Computer Vision

James L. Crowley

M2R MoSIG

Fall Semester
3 oct 2019

Lesson 1

Visual Perception in Man and Machine

Lesson Outline:

1 The Physics of Light

1.1 Photons and the Electro-Magnetic Spectrum

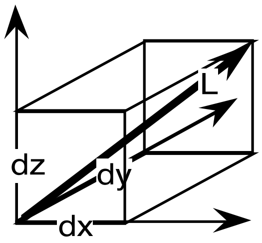
Electromagnetic energy is propagated as a resonant oscillation of electric and magnetic potential (Photons). This oscillation is described by Maxwell's equations. The magnetic field is strength determined the rate of change of the electric field, and the electric field strength is determined by the rate of change of the magnetic field.

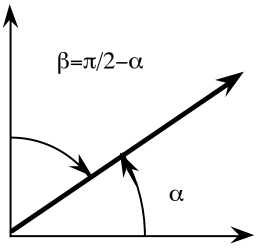
The Frequency of the oscillation determines the wavelength of the photon.

The photon is characterized by
1) a direction of propagation , $\vec{D}$,
2) a direction of oscillation (a polarity), and
3) a wavelength, λ, and its dual, a frequency, f: $\lambda = \frac{1}{f}$

The direction of propagation and direction of oscillation can be represented as vectors of cosine angles. For example the direction of propagation can be represented as D.

$$\vec{D} = \begin{pmatrix} \cos(\alpha) \\ \cos(\beta) \\ \cos(\gamma) \end{pmatrix} = \begin{pmatrix} \Delta x / L \\ \Delta y / L \\ \Delta z / L \end{pmatrix}$$





Photons are created and absorbed by changes in the orbits of electrons. Absorption and creation of photons are probabilistic (non-deterministic) events.

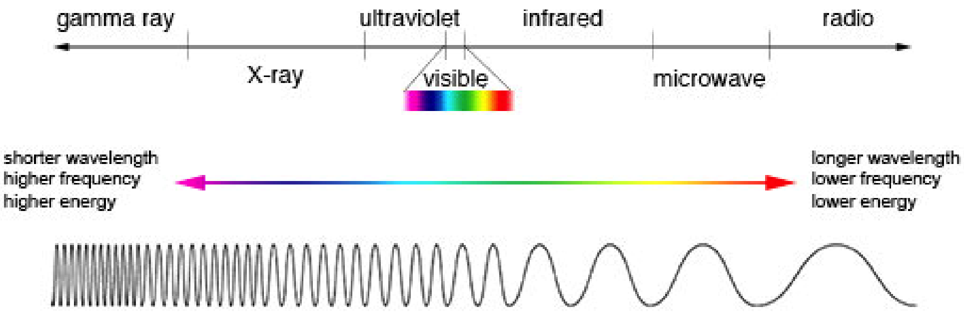
Photons sources generally emit photons over a continuum of directions (a beam) and continuum of wavelengths (a spectrum). The beam intensity is measured in watts, (Joules/sec) and is equivalent to Photons/Meter^2.

Radiant flux a measure of the total "amount" of light in the visible spectrum emitted by a source. The unit of radiant flux can be thought of as a measure of the number of photons of visible light within a beam or angle, or emitted from some source.

Luminous flux (Lumens) is radiant flux adjusted for the wavelengths and sensitivity of light sensed by the human retina. 1W = 683 lumens.

Luminous flux measures the number of photons perceivable by a human.

The spectrum gives the probability of a photon having a particularly wavelength, S(λ). Visible light is only a very small part of the electromagnetic spectrum.



