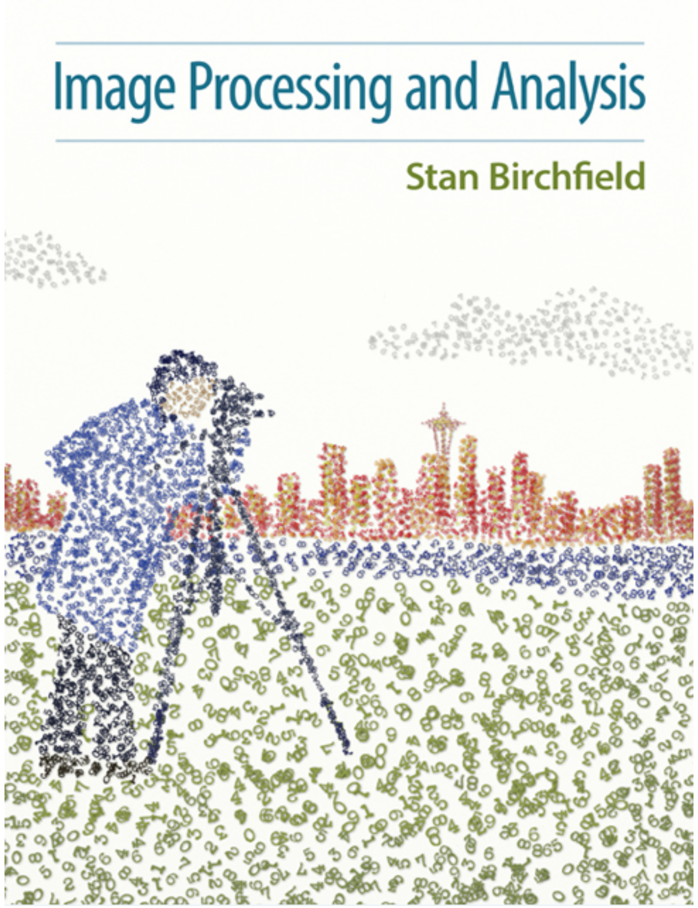
University of Stavanger

Prof. Kjersti Engan

ELE510 Image processing and computer vision

Fundamentals of Imaging, (Chap 2 Birchfield) 2020



Parts in course presentations is material from Cengage learning. It can be used for teaching, and it can be share it with students on access controlled web-sites (Canvas) for use in THIS course. **But it should not be copied or distributed further in any way** (by you or anybody).

© 2018 Cengage Learning®. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

Human visual system

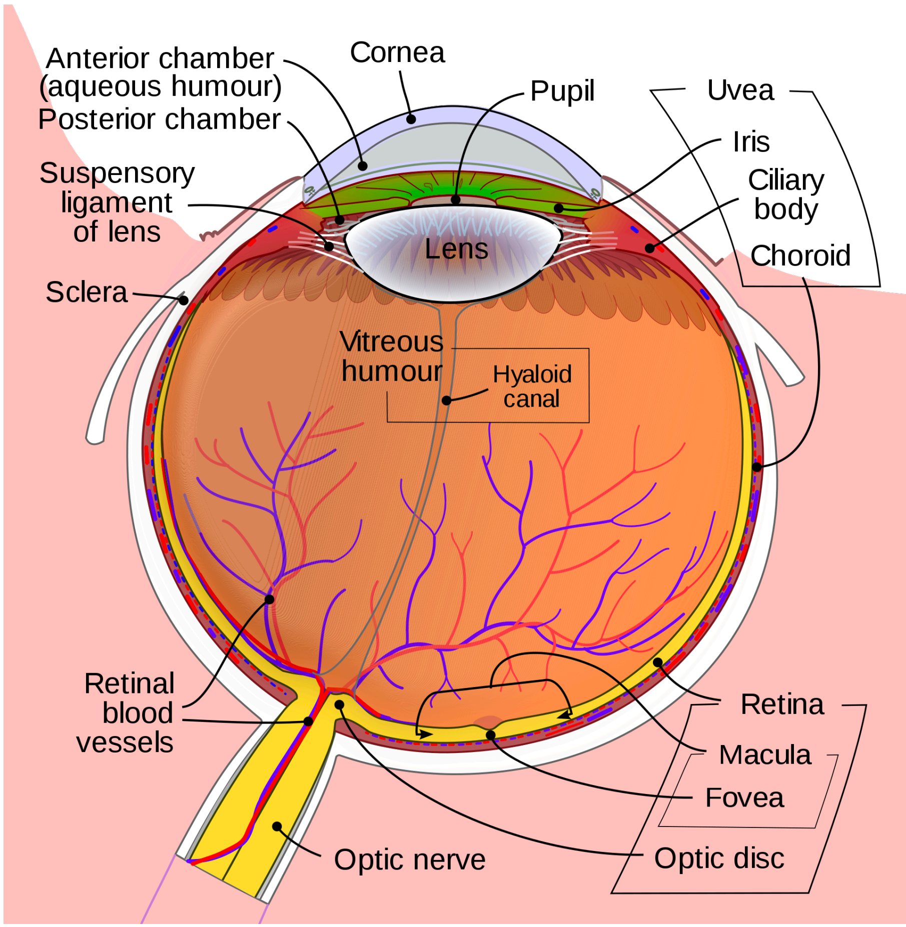
The optics of the eye create a 2D image of the visual world on the **retina** (similar to image sensor in camera), which translates that image into electrical neural impulses to the brain to create visual perception.

Photoreceptor: The most fundamental component of any natural vision system. There are two types of photoreceptors in the retina:

