Computer Vision

James L. Crowley

M2R MoSIG option GVR | Fall Semester
1 October 2009

Lesson 2

Color Perception in Man and Machine

<u>Lesson Outline</u>:

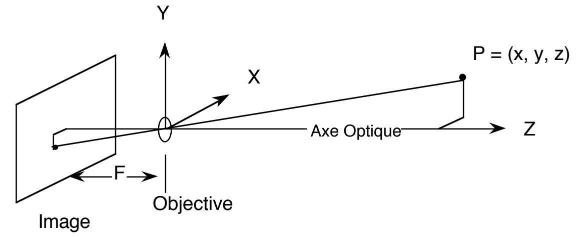
This lessons uses drawings from "The Brain from Top to Bottom" by Bruno Dubuc of McGill University. http://thebrain.mcgill.ca/) (published under "Copy-left").
as well as from WikiPedia.

1 The Camera Model

A "camera" is a closed box with an aperture (a "camera obscura"). Photons are reflected from the world, and pass through the aperture to form an image on the retina. Thus the camera coordinate system is defined with the aperture at the origin.

The Z (or depth) axis runs perpendicular from the retina through the aperture.
The X and Y axes define coordinates on the plane of the aperture.

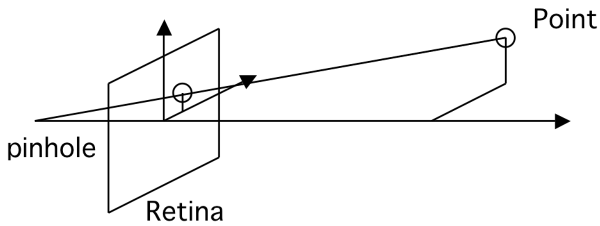
1.1 The Pinhole Camera



Points in the scene are projected to an "up-side down" image on the retina.

This is the "Pin-hole model" for the camera.

The scientific community of computer vision often uses the "Central Projection Model". In the Central Projection Model, the retina is placed in Front of the projective point.



We will model the camera as a projective transformation from scene coordinates, S, to image coordinates, i.

$$\vec{Q}^i = \mathbf{M}_s^i \vec{P}^s$$

This transformation is expressed as a 3x4 matrix:

$$M_s^i = \begin{pmatrix} m_{11} & m_{12} & m_{13} & m_{14} \\ m_{21} & m_{22} & m_{23} & m_{24} \\ m_{31} & m_{32} & m_{33} & m_{34} \end{pmatrix}$$

composed from 3 transformations between 4 reference frames.

1.2 Extrinsic and Intrinsic camera parameters

The camera model can be expressed as a function of 11 parameters.
These are often separated into 6 "extrinsic" parameters and 5 "intrinsic" parameters:

Thus the "Extrinsic" parameters of the camera describe the camera position and orientation in the scene. These are the six parameters:

Extrinsic Parameters = $(x, y, z, \theta, \varphi, \gamma)$

The intrinsic camera parameters express the projection to the retina, and the mapping to the image. These are :

F : The "focal" length
C_x, C_y : the image center (expressed in pixels).
D_x, D_y : The size of pixels expressed in pixels/mm.

The effect of these parameters may be separated by decomposing the projection matrix, M_s^i , into a composition of three matrices : $M_s^i = C_r^i P_c^r T_s^c$

$$\vec{Q} = M_s^i \vec{P} = C_r^i P_c^r T_s^c \vec{P}$$

1.3 Coordinate Systems

This transformation can be decomposed into 3 basic transformations between 4 reference frames. The reference frames are:

Coordinate Systems:
Scene Coordinates:
Point Scène: $P^s = (x_s, y_s, z_s, 1)^T$

Camera Coordinates:
external world: $P^c = (x_c, y_c, z_c, 1)^T$
Retina: $Q^r = (x_r, y_r, 1)^T$

Image Coordaintes
Image: $Q^i = (i, j, 1)^T$

The transformations are represented by Homogeneous projective transformations.

$$\vec{Q}^i = C_r^i \vec{Q}^r \qquad \vec{Q}^r = P_c^r \vec{P}^c \qquad \vec{P}^c = T_s^c \vec{P}^s$$

These express

1) A translation/rotation from scene to camera coordinates: T_s^c
2) A projection from scene points in camera coordinates to the retina: P_c^r
3) Sampling scan and A/D conversion of the retina to give an image: C_r^i