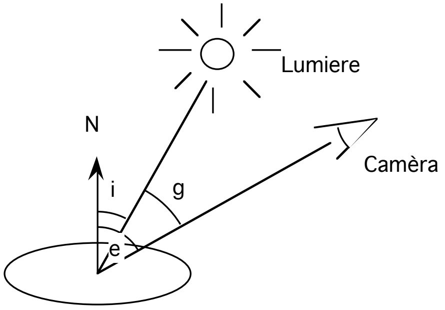
Radio Waves > 1 m
(note : 1 Hz ≈ 300,000 km, 100 Mhz ≈ 3 m, 1 Mhz ≈ 3 cm)
Micro waves: 0.1 cm to 1 m (10^{-3} m to 1 m)
Infrared: 7 x 10^{-7} to 10^{-3}
Visible: 400 nm to 700 nm (4 x 10^{-7} m to 7 x 10^{-7} m)
Ultraviolet: 10 nm to 400 nm (4x10^{-7} m to 10^{-8} m)
X rays: 0.01 nm to 10 nm (10^{-8} m to 10^{-11}m)
Gamma Rays: ≤ 0.01 nm (≤ 10^{-11} m)

1.2 Albedo and Reflectance Functions

The albedo of a surface is the ratio of photons emitted over the photons received. (Emitted radiant flux over incident (received) radiant flux)

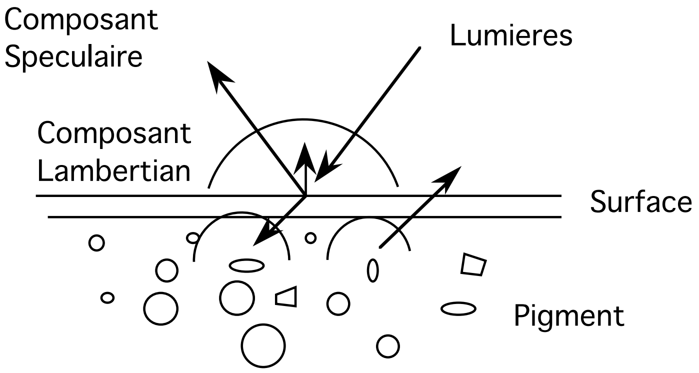
Albedo is described by a Reflectance Function, R(-).

$$\text{Albedo:} \quad R(i, e, g, \lambda) = \frac{\text{Number of photons emitted}}{\text{Number of photons received}}$$



The parameters are
- i: The incident angle (between the photon source and the normal of the surface).
- e: The emittance angle (between the camera and the normal of the surface)
- g: The angle between the Camera and the Source.
- λ: The wavelength

For most materials, when photons arrive at a surface, some fraction of the photons are rejected by an interface layer (determined by the wavelength). The remaining photons penetrate the material and are absorbed by molecules near the surface (pigments).



Most reflectance functions can be modeled as a weighted sum of two components: a specular component, $R_S()$ and a Lambertian component, $R_L()$.

$$R(i, e, g, \lambda) = a \cdot R_S(i, e, g, \lambda) + (1\text{-}a) \cdot R_L(i, \lambda)$$

Specular Reflection

The Specular reflection function is given by

$$R_S(i, e, g, \lambda) = \begin{cases} 1 & \text{if } i=e \text{ and } i+e=g \\ 0 & \textit{otherwise} \end{cases}$$

A mirror is an example of a specular reflector. All (or almost all) of the photons arriving at a mirror are reflected at the surface with no change in spectrum.

Lambertian Reflection

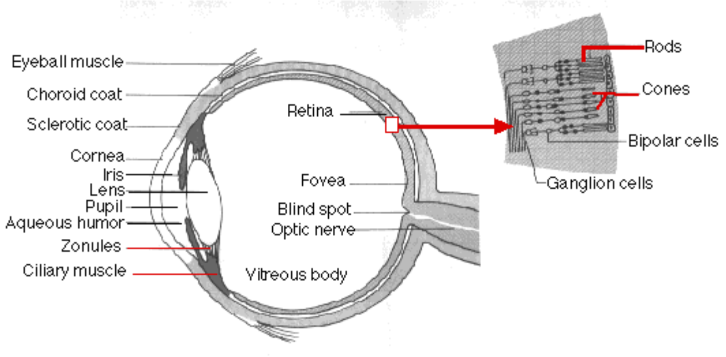
A Lambertian reflection function describes the photons that penetrate and escape from the material.

$$R_L(i, \lambda) = P(\lambda) \cdot \cos(i)$$

White paper and fresh snow are examples of Lambertian reflectors. Pigments in the material can change the color of the incident light by emitting photons at a frequency that is different from the incident photons that enter the material.

