

Espaces couleur adaptés au traitement d'image

Jean Serra

Centre de Morphologie Mathématique - Ecole des Mines de Paris
35, rue Saint-Honoré, 77305 Fontainebleau, France
serra@cmm.ensmp.fr

Abstract

Le traitement des images en couleur est devenu un important domaine d'étude et d'applications. La représentation de l'espace couleur RGB à l'aide de coordonnées polaires peut parfois simplifier les choses en révélant des caractéristiques des images peu visibles en coordonnées cartésiennes. La littérature décrit plusieurs espaces de ce type (HSV, HLS...), mais il ont en général été conçus pour des applications graphiques et ne conviennent guère aux traitements quantitatifs.

Après avoir décrit leurs faiblesses, nous proposons trois conditions à remplir par toute représentation des couleurs qui se veut quantitative; à savoir que les paramètres d'intensité (brillance et saturation) soient des normes, qu'elles soient indépendantes, et que les coordonnées polaires se calculent à partir de RGB de manière réversible.

Nous étudions ensuite trois représentations qui vérifient ces contraintes. La norme euclidienne L_2 tout d'abord, puis la norme plus "morphologique" L_1 et enfin une amélioration du système HLS où la saturation est remplacée par la norme max-min.

1 Introduction

Dans le domaine de la couleur, plus d'une vingtaine de représentations électroniques co-existent, avec trois ou quatre valeurs de base selon le cas, comme par exemple (LAB), (HSV), (HLS), (HIS), (LST), (CMYK) etc.... Ces représentations sont le plus souvent imposées par telle ou telle technologie, ce qui conduit d'ailleurs à les multiplier. Ainsi, le RGB des ordinateurs et des tubes de

télévision, qui n'est pas bien adapté à la transmission hertzienne, se voit remplacé par le triplet YUV (une luminance et deux chrominances) pour les standards de télévision européens, ou YIQ si l'on habite aux Etats-Unis. Mais comme d'autre part, la perception humaine est moins sensible aux modes cartésiens comme RVB ou YUV qu'aux représentations en coordonnées polaires, de type "brillance, saturation, teinte", ce sont ces dernières que l'on propose à l'utilisateur pour faire ses réglages (la saturation). Au total, on utilise finalement, dans les postes de télévision grand public, trois représentations différentes à divers moments de la chaîne de traitement !

Or lorsque nous nous proposons d'effectuer des traitements sur les images électroniques de couleur, de telles contraintes technologiques n'ont plus de raison d'être. On peut soit garder le triplet RGB, parce qu'il est maintenant devenu un standard informatique, soit le remplacer par tout autre système qui lui corresponde de manière réversible, dès que cet autre système est mieux adapté à ce que l'on veut faire. Par rapport à un temps de calcul moyen en traitement d'image, l'étape qui consiste à quitter RGB, ou à y revenir, reste d'une durée le plus souvent négligeable.

Dans cet exposé, nous partons donc de la donne de RGB considérée comme première, c'est à dire que nous ne nous interrogeons pas sur sa pertinence à représenter l'espace physique, avec ses illuminants, ses objets, leurs réflectances et leurs absorptions. Nos questions sont au contraire tournées vers l'aval: si nous voulons remplacer RGB par un triplet de type (brillance, saturation, teinte) beaucoup plus proche de la vision cérébrale, que

c

choisir ? Quels sont nos buts ? Quels impératifs, non technologiques, mais logiques ceux-là, la quantification impose-t-elle ? Les deux principales représentations polaires existantes donnent-elles satisfaction sur ces points ?

C'est ce que nous allons voir, en reprenant partiellement l'étude [1], qui elle même prolonge le chapitre quatre de la thèse [3] d'Allan Hanbury.

2 Quelques incohérences

Les deux triplets classiques (brillance, saturation, teinte) se nomment HSV et HLS et leur expression, inversible, en fonction de RGB s'apparente à un changement de coordonnées "cartésien → polaire". Le système HSV est défini comme suit

$$\left\{ \begin{array}{l} v = \text{Max}(r, g, b) \\ s_c = \frac{\text{Max}(r, g, b) - \text{Min}(r, g, b)}{\text{Max}(r, g, b)} \\ h = \frac{g-b}{(\text{Max} - \text{Min})} \quad \text{si } r = \text{Max} \\ \quad = \frac{b-r}{(\text{Max} - \text{Min})} + 2 \quad \text{si } g = \text{Max} \\ \quad = \frac{r-g}{(\text{Max} - \text{Min})} + 4 \quad \text{si } b = \text{Max} \end{array} \right. \quad (1)$$

Dans toutes ces notations, nous conservons les majuscules pour désigner les axes (RGB, HSV, etc..) et les minuscules pour les coordonnées des points courants. Dans le système (1), r, g et b varient de zéro à un, et il en va de même pour v et s_c , alors que h varie de -1 à cinq. Mais h , approximation d'un angle, est une coordonnée curviligne du cercle unité, donc définie à 2π près. Il faut décaler la valeur de h donnée par (1), et la multiplier comme il se doit pour la placer dans son domaine de variation, comme par exemple (0-255). Le zéro de h dans le système (1) correspond à la teinte du vecteur purement rouge.

On retrouve pour le système (1) les indéterminations classiques des coordonnées polaires: quand v vaut zéro, alors s est indéterminé, et quand s vaut zéro, alors h est à son tour indéterminé.

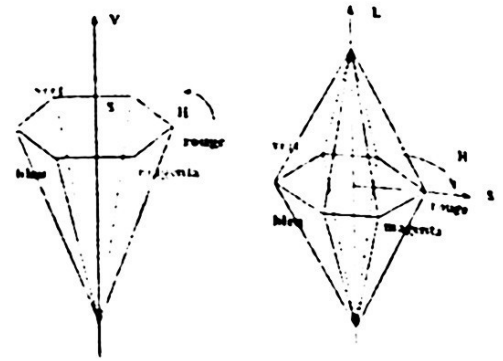


Figure 1: Système HSV (à gauche) et HLS (à droite).

