



Fundamentals of Multimedia

2nd Edition 2014
+Supplementary material

Chapter 5 :
Fundamental Concepts in Video

فهرست مطالب این فصل

- The principal notions needed to understand video
- How video challenges impact multimedia applications
- Analog video
- Digital video
- Video display interfaces
- 3D video

Video

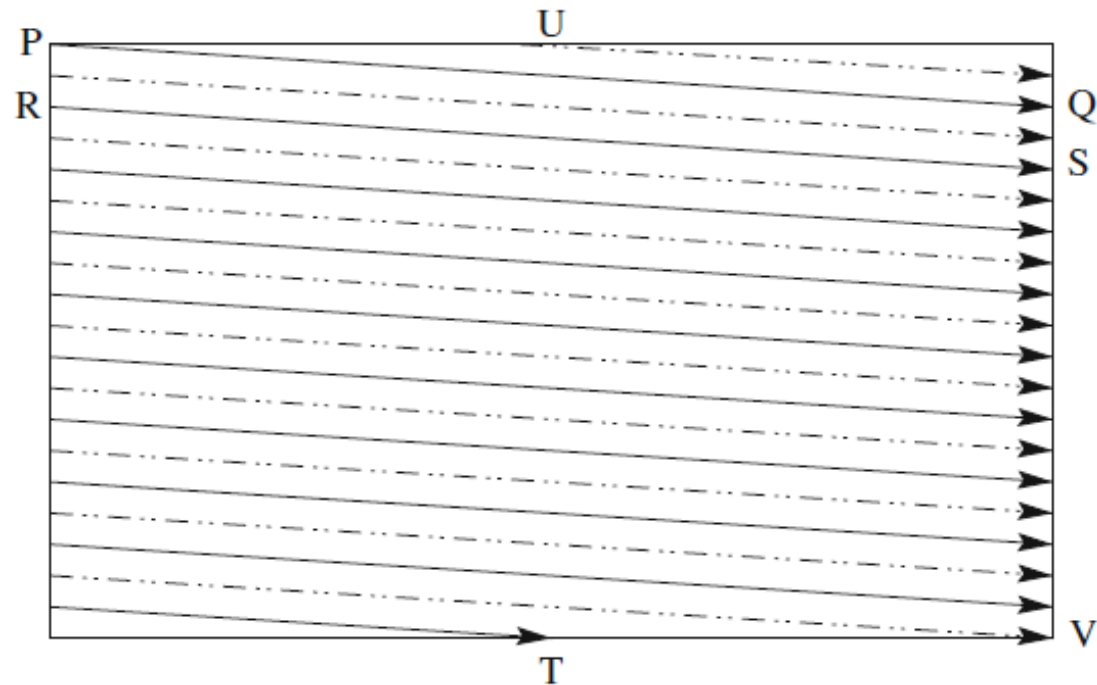
- Since video is created from a variety of sources, we begin with the signals themselves
- **Analog video** is represented as a continuous (time-varying) signal
- **Digital video** is represented as a sequence of digital images.

5.1 Analog Video

- An analog signal $f(t)$ samples a time-varying image. So-called *progressive* scanning traces through a complete picture (a frame) row-wise for each time interval.
- A high-resolution computer monitor typically uses a time interval of $1/72$ s.
- In TV and in some monitors and multimedia standards, another system, *interlaced* scanning, is used.
- Here, the odd-numbered lines are traced first, then the even-numbered lines.
- This results in “odd” and “even” *fields*—two fields make up one frame.

5.1 interlacing

Fig. 5.1 Interlaced raster scan



- The odd lines (starting from *I*) end up at the middle of a line at the end of the odd field, and the even scan starts at a half-way point.
- First the solid (odd) lines are traced—*P* to *Q*, then *R* to *S*, and so on, ending at *T*
- Then the even field starts at *U* and ends at *V*.
- The scan lines are not horizontal because a small voltage is applied, moving the electron beam down over time.

5.1 interlacing

- ▶ **Interlacing** was invented because, when standards were being defined, it was difficult to transmit the amount of information in a full frame quickly enough to avoid flicker, the double number of fields presented to the eye reduces the eye perceived flicker.
- The jump from Q to R and so on in Fig. 5.1 is called the **horizontal retrace**, during which the electronic beam in the CRT is blanked.
- The jump from T to U or V to P is called the **vertical retrace**.