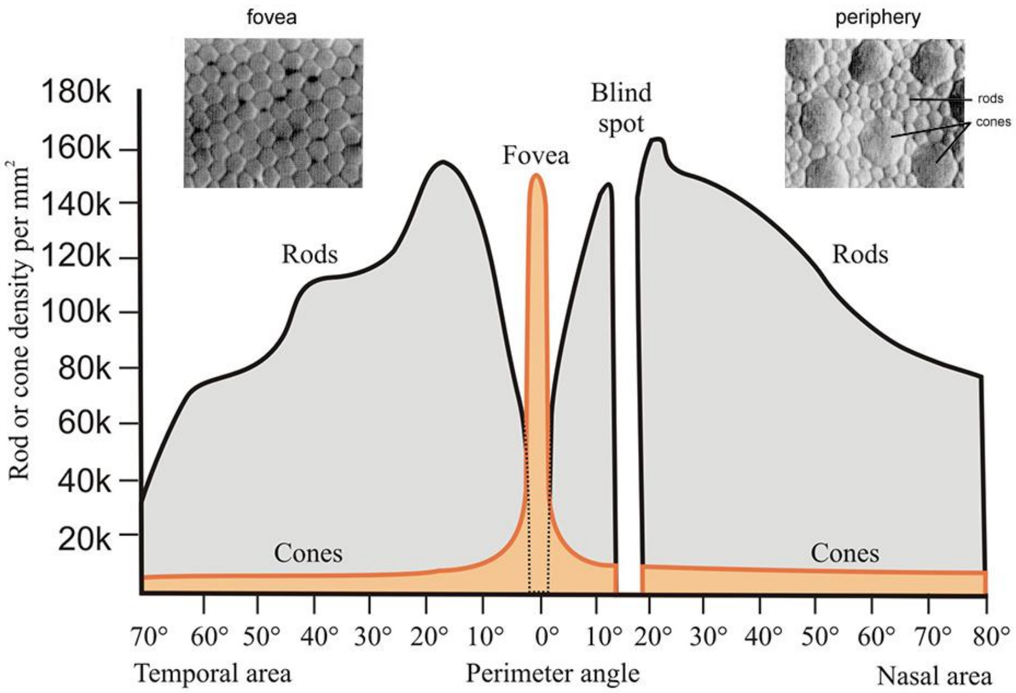
Rods: sensitive to low levels of light

Cones: normal everyday light.

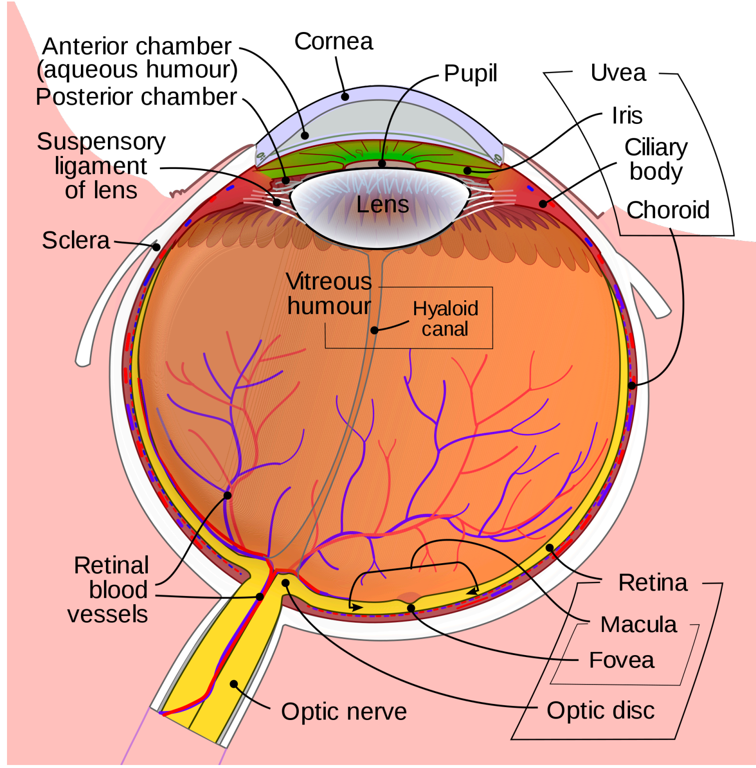
3 types of cones give color



By Rhcastilhos. And Jmarchn. - Schematic_diagram_of_the_human_eye_with_English_annotations.svg, CC BY-SA 3.0, https://commons.wikimedia.org/w/index.php?curid=1597930

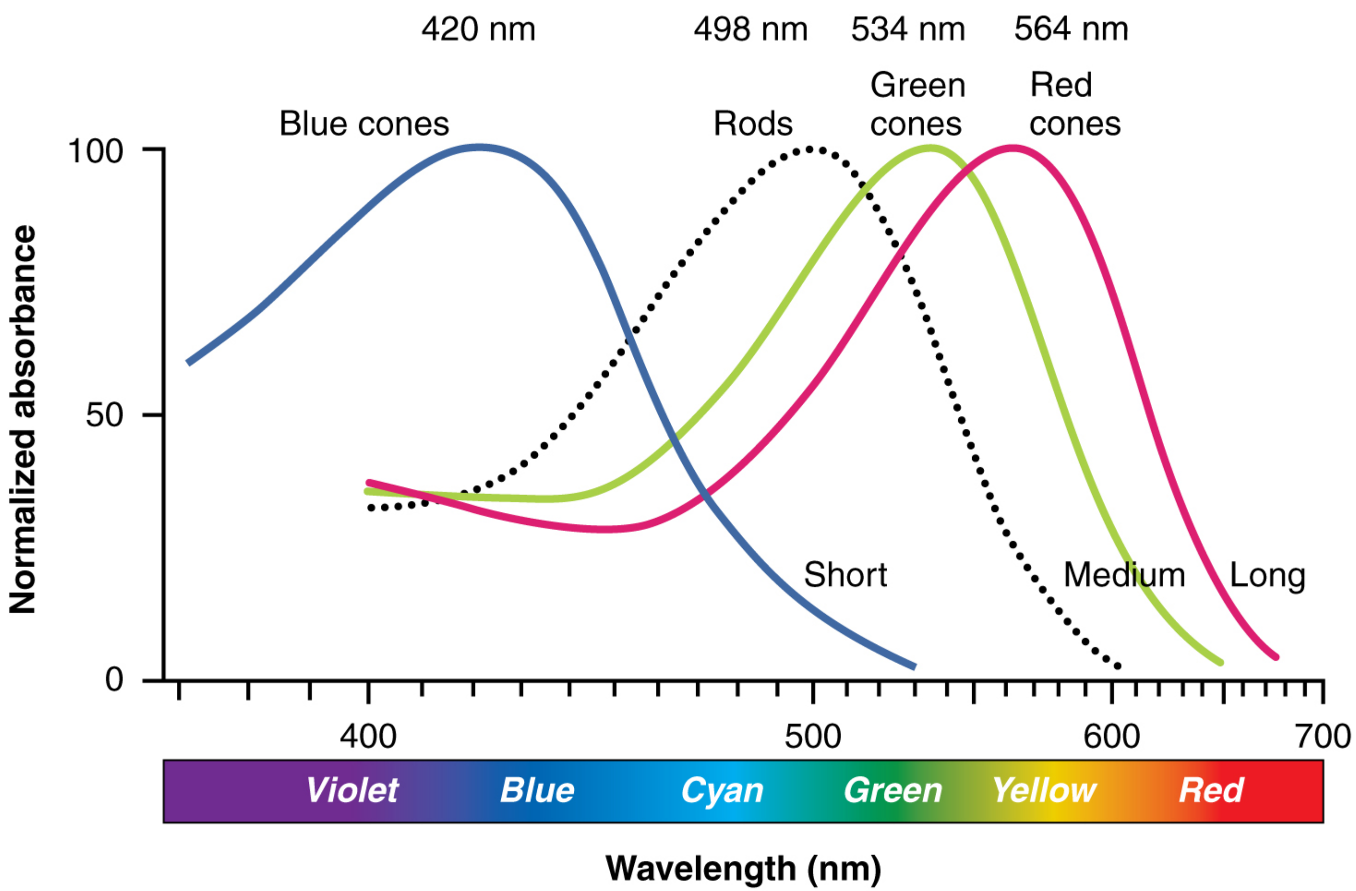


In fovea, the cones are tightly packed, as a regular sampling array, with cells spaced at approximately 2,5 μm. This is approximately the same spacing as between pixels in a typical camera sensor.



