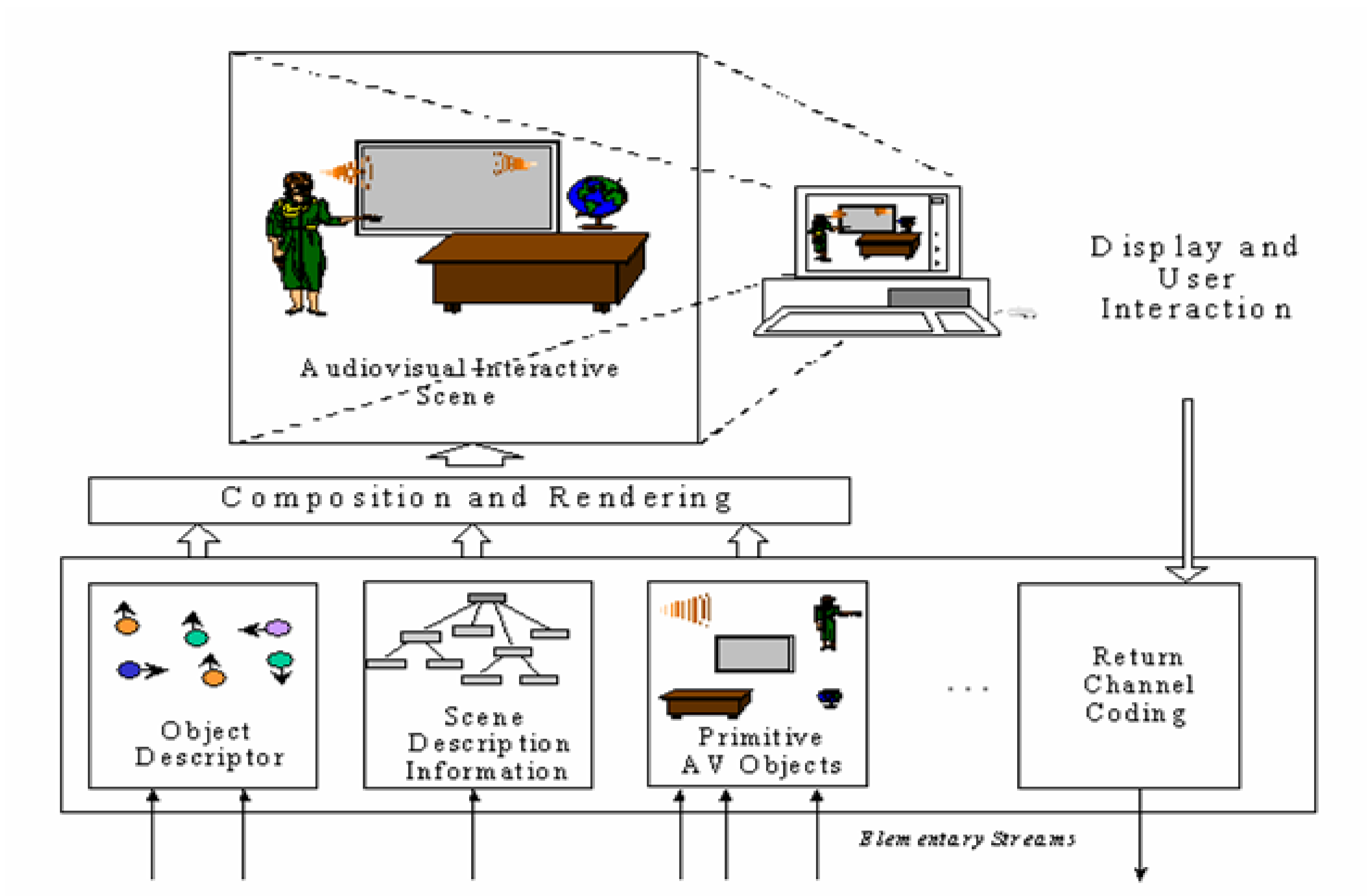


ISO N3536 MPEG4

Composition (cont.)

