Digital Video Processing and Analysis

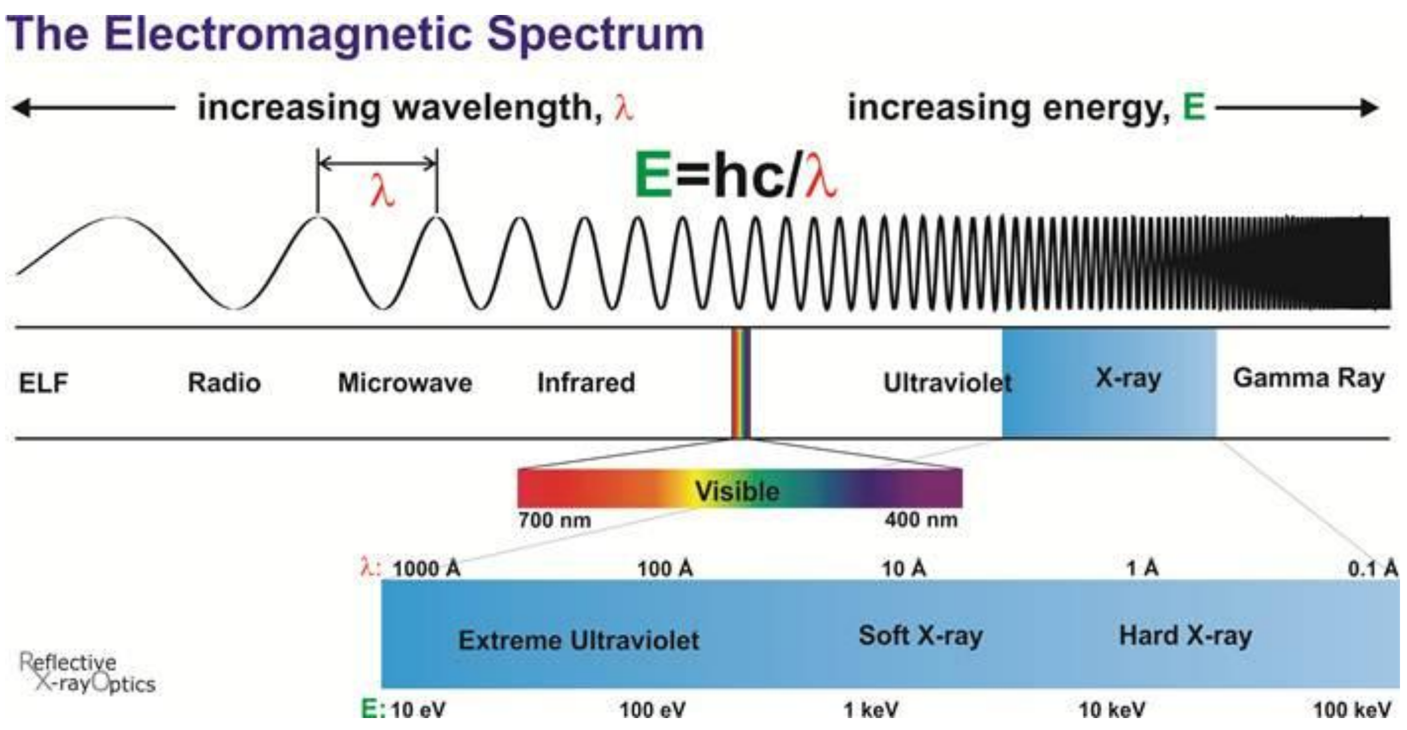
: Video Basics

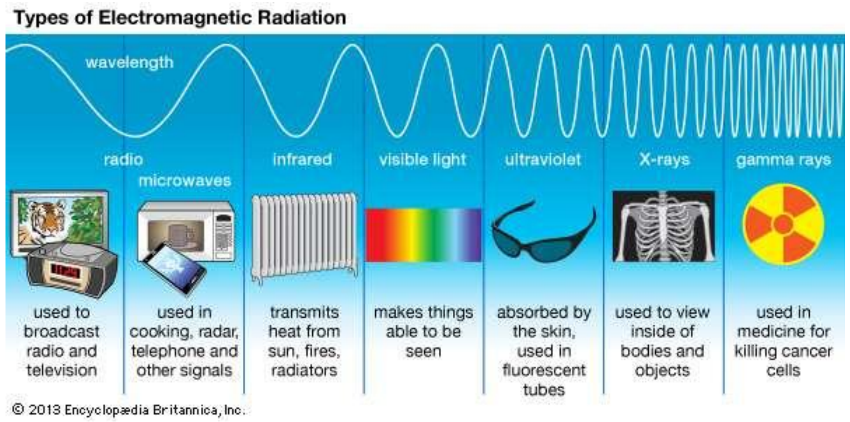
Cheolkon Jung

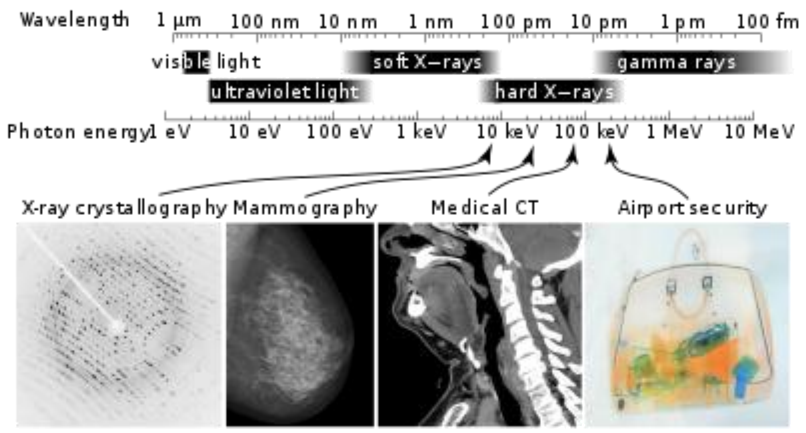
School of Electronic Engineering

Xidian University, China

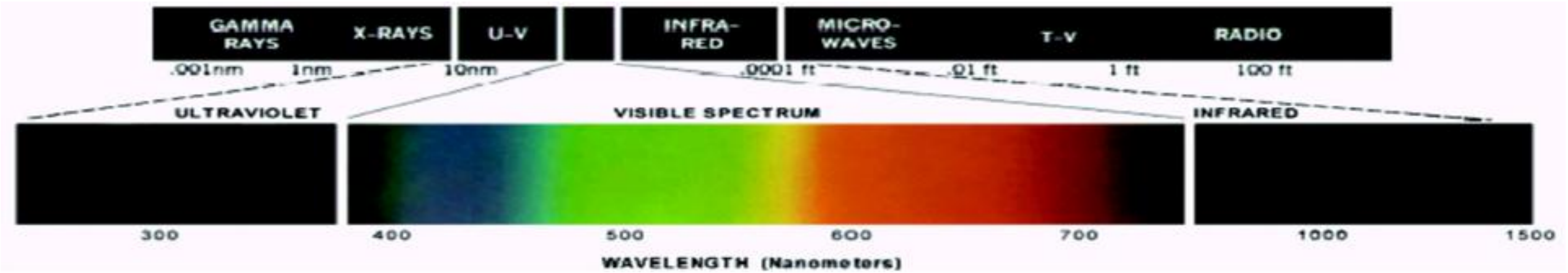
Light and Color







Light and Color



- Visible light is part of electromagnetic wave
- Light intensity distribution: a_i: frequency response

$C(\mathbf{X}, t, \lambda)$
Specifies light intensity
at wavelength λ, spatial location $\mathbf{X}$, and time t

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

Red: 570 nm, Green: 535 nm, Blue: 445 nm

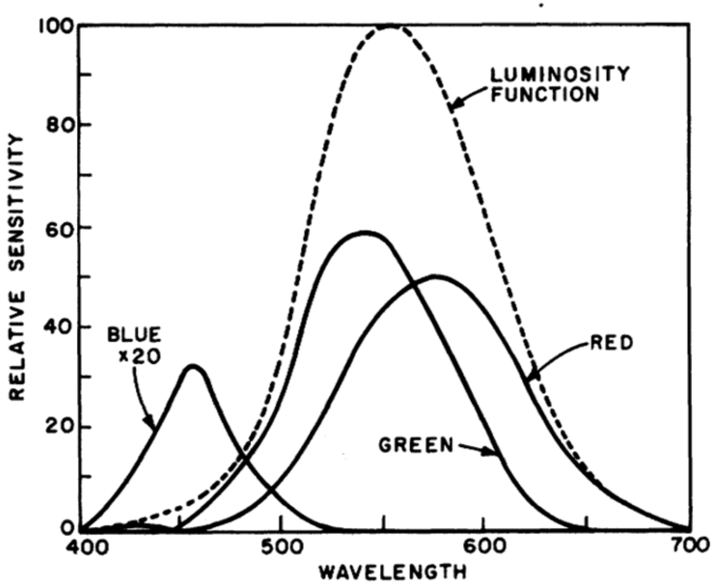
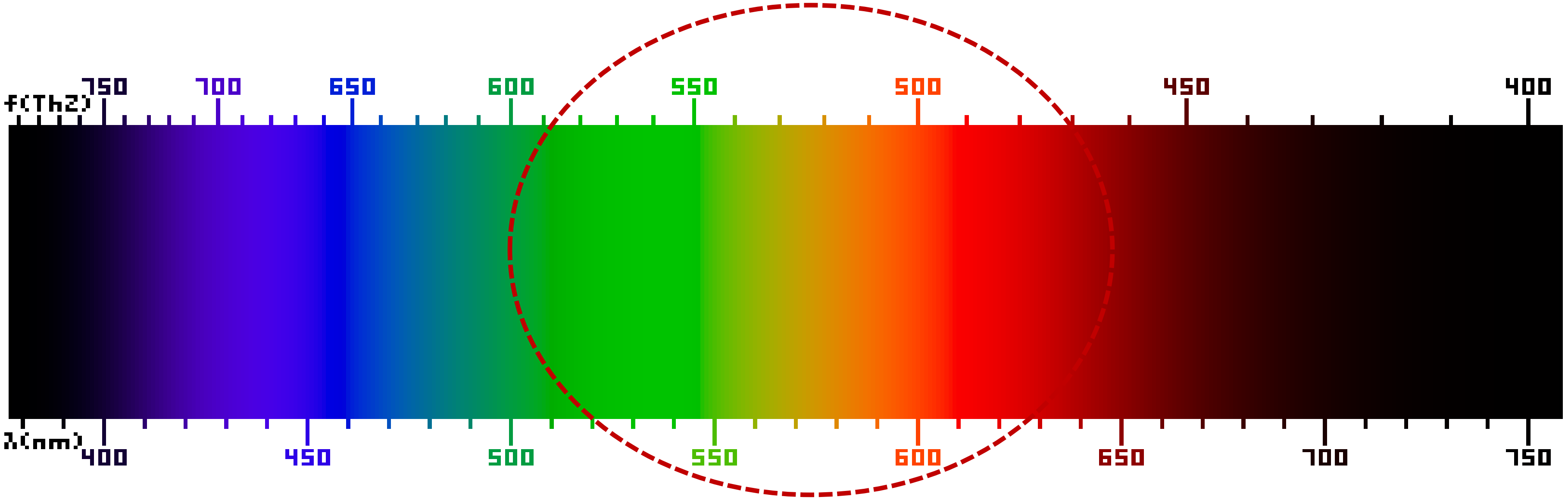


Fig. 1
Trireceptor eye response. The sum of these response characteristics constitutes the total sensitivity, or luminosity, response curve as indicated by the dotted line (reproduced for clarity in Fig. 2). Note that the blue response curve is magnified in Fig. 1 by a factor of twenty.

Light and Color

- Luminous efficiency function: the average spectral sensitivity of human visual perception of brightness.
- Luminance: a photometric measure of the luminous intensity per unit area of light, related to the incoming light spectrum

$$Y = \int C(\lambda) a_y(\lambda) d\lambda$$



Light Sources

- 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.

$$E(x, y, z, \lambda)$$

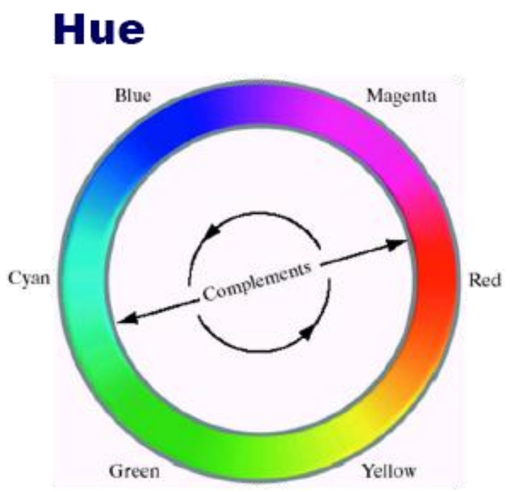
$$c(x, y, z, \lambda) = E(x, y, z, \lambda) \cdot r(x, y, z, \lambda)$$

Camera$(c(x, y, z, \lambda))$ = 

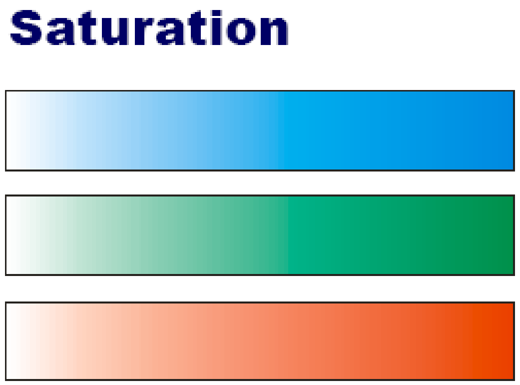
Courtesy of Onur Guleryuz

Human Perception of Colors

- Color sensation is characterized by
  - **Luminance** (brightness)
  - **Chrominance**
    - Hue (color tone) – informally called color
    - Saturation (color purity)



