

Vision par Ordinateur

James L. Crowley

M2R IVR

Premier Bimestre 2006/2007

Séance 2

10 octobre 2006

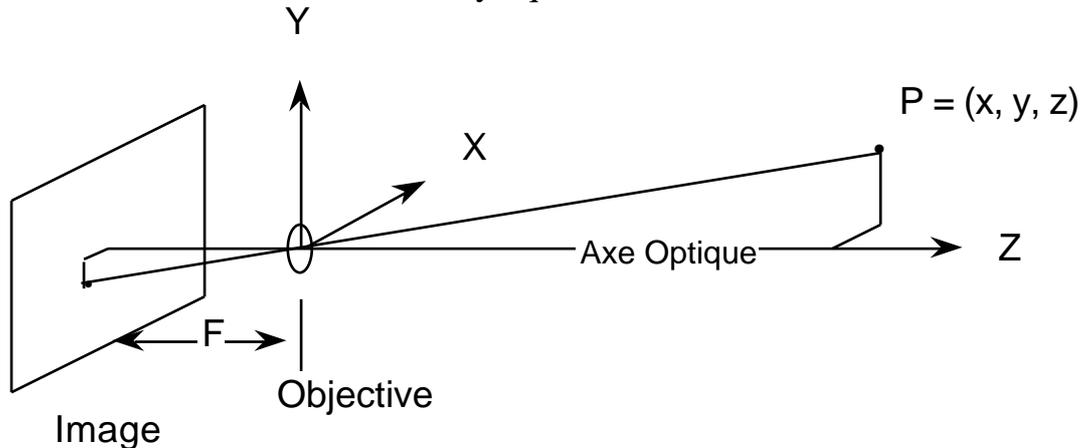
Calibrage, Reflection et Couleur

Plan de la Séance :

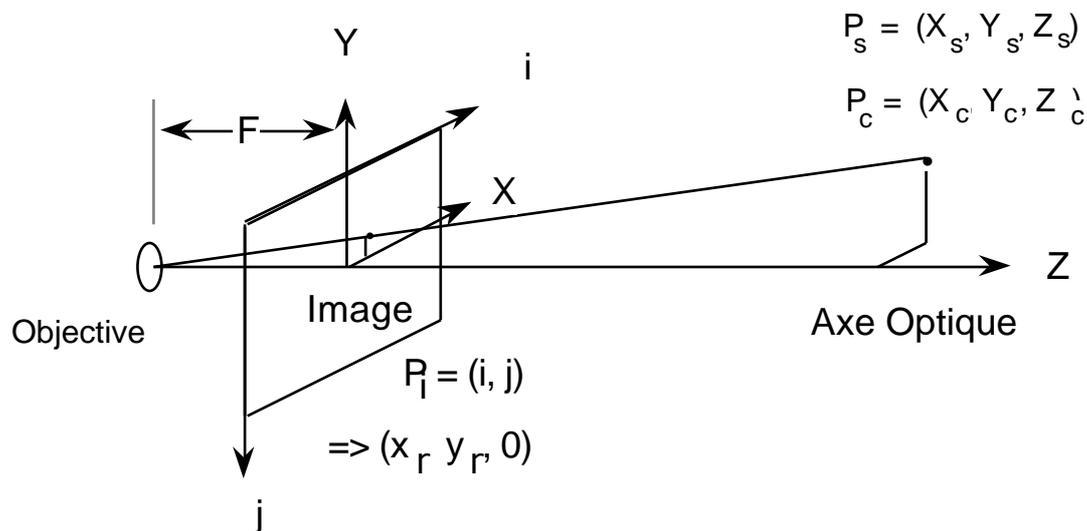
Modèle de la Caméra	2
Les Repères	2
Transformations entres reperes.....	3
La Composition de la Projection Scène - Image	3
La Calibrage.....	4
Dérivation en Notation Tensorielle.....	5
Expression classique en produit croisé de matrices.....	6
Homographie entre un plan et une image.....	8
Couleur et Lumière.....	9
La Spectre	9
La Direction d'un rayon.....	10
Réflexion.....	11
Les fonctions de reflection.....	12
Le Modèle de réflexion di-chromatique.....	13
Les Espace de la Couleur	14
La Perception de la couleur.....	14
Les Espaces de la Couleur	18
l'espace RVB.....	18
l'espace CMY.....	19
L'espace YIQ.....	19
L'espace TLS.....	21
La Distribution de la Couleur d'une Surface	22

Modèle de la Caméra

Modèle de la Caméra Géométrie Physique



Modèle Mathématique: Projection Centrale



Les Repères

Coordonnées de la Scène :

$$\text{Point Scène : } P^s = (x_s, y_s, z_s, 1)^T$$

Coordonnées de la Caméra :

$$\text{Point Caméra : } P^c = (x_c, y_c, z_c, 1)^T$$

$$\text{Point Image : } P^r = (x_r, y_r, 1)^T$$

Coordonnées de l'Image :

$$\text{Point Image : } P^i = (i, j, 1)^T$$

Nota : En coordonnées homogènes, les points sont invariant au multiplication par un constant.

$$(i, j, 1)^T = w \cdot (i, j, 1)^T = (wi, wj, w)^T$$

Transformations entres reperes

La modèle de la caméra est composé d'une composition de transformations.

