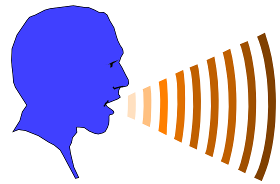
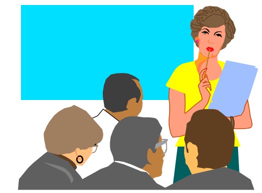
   

Image and Video Processing

Image Formation and Representation

Yao Wang
Tandon School of Engineering, New York University

What will you lean

- How do we perceive color?
- How to capture, display, print in color?
- What are primary colors?
- How are digital images specified?
  - Color resolution vs. spatial resolution vs. temporal resolution.
- What are some standard image/video formats?

Outline

- Color perception
- Color production using primary colors
- Color specification
- Color image representation
- Image capture and display
- 3D->2D Projection
- Video format (SD, HD, UHD)

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

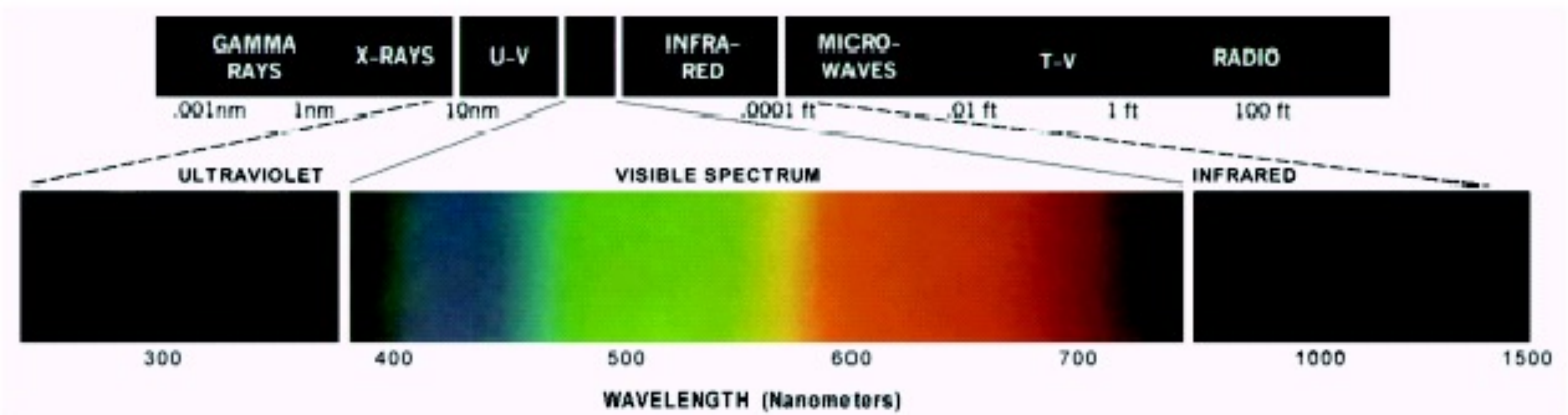
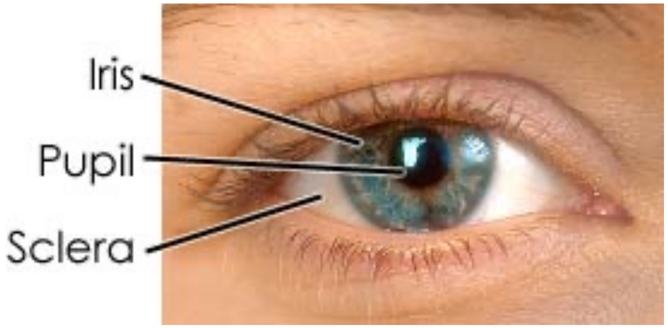
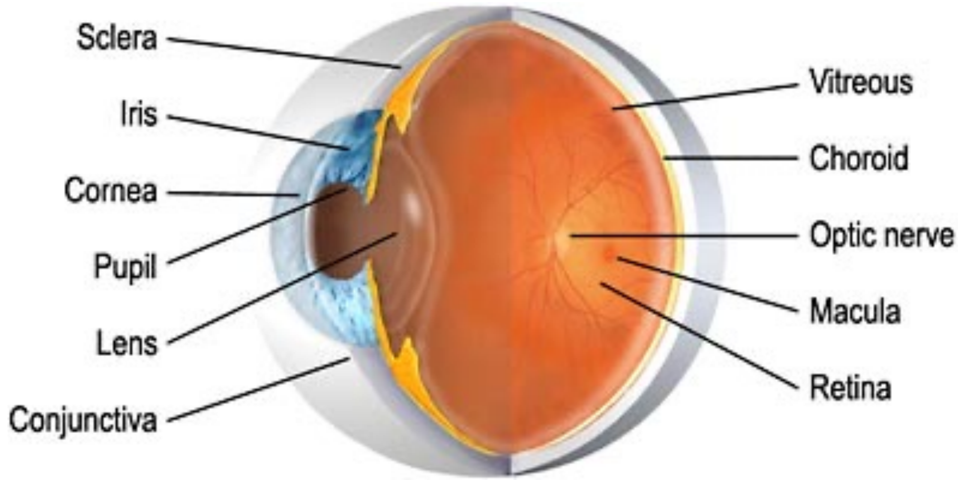


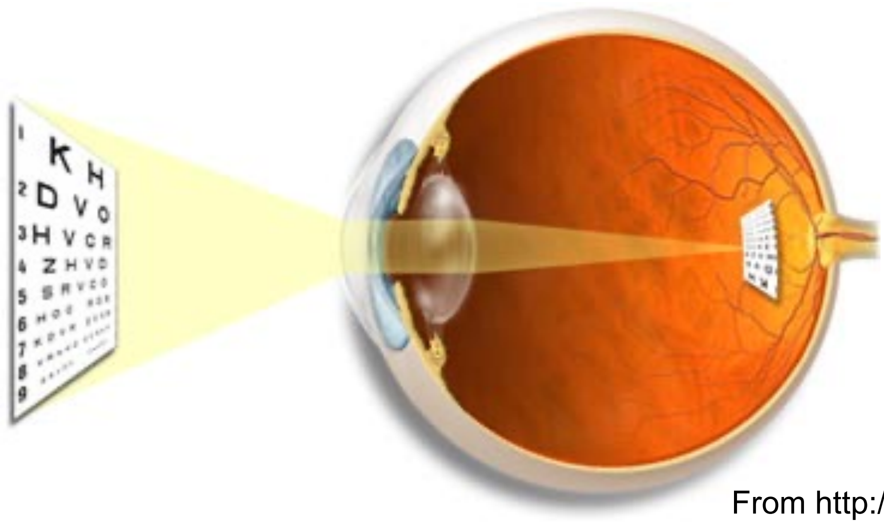
FIGURE 6.2 Wavelengths comprising the visible range of the electromagnetic spectrum. (Courtesy of the General Electric Co., Lamp Business Division.)

from [Gonzalez2008]

