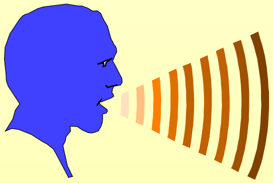
   

EL612

Video Processing

Lecture I
Basics of Video

Yao Wang
Polytechnic University, Brooklyn, NY11201
yao @vision.poly.edu

Outline

- Color perception and specification
- Video capture and display
- Analog raster video
- Analog TV systems
- Digital video

Color Perception and Specification

- Light -> color perception
- Human perception of color
- Type of light sources
- Trichromatic color mixing theory
- Specification of color
  - Tristimulus representation
  - Luminance/Chrominance representation
- Color coordinate conversion

Light is part of the EM wave

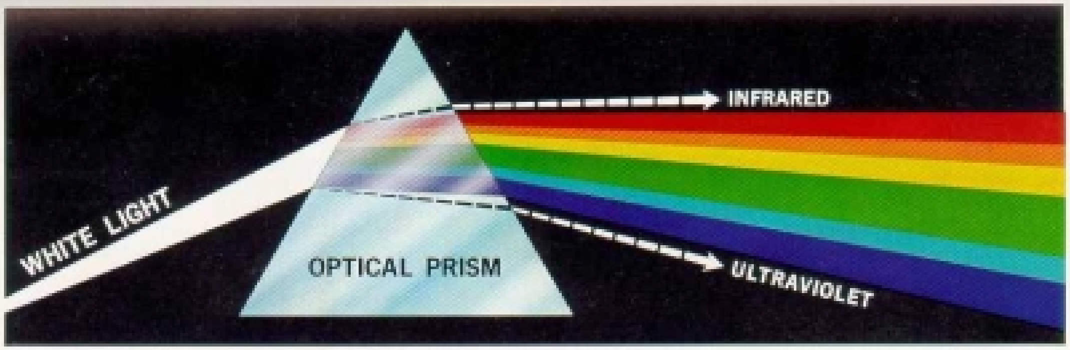


Plate I. Color spectrum seen by passing white light through a prism. (Courtesy of General Electric Co., Lamp Business Division.)

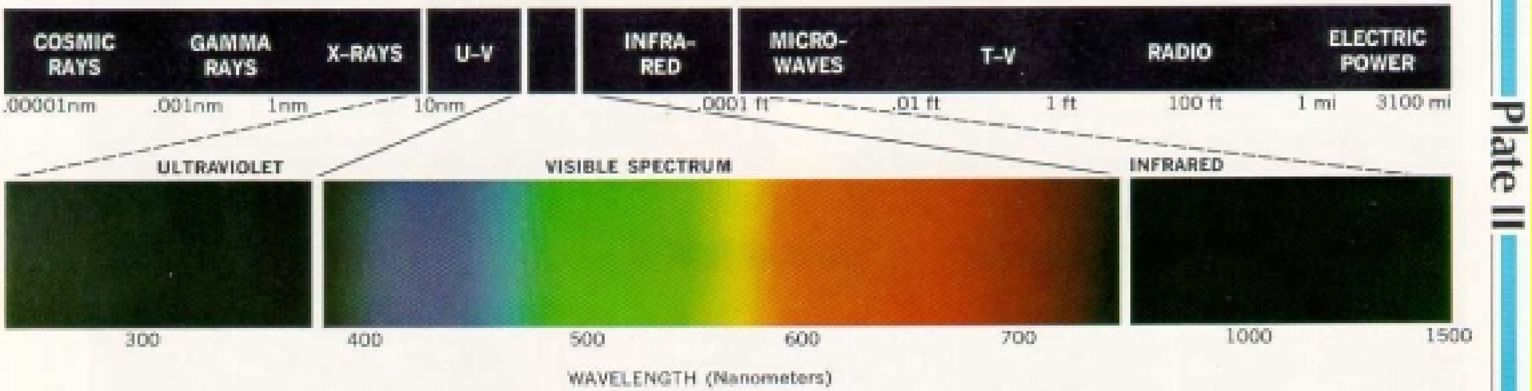


Plate II. A section of the electromagnetic energy spectrum showing the range of wavelengths comprising the visible spectrum. (Courtesy of General Electric Co., Lamp Business Division.)

From [Gonzalez92]

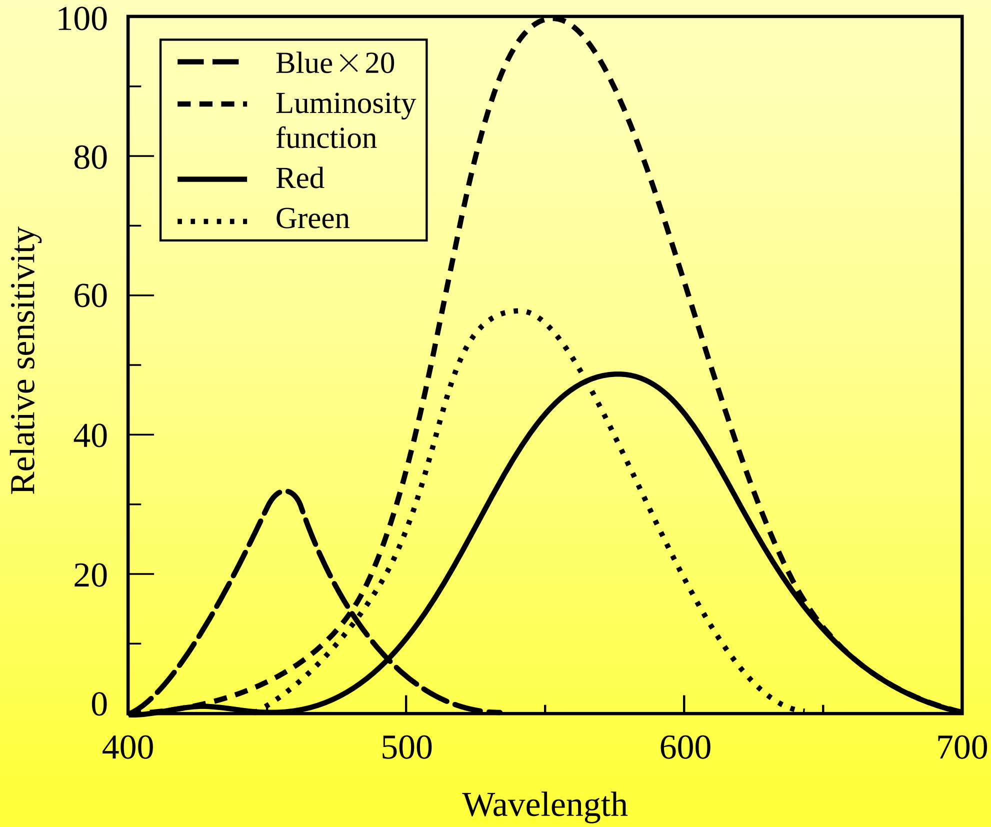
Illuminating and Reflecting Light

- Illuminating sources:
  - emit light (e.g. the sun, light bulb, TV monitors)
  - perceived color depends on the emitted freq.
  - follows additive rule
    - R+G+B=White
- Reflecting sources:
  - reflect an incoming light (e.g. the color dye, matte surface, cloth)
  - perceived color depends on reflected freq (=emitted freq-absorbed freq.)
  - follows subtractive rule
    - R+G+B=Black