Fig. 5.2 Interlaced scan produces two fields for each frame: **a** the video frame; **b** Field 1; **c** Field 2; **d** difference of fields

5.1.1 NTSC Video

- NTSC stands for (National Television System Committee of the U.S.A)
- The NTSC TV standard is mostly used in North America and Japan.
- It uses a familiar 4:3 *aspect ratio* (i.e., the ratio of picture width to height) and 525 (interlaced) scan lines per frame at 30 frames per second (fps).
- Figure 5.4 shows the effect of “vertical retrace and sync” and “horizontal retrace and sync” on the NTSC video raster.

More exactly, for historical reasons NTSC uses 29.97 fps—or, in other words, 33.37 ms per frame. NTSC follows the interlaced scanning system, and each frame is divided into two fields, with 262.5 lines/field. Thus the horizontal sweep frequency is $525 \times 29.97 \approx 15,734$ lines/s, so that each line is swept out in $1/15,734 \text{ s} \approx 63.6 \mu\text{s}$.

Fig. 5.3 Electronic signal for one NTSC scan line

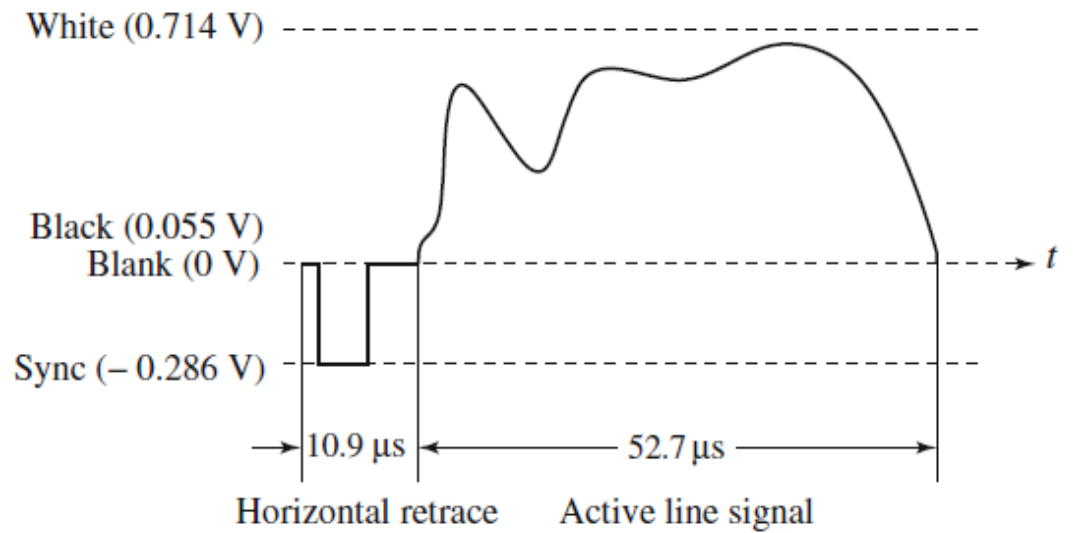
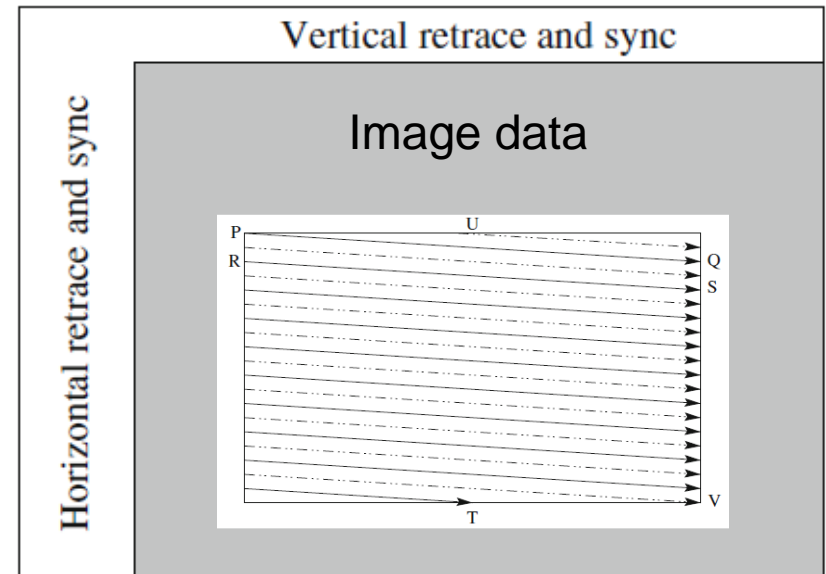


Fig. 5.4 Video raster, including retrace and sync data



The jump from Q to R and so on in Fig. 5.1 is called the *horizontal retrace*, during which the electronic beam in the CRT is blanked. The jump from T to U or V to P is called the *vertical retrace*.