By Rhcastilhos. And Jmarchn. - Schematic_diagram_of_the_human_eye_with_English_annotations.svg, CC BY-SA 3.0, https://commons.wikimedia.org/w/index.php?curid=1597930

Spectral distribution for the human eye, rods and cones



Human Visual Perception

- The human visual system can respond to levels of light ranging 14 orders of magnitude.
- **Luminous efficiency function (LEF):** captures the relative sensitivity of the visual system to different wavelengths.
- **Photopic LEF:** corresponds to normal light levels where the cones dominate due to the saturation of the rods.
- **Scotopic LEF:** corresponds to low light levels where the rods dominate due to the lack of sensitivity of the cones.

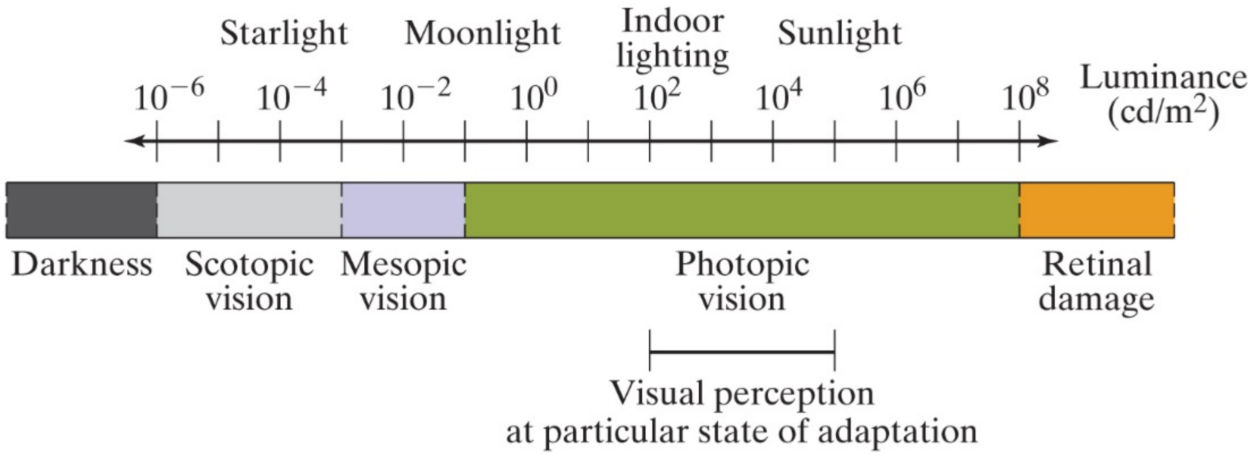
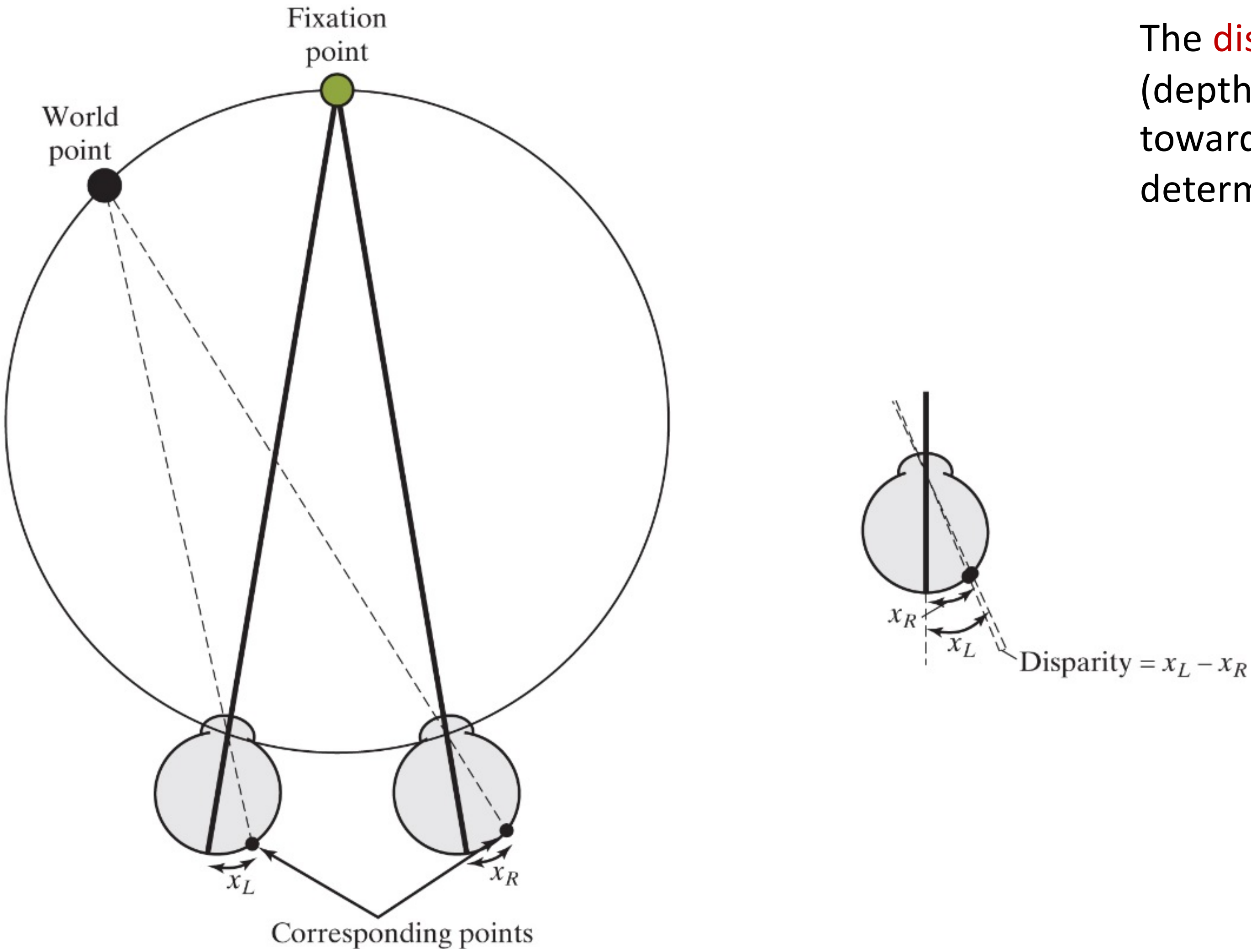


Figure 2.5 Retinal disparity is defined as the distance between corresponding points on the two retinas, after the retinas have been overlaid on top of one another and rotated so that their optical axes are coincident. Based on B. A. Wandell. *Foundations of Vision*. Sunderland, Mass., Sinauer Associates, Inc., 1995.

