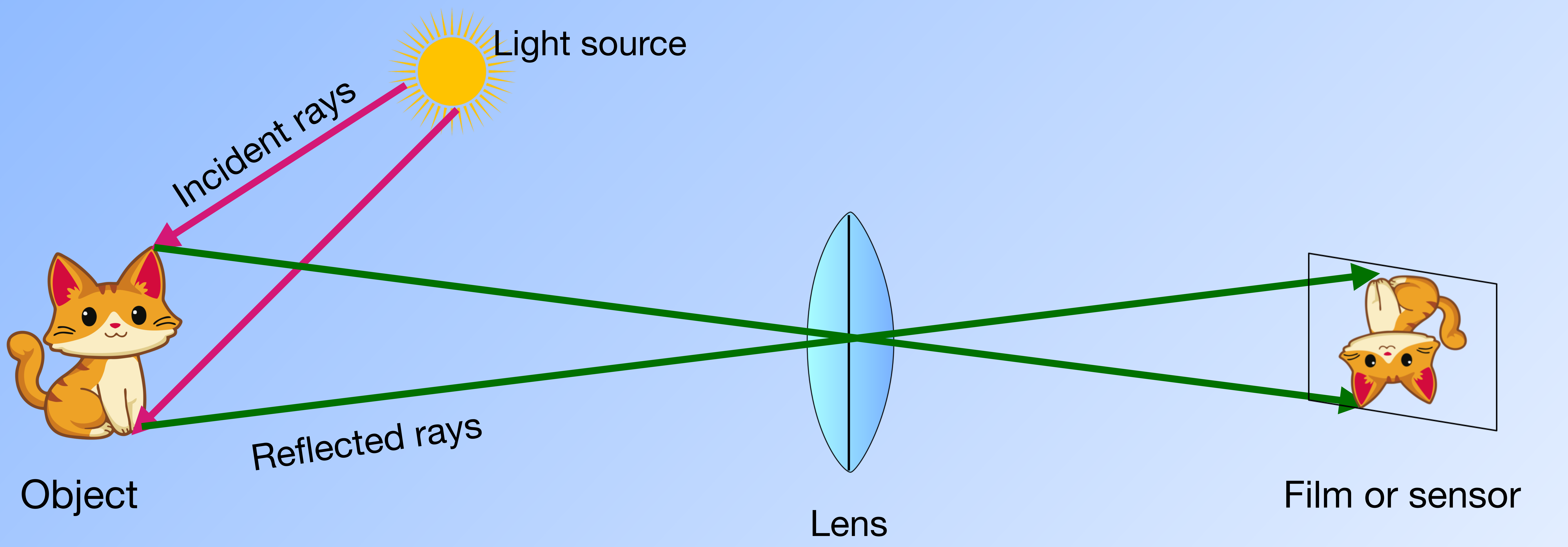
AI61201: Visual Computing With AI/ML

Module 1: Image Formation and Representation

Dr. Somdyuti Paul

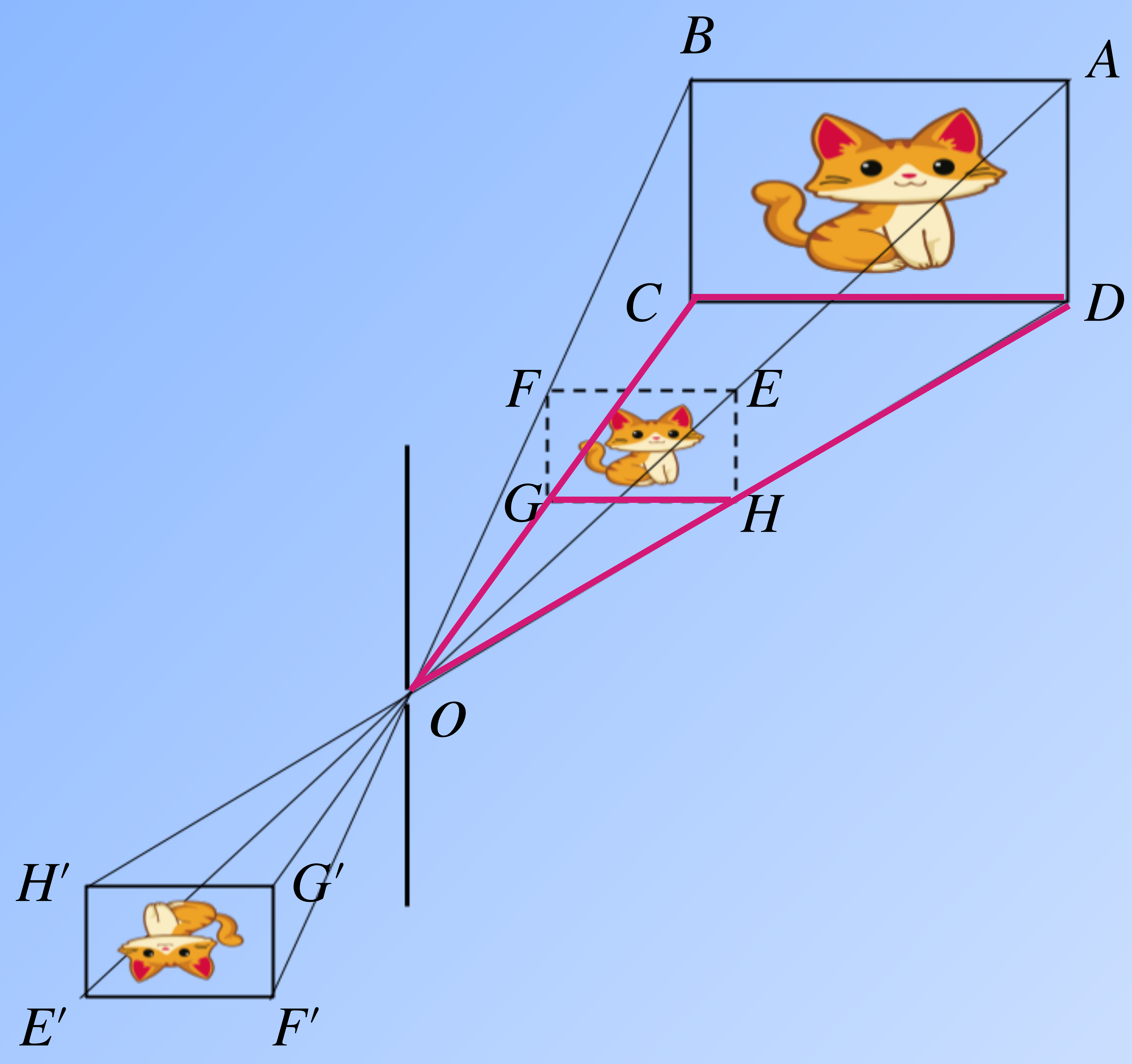
Image Formation

- An photographic image is formed when visible light is focused by a lens and onto a film or sensor.
- The process of image formation involves a projection from the 3D world onto a 2D image plane.



Perspective Projection

In perspective projection, 3D points in the real world are projected on the 2D image plane in such a way that farther an object is from the lens, the smaller its image is.