Transformation entre repère Caméra et repère Scène :

$$P^c = T_s^c P^s$$

Matrice de Projection ${}^r_c P$ du repère Caméra vers le repère Rétine

$$Q^r = P_c^r P^c$$

Transformation entre repère image et repère caméra

$$Q^i = C_r^i P^r$$

La Composition de la Projection Scène - Image

$$Q^i = C_r^i P_c^r T_s^c P^s = M_s^i P^s$$

$$\begin{matrix} w \ i \\ w \ j \\ w \end{matrix} = M_s^i \begin{matrix} x_s \\ y_s \\ z_s \\ 1 \end{matrix}$$

et donc

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

ou bien

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

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

Projection inverse, du repère image vers le repère scène :

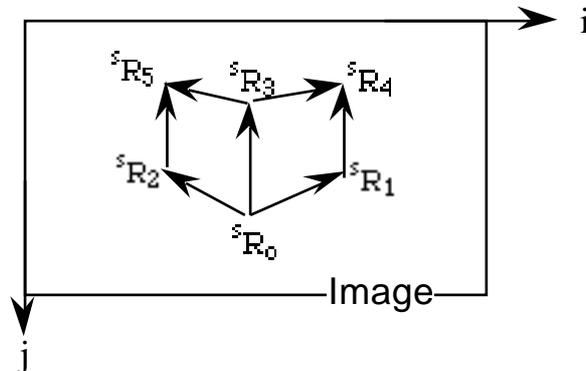
$$P^s = T_c^s P_c^r C_r^i Q^i$$

Cependant, l'inversion de la transformation perspective P_c^r implique la connaissance de la profondeur, z_c , pour chaque pixel.

La Calibrage

Comment obtenir M_s^i ? Par calibrage.

On définit un mire composé de points pour lesquels les positions sont connues R_k^s .



La matrice M_s^i est composé de $3 \times 4 = 12$ coefficients. Cependant, M_s^i est homogène, avec rang $12 - 1 = 11$.

On détermine une correspondance entre les points dans le scènes (R_k^s) et leurs images (P_k^i). Chaque point donne deux équations pour les 11 inconnus.

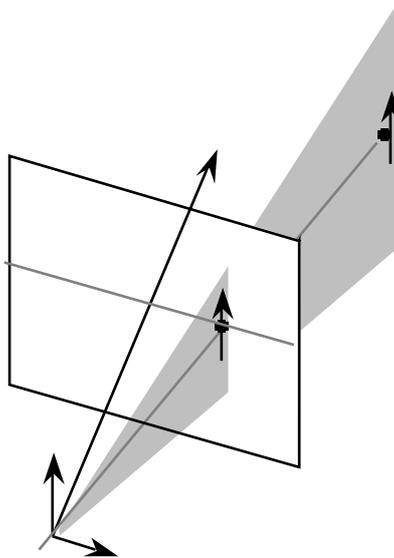
Il faut au moins $5 \frac{1}{2}$ correspondances pour les 11 inconnus.

Pour chaque point de calibrage R_k^s et sa projection P_k^s , on peut écrire :

$$i_k = \frac{w_k i_k}{w_k} = \frac{M_s^1 \cdot R_k^s}{M_s^3 \cdot R_k^s} \quad j_k = \frac{w_k j_k}{w_k} = \frac{M_s^2 \cdot R_k^s}{M_s^3 \cdot R_k^s}$$

Avec lequel on peut déterminer 2 équations pour les 11 inconnus

$$(M_s^1 \cdot R_k^s) - i_k (M_s^3 \cdot R_k^s) = 0 \quad (M_s^2 \cdot R_k^s) - j_k (M_s^3 \cdot R_k^s) = 0$$



L'équation $(M_s^1 \cdot R_k^s) - i_k (M_s^3 \cdot R_k^s) = 0$ est une plan passant par l'origine de la caméra est la colone $i=i_k$.

L'équation $(M_s^2 \cdot R_k^s) - j_k (M_s^3 \cdot R_k^s) = 0$ est une plan passant par l'origine de la caméra est la ligne $j=j_k$.

Dérivation en Notation Tensorielle

Pour N points (non-coplanaires) on peut écrire un système de 2N équations :

$$A M_s^i = 0.$$

Avec notation tensorielle, les équations de la calibrage sont explicites.

soit $P^i = \begin{matrix} w_i \\ w_j \\ w \end{matrix}$ On écrit : $P^i = M_s^i R^s$

Avec K points de la scène R_k^s et leur correspondance de l'image P_k^i on peut écrire

$$P_k^i = M_s^i R_k^s$$

On ne connaît pas P_k^i mais $i w = P_k^1/P_k^3$ et $j w = P_k^2/P_k^3$. Donc Pour chaque point k, il y deux equations independants, $i=1, 2$.

$$\begin{matrix} p^1 \\ p^2 \\ p^3 \end{matrix} = \begin{matrix} w_i \\ w_j \\ w \end{matrix} \quad \text{donc} \quad \begin{matrix} i \\ j \\ 1 \end{matrix} = \begin{matrix} p^1/p^3 \\ p^2/p^3 \\ 1 \end{matrix}$$

et $P_k^3 = M_s^3 R_k^3$

On forme la matrice A for $n=1, \dots, N$ points avec $N > 5$ points (i_n, j_n, R_n^s)
(Le rank de A est 11.)

$$\mathbf{A} = \begin{array}{cccccccccccc}
 R_1^1 & R_1^2 & R_1^3 & 1 & 0 & 0 & 0 & 0 & -i_1 R_1^1 & -i_1 R_1^2 & -i_1 R_1^3 & -i_1 \\
 0 & 0 & 0 & 0 & R_1^1 & R_1^2 & R_1^3 & 1 & -j_1 R_1^1 & -j_1 R_1^2 & -j_1 R_1^3 & -j_1 \\
 \dots & & & & & & & & & & & \\
 R_N^1 & R_N^2 & R_N^3 & 1 & 0 & 0 & 0 & 0 & -i_N R_N^1 & -i_N R_N^2 & -i_N R_N^3 & -i_N \\
 0 & 0 & 0 & 0 & R_N^1 & R_N^2 & R_N^3 & 1 & -j_N R_N^1 & -j_N R_N^2 & -j_N R_N^3 & -j_N
 \end{array}$$

