Chapter 6 & Chapter 7
Digital Video

CS3570

Department of Computer Science
National Tsing Hua University

Film, Television, Movies, and Video

- Video is fascinating, and it is complicated.
- To understand the science of digital video, you need to understand sampling and resolution, mathematical filters and transforms, color models, and compression—topics covered in the first five chapters of this book
- This chapter includes information about analog video, television technology, and film—areas that are not strictly digital video, but that are such close relatives

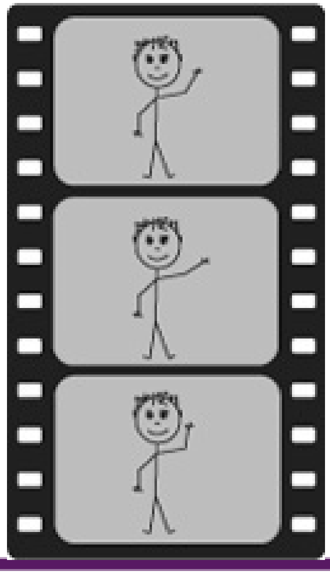
Video, Film, and Television Compared

- ***Movie*** : a story told with moving images and sound
- The word ***motion picture*** and ***movie*** are the same thing
- 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

Frame rate

- Film and television create moving pictures—by a fast sequence of images, called **frames**
- The speed at which images are shown is the ***frame rate***
- a frame rate of about 40 frames per second is needed in order for successive images to be perceived as smooth motion with no flicker
- Sprocket holes—also called ***perforations***—are holes on the sides of the film used to pull the film through the projector

4-perf 35mm film



Standard film aspect ratios

- Silent movies and early sound movies were shot mostly on 16 mm film, introduced by Eastman Kodak in 1923.
- ***Aspect ratio*** is the ratio of the width to the height of a frame, expressed as *width:height*.
- IMAX movies are shot on 70 mm film with aspect ratio of 1.43:1
- IMAX movies are on very large screens, so the frames have to be enlarged more than they are in standard movie projection

Academy aspect ratio 1.33:1

Widescreen, 1.85:1

Anamorphic widescreen, 2.39:1

Standard-definition television

- In the beginning, television was transmitted as an analog signal. In comparison to the newer HDTV, we now sometimes refer to this as ***SDTV*** (***standard-definition television***).
- SDTV was broadcast through radio waves by land-based broadcast stations
- Direct Broadcast Satellite (DBS) is received directly in the home, which must be equipped with a satellite dish

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 × 1080 using interlaced scanning(**1080i**), 1920 × 1080 using progressive scanning(**1080p**), or 1280 × 720 using progressive scanning (**720p**)
- Digital encoding is not part of this definition, and, historically, HDTV was not always digital

Video Standards

- Three main standards emerged in the early days of analog television
  - ***NTSC*** (developed by the ***National Television Systems Committee***)
  - ***PAL*** (***Phase Alternating Line***)
  - ***SECAM*** (***Système Electronique Couleur Avec Mémoire***)
- These began as analog standards that have evolved to cover digital video as well.

Video Standards

- NTSC governs standards in North America, Japan, **Taiwan**, and parts of the Caribbean and South America
- NTSC was instrumental in helping the television industry move from monochrome transmission to color
- In 1967 PAL was adopted for color television broadcasts in the United Kingdom and Germany.
- PAL has a number of variants that are now used in Europe, Australia
- SECAM was developed in France and accepted for color broadcasting in 1967.
- It was later adopted by other countries in Eastern Europe.

Video standards

- The ***ITU*** (***International Telecommunication Union***) is an organization within the United Nations System to develop standards for global telecommunications
  - ITU-R (radio communication)
  - ITU-T (telecom standardization)
  - ITU-D (telecom development)
  - ITU-R BT.601, which gives specifications for storage and communication
  - ITU-R BT.709, developed for HDTV

Digital television (DTV)

- In the 1990s, the development of international standards for the transmission of ***digital television*** (***DTV***) became a hot topic
- Three main standards organizations for DTV

	ATSC	DVB	ISDB
Origin	United States	Europe	Japan
video compression		MPEG-2 main profile	
audio compression	Dolby AC-3	MPEG-2 or Dolby AC-3	MPEG-2 AAC
transmission type	8-vestigial sideband	COFDM (coded orthogonal frequency division multiplexing)	bandwidth segmented transmission of COFDM
bit rate	19.4 Mb/s	3.7–31.7 Mb/s	4–21.5 Mb/s

Standards for DTV

- ***ATSC*** (***Advanced Television Systems Committee***) is an international nonprofit organization that develops standards for digital television
- ATSC developed DTV standards for the United States and Canada (**Taiwan** and south Korea have been adopted the standards)
- In Europe, standards for digital television were developed by ***DVB*** (***Digital Video Broadcasting Project***).
- DVB standards are divided into terrestrial (DVB-T), satellite (DVB-S), and handheld (DVB-H).
- Standards for digital video in Japan go by the name of ***ISDB*** (***Integrated Services Digital Broadcasting***).

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

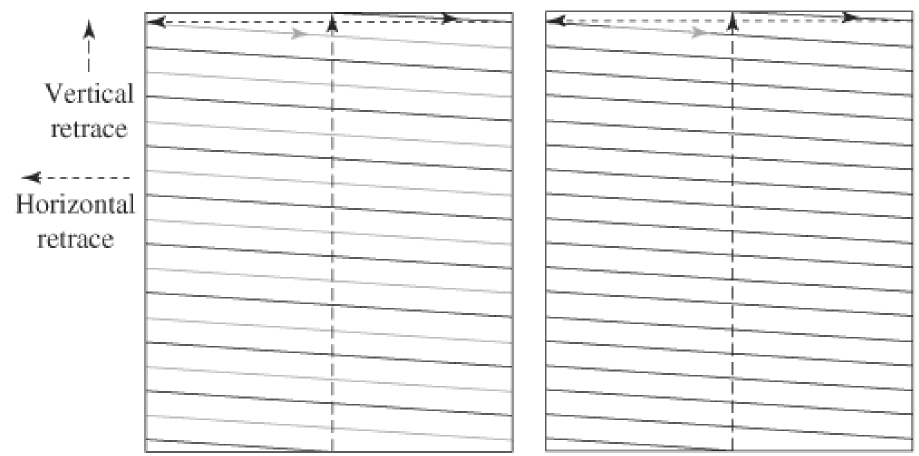
- The scanning process is a movement from left to right and top to bottom.
- When the scanner has finished with one line, it moves back to the left to start another in a motion called ***horizontal retrace***.
- ***Vertical retrace*** takes the scanner from the bottom of the monitor to the top again.
- 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