- Retina contains photo receptors
  - *Cones*: **day vision**, perceive color tone
    - Red, green, and blue cones
    - Tri-receptor theory of color vision
  - *Rods*: **night vision**, perceive brightness only



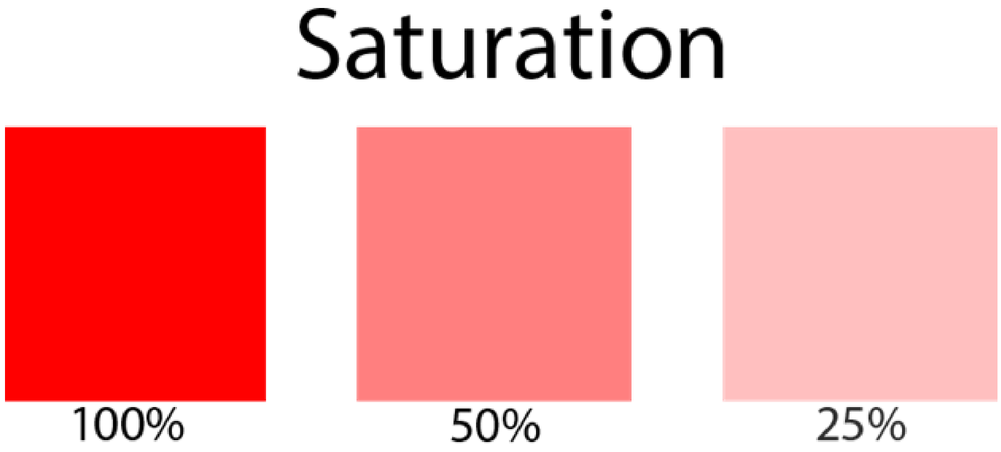
Human Perception of Colors

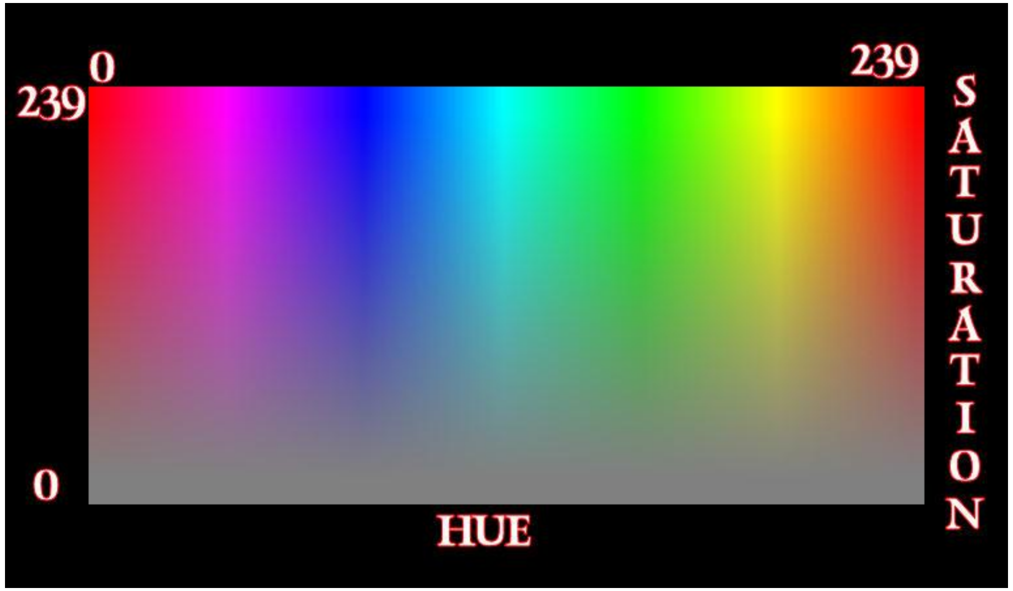

No Saturation


Low Saturation


Full Saturation

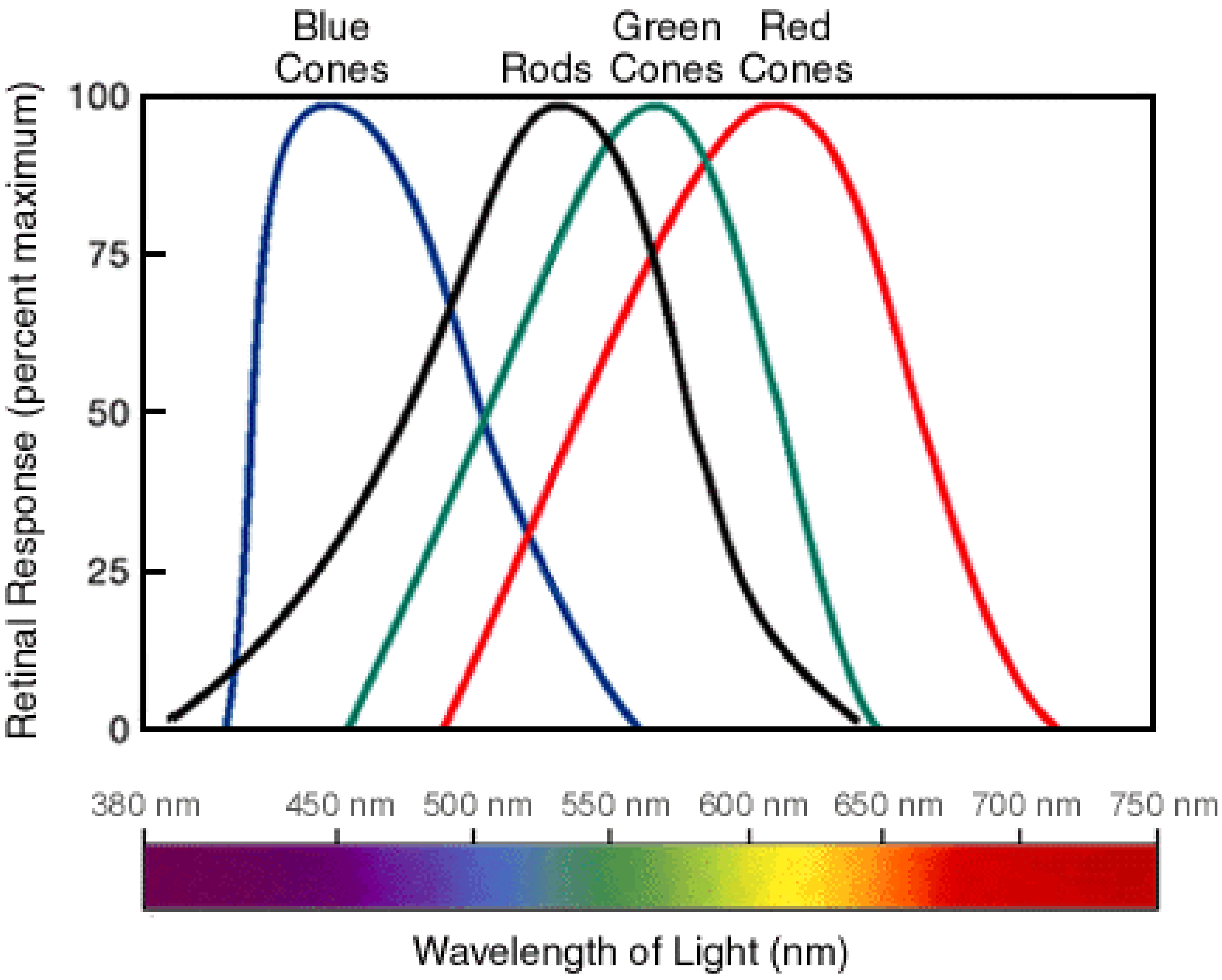

Over Saturation



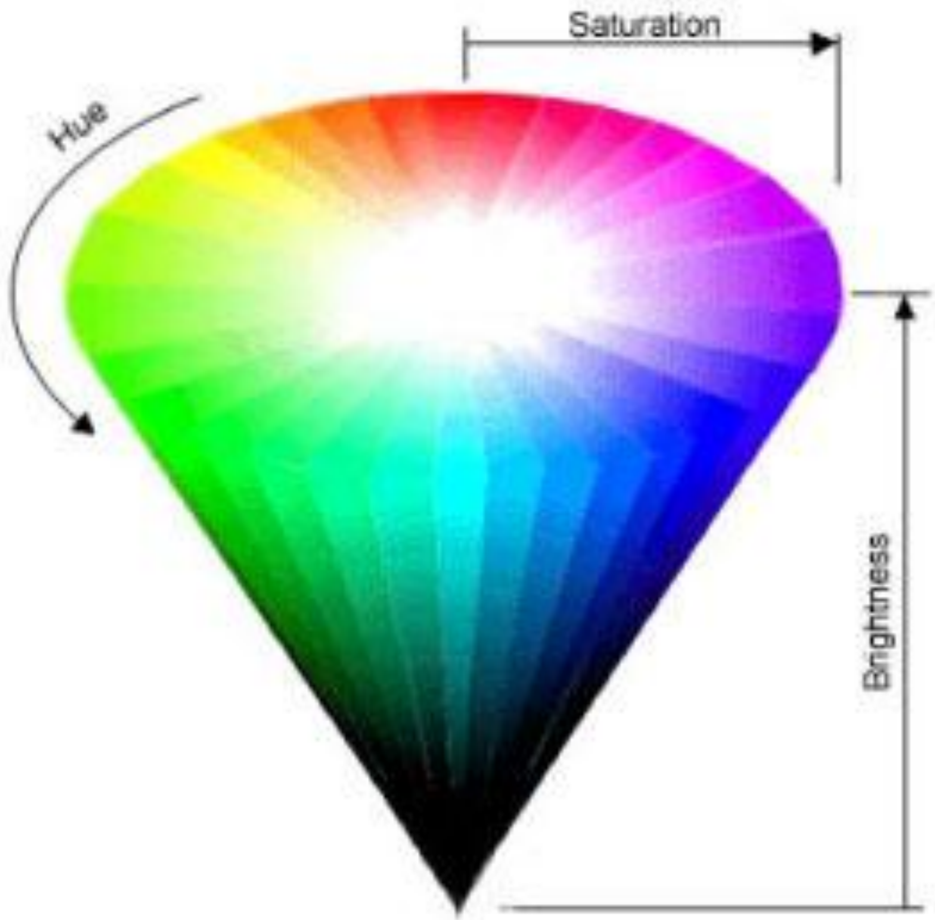


Human Perception of Colors

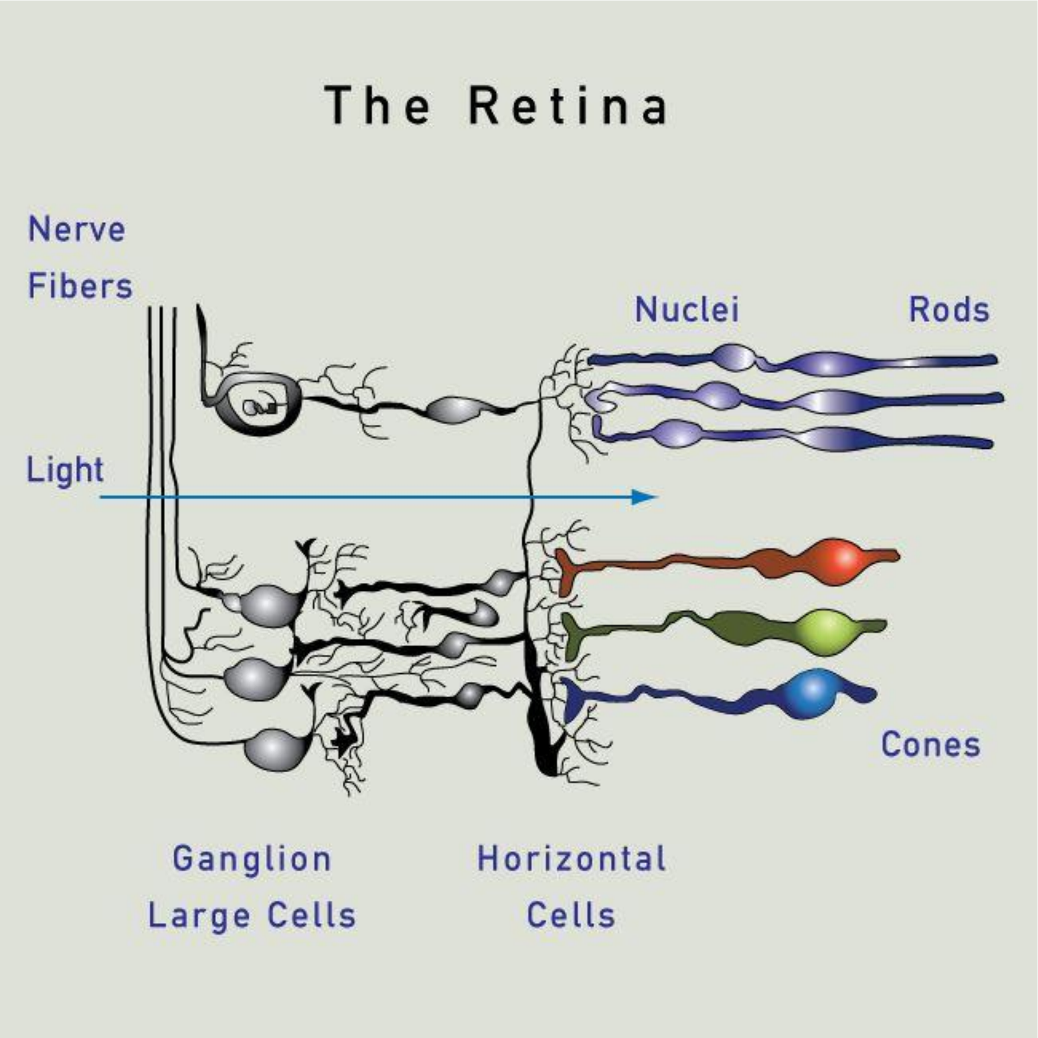
Thee photo receptors

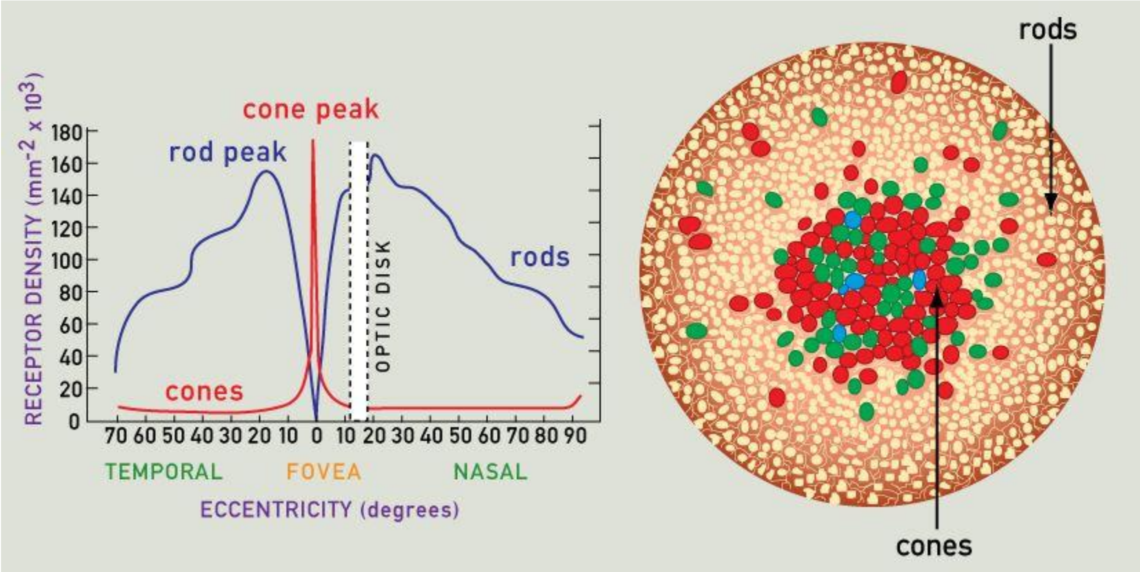


HSI color space



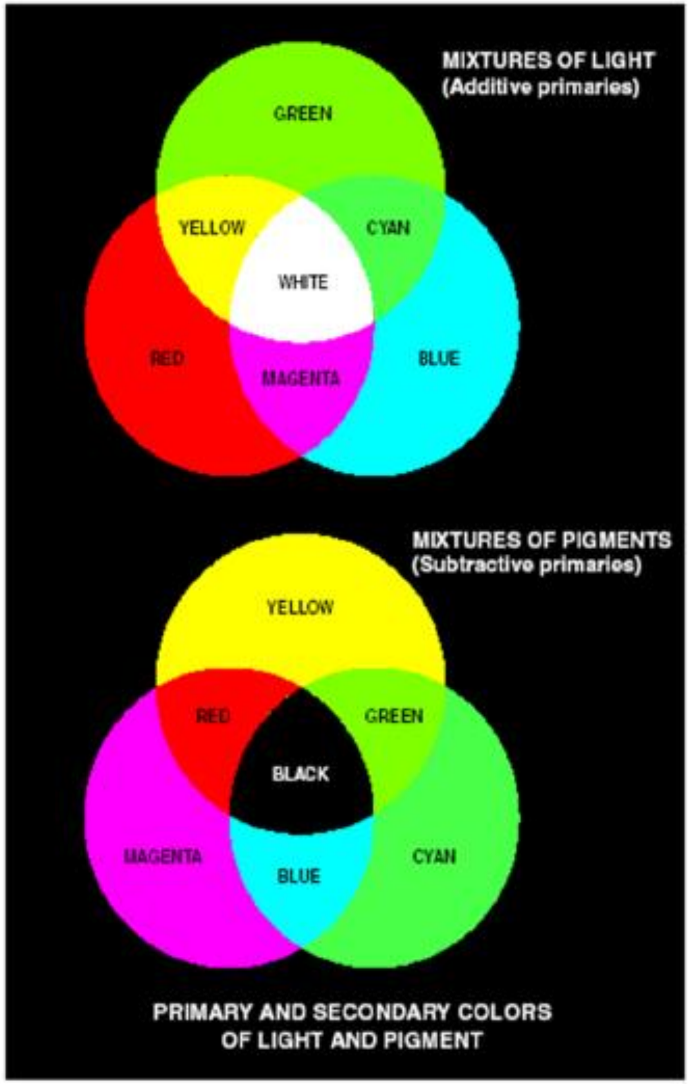
Human Perception of Colors





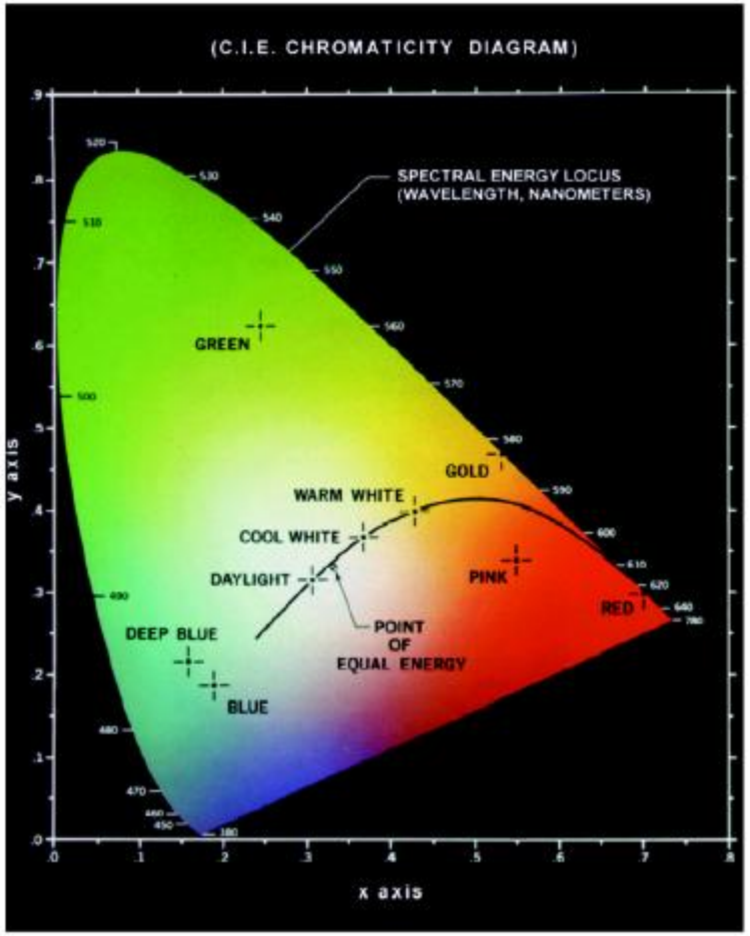
Illuminating and Reflecting Light Sources

- Illuminating sources:
  - Emit light
    - e.g. sun, light bulb, TV monitors
  - Perceived color depends on light intensity distribution
  - Follow *additive rule*
    - R+G+B=White
- Reflecting sources:
  - Reflect incoming light
    - e.g. moon, color dye, cloth
  - Perceived color is given by incoming color – absorbed color
  - Follow *subtractive rule*
    - cyan absorbs red
    - magenta absorbs green
    - yellow absorbs blue
    - C+M+Y=Black



Illuminating and Reflecting Light Sources

- **Trichromatic color mixing theory**
