Computer Vision

James L. Crowley

M2R MoSIG option GVR | Fall Semester
12 October 2012

Lesson 3

Visual Perception in Man and Machine

Lesson Outline:

1 The Physics of Light

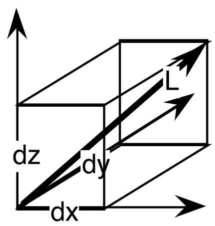
1.1 Photons and the Electo-Magnetic Spectrum

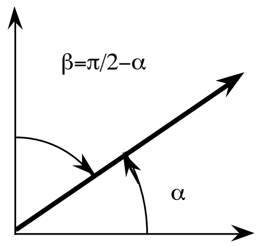
A photon is a resonant electromagnetic oscillation.
The resonance is described by Maxwell's equations.
The magnetic field is strength determined the rate of change of the electric field, and the electric field strength is determined by the rate of change of the magnetic field.

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

Direction of propagation and direction of polarity can be represented as a vector of Cosine angles.

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





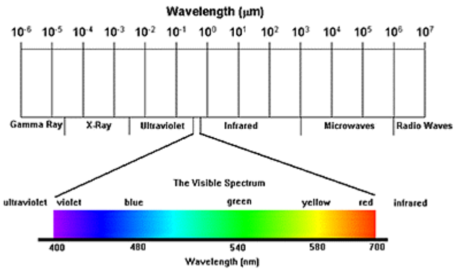
Photon propagation is a probabilistic phenomenon, described by Quantum Chromo-Dynamics. Photons are created and absorbed by abrupt changes in the orbits of electrons. Absorption and creation are probabilistic (non-deterministic) events.

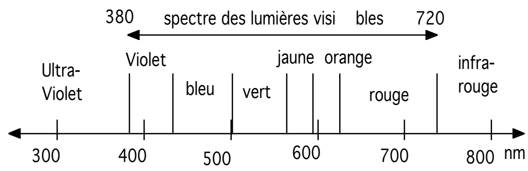
Photons sources generally emit photons over a continuum of directions (a beam) and continuum of wavelengths (spectrum). The beam intensity is measured in Lumens, and is equivalent to Photons/Meter^2.

A lumen a measure of the total "amount" of visible light emitted by a source.
The lumen can be thought of as a measure of the total photons of visible light in some defined beam or angle, or emitted from some source.

The beam spectrum gives the probability of a photon having a particularly wavelength, S(λ).

The human eye is capable of sensing photons with a wavelength between 380 nanometers and 720 nanometers.



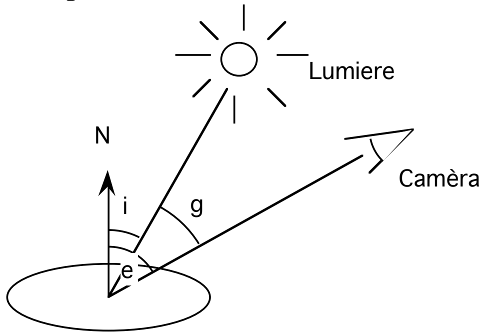


Perception is a probabilistic Phenomena.

1.2 Albedo and Reflectance Functions

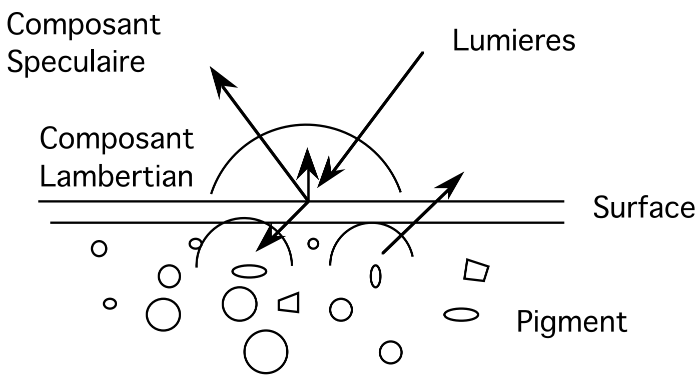
The albedo of a surface is the ratio of photons emitted over photons received. Albedo is described by a Reflectance function

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



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 percentage are rejected by an interface layer (determined by the wavelength). The remainder penetrate and are absorbed by molecules near the surface (pigments).