Human Perception of Color

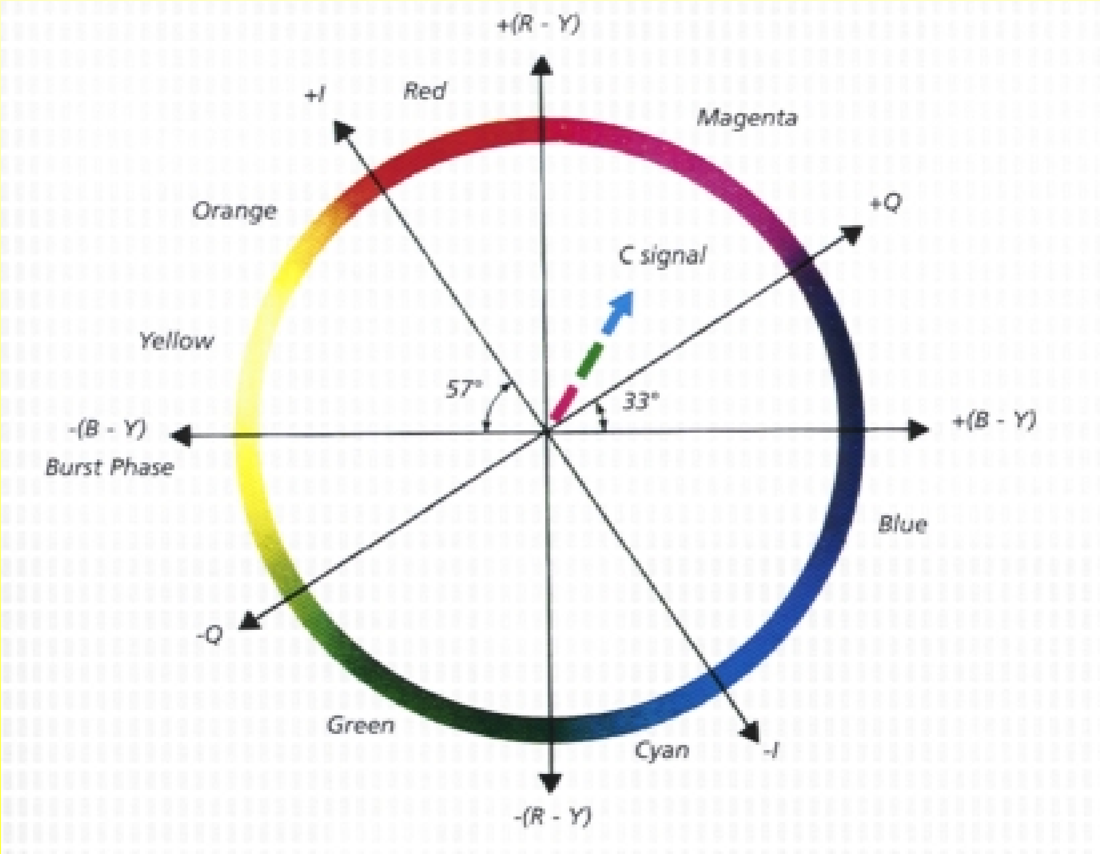
- Retina contains photo receptors
  - Cones: day vision, can perceive color tone
    - Red, green, and blue cones
    - Tri-receptor theory of color vision [Young1802]
  - Rods: night vision, perceive brightness only
- Color sensation is characterized by
  - Luminance (brightness)
  - Chrominance
    - Hue (color tone)
    - Saturation (color purity)

Frequency Responses of Cones and the Luminous Efficiency Function



$$C_i = \int C(\lambda) a_i(\lambda) d\lambda, \quad i = r, g, b, y$$

Color Hue Specification



Trichromatic Color Mixing

- Trichromatic color mixing theory
  - Any color can be obtained by mixing three primary colors with a right proportion

$$C = \sum_{k=1,2,3} T_k C_k, \quad T_k : \text{Tristimulus values}$$

- Primary colors for illuminating sources:
  - Red, Green, Blue (RGB)
  - Color monitor works by exciting red, green, blue phosphors using separate electronic guns
- Primary colors for reflecting sources (also known as secondary colors):
  - Cyan, Magenta, Yellow (CMY)
  - Color printer works by using cyan, magenta, yellow and black (CMYK) dyes

Color Mixing

Plate III

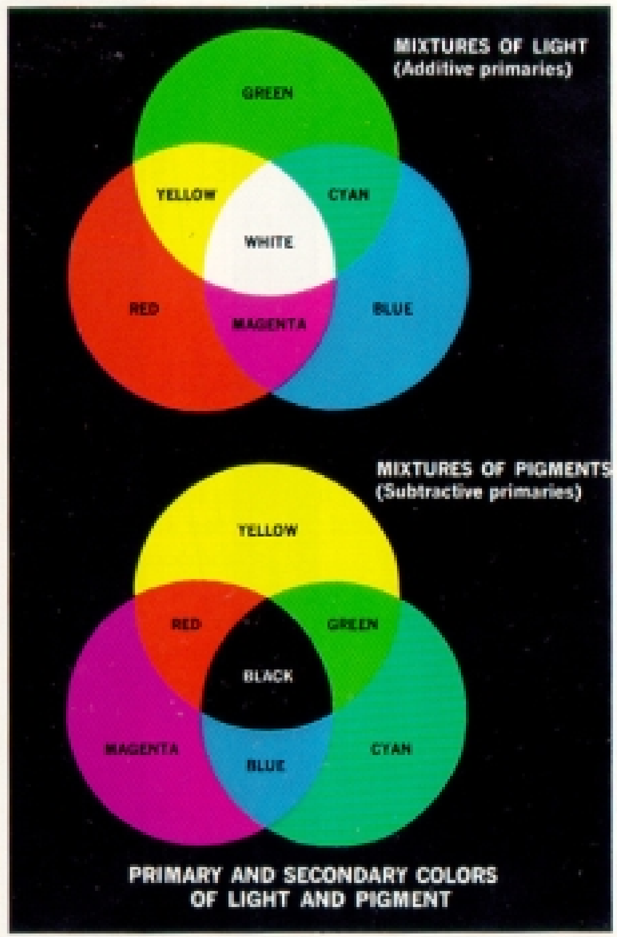


Plate III. Primary and secondary colors of light and pigments. (Courtesy of General Electric Co., Lamp Business Division.)