When expressed in homogeneous coordinates, these transformations are composed as matrix multiplications.

$$\vec{Q} = M_s^i \vec{P} = C_r^i P_c^r T_s^c \vec{P}$$

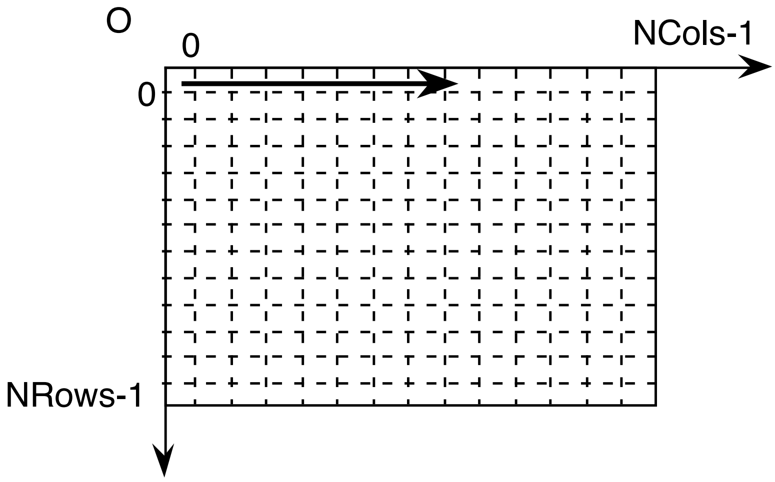
We will use "tensor notation" to keep track of our reference frames

last week we saw P_c^r and T_s^c. Let us look at C_r^i

1.4 From the Retina to Digitized Image

The "intrinsic parameters of the camera are F and C_x, C_y, D_x, D_y

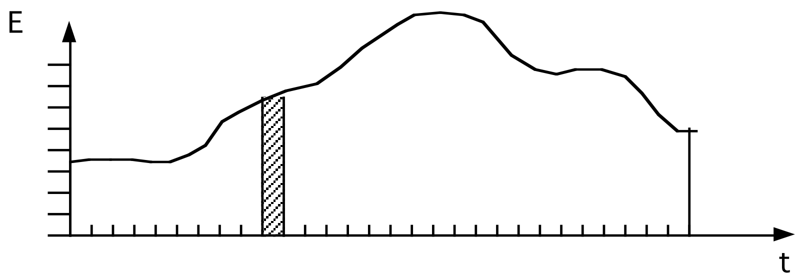
The image frame is composed of pixels (pic<u>ture</u> el<u>ements</u>)



Note that pixels are not necessarily square.

Typical image sizes VGA : 640 x 480

<u>Sampling and A/D Conversion.</u>



The mapping from retina to image can be expressed with 4 parameters:

C_x, C_y : the image center (expressed in pixels).
D_x, D_y : The size of pixels expressed in pixels/mm.

$i = x_r D_i$ (mm · pixel/mm) $+ C_i$ (pixel)
$j = y_r D_j$ (mm · pixel/mm) $+ C_j$ (pixel)

Transformation from retina to image :

$$Q^i = \mathbf{C}_r^i \; Q^r$$

$$\begin{pmatrix} i \\ j \\ 1 \end{pmatrix} = \begin{pmatrix} D_i & 0 & C_i \\ 0 & D_j & C_j \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} x_r \\ y_r \\ 1 \end{pmatrix}$$

That is:

$$\begin{pmatrix} wi \\ wj \\ w \end{pmatrix} = \begin{pmatrix} D_i & 0 & C_i \\ 0 & D_j & C_j \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} wx_r \\ wy_r \\ w \end{pmatrix}$$