Let the lens center O bet taken as the origin of the coordinate system, the coordinates of point A and E be (X, Y, Z) and (x, y), respectively.

$$OG = f, OC = Z$$

Using triangle similarity, we have

$$\Delta OGH \sim \Delta OCD$$

$$\frac{GH}{OG} = \frac{CD}{OC} \implies \frac{x}{f} = \frac{X}{Z}$$

$$\Delta OGF \sim \Delta OCB$$

$$\frac{GF}{OG} = \frac{CB}{OC} \implies \frac{y}{f} = \frac{Y}{Z}$$

Thus, $(x, y) = \dfrac{f}{Z}(X, Y)$

Image Sampling

- Photons impinging on a sensor are converted to a voltage signal via the photovoltaic effect.
- Sampling refers to the process involved in converting the analog signal captured by the sensors to a digital signal.
- Sampling involves taking uniformly spaced samples from the analog signal.
- Improper sampling of the signal leads to appearance of Moiré patterns or aliasing artifacts.

Aliased images

Properly sampled images

Image Sampling

- The number of samples should be large enough to enable a faithful representation of the original signal - i.e. avoid *aliasing*.
- To avoid aliasing, the sampling frequency should be at least twice the highest frequency present in the input signal, i.e. $f_s \geq 2f_{\text{max}}$, where f_s is the Nyquist sampling rate.
- When sensor arrays are used for image acquisition, the number of sensor elements governs the sampling rate achieved.

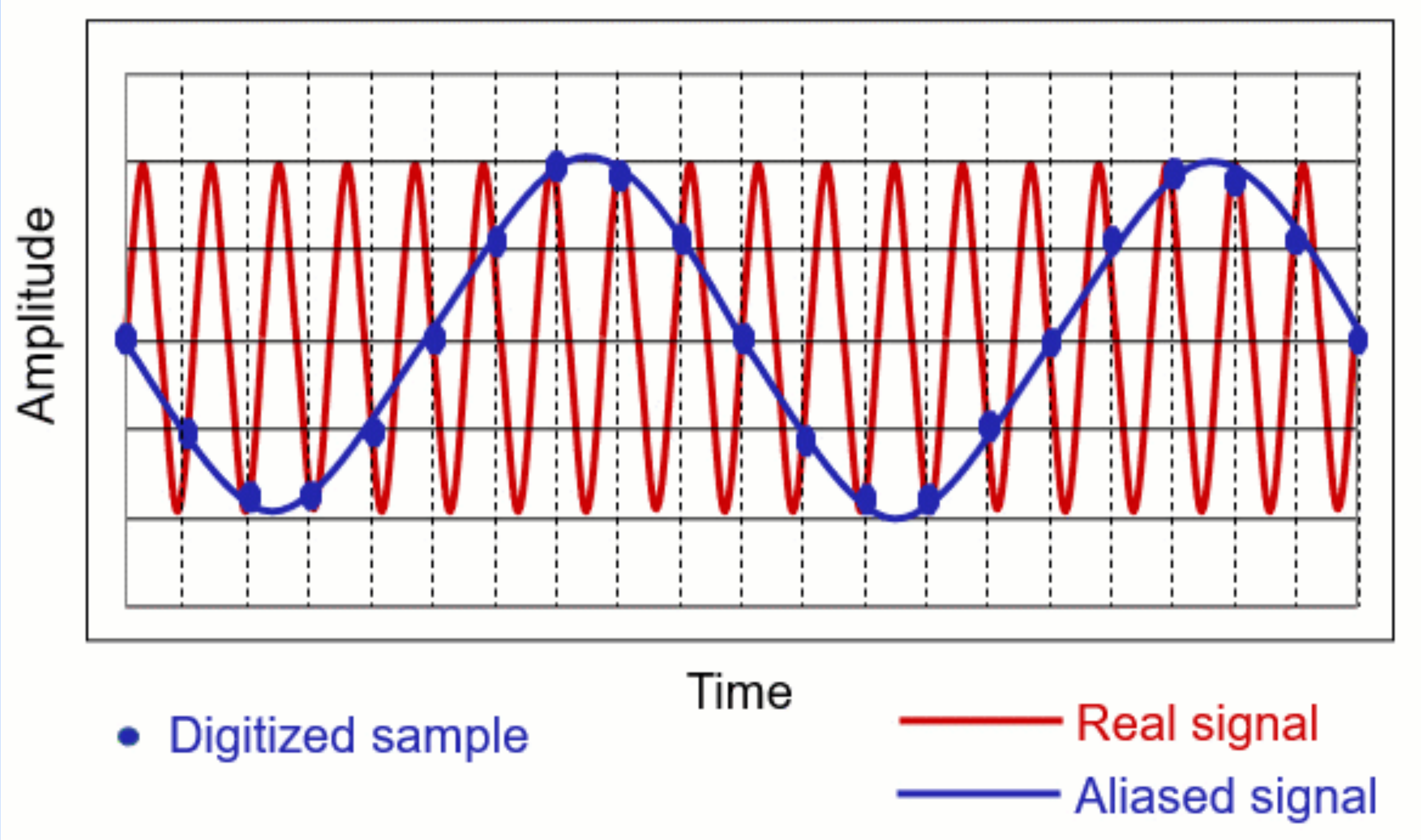


Image Quantization

- An analog signal can take on continuous values in a given range.
- Quantization refers to limiting the number of permissible levels that the signal can take.
- N bits can represent 2^N intensity levels.
- The more the number of quantization levels, the more shades or intensity values could be represented.
- After both sampling and quantization, the analog image signal is converted to a digital image signal, composed indivisible elements called *pixels*.
- Each pixel is associated with a discrete location and a discrete value.