From [Gonzalez92]

red

Green

Blue

Color Representation Models

- Specify the tristimulus values associated with the three primary colors
  - RGB
  - CMY
- Specify the luminance and chrominance
  - HSI (Hue, saturation, intensity)
  - YIQ (used in NTSC color TV)
  - YCbCr (used in digital color TV)
- Amplitude specification:
  - 8 bits for each color component, or 24 bits total for each pixel
  - Total of 16 million colors
  - A true RGB color display of size 1Kx1K requires a display buffer memory size of 3 MB

Color Coordinate Conversion

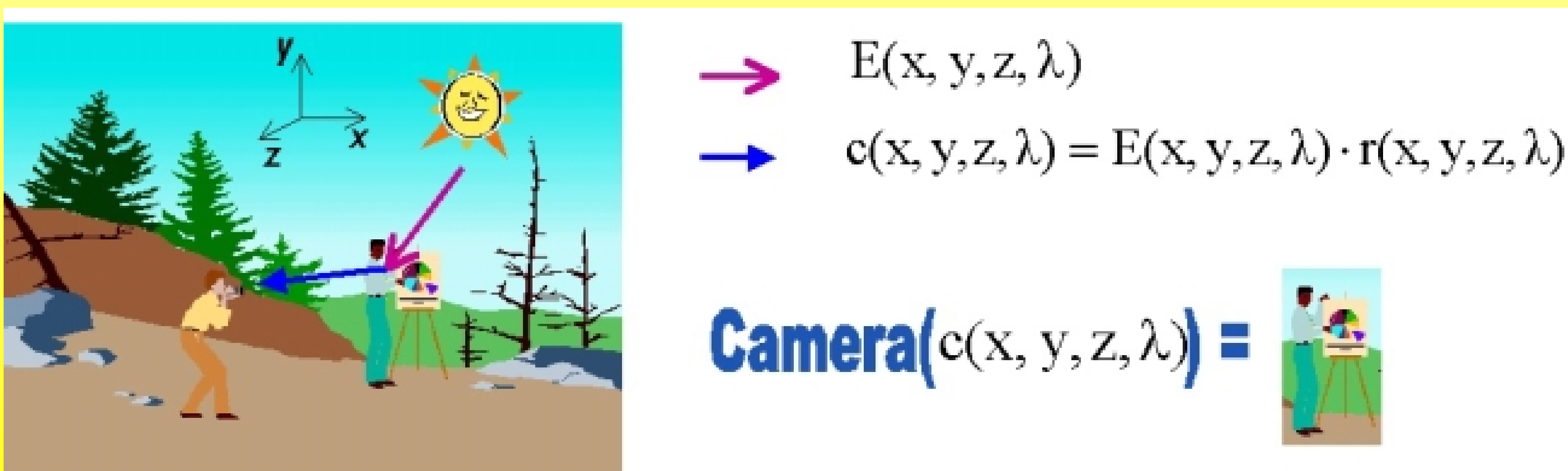
- Conversion between different primary sets are linear (3x3 matrix)
- Conversion between primary and XYZ/YIQ/YUV are also linear
- Conversion to LSI/Lab are nonlinear
  – LSI and Lab coordinates
    - coordinate Euclidean distance proportional to actual color difference
- Conversion formulae between many color coordinates can be found in [Gonzalez92]

Video Capture and Display

- Light reflection physics
- Imaging operator
- Color capture
- Color display
- Component vs. composite video

Video Capture

- For natural images we need a light source *(λ: wavelength of the source)* .
  - $E(x, y, z, \lambda)$: incident light on a point *(x, y, z world coordinates of the point)*
- Each point in the scene has a reflectivity function.
  - $r(x, y, z, \lambda)$: reflectivity function
- Light reflects from a point and the reflected light is captured by an imaging device.
  - $c(x, y, z, \lambda) = E(x, y, z, \lambda) \times r(x, y, z, \lambda)$: reflected light.



Courtesy of Onur Guleryuz

More on Video Capture

- Reflected light to camera
  - Camera absorption function

$$\bar{\psi}(\mathbf{X},t)=\int C(\mathbf{X},t,\lambda)a_c(\lambda)d\lambda$$

  - Projection from 3-D to 2-D

$$\mathbf{X}\underset{P}{\rightarrow}\mathbf{x}$$

$$\psi(P(\mathbf{X}),t)=\bar{\psi}(\mathbf{X},t) \quad \text{or} \quad \psi(\mathbf{x},t)=\bar{\psi}(P^{-1}(\mathbf{X}),t)$$

    - The projection operator is non-linear
      - Perspective projection
      - Othographic projection

How to Capture Color

- Need three types of sensors
- Complicated digital processings are incorporated in advanced cameras

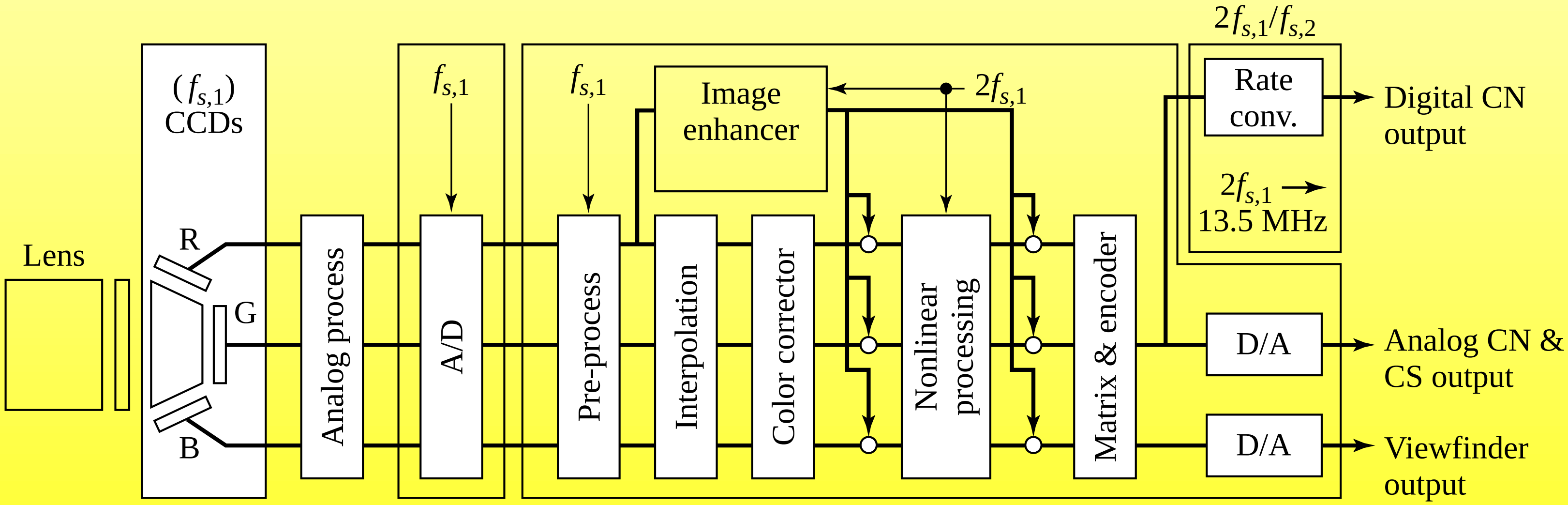


Figure 1.2 Schematic block diagram of a professional color video camera. Reprinted from Y. Hashimoto, M. Yamamoto, and T. Asaida, Cameras and display systems, *IEEE* (July 1995), 83(7):1032–43. Copyright 1995 IEEE.