  - Most colors can be obtained by mixing three primary colors with a right proportion
- Primary colors for illuminating sources
  - Red, Green, Blue (RGB)
  - **Color monitor** works by exciting red, green, blue phosphors using separate electronic guns
  - Most display systems use RGB primary
- Primary colors for reflecting sources
  - Cyan, Magenta, Yellow (CMY)
  - **Color printer** works by using cyan, magenta, yellow and black (CMYK) inks
  - Can render the black color more truthfully



RGB Color Mixing

red

Green

Blue

Color Representations

- Specify tristimulus values associated with three primary colors
  - RGB
  - CMY

  => Color production
- Specify the luminance and chrominance
  - HSI (Hue, saturation, intensity)
  - YIQ (used in NTSC color TV)
  - YCbCr (used in digital color TV)

  => Defining other primaries and numerical specification of color
- 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 1000x1000 requires a display buffer, whose memory size is 3 MB

Video Imaging Principles

- Light intensity distribution in the 3D world:

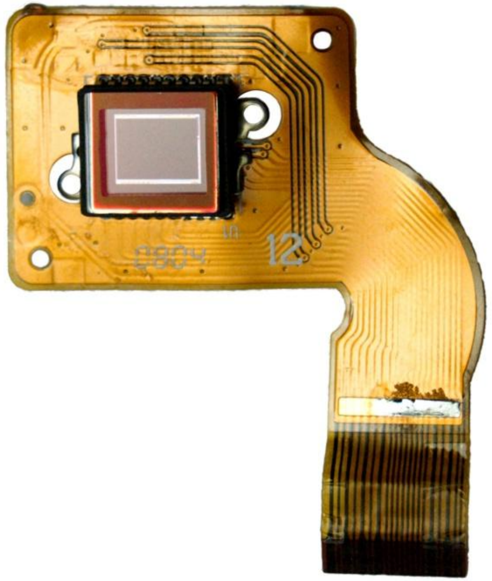
$$\bar{\psi}(\mathbf{X}, t) = \int_0^\infty C(\mathbf{X}, t, \lambda) a_c(\lambda) d\lambda$$

- Projected image from the 3D image (**X**: 3D coordinate; **x**: 2D coordinate):

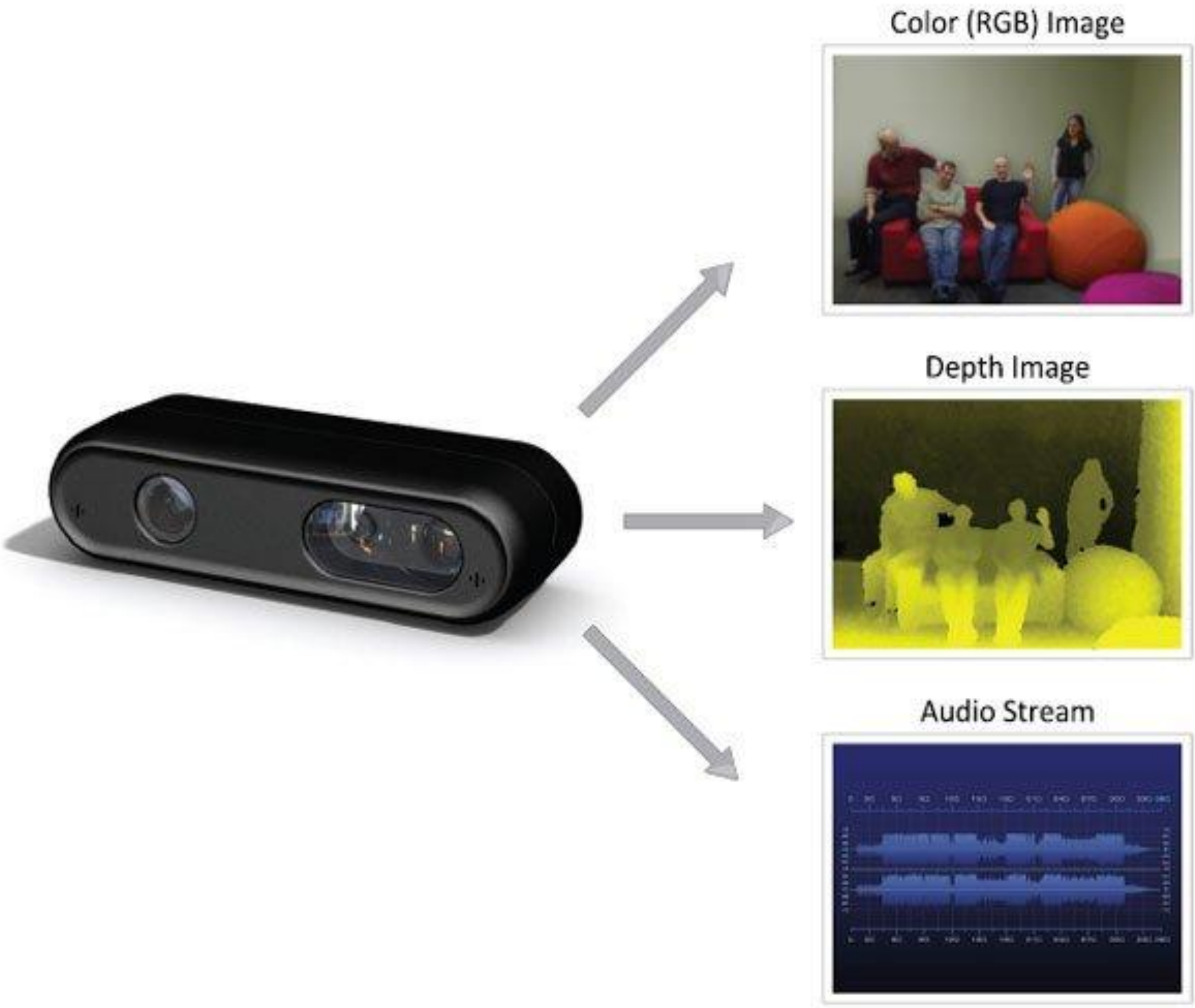
$$\psi(\mathbf{x}, t) = \bar{\psi}\left(\mathcal{P}^{-1}(\mathbf{x}), t\right)$$