Raster 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 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

Native resolution

- A 720p television doesn't necessarily have 1280×720 pixels.
- It accepts a signal with 1280×720 pixels per frame and displays them with progressive scanning.It may in fact have a different ***native resolution***.
- For each frame, the ***logical pixels***—pieces of information saved and transmitted in a video signal—have to be mapped to the ***physical pixels***—points of light on the video display

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

Video connections

- In analog video, the color information can be sent in one of three types of *analog* ***video transmission formats***—component, S-video, or composite form

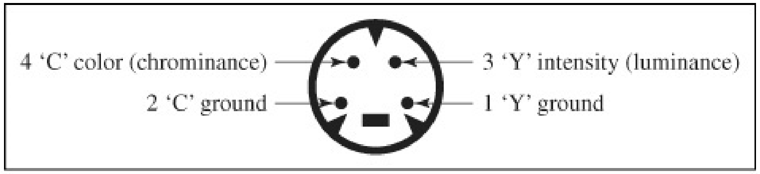
TABLE 6.3 Standards for Analog Video Recording Equipment

Video Format	Year Introduced	Color Transmission Format	Horizontal Resolution	Tape Width	Quality
VHS	1976	composite	~240	½" (12.5 mm)	consumer
Betamax	1976	composite	~240	½" (12.5 mm)	consumer
8mm (Video 8)	1984	composite	~240–300	8 mm	consumer
S-VHS	1987	S-video	~400–425	½" (12.5 mm)	high-end consumer
Hi-8	1998	S-video	~400–425	8 mm	high-end consumer
U-Matic	1971	composite	~250–340	¾" (18.75 mm)	professional
M-II	1986	component	~400–440	½" (12.5 mm)	professional
Betacam	1982	component	~300–320	½" (12.5 mm)	consumer
Betacam SP	1986	component	~340–360	½" (12.5 mm)	professional

video transmission formats

- In ***component video*** a separate signal is sent for each part of the three luminance/chrominance components
- Component video has three separate paths for the information and three connectors at the end.

- ***S-video*** uses two data paths: one for the luminance and one for the two chrominance
- An S-video jack has one connection at the display end, with two channels of information carried through the connection



video transmission formats

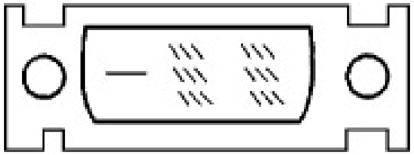
- ***Composite video*** is a video signal that is sent on just one channel.

- Compositing the signal makes it possible to use just one broadcast channel through the airways or one physical connection from device to device
  - Disadvantage to this technology is that crosstalk can occur between the color and luminance components, making composite video the lowest quality of all the alternatives

Digital video transmission format

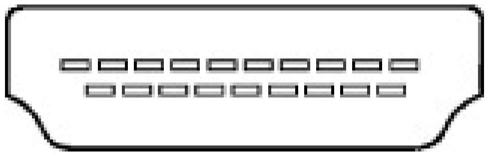
- There are two main types of digital video transmission format: ***DVI (digital video interface)****,* and ***HDMI (high definition multimedia interface)***
- *DVI* connects an uncompressed digital video source (e.g., from a video card) to a digital display device
- There are three basic DVI formats:
  - 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

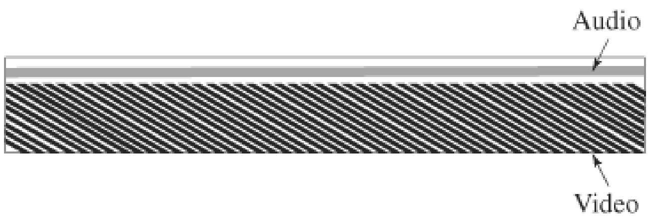
- *HDMI* is an audio/video connection for transmitting uncompressed digital data.
- It is backward compatible with DVI and accommodates audio data on the signal
- HDMI connections can apply ***HDCP (high bandwidth digital content protection)*** to signal
    - HDCP is a digital copy protection protocol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



HDMI interface

Videotape

- Videotape is different from film. Instead of recording a whole frame in a rectangle, a video camera records an image line-by-line, on a magnetized piece of plastic
- The audio track lies in a straight line along the edge, and the video information is written diagonally on the tape



- The number of horizontal lines in a frame is called the ***vertical resolution*** of an image

Digital Video Cameras

- Like analog video cameras, digital video cameras move across an image line-by-line, detecting light coming in through the lens
- The ***DV*** standard was released in 1999 in a document known as the ***Blue Book***, which is now called IEC 61834.
- New digital video standards have continued to emerge, especially with the advent of **HDTV**
- The first of the HDTV videotape standards was D6, an uncompressed format that uses 4: 2: 2 chrominance sub-sampling to achieve a bit rate of about 1.2 Gb/s

Digital Video Cameras

- **HDV** is a high-definition format that is affordable at the consumer level.
- JVC and Sony launched this format, and now Canon, Sharp, and Panasonic make HDV cameras as well
- Data rates for HDV are 19.7 Mb/s for 720p and 25 Mb/s for 1080i.
- HDV advertised as 1080i does not have the full 1920 vertical resolution; it's actually 1440×1080. There is currently no 1080p in the HDV format

Digital Video Cameras

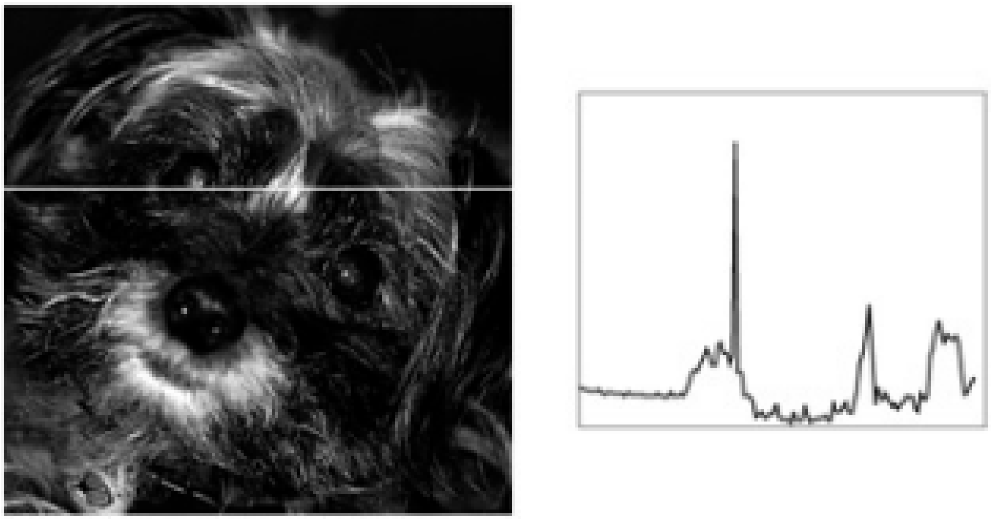
- HDV achieves the higher data rate by using MPEG-2 compression. HDV uses the same type of tapes as are used for the popular mini-DV cameras
- **AVCHD** became popular with Sony and Panasonic around 2006.
- First and second generation AVCHD cameras were 1440×1080; the third generation were 1920×1080