256 gray levels

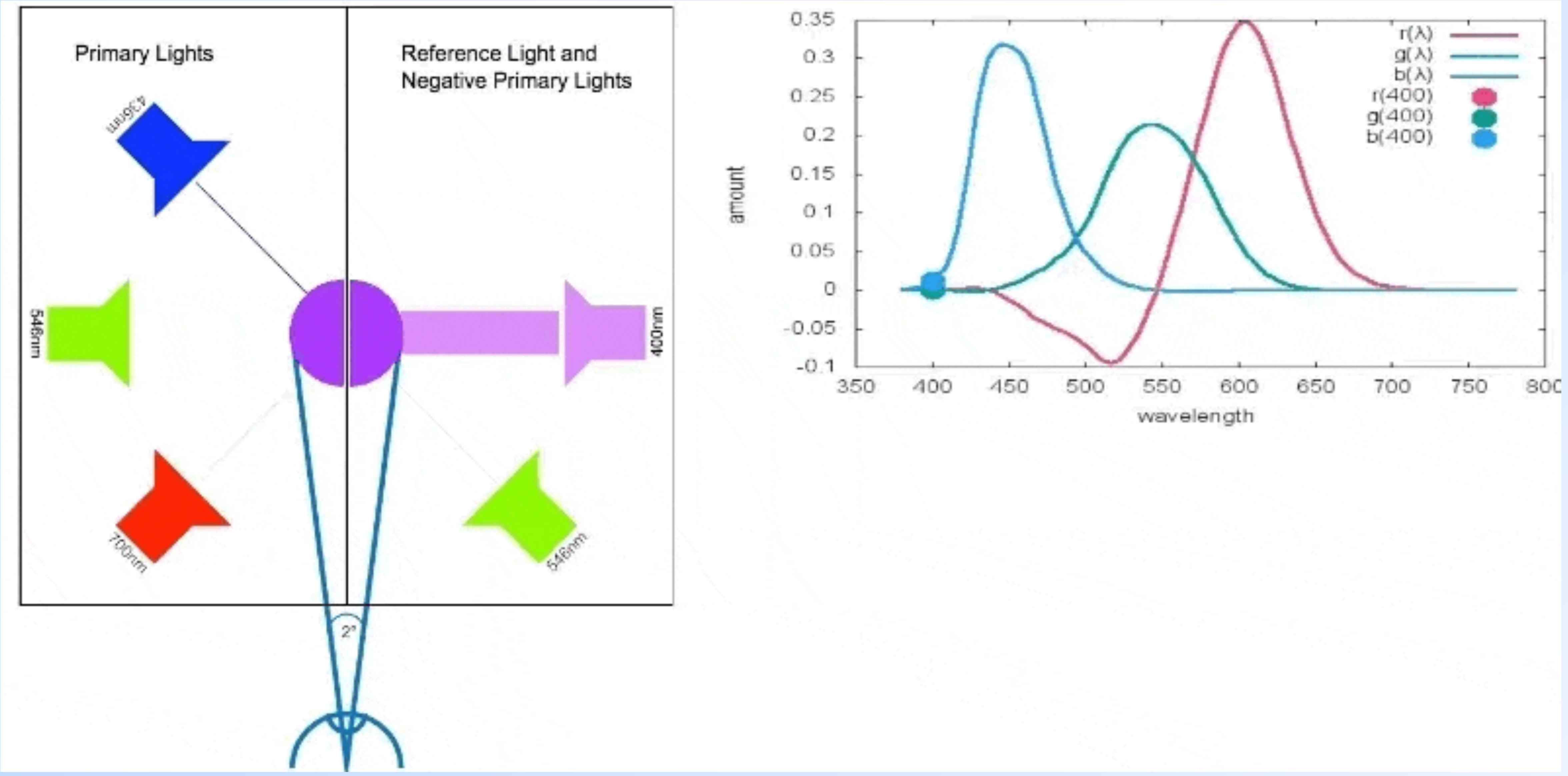
8 gray levels

4 gray levels

2 gray levels

Tri-chromatic Theory and Color Matching

- Color is an important aspect of any visual data.
- Color images are formed by the interaction of three primary colors: red (R), green (G) and blue (B).
- All monochromatic colors can be reproduced by mixing three suitably chosen primaries.
- The recipe for mixing the primaries was derived through color matching experiments (CIE 1931).
- This experiment was used to derive the color matching functions (CMFs), denoted by $r(\lambda)$, $g(\lambda)$ and $b(\lambda)$, which provide the rule to infer the components of any possible color.



Experimental Setup of Color Matching Experiment using R, G and B primaries

Tri-chromatic Theory and Color Matching

- The color matching function used the following wavelengths as primaries:

  R: 700 nm (red)

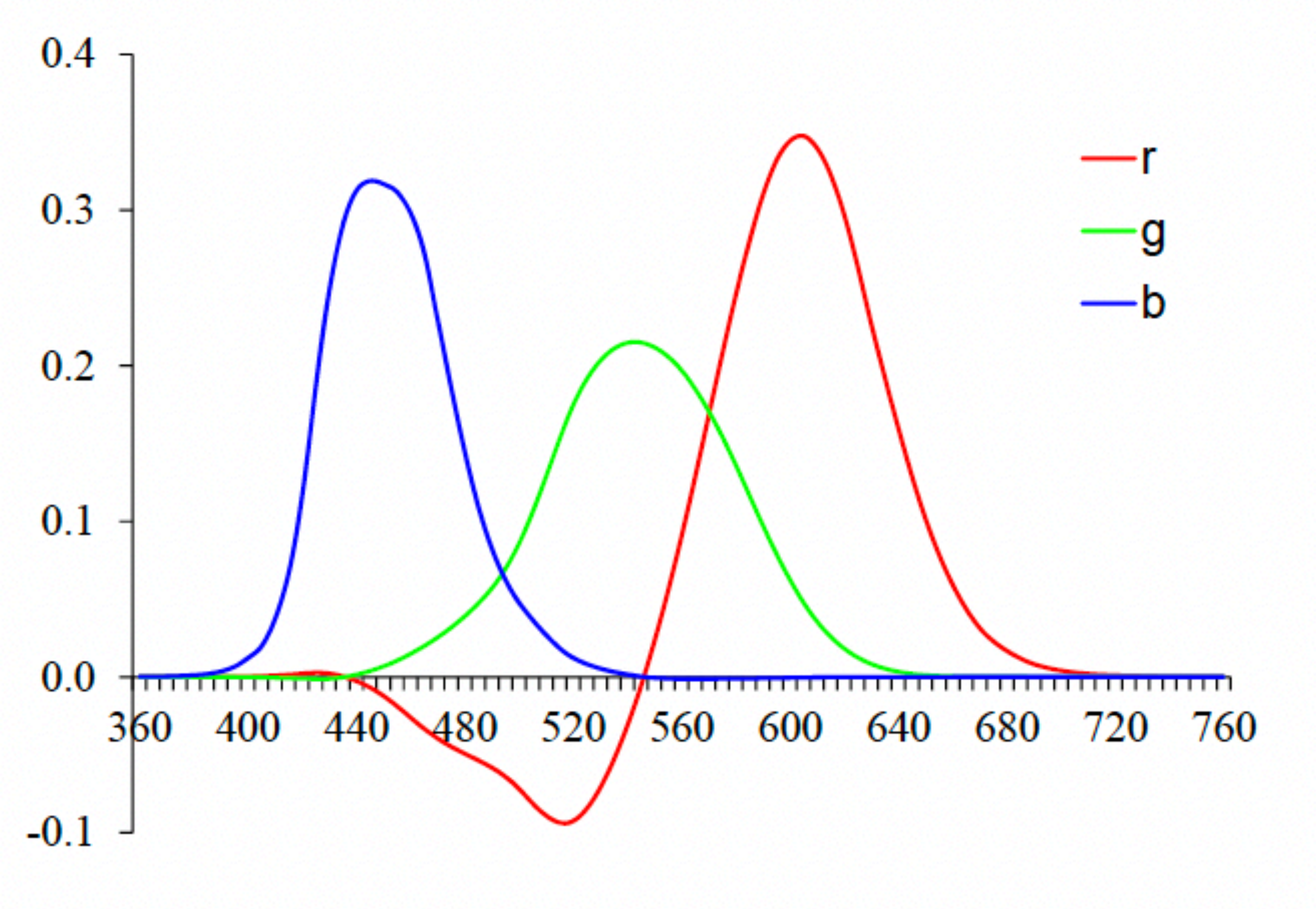
  G: 546.1 nm (green)

  B: 435.8 nm (blue)

- Once the matching functions are determined, the composition of a color having wavelength can be determined as :

  $U(\lambda_0) = r(\lambda_0)R + g(\lambda_0)G + b(\lambda_0)B$

- However, to match certain colors in the blue-green range using these three primaries, a negative amount of red primary were to be added, which seems counterintuitive.