Eye Anatomy

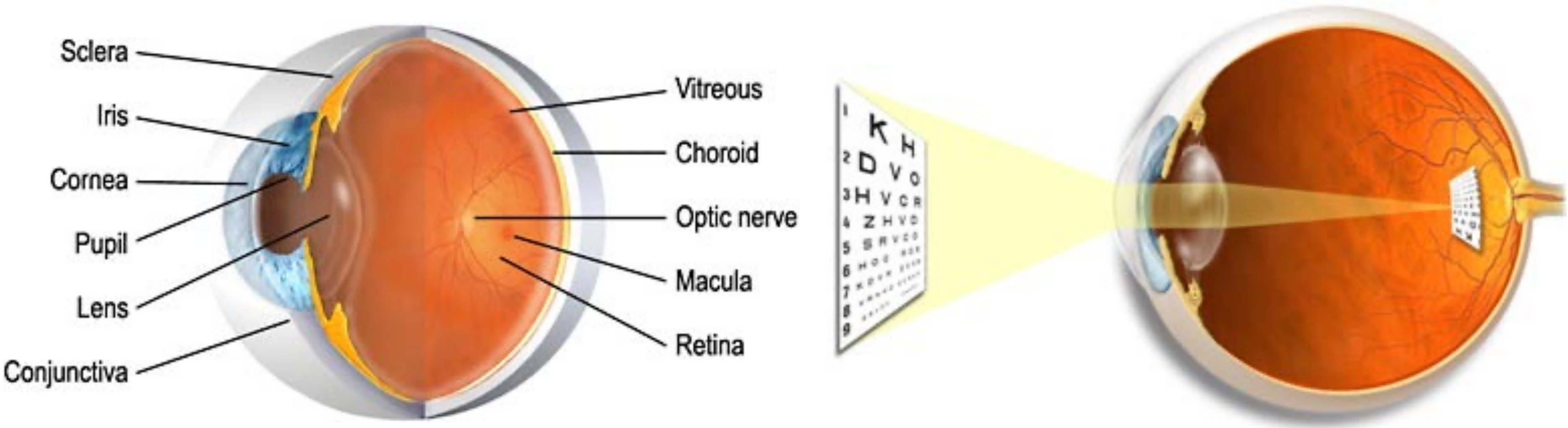






From http://www.stlukeseye.com/Anatomy.asp

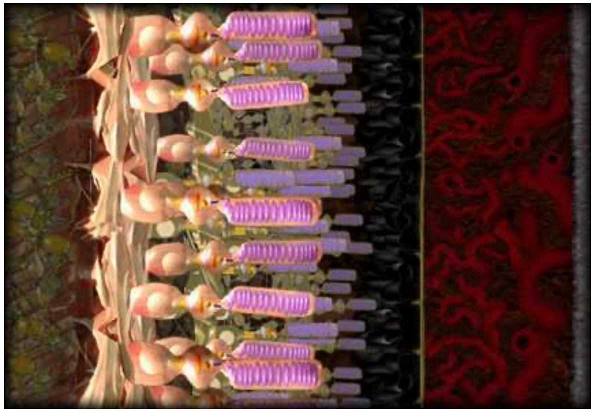
Eye vs. Camera



Camera components	Eye components
Lens	Lens, cornea
Shutter	Iris, pupil
Film	Retina
Cable to transfer images	Optic nerve send the info to the brain

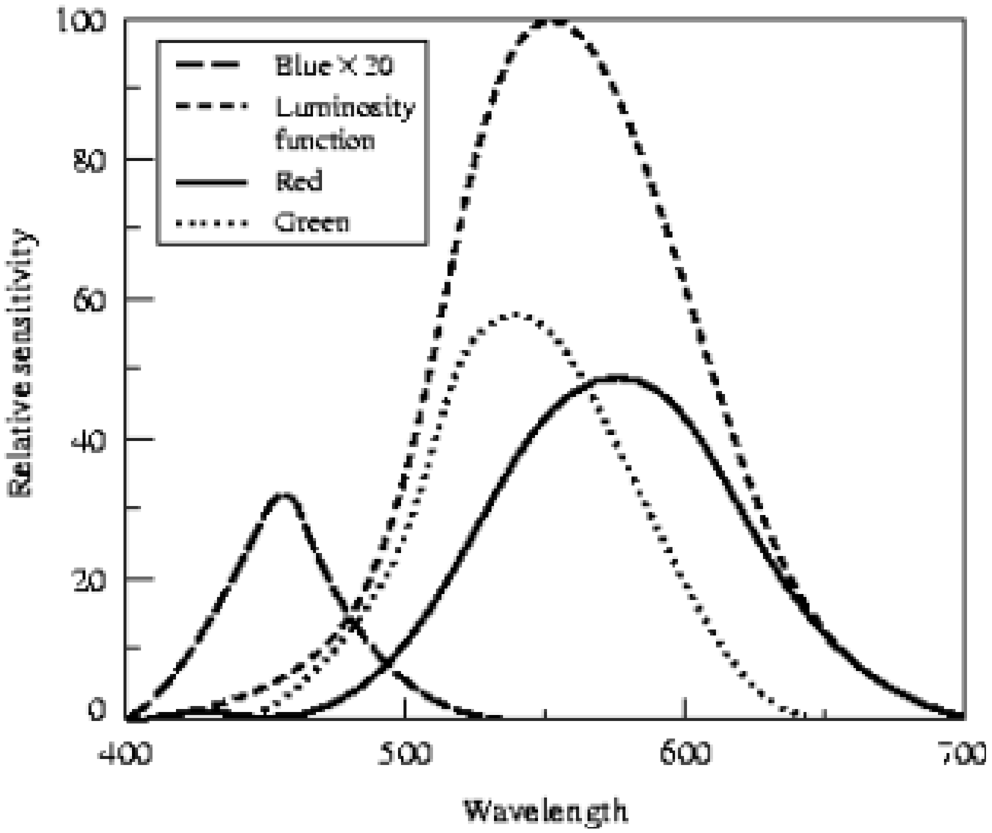
Photo Receptors in the Retina

- Rods: perceive brightness only, extremely sensitive even at night
- Cones: perceive color tone
  - Red (560-580nm), green (530-540nm), and blue (420-440nm) cones
  - Different cones have different frequency responses (each cone like a filter!)
  - Tri-receptor theory of color vision [Young1802]
- More rods (120 million) than cones (6 million)



From http://www.macula.org/anatomy/retinaframe.html

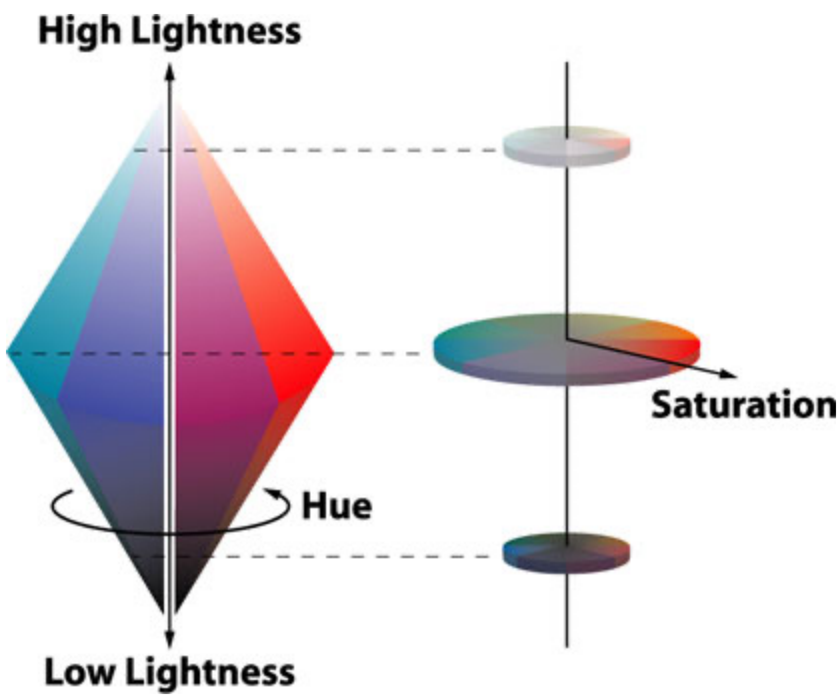
Frequency Responses of Cones and the Luminous Efficiency Function

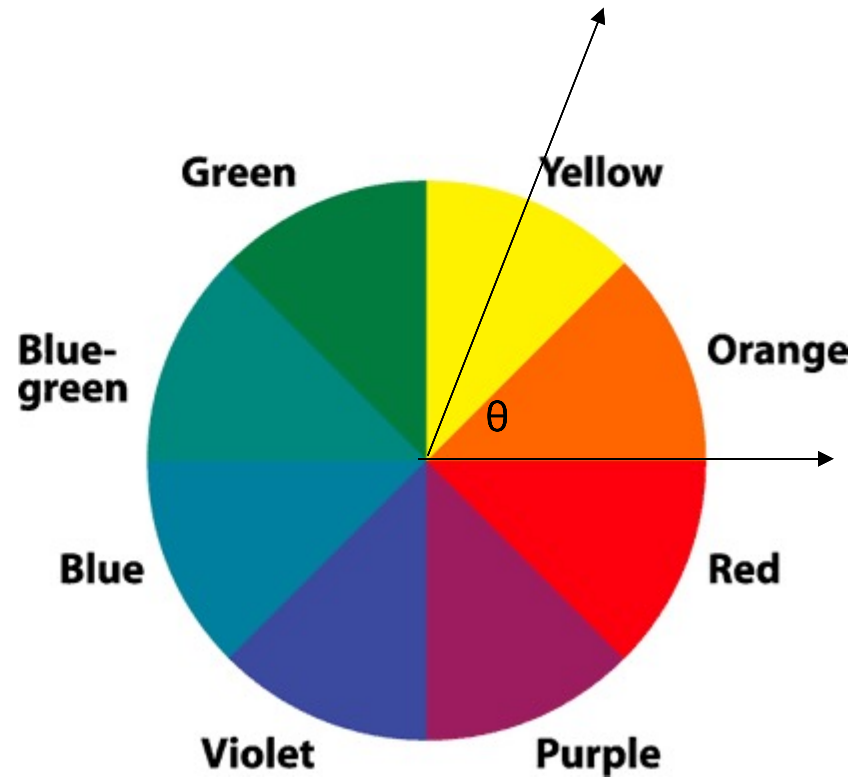


$$C_i = \int C(\lambda) a_i(\lambda) d\lambda, \quad i = r, g, b, y$$

Three Attributes of Color

- Luminance (brightness)
- Chrominance
  - Hue (color tone) and Saturation (color purity)
- Represented by a "color cone"





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 (emits all visible frequency)
- Reflecting sources (secondary light):
  - 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 (absorbs all frequency)