- Encode objects in separate channels
  - encode using most efficient mechanism
  - transmit each object in a separate *stream*
- Composition takes place at the decoder, rather than at the encoder
  - requires a binary scene description (BIFS)
- BIFS is low-level language for describing:
  - hierarchical, spatial, and temporal relations

MPEG-4 Rendering



Interaction as Objects

- Change colors of objects
- Toggle visibility of objects
- Navigate to different content sections
- Select from multiple camera views
  - change current camera angle
- Standardizes content and interaction
  - e.g., broadcast HDTV and stored DVD

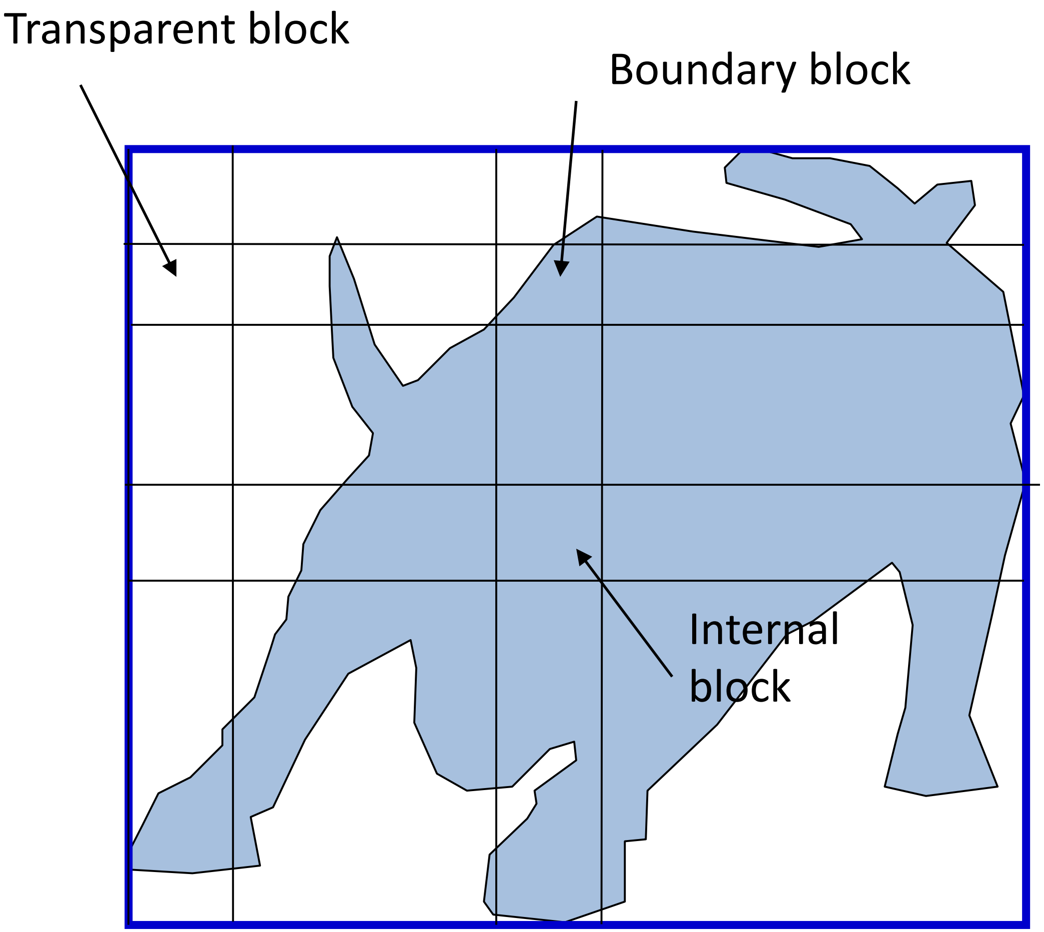
Video Coding in MPEG-4

- Support for 4 types of video coding:
  - Video Object Coding
    - For coding of natural and /or synthetic originated, rectangular or arbitrary shaped video objects.
  - Mesh Object Coding
    - For visual objects represented with a mesh structure.
  - Model-based Coding
    - For coding of a synthetic representation and animation of a human face and body.
  - Still Texture Coding
    - For wavelet coding of still textures.