Most reflectance functions can be modeled as a weighted sum of two components: A Lambertian component and a specular component.

$$R(i, e, g, \lambda) = c\, R_S(i, e, g, \lambda) + (1 - c)\, R_L(i, \lambda)$$

Specular Reflection

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

An example of a specular reflector is a mirror.
All (almost all) of the photons are reflected at the interface level with no change in spectrum.

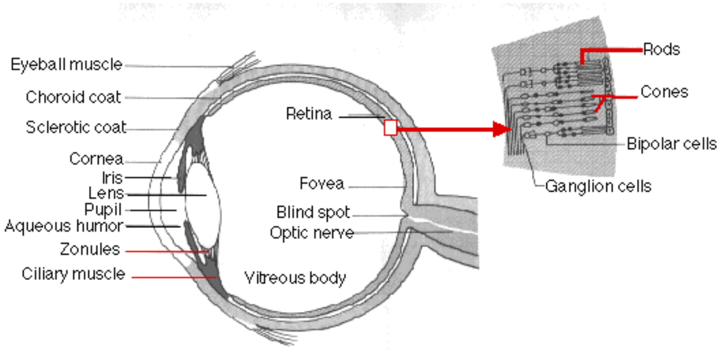
Lambertion Reflection

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

Paper, and fresh snow are examples of Lambertian reflectors.

2 The Human Visual System

2.1 The Human Eye

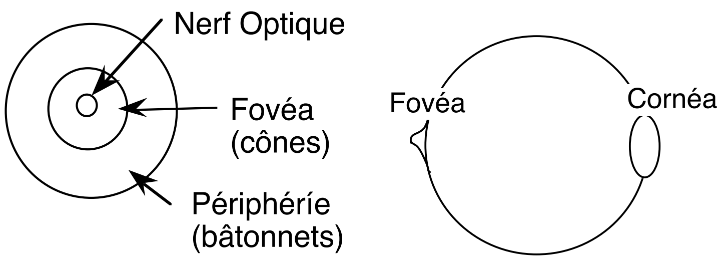


The human eye is a spherical globe filled with transparent liquid.
An opening (iris) allows light to enter and be focused by a lens.
Light arrives at the back of the eye on the Retina.

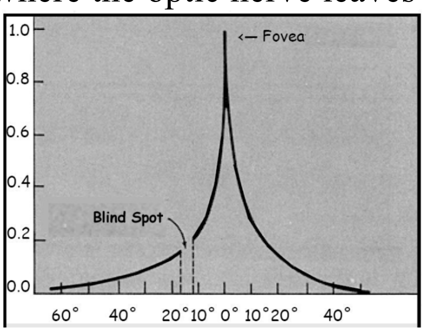
2.2 The Retina

The human retina is a tissue composed of a 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



The cones are distributed over a non-uniform region in the back of the eye.
The density of cones decreases exponentially from a central point.
The fovea contains a "hole" where the optic nerve leaves the retina.



The central region of the fovea is concentrates visual acuity and is used for recognition and depth perception. The peripheral regions have a much lower density of cones, and are used for to direct eye movements.

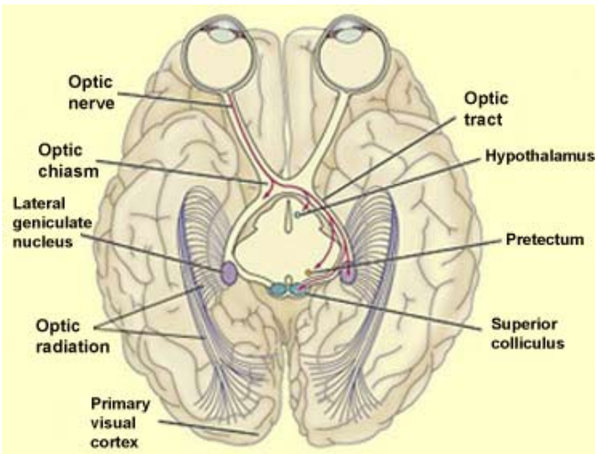
The eye perceives only a small part of the world at any instant. However, the muscles rotate the eyes at

The optical nerves leave the retina and are joined at Optic Chiasm.
Nerves then branch off to the Lateral Geniculate Nucleus (LGN) and the Superior Colliculus.