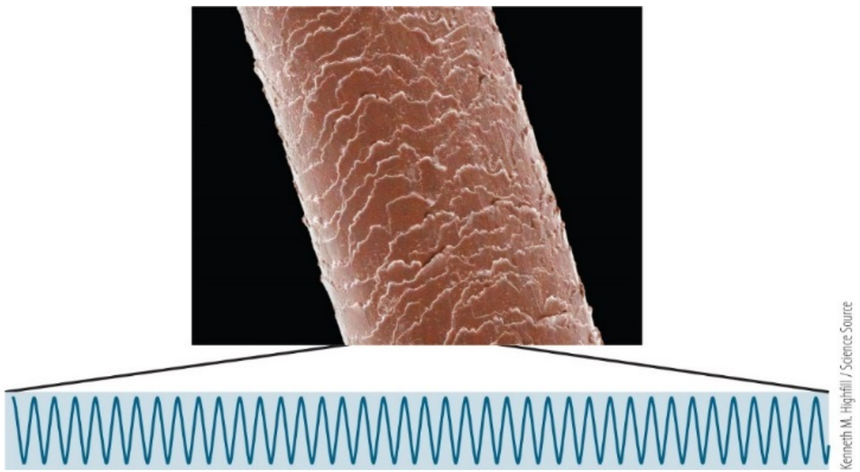


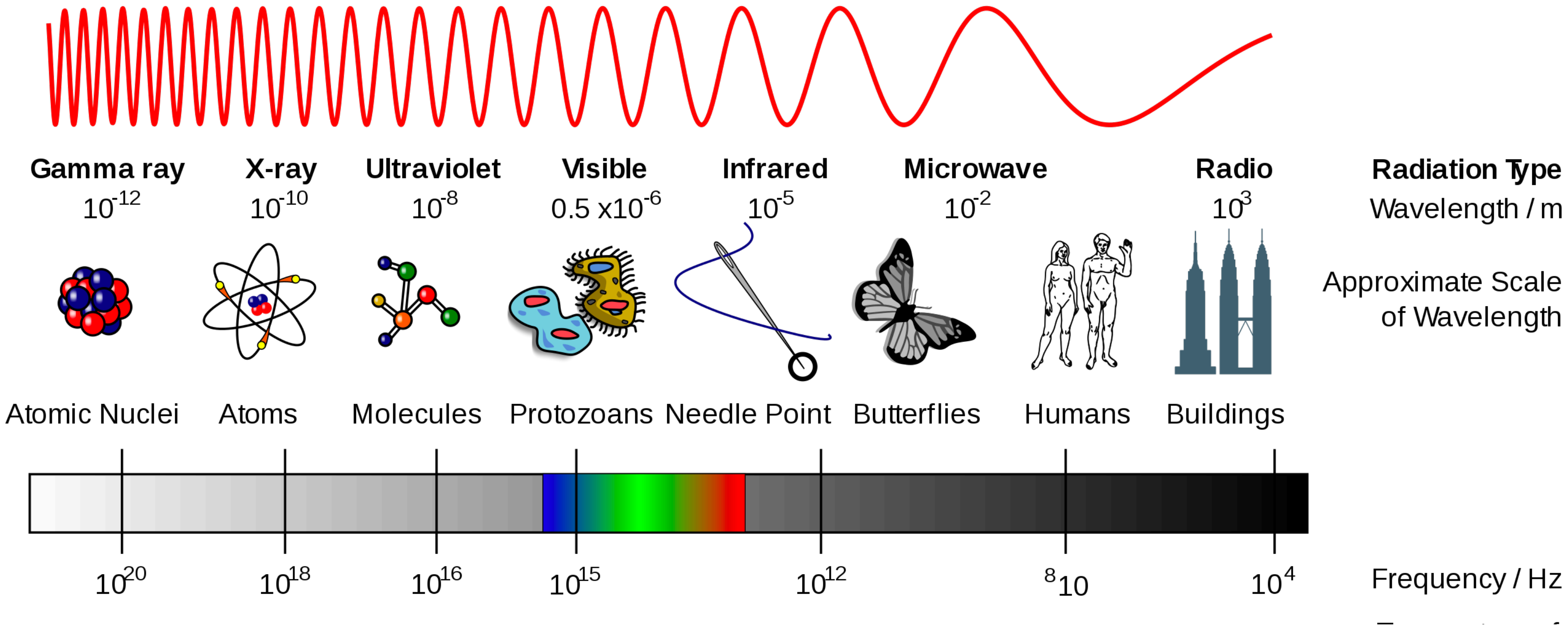
The disparity is related to the distance (depth) to the point and it is a key toward 3D perception and determining the distance to an object.

Light and the Electromagnetic spectrum

- Light is an electromagnetic wave traveling through space
- **Wavelength λ (measured in meters):** the distance between successive peaks in the sinusoid of a wave.
- **Frequency *ν* (measured in hertz):** inversely related to the wavelength.
  - **λ** times ***ν*** is the speed of light in the medium.

Figure 2.19 The wavelength of visible light is about 1/100th the diameter of a human hair.





By Inductiveload - http://commons.wikimedia.org/wiki/File:EM_Spectrum_Properties.svg, CC BY-SA 3.0,
https://commons.wikimedia.org/w/index.php?curid=24646522