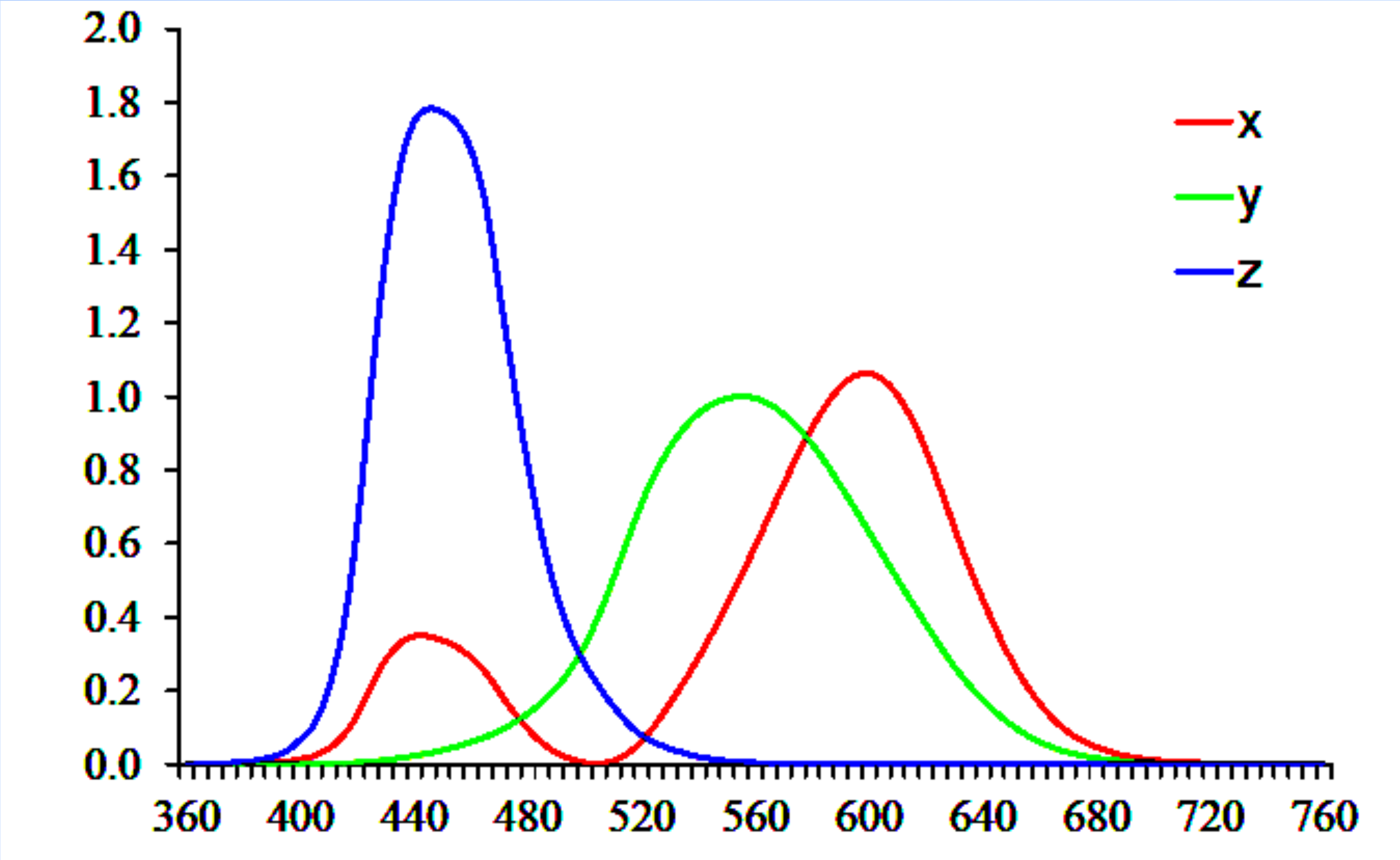
RGB Color Matching Functions

Tri-chromatic Theory and Color Matching

- To circumvent this problem, a new color space called XYZ was developed.
- The transformation from the RGB primaries to the XYZ primaries are given as follows:

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \begin{bmatrix} 0.4124564 & 0.3575761 & 0.1804375 \\ 0.2126729 & 0.7151522 & 0.0721750 \\ 0.0193339 & 0.1191920 & 0.9503041 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix}$$

- The color matching functions using the transformed primaries X, Y, Z have strictly positive values.
- However, X, Y, Z are not real primaries.
- The color matching functions $x(\lambda)$, $y(\lambda)$ and $z(\lambda)$ can be reasonably approximated by sum of Gaussians.



XYZ Color Matching Functions

$$x(\lambda) = 1.065\exp\left(-\frac{1}{2}\left(\frac{\lambda - 595.8}{33.8}\right)^2\right) + 0.366\exp\left(-\frac{1}{2}\left(\frac{\lambda - 446.8}{19.44}\right)^2\right)$$

$$y(\lambda) = 1.065\exp\left(-\frac{1}{2}\left(\frac{\ln\lambda - \ln 556.3}{0.0075}\right)^2\right)$$

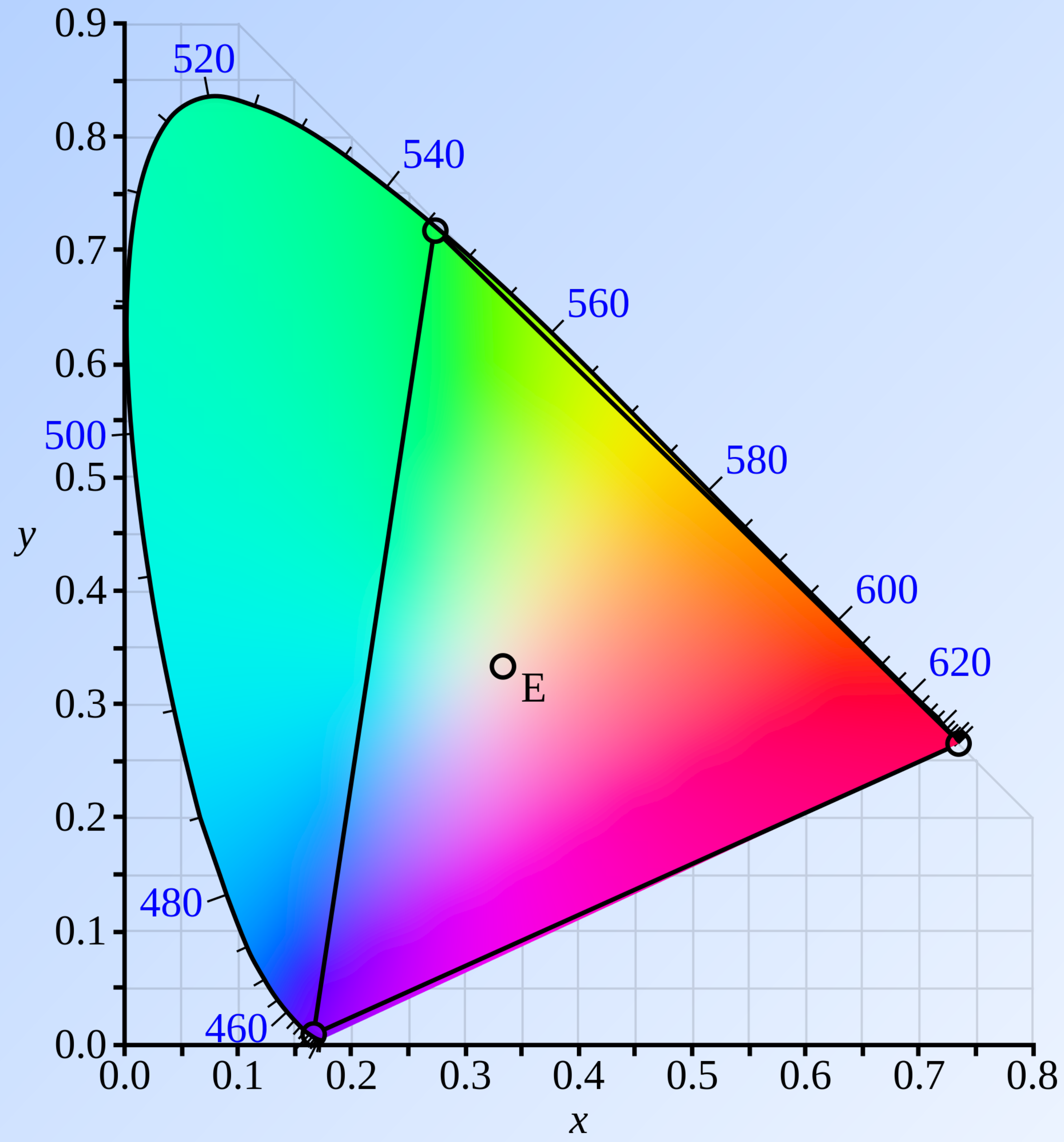
$$z(\lambda) = 1.839\exp\left(-\frac{1}{2}\left(\frac{\ln\lambda - \ln 449.8}{0.051}\right)^2\right)$$