Nerves branch out from the LGN to provide "retinal maps" to the different visual cortexes as well as the "Superior Colliculus".

Surprisingly, 80% of the excitation of the LGN comes from the visual cortex!
The LGN seems to act as a filter for visual attention.

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



2.3 The Superior Colliculus

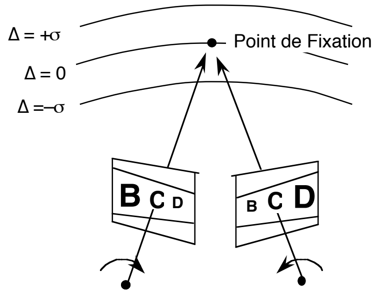
The first visual filter is provided by fixation, controlled by the Superior Colliculus.
The Superior Colliculus is a Feed-Forward (predictive) control system for binocular fixation. The Superior Colliculus is composed of 7 layers receiving stimulus from the frontal cortex, the lateral and dorsal cortexes, the auditory cortex and the retina.

2.4 Vergence and Version

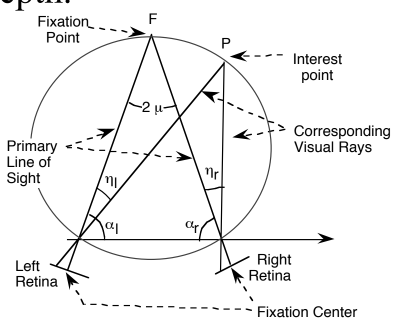
At any instant, the human visual system focuses processing on a small region of 3D space called the Horopter.

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.

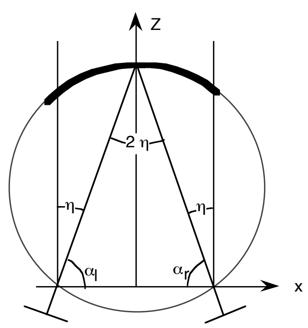
The horopter is controlled by the Superior Colliculus, and can move about the scene in incredibly rapid movements (eye scans). Scanning the horopteur allows the cortex to build up a composite model of the external world.



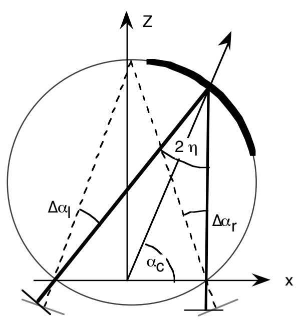
Eye movements can decomposed into "Version" and "Vergence".
Version perceives relative direction in head centered coordinates.
Vergence perceives relative depth.



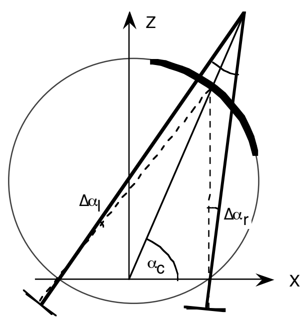
Vergence and version are described by the Vief-Muller Circle.
Version (angle) is the sum of the eye angles.
Vergence (depth) is proportional to difference.



Symmetric Vergence



Vergence

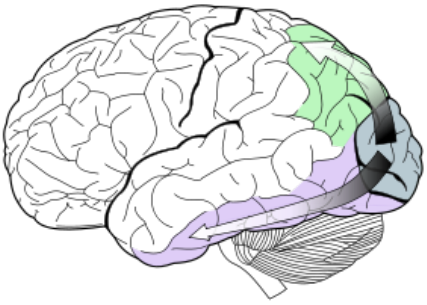


Version

Vergence and version are redundantly controlled by retinal matching and by focusing of the lenses in the eyes (accommodation).