2 The Human Visual System

2.1 The Human Eye

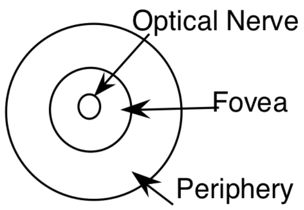


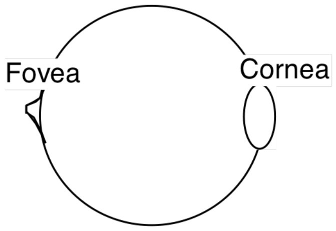
The human eye is a spherical globe filled with transparent liquid.
An opening (iris) allows light to enter and be focused by a lens.
Light arrives at the back of the eye on the retina.

2.2 The Retina

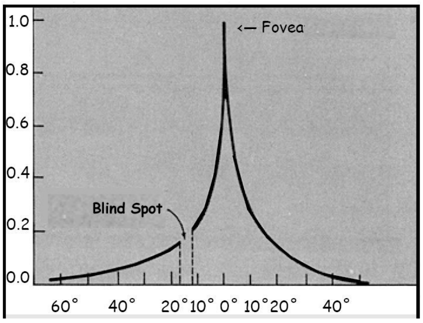
The human retina is a tissue composed of rods, cones and bi-polar cells.
Cones are responsible for daytime vision.
Rods provide night vision.
Bi-polar cells perform initial image processing in the retina.

Fovea and Peripheral regions





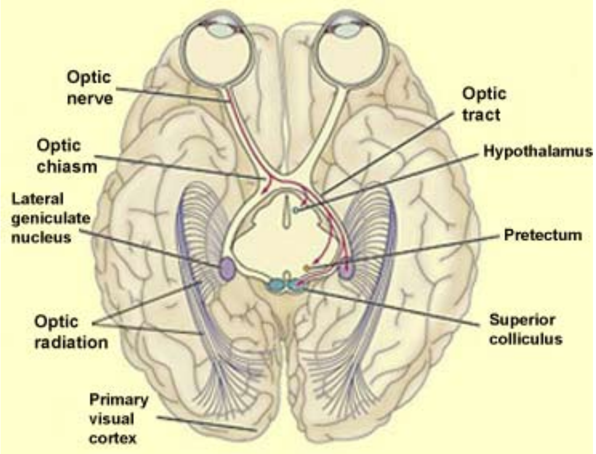
Cones are the primary visual receptors. A human eye contains approximately 6 million cones distributed exponentially over the retina. The density of cones increases exponentially to a central point referred to as the fovea. The retina contains a "hole" near the fovea where the optic nerve leaves the retina.



The fovea provides high visual acuity and is used for recognition and depth perception.

The peripheral regions have a much lower density of cones, and are used to provide shape context and to direct eye movements.

The fovea perceives only a small part of the world at any instant in time. However, the eye muscles can rotate the eyes at very high accelerations. This allows the eyes to rapidly scan to multiple parts of a scene. This motion is called a saccade.
The visual system stops processing during such saccades.



Primary organs of the human visual system.

The rods and cones feed signals to the optical nerve. The optical nerves leave the left and right retina via holes near the fovea to be joined at the Lateral Geniculate Nucleus (LGN) and the Superior Colliculus (SC).

The SC acts as both a relay station to communicate retinal maps to multiple parts of the visual cortex, and as an attention filter, to suppress unattended information. The LGN provides filtered "retinal maps" to the different visual cortexes as well as to the SC.

The LGN acts as a filter for **visual attention**, surpressing information that the system is not “attending” to (looking at). Surprisingly, 80% of the excitation of the LGN comes from the visual cortext and other areas of the brain.

In fact, the entire visual system can be seen as succession of filters.

2.3 The visual system is composed of filters

The entire visual system can be seen as succession of filters. The first visual filter is the horopter. This is the small region of 3D space that projects to the same position in both the left and right retinas. This is followed by the iris, that restricts the direction and focus of the photons entering the eye. The rods and cones of the retina filter photons based on wavelength. The neurons of the retina are filter visual stimuli with an initial point spread function.