Video Display

- CRT vs LCD
- Need three light sources projecting red, green, blue components respectively

Component vs. Composite Video

- Component video
  - Three color components captured/saved/displayed separately
  - Stored in either RGB or YIQ (YCrCb)
  - New digital video format
- Composite video
  - Convert RGB to YIQ (YUV)
  - Multiplexing YIQ into a single signal
  - Chrominance components are band-limited and multiplexed to the upper band of Y
  - Used in all analog video devices (analog TV, VHS tapes, etc.)

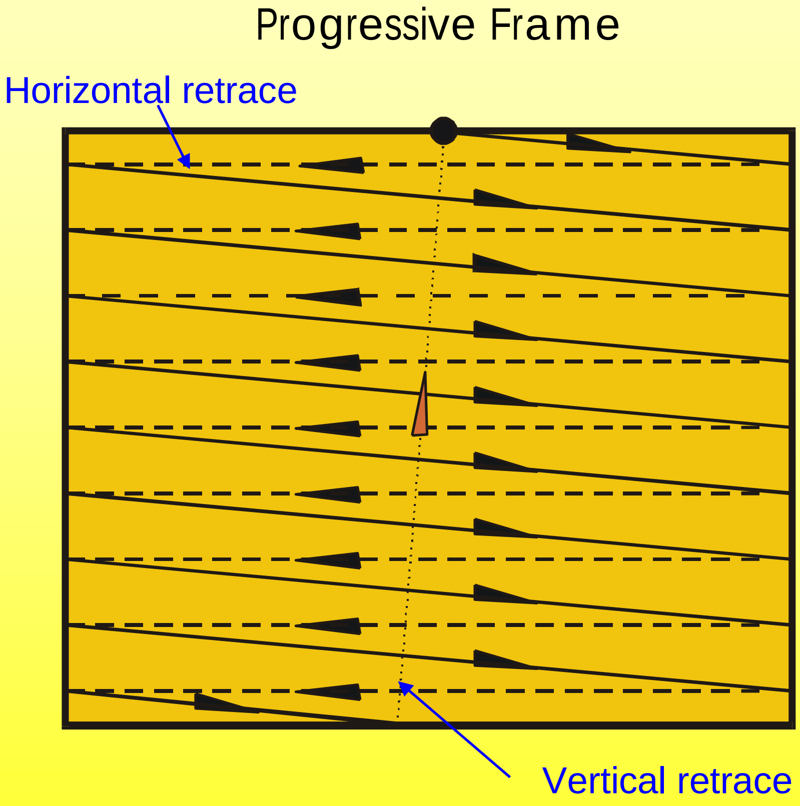
Analog Video

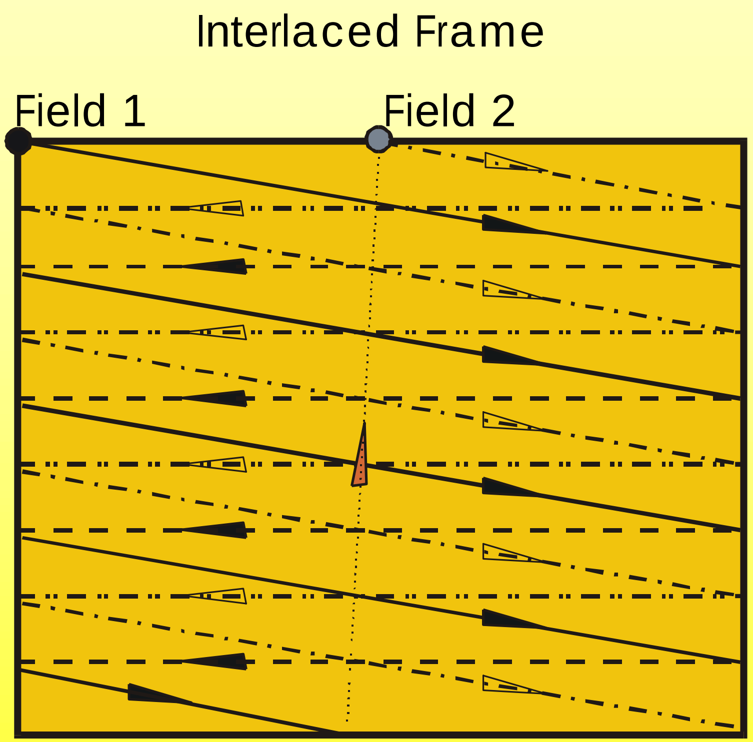
- Video raster
- Progressive vs. interlaced raster
- Analog TV systems

Raster Scan

- Real-world scene is a continuous 3-D signal (temporal, horizontal, vertical)
- Analog video is stored in the raster format
  - Sampling in time: consecutive sets of frames
    - To render motion properly, >=30 frame/s is needed
  - Sampling in vertical direction: a frame is represented by a set of scan lines
    - Number of lines depends on maximum vertical frequency and viewing distance, 525 lines in the NTSC system
  - Video-raster = 1-D signal consisting of scan lines from successive frames