2.5 The Visual Cortex

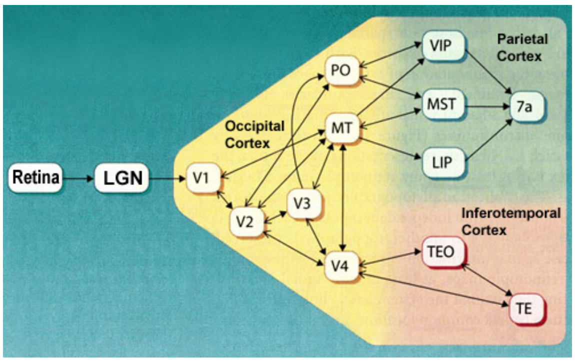
Retinal maps are relayed through the LGN to the primary visual cortex, where they propagate through the Dorsal and Lateral Visual pathways.

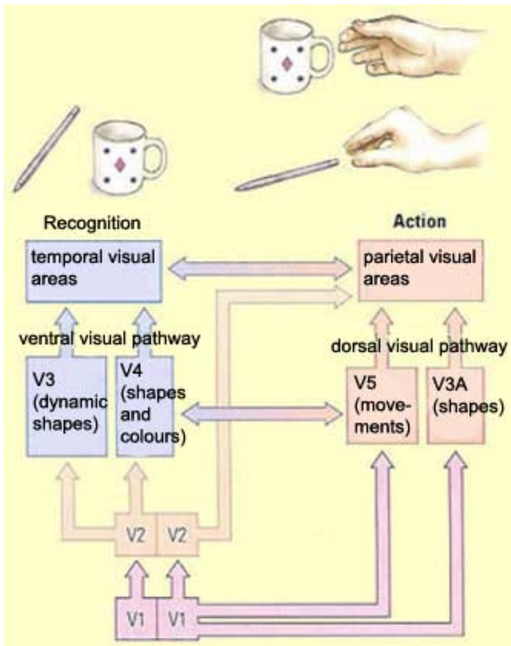


Dorsal visual pathway (green) is the "action pathway".
It controls motor actions. Most of the processing is unconscious.
It makes use of spatial organization (relative 3D position), including depth and direction information from the Superior Colliculus.

The ventral visual pathway (purple) is used in recognizing objects.
It makes use of color and appearance.

These two pathways are divided into a number of interacting subsystems (visual areas).





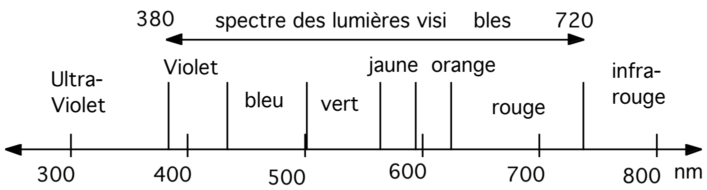
Most human actions require input from both pathways. For example, consider the task of grasping a cup. The brain must recognize and locate the cup, and direct the hand to grasp the cup.

3 Color Spaces and Color Models

3.1 Color Perception

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

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.

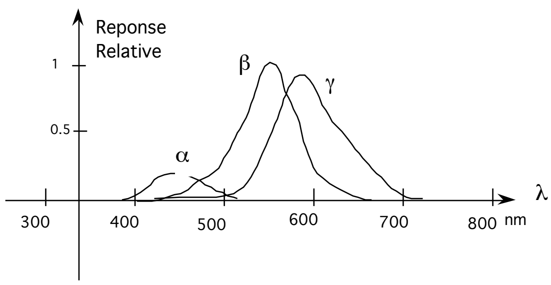


Rods are responsible for perception of very low light levels and provide night vision. Rods employ a very sensitive pigment named "rhodopsin".

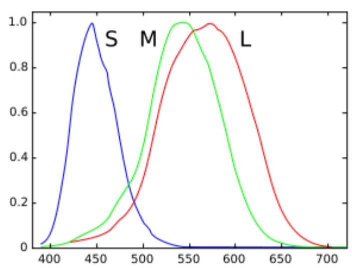
Rodopsin is sensitive to a large part of the visible spectrum of with a maximum sensitivity around 510 nano-meters.