Video Object Coding

- Video Object (VO)
  - Arbitrarily shaped video segment that has a semantic meaning.
- Video Object Plane (VOP)
  - 2D snapshot of a VO at a particular time instance.
- Coding of VOs: 3 "elements"
  - Shape
    - Rectangularly shaped VO.
    - Arbitrarily shaped VO.
  - Motion
  - Texture

Shape Coding

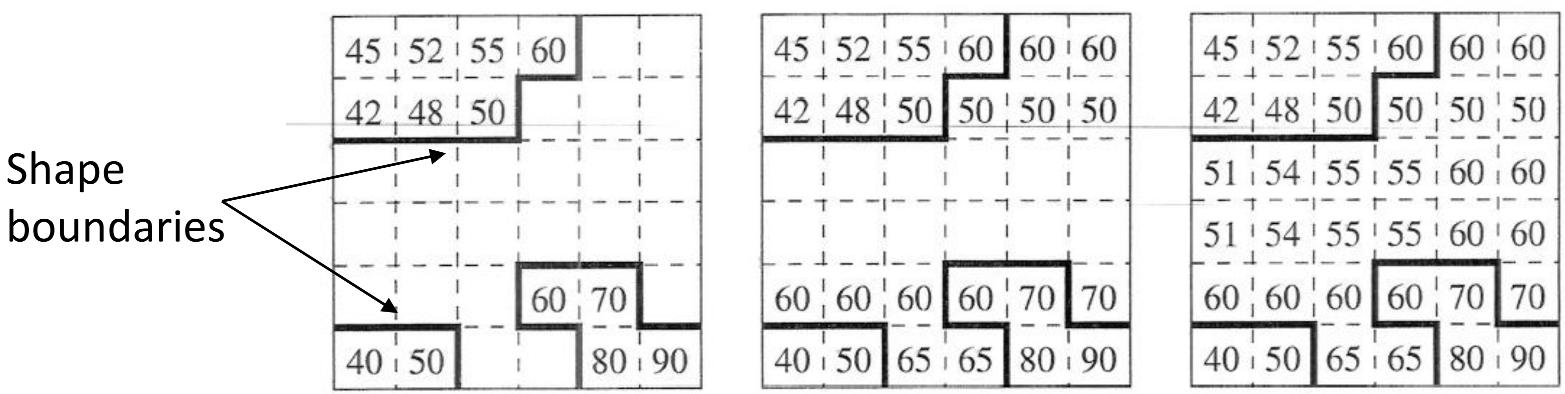


Shape coding:

- Bitmap image of a shape – alpha plane
    - Binary alpha plane.
    - Grayscale alpha plane.
- Binary alpha plane – shape information only.
- Grayscale alpha plane – shape and transparency information.
- Inter and Intra coding for the binary shapes.

Motion Compensation

- We have to deal with shaped objects.
- Motion estimation for internal blocks uses similar schemes as MPEG-1 and 2.
- For the boundary blocks, we first do “padding”, and then do motion estimation and compensation.