5.1.1 NTSC Video

- Figure 5.4 shows the effect of “**vertical retrace and sync**” and “**horizontal retrace and sync**” on the NTSC video raster.
- **Blanking information** is placed into 20 lines reserved for control information at the beginning of each field.
- Hence, the number of **active video lines** per frame is only 485.
- Similarly, almost 1/6 of the raster at the left side is blanked for horizontal retrace and sync.
- The nonblanking pixels are called **active pixels**.
- Image data is not encoded in the blanking regions, but other information can be placed there, such as **V-chip information**, **stereo audio** channel data, and **subtitles** in many languages.

5.1.1 NTSC Video

- NTSC video is an analog signal with no fixed horizontal resolution.
- Therefore, we must decide how many times to sample the signal for display.
- Each sample corresponds to one pixel output.
- A *pixel clock* divides each horizontal line of video into samples.
- The higher the frequency of the pixel clock, the more samples per line.
- Different video formats provide different numbers of samples per line, as listed in Table 5.1.

Table 5.1 Samples per line for various analog video formats

Format	Samples per line
VHS	240
S-VHS	400–425
Beta-SP	500
Standard 8 mm	300
Hi-8 mm	425

NTSC Color Scheme

NTSC uses the YIQ color model. We employ the technique of *quadrature modulation* to combine (the spectrally overlapped part of) I (in-phase) and Q (quadrature) signals into a single chroma signal C [3,1]:

$$C = I \cos(F_{sc}t) + Q \sin(F_{sc}t) \quad (5.1)$$

RGB to YUV

$$\begin{aligned} Y &= (0.257 * R) + (0.504 * G) + (0.098 * B) + 16 \\ Cr = V &= (0.439 * R) - (0.368 * G) - (0.071 * B) + 128 \\ Cb = U &= -(0.148 * R) - (0.291 * G) + (0.439 * B) + 128 \end{aligned}$$

سه مدل مختلف نمایش رنگ

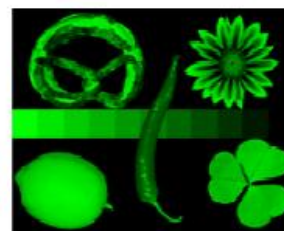
RGB, CMY, YUV



Original Image



Red Color Component



Green Color Component



Blue Color Component



Original Image



Cyan Color Component



Magenta Color Component



Yellow Color Component



Original Image



Luminance (Y) Component



Chrominance (U) Component



Chrominance (V) Component

نحوه مدولاسیون NTSC

- اطلاعات روشنایی (Y) و رنگ (مولفه‌های I و Q یا U و V) بشکل زیر بر روی سیگنال رادیویی مدوله می شوند.

