

ECOLE NATIONALE SUPÉRIEURE DES MINES DE SAINT-ETIENNE

RAPPORT DE STAGE DE DEA

Distance entre images

Stéphane ALBIN

Saint-Etienne, Septembre 1998

Remerciements

Tout d'abord, je tiens à remercier Bernard Péroche, mon maître de stage et responsable du LISSE, pour m'avoir fait découvrir le monde de la recherche. Sa patience et sa disponibilité n'ont jamais failli malgré mes nombreuses questions saugrenues. Je le remercie aussi chaleureusement de m'avoir associé à l'écriture d'un article bien que ma participation soit infinitésimale.

Je remercie Gilles Rougeron pour son aide précieuse et sa compétence en matière de couleur. Il a, de plus, le grand mérite d'avoir jeté les bases d'une distance entre images.

Je remercie Laurent Carraro pour son concours dès que je butais sur un problème mathématique.

Je remercie Alain Trémeau et Ken Knoblauch de l'IIV Saint-Etienne, l'un pour m'avoir accepté comme co-auteur d'une publication, l'autre pour m'avoir fait entrevoir les méandres de la théorie de la détection.

Je remercie tous ceux qui ont bien voulu consacrer de leur temps en période estivale pour se prêter à mes tests.

Je remercie Jacques pour ses discussions toujours engagées, sa disponibilité malgré une fin de thèse imminente.

Je remercie Marc pour sa culture sans limites et la qualité du système qu'il maintient.

Je remercie toute l'équipe *SIMADE* et plus particulièrement Marie-Line (que ferions-nous sans elle ?), Dominique, Grégory, Helymar, Jean-Claude, Jean-Michel, Marion et Pierre, mon compagnon de DEA, pour leur bonne humeur et leur sourire inamovible.

Je remercie JeT, Nad, Seb et Toine pour leur compagnie électronique.

Je remercie mon père pour son soutien journalier. Qu'il sache que je ne l'oublie pas.

Enfin, je remercie Vanessa, Angélique, Stéphanie, Marc, Philippe, Thierry, Alex et Gilles pour leur amitié de tous les jours, sincère et partagée.

Mots-clés

Distance entre images — Vision — Psycho-visuel — Modèles de vision — Arbre quaternaire — Échantillonnage — Monte-Carlo

Résumé

Ce document est le mémoire d'un stage de DEA. Il présente l'élaboration d'une **distance entre images**. Outre les phénomènes visuels auxquels nous sommes sensibles et la façon de les modéliser, nous étudions, ici, une méthode permettant l'accélération des calculs de distance.

Notre solution repose sur la technique de l'**échantillonnage**. Nous montrons d'abord qu'un nombre réduit d'individus suffit pour connaître une valeur de distance très proche de la valeur idéale. Ensuite, nous proposons l'utilisation d'un **arbre quaternaire** pour reconstruire une carte de distances en un temps inférieur à la méthode globale où tous les pixels sont calculés.

Enfin, une procédure de test auprès d'observateurs volontaires a été organisée. Nous montrons les résultats et leur corrélation avec les valeurs que nous calculons.

Keywords

Distances between images — Vision — Vision models — Quad-tree — Sampling — Monte-Carlo

Abstract

This document reports the work of the training period of a DEA. It expounds the development of a **distance between two images**. Besides the visual phenomena and the way to model them, we study a method to compute a distance faster.

Our solution is based on a **sampling technique**. First, we show that a reduced number of samples is sufficient to get a good value. Then, we suggest the use of a **quad-tree** to build a distance map faster than computing all the pixels.

Finally, the results of a test procedure with some observers are shown. We analyze their correlation with our distance.