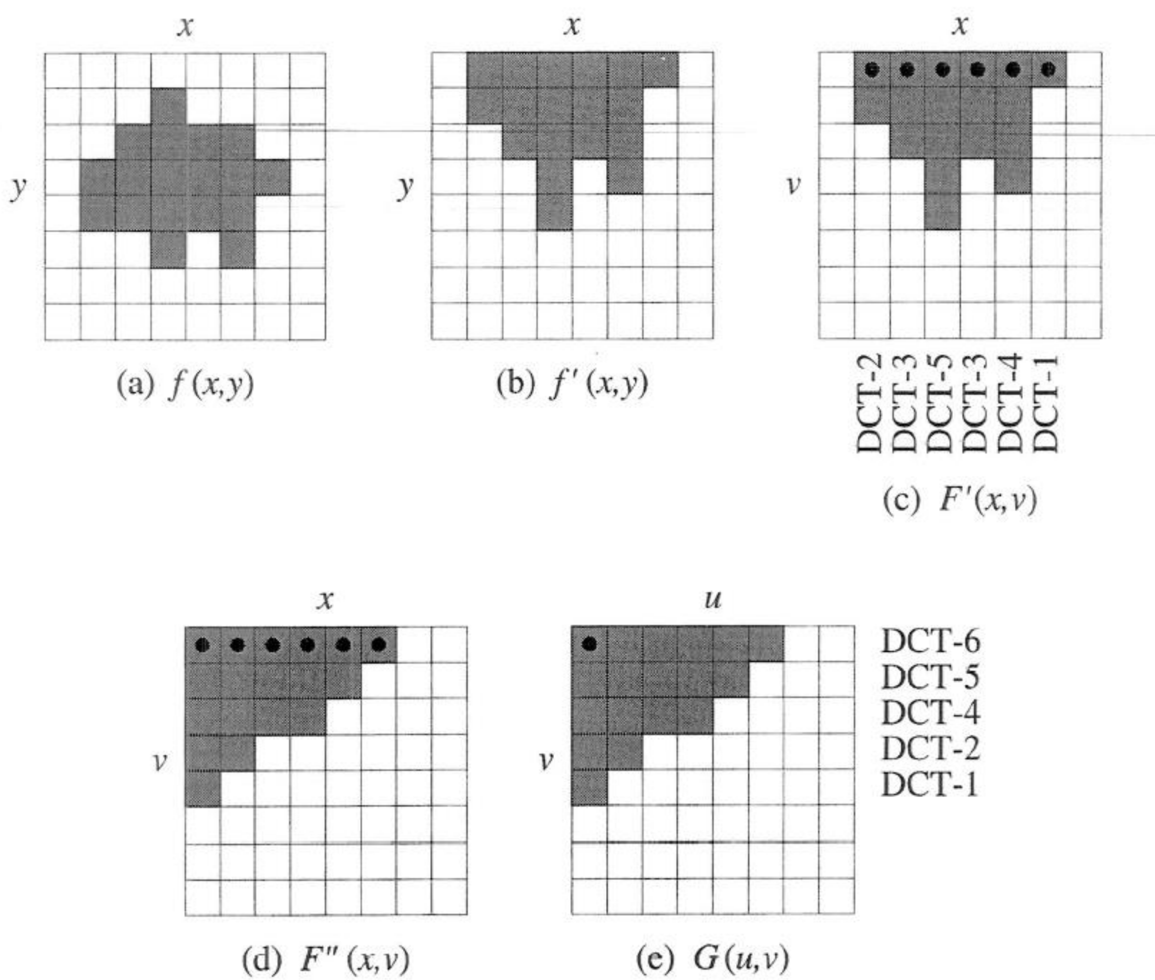


(a)

(b) Horizontal Padding

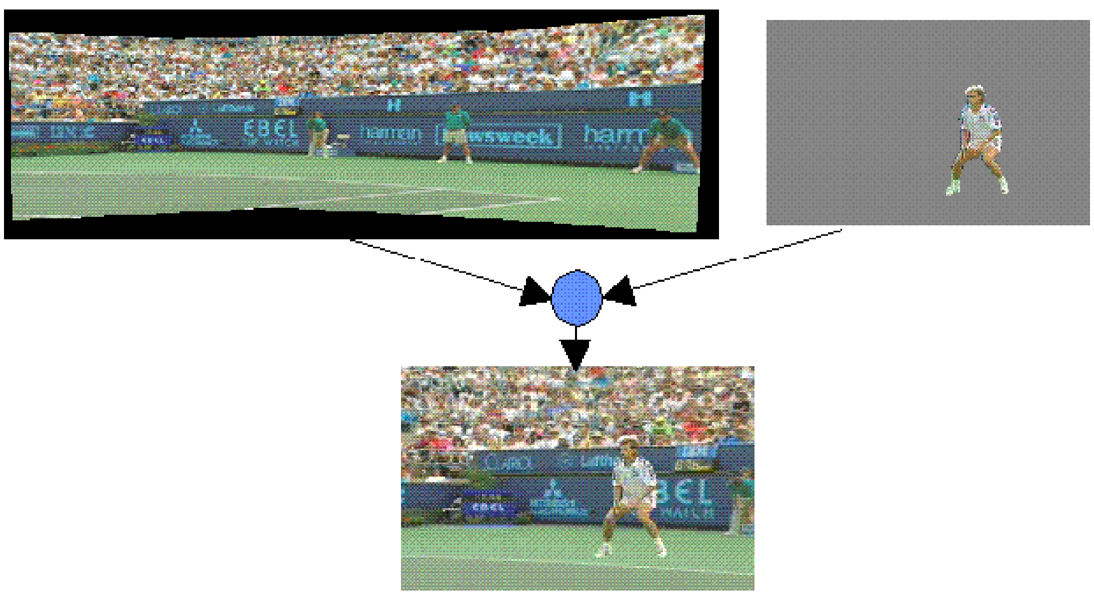
(c) Vertical Padding

Shape Adaptive DCT in Texture Coding



Sprite Coding

- Sprite coding is use for encoding a scene with large static background with small foreground objects.
- Background is coded only once at the beginning of the sequence as an Intra-VOP.
- It uses global motion parameters to manipulate the background.