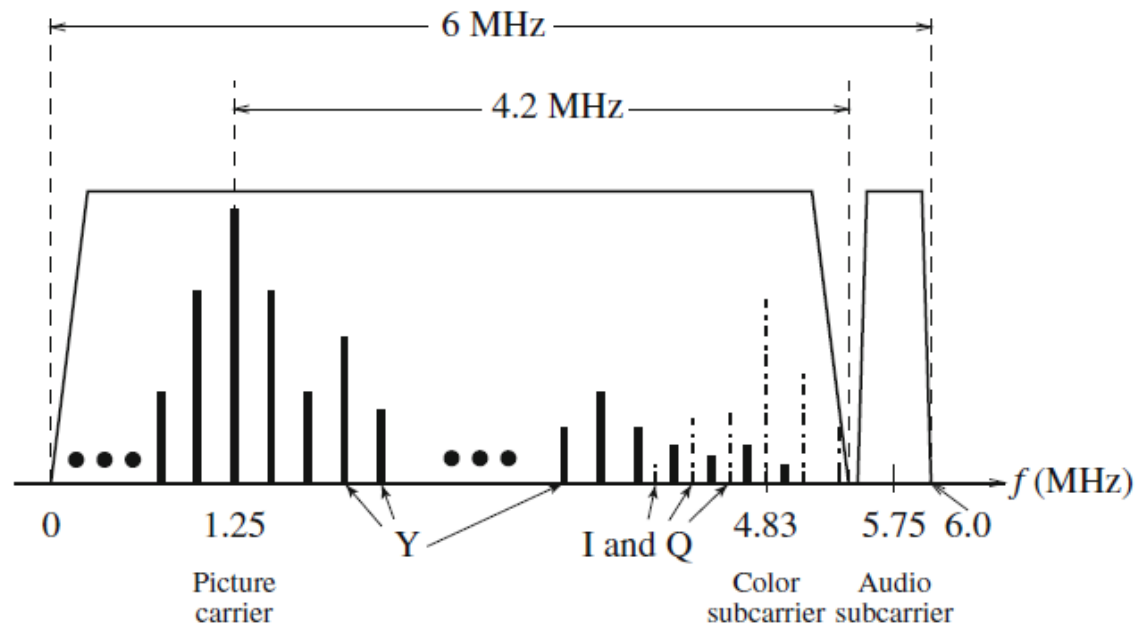


Fig. 5.5 Interleaving Y and C signals in the NTSC spectrum

5.1.2 PAL Video

- PAL (**Phase Alternating Line**) is a TV standard originally invented by German scientists.
- This important standard is widely used in Western Europe, China, India, and many other parts of the world.
- Because it has higher resolution than NTSC, the visual quality of its pictures is generally better.

Table 5.2 Comparison of analog broadcast TV systems

TV system	Frame rate (fps)	Number of scan lines	Total channel width (MHz)	Bandwidth allocation (MHz)		
				<i>Y</i>	<i>I</i> or <i>U</i>	<i>Q</i> or <i>V</i>
NTSC	29.97	525	6.0	4.2	1.6	0.6
PAL	25	625	8.0	5.5	1.8	1.8
SECAM	25	625	8.0	6.0	2.0	2.0

5.1.3 SECAM Video

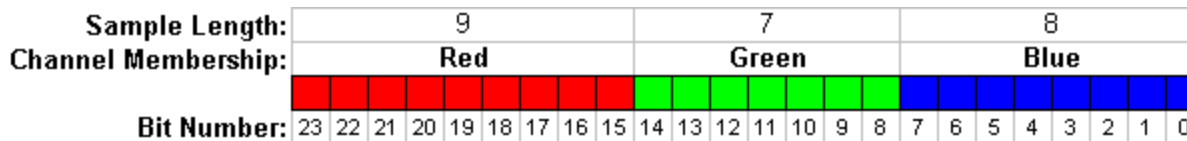
- SECAM, which was invented by the French, is the third major broadcast TV standard.
- SECAM stands for *Système Electronique Couleur Avec Mémoire*.
- SECAM and PAL are similar, differing slightly in their color coding scheme.

5.2 Digital Video

- The advantages of digital representation for video:
 - Storing video on digital devices or in memory, ready to be processed (noise removal, cut and paste, and so on) and integrated into various multimedia applications.
 - Direct access, which makes nonlinear video editing simple.
 - Repeated recording without degradation of image quality.
 - Ease of encryption and better tolerance to channel noise.

Sampling

- A **sample** is an intersection of channel and a pixel
- The diagram below depicts a 24-bit pixel, consisting of 3 samples for Red (channel) , Green (channel) , and Blue (channel) .
- In this particular diagram, the Red sample occupies 9 bits, the Green sample occupies 7 bits and the Blue sample occupies 8 bits, totaling 24 bits per pixel



- A sample is related to a subpixel on a physical display.

Chroma Subsampling

- چون چشم آدمی به رنگ کمتر از روشنایی حساس است، روش نمونه‌برداری رنگ مطرح شده است که در آن اطلاعات رنگ برای بخشی از پیکسل‌ها ارسال می‌شود.