1.5 The Complete Camera Model

$$\boxed{P^i = \mathbf{C}_r^i \; \mathbf{P}_c^r \; \mathbf{T}_s^c \, P^s = \mathbf{M}_s^i \; P^s}$$

$$\begin{bmatrix} w\,i \\ w\,j \\ w \end{bmatrix} = \mathbf{M}_s^i \begin{bmatrix} x_s \\ y_s \\ z_s \\ 1 \end{bmatrix}$$

and thus

$$i = \frac{w\,i}{w} = \frac{M_s^1 \cdot P^s}{M_s^3 \cdot P^s} \qquad\qquad j = \frac{w\,j}{w} = \frac{M_s^2 \cdot P^s}{M_s^3 \cdot P^s}$$

or

$$i = \frac{w\,i}{w} = \frac{M_{11} X_s + M_{12} Y_s + M_{13} Z_s + M_{14}}{M_{31} X_s + M_{32} Y_s + M_{33} Z_s + M_{34}}$$

$$j = \frac{w\,j}{w} = \frac{M_{21} X_s + M_{22} Y_s + M_{23} Z_s + M_{24}}{M_{31} X_s + M_{32} Y_s + M_{33} Z_s + M_{34}}$$

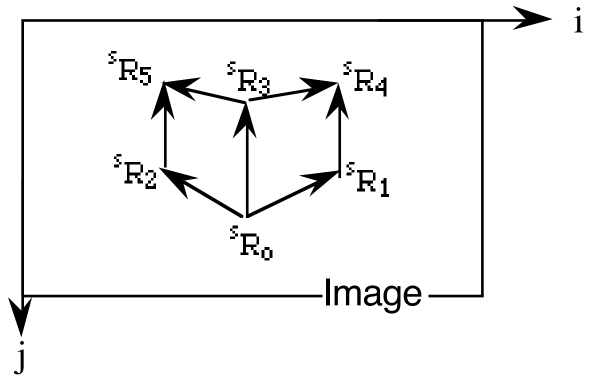
1.6 Calibrating the Camera

How can we obtain $\mathbf{M}_s^i$? By a process of calibration.

Observe a set of at least 6 non-coplanar points whose position in the world is known.

R^s_k for k=0,1,2,3,4,5 (s are the scene coordinate axes s=1,2,3)

For example, we can use the corners of a cube. Define the lower front corner as the origin, and the edges as unit distances.



The matric $\mathbf{M}_s^i$ is composed of 3x4=12 coefficients. However because, $\mathbf{M}_s^i$ is in homogeneous coordinates, the coordinate m_{34} can be set to 1.

Thus there are 12-1 = 11.
We can determine these coefficients by observing known points in the scene. (R^s_k).
Each point provides two coefficients. Thus, for 11 coefficients we need at least $5\frac{1}{2}$ points. With 6 points the system is overconstrained.

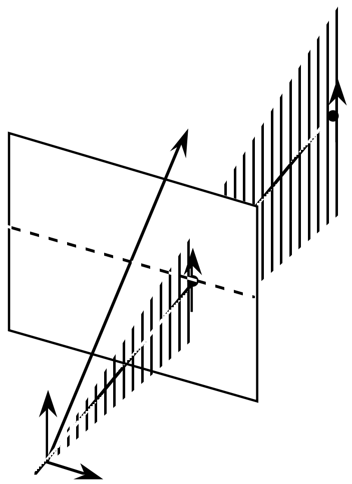
For each known calibration point R^s_k given its observed image position P^s_k, we can write:

$$i_k = \frac{w_k\, i_k}{w_k} = \frac{M^1_s \cdot R^s_k}{M^3_s \cdot R^s_k} \qquad j_k = \frac{w_k\, j_k}{w_k} = \frac{M^2_s \cdot R^s_k}{M^3_s \cdot R^s_k}$$

This gives 2 equations for each point.

$$(M_s^1 \cdot R^s{}_k) - i_k\,(M_s^3 \cdot R^s{}_k) = 0 \qquad (M_s^2 \cdot R^s{}_k) - j_k\,(M_s^3 \cdot R^s{}_k) = 0$$

Each pair of equations corresponds to the planes that pass though the image row and the image column of the observed image point $P^s{}_k$



The equation $(M_s^1 \cdot R^s{}_k) - i_k\,(M_s^3 \cdot R^s{}_k) = 0$ is the vertical plane that includes the projective center through the pixel $i=i_k$.

The equation $(M_s^2 \cdot R^s{}_k) - j_k\,(M_s^3 \cdot R^s{}_k) = 0$ is the horizontal plane that includes the projective center and the row $j=j_k$.