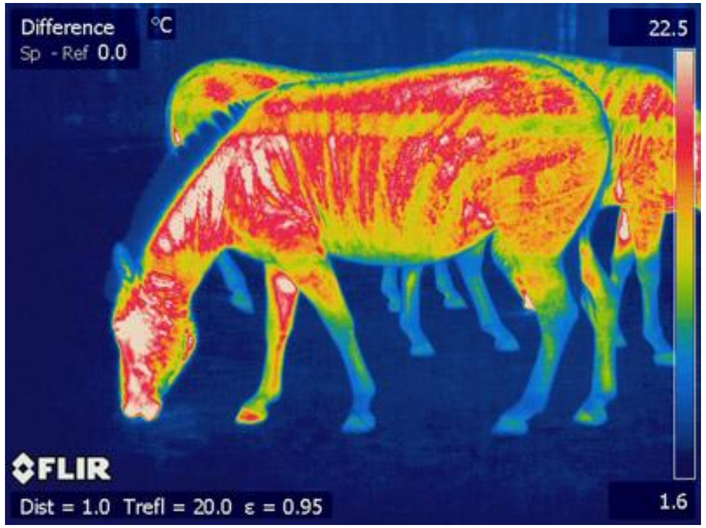
- $a_c(\lambda)$: spectral absorption function
  - Luminous efficiency function: luminance image, monochrome image.
  - Color matching function: color image

- Sensors for imaging:
  - Visible sensor: Luminance, Color (Reflective light => Electric signals)
  - Invisible frequency range sensor: X-ray imaging, Infra-red imaging
  - Range camera, Depth camera

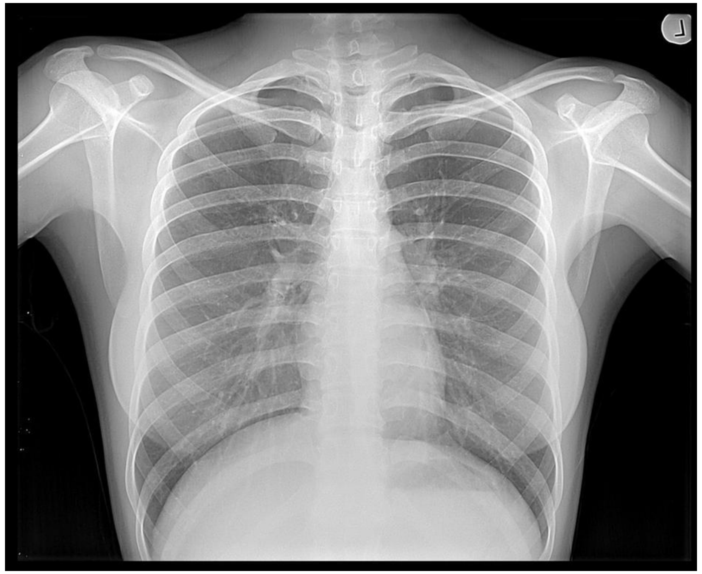


Video Imaging Principles





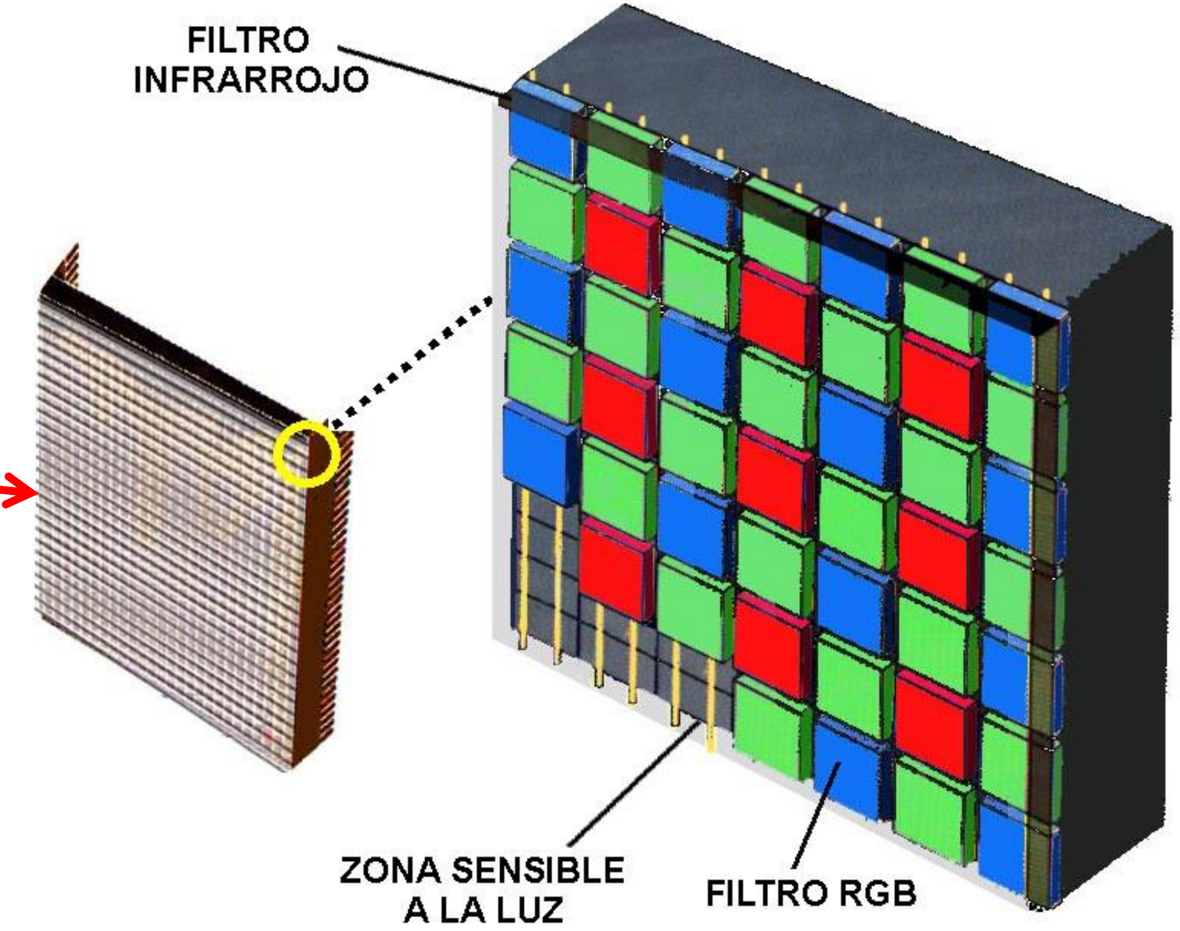
Thermal



X-ray

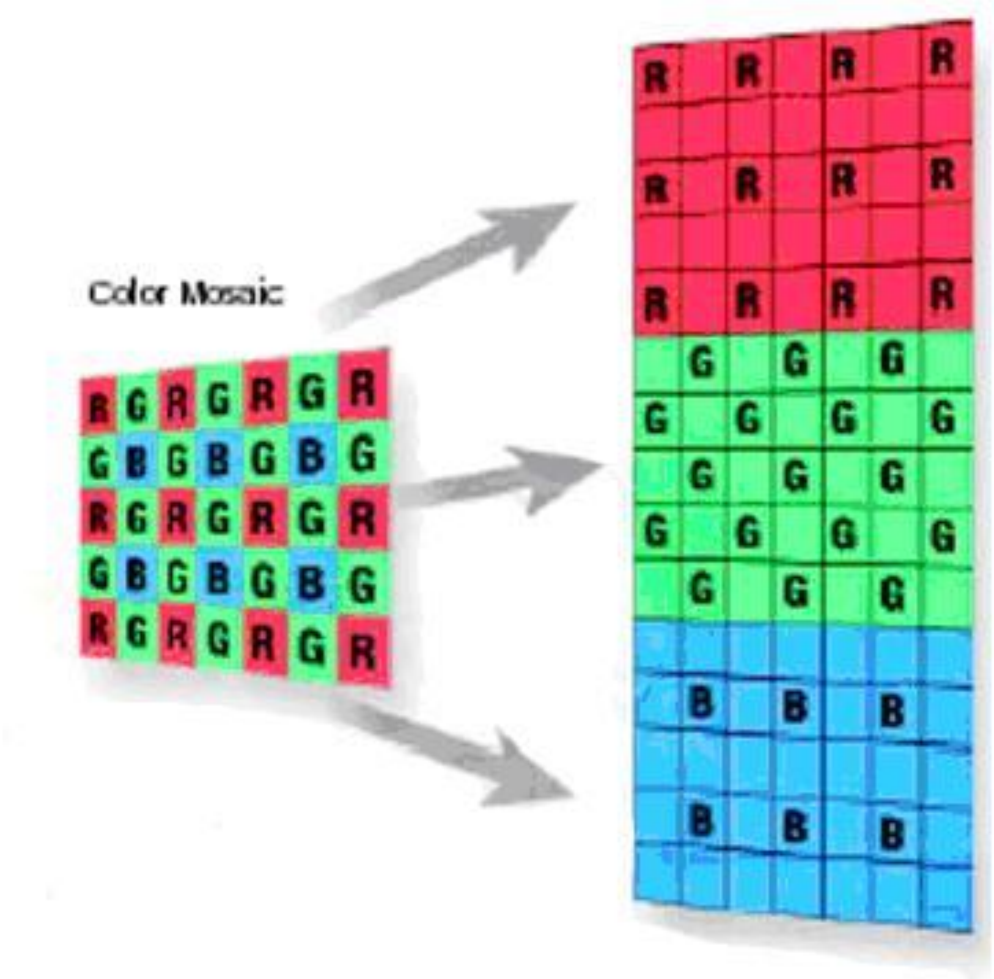
Video Cameras

- Solid-state sensors
  - Charge coupled device (CCD)

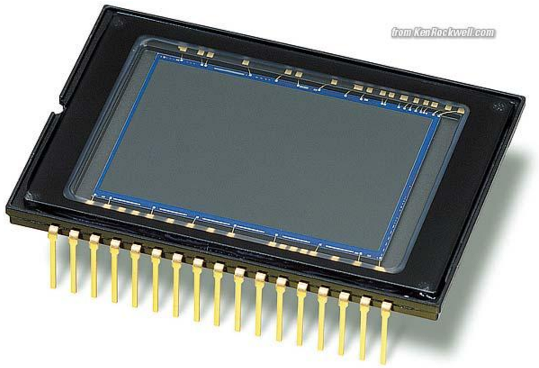


Video Cameras

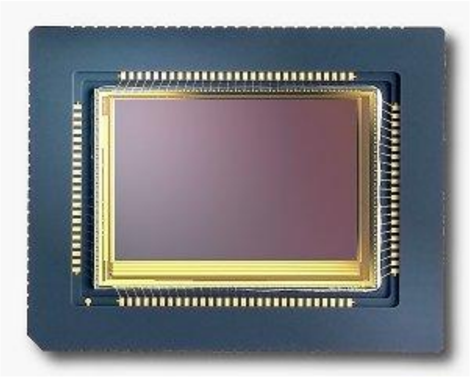
- Image demosaicing:



Video Cameras

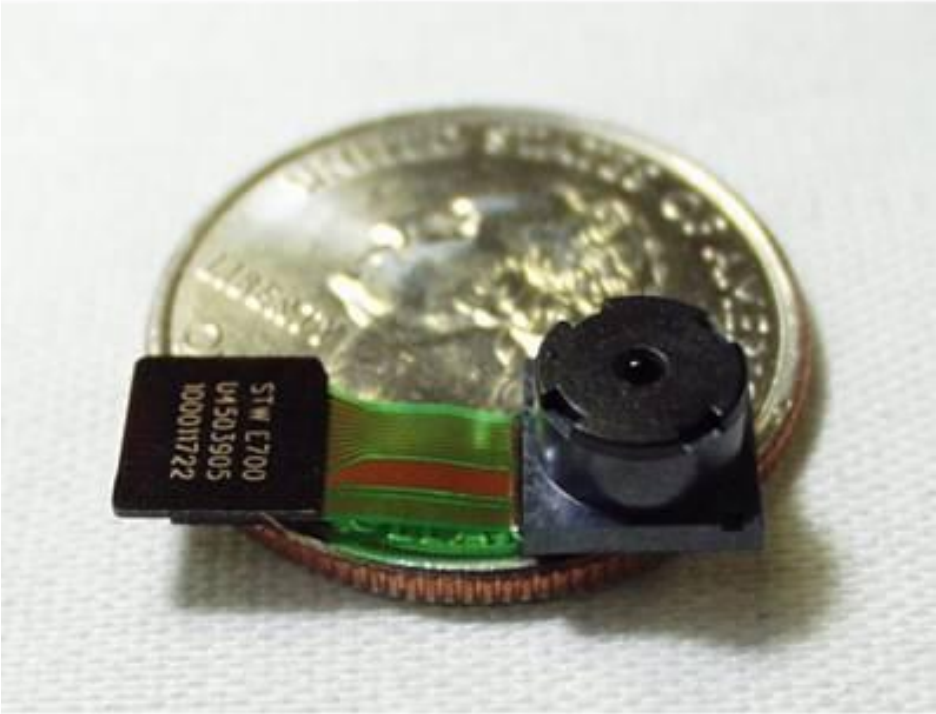


CCD



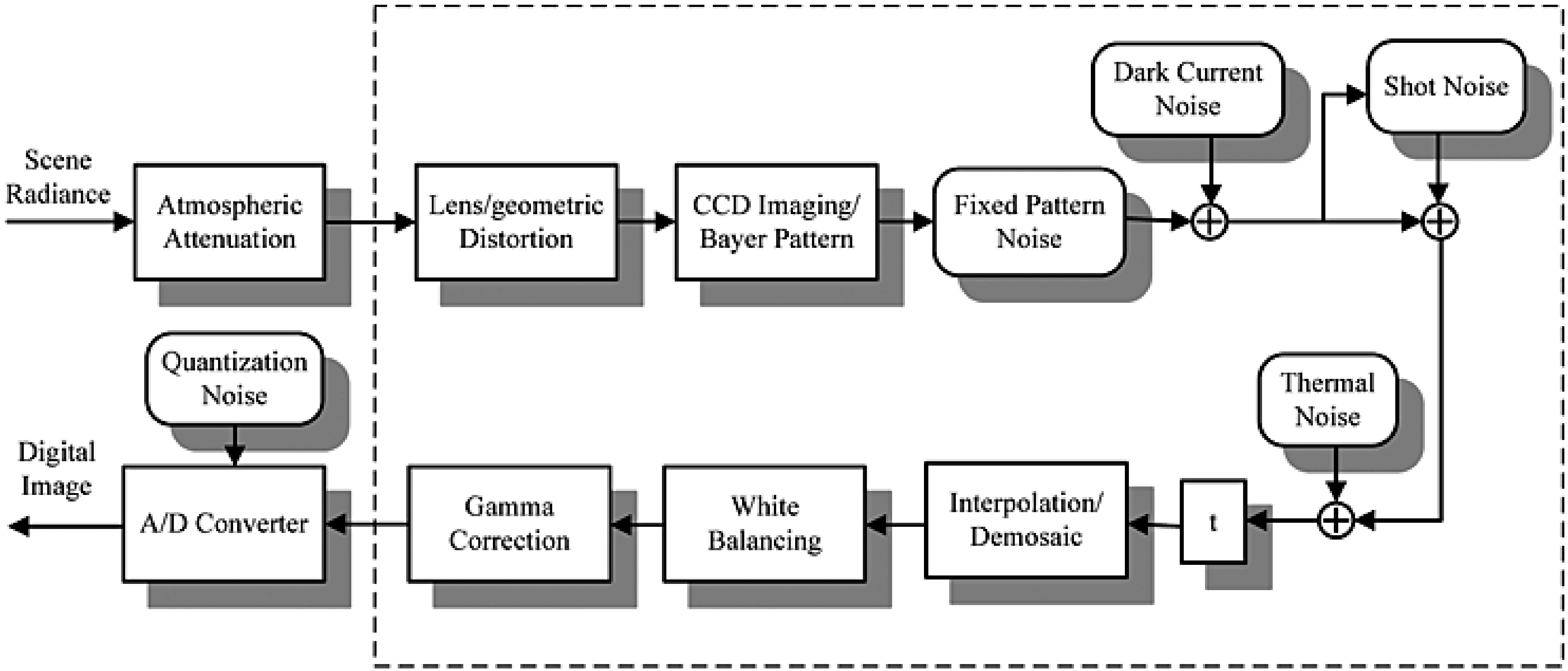
CMOS

Video Cameras



CMOS

Video Cameras



Video Displays

- Use three light sources emitting red, green, blue components:

1. CRT (cathode ray tube): high resolution, thick
2. **LCD** (liquid crystal display): restricted viewing angle
3. PDP (plasma display panel): dynamic false contour, high power consumption
4. **OLED** (Organic Light Emitting Diodes): active lighting, bright and clear

- More details on their operating principles can be found in Y. Hashimoto, M. Yamamoto, and T. Asaida, “Cameras and display systems,” Proc. IEEE, vol. 83, pp. 1032-1043, July 1995