Tri-chromatic Theory and Color Matching

- The values of X, Y and Z are normalized by dividing by the absolute intensity to give the chromaticity co-ordinates.

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

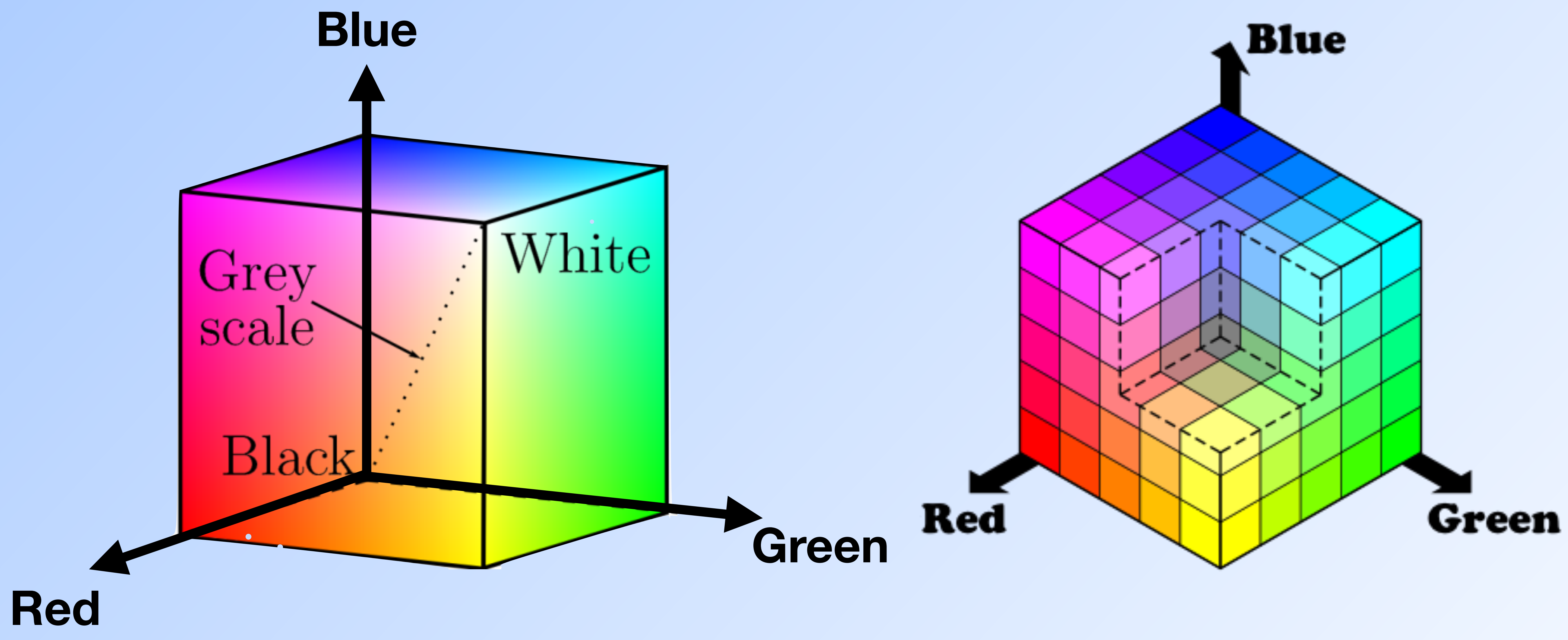
- The chromaticity coordinates are used to specify the gamut of colors perceivable by humans.
- The chromaticity diagram is widely used to perform color mixing and color calibration of color printing and color display devices.



Chromaticity diagram and color gamut

Color Images

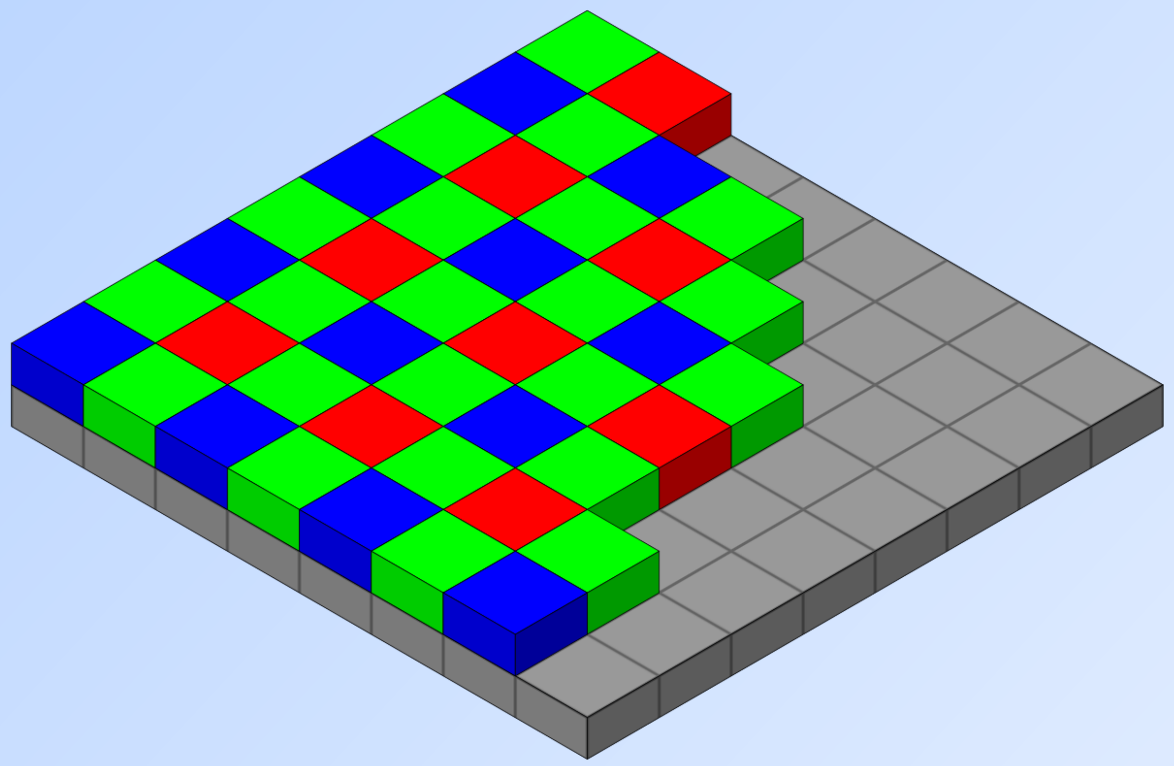
- Color images are formed by the interaction of three primary colors: red (R), green (G) and blue (B).
- While each pixel of a grayscale image has a single intensity value, each pixel of a color image is a vector having three values, corresponding to the R, G and B components.
- Thus, 2^{3N} different colors can be represented using *N* bits per color channel.