History of HDTV digital videotape

Type	Year	Maker	Compression	Approximate Video Bit Rate in Mb/s	Chrominance Subsampling	Tape Size
D5 HD	1994	Panasonic	motion-JPEG, 5:1	270	4:2:2	½"
D6	1995		none	1188	4:2:2	¾"
D9 (Digital S)	1996	JVC	MPEG-2	50	4:2:2	½"
D11 HDCAM	1997	Sony	M-JPEG-type	140	3:1:1	½"
D12 DVCPRO HD	2000	Panasonic	combination of codecs	100	4:2:2	½"
HDCAM SR	2003	Sony	MPEG-4 studio profile	440 or 880	4:4:4	½"
HDV	2003	JVC, Sony, Canon, Sharp	MPEG-2	17 or 25	4:2:0	¼"
XDCAM	2003	Sony	MPEG-2, DV25, and MPEG4	18, 35, or 50	4:2:0	optical disk
AVCHD	2006	Sony, Panasonic, Canon	MPEG-4 H.264	12–24	4:2:0	DVD disk, hard disk, memory stick, or internal flash
XDCAM EX	2007	Sony	MPEG-2 long GOP	25 or 35	4:2:2	SxS memory card
Red 3K Scarlet, 4K, One, 5K Epic	2006–2008	Red	uncompressed or compressed as Redcode raw	28–36 MB/s	4:2:2	hard disk or flash memory

Analog Video Resolution and Bandwidth

- A line from an image can be represented by a waveform where the amplitude of the wave changes in relation to how the colors of the image change across the line



Line of a grayscale image as an analog waveform

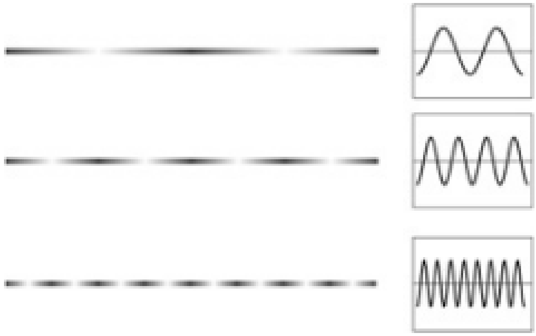
Analog Video Resolution and Bandwidth

- In digital video, discrete pixel positions across one line of the image are sampled and encoded in 0s and 1s.
- Video is a time-dependent medium. Thus, the vertical resolution of video has to be set at a specific value, both in number of lines per frame and in the time it takes for one line to be transmitted.
- The digital video signal itself has a specific resolution in logical pixels in the horizontal and vertical directions
- Analog video, in contrast, has no precise horizontal resolution, but there *are* limits that depend on the video equipment and signal bandwidth.
- How horizontal resolution for analog video is relates to frequency

Department of Computer Science
National Tsing Hua University

Frequency of color change

- Consider the three video lines. Each is represented by a waveform on the right. The one that changes from black to white most frequently has the most detail, and in that sense it has the highest resolution
- A video camera has technological limits – how fast it can detect changing colors and record that signal



- The rate at which the camera can do this limits its horizontal resolution
- The faster the camera device can alternate between neighboring colors, the higher its horizontal resolution.

Analog video bandwidth

- Horizontal resolution of analog video is directly related to bandwidth
- **Bandwidth** is defined as the number of times a signal can change per unit time.
- Thus, it is measured in cycles per second, which is also a frequency measure
- According to the NTSC standard, an analog video signal is allocated a bandwidth of 6.0 MHz, 4.2 MHz for luminance and 1.5 for chrominance, and the remainder for audio (Table 6.6)

Standards for analog video

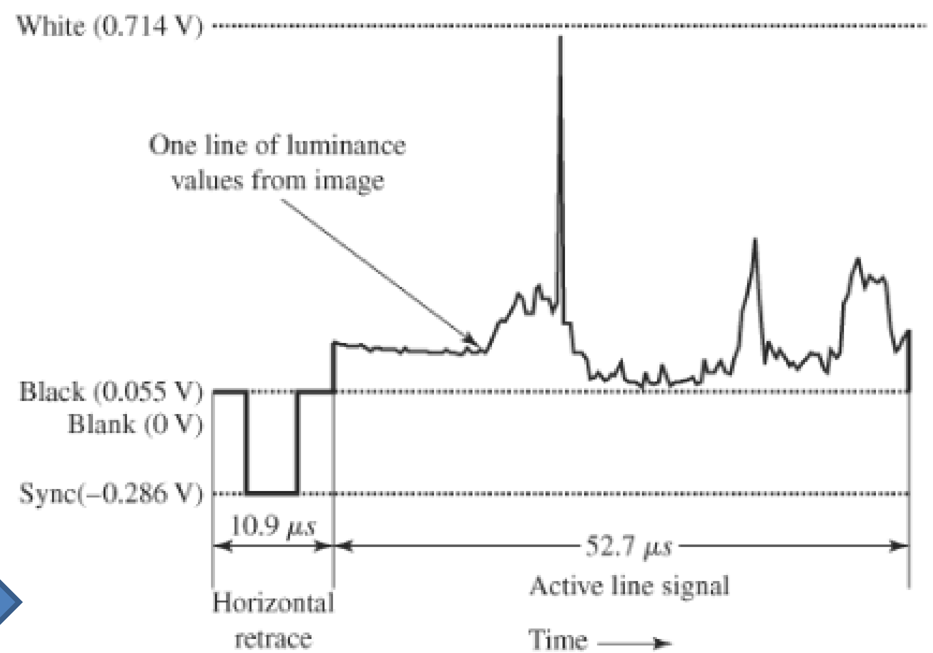
TABLE 6.6 NTSC, PAL, and SECAM Standards for Analog Television Composite Signals