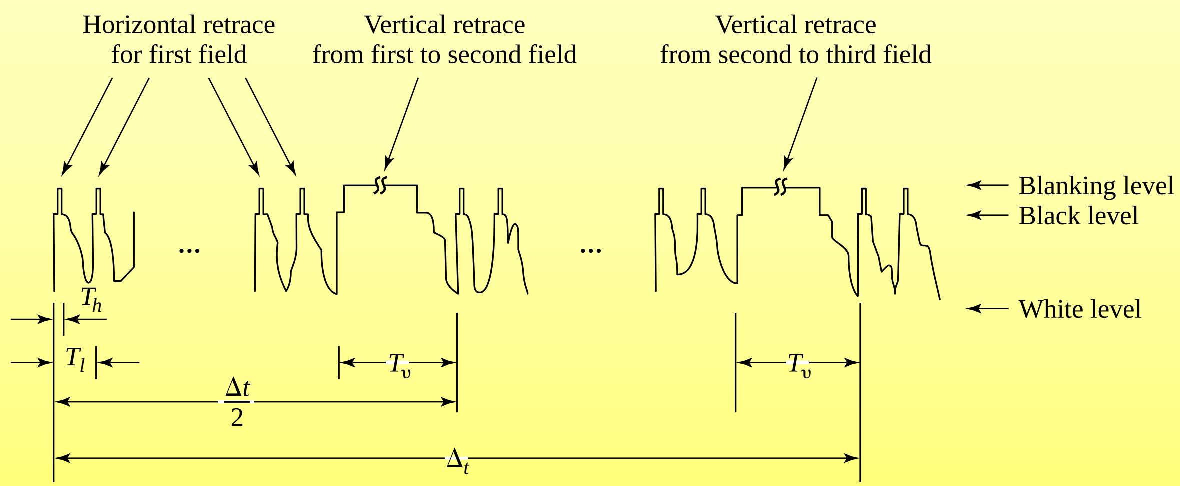
Progressive and Interlaced Scans



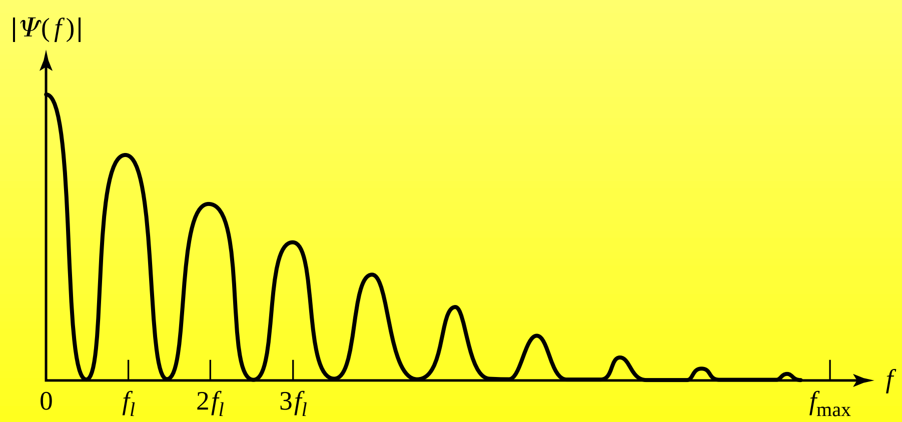


Interlaced scan is developed to provide a trade-off between temporal and vertical resolution, for a given, fixed data rate (number of line/sec).

Waveform and Spectrum of an Interlaced Raster

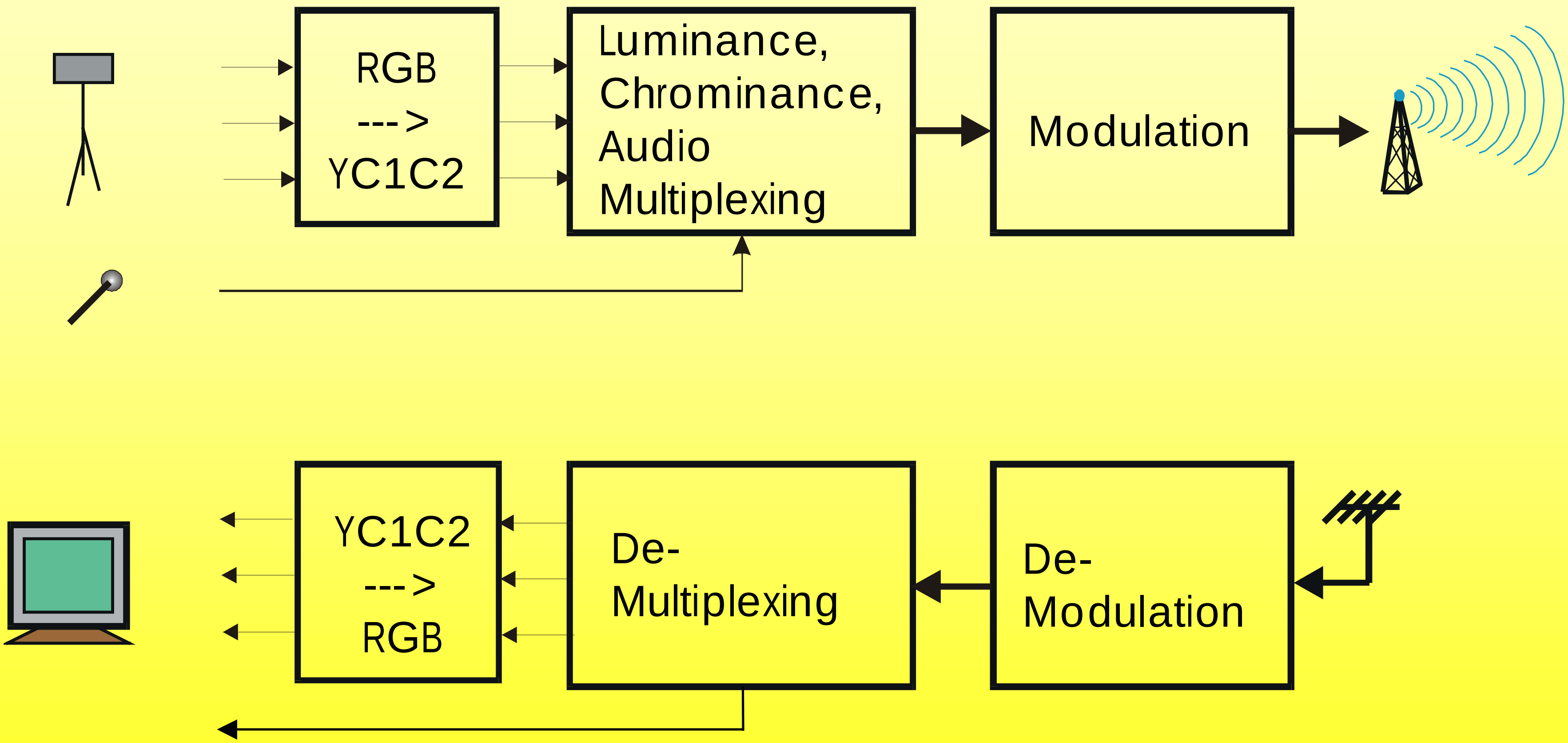


(a)



(b)