L'espace HLS dérive du précédent, mais il est plus courant: on le rencontre dans les logiciels grand public comme *paint shop pro*. Il se définit par les équations suivantes:

$$\left\{ \begin{array}{l} l = \frac{\text{Max}(r, g, b) + \text{Min}(r, g, b)}{2} \\ s_c = \frac{\text{Max} - \text{Min}}{\text{Max} + \text{Min}} \quad \text{si } l \leq 0.5 \\ s_c = \frac{\text{Max} - \text{Min}}{2 - \text{Max} - \text{Min}} \quad \text{si } l \geq 0.5 \\ h = \text{identique à celui de HSV} \end{array} \right. \quad (2)$$

On comprend aisément l'interprétation géométrique des représentations (1) et (2). Si l'on prend pour axe vertical la diagonale du cube unité RGB, celui-ci ressemble, grosso modo, à deux pyramides hexagonales accolées par la base. C'est cette forme que simulent les deux structures HSV et HLS de la Fig.1.

Or, avec les pyramides de la Fig.1, beaucoup de coordonnées cessent d'être valides. Quand la luminance tend vers zéro sur l'axe V, on ne peut pas trouver grande saturation, ni sur l'axe L pour les valeurs proches du zéro ou proches du un. Le moindre calcul risque conduire à des résultats qui sortent du domaine de définition des variables, et l'on doit vérifier la validité toutes les opérations que l'on effectue. C'est pour éviter d'avoir à le faire en permanence que l'on a transformé le cône en cylindre, sans trop prendre garde aux risques



Figure 2: Les quatre images sont des coupes de l'espace couleur par le plan contenant l'axe des gris (diagonale du cube RGB) et la droite du plan chromatique qui oppose le rouge (0°) au vert (180°).

qui allaient en découler pour l'analyse d'images. De fait, les jolis dessins de la figure 1, que l'on trouve dans tous les bons manuels, ne correspondent pas du tout aux systèmes qu'ils sont censés illustrer. Ces jolis dessins représenteraient les équations si l'on n'y avait pas divisé la saturation par la luminance. En ne le faisant pas, on a implicitement remplacé les cônes de la figure 1 par des cylindres qui ne leur sont pas du tout équivalents.

La figure 2 montre deux coupes verticales sur les volumes HSV (Fig. 2a) et HLS (Fig. 2c) de la Fig. 1. On a placé à côté les mêmes coupes telles que les systèmes d'équations (1) et (2) la reconstruisent (Fig. 2b et 2d respectivement). Lorsqu'on place un seuil sur la saturation cylindrique selon les deux verticales claires de la Fig. 2d, on traverse des zones très colorées vers le centre de l'image, mais qui ne le sont plus en haut et en bas. Pour réaliser un seuillage qui représente un degré de colorisation constant, il faut se placer en saturation conique. C'est ce qui est fait en Fig. 2a pour HSV, de manière visuellement satisfaisante. Mais alors le même seuil s'exprime en mode cylindrique comme une *hyperbole*, conformément aux formules (1) !

Les formules de passage des saturation cylindrique s_c des systèmes (1) et (2) vers leurs version coniques s ne sont pourtant pas très compliquées; les voici:

$$\begin{aligned} s &= s_c \cdot v & (HSV) \\ s &= s_c \left[1 - 2 \left| \frac{1}{2} - l \right| \right] & (HLS). \end{aligned} \quad (3)$$



Figure 3: Un tableau de J. Miro, avec inversion du bas, et trois variantes de sa saturation

Au plan de la signification physique, le passage en cylindrique revient à donner de l'importance à des aspects de l'image qui ne le méritent pas et à en ignorer d'autres plus fondamentaux. On s'en rendra compte en prenant le tableau de Joan Miro, "le chanteur", et en en inversant la moitié basse, c'est à dire en y remplaçant les rouges, les verts et les bleus par leurs compléments à 255 (Fig. 3). Comme une symétrie par rapport au centre du cube RGB ne change pas la distance à sa diagonale, la saturation ne devrait pas changer. En HLS cette symétrie est respectée, mais pas du tout en HSV (Fig. 3b). D'autre part apparaissent comme blancs des pixels RGB qui valent *presque* [1,1,1] (idem pour les noirs avec [000]). Ces petites variations sont grandement amplifiées dans les saturations HSV et HLS du fait que dans les expressions de s_c de (1) et (2) le dénominateur tend alors vers 0 (Fig. 3b et c). L'expansion cylindrique se paye donc cher, avec toutes ces taches parasites qui heureusement disparaissent dans la saturation max-min, beaucoup plus propre (Fig. 3d).

3 Les trois prérequis

On dominera mieux l'analyse critique qui précède si l'on distingue clairement ce que la notion d'espace vectoriel, celle de *norme* et celle d'*indépendance* apportent à la représentation des images en couleur. Dans ce qui suit, l'espace RGB est toujours modélisé par l'espace euclidien R^3 , avec ses projections, son orthogonalité, etc.. mais nous l'équipons successivement de normes différentes, dont *entres autres* la norme euclidienne.

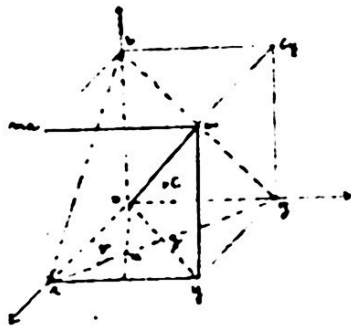


Figure 4: *Cube couleur unité. Les points r, g, b, m, c, y (i.e. rouge, vert, bleu, magenta, cyan, jaune) ont une saturation maximale.*

La notion d'espace vectoriel identifie le point courant $c = (r, g, b)$ au vecteur $\vec{oc}$. Elle définit

- la somme de plusieurs vecteurs comme étant le vecteur de composantes sommes des composantes élémentaires;

- et le produit d'un vecteur par un scalaire en multipliant toutes les composantes de ce vecteur par le scalaire.