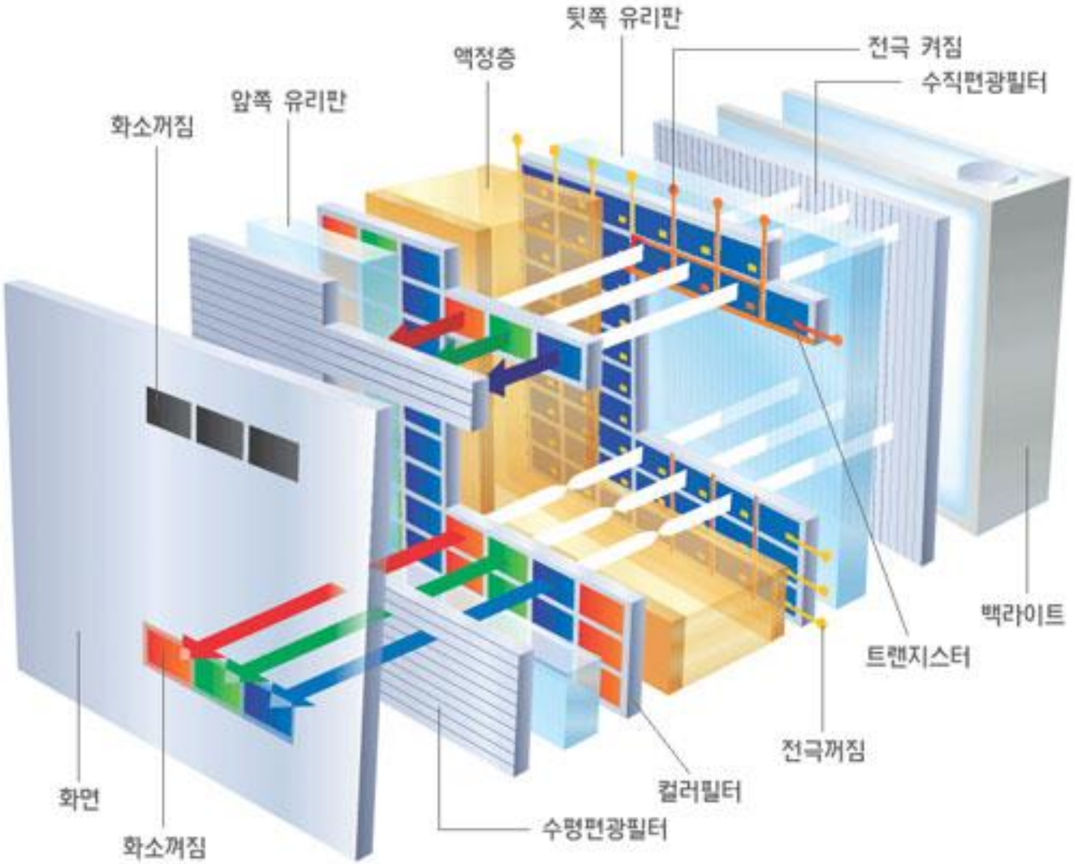
Video Displays

CRT

Cathode ray tube

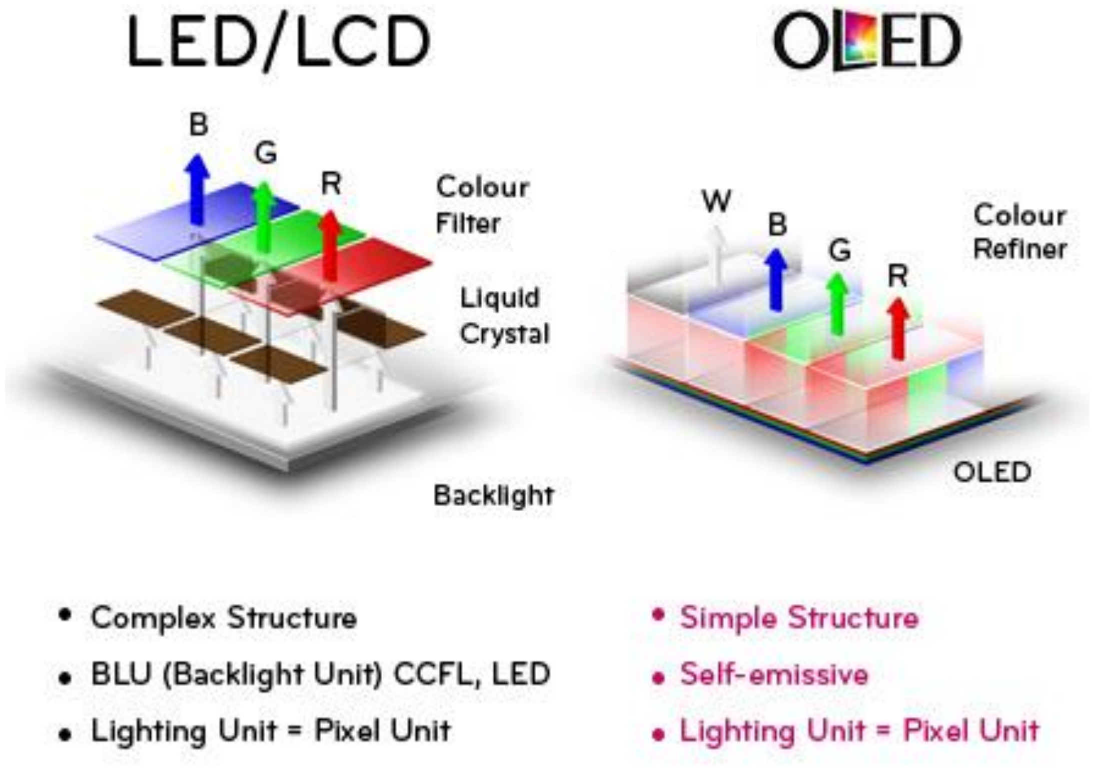
Video Displays

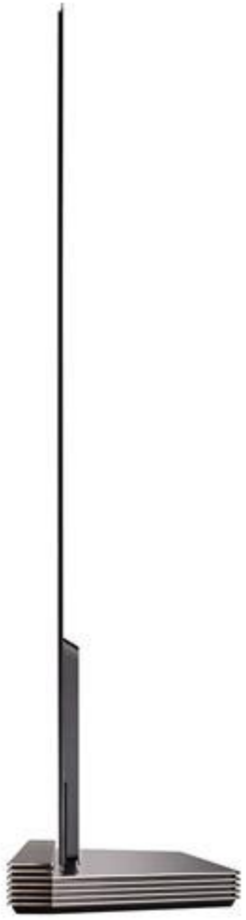
LCD monitor

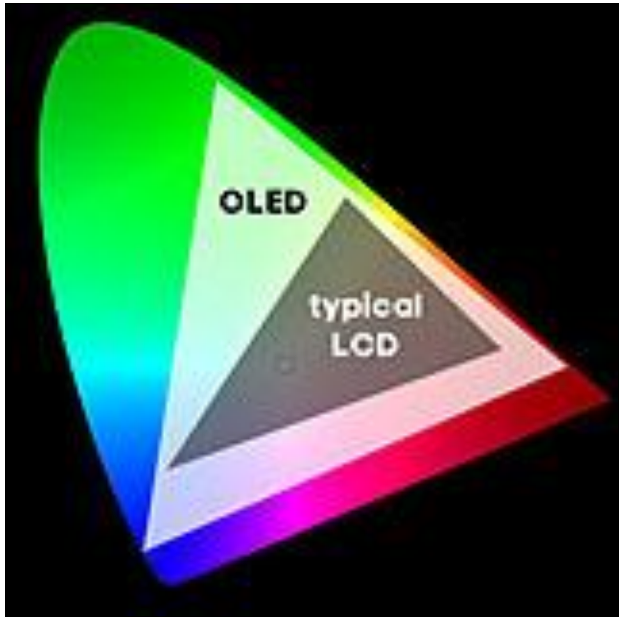


TFT-LCD

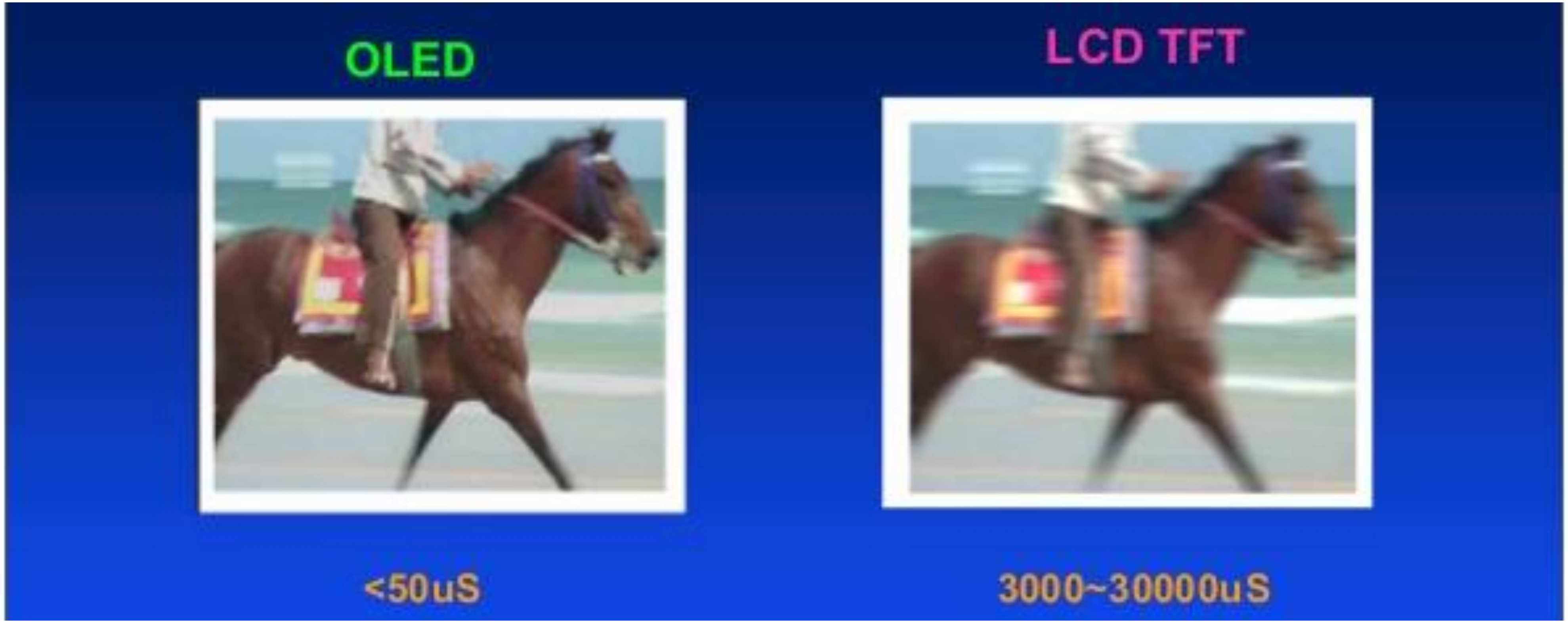
Video Displays





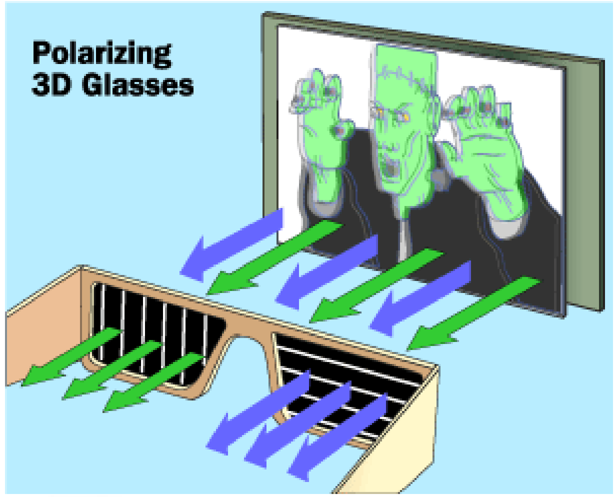


Video Displays



Response time

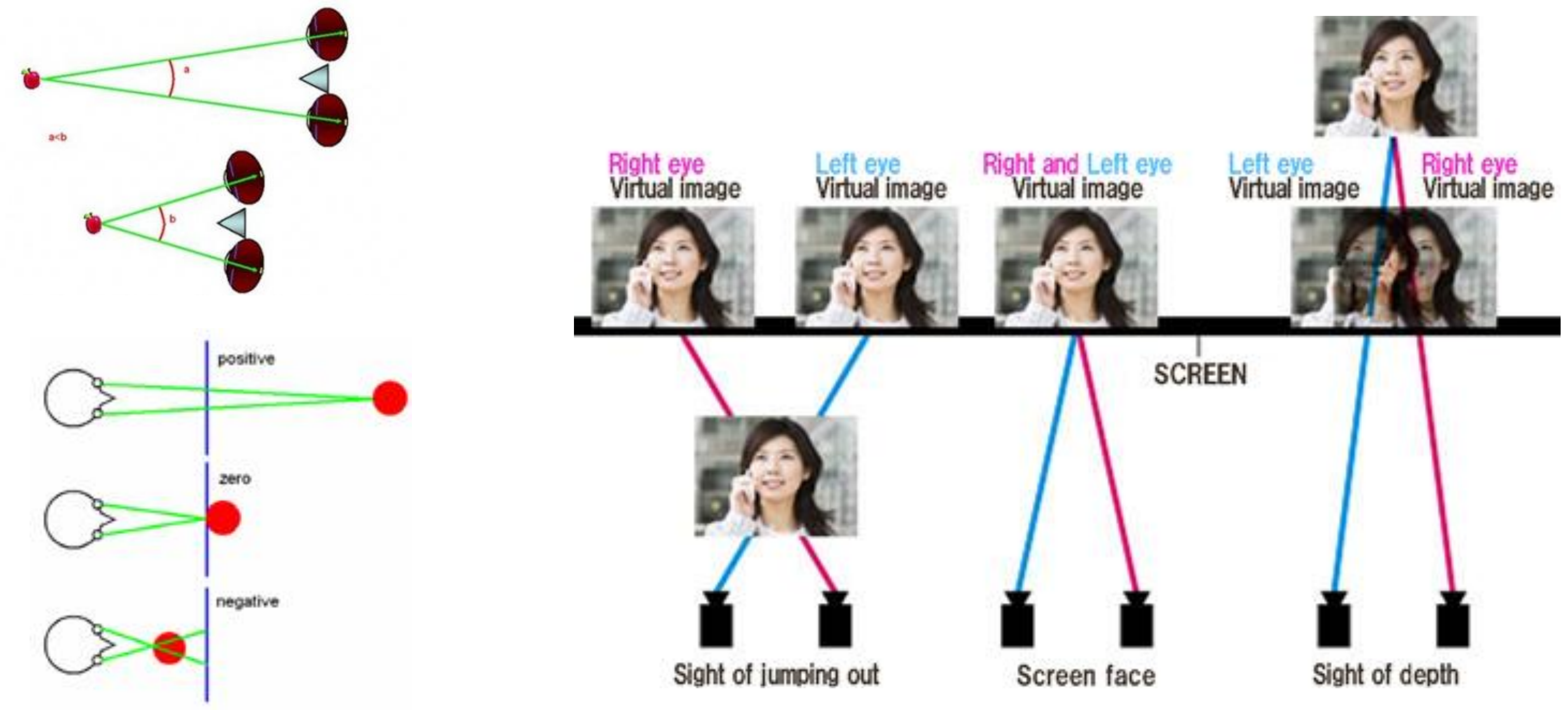
Video Displays



Polarized glasses

Polarized 3DTV

Video Displays



Composite vs Component Videos

- Composite video
  - Convert RGB to YIQ (YCbCr)
  - Multiplexing YIQ into a single signal
  - Chrominance components are band-limited and multiplexed to the upper band of Y
  - Used in cheap analog video devices (analog TV, VHS tapes, etc.)

- S (separate)-video
  - 2 separate signals
  - one for luminance (Y), the other for chrominance (IQ)

- Component video
  - Three color components captured/saved/displayed separately
  - Stored in either RGB or YIQ
  - Natural video format of DVD

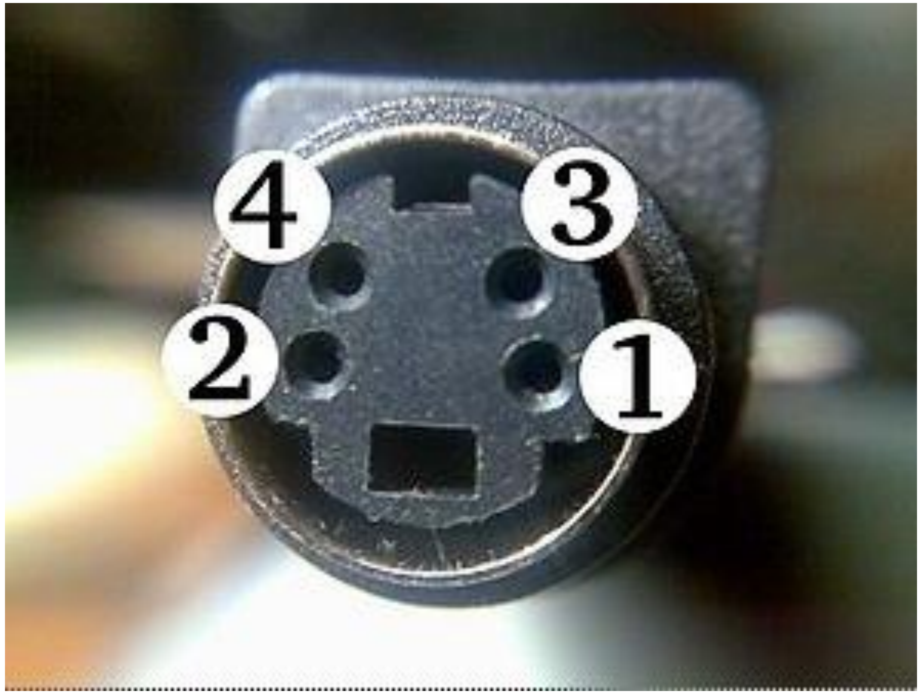
Composite vs Component Videos

- Image quality
  - Component > S-video >> Composite

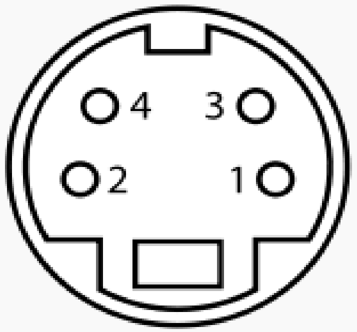
DVD

TV

Composite vs Component Videos



Pin out



Looking at the female connector

Pin 1	GND	Ground (Y)
Pin 2	GND	Ground (C)
Pin 3	Y	Intensity (Luminance)
Pin 4	C	Colour (Chrominance)

S-video

Gamma Correction

- Most of display devices
  - Nonlinear relationship between the input voltage v_d and the displayed color intensity B_d:

$$B_d = v_d^{\gamma_d}$$

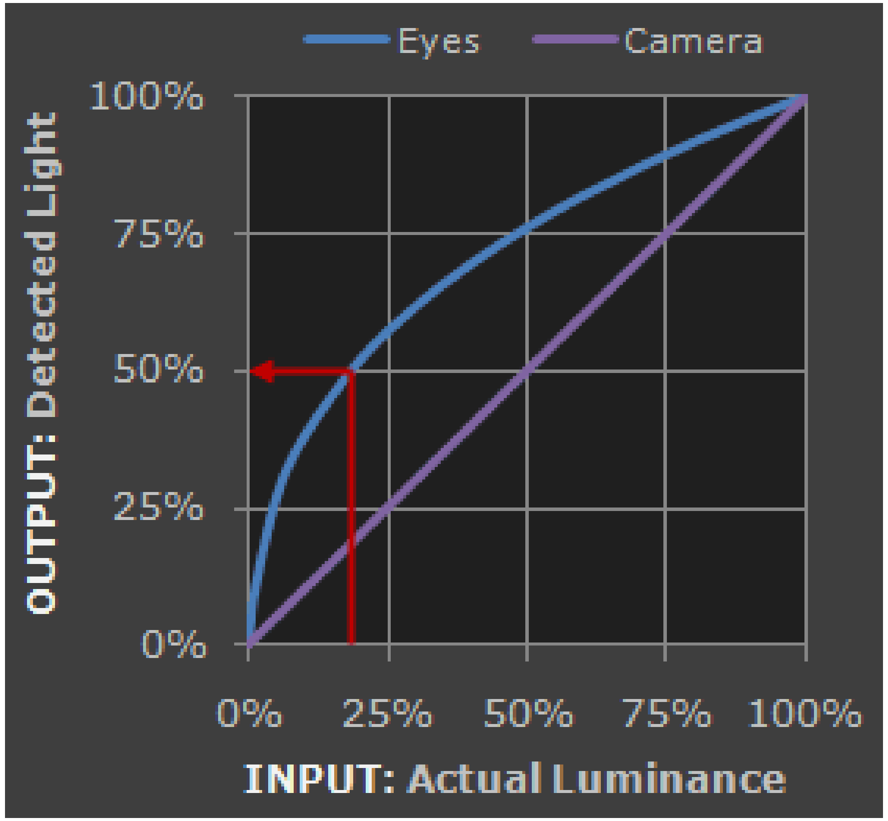
  - CRT display: $\gamma_d = 2.2 \sim 2.5$ => **Gamma effect**

- Gamma Correction
  - v_c: R,G, or B signal captured by the camera

$$v_d = v_c^{\gamma_c/\gamma_d}$$

  - γ_c/γ_d = 2.2 in most TV systems

Linear encoding V_S =	0.0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
Linear intensity I =	0.0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0

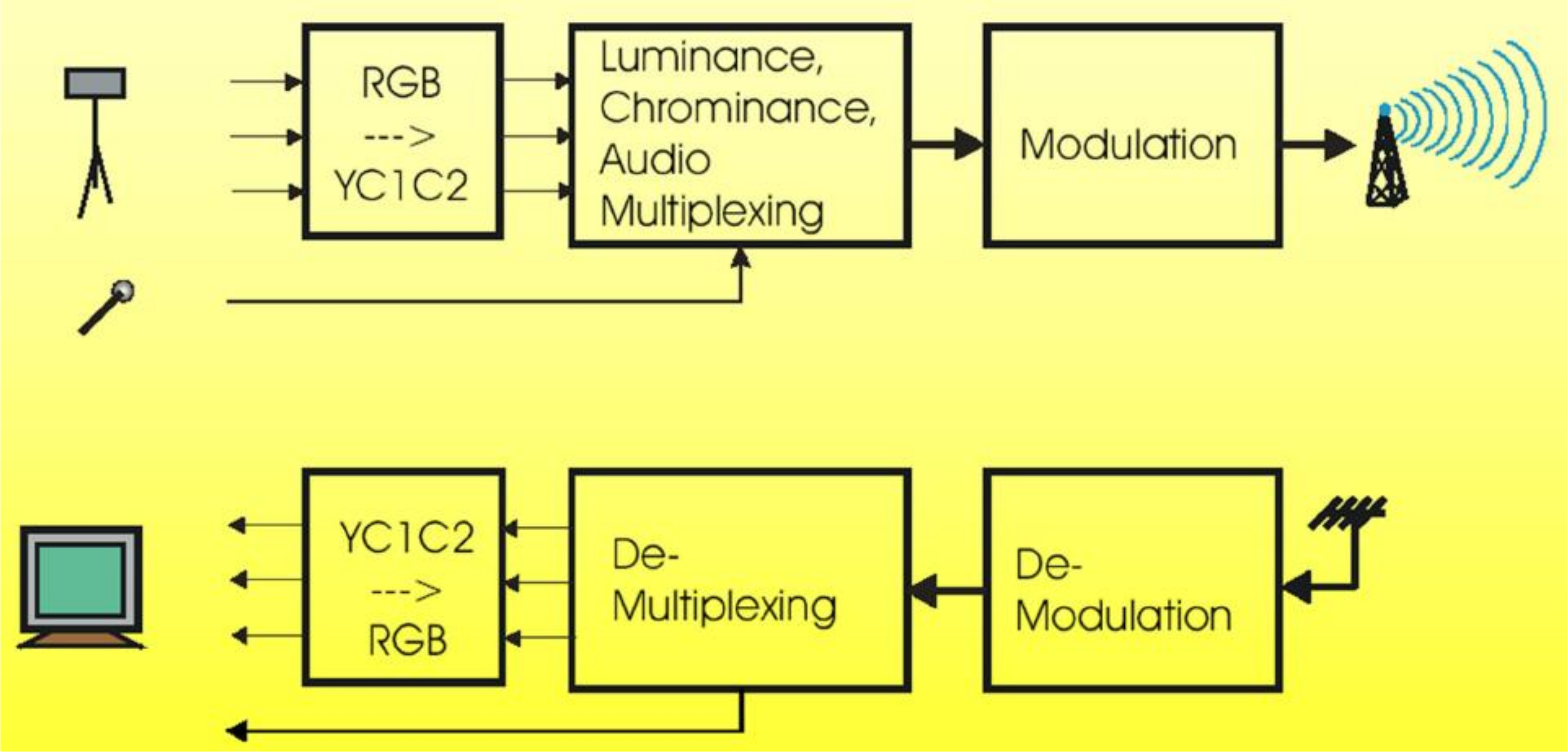


Gamma Correction

Gamma correction results

Analog Color Broadcasting

- Two requirements:
  - Overall bandwidth of a color TV must fit within that of a monochrome TV signal (i.e., 6 MHz in USA)
  - All color signals must be multiplexed into a single composite signal
- Analog color TV systems:



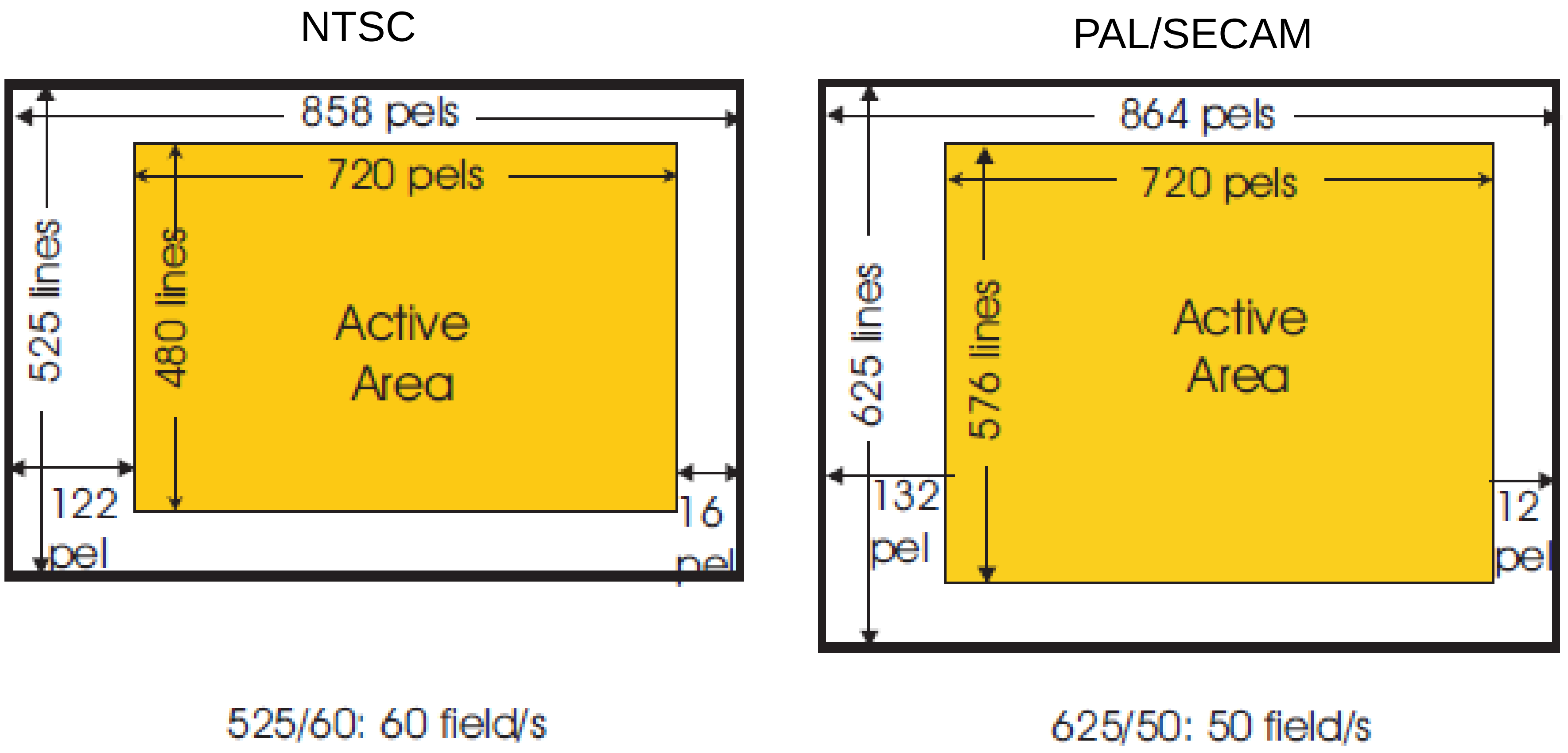
Analog Color Broadcasting

- Three systems:
  - NTSC (USA), PAL (Europe, China), SECAM (Russia)

Table 1.1. Parameters of Analog Color TV Systems

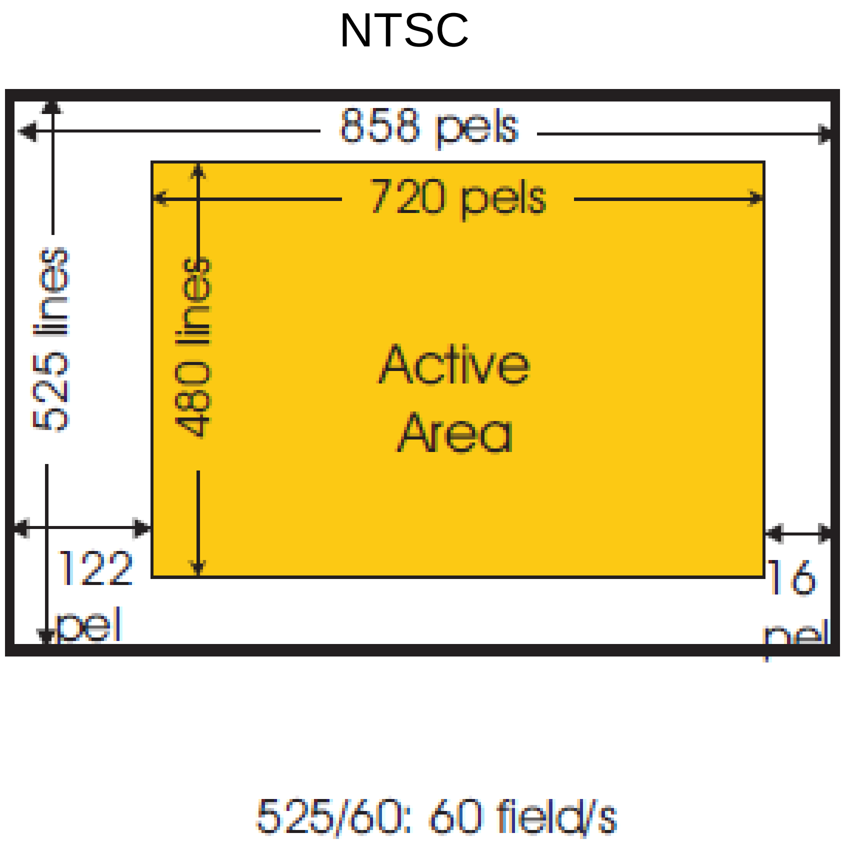
Parameters	NTSC	PAL	SECAM
Field Rate	59.94	50	50
Line Number/Frame	525	625	625
Line Rate (Line/s)	15,750	15,625	15,625
Image Aspect Ratio	4:3	4:3	4:3
Color Coordinate	YIQ	YUV	YDbDr
Luminance Bandwidth (MHz)	4.2	5.0, 5.5	6.0
Chrominance Bandwidth (MHz)	1.5(I),0.5(Q)	1.3(U,V)	1.0 (U,V)
Color Subcarrier (MHz)	3.58	4.43	4.25 (Db),4.41 (Dr)
Color modulation	QAM	QAM	FM
Audio Subcarrier (MHz)	4.5	5.5,6.0	6.5
Composite Signal Bandwidth(MHz)	6.0	8.0,8.5	8.0

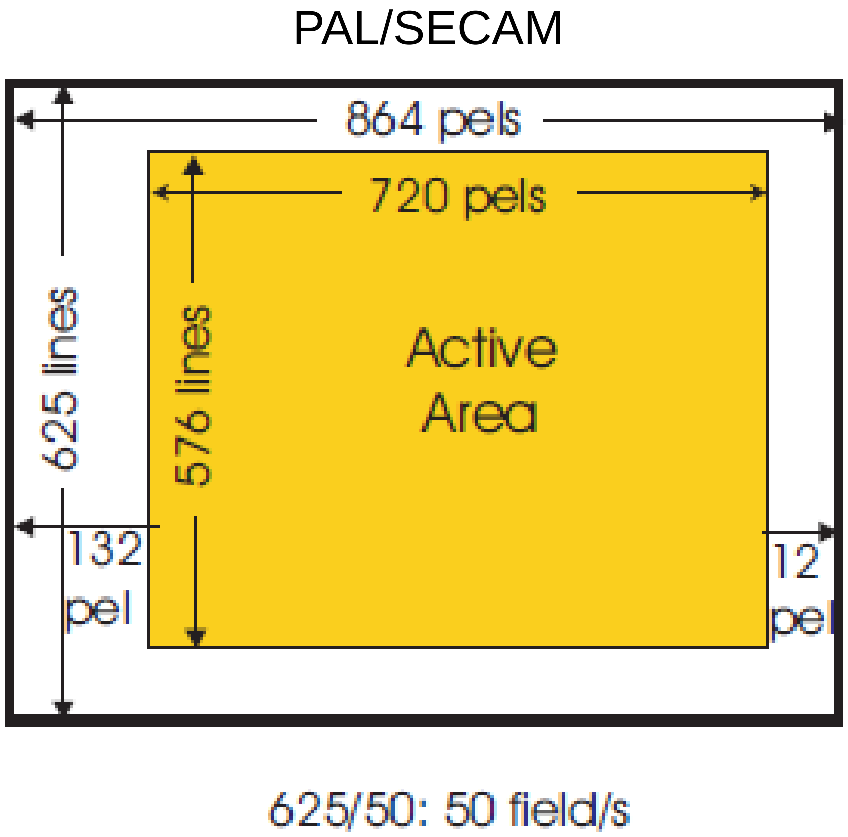
Video Format for Broadcasting



ITU-R BT.601 video format

Video Format for Broadcasting





ITU-R BT.601 video format

Color Space Conversion

- YC_bC_r
  - Scaled version of YUV to the range [0, 255]

- RGB $\rightarrow$ YC_bC_r

$$\begin{bmatrix} Y \\ C_b \\ C_r \end{bmatrix} = \begin{bmatrix} 0.257 & 0.504 & 0.098 \\ -0.148 & -0.291 & 0.439 \\ 0.439 & -0.368 & -0.071 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix} + \begin{bmatrix} 16 \\ 128 \\ 128 \end{bmatrix}$$

To save broadcasting bandwidth

- $YC_bC_r \rightarrow$ RGB

$$\begin{bmatrix} R \\ G \\ B \end{bmatrix} = \begin{bmatrix} 1.164 & 0.000 & 1.596 \\ 1.164 & -0.392 & -0.813 \\ 1.164 & 2.017 & 0.000 \end{bmatrix} \begin{bmatrix} Y-16 \\ C_b-128 \\ C_r-128 \end{bmatrix}$$