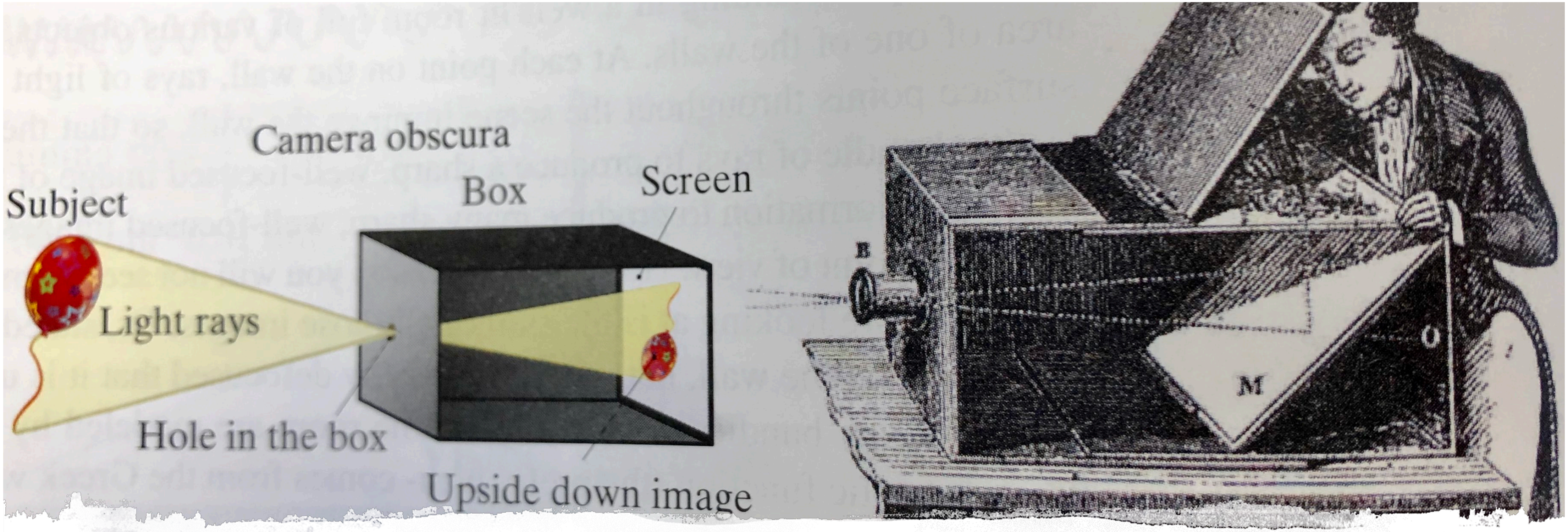


Fig 2.22 Birchfield, «Image processing and Analysis»

Pin-hole camera

- The many bundles of light rays in a room are modeled by the plenoptic function (complete), that is all light rays passing all locations (x,y,z) in all directions (θ, φ) , 5D function.
- Sampling the plenoptic function at one point (x,y,z) gives rise to the pinhole-camera. Allow the light from all angles to pass through one point (the pin-hole)

Pinhole Camera

- **Focal point:** the pinhole through which all rays of light pass.
- **Image plane:** the sensor surface on which the image is formed.
- **Optical axis:** the line through the focal point perpendicular to the image plane.
- **Focal length:** the distance from the focal point to the image plane along this line.
- **Perspective projection:** Light rays from the source reflect off the surface of an object in the scene, travel through the focal point (also called the center of projection), then land on the image plane.

- (x_w, y_w, z_w): the 3D coordinates of the world point.
- (x, y): the 2D coordinates of its projection onto the image.

$$y = f\frac{y_w}{z_w}$$

$$x = f\frac{x_w}{z_w}$$

Perspective projection, pinhole camera

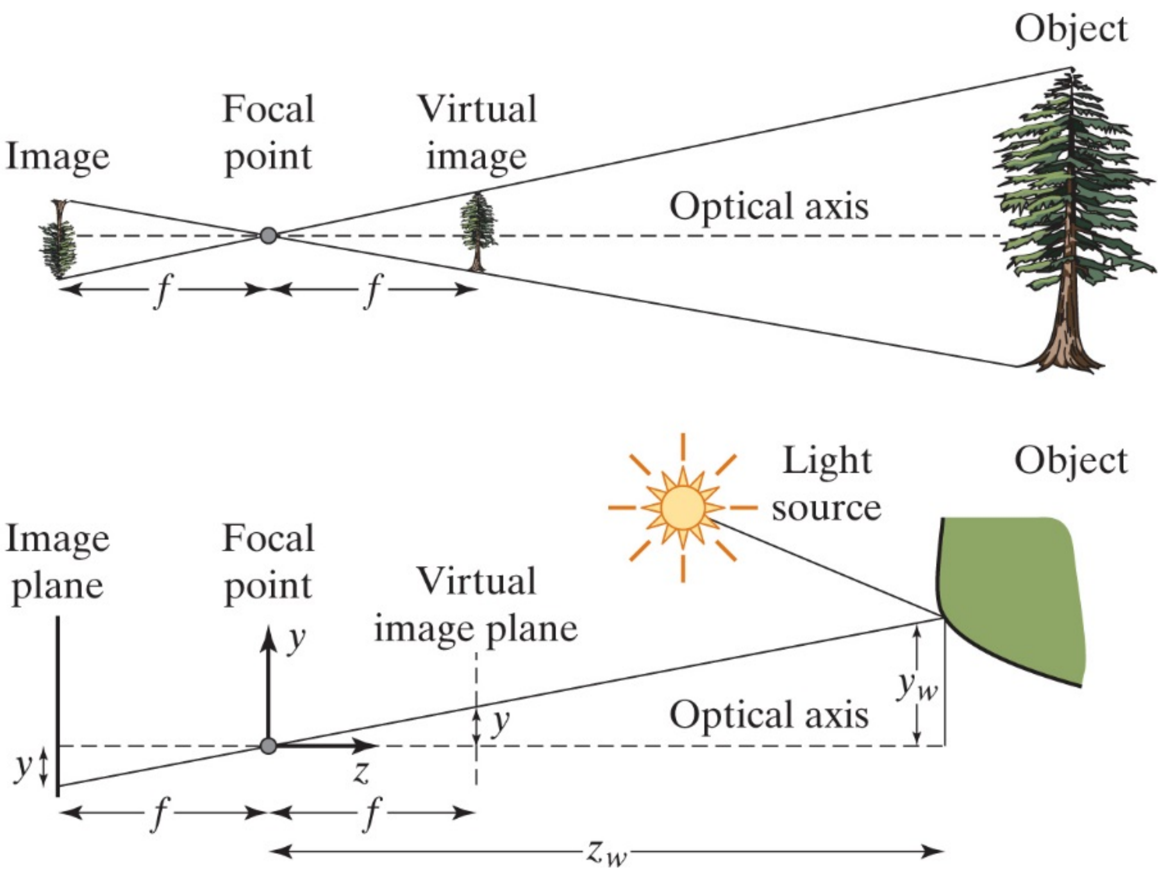


Fig 2.23 Birchfield, «Image processing and Analysis»

Focal point = pinhole
Focal length: distance between focal point and image plane along optical axis (here: f) .

3D->2D
Any point on the light ray will project to same point on the image plane, thus it is impossible to recover distance from the camera without additional information.