Color TV Broadcasting and Receiving



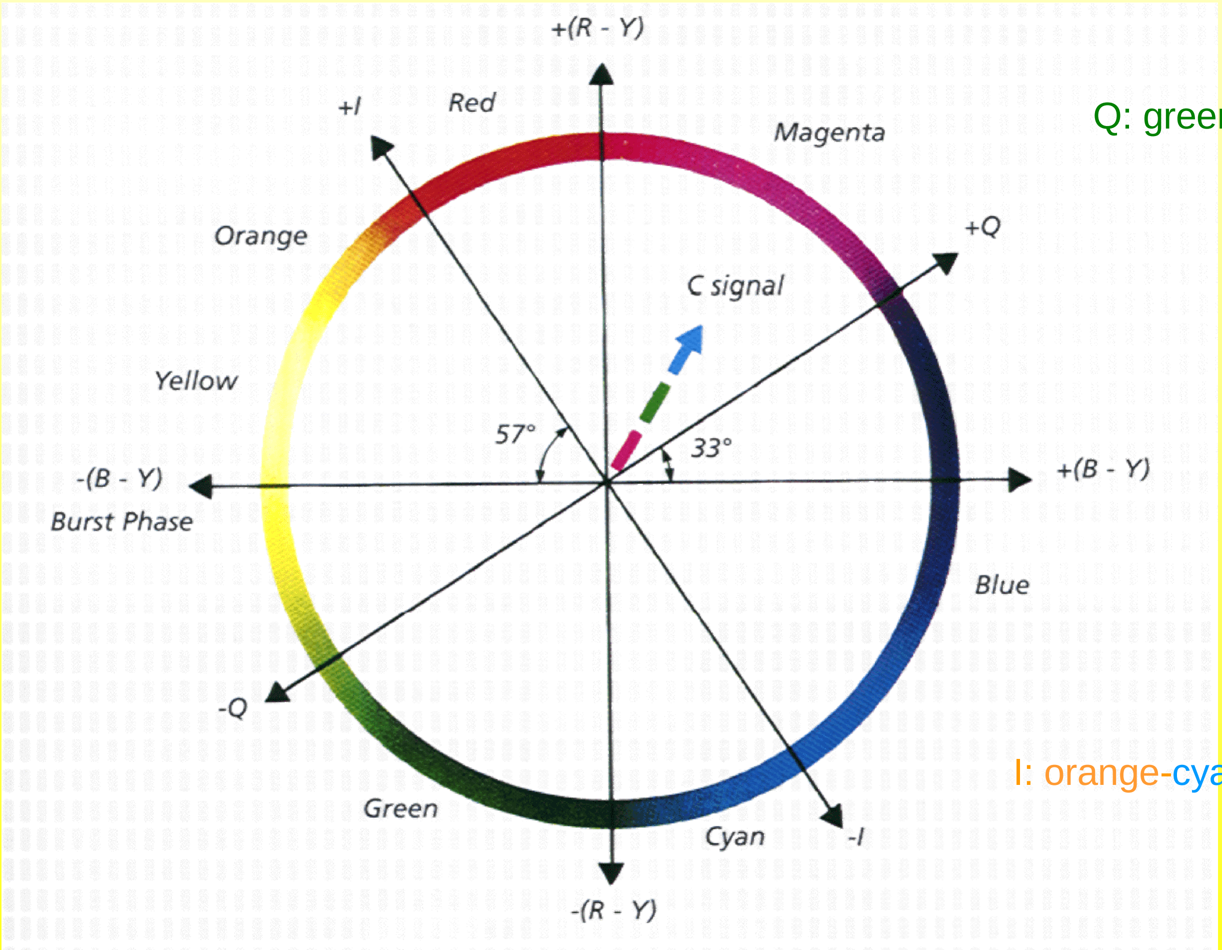
Different Color TV Systems

Parameters	NTSC	PAL	SECAM
Field Rate (Hz)	59.95 (60)	50	50
Line Number/Frame	525	625	625
Line Rate (Line/s)	15,750	15,625	15,625
Color Coordinate	YIQ	YUV	YDbDr
Luminance Bandwidth (MHz)	4.2	5.0/5.5	6.0
Chrominance Bandwidth (MHz)	1.5(I)/0.5(Q)	1.3(U,V)	1.0 (U,V)
Color Subcarrier (MHz)	3.58	4.43	4.25(Db),4.41(Dr)
Color Modulation	QAM	QAM	FM
Audio Subcarrier	4.5	5.5/6.0	6.5
Total Bandwidth (MHz)	6.0	7.0/8.0	8.0

Why not using RGB directly?

- R,G,B components are correlated
  - Transmitting R,G,B components separately is redundant
  - More efficient use of bandwidth is desired
- RGB->YC1C2 transformation
  - Decorrelating: Y,C1,C2 are uncorrelated
  - C1 and C2 require lower bandwidth
  - Y (luminance) component can be received by B/W TV sets
- YIQ in NTSC
  - I: orange-to-cyan
  - Q: green-to-purple (human eye is less sensitive)
    - Q can be further bandlimited than I
  - Phase=Arctan(Q/I) = hue, Magnitude=sqrt (I^2+Q^2) = saturation
  - Hue is better retained than saturation

I and Q on the color circle



Q: green-purple

I: orange-cyan

Conversion between RGB and YIQ

- RGB -> YIQ

$$Y = 0.299\ R + 0.587\ G + 0.114\ B$$
$$I = 0.596\ R - 0.275\ G - 0.321\ B$$
$$Q = 0.212\ R - 0.523\ G + 0.311\ B$$

- YIQ -> RGB

$$R = 1.0\ Y + 0.956\ I + 0.620\ Q,$$
$$G = 1.0\ Y - 0.272\ I - 0.647\ Q,$$
$$B = 1.0\ Y - 1.108\ I + 1.700\ Q.$$

TV signal bandwidth

- Luminance
  - Maximum vertical frequency (cycles/picture-height)= black and white lines interlacing

$$f_{v,\max} = Kf'_{s,y} / 2$$

  - Maximum horizontal frequency (cycles/picture-width)

$$f_{h,\max} = f_{v,\max} \cdot \mathrm{IAR}$$

  - Corresponding temporal frequency (cycles/second or Hz)