Le problème est de minimiser un critère :

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

On utilise les multiplieurs de Lagrange afin d'obtenir le \mathbf{M}_s^i qui minimise \mathbf{C}

Exemple : Soit un cube dont les coins sont détectés à :

$$\begin{array}{lll}
 P_o^L = (101, 221) & P_1^L = (144, 181) & P_2^L = (22, 196) \\
 P_3^L = (105, 88) & P_4^L = (145, 59) & P_5^L = (23, 67)
 \end{array}$$

Par moindre de carré on obtient :

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

Homographie entre un plan et une image

La projection d'un plan vers un autre plan est une transformation projective entre deux plans. Cette transformation s'appelle une "homographie".
L'homographie est bijective.

Elle est bijective est facile à estimer et facile de rectifier.

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

En notation "classique".

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

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

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

En notation "tensorielle"

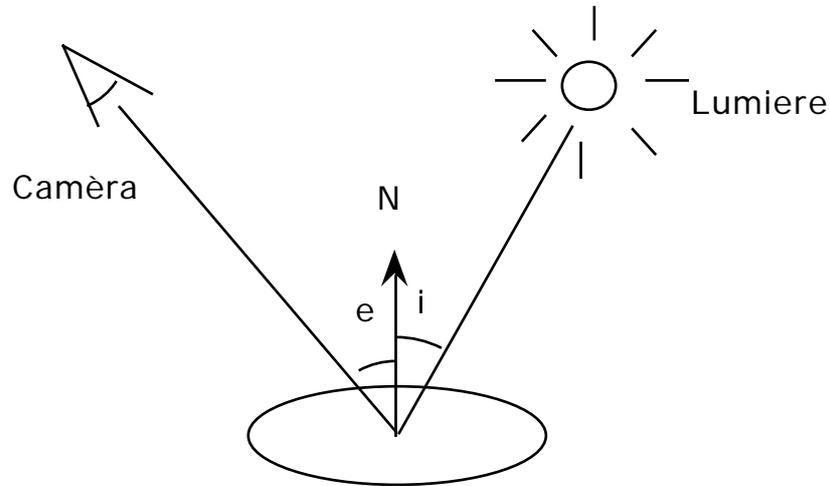
$$Q^B = \mathbf{H}_A^B P^A$$

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

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

Pour calibrer il suffit de noter la position des coins de l'image A dans le repère B, et de formuler 8 équations pour les 8 inconnus du \mathbf{H}_A^B .

Couleur et Lumière



Angles entre :

- e : Caméra et Normale
- i : Rayon incident (entre source et normale)
- g : Caméra et Source

La Spectre

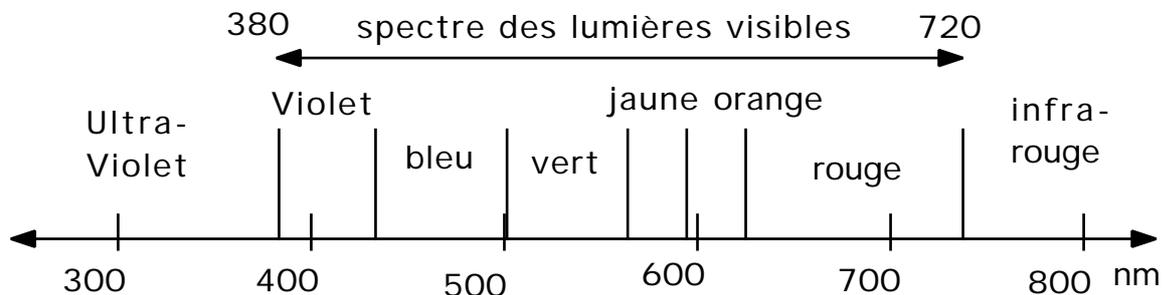
Une source lumière est caractérisé par une spectre :

Il s'agit de l'amplitude (quantité de photons en moyenne) pour chaque

fréquence. La fréquence est décrit par sa longueur d'onde $= \frac{1}{f}$

Les hommes peuvent percevoir les photons entre 380 nm et 720 nm.

Les caméras disposant d'une rétine en silicium peuvent percevoir d'une spectre de photons plus large.

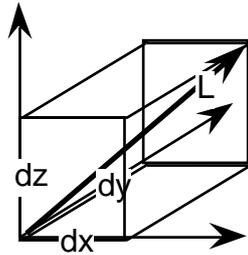


La Direction d'un rayon

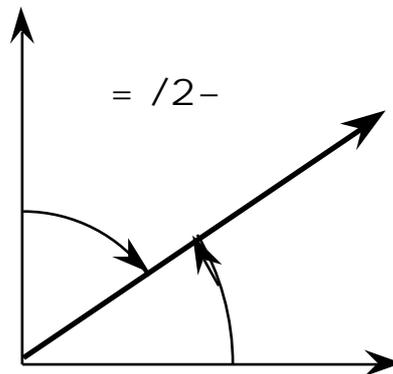
La direction d'une rayone de lumière est définit par une vecteur directeur composé de cosines d'angles.

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

Les cosines de la direction sont la projection de la rayon sur 3 axes, normalisé à une longuer de unité.