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.



Relative Sensitivities

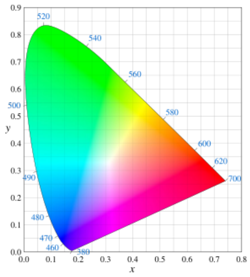


Normalised Sensitivities

Cones provide our chromatique "day vision". 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 difficult to model phenomena known as "color constancy".

For example, yellow is always yellow, despite changes to the spectrum of an ambiant source

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

3.2 Bayer Matrix Retina

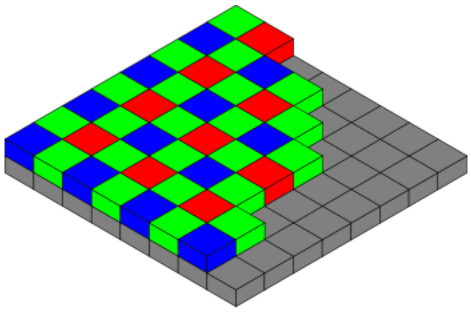
Silicon semiconductors respond to light by emitting photons (Einstein effect), thus generating a charge. A silicon retina is composed of a matrix of individual photocells cells (sensels) that convert photons to positive voltage.

Note that silicon is sensitive to light out into the near infrared (< 1500 Nm). Color filters are used to limit the spectrum of light reaching each photo-cell.

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.

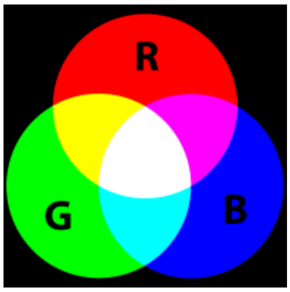
The Bayer mosaic uses twice as many green elements as red or blue to mimic the pigments of the human eye. These elements are referred to as sensor elements, sensels, pixel sensors, or simply pixels;



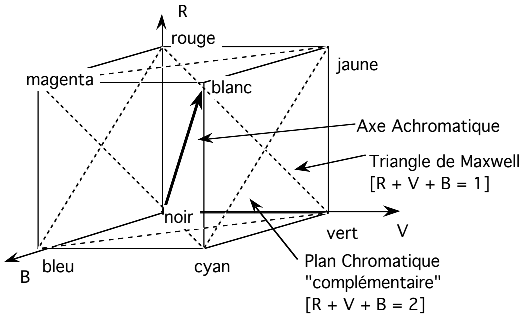
The voltage values on each sensel are converted to numeric values, interpolated and processed to provide image pixels. This step is sometimes called "image reconstruction" in the image processing community, and is generally carried out on the retina. The result is generally an image with colors coded as independent components: RGB.

3.3 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.



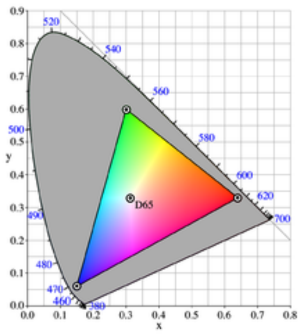
The achromatic axis is R=G=B.
Maxwell's triangle is the surface defined when R+G+B = 1.
A complementary triangle exists when R+G+B = 2.

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

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

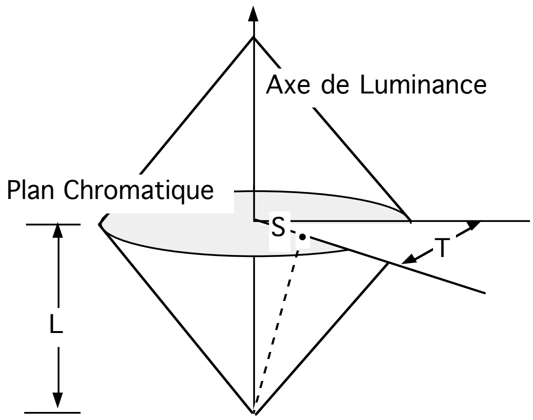
3.4 The HLS color model

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.