Property	NTSC	PAL (version G)	SECAM (version D)
frame rate	29.97	25	25
type of scanning	interlaced	interlaced	interlaced
number of lines	525	625	625
number of active lines	~480	~575	~575
time to display one line (including horizontal retrace)	63.56 μsec	64 μsec	64 μsec
horizontal retrace	10.9 μsec	12 μsec	12 μsec
aspect ratio	4:3	4:3	4:3
color model	YIQ	YUV	YDbDr
luminance bandwidth	4.2 MHz	5.0 MHz	6.0 MHz
chrominance bandwidth	1.5 MHz (I), 0.5 MHz (Q)	1.3 MHz (U), 1.3 MHz (V)	1.0 MHz (Db), 1.0 MHz (Dr)
method of color modulation	QAM	QAM	FM
frequency of color subcarrier	3.58 MHz	4.43 MHz	4.25 MHz (Db), 4.41 MHz (Dr)
frequency of audio subcarrier	4.5 MHz	5.5 MHz	6.5 MHz
composite signal bandwidth	6.0 MHz	8.0 MHz	8.0 MHz

Analog video bandwidth

- Notice that neither the video camera nor the signal bandwidth dictates that there *will* be a certain horizontal resolution, only that the resolution will have to be within the bandwidth limits of the camera and the signal
- It indicates how changing voltages can be used to communicate one line of a video signal

The timing of one line of NTSC analog video



The timing of one line of NTSC analog video

- The drop in voltage at the beginning of the line indicates that the scanner should go back to the beginning of a line on the display device—the horizontal retrace.
- Transmission/display of one line is allotted about 63.56 μsec, with 10.9 μsec of this time taken up by the horizontal retrace.
- When the scanner reaches the end of a frame, it must return to the top left corner. This is called the ***vertical retrace*** (or in some sources, the ***vertical blanking interval***, ***VBI***)

The timing of one line of NTSC analog video

- In Table 6.6, a distinction is made between total lines and ***active lines***.
- For example, the total number of lines for NTSC is 525, but only 480 of these are active lines. The frame rate:

  0.00006356 *s/line* * 525 *lines/frame* = 0.33369 *s/frame* ≈ 29.97 *frame/s*
- ***Active lines*** : lines of data contain information relating to pixel colors
  - Some of the lines of data are reserved for the vertical retrace and other auxiliary information

Bandwidth and frequency rate

- A simplest case
  - a wave of maximum frequency *f*, and a single cycle of that wave. The maximum value of the wave communicates "white" and the minimum communicates "black."
  - Thus, the maximum frequency component of the wave sets a limit on how fast you can communicate information about the change in grayscale value from white to black
  - This is a basic concept—one cycle of a wave with maximum frequency component *f* can communicate two pieces of information yielding a sample rate of 2*f*
  - The sample rate is twice the frequency

Bandwidth and frequency rate

- An NTSC analog video signal is allotted 6 MHz in bandwidth, with 4.2 MHz for luminance information
- A bandwidth of 4.2 MHz can yield 8.4 million samples per second.

  4,200,000 *cycles/s* * 2 *samples/cycle* = 8,400,000 *samples/s*

- The samples for one line have 52.7 μsec allotted for their display.

  8,400,000 *samples/s* * 0.0000527*s/line* = 443 *samples/line*

- For a total resolution of 443 × 480, which is a ratio of about 0.923 to 1, not the 1.33: 1 aspect ratio

Bandwidth and frequency rate

- One way to explain the numbers is that you can't actually capture 480 distinct lines with a video camera, so the effective vertical resolution is less than 480
- By subjective experiments, it was determined that the best you could do would be to get about 70% of the lines.
- Thus, the ***Kell factor***, as it was called, was determined to be about 0.7
- We get 480 * 0.6875, or 330 lines in the vertical direction, yielding a resolution of 440×330

Bandwidth and frequency rate

- For a video signal transmission, let *a* be the aspect ratio, *v* be the number of active lines, *t* be the time to transmit one line, and *k* be the Kell factor. Then the bandwidth of the transmission, *b*, is defined by

$$b = \frac{akv}{2t}$$

- For example, assuming *k* = 0.6875; *a* = 4/3; *v* = 480; and *t* = 52.7 μsec/line. The effective number of active lines

0.6875 * 480 = 330

Bandwidth and frequency rate

- Multiply the resulting value by the aspect ratio to get a vertical resolution.

$$330 * 4/3 \approx 440$$

- A video line is transmitted in *t* seconds. Calculate the number of lines transmitted per second

$$1/t = 1\ \mathit{line}/0.0000527\ s = 18{,}975\ \mathit{lines}/s$$

- Frequency is then computed

$$440 * 18{,}975 = 8{,}349{,}146\ \mathit{samples}/s$$

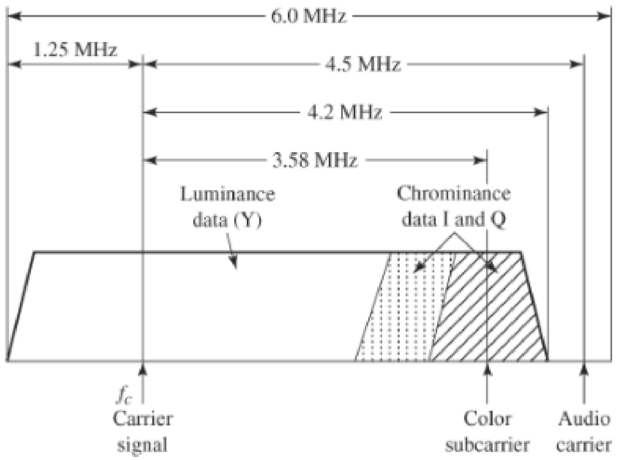
- A cycle of the signal can communicate two samples.

$$8{,}349{,}146/2 \approx 4.17\ \mathrm{MHz}$$

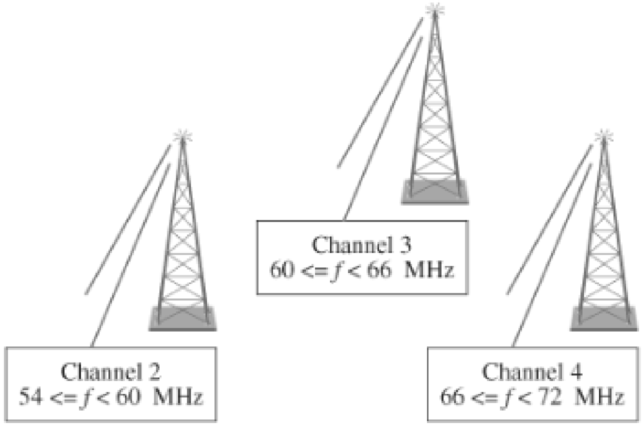
- A bandwidth of 4.17 MHz is close enough to the 4.2 MHz