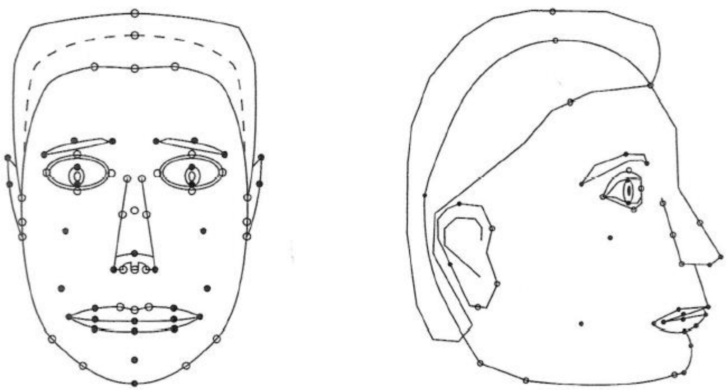


Mesh Coding

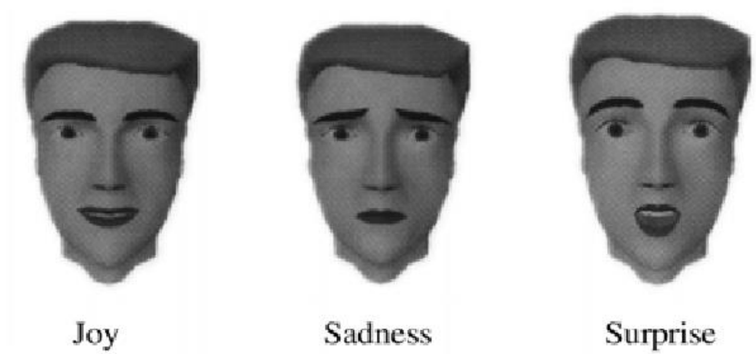
- Mesh
  - Partitioning of an image into polygonal patches.
- MPEG-4 supports 2D meshes
  with triangular patches.
- Benefits of using mesh coding
  - Easy to manipulate an object.
  - Easy to track the motion of a video object after it has been encoded.
- Superior compression

Model Based Coding

- MPEG-4 supports 2 types of models
  - Face object model
    - Synthetic representation of the human face with 3D polygon meshes that can be animated.



The face model



  - Body object model
    - Synthetic representation of a human body with 3D polygon meshes that can be rendered to simulate body movement.

MPEG-4 Conclusion

- A lot of MPEG-4 examples with interactive capabilities
- Content-based Interactivity
  - Scalability
    - Spatial Scalability
    - Temporal Scalability
  - Sprite Coding
- Improved Compression Efficiency
- Universal Accessibility
  - re-synchronization
  - data recovery
  - error concealment

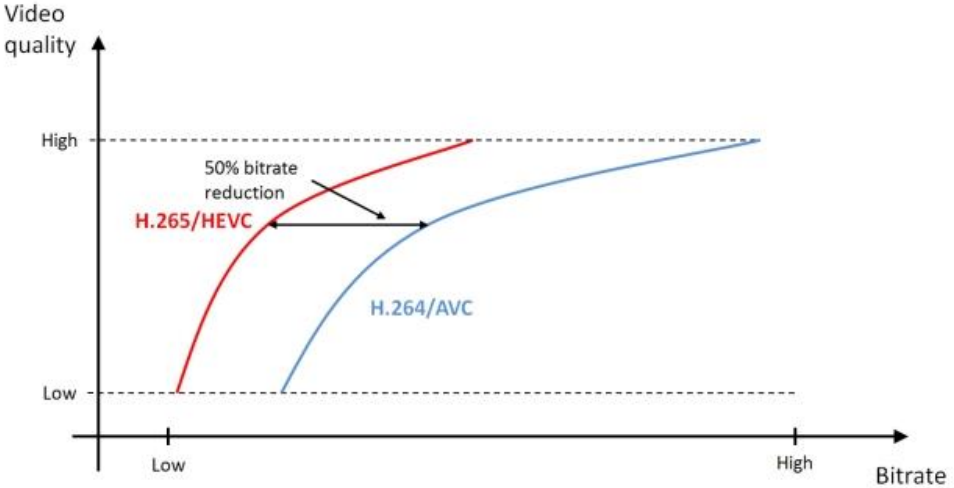
High Efficiency Video Coding (HEVC)