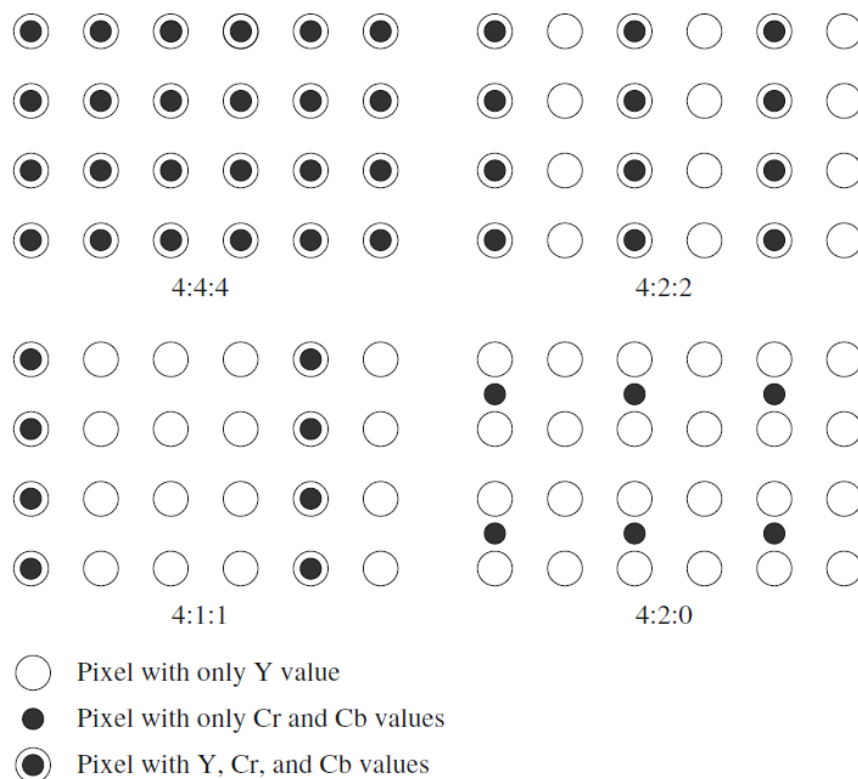


Fig. 5.6 Chroma subsampling



Original without color subsampling. 200% zoom.



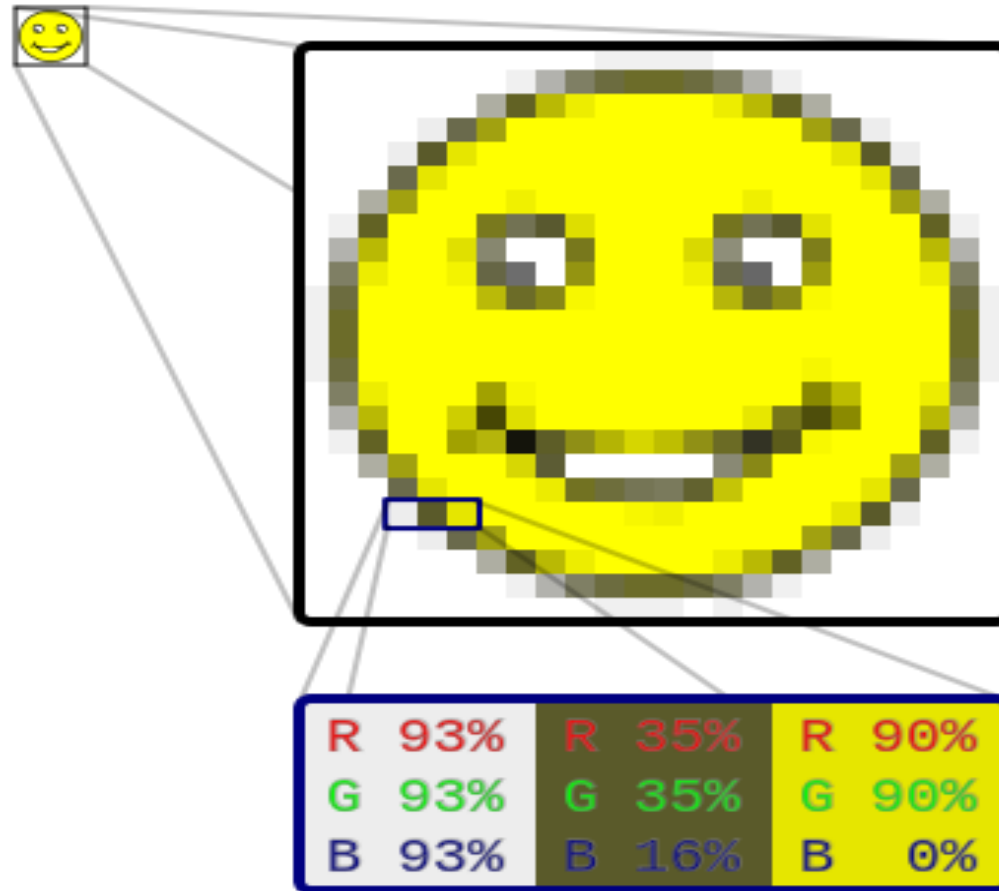
Image after color subsampling (compressed with Sony Vegas DV codec, box filtering applied.)