In tensor notation

given $P^i = \begin{pmatrix} wi \\ wj \\ w \end{pmatrix}$ we write : $P^i = \mathbf{M}_s^i\ R^s$

with k scene points, R_k^S and their image correspondance P_k^i we can write

$$P_k^i = \mathbf{M}\,{}_s^i\ R_k^s$$

with $i\,w = P_k^1/P_k^3$ et $j\,w = P_k^2/P_k^3$ for each image point k, there are two independent equations

$$\begin{pmatrix} p^1 \\ p^2 \\ p^3 \end{pmatrix} = \begin{pmatrix} wi \\ wj \\ w \end{pmatrix} \quad \text{donc} \quad \begin{pmatrix} i \\ j \\ 1 \end{pmatrix} = \begin{pmatrix} p^1/p^3 \\ p^2/p^3 \\ 1 \end{pmatrix}$$

and with $P_k^3 = \mathbf{M}\,{}_s^3\,R_k^3$

$$i = p^1/p^3 = \mathbf{M}_s^1 \; R_k^s \;/\; \mathbf{M}_s^3 \; R_k^s \;\Rightarrow i\,\mathbf{M}_s^3 \; R_k^s - \mathbf{M}_s^1 \; R_k^s = 0$$
$$j = p^2/p^3 = \mathbf{M}_s^2 \; R_k^s \;/\; \mathbf{M}_s^3 \; R_k^s \;\Rightarrow j\,\mathbf{M}_s^3 \; R_k^s - \mathbf{M}_s^2 \; R_k^s = 0$$

We can write this as :

$$\begin{pmatrix} R^1 & R^2 & R^3 & 1 & 0 & 0 & 0 & 0 & -iR^1 & -iR^2 & -iR^3 & -i \\ 0 & 0 & 0 & 0 & R^1 & R^2 & R^3 & 1 & -jR^1 & -jR^2 & -jR^3 & -j \end{pmatrix} \begin{pmatrix} \mathbf{M}_1^1 \\ \mathbf{M}_2^1 \\ \mathbf{M}_3^1 \\ \mathbf{M}_4^1 \\ \mathbf{M}_1^2 \\ \mathbf{M}_2^2 \\ \mathbf{M}_3^2 \\ \mathbf{M}_4^2 \\ \mathbf{M}_1^3 \\ \mathbf{M}_2^3 \\ \mathbf{M}_3^3 \\ \mathbf{M}_4^3 \end{pmatrix} = 0$$

For N non-coplanair points we can write 2N equations.

$$\mathbf{A} \;\; \mathbf{M}_s^i = 0.$$

We then use least squares to minimize the criteria:

$$\mathbf{C} = \| \mathbf{A}\, \mathbf{M}_s^i \|$$

For example, give a cube with observed corners

$P^L{}_o = (101, 221)$ $\quad$ $P^L{}_1 = (144, 181)$ $\quad$ $P^L{}_2 = (22, 196)$
$P^L{}_3 = (105, 88)$ $\quad$ $P^L{}_4 = (145, 59)$ $\quad$ $P^L{}_5 = (23, 67)$

Least squares will give:

$$\mathbf{M}_s^i = \begin{pmatrix} 55.886873 & -79.292084 & 1.276703 & 101.917630 \\ -22.289319 & -17.878203 & -134.345576 & 221.300658 \\ 0.100734 & 0.038274 & -0.008458 & 1.000000 \end{pmatrix}$$

1.7 Alternate Derivation using the Cross product

In classic matrix notation:

$$\vec{P} \times \mathbf{M}_s^i \, \vec{R} = 0$$

The term $\vec{R}$ can be factored to set $\vec{P} \, \vec{R} \times \mathbf{M}_s^i = 0$

This gives $\begin{pmatrix} 0 & -wR^s & jwR^s \\ -wR^s & 0 & -iwR^s \\ wR^s & -wR^s & 0 \end{pmatrix} \begin{pmatrix} \mathbf{M}_s^1 \\ \mathbf{M}_s^2 \\ \mathbf{M}_s^3 \end{pmatrix} = 0$

Where $\vec{R}$ and $\mathbf{M}_s^i$ are vectors. Thus:

$$\begin{pmatrix} 0 & 0 & 0 & 0 & wX & wY & wZ & w1 & -jwX & -jwY & -jwZ & -jw1 \\ -wX & -wY & -wZ & -w & 0 & 0 & 0 & 0 & -iwX & -iwY & -iwZ & -iw1 \\ wX & wY & wZ & w & -wX & -wY & -wZ & -w & 0 & 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} \mathbf{M}_1^1 \\ \mathbf{M}_2^1 \\ \mathbf{M}_3^1 \\ \mathbf{M}_4^1 \\ \mathbf{M}_1^2 \\ \mathbf{M}_2^2 \\ \mathbf{M}_3^2 \\ \mathbf{M}_4^2 \\ \mathbf{M}_1^3 \\ \mathbf{M}_2^3 \\ \mathbf{M}_3^3 \\ \mathbf{M}_4^3 \end{pmatrix} = 0$$

Two of the equations are independent.