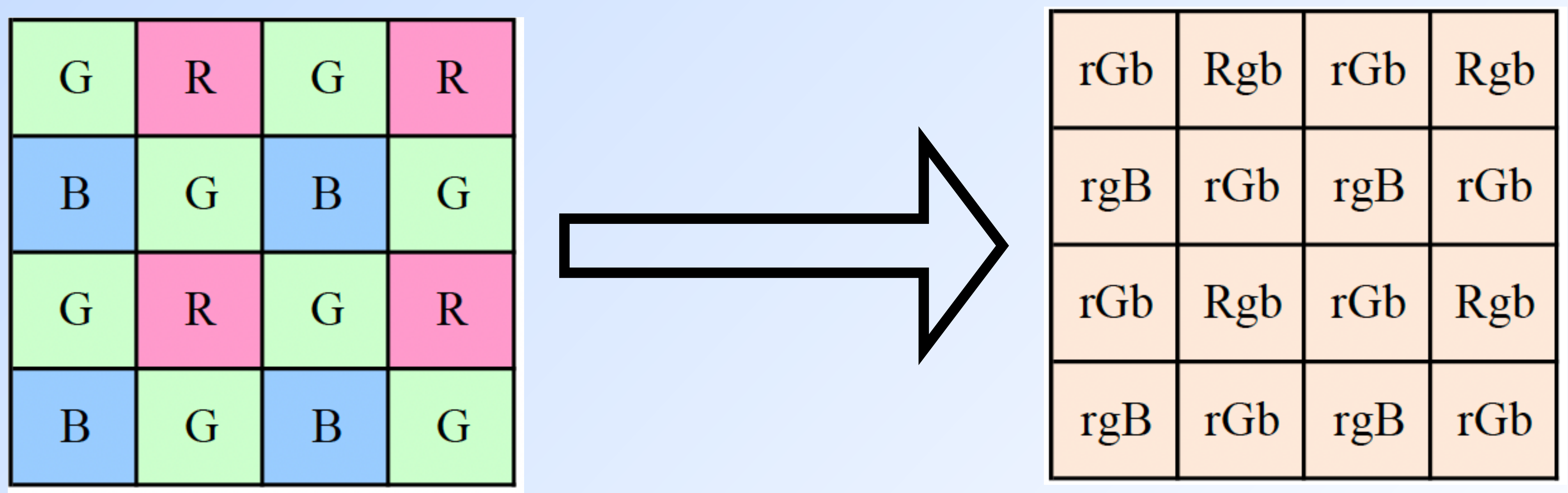
RGB Color Cube

Color Images

- Color images are acquired by masking sensor arrays with red, green and blue filters.
- The Bayer pattern is a commonly used masking pattern for color sensors, where half of the sensors are masked with green filters, and the remaining with red and blue filters.
- The RGB values for all the sensors are interpolated using the actual values recorded by the neighboring sensors.