Par exemple, en 2D

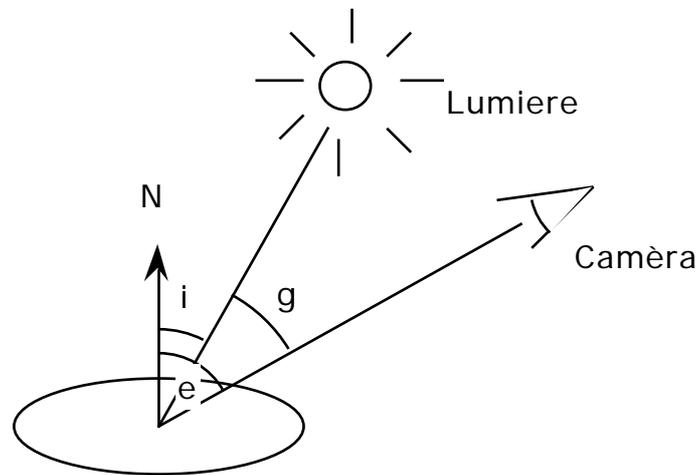


$$D = \begin{pmatrix} x/L \\ y/L \end{pmatrix} = \begin{pmatrix} \cos(\alpha) \\ \cos(\beta) \end{pmatrix} = \begin{pmatrix} \cos(\alpha) \\ \cos(\pi/2 - \alpha) \end{pmatrix} = \begin{pmatrix} \cos(\alpha) \\ \sin(\alpha) \end{pmatrix}$$

Réflexion

L'Albédo d'une Surface

Albédo : propriété de réflectance d'une surface



Angles entre :

e : Caméra et Normale

i : Rayon incident (entre source et normale)

g : Caméra et Source

Les angles sont mesuré en degrés ou en rayons (sur).

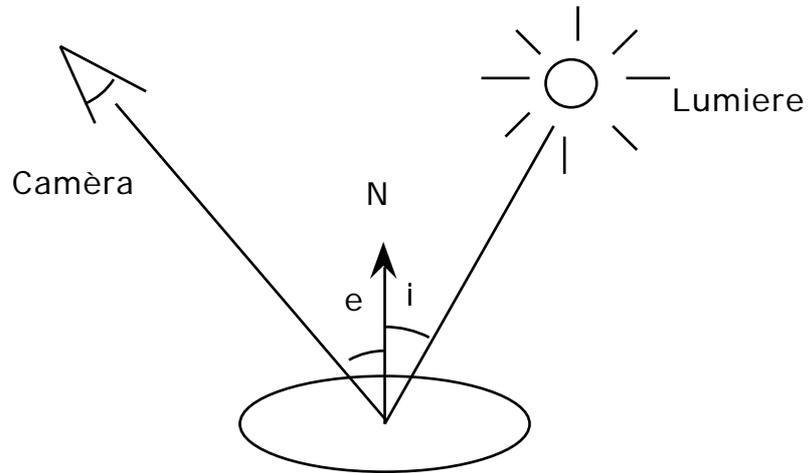
Carte de Réflectance : Ratio

$$R(i, e, g) = \frac{\text{Radiance}}{\text{Irradiance}} = \frac{\text{lumière émise}}{\text{lumière reçue}}$$

Les fonctions de réflexion

Réflexion spéculaire (la cas d'un miroir)

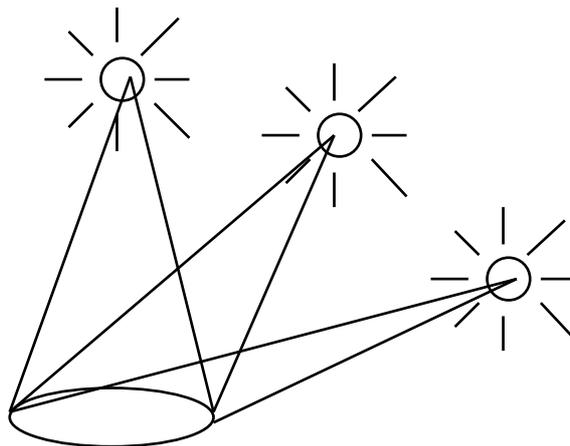
$$R(i, e, g, \lambda) = \begin{cases} 1 & \text{Si } i=e \text{ et } i+e = g \\ 0 & \text{Sinon} \end{cases}$$



v

Réflexion Lambertienne (exemples : neige, papier)

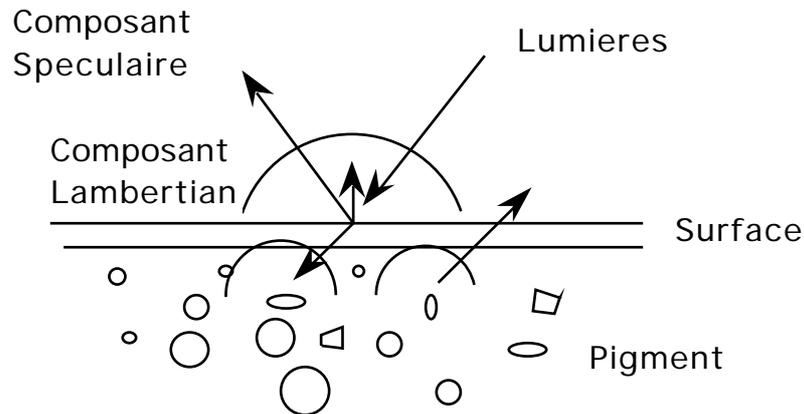
$$R(i, e, g, \lambda) = P(\lambda) \cos(i) \quad (P(\lambda) - \text{Spectre du pigment})$$



Le Modèle de réflexion di-chromatique

L'albédo d'un objet non-métallique peut être approximé par la composition d'une réflexion "spéculaire" et d'une réflexion "lambertienne".

$$R(i, e, g, \lambda) = \alpha_s R_s(i, e, g, \lambda) + (1 - \alpha_s) R_l(i, e, g, \lambda)$$



La composante "spéculaire" a le même spectre que la source de la lumière. $S(\lambda)$.