2 Homography between two planes.

The projection of a plane to another plane is a degenerate case of the the project transform. In this case, the transform is bijective and reduces to a 3 x 3 invertible

This matrix can be used to rectify an image to a perpendicular view.

$$\boxed{Q^B = \mathbf{H}_A^B \; P^A}$$

In classic notation

$$\begin{pmatrix} w\,x_B \\ w\,y_B \\ w \end{pmatrix} = \mathbf{H}_A^B \begin{pmatrix} x_A \\ y_A \\ 1 \end{pmatrix} = \begin{pmatrix} m_{11} & m_{12} & m_{13} \\ m_{21} & m_{22} & m_{23} \\ m_{31} & m_{32} & m_{33} \end{pmatrix} \begin{pmatrix} x_A \\ y_A \\ 1 \end{pmatrix}$$

$$x_B = \frac{w\,x_B}{w} = \frac{m_{11}\,x_A + m_{12}\,y_A + m_{13}}{m_{31}\,x_A + m_{32}\,y_A + m_{33}}$$

$$y_B = \frac{w\,y_B}{w} = \frac{m_{21}\,x_A + m_{22}\,y_A + m_{23}}{m_{31}\,x_A + m_{32}\,y_A + m_{33}}$$

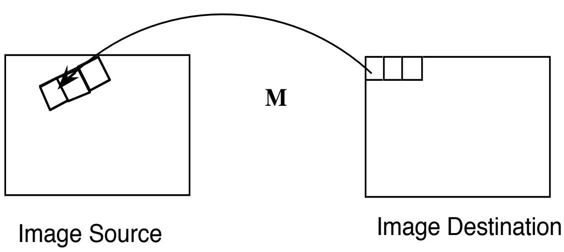
In tensor notation:

$$\boxed{Q^B = \mathbf{H}_A^B \; P^A}$$

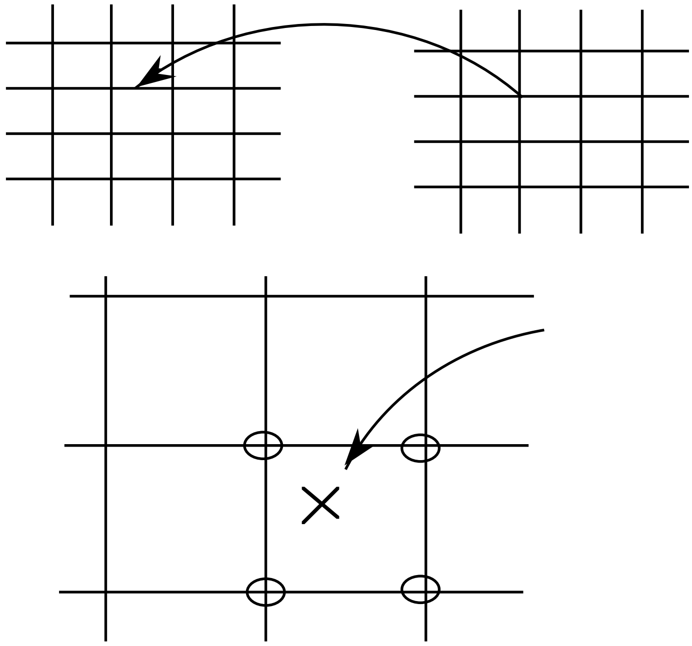
$$\begin{pmatrix} q^1 \\ q^2 \\ q^3 \end{pmatrix} = \mathbf{H}_A^B \begin{pmatrix} p^1 \\ p^2 \\ p^3 \end{pmatrix} = \begin{pmatrix} h_1^1 & h_2^1 & h_3^1 \\ h_1^2 & h_2^2 & h_3^2 \\ h_1^3 & h_2^3 & h_3^3 \end{pmatrix} \begin{pmatrix} p^1 \\ p^2 \\ p^3 \end{pmatrix}$$

$$x_B = \frac{q^1}{q^3} \qquad y_B = \frac{q^2}{q^3}$$

For each pixel in the destination image, (x_d, y_d) compute its position in the source image(x_s, y_s)



Determine the appropriate pixel value (intensity or color) for that positon and put this in teh destination.