A noter que ces deux opérations transforment des vecteurs en vecteurs et non pas des vecteurs en nombres. On montre de plus que tout vecteur admet une décomposition *unique* selon chaque système d'axes. Par exemple, partant du cube unité de coordonnées $0 \leq r \leq 1$, $0 \leq g \leq 1$, $0 \leq b \leq 1$, nous désignons sa diagonale $(0,0,0)$, $(1,1,1)$ comme l'axe *achromatique*. Le plan qui lui est perpendiculaire à l'origine devient alors le *plan chromatique*, qui renferme toute l'information sur la couleur (voir figures 4 et 7). Le point coloré c est alors aussi bien obtenu comme $\vec{oc} = \vec{or} + \vec{og} + \vec{ob}$, que comme la somme $\vec{oc} = \vec{oc_d} + \vec{oc_p}$ où c_d et c_p sont les projections de c sur l'axe achromatique et sur le plan chromatique respectivement.

Peut-on dire alors que les vecteurs c_d et c_p , qui sont orthogonaux, sont indépendants ? La réponse dépend du sens que l'on attribue à l'adjectif "indépendant". S'il porte sur un lien possible entre les deux projections c_d et c_p , alors manifestement celles-ci ne sont pas indépendantes : les points d'intensité lumineuse extrême, très

faible ou très forte, sont peu saturés en couleur. Cela est une simple conséquence de la forme du domaine de variations. Mais l'indépendance peut signifier aussi autre chose, comme par exemple que les paramètres qu'on associe à c_p (saturation, teinte) ne mettent pas en jeu ceux qu'on associe à c_d . Dans ce cas, si deux points colorés *différents* c et c' se projettent sur le même c_p , l'indépendance entraîne qu'ils aient *même saturation et même teinte*. Toute représentation des couleurs adaptée au traitement d'image devra donc vérifier le premier prérequis suivant

1er prérequis : Deux points distincts qui ont la même projection dans le plan chromatique ont les mêmes paramètres chromatiques.

On pourrait aller plus loin et exiger aussi que deux points qui se projettent au même endroit sur l'axe achromatique aient même intensité. Mais cela nous limiterait aux fonctions symétriques de (r, g, b) , excluant les expressions pondérées comme la luminance Y du standard européen de télévision

$$Y = 0,299r + 0,587g + 0,144b \quad (4)$$

où les poids sont proportionnels à la sensibilité de l'œil pour la couleur correspondante.

La seconde idée sur laquelle nous pouvons nous fonder, après celle d'indépendance, est celle de norme. Elle associe à tout vecteur un paramètre, disons α , positif ou nul qui augmente quand le point c s'écarte de l'origine, i.e.

$$\alpha(\lambda c) = \lambda \alpha(c) \quad (\lambda \text{ poids} \geq 0).$$

De plus la norme relie l'addition au sens des vecteurs à celle au sens des nombres par la classique inégalité triangulaire :

$$\alpha(c + c') \leq \alpha(c) + \alpha(c'), \quad (5)$$

qui dit que la norme du vecteur moyen entre c et c' ne saurait être plus grande que la moyenne des normes de c et de c' . Enfin, il est équivalent de dire que le vecteur c est nul ou que sa norme l'est.

$$c = 0 \iff \alpha(c) = 0 \quad (6)$$

Quand cette dernière condition n'est pas remplie, on parle de "semi-norme" seulement. On notera que dans l'inégalité triangulaire (5) les deux symboles "+" n'ont pas le même sens: le premier se rapporte à des vecteurs, le second à des nombres. Il en va de même pour les deux zéros impliqués dans l'équivalence (6).

L'inégalité (5) doit être envisagée comme une manifestation de prudence. Par exemple, deux projections chromatiques grandes mais opposées représentent deux couleurs très saturées, alors que la somme vectorielle de ces deux couleurs est achromatique. Il est donc sage que le paramètre de brillance associé à cette somme ne soit pas supérieur à celui des couleurs de départ. C'est bien ce qui se passe pour la longueur du vecteur c (norme L_2 , ou euclidienne), sa moyenne (norme L_1) ou encore le $\max$ de r, g et b (norme L_∞) et enfin la semi-norme $\delta = \max - \min$. Ces normes ne sont pas toutes aussi pertinentes visuellement ni pour les traitements d'image, et la norme L_∞ , du système HSV, fait relativement peu recette.

Le véritable choix est en fait entre les normes L_1 , L_2 , et $\max - \min$. La seconde conduit à des formules de passage

$$(r, g, b) \rightarrow (\text{luminance, saturation, teinte})$$

qui sont quadratiques, et trigonométriques pour la teinte; aussi ne convient-elle pas très bien au troisième prérequis ci-dessous, plus facile à mettre en oeuvre avec des formules de passage linéaires. En contrepartie, comme la distance associée à la norme L_2 est celle de l'espace euclidien, elle présente les avantages

- vérifier le théorème de Pythagore,
- d'être doublement indépendante vis à vis des projections chromatiques et a-chromatiques,
- et d'être connue de tout le monde.

La norme L_1 a pratiquement déjà fait son apparition dans le traitement des images en couleur, mais sans l'avouer. On la voit par exemple dans [4] et [5] pour l'axe chromatique, mais pas pour le plan achromatique. Elle conduit à une distance moins habituelle que l'euclidienne, mais plus rapide à implémenter et souvent aussi précise. Notons enfin que, s'agissant de variables $r, g, b \geq 0$, toute quantité $\alpha r + \beta g + \gamma b$, avec des poids

$\alpha, \beta, \gamma \geq 0$ est encore une norme L_1 sur l'axe achromatique. Cela conduit au second prérequis suivant

2ème prérequis : Les paramètres d'intensité associés au vecteur couleur courant c et à sa projection c_p doivent être des normes.