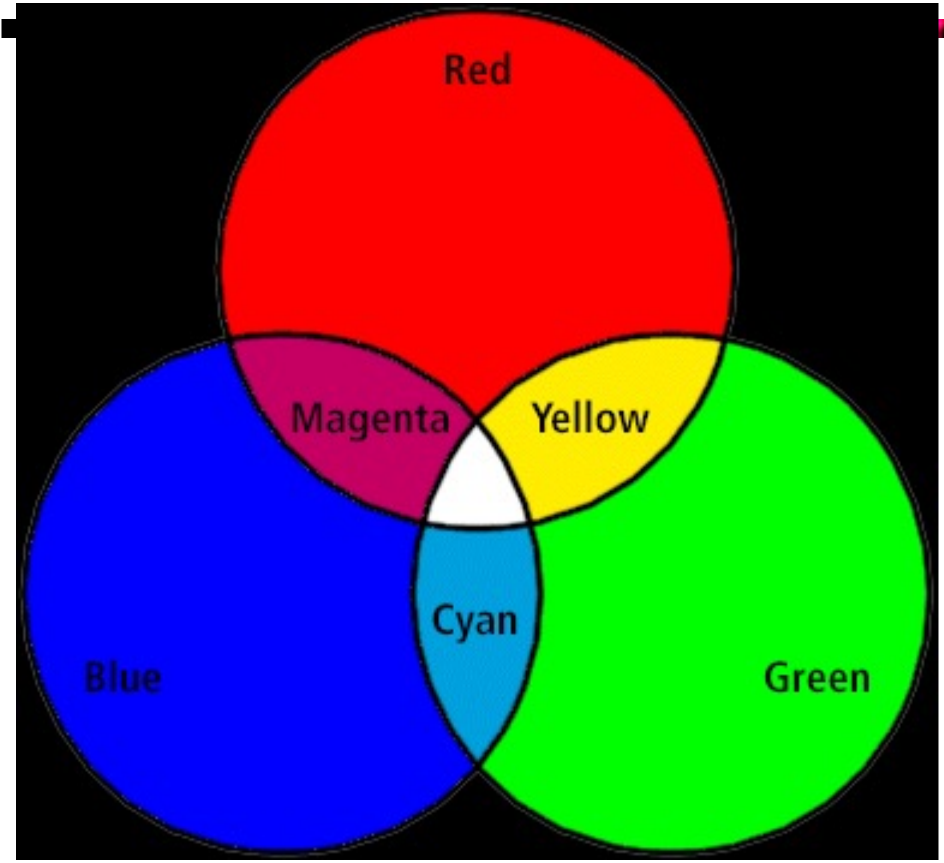
Trichromatic Color Mixing

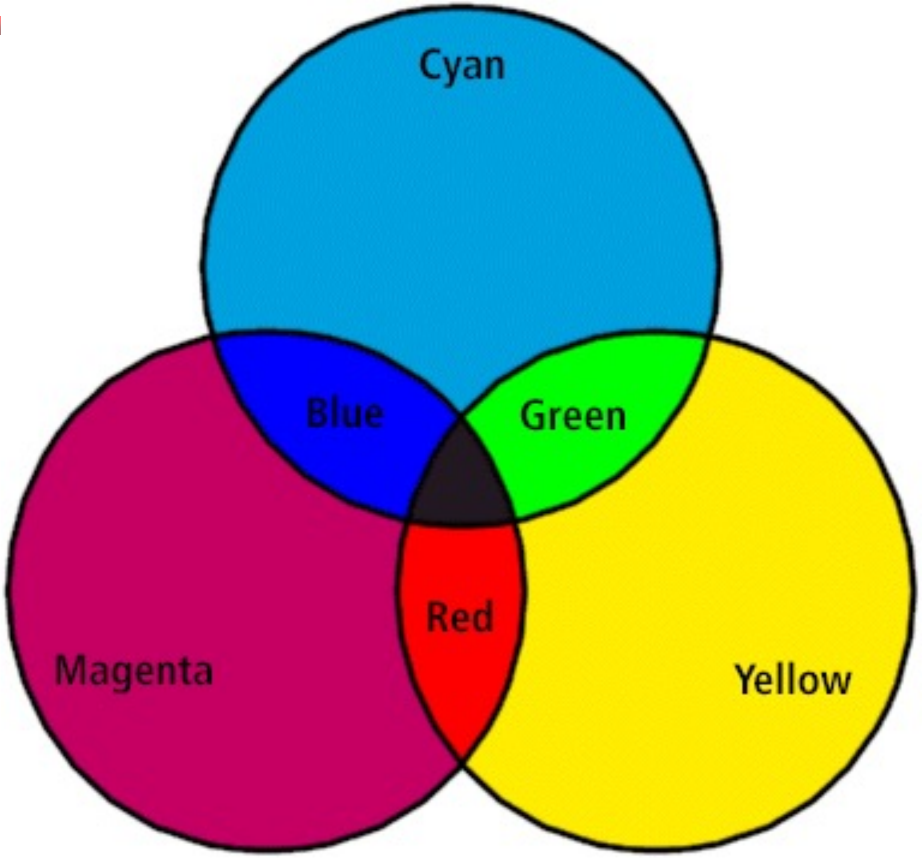
- Trichromatic color mixing theory
  - Most visible color can be obtained by mixing three properly chosen primary colors with a right proportion

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

- 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:
  - Cyan, Magenta, Yellow (CMY)
  - Color printer works by using cyan, magenta, yellow and black (CMYK) dyes
    - More advanced printers use more primary colors to produce a wider range of colors

RGB vs CMY





Magenta = Red + Blue
Cyan = Blue + Green
Yellow = Green + Red

Magenta = White - Green
Cyan = White - Red
Yellow = White - Blue

red

Green

Blue

Color Representation Models

- Specify the tristimulus values associated with the three primary colors
  - RGB, CMY, XYZ
- Specify the luminance and chrominance
  - CIELAB (nonlinear transformation of XYZ, so that perceived difference in luminance and chrominance are more uniformly spaced in LAB coordinates)
  - HSI (Hue, saturation, intensity)
  - YIQ (used in analog NTSC color TV)
  - YCbCr (used in digital color TV)
- Amplitude specification (Standard Dynamic Range):
  - 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
- High dynamic range (HDR) Image: up to 16 bits/component

HSI Color Model

- Hue represents dominant color as perceived by an observer. It is an attribute associated with the dominant wavelength.
- Saturation refers to the relative purity or the amount of white light mixed with a hue. The pure spectrum colors are fully saturated. Pink and lavender are less saturated.
- Intensity reflects the brightness.
- Also know as HSB, where B stands for brightness

Conversion Between RGB and HSI

- Converting from RGB to HSI

$$H = \begin{cases} \theta & \text{if } B \le G \\ 360 - \theta & \text{if } B > G \end{cases} \quad \text{with } \theta = \cos^{-1}\left\{ \frac{\frac{1}{2}\left[(R-G)+(R-B)\right]}{\left[(R-G)^2+(R-B)(G-B)\right]^{\frac{1}{2}}} \right\}$$

$$S = 1 - \frac{3}{(R+G+B)}[\min(R,G,B)]$$

$$I = \frac{1}{3}[R+G+B]$$

- Converting from HSI to RGB

RG sector (0≤H<120)

$$B = I(1-S)$$
$$R = I\left[1 + \frac{S\cos H}{\cos(60-H)}\right]$$
$$G = 1-(R+B)$$

GB sector (120≤H<240)

$$R = I(1-S)$$
$$G = I\left[1 + \frac{S\cos(H-120)}{\cos(60-(H-120))}\right]$$
$$B = 1-(R+G)$$

BR sector (240≤H<360)

$$G = I(1-S)$$
$$B = I\left[1 + \frac{S\cos(H-240)}{\cos(60-(H-240))}\right]$$
$$R = 1-(G+B)$$

From [Gonzalez 2008]

YIQ Color Coordinate System

- YIQ is defined by the National Television System Committee (NTSC) for US analog color TV system
  - Y describes the luminance, I and Q describes the chrominance.
  - A more compact representation of the color.
  - YUV plays similar role in PAL and SECAM.
- Conversion between RGB (analog) and YIQ (analog)

$$\begin{bmatrix} Y \\ I \\ Q \end{bmatrix} = \begin{bmatrix} 0.299 & 0.587 & 0.114 \\ 0.596 & -0.274 & -0.322 \\ 0.211 & -0.523 & 0.311 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix}, \quad \begin{bmatrix} R \\ G \\ B \end{bmatrix} = \begin{bmatrix} 1.0 & 0.956 & 0.621 \\ 1.0 & -0.272 & -0.649 \\ 1.0 & -1.106 & 1.703 \end{bmatrix} \begin{bmatrix} Y \\ I \\ Q \end{bmatrix}$$