NTSC video signal spectrum

- The total bandwidth allotted for an NTSC analog video signal is 6.0 MHz. This includes luminance, chrominance, audio data, and auxiliary information



- How would receivers be able to distinguish one station's frequencies from another, if they're all in the same frequency range?
- Each station gets a band of frequencies—called a ***channel***—with a bandwidth of 6 MHz



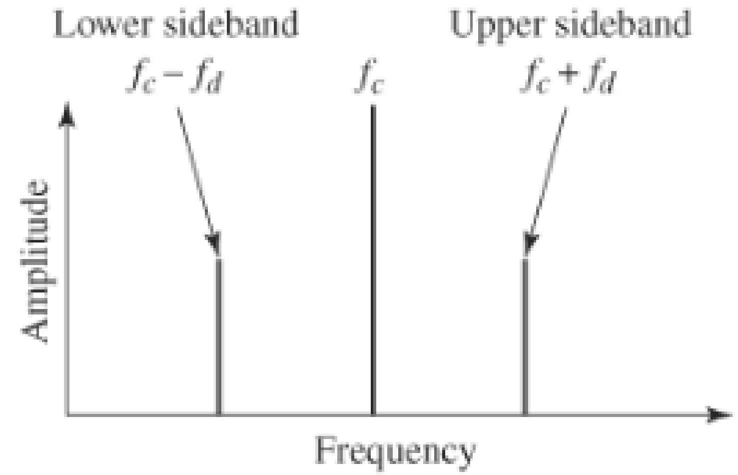
Amplitude modulation

- With amplitude modulation, the amplitude of a carrier wave can be done by a simple multiplication of the sinusoidal functions.
- Let ω_c be the angular frequency of a carrier signal. Let ω_d be the angular frequency of a data signal to be amplitude-modulated onto the carrier signal. Then the function defining the amplitude-modulated wave is

$$\begin{aligned}
&\cos(\omega_c t)(1.0 + \cos(\omega_d t)) \\
&= \cos(\omega_c t) + (\cos(\omega_c t)\cos(\omega_d t)) \\
&= \underbrace{\cos(\omega_c t)}_{\text{carrier frequency}} + \underbrace{\frac{1}{2}\cos(2\pi(f_c + f_d)t)}_{\text{upper sideband}} + \underbrace{\frac{1}{2}\cos(2\pi(f_c - f_d)t)}_{\text{lower sideband}}
\end{aligned}$$

Frequency bands from amplitude modulation

- Notice that the bandwidth of the signal is $b = (f_c + f_d) - (f_c - f_d) = 2f_d$.
- With this method of modulation, if the channel has a bandwidth of $2d_f$, that will be sufficient to carry the signal with frequency f_d.

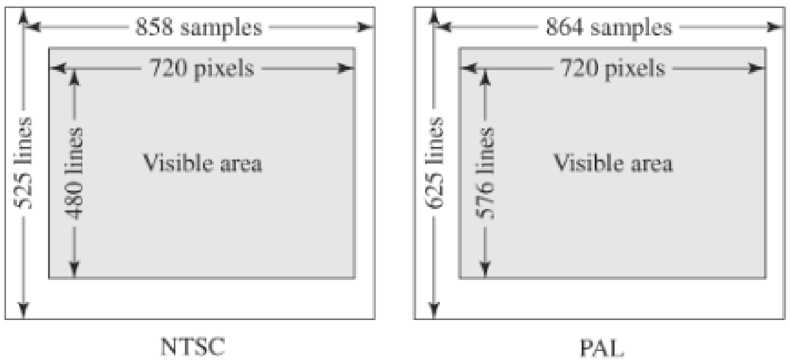


Analog-to-digital video converter

- The converter has to read 858 samples in each of the 525 lines in each of the 29.97 frames. From this we get the sample rate

sampling rate = horizontal resolution * vertical resolution * frames/s
NTSC: (858 * 525) samples/frame * 29.97 frames/s ≈ 13,500,000 samples/s = 13.5 MHz
PAL: (864 * 625) samples/frame * 25 frames/s = 13,500,000 samples/s = 13.5 MHz

- In NTSC analog video, there are 525 lines, 480 of which relate to the actual image. BT.601 stipulates that there be 858 samples in each line, 720 of these corresponding to visible pixels.

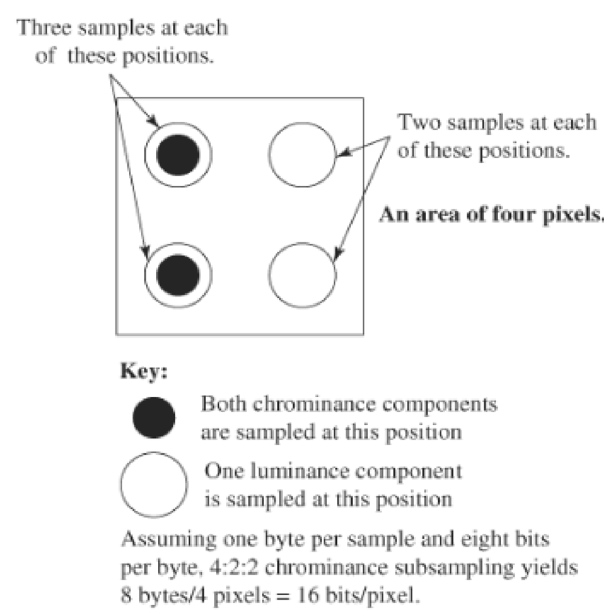


Determine bandwidth and the data rate

- Assume that you have 720 × 480 pixels in a frame, 29.97 frame/s, and 4: 2: 2 subsampling.

$$720 * 480 \ pixels/frame = 345{,}600 \ pixels/frame$$
$$345{,}600 \ pixels/frame * 29.97 \ frames/s = 10{,}357{,}632 \ pixels/s$$
$$10{,}357{,}632 pixels/s * 16 \ bits/pixel = 165{,}722{,}112 \ bits/s \approx 166 \ Mb/s$$

- The difference between this value of 166 Mb/s and the value of 172 Mb/s (D1) can be accounted for by overhead—*e.g.*, error checking, etc



Analog-to-digital video converter

- Bit rate tells us two things
  - how much data would have to be processed per second
  - how much data is generated for each second of video
- A data rate of 172 Mb/s is 21.5 MB/s, which is 77.4 GB per hour of uncompressed video
- Video compression methods are very effective at reducing file size while retaining quality. For DV video, a popular consumer format, is compressed at a rate of almost 5:1.