The upside-down image on the image plane is mathematically equivalent to produce a rightside-up image on the virtual image plane in front of the focal point.

$$x = f\frac{x_w}{z_w} \qquad y = f\frac{y_w}{z_w}$$

When all zw are close in values, i.e. Approximately a constant, we get a linear model:

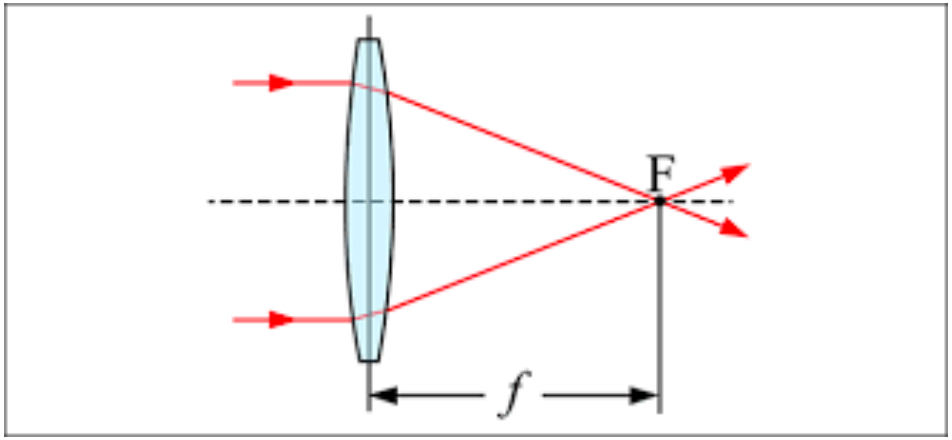
$$x = f\frac{x_w}{z_w} \approx \alpha x_w \qquad y = f\frac{y_w}{z_w} \approx \alpha y_w$$

This is called scaled orthographic projection or weak perspective projection, and is realistic/a good model if the scene lies close to the optical axis or the depth varies litle over the scene.

The ideal pinhole camera is not practical because the pinhole does not let in much light. A practical imaging device replace the pinhole with a lens. The pinhole camera is still important because even real cameras with lenses are well modeled mathematically as pinhole cameras with distortion due to the lens.

Camera with lens

- To build a more practical imaging device, the pinhole is replaced with a **lens**.
  - A simplified lens consists of two spherical surfaces joined so that the centers of the spheres are collinear with the centers of the surfaces.
- **Gaussian optics:** a common assumption in which all light rays are **paraxial** (they form small angles with respect to the optical axis)
- **Lens maker's formula:** also known as the **thin lens formula**.



$$\frac{1}{f} = (n-1)\left(\frac{1}{r_1} - \frac{1}{r_2}\right)$$

f is focal length
r1 and r2 is radii of the spherical surfaces

Camera with lens

- Aperture of a camera is the opening through wich light rays enter the lens on the way to the sensor. The focal length, *f*, relative to the diameter, *d*, of the aperature is called f-number (*f/d*).
- The sensors size (diameter of the aperture) and focal length determine det angular field-of-view (FOV), given by 2θ, where $\tan\theta = {}^{d}/_{2f}$.

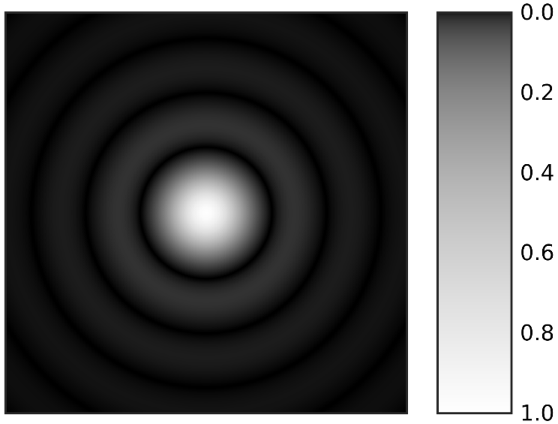
Camera with lens

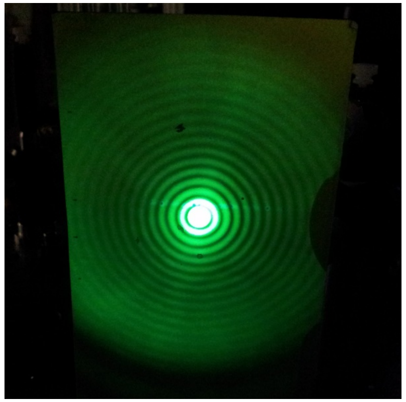
- **Depth of field (DOF):** the range of distances in the scene that form acceptably sharp images.

- **Point spread function (PSF):** specifies the shape that a point will take on the image plane. Also called impulse response (will come back to that later for general transferfunction).

- **Optical transfer function (OTF):** The Fourier transform of the PSF whose magnitude is the **modulation transfer function (MTF)**.

Imaging a point source, the point spread function





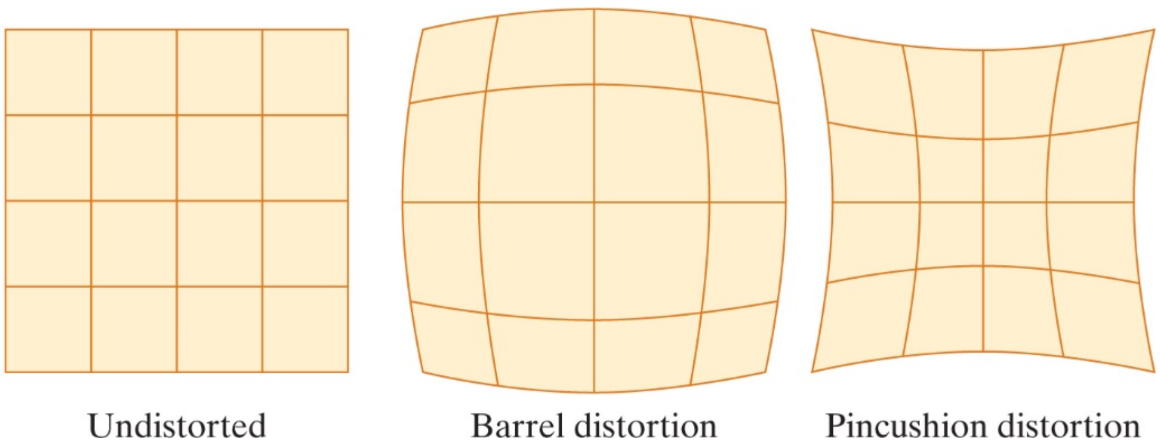
Airy disk:
Created by a laser beam through a pinhole.

Airy disk:
Best focused spot of light that a perfect lens with a circular aperture can make, is limited by the diffraction of light.

Distortion

- **Aberration:** any deviation of the performance of a lens from ideal.
- **Distortion:** arises from the fact that the light rays do not necessarily follow straight lines when passing through the lens.
- Most noticeable for lenses with small focal lengths (fisheye lenses)

Figure 2.26 An undistorted image, barrel distortion, and pincushion distortion.