YUV/YCbCr Coordinate

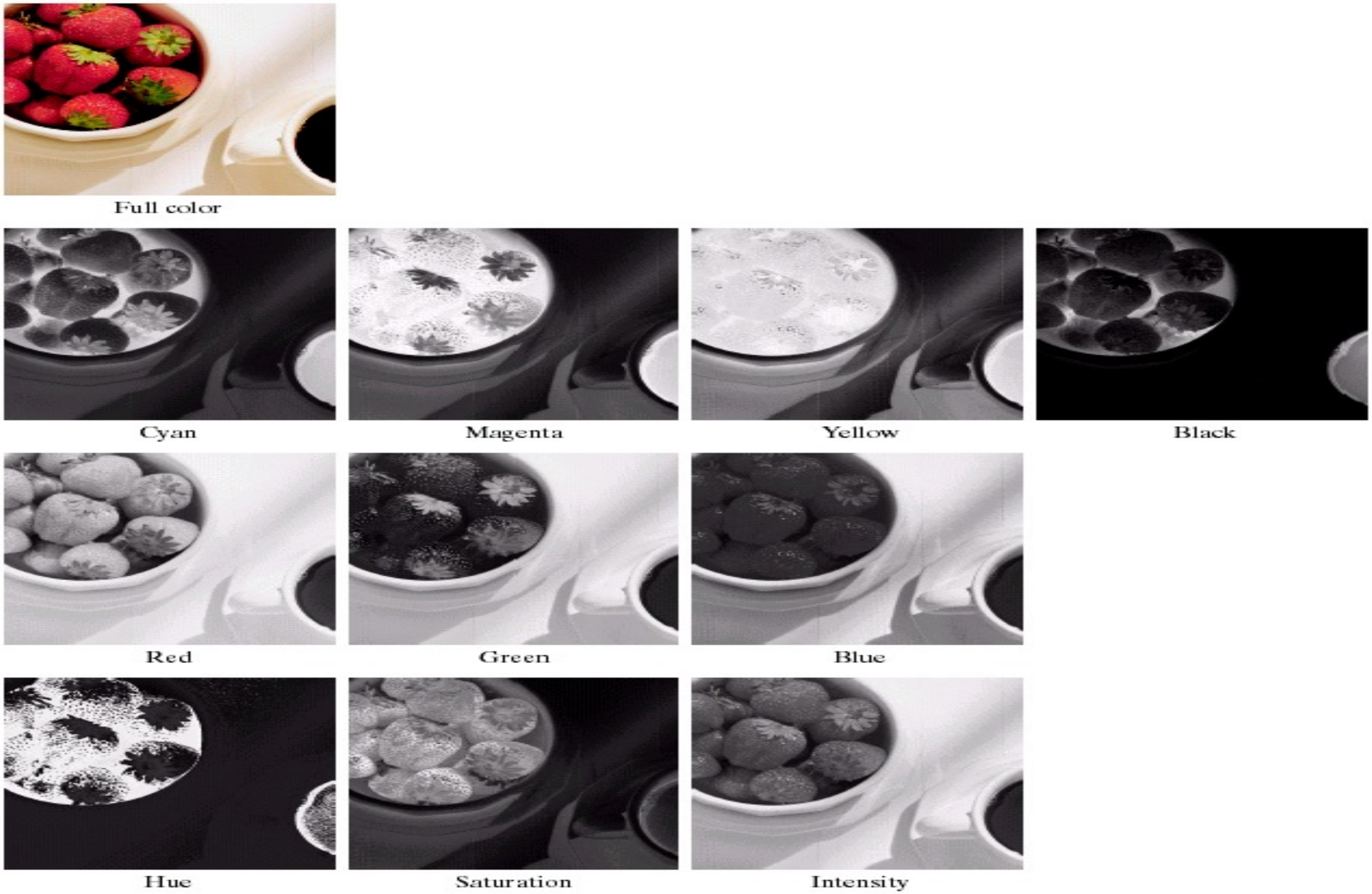
- YUV is the color coordinate used in color TV in PAL system, somewhat different from YIQ.
- YCbCr is the digital equivalent of YUV, used for digital TV, with 8 bit for each component, in range of 0-255

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

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

Conversion from RGB to Y is how we generate a “gray scale” version of a color image

Comparison of Different Color Spaces



Color Coordinate Conversion (summary)

- Conversion between different primary sets are linear (3x3 matrix)
- Conversion between primary and XYZ/YIQ/YUV are also linear
- Conversion to HSI/Lab are nonlinear
  - HSI and Lab coordinates
    - he Euclidean distance between two colors in these HAS and Lab is proportional to actual color difference
- Conversion formulae between many color coordinates can be found in [Gonzalez2008]
- Color space of HDTV by Recommendation 709 can be found in [Woods2012]

Grayscale Image Specification

- Each pixel value represents the brightness of the pixel. With 8-bit image, the pixel value of each pixel is 0 ~ 255
- Matrix representation: An image of MxN pixels is represented by an MxN array, each element being an unsigned integer of 8 bits

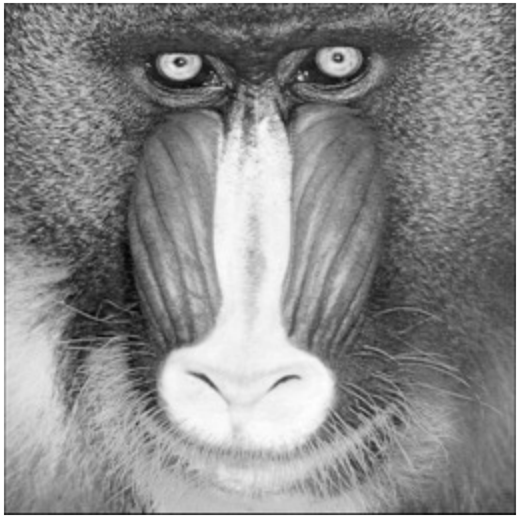
$$M = \begin{bmatrix} 160 & 162 & \cdots & 166 & 154 \\ 162 & 158 & \cdots & 122 & 69 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 60 & 55 & \cdots & 79 & 94 \\ 58 & 55 & \cdots & 99 & 109 \end{bmatrix}$$

Color Image Specification

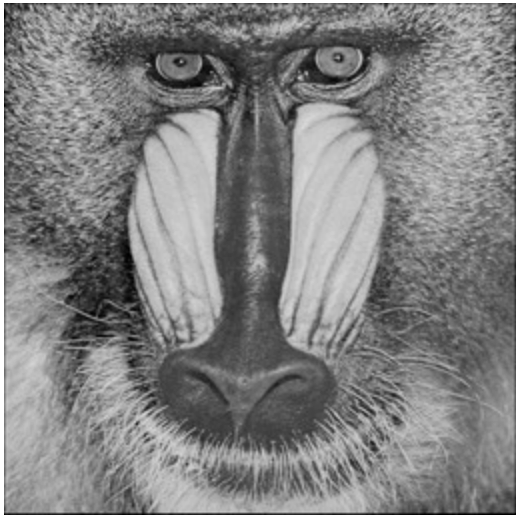
- Three components
  - M = {R, G, B}

$$R = \begin{bmatrix} 73 & \cdots & 87 \\ \vdots & \ddots & \vdots \\ 27 & \cdots & 17 \end{bmatrix}, G = \begin{bmatrix} 66 & \cdots & 98 \\ \vdots & \ddots & \vdots \\ 36 & \cdots & 13 \end{bmatrix}, B = \begin{bmatrix} 31 & \cdots & 61 \\ \vdots & \ddots & \vdots \\ 36 & \cdots & 14 \end{bmatrix}$$

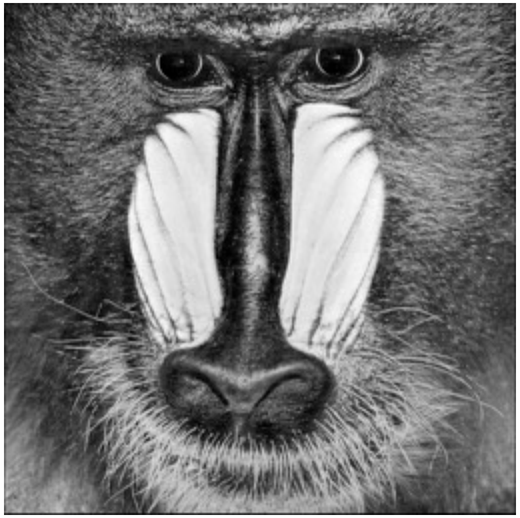
R



G



B



Red nose is brightest!

Blue Cheek is brightest

Image Capture and Display

- Light reflection physics
- Imaging operator
- Color capture
- Color display

Gray and Color Image Capture

- Gray images are captured by a single sensor, sensitive to the entire visible spectrum, similar to the rods
- Color images are captured by having three sensors, each sensitive to a different primary color, similar to the cones
  - Alternatively using a single sensor proceeded by different color filters
- Each sensor or filter has its own frequency response, which may differ from the cones responses in the human retina.