3.5 Color Opponent Model

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

Subjective color perception is provide by "Relative" color and not "absolute" measurements.

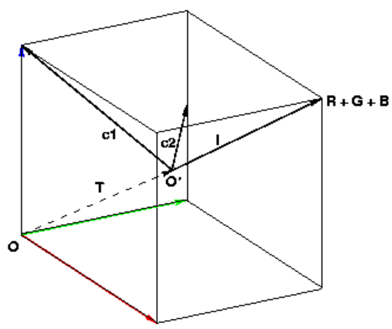
This is commonly modeled using a Color Opponent space.

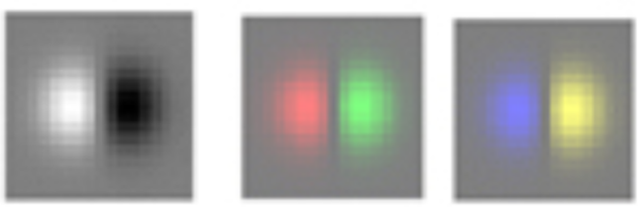
The opponent color theory suggests that there are three opponent channels: red versus green, blue versus yellow, and black versus white (the latter type is achromatic and detects light-dark variation, or luminance).

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 :

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

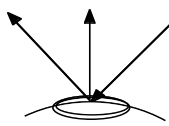




Such a vector can be "steered" to accommodate changes in ambient illumination.

3.6 Separating Specular and Lambertian Reflection.

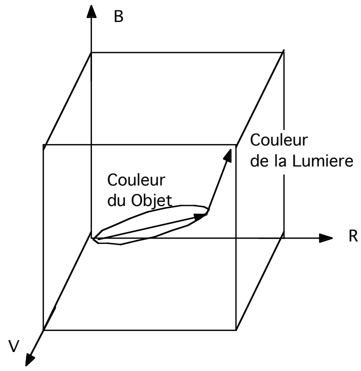
Consider what happens at a specular reflection.



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:

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



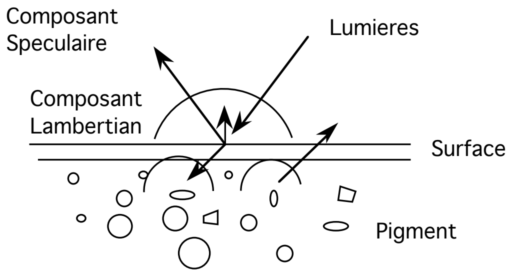
Two clear axes emerge:
One axis from the origin to the RGB of the product of the illumination and the source.
The other axis towards the RGB representing the illumination.

4 Detection and Tracking using Color

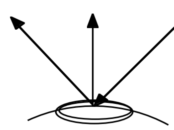
4.1 Object detection by pigment color

Recall the Bichromatic reflection function :

$$R(i, e, g, \lambda) = \alpha R_S(i, e, g, \lambda) + (1-\alpha) R_L(i, \lambda)$$



For Lambertian reflection, the intensity $\| P(i,j) \|$ is generally determined by changes in surface orientation, while color is determined by Pigment.



Thus Luminance captures surface orientation (3D shape) while
Chrominance is a signature for object pigment (identity)

Thus it is often convenient to transform the (RGB) color pixels into a color space that separates Luminance from Chrominance.

$$\begin{pmatrix} L \\ C_1 \\ C_2 \end{pmatrix} \Leftarrow \begin{pmatrix} R \\ G \\ B \end{pmatrix}$$

A popular space for skin detection is normalized R, G.

Luminance: $L = R+G+B$

Chrominance : $c_1 = r = \dfrac{R}{R+G+B}$ $\quad c_2 = g = \dfrac{G}{R+G+B}$

Suppose that these are coded with N values between 0 and $N-1$

$$c_1 = \mathrm{trunc}\left(N \cdot \frac{R}{R+G+B} \right) \qquad c_2 = \mathrm{trun}\left(N \cdot \frac{G}{R+G+B} \right)$$

Luminance normalized RG is often used for skin detection.
Skin pigment is generally always the same color. Luminance can change with pigment density, and skin surface orientation. Chrominance will remain invariant.