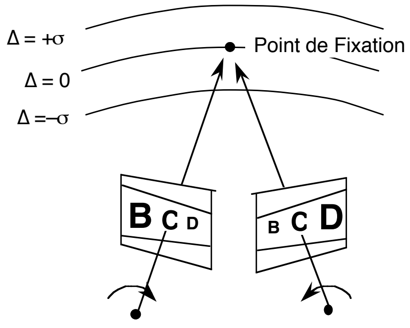
2.4 The Superior Colliculus and the Horopter.

Visual stimuli enter the optical nerves where they are communicated to the LGN and the Superior Colliculus (SC).

The Superior Colliculus control fixation. At any instant, the human visual system focuses processing on a small region of 3D space called the Horopter. The horopter is mathematically defined as the region of space that projects to the same retinal coordinates in both eyes.

The horopter is the locus of visual fixation in the world.

The horopter is controlled by the Superior Colliculus (SC), and can be moved about the scene with rapid movements (saccades). The LGN suppresses visual stimuli during a saccade, so that the movement is not perceptible.



The Superior Colliculus is a Feed-Forward (predictive) control system for binocular fixation, composed of 7 layers of filters that receive stimulus from the frontal cortex, the lateral and dorsal cortexes, the auditory cortex the peripheral region of the retina and other regions of the brain. The layers operate in the activation space of the eye-muscles. The output is a single activation peak that directly drives the movement of the muscles of the two eyes.

2.5 Vergence and Version

The output of the Superior Colliculus is a neural map that directly activates the muscles that rotate the eyes. The direction of gaze can be roughly modeled in a polar coordinate system with angle determined by the sum of individual gaze directions, and depth determined by the difference of gaze directions of the two eyes. Depth and direction are known as "Version" and "Vergence".

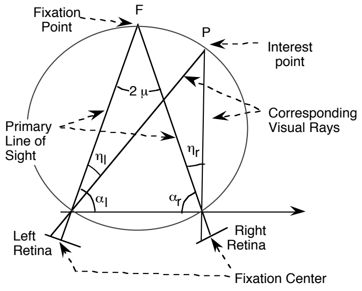
Vergence, or the difference of the individual gaze directions determine the 3D point where the gaze angles intersect thus determines the relative depth of the horopter. Vergence is proportional to the difference in gaze angles.

Version determines the direction of gaze. Version is roughly proportional to the sum (average) of the gaze direction of the two eyes and thus determines the direction to the horopter. The horopter can also be tilted by a small rotation of the eyes referred to as cyclo-torsion.

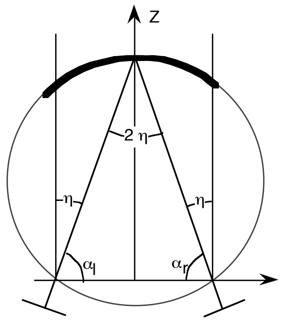
The sum and difference relation of the eye gaze has been known since the early 19th century and is referred to as the Vief-muller circle.

Version (left-right angle) is approximately the sum of the gaze angles. $\alpha_c = \frac{\alpha_L + \alpha_R}{2}$

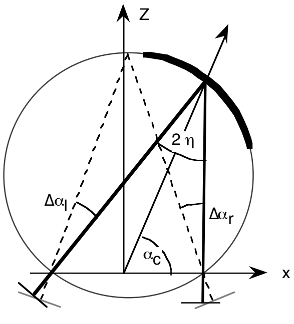
Vergence (depth) is approximately proportional to difference. $\Delta z = \frac{\Delta\alpha_L - \Delta\alpha_R}{2}$



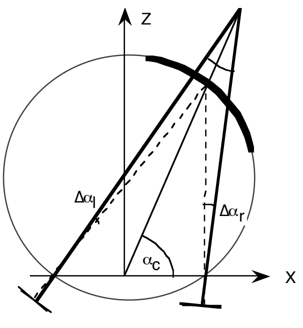
Vergence and version are described by the Vief-Muller Circle.