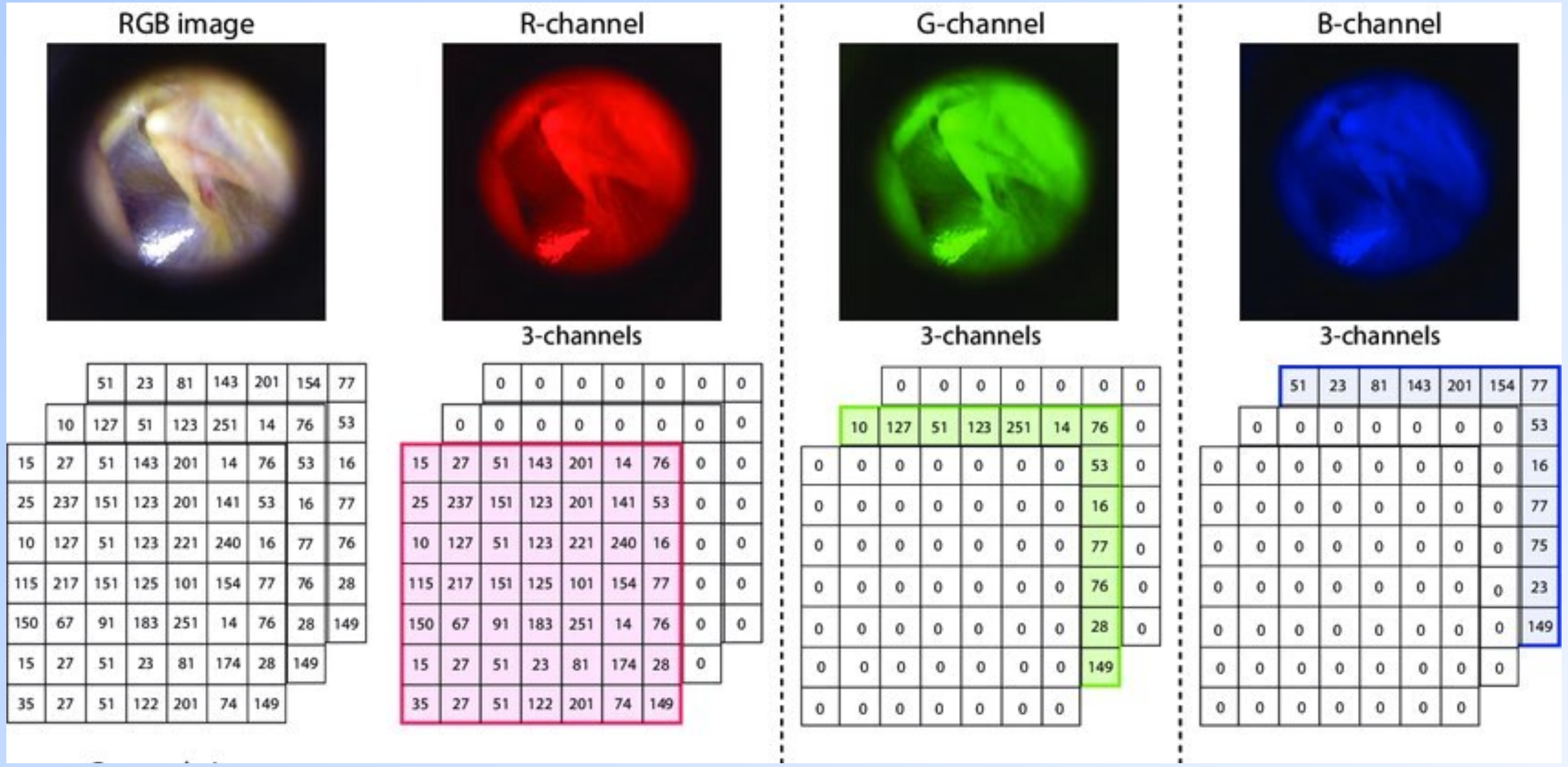
Bayer Pattern



Demosaicing

Digital Image Representation

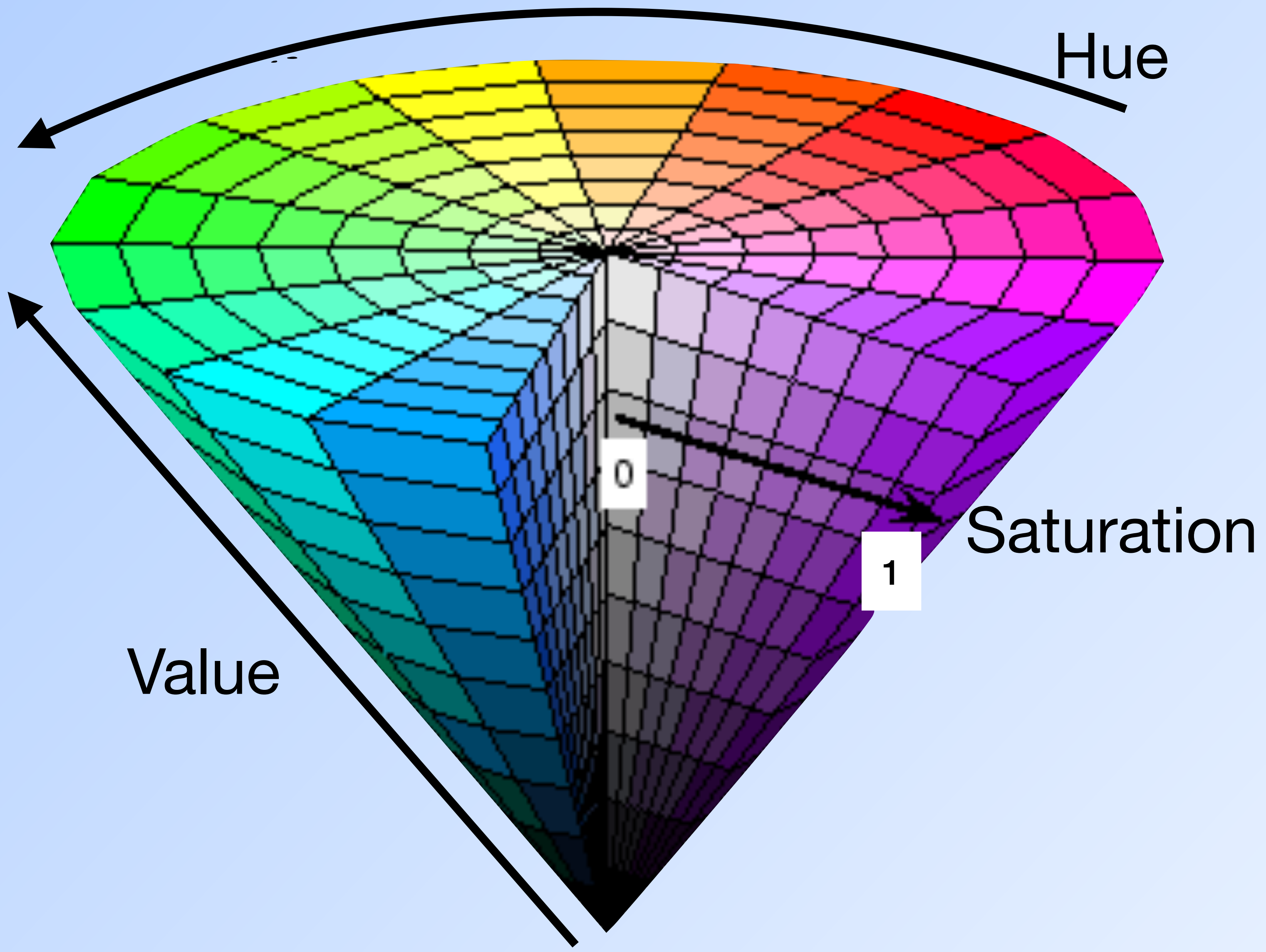
- The dimensions of a digital image is usually expressed in terms of the number of pixels in the horizontal and vertical directions
- A digital image of dimensions $W \times H$, and having N bits per pixel can be represented as a function $f(x, y)$ where:
  $0 \leq x \leq W - 1, 0 \leq y \leq H - 1$, and
  $0 \leq f(x, y) \leq 2^N - 1$

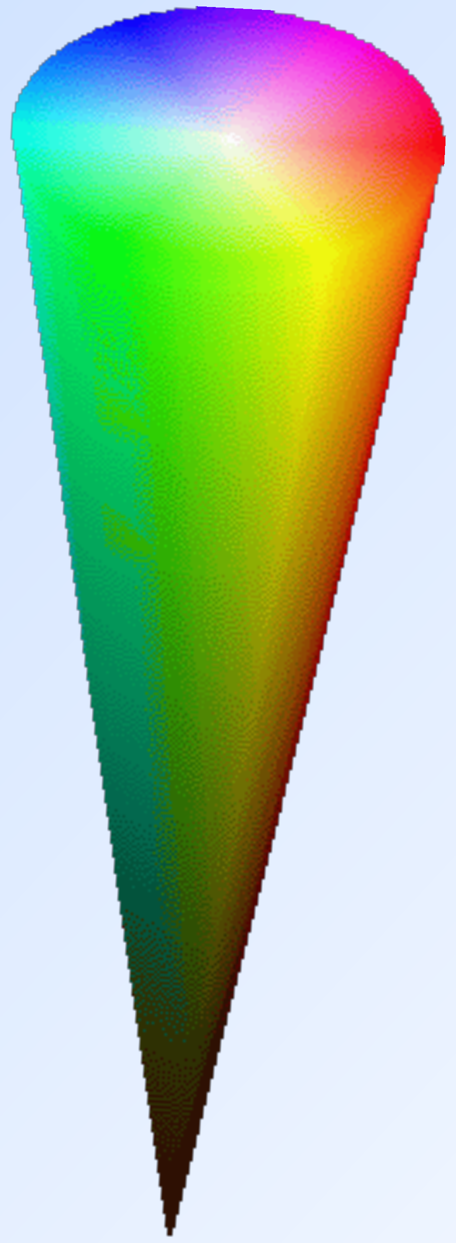


Color image representation

Useful Color Models - HSV

- The RGB color space is not perceptually intuitive.
- Alternative color spaces could be derived by applying non-linear transformations to the RGB color space.
- The HSV color space is one such color space that is better in explaining the human perception of color.
    - Hue (S): refers to the tint/shade of a specific color
    - Saturation (S): refers to the purity of the color
    - Value (V): refers to the brightness of the color.
- Similar color spaces: HSI/HSL.