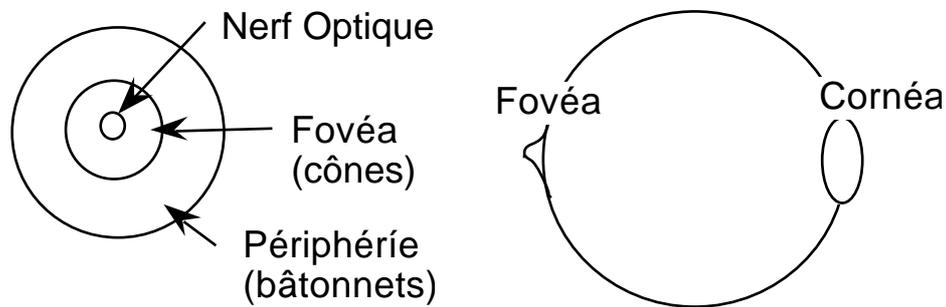
La composante "lambertienne" est "filtrée" par les pigments de la matière $S(\lambda) P(\lambda)$

La composant "luminant" est déterminé par l'orientation de la surface.

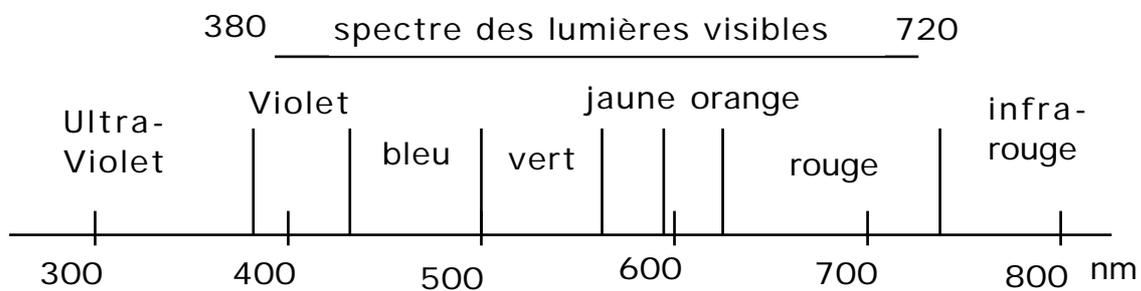
La composant "chrominant" est déterminé par la composition de la spectre de la source et le spectre d'absorption des pigments de la surfaces.

Les Espace de la Couleur

La Perception de la couleur



Spectre de la lumière visible pour l'homme :



Le système humain est composé de bâtonnets et de cônes.

Les bâtonnets sont responsables de la perception achromatique et des lumières atténuées.

Ils sont formés d'une pigmentation, la rhodopsine, sensible à tout le spectre, avec un sensibilité maximum de 510 nm (vert).

Ils sont dans la périphérie de la rétine, et donnent une perception de plus faible résolution.

Les cônes sont responsables de notre vision chromatique.

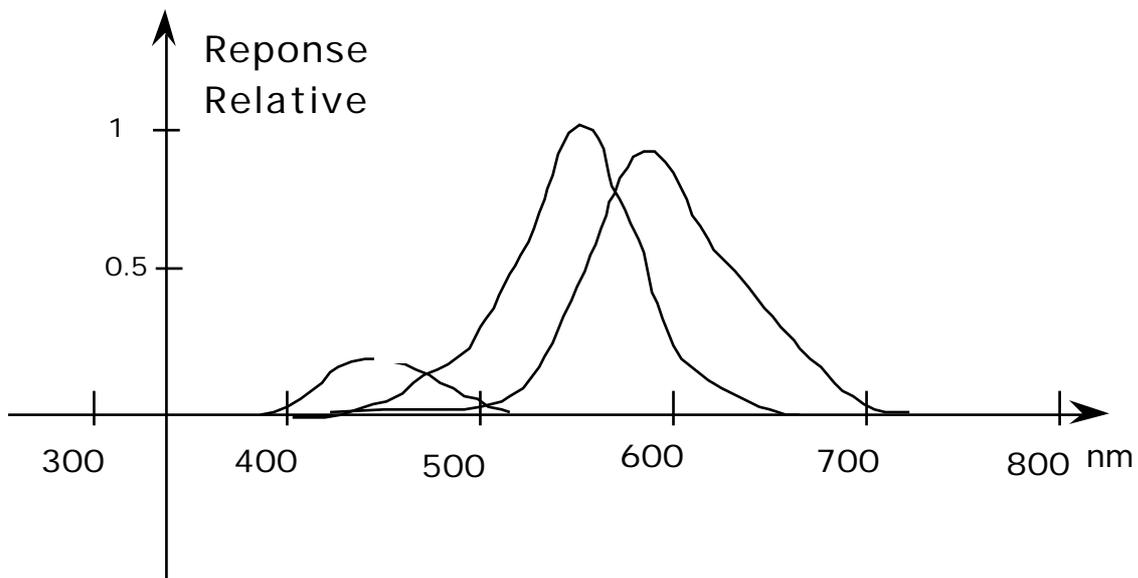
Ils sont concentrés dans une zone appelée la fovéa.