Image and 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.

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

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

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

Courtesy of Onur Guleryuz

More on Image/Video Capture

- Camera absorption function

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

X: 3D position in a world coordinate
$a_c(\lambda)$: the camera sensor frequency response

- Projection from 3-D to 2-D camera plane

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

x: 2D position in camera plane

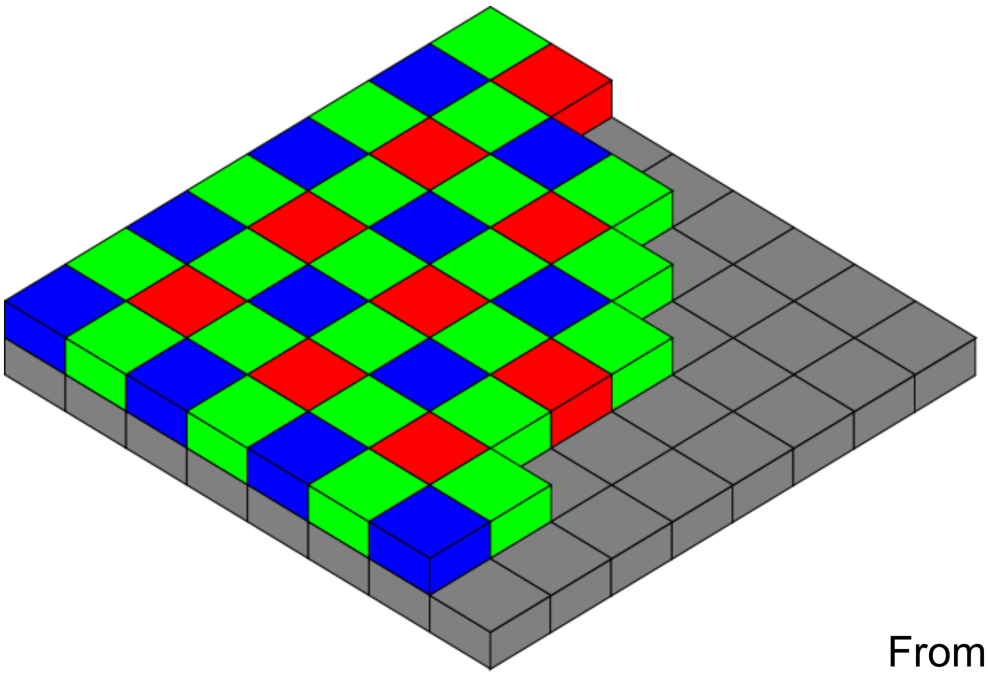
Image function

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

- The projection operator is non-linear
  - Perspective projection
  - Orthographic projection
  - More on this later

Color Imaging Using Color Filter Arrays

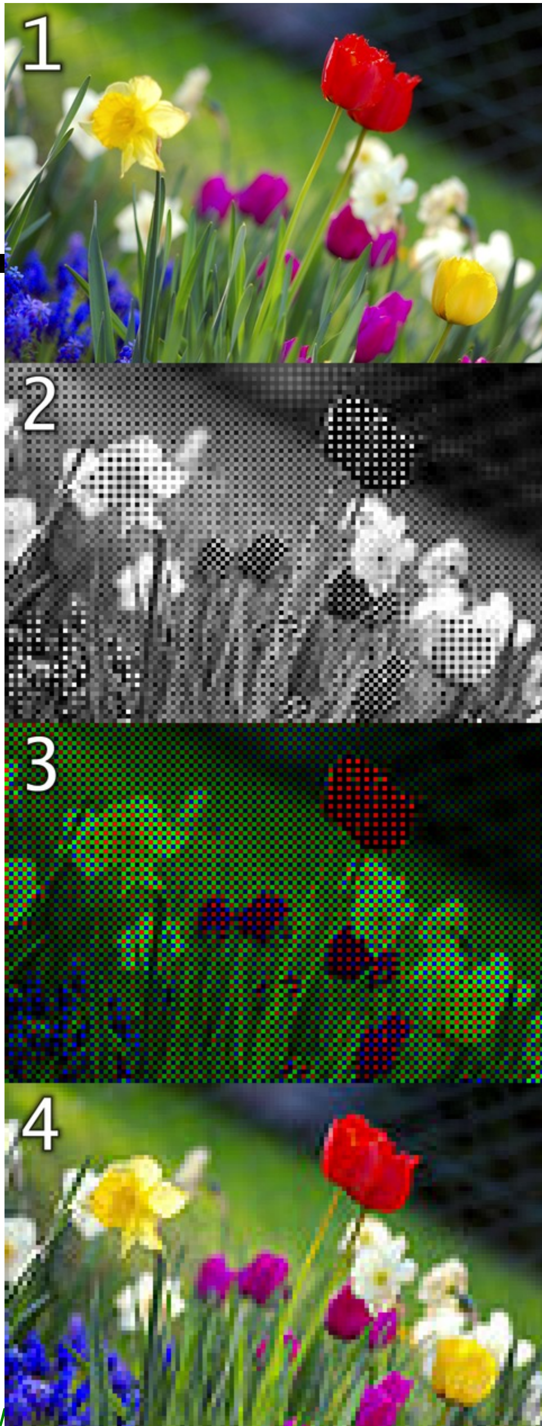
- Single sensor array, with different color filters to separate the RGB primary components
- Sensors: CCD, CMOS



Bayer RGB Pattern
(Each 2x2 pixels contains 2 green, 1 red, 1 blue)

Human eye is more sensitive to the high frequency in the luminance (mostly determined by green)

From http://en.wikipedia.org/wiki/Bayer_filter



From http://en.wikipedia.org/wiki/Bayer_filter

1: original scene
2: output of a 120x80 pixel sensor with a Bayer filter
3: output color coded
4: Reconstructed image after interpolating missing colors (color demosaicing)

Gray and Color Image Display/Printing

- Gray images are displayed by a single light sensitive diode, with intensity proportional to gray level.
- LCD monitors display color images by having three phosphors at slightly shifted positions near each pixel, each generating a different primary color (red, green, blue)
  - Our eye does the interpolation!
- Color images are printed by having three color inks (cyan, magenta, yellow). High end printers use more inks to produce a larger color gamut and more vivid colors.
- The color gamut of display primary colors differs from that of printing primary colors.
  - Some displayed colors cannot be reproduced in printing.

Gamma Correction

- Displayed light intensity is nonlinearly related to the actual intensity following the Gamma rule

$$g = af^r$$

- Gamma correction pre-compensates this nonlinearity inside the camera

$$h = f^{1/r}$$

$$g = ah^r = af$$

- Gamma value depends on the display device, typically gamma~2.2

Digital Video Formats

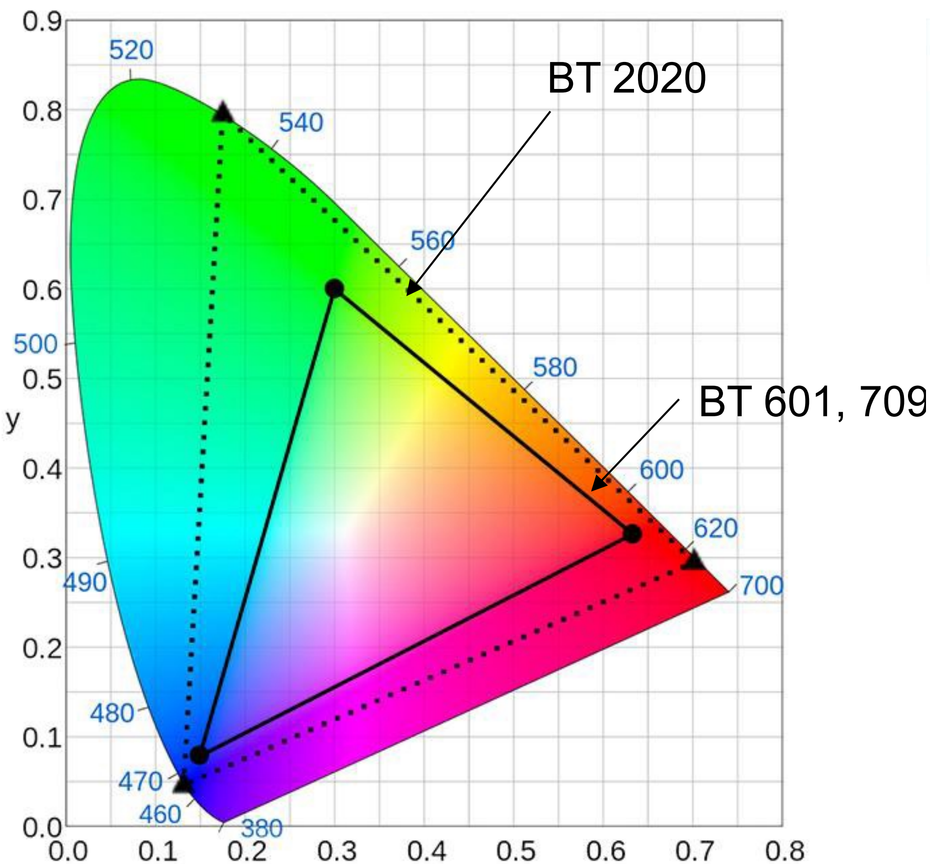
- Standard Definition (SD) video (BT. 601)
  - Resolution: 720x480, 25-30Hz (frames/sec or fps)
  - Interlaced (even or odd lines only per field) or progressive
  - Standard Dynamic Range (SDR): 0.0002 to 100 cd/m2 (nits) in luminance range, represented in 8bits/color
  - Color spec: BT. 709
    - Narrow color range: 33.5% of visible range
- High Definition (HD) video (BT. 709)
  - Resolution increase to 1920x1080, up to 60 Hz (1080p or 2K video)
  - Wide aspect ratio: 16:9 or 2:1
  - Color spec: BT. 709
- Ultra High Definition (UHD) video
  - Resolution further increase to 3840x2160 (4K video) or higher, up to 120Hz
  - High dynamic range (HRD): 0,00005 to 1000 cs/m2, represented by 16bits/color (10-12 bpp for distribution, 14-16bpp for production)
  - Wide color gamut: DCI-P3 (41.8%) and BT.2020 (57.3%).