$$f_{\max} = f_{h,\max} / T'_l = \mathrm{IAR} \cdot Kf'_{s,y} / 2T'_l$$

  - For NTSC, $f_{\max} = 4.2\,\mathrm{MHz}$

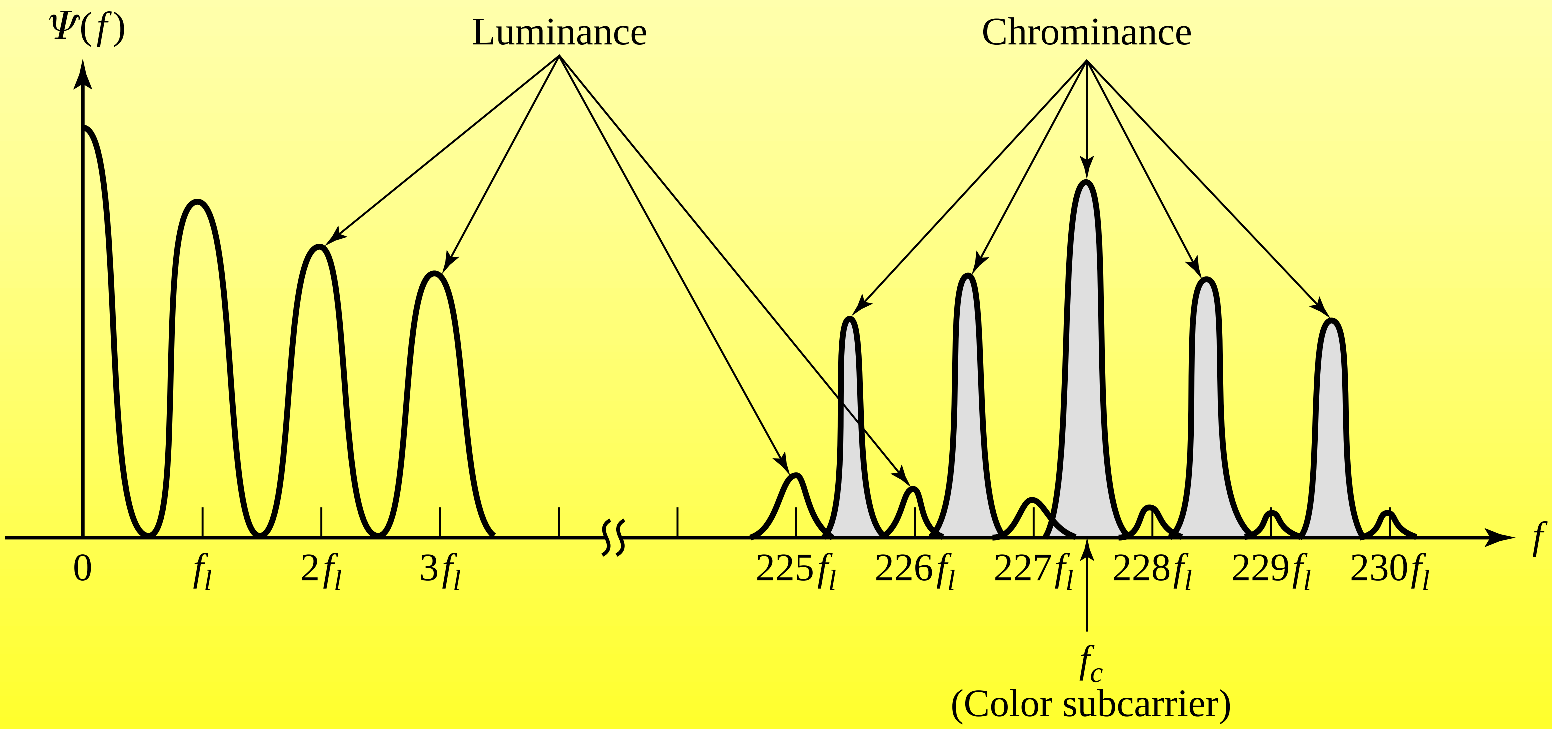
- Chrominance
  - Can be bandlimited significantly
    - I: 1.5 MHz, Q: 0.5 MHz.

Multiplexing of Luminance and Chrominance

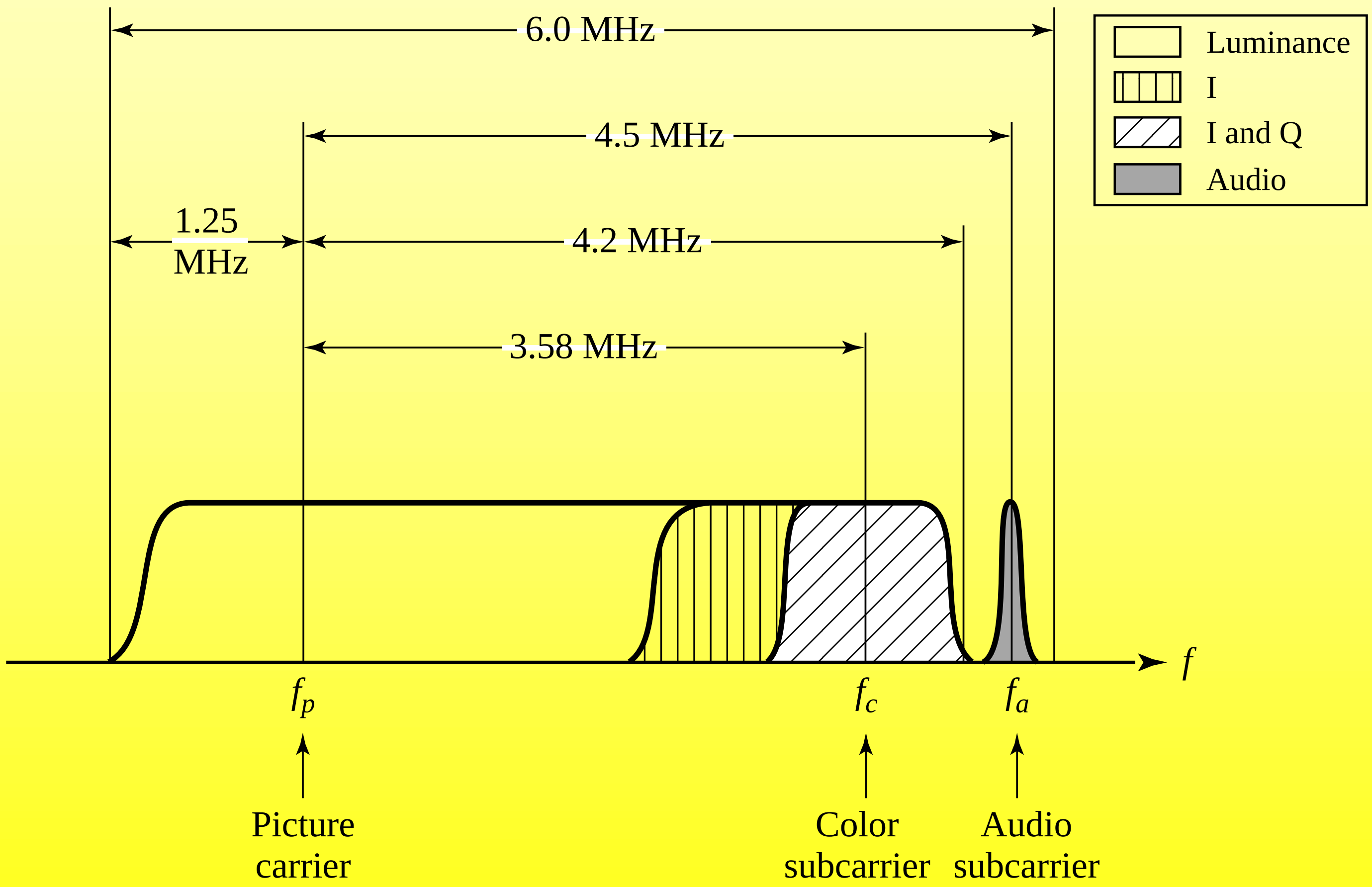
- Chrominance signal can be bandlimited
  – it usually has a narrower frequency span than the luminance and the human eye is less sensitive to high frequencies in chrominance
- Position the bandlimited chrominance at the high end spectrum of the luminance, where the luminance is weak
- The actual position should be such that the peaks of chrominance spectrum interlace with those of the luminance

$$f_c = 455 f_l / 2 \ \ (= 3.58 \ \text{Hz for NTSC})$$

Spectrum Illustration



Multiplexing of luminance, chrominance and audio (Composite Video Spectrum)