Chrominance Sub-sampling

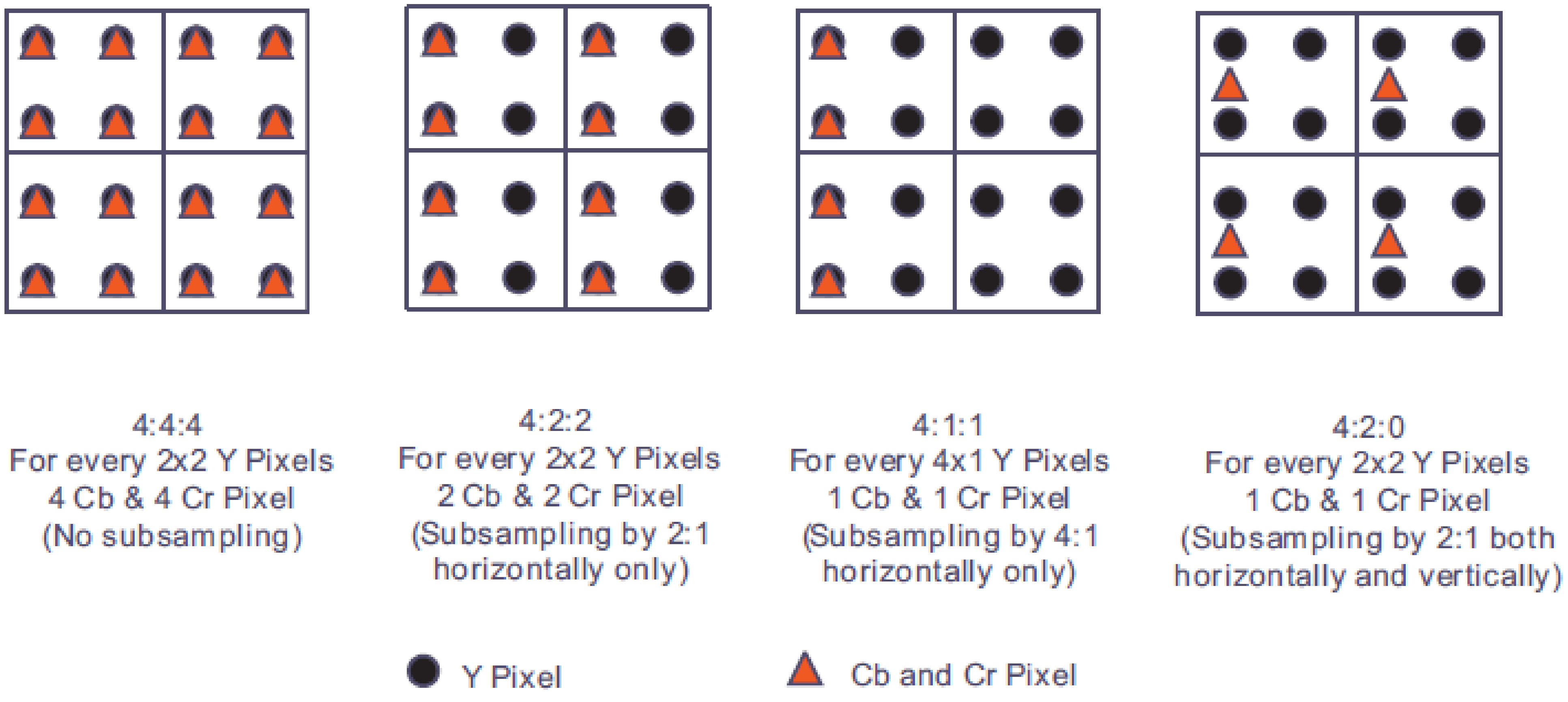
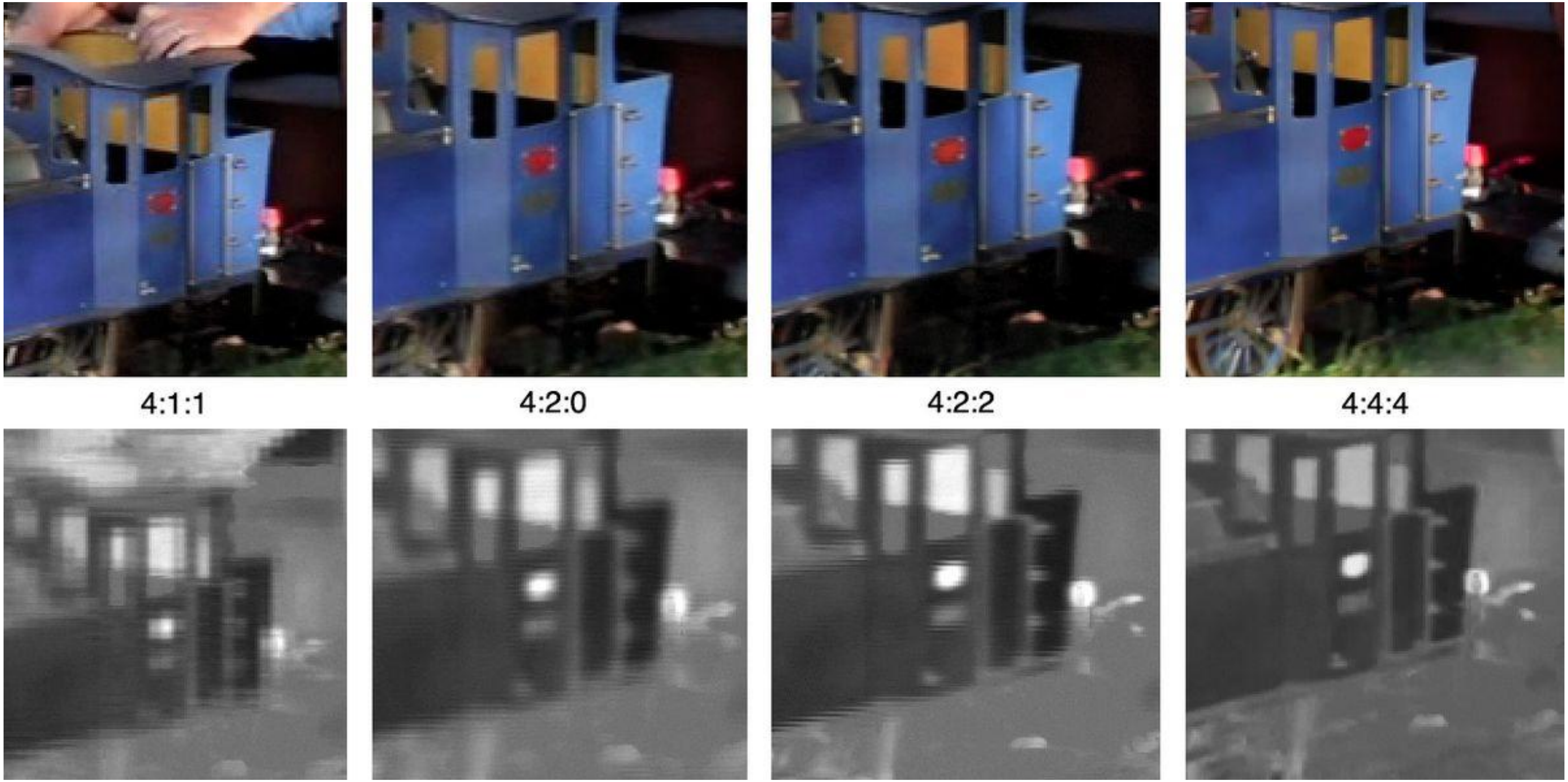
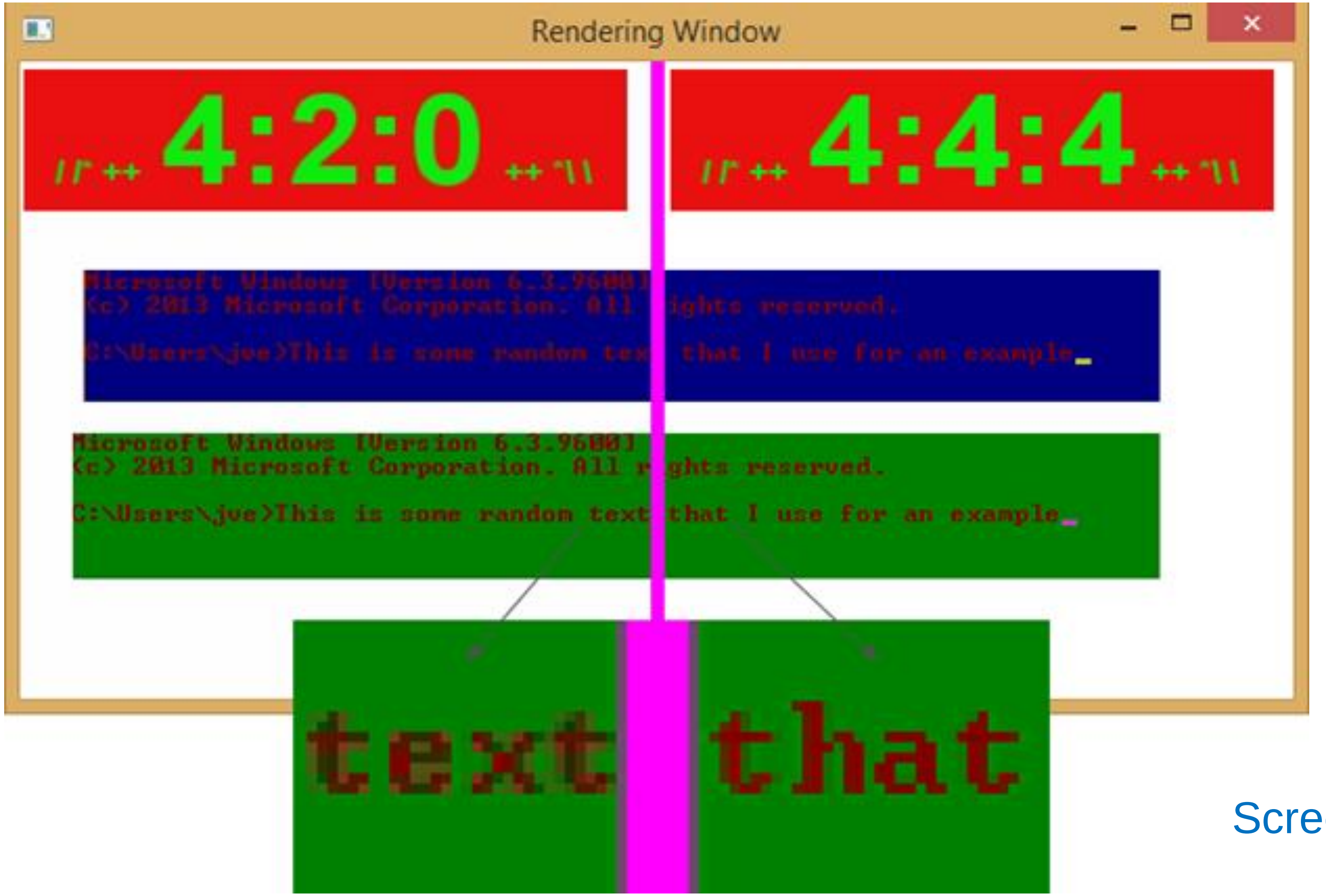


Figure 1.8. BT.601 chrominance subsampling formats. Note that the two adjacent lines in any one component belong to two different fields.

Chrominance Sub-sampling



Chrominance Sub-sampling



Screen contents

Chrominance Sub-sampling

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

H.264

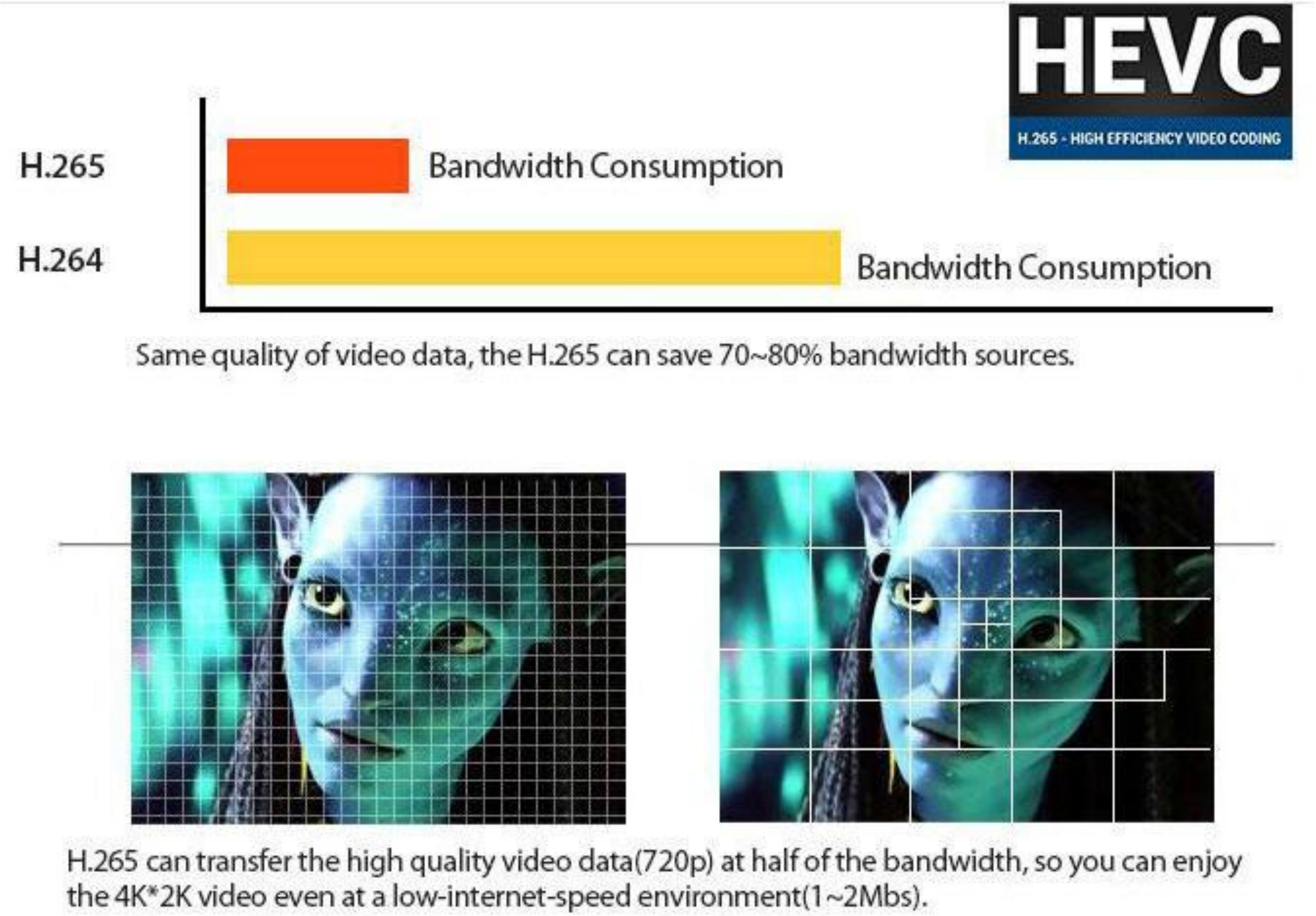
H.264 MPEG-4/AVC Full HD 1080p

Video Compression

- QCIF
  - Width: 176
  - Height: 144
  - Reduced frame rate: 8.33 frames/s
  - 12 bits/pixel (4:2:0)
    - 8 for luminance
    - 4 for chrominance
  - Duration: 12 s

- File size
  - 176x144x8.33x12x12/8 = 3801600 (bytes)

Video Compression



Video Quality Assessment

- MSE

$$\mathrm{MSE} = \frac{1}{N}\sum_{m,n,k}(\psi_1(m,n,k) - \psi_2(m,n,k))^2$$

- PSNR

$$\mathrm{PSNR} = 10\log_{10}\frac{\psi_{\max}^2}{\mathrm{MSE}}$$

  - > 40 dB: excellent, very close to original
  - 30-40 dB: good
  - 20-30 dB: quite poor
  - <20 dB: not acceptable

- Mean Opinion Score (MOS)

  5: Excellent Imperceptible

  4: Good Perceptible but not annoying

  3: Fair Slightly annoying

  2: Poor Annoying

  1: Bad Very annoying

Video Quality Assessment

MSE cannot measure all kinds of distortions, which is useful for Gaussian noise.

Research on **image quality assessment** is in progress!

THANK YOU!