Il y a trois sortes de cônes, distingués par leurs pigments :

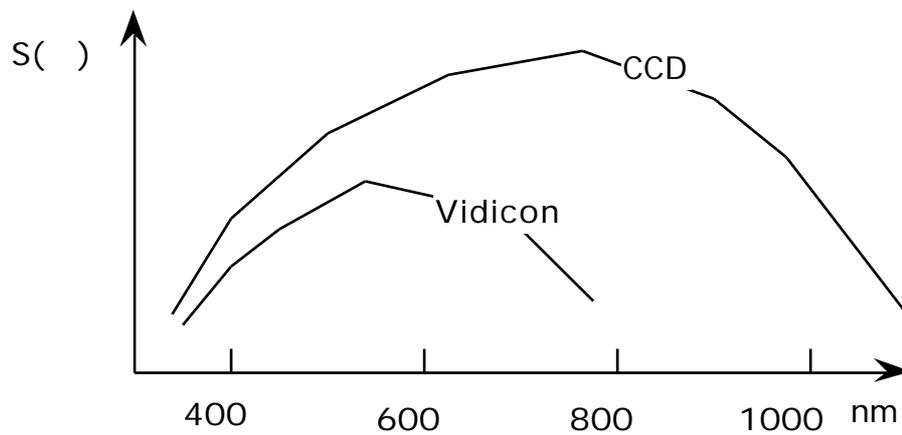
cyanolabe (445 nm),

chlorolabe (535 nm), et

erythrolabe (570 nm)



La transformation entre stimuli de cônes et perception de couleur est un phénomène perceptuel qui n'est pas encore bien modélisé.

La sensibilité des caméras

Intensité de lumière perçue :

elle dépend :

du spectre de la source : $S(\lambda)$

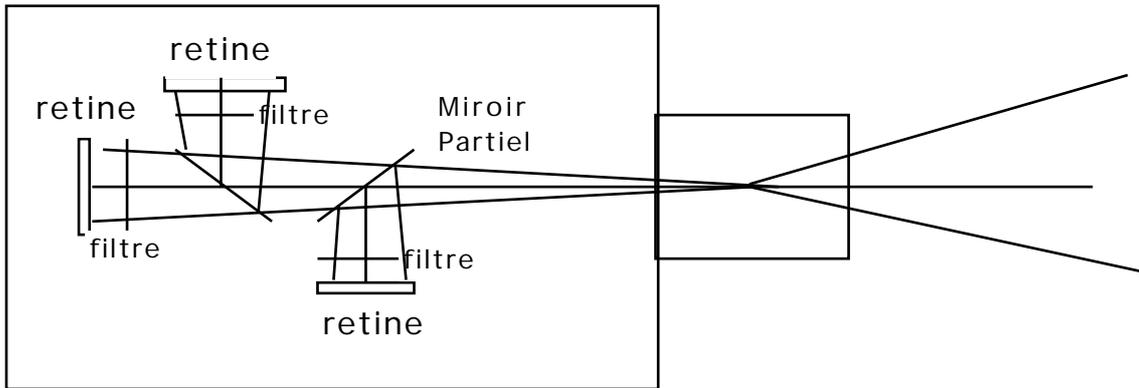
de la réflectance d'un point i, j de la scène vue : $R(i, e, g, \lambda)$

et du spectre de réceptivité de la caméra, $C(\lambda)$

$$p(x, y) = p_0 \int R(i, e, g, \lambda) S(\lambda) C(\lambda) d\lambda$$

p_0 est le Gain. $R(i, e, g, \lambda)$ est le cône "vue" par le pixel x, y .

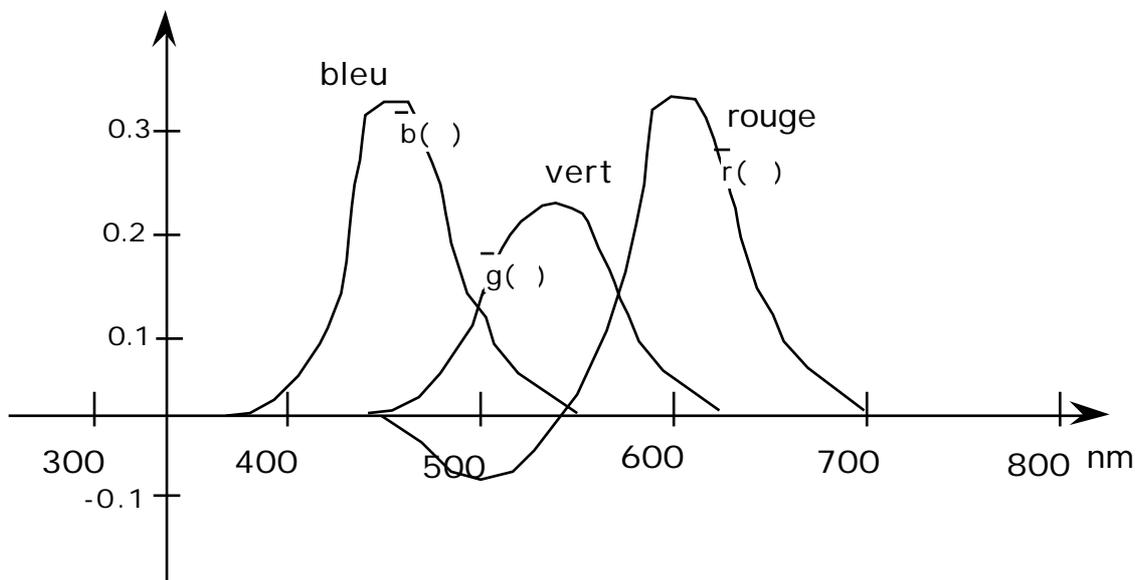
Caméra couleur "classique" :



Les filtres suivent un standard défini par la IIC (International Illumination Commission).