En plus des deux prérequis que nous avons dégagés, il convient d'introduire une troisième contrainte, moins fondamentale et suggérée par la pratique. Quand on parcourt la littérature actuelle sur le traitement des images en couleur, on voit que malgré le nombre élevé des espaces couleur déjà existants, les auteurs proposent fréquemment de nouvelles variantes, de nouvelles représentations, confirmant ce que nous notions dans l'introduction plus haut. Ils ne se sentent plus liés par les représentations "officielles" des images couleurs, telles que YUV, YAB, HLS, etc ... Ils partent de l'image RGB, choisissent l'espace existant qui leur paraît le plus adapté pour le problème qu'ils ont à traiter, ou, sinon, ils en construisent un autre (voir par exemple [4] ou [5]). Ce qui est bien normal : ils vont de toutes façons dépenser du temps machine pour leur traitement d'image, et n'hésitent pas à allonger légèrement ce temps pour avoir une représentation couleur qui leur convienne.

En retour, la seule contrainte que l'on soit en droit de demander à ces traitements d'images en couleur est que le système qui a servi pour les traitements soit *réversible*. En fin de traitement on doit pouvoir restituer une image dans le standard (r, g, b) d'où l'on est parti. Cela nous conduit à proposer le dernier prérequis suivant:

3ème prérequis: Tout système de représentation des images en couleur doit être réversible par rapport au standard RGB.

4 Retour aux critiques

Si nous examinons le système HLS à la lumière des deux premiers prérequis, les raisons de fond des critiques présentées dans [1] deviennent très claires. Dans le système HLS, ni la luminance ni la saturation ne sont des normes, et en plus il n'y a pas d'indépendance entre

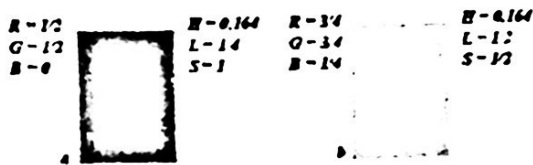


Figure 5: *Dépendance entre luminance et saturation en HLS*

l'axe a-chromatique et le plan chromatique: il semble difficile de faire pire!

Pour montrer l'absence d'indépendance, il suffit de prendre les points $(\frac{1}{2}, \frac{1}{2}, 0)$ et $(\frac{3}{4}, \frac{3}{4}, \frac{1}{4})$, qui se projettent tous deux au même point sur le plan chromatique (cf fig.5), puisque l'on passe de la première couleur (a) à la seconde (b) en ajoutant $R=G=B=1/4$, donc une composante c_d de gris pure. Leur saturations HLS sont données par le rapport $\frac{\max - \min}{\max + \min}$, puisque leurs luminances sont ≤ 0.5 . On trouve pour le premier la valeur 1 et pour le second $1/3$ (ce dernier est désavantagé parce qu'il a une luminance moins basse que l'autre: conséquence du mode cylindrique).

De plus, non seulement la représentation crée une différence artificielle, mais elle ne discrimine pas entre des points qui sont, eux, vraiment différents puisque tous les points de luminance ≤ 0.5 et de $\min = 0$ ont même saturation.

Pour vérifier que la luminance $l = \frac{\max + \min}{2}$ ne vérifie pas l'inégalité triangulaire on peut prendre le couple de points $c = (1/2, 1/2, 0)$ et $c' = (0, 1/2, 1/2)$ dont les luminances HLS, égales, valent $1/4$, alors que celle de $c + c'$ vaut $3/4$ (cf fig.6) ! Que peut valoir une luminance qui invente des sur-brillances sans qu'on le lui demande ? Quand les auteurs la remplacent par la moyenne arithmétique des trois composantes, ils ne donnent pas toujours leurs raisons (cf.[4] ou [5] par exemple) mais on les comprend aisément...

Enfin, la saturation HLS n'est pas non plus une norme car les deux points $c = (1/3, 2/3, 1/3)$ et $c' = (2/3, 1/3, 1/3)$ ont pour commune saturation HLS la valeur $1/3$ alors que celle de leur somme vaut 1 (par

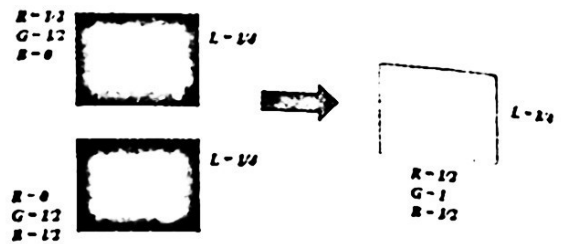


Figure 6: *Comme la luminance HLS n'est pas une norme, elle crée des sur-brillances.*

contre le terme $\max - \min$ reste le même): la saturation de $c + c'$ est hypertrophiée parce que ce point est plus lumineux que c et c' .

Dans ces conditions, que faire ? Deux stratégies sont envisageables. On peut tout d'abord se tourner vers les deux normes L_1 et L_2 qui sont des outils de base de l'espace euclidien. C'est ce que nous allons faire, mais nous reviendrons ensuite vers l'espace HLS pour voir comment l'amender pour le rendre acceptable. On verra que curieusement, il faut associer la quantité $\max - \min$ à la saturation, mais pas à la luminance. Toutefois, avant d'étudier ces trois normes, nous allons brièvement décrire notre espace de travail.

5 L'espace de travail

Nous partons ici de l'hypothèse que pour définir la teinte, les trois couleurs de la base RGB ont autant d'importance, et que par conséquent, la grande diagonale du cube unité est l'axe achromatique. L'information chromatique est alors entièrement portée par le plan chromatique perpendiculaire à cet axe à l'origine. Tout vecteur c du cube unité r, g, b se décompose en la somme vectorielle de sa projection c_d sur la diagonale et c_p sur le plan chromatique

$$c = c_p + c_d \quad (7)$$

et l'unicité de cette décomposition montre que toute l'information chromatique est portée par la composante c_p . Pour une meilleure lisibilité, tous les vecteurs projetés sur le plan chromatique sont affectés de l'indice p .

Ainsi, r_p représente la projection du vecteur rouge pur $(\sqrt{2}, 0, 0)$ et g_p celle du vert pur $(0, \sqrt{2}, 0)$. Les coordonnées de ces vecteurs, exprimées dans la base (r, g, b) sont indiquées en figure 7.

Les symboles marqués de l'indice p représentent les projections orthogonales sur le plan chromatique (Ex r_p est la projection du point r dans l'espace).