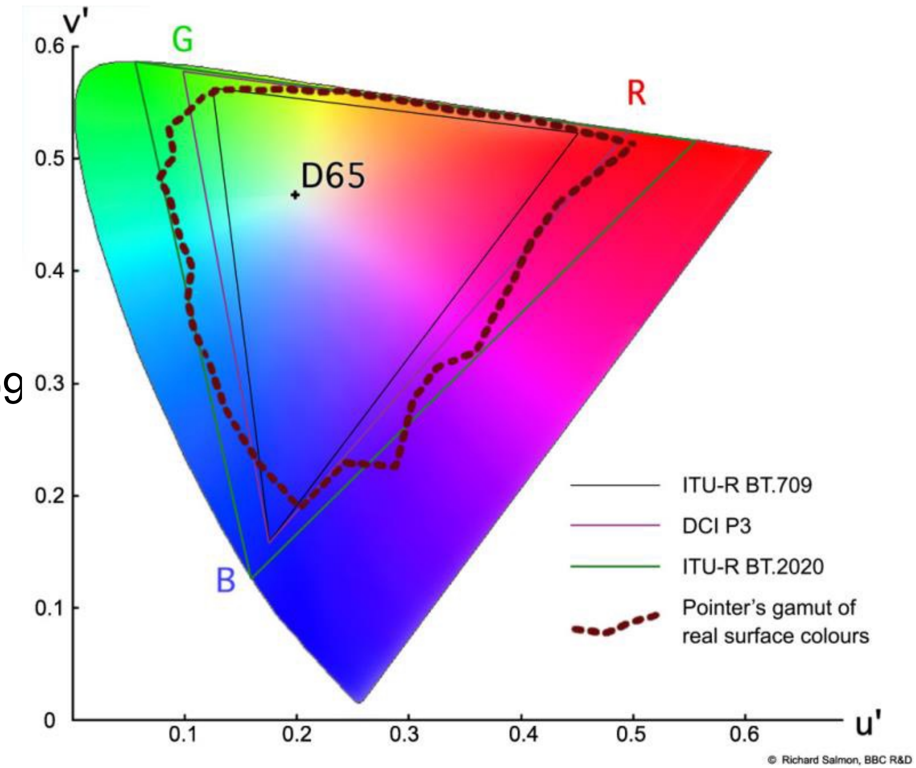
Primary Color Specifications

- CIE 1931: analog TV (interlaced)
- BT.601: Used in SDTV (interlaced or progressive) and HDTV (now mostly progressive)
- DCI-P3: Used in cinema presentation
- BT.2020: Specified for UHD, covers 99.9% of Pointer's gamut

- BT.2020 primaries are maximum saturation pure colors, created by extremely narrow spectral slices of light energy: Red (630nm), Green (532 nm), Blue (467nm)
- Recall CIE RGB primaries are Red (700nm), Green (546.1nm), Blue (435.8nm)

Color Gamut Comparison





From [Schulte2016]

Color Calibration

- Each camera/display may use R,G,B primaries somewhat different
- Captured images also depend on scene lighting
- Using a reference white, rescale the RGB values so that “white” will have equal R,G,B values
  - Important when the images are captured under non-white illumination
- BT. 709 uses daylight illuminate D_65 as standard white

SDR vs. HDR

From: http://files.spectracal.com/Documents/White%20Papers/HDR_Demystified.pdf

HDR TV: captured using HDR camera and displayed by HDR display
HDR photography: computed HDR image using multiple captured images using SDR cameras

Summary: What should you know?

- What is light? Difference between illuminating and reflecting light? What are the attributes used to describe the color?
- How human perceive color? Functions of cones and rods in the retina?
- How to produce different colors ? What are primary colors? How to judge the “goodness” of color primaries?
- How to represent a color? Different color models? Meaning of chromaticity?
- How to capture a color image?
- How to display a color image?
- How to print a color image?
- How to specify a color image digitally?
- What are some standard image formats?
- You should be able to answer all the questions in this slide as a review of this lecture. Similarly for following lectures.

Reading Assignments

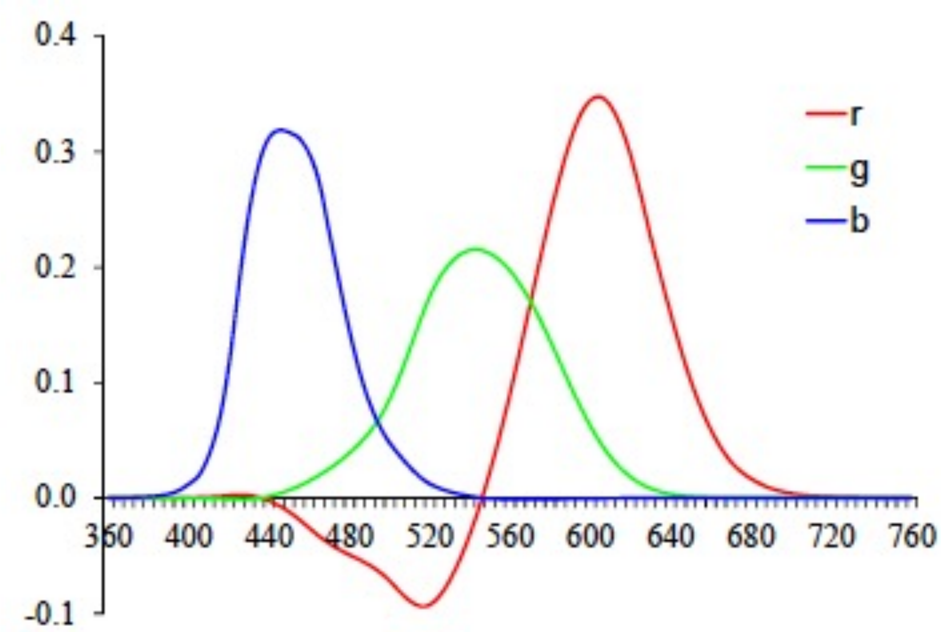
- [Szeliski2021] Richard Szeliski, Computer Vision: Algorithms and Applications. Ch. 2. (Sec. 2.2.1, 2.3)
- Optional: [Schulte2016] Tom Schulte, Joel Barsotti, "HDR Demystified – Emerging UHDTV Systems," Mar. 2016. http://files.spectracal.com/Documents/White%20Papers/HDR_Demystified.pdf

- Other reference:
- [Wang2002] Wang et al, Digital video processing and communications. Chap 1 (Sec. 1.3-1.4 optional)
- [Gonzalez2008] Gonzalez & Woods, "Digital Image Processing", Prentice Hall, 2008, 3rd ed. Chapter 3 (Section 6.1 – 6.2)
- [BT2020] "BT.2020 : Parameter values for ultra-high definition television systems for production and international programme exchange". International Telecommunication Union. 2012-08-23. Retrieved 2014-08-31.