Symmetric Vergence

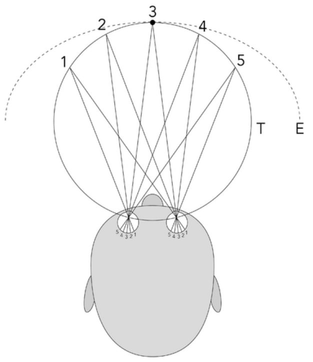


Vergence



Version

Vergence and version are redundantly controlled by retinal matching and by focusing of the lenses in the eyes (accommodation). Note that human's lose the ability to accommodate (focus the lens) around the age of 45 as the lens becomes rigid.

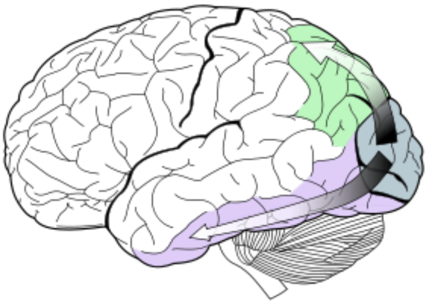


T - Theoretical (mathematical) horopter

E - Empirical (observed) horopter - A twisted cubic surface (larger than T).

2.6 The Visual Cortex

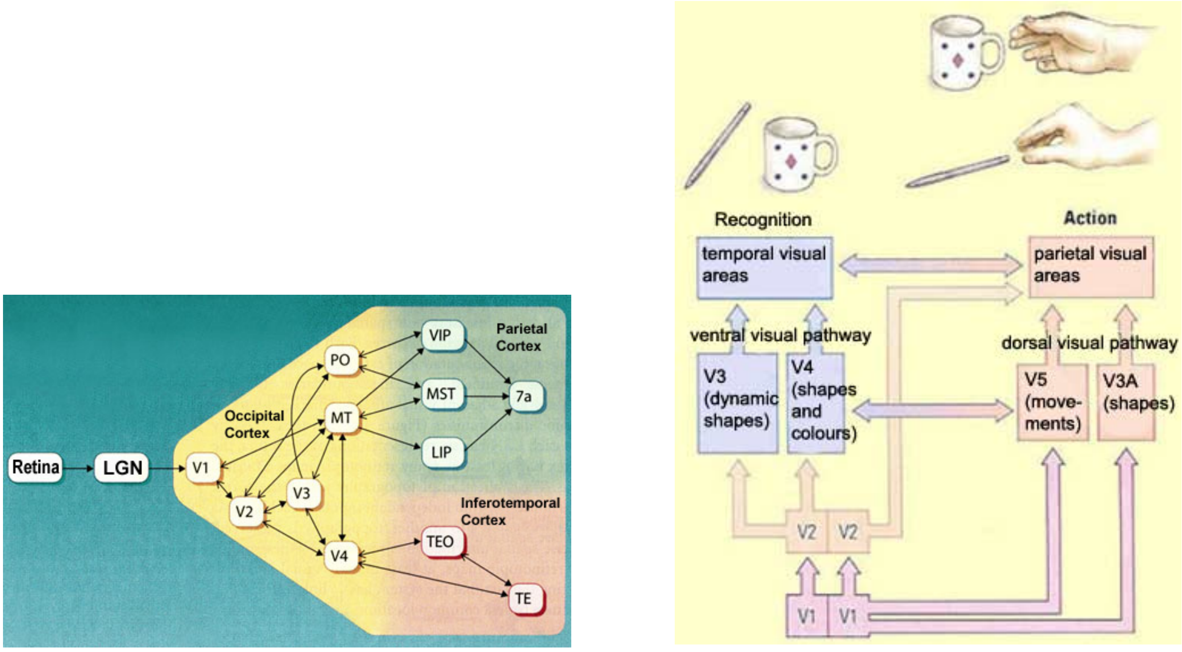
Retinal maps are relayed through the LGN to the primary visual cortex, where they propagate through the Dorsal and Lateral Visual pathways. Only activations that are not suppressed by the LGN are relayed.