Eléments remarquables du plan chromatique
$c_p = \text{proj. de } c = (2r - g - b) / 3,$ $(2g - b - r) / 3, (2b - r - g) / 3$
$r_p = (2/3, -1/3, -1/3)$ $= \text{proj. de } r = (1, 0, 0)$
$g_p = (-1/3, 2/3, -1/3)$ $= \text{proj. de } g = (0, 1, 0)$
$s_p = s_p(\theta)$ $= \text{proj. de } [(\sqrt{2} - x(\theta), x(\theta), \theta)]$
rayon du cercle : $\frac{\sqrt{6}}{3}$
Hexagone = proj. du cube couleur unité
Triangle $o r_p y_p =$
proj. des points du cube où $r \geq g \geq b$

De plus, le demi plan limité par $y_p b_p$ et contenant r_p est la projection du demi espace $r \geq g$, limité par $01y$ et qui contient r . De même, le demi plan limité par $r_p c_p$ et contenant g_p est la projection du demi espace $g \geq b$, limité par $01c_y$ et contenant g . Enfin, les points tels que $r + b - 2g = 0$ forment le plan passant par l'axe achromatique et la droite $r + b = 0$ du plan orb. Ce plan coupe le plan chromatique suivant la parallèle à $r_p b_p$ passant par o .

Equipons le plan chromatique de coordonnées polaires, en prenant le vecteur r_p pour origine des angles. Le sens positif est celui indiqué sur l'angle θ de la fig.7. Tout point (i.e. tout vecteur) c du cube unité $[01] \otimes [01] \otimes [01]$ s'exprime de manière équivalente par ses coordonnées cartésiennes r, g, b ou spatio-polaires $(\|c_p\|, \|c_p\|, \theta)$.

Toutefois, cette équivalence ne conduit pas aux mêmes équations selon que l'on munit l'espace de la norme L_2 ou L_1 . Ainsi, il vient pour la norme L_2

$$\|c\|^2 = \|r\|^2 + \|g\|^2 + \|b\|^2 \quad (8)$$

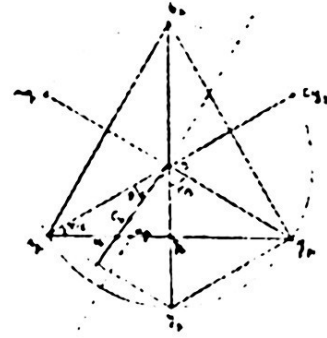


Figure 7: Plan chromatique

alors que pour L_1 , on a

$$|c| = |r| + |b| + |g| \quad (9)$$

Les différences sont suffisamment considérables pour qu'on étudie plus précisément les avantages et les inconvénients des deux approches. Pour l'une et l'autre, nous imposons aux domaines de variations de chaque variable d'être le segment $[01]$. On sera ainsi en conformité avec la technique informatique pour laquelle une image de 3k bits RGB se convertit en 3k bits HLS, et inversement.

6 Cadre de la norme L_2

On peut penser qu'à priori elle est la mieux adaptée, puisqu'elle repose sur le théorème de Pythagore, qui interprète la norme en termes de longueur des vecteurs. De plus, le produit scalaire qui l'accompagne est un outil indispensable pour calculer les angles, ce dont nous avons besoin.

Les formules de passage à partir de l'espace RGB s'établissent facilement. Appelons m_2 et s_2 les normes réduites à $[01]$ de la projection sur l'axe achromatique et sur le plan chromatique, et h_2 l'angle θ , réduit lui aussi. Il vient par application de la relation (8) et de la

fig.7 :

$$m_2 = \frac{1}{\sqrt{3}} [r^2 + g^2 + b^2]^{1/2}$$

$$s_2 = \frac{3}{2} [(2r - g - b)^2 + (2g - b - r)^2 + (2b - r - g)^2]^{1/2}$$

Le calcul de l'angle θ dérive du produit scalaire des vecteurs c_p et r_p de la fig.7, selon la relation établie dans [1]

$$\begin{aligned} h_2 &= \frac{2}{\pi} \theta = \text{Arc cos} \left[\frac{r_p \cdot c_p}{\|r_p\| \cdot \|c_p\|} \right] \\ &= \text{Arc cos} \frac{r - g/2 - b/2}{(r^2 + g^2 + b^2 - rg - rb - bg)^{1/2}} \end{aligned} \quad (11)$$

Formellement parlant, le problème est résolu. Les trois variables m_2 , s_2 et h_2 s'expriment en fonction de r , g et b et varient entre 0 et 1. Mais malgré l'équivalence théorique des deux systèmes, le passage inverse n'est pas simple. C'est sans doute la raison pour laquelle on n'utilise pas les variables m_2 , s_2 et h_2 en pratique, alors qu'elles sont les plus cohérentes avec les traitements d'image qui font intervenir des distances euclidiennes (les filtres médians vectoriels par exemple).

Toutefois, les traitements morphologiques, avec leurs sup et leurs inf semblent a priori mieux adaptés au cadre L_1 , qui s'exprime lui aussi à l'aide d'opérateurs de min et de max.

7 Luminance et saturation en norme L_1

Dans cette section nous gardons le même espace vectoriel que précédemment, avec la décomposition $c = c_p + c_d$ mais en affectant aux vecteurs de l'espace la norme L_1 , et non plus la norme euclidienne classique L_2 . La projection c_d du vecteur c sur l'axe achromatique a pour coordonnées sur les trois axes la même valeur $(r + g + b) / \sqrt{3}$. S'agissant de nombres r, g et $b \geq 0$, la norme L_1 du vecteur c_d n'est autre que la somme des trois composantes r, g et b . Comme de plus nous désirons conserver le domaine de variation $[0, 1]$, nous prenons donc comme variable d'intensité lumineuse la

moyenne arithmétique m des trois couleurs de c :

$$(10) \quad m_1 = \frac{1}{3} (r + g + b) \quad (12)$$

On notera que si deux points c et c' se projettent au même point c_d de l'axe achromatique, alors $|cc'| = 0$, et par application de l'inégalité triangulaire on trouve $m_1(c) = m_1(c')$.