Supplementary Material:

Color Primary Specification Using CIE XYZ Chromaticity

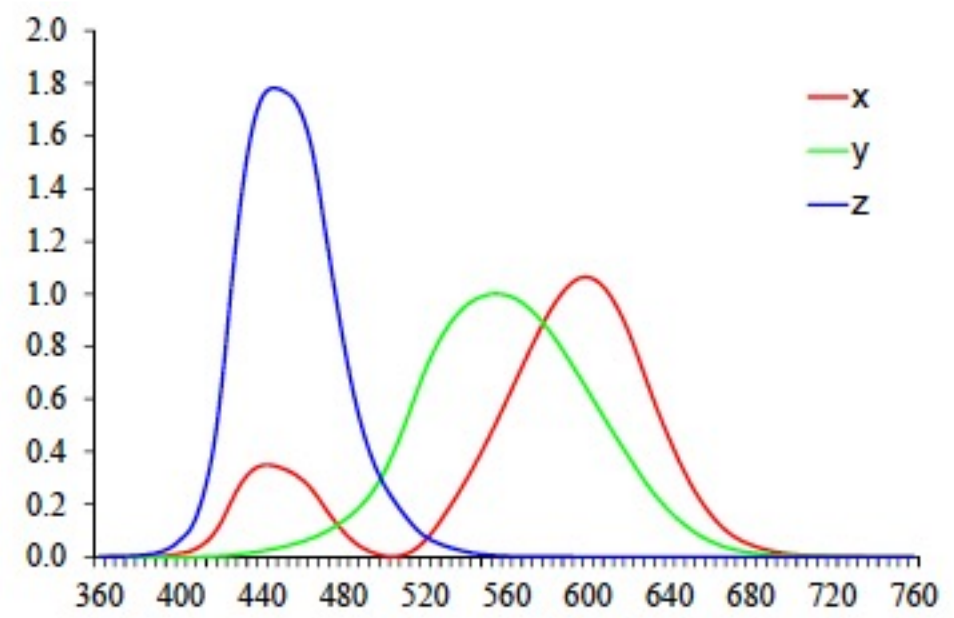
CIE 1931 RGB Primaries



from [Szeliski2010]

- Monochromatic primaries: Red (700nm), Green (546.1nm), Blue (435.8nm)
- Color matching functions: specify the amount of each primary to match a particular monochromatic color
- Need “negative amount” of red to match some colors: add this amount of red to the test color to match the sum of remaining primaries

CIE 1931 XYZ Primaries



from [Szeliski2010]

- XYZ do not correspond to real color primaries (Y=luminance, Z=blue, X: mixed). They are *imaginary* primary colors.
- Using these color matching functions can however represent any single spectral color using non-negative tristimulus values

Tristimulus Values and Chromaticity

- Tristimulus values
  - The amounts of the three primary colors needed to form any particular color are called the **tristimulus values**, denoted by X, Y, Z (or any other three symbols).
  - For any color with spectrum $S(\lambda)$

$$X = \int S(\lambda)x(\lambda)d\lambda, \quad Y = \int S(\lambda)y(\lambda)d\lambda, \; Z = \int S(\lambda)z(\lambda)d\lambda$$

- Trichromatic (or chromaticity) coefficients

$$x = \frac{X}{X+Y+Z}, \quad y = \frac{Y}{X+Y+Z}, \quad z = \frac{Z}{X+Y+Z}.$$

  - Only two chromaticity coefficients are necessary to specify the chrominance of a light.

$$x + y + z = 1$$

Conversion between CIE 1931 RGB and XYZ

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \frac{1}{0.17697} \begin{bmatrix} 0.49 & 0.31 & 0.20 \\ 0.17697 & 0.81240 & 0.01063 \\ 0.00 & 0.01 & 0.99 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix}$$

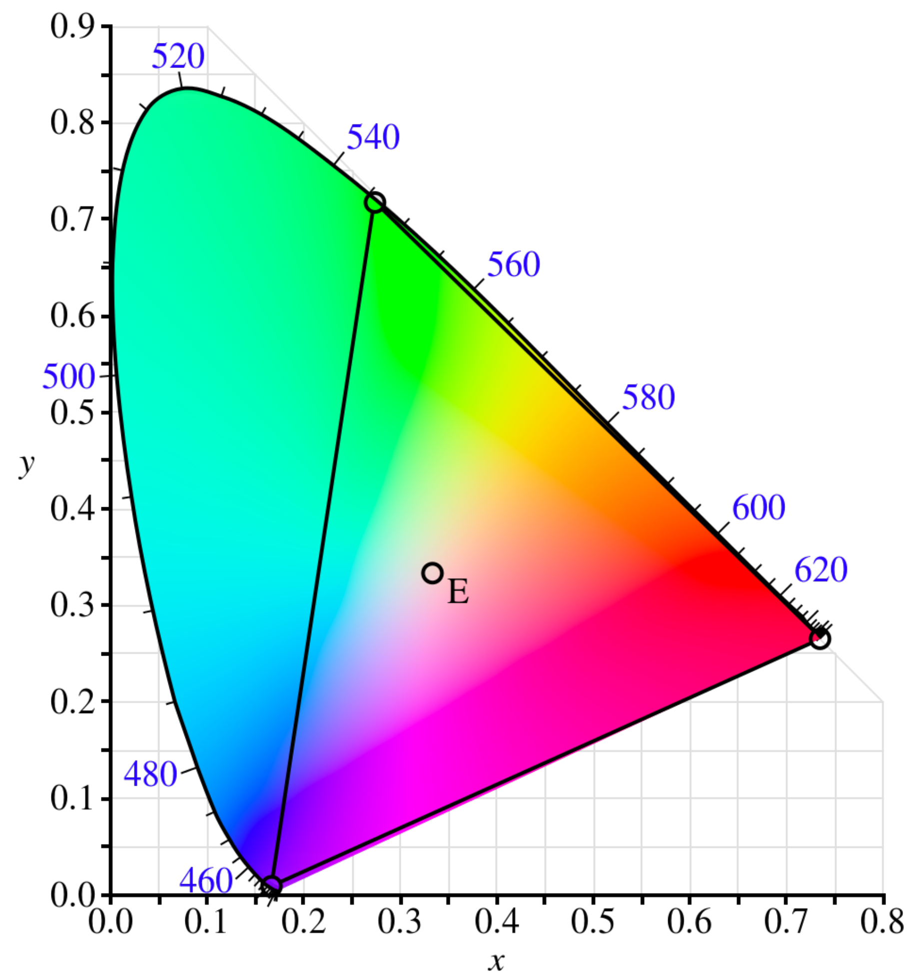
from [Szeliski2010]

The three columns in the matrix are the tristimulus values of the R,G,B primaries defined in terms of CIE XYZ primaries. From these, you could derive the chromaticity coefficients of RGB primaries and correspondingly locate them in the CIE chromaticity diagram.

Ex: Red: x=0.49/(0.49+0.17697+0)=0.735, y=0.17697/(0.49+0.17697+0)=0.265.

CIE 1931 XYZ Chromaticity Diagram

- Shows all visible colors by humans.
- Colors on the boundary: spectrum colors, highest saturation.
- Mixing any three colors in the visible range can generate colors in the triangle formed by these three points.
- Not all visible colors can be reproduced by CIE RGB primaries.
- The triangle on the right shows the color gamut of CIE RGB primary.
- Printing gamut using CMY is smaller than display gamut using RGB.



from [https://commons.wikimedia.org/wiki/File:CIE1931xy_CIERGB.svg]