Thus we can use $\vec{c} = \begin{pmatrix} r \\ g \end{pmatrix}$ as a "signature for detecting skin.

4.2 Histograms

A histogram is a table of frequency of occurrence. We can use histograms to estimate probability densities for integer valued features.

Assume integer x from a bounded set of values, such that $x \in [x_{min}, x_{max}]$,
the probability that a random observation X takes on x is

$$P(X=x) = \frac{1}{M} h(x)$$

The validity of this depends on the ratio of the number of sample observations M and the number of cells in the histogram Q=N

This is true for vectors as well as values.
For a vector of D values $\vec{x}$ the table has D dimensions. $h(x_1, x_2, \ldots, x_D) = h(\vec{x})$

The average error depends on the ration $Q=N^D$ and M. : $E_{ms} \sim O(\frac{Q}{M})$

We need to assure that $M >> Q = N^d$
As a general rule : M should be greater than $8N^d$

4.3 Color Skin Detection

We can use statistics chrominance to build a very simple skin detector.

To use a Bayesian approach we need to represent the probability for each possible chrominance. We can estimate probability for chrominance with a histogram calculated from a set of training images.

Suppose the training images are composed of M pixels $\{P_m\}$.
Suppose we project these into a set of M chrominance pixels. $S = \{\vec{c}_m\}$

We then allocate a 2D table : $h(\vec{c})$ of size N x N.

For example, for skin chrominance, N=32 seems to work well.
Let *$h(\vec{c})$* be a 32 x 32 table. Q = 32 x 32 = 1024 cellules

For each pixel, (i,j) possibly from S,

$$\underset{\vec{c}_m \in S}{\forall} h(\vec{c}_m) = h(\vec{c}_m) + 1$$

For M pixels in the training data, the histogram $h(\vec{c})$ of chrominance gives an estimate of the probability for a chrominance value within the data (or in similar data).

$$p(\vec{c}) = \frac{1}{M} h(\vec{c})$$

Important. The number of pixels, M, should be much larger than $Q = N^2$

We also can apply this to learn the probability chrominance for a target.

Suppose that we mark all pixels that belong to the target in the training data to obtain a subset $T \subset S$ composed of M_k target pixels.

We can learn a second histogram:

$$\underset{\vec{c}_m \in T}{\forall} h_k(\vec{c}_m) = h_k(\vec{c}_m) + 1$$

then the probability of observing a chrominance value $\vec{c}$ given the target is

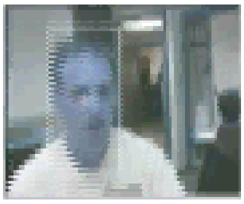
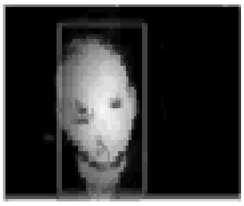
$$p(\vec{c} \mid \text{target}) = \frac{1}{M_k} h_k(\vec{c})$$

Because the target samples are a subset of the training data, the probability of a target pixel is

$$p(\text{target}) = \frac{M_k}{M}$$

From Bayes rule:

$$p(\text{target} \mid \vec{c}(i,j)) = \frac{p(\vec{c}(i,j) \mid \text{target})\,p(\text{target})}{p(\vec{c}(i,j))} = \frac{\frac{1}{M_k} h_k(\vec{c}(i,j)) \frac{M_k}{M}}{\frac{1}{M} h(\vec{c}(i,j))} = \frac{h_k(\vec{c}(i,j))}{h(\vec{c}(i,j))}$$

We can use this to convert each color pixel *c(i,j)* to a probability, *p(i,j)*, by table lookup.

$$p(i,j) = p(\text{target} \mid \vec{c}(i,j)) = \frac{h_k(\vec{c}(i,j))}{h(\vec{c}(i,j))}$$