La saturation s_1 se définit comme étant proportionnelle à la norme L_1 du vecteur c_p , c'est à dire à

$$\sigma_1 = \frac{1}{3} [|2r - g - b| + |2g - b - r| + |2b - r - g|] \quad (13)$$

à une constante multiplicative près, qui assure l'intervalle $[0, 1]$ comme domaine de variation. Comme la saturation est maximale quand deux couleurs prennent une valeur extrême et la troisième la valeur extrême opposée, on voit que l'expression (13) admet 4/3 pour maximum. On définira donc s_1 par l'équation:

$$s_1 = \frac{3}{4} \sigma_1 = \frac{1}{4} [|2r - g - b| + |2g - b - r| + |2b - r - g|] \quad (14)$$

La saturation s_1 varie de 0 à 1 et vaut zéro si et seulement si $r = g = b$, c'est à dire si le point c est sur l'axe achromatique.

Pour faire disparaître les valeurs absolues de la relation (14), qui sont liées au choix de la norme L_1 , il faut mettre en évidence le maximum, la médiane et le minimum de (r, g, b) , qu'on notera max, mid et min. Comme l'expression (14) est symétrique en r, g, b , il suffit d'adopter une convention quelconque, par exemple

$$1 \geq r \geq g \geq b \geq 0 \quad (15)$$

et de remplacer en fin de calcul r par max, g par mid et b par min. Quand l'ordre (15) est vérifié, le premier terme de (14) est alors positif et le dernier négatif. En revanche, le terme central, dont le signe peut changer, impose de distinguer deux cas :

a) si $b + r \geq 2g$, ou aussi bien $m_1 \geq g$, il vient

$$s_1 = \frac{1}{2} [(r - g) + (r - b)] = \frac{3}{2} (r - m_1) \quad (16)$$

b) si $b + r \leq 2g$, ou encore $m_1 \leq g$, on a

$$s_1 = \frac{1}{2} [(r - b) + (g - b)] = \frac{3}{2} (m_1 - b) \quad (17)$$

Pour $b + r = 2g$, on trouve en particulier, pour les 2 formes :

$$\max - \min = r - b = \sigma_1 = \frac{4}{3} s_1$$

Ces égalités restent encore vérifiées aux points de saturation maximale $r = (1, 0, 0)$ et $y = (1, 1, 0)$.

La teinte θ , puisqu'elle est un angle, s'exprime de manière identique en norme L_1 ou L_2 , et l'équation (11) reste applicable.

En résumé, malgré la présence de deux cas de figure pour la saturation, les équations de passage (12), et (16) (17) forment un système linéaire, donc sensiblement plus simple que celui de la norme L_2 . La complexité du calcul arrive avec la teinte, que nous allons donc chercher à représenter plus commodément.

8 Calcul simplifié de la teinte en norme L_1

L'approximation qui suit reprend dans les grandes lignes la démarche présentée dans [7] pour justifier le calcul de la teinte dans le système HLS [9]. Mais dans les grandes lignes seulement, car notre saturation s_1 prend deux formes différentes (dont aucune n'est celle proposée dans [7]) et ne se présente pas comme un rapport.

Dans un premier temps, nous nous limitons aux vecteurs tels que $r > g \geq b$ et de saturation non nulle. Leurs projections sur le plan chromatique délimitent le triangle $ox_p y_p$ de la fig. 7. Puisque l'origine des teintes est le vecteur r_p , l'angle θ de la figure varie donc de 0 à $\pi/3$, en radians, ou conventionnellement si nous nous plaçons en variables réduites, de 0 à 1. Pour approximer θ , nous pouvons partir de l'équation de la teinte en système HLS, à savoir de $h = \frac{g-b}{\max-\min}$ et la transposer *mutatis mutandis*, c'est à dire en remplaçant $\max-\min$ par la saturation (à un facteur près) et en tenant compte de la dualité des deux expressions de la saturation.

Plus précisément, comme la droite $r_p g_p$ est parallèle à la droite rg du plan $b = 0$, lorsque θ varie de 0 à 1, le point v du plan $b = 0$ décrit la demi diagonale rz et sa projection v_p décrit le segment $r_p z_p$ du plan chromatique. Le point x , intersection du plan $r + b - 2g = 0$ (qui contient l'axe achromatique) et de la droite rg , sépare les deux zones de définition différentes de la saturation. En projection, ce point $x = (\frac{2}{3}, \frac{1}{3}, 0)$ donne le point x_p , qui correspond à la valeur $\theta = \frac{1}{2}$.

Puisque la valeur $\max - \min$ du système HLS correspond à la saturation s_1 , posons donc

$$\varphi = \frac{k}{s_1} (g - b) \quad (\text{avec } r + b \geq 2g) \quad (18)$$

On trouve bien $\theta^* = 0$ pour $r = (1, 0, 0)$. Le facteur k est déterminé par la condition d'avoir $\theta^* = \frac{1}{2}$ en $x = (\frac{2}{3}, \frac{1}{3}, 0)$, qui donne $k = \frac{3}{4}$. On notera en passant que $\frac{3}{4}$ n'est autre que le module σ_1 , i.e. non réduit du vecteur r, g, b projeté dans le plan chromatique (rel. 13).

Quand $r + b \leq 2g$, la dualité suggère de remplacer $g - b$ par $r - g$, et φ par $1 - \varphi$, soit

$$\varphi = 1 - \frac{3}{4} \frac{r - g}{s_1} \quad (\text{avec } r + b \leq 2g) \quad (19)$$

De fait, au point x_p , la saturation s_1 prend la même valeur $\frac{1}{2}$ dans les deux modes, et l'on trouve $\varphi = \frac{1}{2}$. Enfin, à l'extrémité du parcours de θ , on voit en prenant le point y , ou le point z , que $\varphi = 1$.