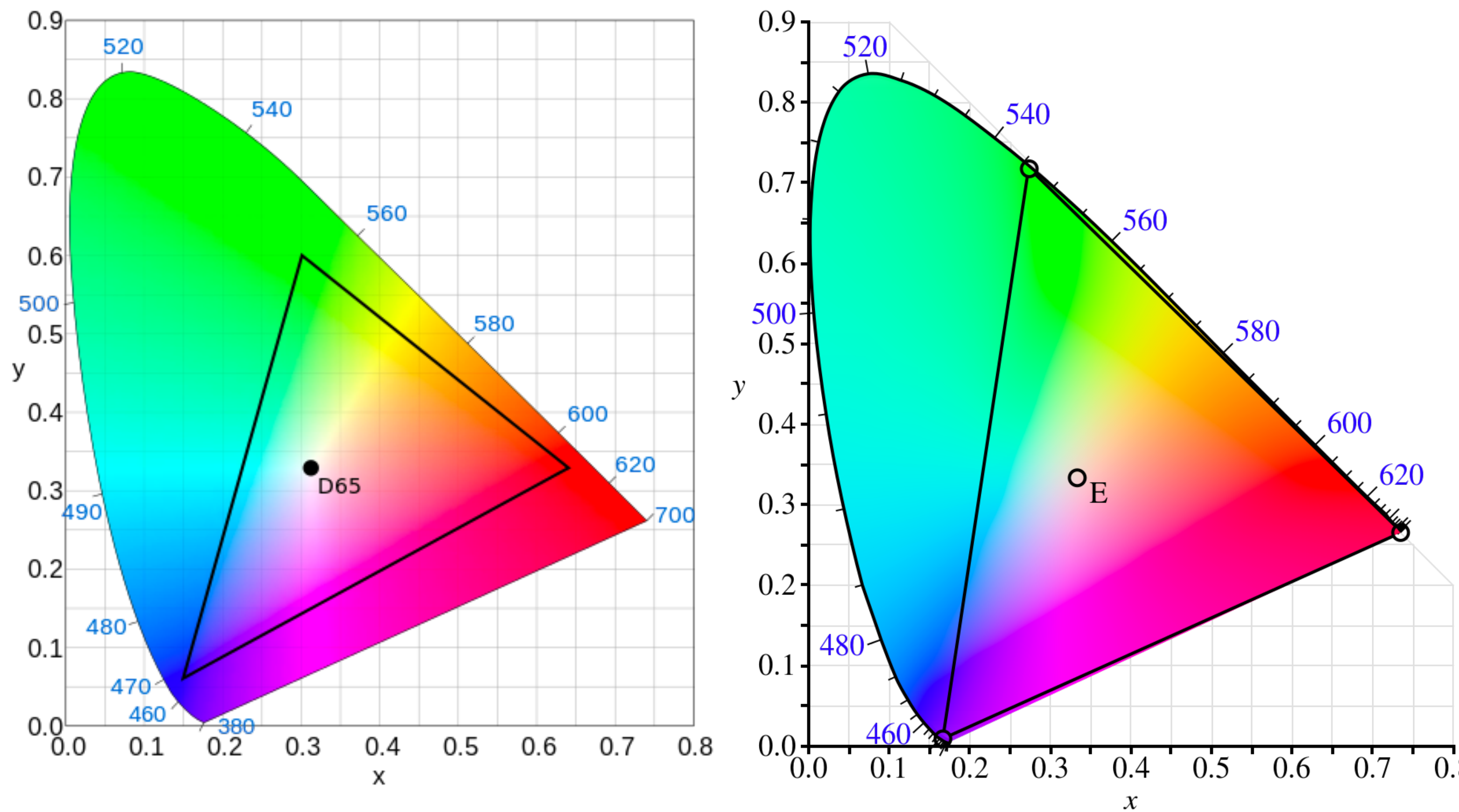
RGB Color Primary Defined by ITU Rec. 709 for Digital TV (BT 709)

$$\begin{bmatrix} R \\ G \\ B \end{bmatrix} = \begin{bmatrix} 3.240479 & -1.537150 & -0.498535 \\ -0.969256 & 1.875992 & 0.041556 \\ 0.055648 & -0.204043 & 1.057311 \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \end{bmatrix}$$

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \begin{bmatrix} 0.412453 & 0.357580 & 0.180423 \\ 0.212671 & 0.715160 & 0.072169 \\ 0.019334 & 0.119193 & 0.950277 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix}$$

The three columns in the RGB-> XYZ conversion are the tristimulus values of the R,G,B primaries defined in terms of CIE XYZ primaries. From these, you could derive the chromaticity coefficients of RGB primaries and correspondingly locate them in the CIE diagram

BT.709 (left) vs. CIE 1931 (right) Color Gamut



https://en.wikipedia.org/wiki/Rec._709#/media/File:CIExy1931_Rec_709.svg

New Color Space for UHD: BT.2020

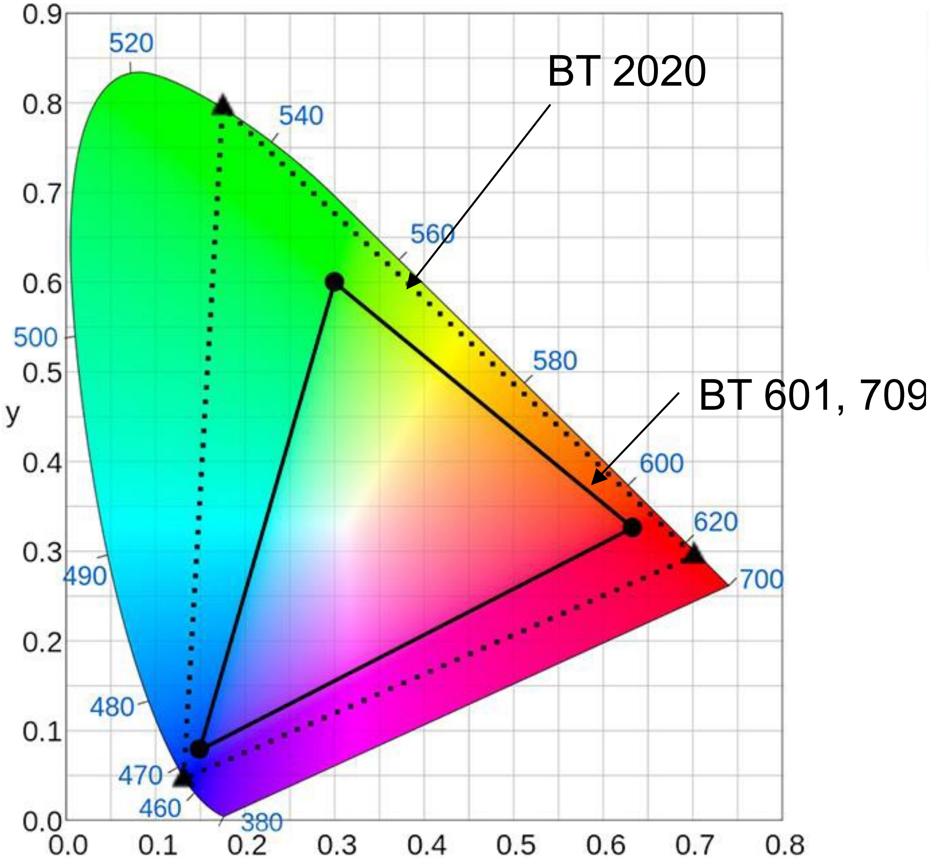
Chromaticity coordinates (CIE, 1931)	x	y
Red primary (R)	0.708	0.292
Green primary (G)	0.170	0.797
Blue primary (B)	0.131	0.046
Reference white (D65)	0.3127	0.3290

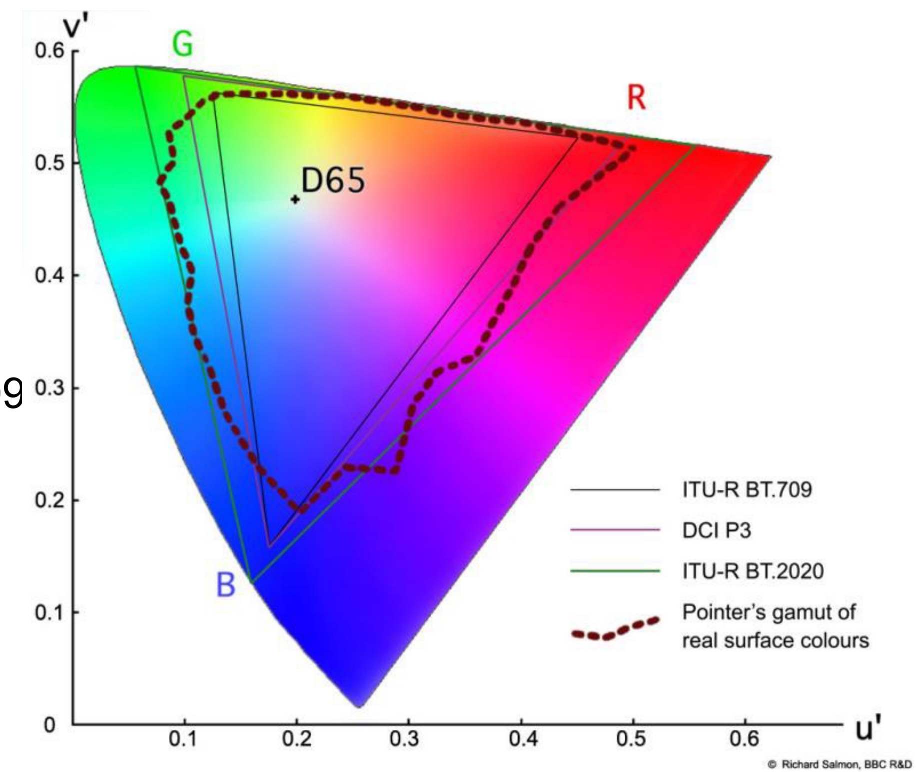
BT.2020 primaries are maximum saturation pure colors, created by extremely narrow spectral slices of light energy: Red (630nm), Green (532 nm), Blue (467nm)

Recall CIE RGB primaries are Red (700nm), Green (546.1nm), Blue (435.8nm)

From [BT2020]

Color Gamut Comparison





From [Schulte2016]

UHD Alliance Definition of UHD TV

- **Resolution**: 3840x2160 for content, distribution, and playback displays.
- **Color bit depth**: 10 bits minimum for content and distribution, 10 bits for playback displays.
- **Color representation**: BT.2020 for content, distribution, and playback displays.
- **Mastering display**: 0.03- 1,000 nits
- **Playback display:** 0.05-1,000 nits, or 0.0005-540nits
- UHD content backwards compatible with SDR displays