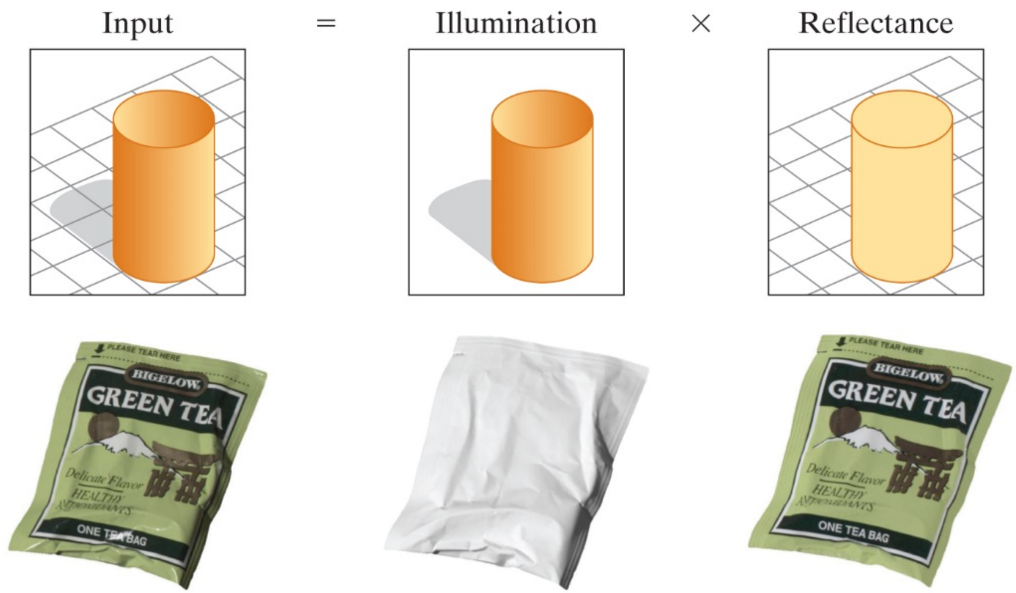
Simplified Imaging Model

$$\boxed{E = L\frac{\pi}{4}\left(\frac{d}{f}\right)^2 cos^4\,\Theta}$$

- **Irradiance E** [N/m^2] is the strength of electromagnetic radiation on a surface. For a sensor, proportional to the radiance, L.
- The **irradiance** *E* on the image sensor is the product of a lighting function Λ and a reflectance function *R*.

$$E(x, y, \lambda) = \Lambda(x, y, \lambda)R(x, y, \lambda)$$

- In **intrinsic images**, multiple images of a static scene under different imaging conditions can be used to estimate the reflectance or other properties in the scene.



© 2009 IEEE. Reprinted, with permission, from Roger Grosse, Micah K. Johnson, Edward H. Adelson, William T. Freeman, "Ground truth dataset and baseline evaluations for intrinsic image algorithms," 2009 IEEE 12th International Conference on Computer Vision, pp. 2335–2342.

Image Acquisition

- The image pixel value $I(x, y)$ can be modeled as the integration of the irradiance function over the area of the pixel and over all wavelengths, after first multiplying by the sensitivity (the sensors sensitivity to a particular wavelength) function:

$$I(x, y) = \varphi\left(\int\int\int E(x', y', \lambda')s(\lambda')\, dx'\, dy'\, d\lambda'\right)$$

- **Sampling:** converts the continuous irradiance function into a discrete function defined only over the rectangular lattice of integer (x, y) coordinates.

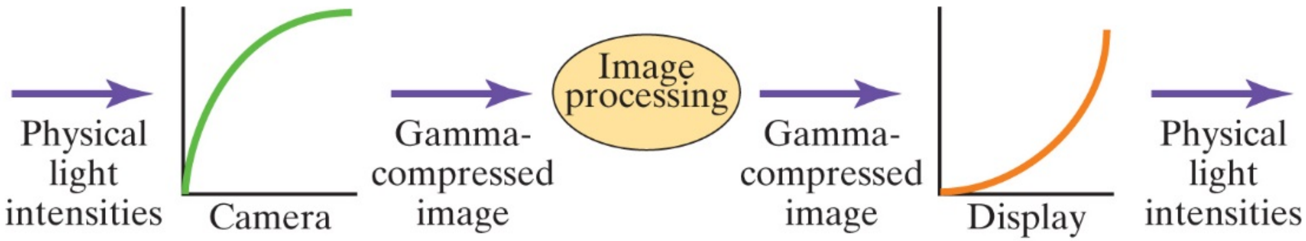
- **Quantization:** assigns a discrete gray level to every pixel in order to represent its value in digital form.

Gamma compression

- **Gamma Compression:** transforms the linear, physical light intensity into a perceptually uniform quantity, so that the pixel values in a digital image are not directly proportional to the amount of light collected by the sensor.

$$L = cV^{\gamma} + b$$

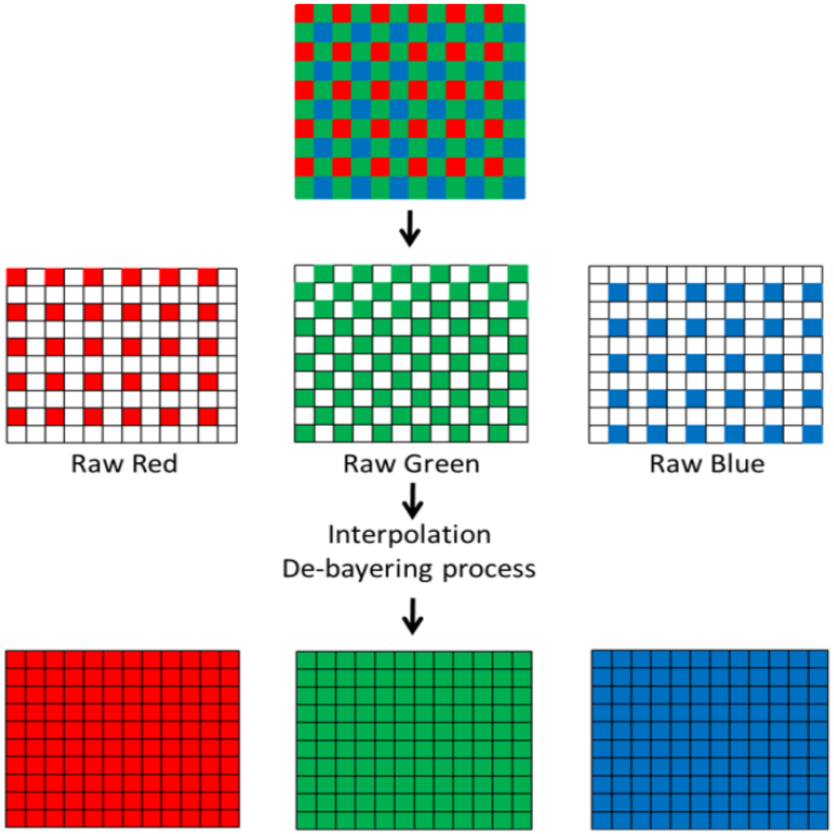
Figure 2.29 Linear light intensities are gamma compressed by the camera into perceptually uniform quantities, which are then gamma expanded by the display.