Digital Video

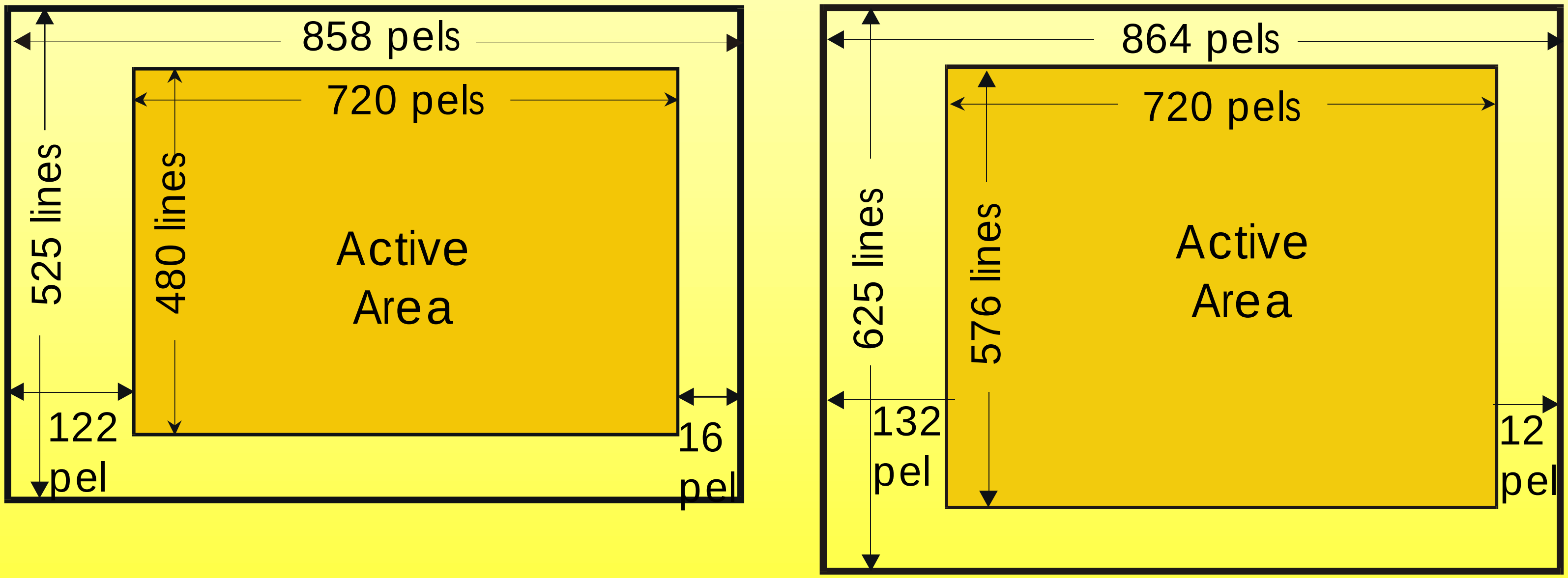
- Digital video by sampling/quantizing analog video raster → BT.601 video
- Other digital video formats and their applications

Digitizing A Raster Video

- Sample the raster waveform = Sample along the horizontal direction
- Sampling rate must be chosen properly
  - For the samples to be aligned vertically, the sampling rate should be multiples of the line rate
  - Horizontal sampling interval = vertical sampling interval
  - Total sampling rate equal among different systems

$$f_s = 858 f_l \text{(NTSC)} = 864 f_l \text{(PAL/SECAM)} = 13.5 \text{ MHz}$$

BT.601* Video Format



525/60: 60 field/s

625/50: 50 field/s

* BT.601 is formerly known as CCIR601

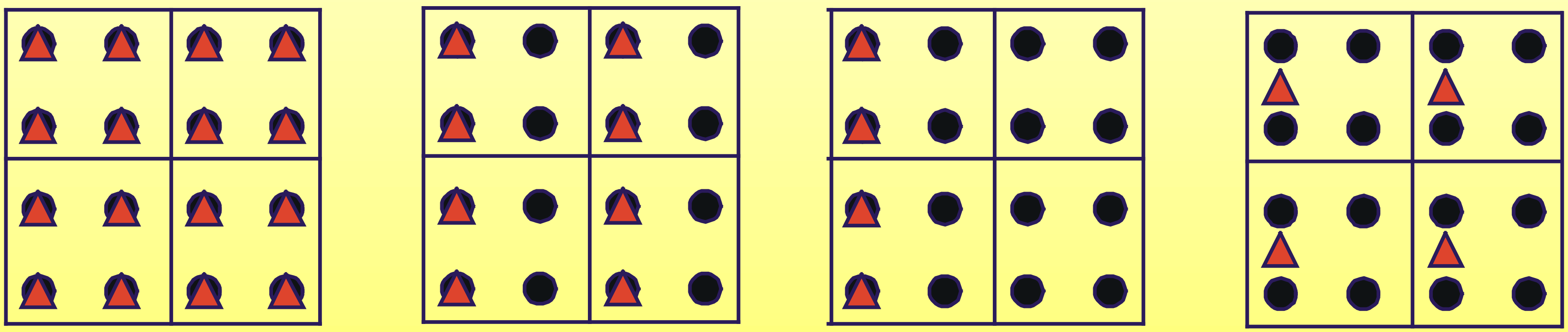
RGB <--> YCbCr

$$Y\_d = 0.257\, R\_d + 0.504\, G\_d + 0.098\, B\_d + 16,$$
$$C\_b = -0.148\, R\_d - 0.291\, G\_d + 0.439\, B\_d + 128,$$
$$C\_r = 0.439\, R\_d - 0.368\, G\_d - 0.071\, B\_d + 128,$$

$$R\_d = 1.164\, Y\_d' + 0.0\, C\_b' + 1.596\, C\_r',$$
$$G\_d = 1.164\, Y\_d' - 0.392\, C\_b' - 0.813\, C\_r',$$
$$B\_d = 1.164\, Y\_d' + 2.017\, C\_b' + 0.0\, C\_r',$$

$$Y\_d' = Y\_d - 16,\ C\_b' = C\_b - 128,\ C\_r' = C\_r - 128$$

Chrominance Subsampling Formats



4:4:4
For every 2x2 Y Pixels
4 Cb & 4 Cr Pixel
(No subsampling)

4:2:2
For every 2x2 Y Pixels
2 Cb & 2 Cr Pixel
(Subsampling by 2:1 horizontally only)

4:1:1
For every 4x1 Y Pixels
1 Cb & 1 Cr Pixel
(Subsampling by 4:1 horizontally only)

4:2:0
For every 2x2 Y Pixels
1 Cb & 1 Cr Pixel
(Subsampling by 2:1 both horizontally and vertically)

● Y Pixel

▲ Cb and Cr Pixel

Digital Video Formats

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