Ils sont des fonctions de λ : $b(\lambda)$, $g(\lambda)$, et $r(\lambda)$.

Note : ils sont composés d'un autre ensemble de fonctions, $x(\lambda)$, $y(\lambda)$, $z(\lambda)$. Ceci permet la valeur négative de $r(\lambda)$.



Un pixel "couleur" est défini par :

$$P(x,y) = \begin{pmatrix} R(x,y) \\ V(x,y) \\ B(x,y) \end{pmatrix}$$

où :

$$R(x,y) = r_0 \int_0^d R(i, e, g, \lambda) S(\lambda) C(\lambda) r(\lambda) d\lambda$$

$$V(x,y) = g_0 \int_0^d R(i, e, g, \lambda) S(\lambda) C(\lambda) g(\lambda) d\lambda$$

$$B(x,y) = b_0 \int_0^d R(i, e, g, \lambda) S(\lambda) C(\lambda) d\lambda$$

ou bien :

$$R = r_0 \int_0^d R(i, e, g, \lambda) S(\lambda) C(\lambda) r(\lambda) d\lambda$$

$$V = g_0 \int_0^d R(i, e, g, \lambda) S(\lambda) C(\lambda) g(\lambda) d\lambda$$

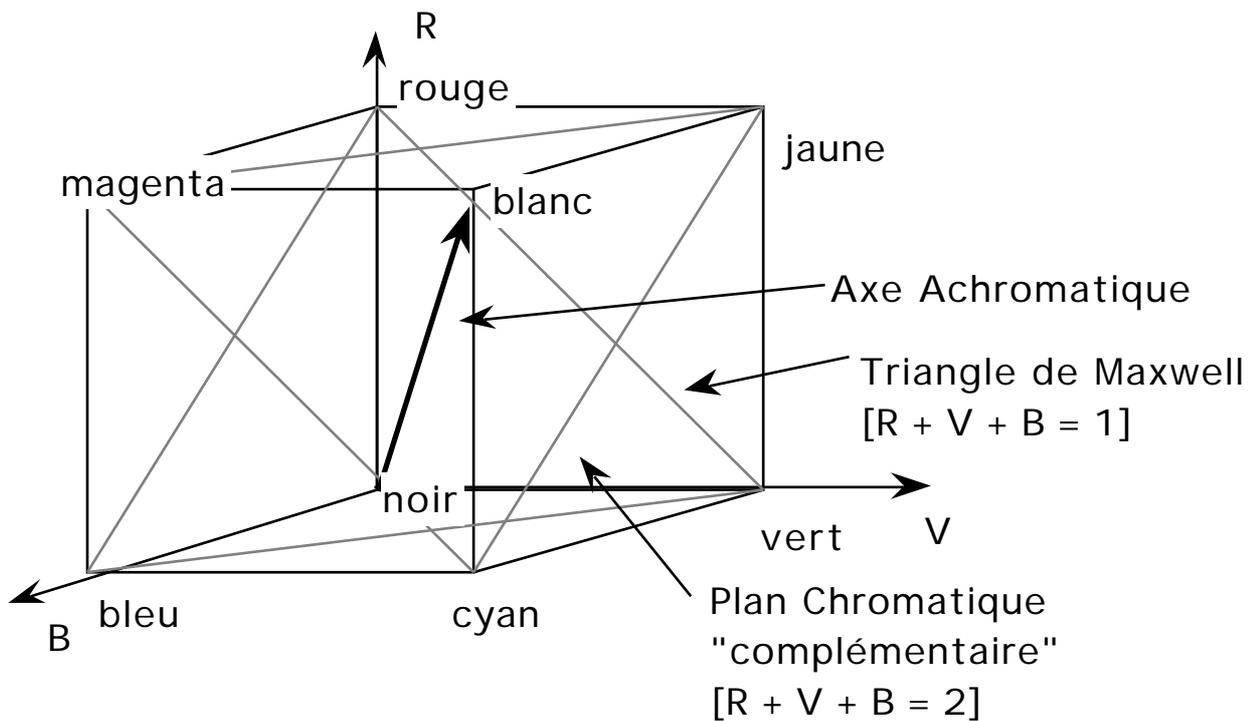
$$B = b_0 \int_0^d R(i, e, g, \lambda) S(\lambda) C(\lambda) b(\lambda) d\lambda$$

Donc, la couleur est une projection de la spectre sur les spectres des récepteurs. On ne peut pas percevoir la spectre d'une manière directe.

Les Espaces de la Couleur

l'espace RVB

Les signaux d'une caméra vidéo nous donnent un espace RVB. RVB est un espace ADDITIF.



L'axe achromatique est $R = V = B$

L'espace CMY

Une page est imprimé avec les pigments.
Les pigments sont "soustractifs". Ils absorbe les photons.

On utilise l'espace : CMY : "Cyan-Magenta-Yellow".

$$\begin{array}{l} C \\ M \\ Y \end{array} = \begin{array}{l} R_{\max} \\ V_{\max} \\ B_{\max} \end{array} - \begin{array}{l} R \\ V \\ B \end{array}$$

CMYK est une variation de CMY dans lequel on ajoute du noir (k).