High Efficiency Video Coding (HEVC) is the latest generation video compression standard.

- This standard was developed by the **ISO/IEC** MPEG and **ITU-T** VCEG, through their JCT-VC

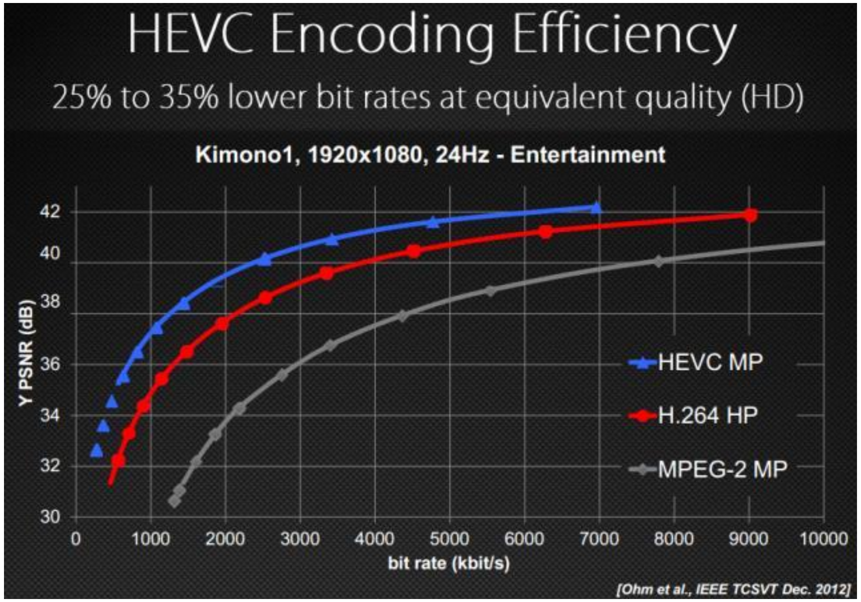
- HEVC is also known as ISO/IEC 23008-2 MPEG-H Part 2 and ITU-T **H.265**

- Have a bit rate reduction of 50% at the same subjective image quality compared to the H.264/MPEG-4 AVC High profile



High Efficiency Video Coding (HEVC)

- It can support 8K UHD and resolutions up to 8192×4320
- HEVC is said to double the data compression ratio compared to H.264/MPEG-4 AVC at the same level of video quality



Subjective video performance comparison[61]

Video coding standard	Average bit rate reduction compared to H.264/MPEG-4 AVC HP			
	480p	720p	1080p	4K UHD
HEVC	52%	56%	62%	64%

H.265/HEVC/MPEG-H Part 2

- Main drivers
  - Get **Low bitrate target – target 2:1** over H.264
  - Cheat your eyes – how much can you cut bits and still see the same quality
  - Improve **resolutions (8K by 4K and 4K by 2K) and frame rates**
  - Launch **1080p50/60 services** to compete against BluRay
  - Expect <10x more computational complexity and 2x-3x (decode)

H.265

- Derived from H.264
  - More modes, tools and more interdependencies
  - More efficient search algorithms
  - More complex intra-prediction
  - Macroblocks vs Partitions

- AVC
- 16x16 macro-blocks
- 8x8 and 4x4 transform sizes

- HEVC
- Coding unit size 64x64 to 8x8
- 32x32, 16x16, 8x8 and 4x4 transform sizes

Summary

- Standard formats for analog and digital video
- Main properties of analog and digital video
- Deinterlacing
- Different kinds of codecs
- MPEG compression
- Motion estimation
- Main features of MPEG-4
- H.265