On peut réduire à une seule les deux équations (18) et (19) qui définissent φ en se calant par rapport à l'élément critique $r + b - 2g$, on trouve alors pour les deux expressions

$$b + r - 2g = s_1 (1 - 2\varphi) \quad (20)$$

relation qui montre l'équivalence

$$\begin{aligned} b + r - 2g \geq 0 &\iff 0 \leq \varphi \leq \frac{1}{2} \\ b + r - 2g \leq 0 &\iff \frac{1}{2} \leq \varphi \leq 1 \end{aligned}$$

9 Formules de passage en norme L_1

Les résultats précédents conduisent aux formules de passage $(r, g, b) \rightarrow (m_1, s_1, \varphi)$ suivantes

$$\begin{cases} m_1 = \frac{1}{3}(r + g + b) \\ s_1 = \frac{1}{2}(2r - g - b) = \frac{3}{2}(r - m_1) \text{ si } b + r \geq 2g \\ s_1 = \frac{1}{2}(r + g - 2b) = \frac{3}{2}(m_1 - b) \text{ si } b + r \leq 2g \\ \varphi = \frac{1}{2} - \frac{b+r-2g}{rs_1} \end{cases} \quad (21)$$

Quand $b + r \geq 2g$, ou, de manière équivalente, quand $0 \leq \varphi \leq \frac{1}{2}$, elles s'inversent en

$$\begin{cases} r = m_1 + \frac{2}{3}s_1 \\ g = m_1 - \frac{1}{3}s_1 + \frac{2}{3}s_1\varphi \\ b = m_1 - \frac{1}{3}s_1 - \frac{2}{3}s_1\varphi \end{cases} \quad (22)$$

et pour $b + r \leq 2g$ (ou $1/2 \leq \varphi \leq 1$) en

$$\begin{cases} r = m_1 + s_1 - \frac{2}{3}s_1\varphi \\ g = m_1 - \frac{1}{3}s_1 + \frac{2}{3}s_1\varphi \\ b = m_1 - \frac{2}{3}s_1 \end{cases} \quad (23)$$

Le domaine de définition $1 \geq r \geq g \geq b \geq 0$ (diminué de l'axe achromatique) correspond au tétraèdre $oryw$ de la figure 2. Les coefficients du système (21) ont été choisis pour que $m_1 s_1$ et φ varient eux aussi entre 0 et 1, ce qui ne veut pas dire qu'ils correspondent tous à des points du tétraèdre. On vérifie toutefois sans peine l'équivalence

$$1 \geq r \geq g \geq b \geq 0 \iff \frac{2}{3}s_1 \leq m_1 \leq 1 - \frac{2}{3}s_1$$

En pratique, cette condition n'est pas très gênante car les opérations que l'on fait subir aux vecteurs c , exprimés en $(m_1 s_1 \varphi)$ font rarement sortir du cube de définition de RGB.

Reste à étudier le cas où le vecteur couleur c se trouve sur l'axe achromatique, c'est à dire où $r = g = b$. Le système (21) n'a plus de sens, puisqu'on y introduit une division par zéro, et doit être remplacé par

$$\begin{aligned} m_1 &= \frac{1}{3}(r + g + b) \\ s_1 &= \frac{1}{2}(2r - g - b) = \frac{1}{2}(r + g - 2b) \\ s_1(1 - 2\varphi) &= b + r - 2g \end{aligned}$$

qui montre que φ est indéterminé. Cela ne veut pas dire qu'on n'arrive pas à trouver la couleur de c , mais que l'intensité chromatique $s_1(1 - 2\varphi)$ est nulle.

10 Passage au cube digital complet

Respectivement à la couleur, l'espace du cubeRGB est divisé en six secteurs délimités par les six plans contenant l'axe achromatique d'une part et d'autre part l'un des 3 axes r, g ou b et l'une des diagonales des carrés dans $b = 0, r = 0$ ou $g = 0$ (typiquement : l'axe achromatique, l'axe or et la diagonale oy).

On passe d'un secteur à un autre en rajoutant à φ la quantité

$$\lambda(c) = \begin{cases} 0 & \text{si } r > g \geq b \\ 1 & \text{si } g \geq r > b \\ 2 & \text{si } g > b \geq r \\ 3 & \text{si } b \geq g > r \\ 4 & \text{si } b > r \geq g \\ 5 & \text{si } r \geq b > g \end{cases} \quad (24)$$

La teinte est alors approximée par la relation

$$h_1 = (\lambda + \varphi)k$$

de même structure que dans le système HLS. Le coefficient k détermine l'unité de travail ; il vaut 60 si l'on s'exprime en degrés. En imagerie digitale, où les couleurs sont définies sur 3x8 bits, la première idée est de réserver 8 bits à la teinte. Quand on voit l'importance des erreurs d'arrondi sur la teinte [3], p. 72), il paraît bien plus sage de mesurer celle-ci sur 252 niveaux, ce qui revient à prendre $k = 42$ et de travailler avec des entiers.

Parallèlement au passage de φ à h_1 , il convient de ré-exprimer la moyenne et la saturation en termes de $\min, \text{mid}$ et $\max$. Cet exercice de ré-écriture conduit à remplacer le système (21) par le suivant, valable pour

les images de 3×8 bits

$$\begin{aligned} m_1 &= \frac{1}{3} (max + mid + min) \\ s_1 &= \frac{3}{2} (max - m_1) \quad \text{si } max + min \geq 2mid \\ s_1 &= \frac{3}{2} (m_1 - min) \quad \text{si } max + min \leq 2mid \\ h_1 &= 42 \left[\lambda + \frac{1}{2} - (-1)^{\lambda(c)} \frac{max + min - 2mid}{2s_1} \right] \end{aligned} \quad (25)$$

valable pour r, g et b définis sur 8 bits. Les coordonnées m_1 et s_1 varient alors sur 8 bits et h_1 sur 252 niveaux. Pour le passage inverse, la valeur de h_1 détermine, via λ , l'ordre entre r, g et b et la valeur de 42φ . Selon que cette dernière est ≤ 24 ou ≥ 24 , on prendra le système (22) ou (23) remplaçant r, g, b par les variables classées dans l'ordre donné par λ , et φ par 42φ . Par continuité, les $\geq$ ou $>$, dans l'équation (24) de λ sont équivalents.