HSV Color Cone

Useful Color Models - HSV

Normalize R,G,B values

$$R' = \frac{R}{255}, G' = \frac{G}{255}, B' = \frac{B}{255}$$

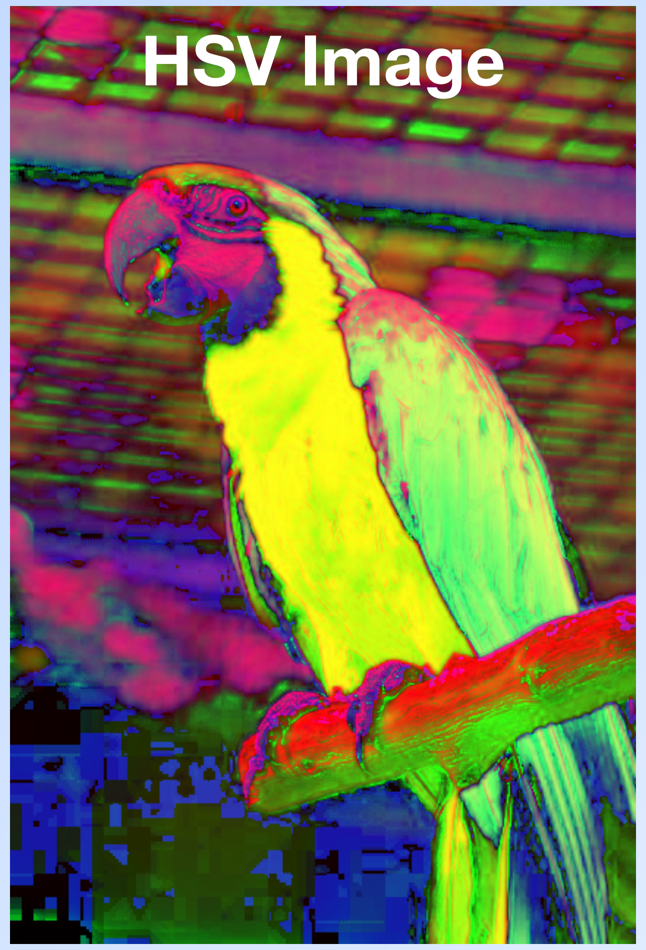
Calculate Value: $V = \max(R', G', B'),\ V \in [0,\ 1]$

Calculate Saturation: $S = \begin{cases} 0 & \text{if } V = 0 \\ \dfrac{V - \min(R', G', B')}{V} & \text{otherwise} \end{cases},\ S \in [0,\ 1]$

Calculate Hue: $H' = \begin{cases} \dfrac{(G' - B')}{(V - \min(R', G', B'))} & \text{if } V = R' \\ 2 + \dfrac{(B' - R')}{(V - \min(R', G', B'))} & \text{if } V = G' \\ 4 + \dfrac{(R' - G')}{(V - \min(R', G', B'))} & \text{if } V = B' \end{cases}$

$$H = H' \times 60^\circ$$

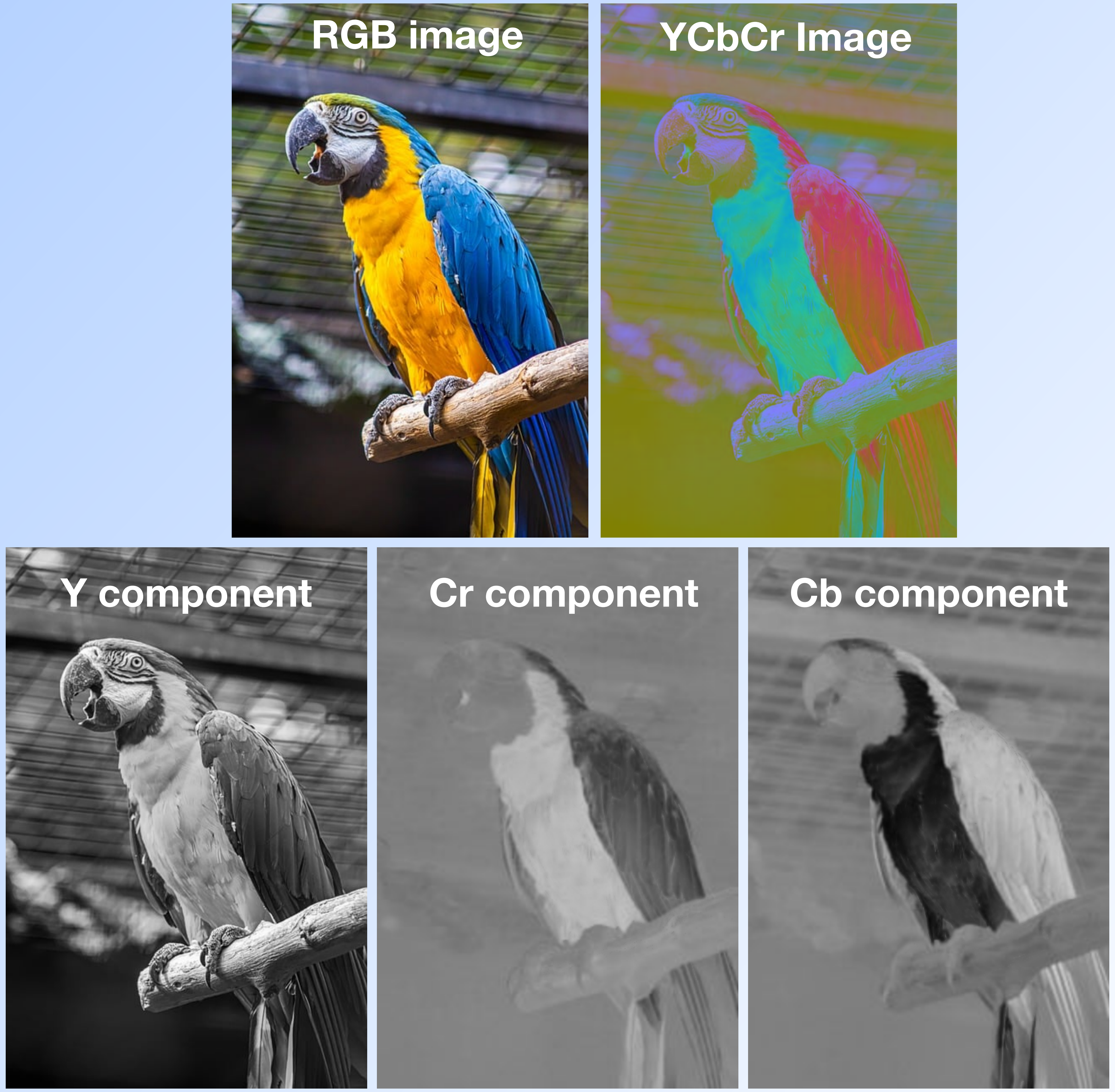
if $H < 0$ then $H = H + 360^\circ,\ H \in [0,\ 360^\circ)$



Useful Color Models - YCbCr

- The RGB color space is also highly correlated.
- Alternative color spaces such as YUV/YCbCr were developed to decouple the luminance information from the chrominance (color) information.
- In these color spaces, Y stands for the luma channel and U, V or Cb. Cr stand for the chroma channels.
- This separation enables representation of the luma component at higher resolutions as compared to chroma components to save bandwidth.
- The conversion of RGB to YCbCr can be performed as:

$$Y = 0.299 \cdot R + 0.587 \cdot G + 0.114 \cdot B$$
$$Cr = 0.713(R - Y) + 128$$
$$Cb = 0.564(B - Y) + 128$$



Common Image Formats

Images in the following formats are common in visual computing tasks:

Image Format	Extension	Brief Description
Graphics Interchange Format	.gif	Supports lossless compression, limited to 256 colors, supports animation
Portable Network Graphics	.png	Successor to GIF, supports lossless image compression, supports true color (16 million colors), does not support animation
Tagged Image File Format	.tiff/.tif	Supports both lossless and lossy image compression, supports 256 color as well as true color palettes, does not support animation
Bit Mapped Pixel	.bmp	Supports both lossless and lossy compression, various color depths. Somewhat specific to Windows OS.
Joint Photographic Experts Group	.jpg/.jpeg	The prevailing standard of lossy image compression.
WebP	.webp	Supports both lossless and lossy compression, transparency and animation. Offers improved compression performance over jpeg
Scalable Vector	.svg	Supports distortionless scaling, editing, and animation