The dorsal visual pathway (green) is the "action pathway". It controls motor actions and spatial organization of perception (relative 3D position), expressed as depth and direction of gaze as relayed by the Superior Colliculus.

The ventral visual pathway (purple) is uses color, appearance and motion to recognize phenomena and objects. (Phenomena are anything that can be perceived, regardless whether it is a solid objects or has a name. Clouds, rivers, rain, day are phenomena.)

These dorsal and ventral pathways are divided into a number of interacting subsystems (visual areas). Most human actions require input from both pathways. For example, consider the task of grasping a cup. The brain must recognize and locate the cup, and direct the hand to grasp the cup.



Architectural models of the human visual system

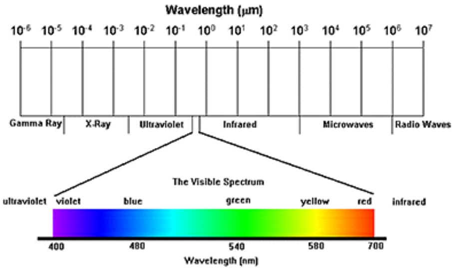
As with any model, this description is a simplification of a much more complex process.

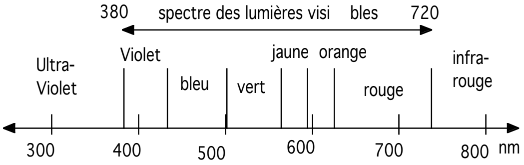
Biological theories posit that this architecture is the result of genetic programming, by a process of evolution. Minor variations that bring survival value are more likely to be reproduced. Variations that degrade survival are not.

Thus, biology tells us that while the weights of individual neurons may be affected (learned) by experiences during development, the architecture is determined by human genetics.

3 Color Spaces and Color Models

The human eye is capable of sensing photons with wavelengths from 380 nanometers to 720 nanometers.

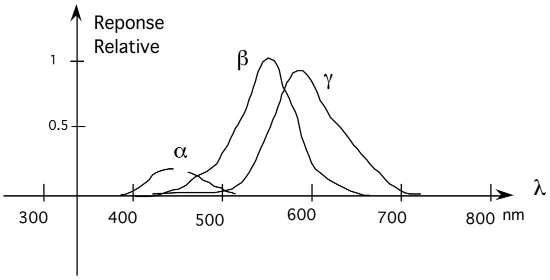




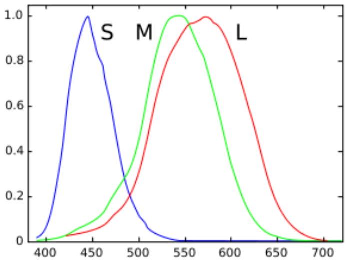
Perception of a photon by the retina is a probabilistic phenomena, with a fraction of the photons detected at each frequency.

3.1 Color Perception by Cones

The human retina is a tissue composed of rods, cones and bi-polar cells.
Cones provide chromatique "day vision".



Relative Sensitivities

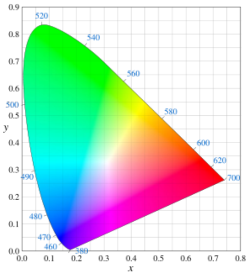


Normalised Sensitivities

Human Cones employ 3 pigments:

cyanolabe α 400–500 nm peak at 420–440 nm
chlorolabe β 450–630 nm peak at 534–545 nm
erythrolabe γ 500–700 nm peak at 564–580 nm

Perception of cyanolabe is low probability, hence poor sensitivity to blue.
Perception of chlorolabe and erythrolabe are more sensitive.



The three pigments give rise to a color space shown here (CIE model).

Note, these three pigments do NOT map directly to color perception.

Color perception is MUCH more complex, and includes a phenomena known as "color constancy" whereby the same color is perceived independent of ambient illumination. For example, yellow is always yellow, despite changes to the spectrum of an ambient source

Many color models have been proposed but each has its strengths and weaknesses. All models are approximations.