The appropriate value is determined by interpolation:

zeroth order:	Nearest neighbor
First order:	Lineare or bilinear interpolation
second order	Cubic spline.

3 The Physics of Light

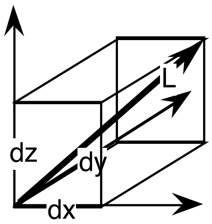
3.1 Photons and the Electo-Magnetic Spectrum

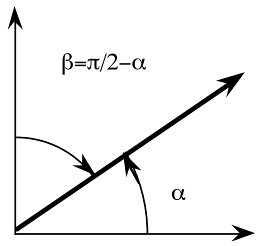
A photon is a resonant electromagnetic oscillation.
The resonance is described by Maxwell's equations.
The magnetic field is strength determined the rate of change of the electric field, and the electric field strength is determined by the rate of change of the magnetic field.

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.

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





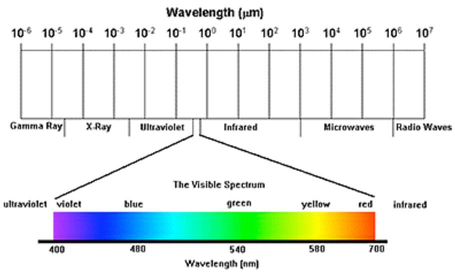
Photon propagation is a probabilistic phenomenon, described by Quantum Chromo-Dynamics. According to Feynman, photons "probably" travel in a straight line at "probably" the speed of light.

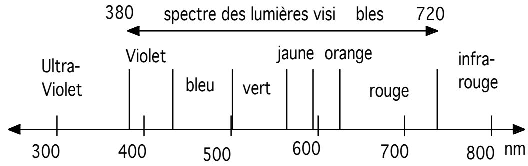
Photons can be created and absorbed by abrupt changes in the orbits of electrons. Absorption and creation are probabilistic (non-deterministic) events.

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.
The beam spectrum is 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.



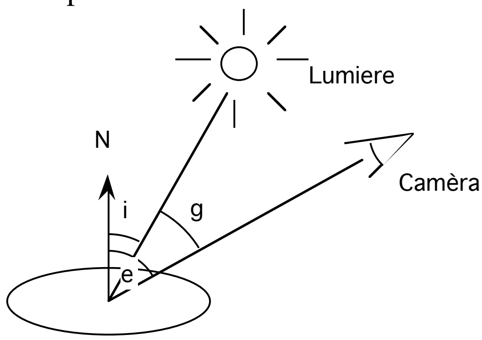


Perception is a probabilistic Phenomena.

3.2 Albedo and Reflectance Functions

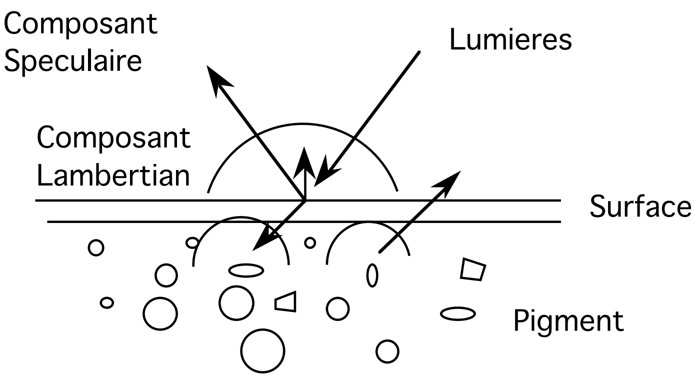
The albedo of a surface is the ratio of photons emitted over photons received.
Albedo is described by a Reflectance function

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



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.