11 La norme $max - min$

Nous revenons maintenant au système HLS en nous proposant de l'améliorer. A vrai dire, il ne faut pas beaucoup le modifier pour le rendre compatible avec les trois prérequisites. Il suffit de passer de sa version cylindrique à la version conique, ce qui revient à remplacer la saturation HLS par la fonction $max - min$. Comme la propriété qu'a la quantité $max - min$ d'être une semi-norme n'est en général pas connue, nous allons tout d'abord la démontrer.

Proposition 1 Dans un espace vectoriel E de dimension finie n et où le vecteur x a pour coordonnées $\{x_i, 1 \leq i \leq n\}$, la quantité

$$\delta(x) = max \{x_i, 1 \leq i \leq n\} - min \{x_i, 1 \leq i \leq n\}$$

est une semi-norme.

Proof. Pour tout $x \in E$, on a manifestement $\delta(\lambda x) = \lambda \delta(x)$. Reste à prouver l'inégalité triangulaire associée à tout couple de points (x, x') de E par la relation

$$\delta(x + x') \leq \delta(x) + \delta(x') \quad (26)$$

Nous allons d'abord montrer que

$$max(x + x') \leq max(x) + max(x')$$

Prenons pour $n^{\text{ième}}$ coordonnée celle qui maximise $x + x'$, i.e. $max(x + x') = x_n + x'_n \geq x_i + x'_i, 1 \leq i < n$. On ne peut évidemment pas avoir $x_n < x_i$ et $x'_n < x'_i$, avec $1 \leq i < n$. Supposons donc, par exemple que $max(x) = x_n$ et $max(x') = x'_j$ pour un indice $j \in [1, n]$. Il vient alors

$$max(x + x') = x_n + x'_n \leq x_n + x'_j = max(x) + max(x').$$

On prouve exactement de la même manière que $min(x + x') \geq min(x) + min(x')$, d'où par différence l'inégalité (26), ce qui termine la démonstration. ■

La semi-norme $\delta = max - min$ vérifie manifestement le 2ème prérequisite sur l'indépendance par projection sur l'axe chromatique. Rajouter à x un vecteur parallèle à la première diagonale revient à rajouter une constante (algébrique) à chacune de ses composantes, ce qui ne modifie pas la valeur $max - min$. En particulier, dans l'espace RGB le vecteur c et sa projection c_p sur le plan chromatique, de coordonnées $(2r - g - b/3, 2g - b - r/3, 2b - r - g/3)$ ont le même $max - min$.

En revanche, cette semi-norme n'est pas invariante par projection sur la diagonale principale, à la différence de L_1 et de L_2 . Plus le point c se rapproche de cette diagonale, plus sa norme δ diminue, et elle devient nulle sur l'axe achromatique. La projection c_d du vecteur courant c sur cet axe a donc toujours 0 pour la semi-norme δ . On ne peut par conséquent pas bâtir une représentation de tous les points de l'espace fondée sur cette seule semi-norme, qui est aveugle aux luminances. Mais en contre-partie, on voit aussi que pour tous les vecteurs du plan chromatique (et uniquement pour eux), la quantité $max - min$ devient une norme. C'est pourquoi on l'utilisera pour la saturation en l'associant à L_1 ou L_2 sur l'axe achromatique, avec de possibles pondérations des coordonnées.

12 Conclusion

Les normes L_1, L_2 et $max - min$ que nous avons étudiées garantissent toutes le premier prérequisite d'indépendance: deux points c et c' différents mais de même projection c_p sur le plan chromatique ont la

même saturation et la même teinte. Comme la propriété reste vraie quelle que soit la norme choisie sur l'axe a -chromatique, on peut très bien remplacer m_2 dans la rel.(10), ou m_1 dans la rel.(21) par une moyenne pondérée quelconque m (celle-ci est encore une norme):

$$m(r, g, b) = \alpha r + \beta g + \gamma b \quad \text{où} \quad \alpha + \beta + \gamma = 1,$$

ou même par un estimateur non linéaire, pourvu que celui-ci soit une norme. Le cercle des teintes restera établi en donnant le même poids aux trois couleurs fondamentales, ainsi que la saturation, même si ce n'est plus le cas de l'intensité qui remplace m_1 ou m_2 .

Pour les teintes, les résultats sont différents. Nous arrivons, aux termes de l'analyse, à une unique fonction mathématique, donnée sous sa forme exacte en relation (11), et sous deux formes approchées, avec la teinte du système HLS et celle h_1 de la relation (25). Les essais que nous avons pu faire montrent que les deux approximations sont à peu près satisfaisantes, avec un avantage pour h_1 , surtout si l'on discrétise les teintes selon un multiple de six.

References

- [1] A. Hanbury J.Serra : A 3D-polar coordinate colour representation suitable for Image Analysis (rapport technique Ecole des Mines N-37/ 02/MM), 39 p. novembre 2002 (soumis pour publication à *JVCIR*)
- [2] A. Hanbury The taming of the Hue, Saturation and Brightness Colour Space (rapport technique Ecole des Mines N-18/02/MM), juin 2002
- [3] A. Hanbury : *Morphologie Mathématique sur le cercle unité*, Thèse Ecole des Mines de Paris, France, 2002.
- [4] V. Caselles, B. Coll and J.M. Morel : Geometry and Color in Natural Images. *J.M.I.V.* 16 : 89-105, 2002.
- [5] T. Gevers and A.W.M. Smeulders A comparative Study of Several Color Models for Color Image Invariant Retrieval (Tech. report Univ. of Amsterdam, 2002)
- [6] T. Carron : *Segmentation d'images couleurs dans la base teinte-luminance-saturation : approche numérique et symbolique*. Thèse, Université de Savoie (France) 1995
- [7] H. Levkowitz and G.T. Herman : GLHS : a generalized lightness, hue and saturation color model, *CVGIP*, 55(4) : 271-285, 1993.
- [8] J. Serra: Espaces couleur et traitement d'images (rapport technique Ecole des Mines N-34/02/MM), octobre 2002
- [9] A.R. Smith : Color gammet transform pairs, *Computer Graphics*, 12(3) : 12-19, 1978.