3.2 Nocturnal Perception by Rods

Rods provide night vision. Night vision is achromatique. It does not provide color perception. Night vision is low acuity - Rods are dispersed over the entire retina.

Rods are responsible for perception of very low light levels and provide night vision. Rods employ a very sensitive pigment named "rhodopsin".

Rodopsin is sensitive to a large part of the visible spectrum of with a maximum sensitivity around 510 nanometers.

Rhodopsin sensitive to light between 0.1 and 2 lumens, (typical moonlight) but is destroyed by more intense lights.

Rhodopsin can take from 10 to 20 minutes to regenerate.

3.3 Bayer Matrix Retina

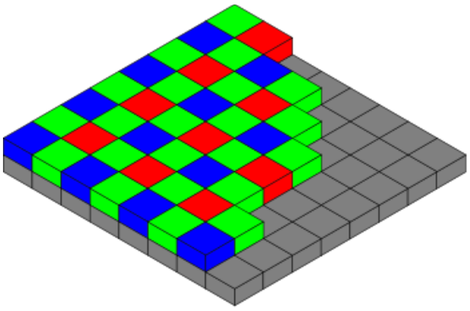
Silicon semiconductors respond to light by emitting photons (the Einstein effect), thus generating a charge. A silicon retina is composed of a matrix of individual photocells cells (sensels) that convert photons to positive voltage.

Note that silicon is sensitive to light into the near infrared (< 1500 Nm). Color filters are used to limit the spectrum of light to the visible spectrum.

Most modern digital cameras employ a Bayer Mosaic Retina, named after its inventor, Bryce E. Bayer of Eastman Kodak who patented the design in 1976.

A Bayer filter mosaic is a color filter array (CFA) for arranging RGB color filters on a square grid of photosensors. The filter pattern is 50% green, 25% red and 25% blue, hence is also called RGBG, GRGB, or RGGB.

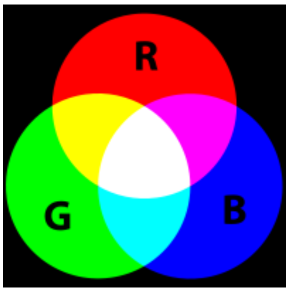
The Bayer mosaic uses twice as many green elements as red or blue to mimic the pigments of the human eye. These elements are referred to as sensor elements, sensels, pixel sensors, or simply pixels;



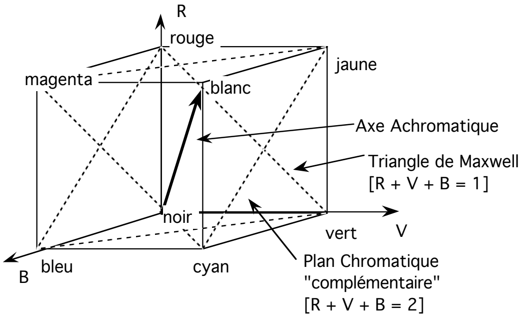
The voltage values on each sensel are converted to numeric values, interpolated and processed to provide image pixels. This step is sometimes called "image reconstruction" in the image processing community, and is generally in the camera as each image is recorded. The result an image with colors coded as independent components, typically Red-Green-Blue (RGB) or Hue-Luminance and Saturation (HLS).

3.4 The RGB Color Model

RGB is one of the oldest color models, originally proposed by Isaac Newton. This is the model used by most color cameras.



The RGB model "pretends" that Red, Green and Blue are orthogonal (independent) axes of a Cartesian space.



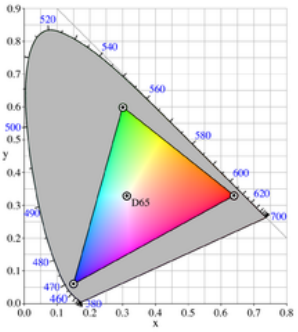
The achromatic axis is R=G=B.
Maxwell's triangle is the surface defined when R+G+B = 1.
A complementary triangle exists when R+G+B = 2.