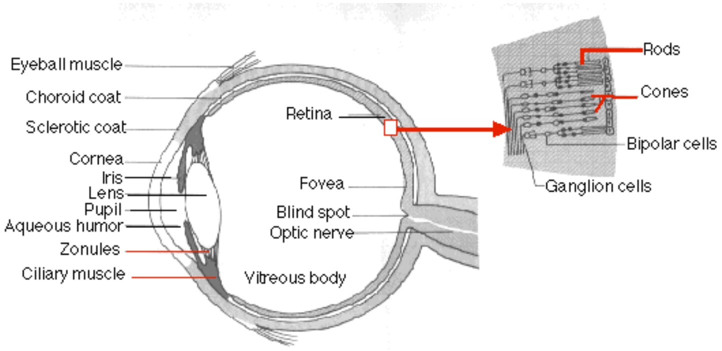
Lambertion Reflection

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

Paper, and fresh snow are examples of Lambertian reflectors.

4 The Human Visual System

4.1 The Human Eye

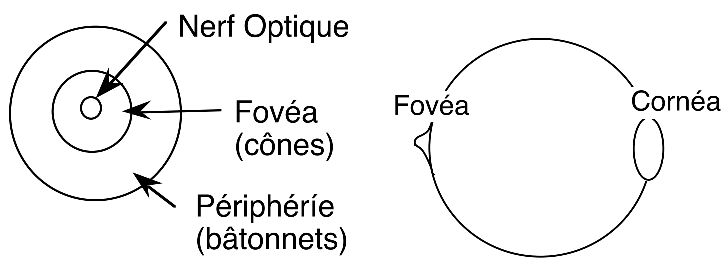


The human eye is a spherical globe filled with transparent liquid.
An opening (iris) allows light to enter and be focused by a lens.
Light arrives at the back of the eye on the Retina.

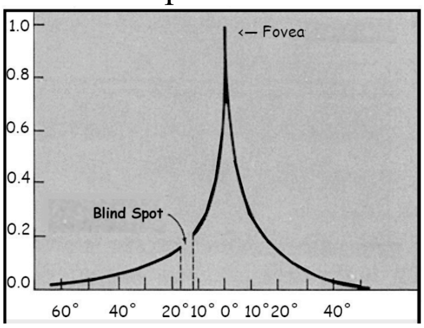
4.2 The Retina

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.

<u>Fovea and Peripheral regions</u>



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.

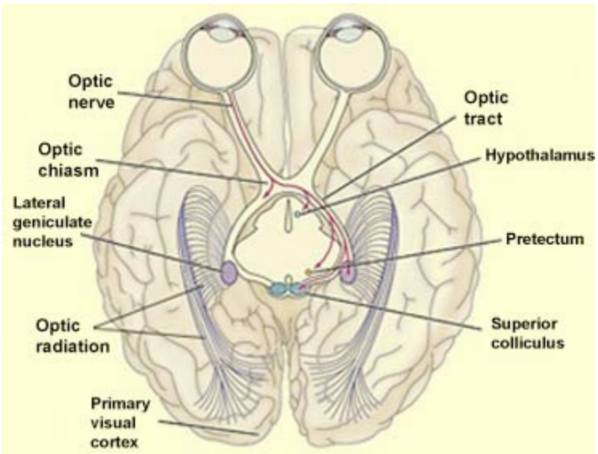
The eye perceives only a small part of the world at any instant. However, the muscles rotate the eyes at

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.



The Superior Colliculus

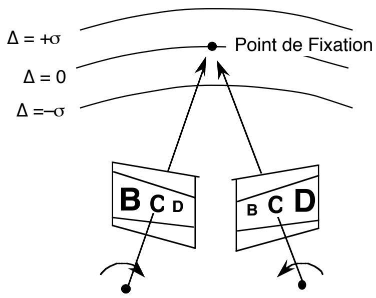
The first visual filter is provided by fixation, controlled by the Superior Colliculus.
The Superior Colliculus is a Feed-Forward (predictive) control system for binocular fixation. The Superior Colliculus is composed of 7 layers receiving stimulus from the frontal cortex, the lateral and dorsal cortexes,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

4.3 Vergence and Version

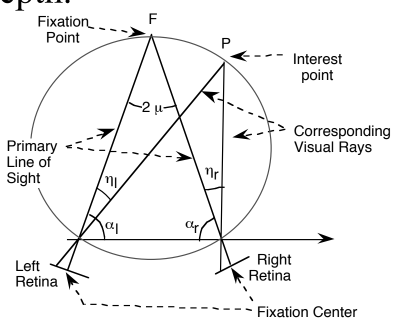
At any instant, the human visual system focuses processing on a small region of 3D space called the Horopter.