Digital Video Distribution Media

- Overview of types of CDs and DVDs

TABLE 6.7 Digital Video Distribution Media

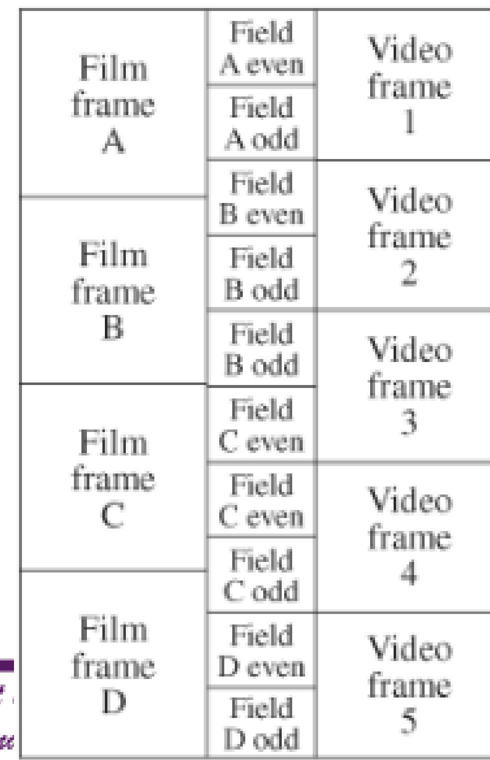
Format	VCD	SVCD	DVD	HD-DVD	Blu-ray
NTSC Resolution	352 × 240	480 × 480	720 × 480	1920 × 1080	1920 × 1080
Video Compression	MPEG-1	MPEG-2	MPEG-2	MPEG-2, MPEG-4 AVC, SMPTE-VC1	MPEG-2, MPEG-4 AVC, SMPTE-VC1
Audio Compression	MP1	MP1	PCM, DD, DTS Surround	PCM, DD, DD^{+}, DD, TrueHD, DTS, DTS-HD	PCM, DD, 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



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.
  - This second situation arises when HDTV is transmitted in 1080i format and received by a television that is 1080p.

Department of Computer Science
National Tsing Hua University

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

Digital video file

File Extension	Container File Type	Characteristics
.avi	Audio/Video Interleaved	A type of RIFF file designed for Windows Media. Created for the PC platform, now used on Mac and Linux also. Can be uncompressed, or compressed with a variety of codecs, including DivX, Cinepak, Indeo, MJPEG, or DV.
.mov (sometimes *.qt*)	QuickTime Movie	A multimedia container file framework created by Apple; it stores different media—including video, sound, and text—in different tracks. Cross-platform, for Mac, PC, and Linux. Accommodates a variety of audio and video codecs, including Sorensen, MPEG, Cinepak, and DivX.
.mpg (also *.mpeg, .m1v, .m2v, .m2t, .mp4, .mpv2*)	MPEG	A file that has been compressed with some version of the MPEG codec. MPEG-1 is greatly compressed with small resolution, for use on CD or web. MPEG-2 is the standard for video on DVD. MPEG-4 has highest compression rate and serves as a container file, modeled after QuickTime.
.flv, .f4v, .f4p, .f4a, .f4b	Flash video	Encoded audio and video streams playable by the Flash player, which can exchange audio, video, and data over RTMP (real-time messaging protocol) connections with Adobe Flash Media Server. Now used widely on the web.
.ogg	Ogg (by Xiph Foundation)	An open-source format, good for internet streaming.
.rm and *.rmvb*	Real Media	A proprietary file format with accompanying codecs, developed by RealNetworks; it works on multiple platforms, including Windows, Mac, Linux, and Unix.
.wmv (sometimes *.asf*)	Microsoft Windows Media	Originally Microsoft proprietary codec for Windows Media Player, standardized by SMPTE; uses Advanced Systems format (a container format) sometimes with .asf suffix. Uses its own codec.

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)

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…

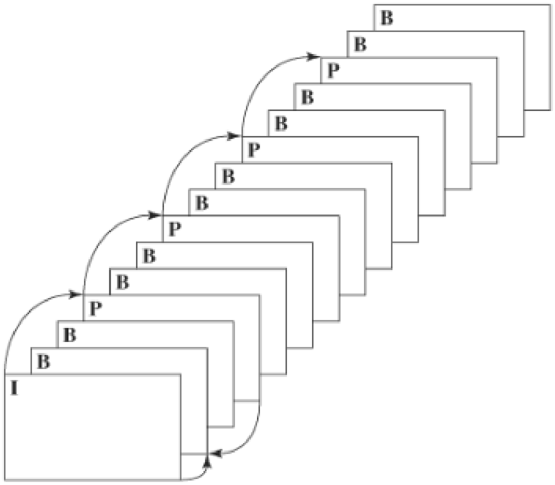
The main steps of MPEG compression

ALGORITHM 7.1

```
algorithm mpeg
/*Input: A sequence of frames of digital video
  Output: The same sequence, compressed*/
{
   Divide the sequence of frames into groups of pictures (GOPs), identifying I, P,
   and B frames.
   Divide each frame into macroblocks.
   Identify I frames (intraframes, to be compressed spatially with JPEG), P frames
   (forward prediction frames, to be compressed temporally relative to a preceding
   I or P frame), and B frames (bidirectional frames, to be compressed temporally
   relative to preceding and following I and/or P frames).
   For each P and B frame, compare the frame to the related I or P frame to determine
   a motion vector (or more than one motion vector).
   Record differential values for P and B frames (the difference between the expected
   pixel value, adjusted by motion compensation, and the actual value).
   For all frames, compress with JPEG compression
     transform data to the frequency domain with DCT
     arrange in zigzag order
     quantize
     apply entropy encoding (e.g., Huffman encoding)
}
```