For printers (subtractive color) this is converted to CMY (Cyan, Magenta, Yellow).

$$\begin{pmatrix} C \\ M \\ Y \end{pmatrix} = \begin{pmatrix} R_{\max} \\ G_{\max} \\ B_{\max} \end{pmatrix} - \begin{pmatrix} R \\ G \\ B \end{pmatrix}$$

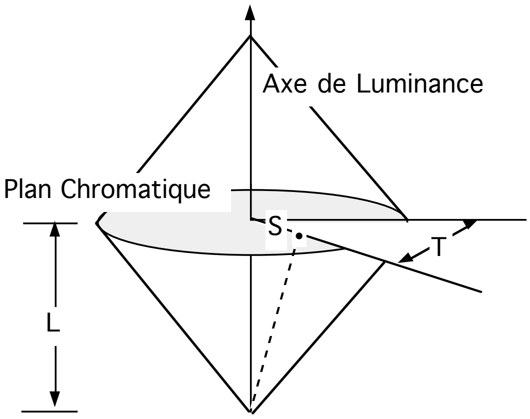
3.5 The HLS color model

The RGB model only captures a small part of visible colors:



Painters and artists generally use the HLS: Hue Luminance Saturation model.

HLS is a polar coordinate model for and hue (perceived color) and saturation.
The polar space is placed on a third axis. The size of the disc corresponds to the range of saturation values available.



One (of many possible) mappings from RGB:

Luminance : $L = (R + B + B)$

Saturation : $1 - 3*\min(R, G, B)/L$

Hue : $x = \cos^{-1}\left(\dfrac{\frac{1}{2}(R-G)+(R-B)}{\sqrt{(R-G)^2+(R-B)(G-B)}}\right)$

if B>G then H = x else H = 2π–x.

3.6 Color Opponent Model

Color Constancy: The subjective perception of color is independent of the spectrum of the ambient illumination.

Subjective color perception is provided by "Relative" color and not "absolute" number of photons.

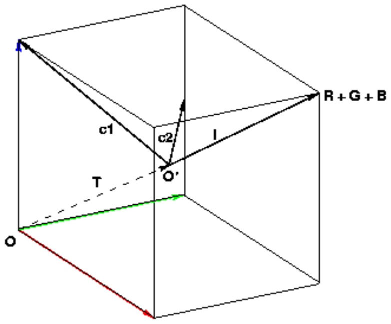
This is commonly modeled using a Color Opponent space.

The opponent color theory suggests that there are three opponent channels: red versus green, blue versus yellow, and black versus white (the latter type is achromatic and detects light-dark variation, or luminance).

This can be computed from RGB by the following transformation:
Luminance : L = R+G+B
Chrominance: C1 = (R-G)/2
C2 = B – (R+G)/2

as a matrix :

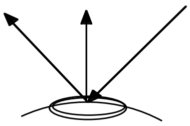
$$\begin{pmatrix} L \\ C_1 \\ C_2 \end{pmatrix} = \begin{pmatrix} 1 & 1 & 1 \\ 1 & -1 & 0 \\ -0.5 & -0.5 & 1 \end{pmatrix} \begin{pmatrix} R \\ G \\ B \end{pmatrix}$$



Such a vector can be "steered" to accommodate changes in ambient illumination.

3.7 Separating Specular and Lambertian Reflection.

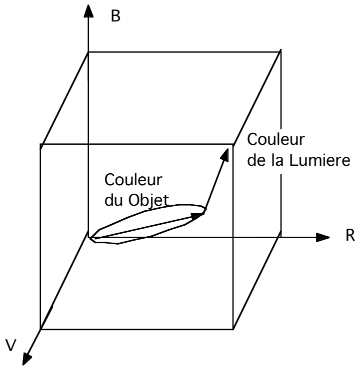
Consider what happens at a specular reflection.



The specularity has the same spectrum as the illumination.
The rest of the object has a spectrum that is the product of illumination and pigments.

This scan be seen in a histogram of color:

$$\forall \vec{C}(i,j) \ : \ H(\vec{C}(i,j)) = H(\vec{C}(i,j)) + 1$$



Two clear axes emerge:
One axis from the origin to the RGB of the product of the illumination and the source.
The other axis towards the RGB representing the illumination.