L'espace YIQ

Le norme TV NTSC des États-Unis utilise l'encodage Y I Q :

$$\begin{array}{l} Y \\ I \\ Q \end{array} = \begin{array}{lll} 0.30 & 0.59 & 0.11 \\ 0.60 & -0.28 & -0.32 \\ 0.21 & -0.52 & -0.31 \end{array} \begin{array}{l} R \\ G \\ B \end{array}$$

Approximation de :

Y = Luminance

I = rouge - cyan

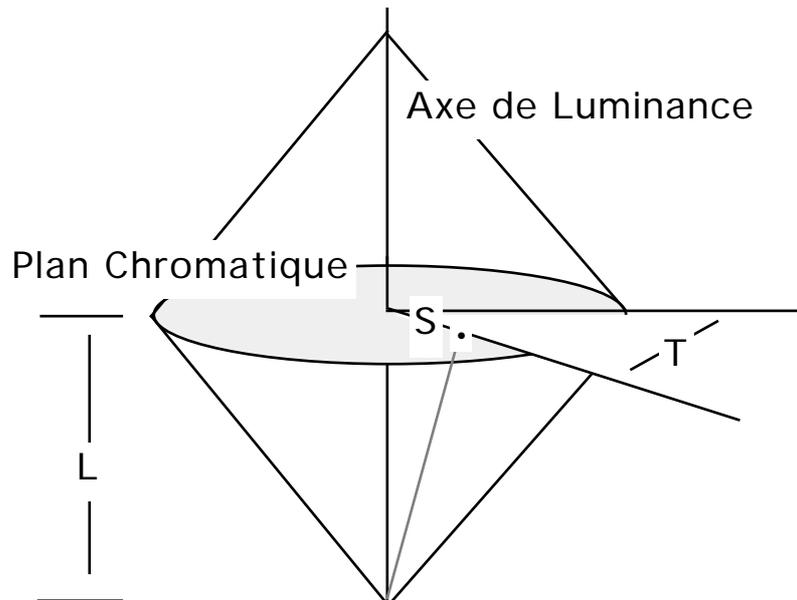
Q = magenta - vert

Les téléviseurs noir et blanc n'affiche que Y.

L'espace TLS

(en anglais Hue Luminance Saturation - HLS)

Pour mieux modéliser la perception de la couleur, nous pouvons transformer l'espace RVB dans l'espace TLS, composé de teinte, luminance, et saturation :



luminance : $L = (R + V + B)$

Normalisation des couleurs :

$$r = R / L$$

$$v = V / L$$

$$b = B / L$$

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

teinte :
$$x = \text{Cos}^{-1} \frac{\frac{1}{2} [(R-V) + (R-B)]}{\sqrt{(R-V)^2 + (R-B)(V-B)}}$$

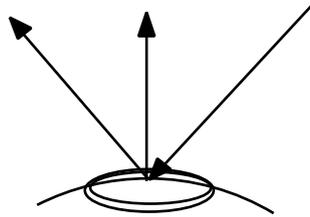
puis si $b < v$ alors $T = x$
 si $b > v$ alors $T = 2 - x$

La Distribution de la Couleur d'une Surface

Qu'est-ce que la couleur d'un reflet ?

Réponse : la couleur de la source.

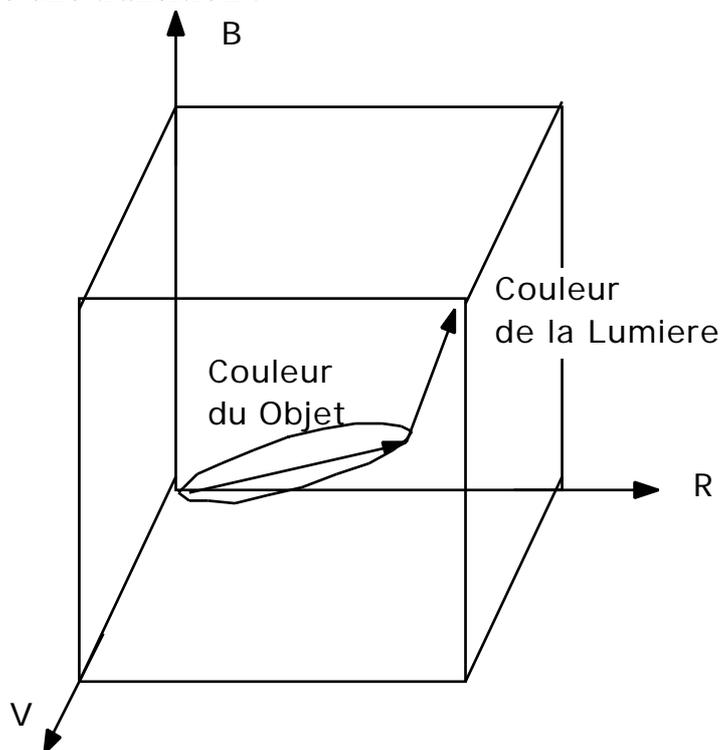
Qu'est-ce qui fait la frontière d'un reflet ?



Le corps d'une surface a la couleur de ses pigments.

Le reflet a la couleur de la source.

La frontière donne une transition :



On peut détecter les deux axes à partir d'un histogramme de couleur.

Un histogramme de couleur est un tableau listant la fréquence d'occurrence pour chaque triple (R, G, B) dans l'image.

On alloue un tableau 3D de taille N_h (exemple $32 \times 32 \times 32 = 32 \text{ K}$ cellules) :

$h(R, V, B)$. Pour chaque pixel $C = c(i, j) = \begin{matrix} R(i, j) \\ V(i, j) \\ B(i, j) \end{matrix}$, on incrémente la cellule de

l'historgramme qui correspond à (R, V, B)

$$h(\mathbf{R}, \mathbf{V}, \mathbf{B}) := h(\mathbf{R}, \mathbf{V}, \mathbf{B}) + 1$$