The 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



The 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

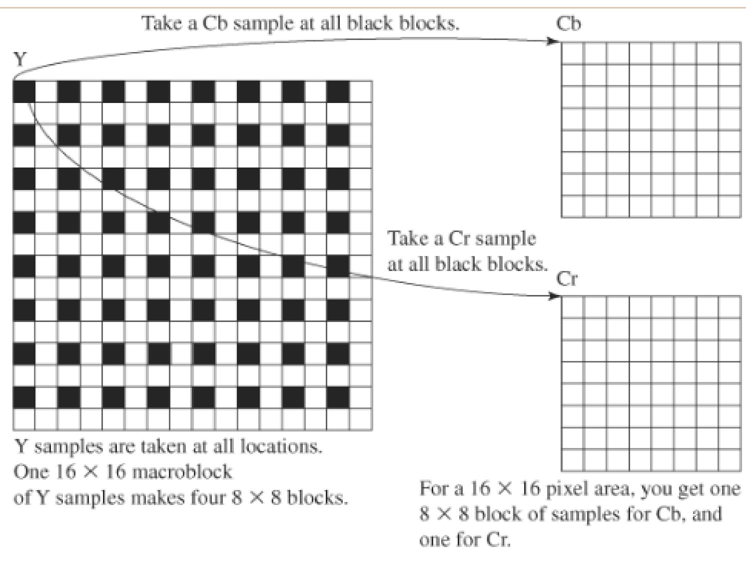


Figure 7.22 Macroblock for 4: 2: 0 subsampling

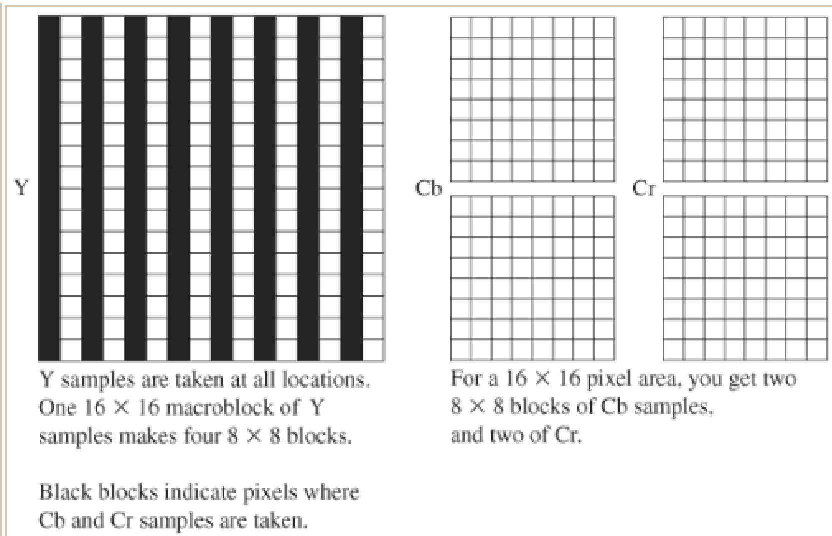
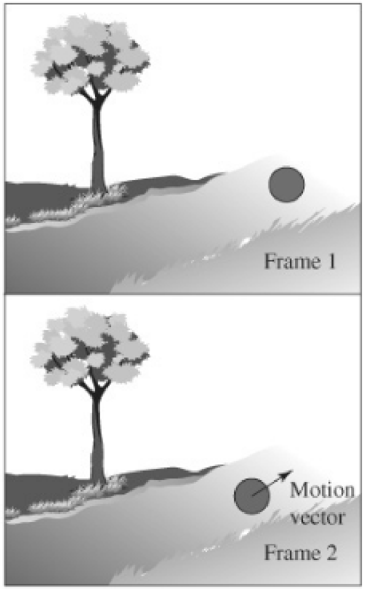


Figure 7.23 Macroblock for 4: 2: 2 subsampling

The 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***.



The 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 $\boldsymbol{T}_{x,y}$.
- ***mean absolute difference (MAD):*** use the average of each pixel's color difference from its corresponding pixel in the reference block.

$$MAD(i,j) = \frac{1}{N^2}\sum_{a=0}^{N-1}\sum_{b=0}^{N-1}|T(x+a,y+b)-R(x+i+a,y+j+b)|$$

where *N* is the size of the macroblock

The main steps of MPEG compression

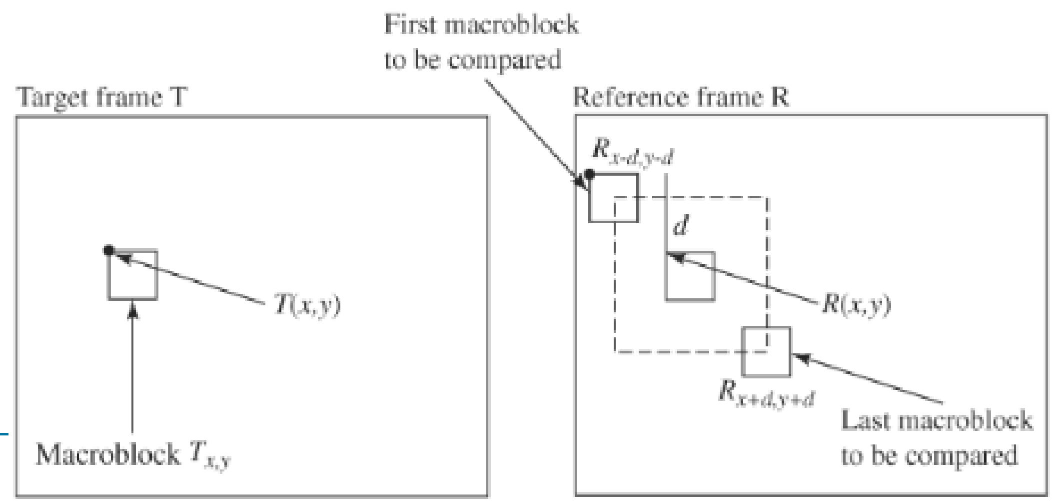
- Motion Estimation algorithm

ALGORITHM 7.2

```
algorithm motion_estimation
/*Input: Two video frames, the target frame T and reference frame R; the origin (x, y)
  of a macroblock to be matched
  Output: Motion vector (p, q)*/
{
/*Let MAX_NUM be a value larger than any anticipated MAD*/
```

$$/\text{*Let } MAD(i, j) = \frac{1}{N^2}\sum_{a=0}^{N-1}\sum_{b=0}^{N-1} |T(x + a, y + b) - R(x + i + a, y + j + b)| \text{*}/$$