What is Raster Graphics?

- a raster graphics image is a dot matrix data structure representing a generally rectangular grid of pixels, or points of color, viewable via a monitor, paper, or other display medium. (=Bitmap)
- A *raster* is technically characterized by the width and height of the image in pixels and by the number of bits per pixel (a color depth, which determines the number of colors it can represent)
- Most computer images are stored in raster graphics formats.
- Raster graphics are *resolution dependent*, meaning they cannot scale up to an arbitrary resolution without loss of **apparent quality**. This property contrasts with the capabilities of **vector graphics**, which easily scale up to the quality of the device rendering them.
- http://vector-conversions.com/vectorizing/raster_vs_vector.html
- <https://99designs.com/designer-blog/2011/05/02/vector-vs-raster-images/>

What is Raster Graphics?

- The smiley face in the top left corner is a raster image. When enlarged, individual pixels appear as squares. Zooming in further, they can be analyzed, with their colors constructed by adding the values for red, green and blue.



5.2.2 استانداردهای ویدئو دیجیتال

- **NTSC** compatible format: One of the most important standards it has produced is **CCIR-601** for component digital video. This standard has since become standard ITU-R **Rec. 601**.
 - **CIF** stands for *Common Intermediate Format*, specified by the International Telegraph and Telephone Consultative Committee (**CCITT**).
 - The idea of CIF, which is about the same as VHS quality, is to specify a format for lower bitrate. CIF uses a progressive (noninterlaced) scan.
 - CIF is a compromise between NTSC and PAL, in that it adopts the NTSC frame rate and half the number of active lines in PAL.
 - **QCIF** stands for Quarter-CIF, and is for even lower bitrate.
-
- The **CCIR** is the **C**onsultative **C**ommittee for **I**nternational **R**adio.

5.2.2 استانداردهای ویدئو دیجیتال

The NTSC version has 525 scan lines, each having 858 pixels (with 720 of them visible, not in the blanking period). Because the NTSC version uses 4:2:2, each pixel can be represented with two bytes (8 bits for Y and 8 bits alternating between Cb and Cr). The Rec. 601 (NTSC) data rate (including blanking and sync but excluding audio) is thus approximately 216 Mbps (megabits per second):

$$525 \times 858 \times 30 \times 2 \text{ bytes} \times 8 \frac{\text{bits}}{\text{byte}} \approx 216 \text{ Mbps}$$

Table 5.3 ITU-R digital video specifications

	Rec. 601 525/60 NTSC	Rec. 601 625/50 PAL/SECAM	CIF	QCIF
Luminance resolution	720×480	720×576	352×288	176×144
Chrominance resolution	360×480	360×576	176×144	88×72
Color subsampling	4:2:2	4:2:2	4:2:0	4:2:0
Aspect ratio	4:3	4:3	4:3	4:3
Fields/sec	60	50	30	30
Interlaced	Yes	Yes	No	No

5.2.3 High-Definition TV

- The introduction of wide-screen movies brought the discovery that viewers seated near the screen enjoyed a level of participation (sensation of immersion) not experienced with conventional movies.
- Apparently the exposure to a greater field of view, especially the involvement of peripheral محيطي vision, contributes to the sense of “being there.”
- The main thrust of High-Definition TV (**HDTV**) is not to increase the “definition” in each unit area, but rather to increase the visual field, especially its width.
- First-generation HDTV was based on an analog technology developed by Sony and NHK in Japan in the late 1970s.