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.

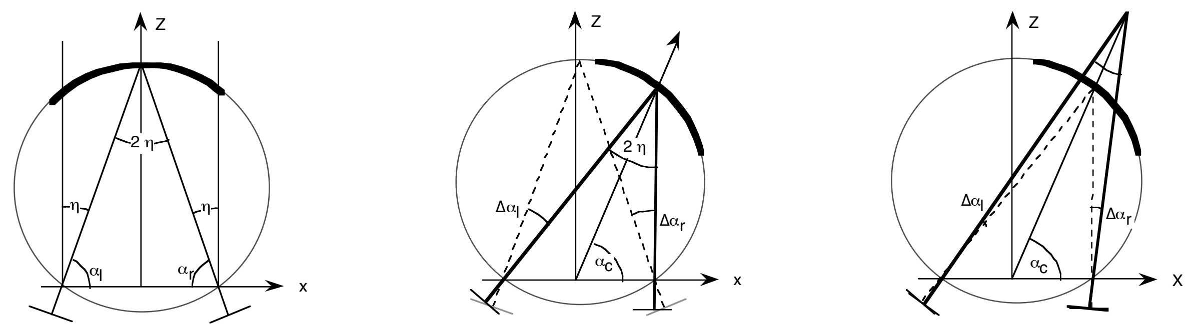
The horopter is controlled by the Superior Colliculus, and can move about the scene in incredibly rapid movements (eye scans). Scanning the horopteur allows the cortex to build up a composite model of the external world.



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.

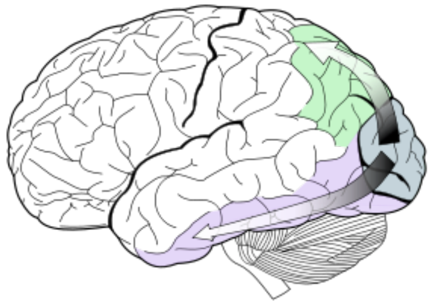


Symmetric Vergence Vergence Version

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

4.4 The Visual Cortex

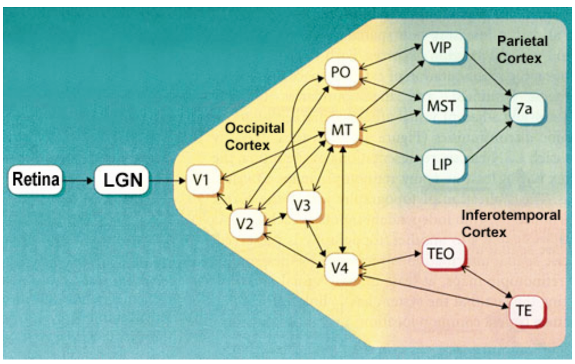
Retinal maps are relayed through the LGN to the primary visual cortex, where they propagate through the Dorsal and Lateral Visual pathways.

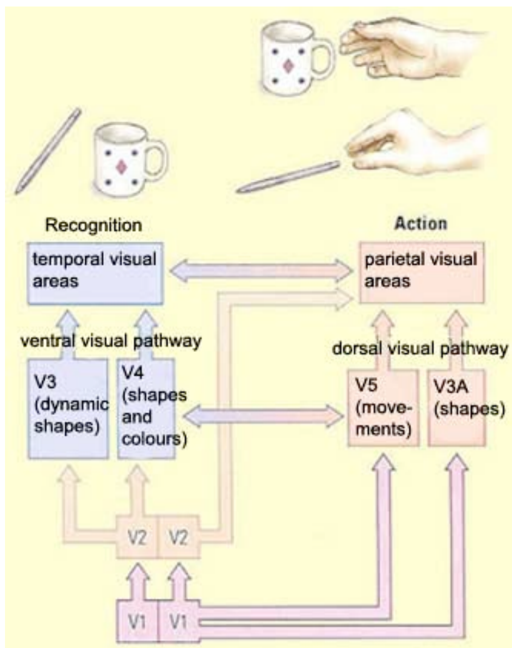


Dorsal visual pathway (green) is the "action pathway".
It controls motor actions. Most of the processing is unconscious.
It makes use of spatial organization (relative 3D position), including depth and direction information from the Superior Colliculus.

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.