Camera sensors

- Two most common digital sensors, both with a dense array of photodiodes (typically spaced 1 to 5 µm apart) converting light photons to electrons (remember: cones in the fovea, approximately 2,5 µm)
  - CCD (charge-coupled device), older technology
  - CMOS (complementary metal-oxide semiconductor), have more possibilities, individual pixels can be read, less expensive, less power ..
- Color
  - 3CCD.   Trichoric prism split light in three separate beams, sensed by three seperate sensors. Used in professional and high end, expensive
  - Color Filter Array / mosaic (CFA/CFM) is less expensive. Much used Bayer filter, each pixel sense one color.

Bayer Color Filter

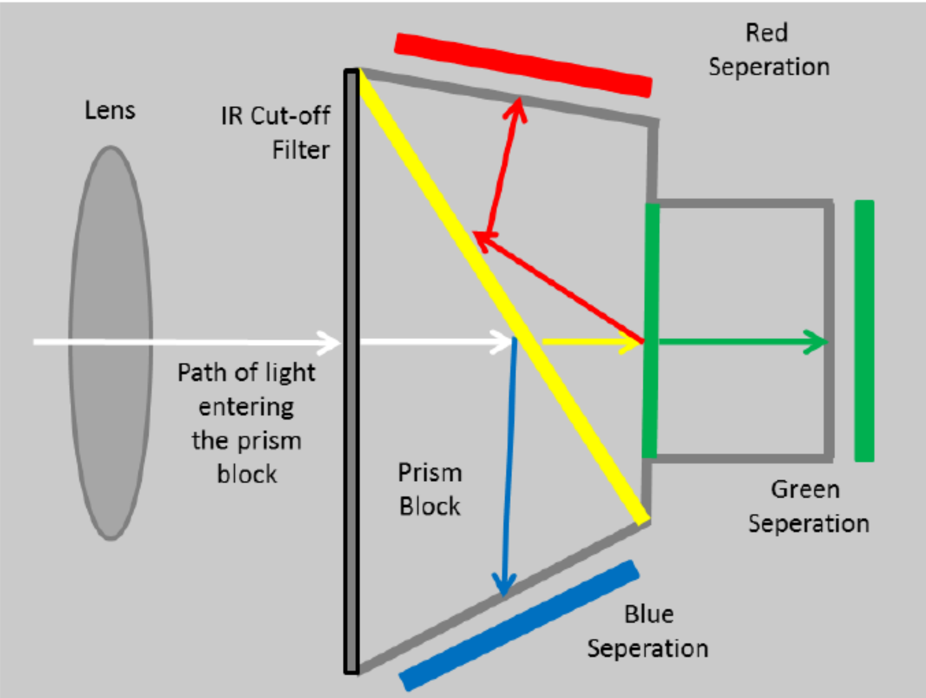


In this solution the raw primary color images are not dense, solution in-between pixels are missing.

Full color information is only possible by using *debayering algorithms* which interpolate the missing information from neighboring pixels and provide an estimation.
This involves image processing that is implemented on the same chip as the image sensor.
This process can lead to false representation of image quality.

From: White Paper, JAI, “How does prism technology help to achieve superior color image quality”.

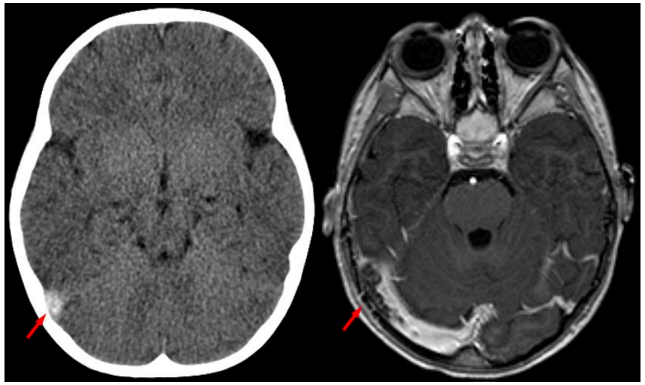
CMOS Machine Vision Camera Pixel architecture - Prism technology



- Inside a prism-based camera, the photons entering the optical system propagate through the prism before interacting with the sensors.
- The short wavelength, i.e., blue is separated first, then red while the green passes through the prism block. This ensures minimum loss of signal strength

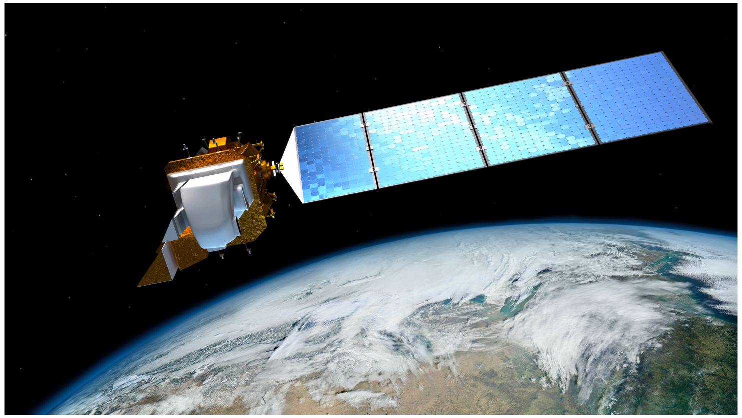
From: White Paper, JAI, "How does prism technology help to achieve superior color image quality".

Other Imaging Modalities



- **X-ray radiography:** produces images by transmission rather than reflection. A generator emits X-rays toward an object of interest, and a detector measures the photons that make it to the other side (not absorbed).
- **Computed tomography (CT) (3D structure, X-ray):** refers to imaging by slices. The term *axial* refers to the horizontal plane through the human body when standing upright, and since the slices are parallel to this plane, the approach is also known as **computed axial tomography (CAT).**
- **Magnetic resonance imaging (MRI):** safer than X-ray because it does not use ionizing radiation.
  - MRI uses powerful magnets to align the hydrogen nuclei (that is, protons) in the water molecules.
  - A radio frequency (RF) signal pulse at the resonance frequency is emitted that systematically alters the alignment of the nuclei by flipping the spin of the protons.

Remote Sensing



- Cameras are often attached to aircraft or satellites for remote sensing of the earth for applications in meteorology, agriculture, surveillance, and geology.
- These cameras typically sense multiple frequencies simultaneously.
- A multispectral sensor senses a small number of frequencies, typically 5 to 7. A hyperspectral sensor senses a much larger range of frequencies.
- LANDSAT – gathered satellite imagery of the earths surface. Now use 8 spectral bands (earlier 7) typ. RGB + infrared bands.
- SAR Synthetic aperture radar. Usually captured from aircraft or spacecraft.