5.2.3 High-Definition TV

- Multiple sub-Nyquist Sampling Encoding (MUSE) was an improved NHK HDTV with **hybrid** analog/digital technologies that was put in use in the 1990s.
- It has 1,125 scan lines, interlaced (60 fields per second), and a 16:9 aspect ratio. (compare with NTSC 4:3 aspect ratio, see slide 8)
- In 1987, the FCC decided that HDTV standards must be compatible with the existing NTSC standard and must be confined to the existing Very High Frequency (**VHF**) and Ultra High Frequency (**UHF**) bands.

5.2.4 Ultra High Definition TV (UHD TV)

- UHD TV is a new development—a new generation of HDTV!
- The standards announced in 2012
- The aspect ratio is 16:9.
- The supported frame rate has been gradually increased to 120 fps.

5.3 Video Display Interfaces

- We now discuss the interfaces for video signal transmission **from** some output devices (e.g., set-top box, video player, video card, and etc.) **to** a video display (e.g., TV, monitor, projector, etc.).
- There have been a wide range of video display interfaces, supporting video signals of different formats (analog or digital, interlaced or progressive), different frame rates, and different resolutions
- We start our discussion with
 - **analog** interfaces, including Component Video, Composite Video, and S-Video,
 - and then **digital** interfaces, including DVI, HDMI, and DisplayPort.

5.3.1 Analog Display Interfaces

- Analog video signals are often transmitted in one of three different interfaces:
 - Component video,
 - Composite video, and
 - S-video.
- Figure 5.7 shows the typical connectors for them



Fig. 5.7 Connectors for typical analog display interfaces. From left to right: Component video, Composite video, S-video, and VGA

5.3.1 Analog Display Interfaces

- **Component Video**
- Higher end video systems, such as for studios, make use of three separate video signals for the red, green, and blue image planes.
- This is referred to as **component** video.
- This kind of system has three wires (and connectors) connecting the camera or other devices to a TV or monitor.

5.3.1 Analog Display Interfaces

- **Composite Video**
- When connecting to TVs or VCRs, composite video uses only one wire (and hence one connector, such as a BNC connector at each end of a coaxial cable or an RCA plug at each end of an ordinary wire), and video color signals are mixed, not sent separately.
- The audio signal is another addition to this one signal.

5.3.1 Analog Display Interfaces

- **S-Video**

- As a compromise, S-video (separated video, or super-video, e.g., in S-VHS) uses two wires: one for luminance and another for a composite chrominance signal.
- The reason for placing luminance into its own part of the signal is that black-and white information is most important for visual perception.
- As noted in the previous chapter, humans are able to differentiate spatial resolution in the grayscale (“black and-white”) part much better than for the color part of RGB images.
- Therefore, color information transmitted can be much less accurate than intensity information.
- We can see only fairly large blobs (نقاط) of color, so it makes sense to send less color detail.

5.3.1 Analog Display Interfaces

- **Video Graphics Array (VGA)**
- The Video Graphics Array (VGA) is a video display interface that was first introduced by IBM in 1987, along with its PS/2 personal computers. It has since been widely used in the computer industry with many variations, which are collectively referred to as VGA.
- The initial VGA resolution was 640×480 pixels.
- The VGA video signals are based on analog component RGBHV (red, green, blue, horizontal sync, vertical sync).

5.3.2 Digital Display Interfaces

- Given the rise of digital video processing and the monitors that directly accept digital video signals, there is a great demand toward video display interfaces that transmit digital video signals.
- Such interfaces emerged in 1980s (e.g., Color Graphics Adapter (CGA))
- Today, the most widely used digital video interfaces include Digital Visual Interface (DVI), High-Definition Multimedia Interface (HDMI), and Display Port, as shown in Fig. 5.8.



Fig. 5.8 Connectors of different digital display interfaces. From left to right: DVI, HDMI, DisplayPort

5.3.2 Digital Display Interfaces

- **Digital Visual Interface (DVI)**
- Digital Visual Interface (DVI) was developed by the *Digital Display Working Group* (DDWG) for transferring digital video signals, particularly from a computer's video card to a monitor.
- It carries uncompressed digital video and can be configured to support multiple modes, including DVI-D (digital only), DVI-A (analog only), or DVI-I (digital and analog).
- The support for analog connections makes DVI backward compatible with VGA (though an adapter is needed between the two interfaces).
- The DVI allows a maximum 16:9 screen resolution of 1920×1080 pixels.