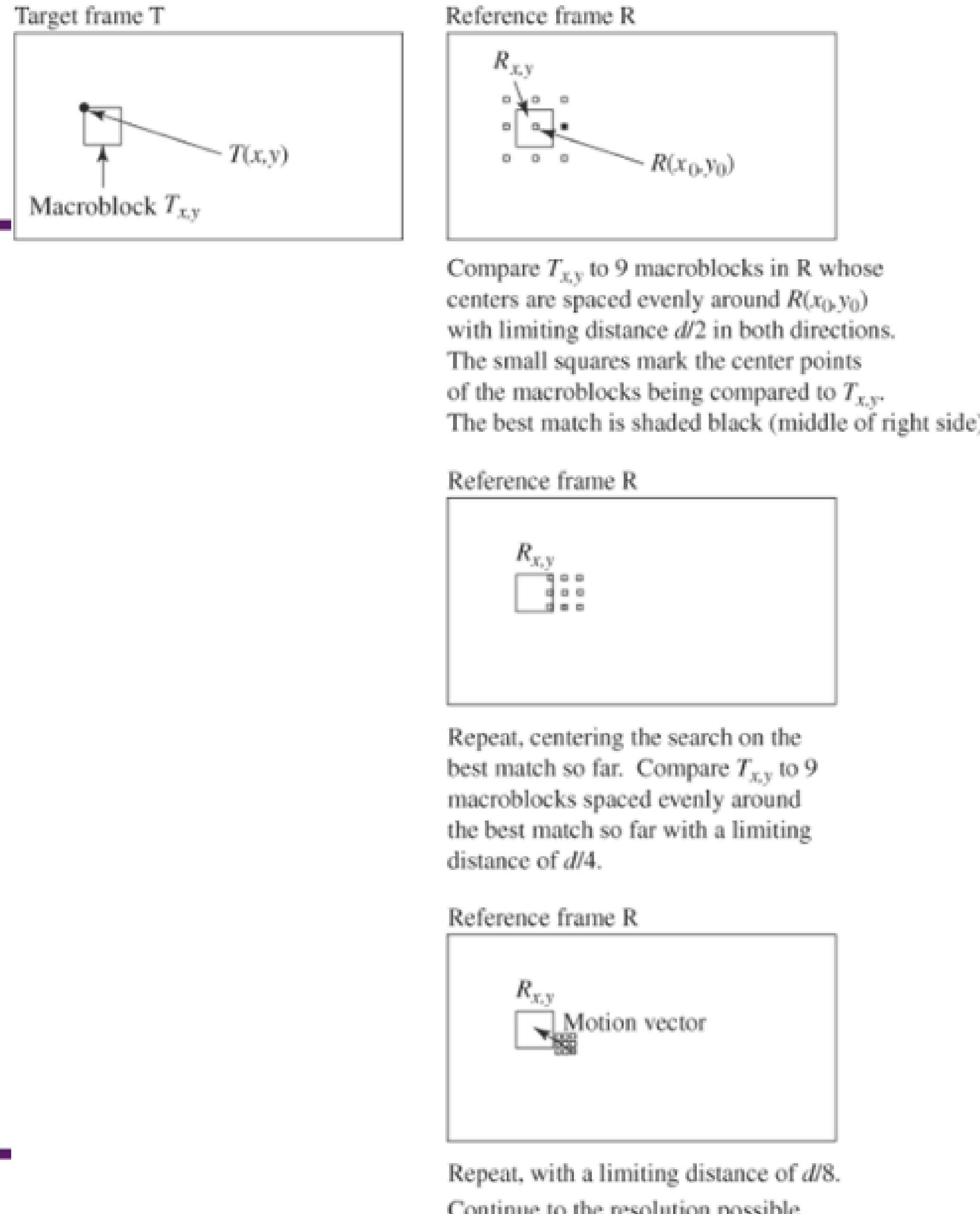
```
   min = MAD(0,0);
   p = 0;
   p = 0;
   for i = -d to d
     for j = -d to d {
       avg = MAD (i,j;)
       if (avg < min) {
          min = avg;
          p = i;
          q = j;
       }
     }
}
```



Compare macroblock $T_{x,y}$ to every macroblock $R_{x+i,y+j}$ for $-d \le i \le d$ and $-d \le j \le d$. The dotted line goes through the centers of the most-distant macroblocks to be compared.

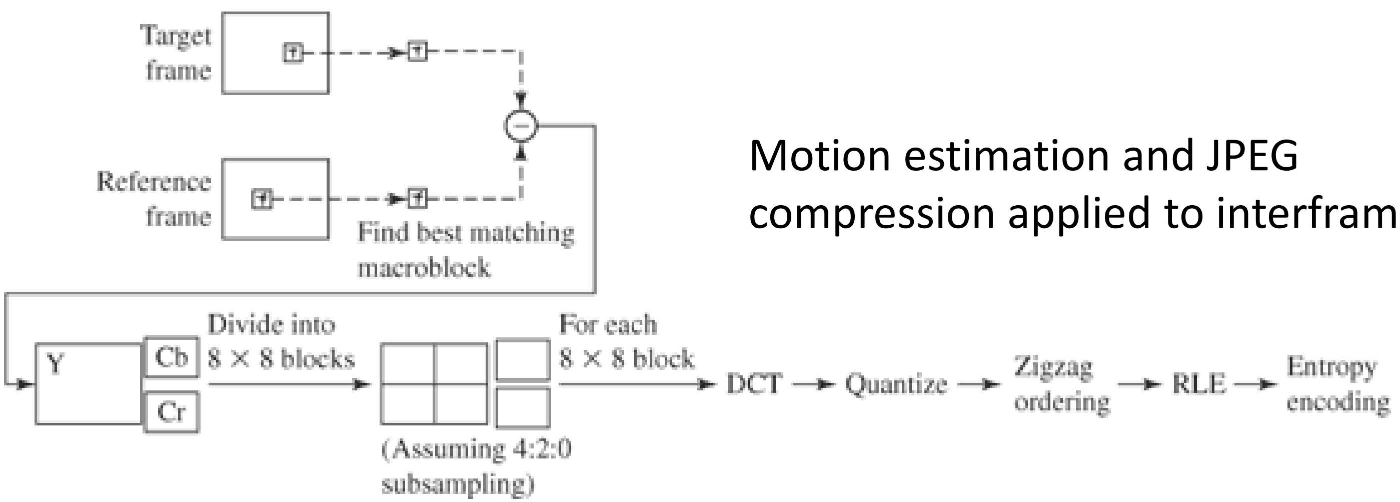
Motion estimation, full search

- 2D logarithmic motion estimation



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.



Motion estimation and JPEG compression applied to interframes