5.3.2 Digital Display Interfaces

- **High-Definition Multimedia Interface (HDMI)**
- HDMI is a newer digital audio/video interface developed to be backward-compatible with DVI.
- HDMI, however, differs from DVI in the following aspects:
 1. HDMI does not carry analog signal and hence is not compatible with VGA.
 2. DVI is limited to the RGB color range (0–255).
 3. HDMI supports digital audio, in addition to digital video.
- The HDMI allows a maximum screen resolution of 2560×1600 pixels.

5.3.2 Digital Display Interfaces

- **Display Port**
- Display Port is a digital display interface. It is the first display interface that uses packetized data transmission, like the Internet or Ethernet
- Display Port can achieve a higher resolution with fewer pins than the previous technologies.
- The use of data packets also allows Display Port to be extensible, i.e., new features can be added over time without significant changes to the physical interface itself.
- Display Port can be used to transmit audio and video simultaneously, or either of them.
- Compared with HDMI, Display Port has slightly more bandwidth, which also accommodates multiple streams of audio and video to separate devices.

5.4 3D Video and TV

- the rapid progress in the research and development of 3D technology and the success of the 2009 film *Avatar* have pushed 3D video to its peak.
- The main advantage of the 3D video is that it enables the experience of immersion be there, and really Be there!
- Increasingly, it is in movie theaters, broadcast TV (e.g., sporting events), personal computers, and various handheld devices.

5.4.1 Cues (اشارات) for 3D Percept

- The human vision system is capable of achieving a 3D percept by utilizing multiple cues.
- They are combined to produce optimal (or nearly optimal) depth estimates.
- When the multiple cues agree, this enhances the 3D percept.
- When they conflict with each other, the 3D percept can be hindered. Sometimes, illusions can arise.

Monocular Cues

- The monocular cues that do not necessarily involve both eyes include:
 - **Shading**—depth perception by shading and highlights
 - **Perspective** منظور scaling—converging parallel lines with distance and at infinity
 - **Relative size**—distant objects appear smaller compared to known same-size objects not in distance
 - **Texture gradient** نسيج —the appearance of textures change when they recede يتراجع in distance
 - **Blur gradient** —ضبابي objects appear sharper at the distance where the eyes are focused, whereas nearer and farther objects are gradually blurred
 - **Haze—due** تشوش to light scattering by the atmosphere, objects at distance have lower contrast and lower color saturation
 - **Occlusion** إطباق —a far object occluded by nearer object(s)
 - **Motion parallax** —induced by object movement and head movement, such that nearer objects appear to move faster.
- Among the above monocular cues, it has been said that Occlusion and Motion parallax are more effective.

Binocular Cues

- The human vision system utilizes effective binocular vision, i.e., **stereo vision** or **stereopsis** (Greek word "stereos" which means firm or solid).
- Our left and right eyes are separated by a small distance, on average approximately 2.5 inches, or 65mm, which is known as the *interocular distance*.
- As a result, the left and right eyes have slightly different views, i.e., images of objects are shifted horizontally.
- The amount of the shift, or disparity, is dependent on the object's distance from the eyes, i.e., its depth, thus providing the binocular cue for the 3D percept.
- The horizontal shift is also known as *horizontal parallax*.
- The *fusion* of the left and right images into single vision occurs in the brain, producing the 3D percept.
- Current 3D video and TV systems are almost all based on **stereopsis** because it is believed to be the most effective cue.

5.4.2 3D CameraModels

- **Simple Stereo Camera Model**

We can design a simple (artificial) stereo camera system in which the left and right cameras are identical (same lens, same focal length, etc.); the cameras' optical axes are in **parallel**, pointing at the Z-direction, the scene depth

- **Toed-in Stereo Camera Model**

Human eyes can be emulated by so-called Toed-in Stereo Cameras, in which the camera axes are usually **converging** and not in parallel.

One of the complications of this model is that objects at the same depth (i.e., the same Z) *in the scene no longer yield the same disparity* (تفاوت)

In other words, the “disparity planes” are now curved.

Objects on both sides of the view appear farther away than the objects in the middle, even when they have the same depth Z .

5.4.3 3DMovie and TV Based on Stereo Vision

- **3D Movie Using Colored Glasses**
- **3D Movies Using Circularly Polarized Glasses**
- **3D TV with Shutter Glasses**



End of Chapter 5