Table des matières

Table des figures	iv
Liste des tableaux	v
Introduction	1
1 Travaux préliminaires	2
1.1 Cadre d'étude	2
1.2 Champs visuels	2
1.3 L'espace colorimétrique <i>LLAB</i>	4
1.4 Calcul de distance	5
1.5 Résultats	5
1.6 <i>LLAB</i> est-il bien approprié ?	5
1.6.1 Les espaces basiques	6
1.6.2 L'espace <i>CIECAM97</i>	6
1.6.3 L'espace <i>LAB</i>	6
2 Le système visuel humain	8
2.1 L'œil	9
2.2 La rétine	10
2.3 Transmission de l'information au cerveau	11
2.4 Les phénomènes importants de la vision	12
2.4.1 L'accommodation	12
2.4.2 L'adaptation à la lumière	12
2.4.3 L'acuité et la sensibilité	13
2.4.4 La transduction visuelle	13
2.4.5 La vision des couleurs	13
2.4.6 Le rôle des fréquences et orientations	14
2.4.7 L'aberration chromatique	16
2.4.8 Le masquage	17
2.4.9 La fonction psychométrique	17

3	Les modèles de vision	19
3.1	Le modèle de Sarnoff, un algorithme séquentiel	19
3.1.1	Améliorations	22
3.2	Le modèle de Daly, <i>Visual Difference Predictor</i>	23
3.2.1	Amplitude non linéaire	24
3.2.2	Fonction de sensibilité au contraste	24
3.2.3	Mécanismes de détection	24
3.3	D'autres approches	28
3.3.1	Distance perceptuelle à base d'ondelettes	28
3.3.2	Un tracé de rayons fréquentiel	28
3.3.3	Le modèle de Watson	29
4	Accélération des calculs : sonder l'image	30
4.1	Échantillonnage	30
4.1.1	Précision	30
4.1.2	Temps de calcul	31
4.2	Reconstruction de l'image de distances	33
4.2.1	Nombre d'échantillons	33
4.2.2	Quelle condition ?	34
4.3	Une pseudo-distance	35
5	Expérimentation	38
5.1	Trois tests	38
5.2	Les images	39
5.3	Résultats	45
5.3.1	Première expérience	45
5.3.2	Deuxième expérience	47
5.3.3	Troisième expérience	49
5.3.4	Conclusion	50
6	Perspectives	53
6.1	Un modèle de vision	53
6.1.1	Système optique	53
6.1.2	Fréquences et orientations	53
6.1.3	Fonction psychométrique	54
6.2	La dispersion des erreurs	54
6.3	Un tracé de rayons incrémental	55
	Conclusion	57
	Bibliographie	60
	Manuel d'utilisation	61

Table des figures

1.1	Champs visuels associés à chaque pixel.	3
1.2	Distribution des poids dans le champ visuel.	4
1.3	Segmentation du champ visuel.	4
2.1	Le système visuel humain	8
2.2	Coupe transversale de l'œil	9
2.3	Schéma d'une section de rétine	10
2.4	Répartition spatiale des photo-récepteurs dans la rétine	11
2.5	Réponse d'un bâtonnet en fonction de l'intensité de son stimulus	14
2.6	CSF mesurée par Watanabe et al.	15
2.7	CSF mesurée par Campbell et al.	15
2.8	Aberration chromatique.	16
2.9	Le phénomène de masquage	17
2.10	Fonction psychométrique	18
3.1	Diagramme d'utilisation d'un modèle de vision	20
3.2	Le modèle de Sarnoff	21
3.3	Le modèle de Daly	24
3.4	Fonction de sensibilité au contraste en deux dimensions	25
3.5	Cortex Transform : filtre fréquentiel	26
3.6	Cortex Transform : composante radiale du filtre fréquentiel	26
3.7	Cortex Transform : orientations	27
3.8	Cortex Transform	27
3.9	Le modèle de Watson	29
4.1	Convergence du sondage d'une image.	32
4.2	Image de distances : Méthode adaptative	33
4.3	Condition de subdivision	35
4.4	Points effectivement calculés	37
5.1	Troisième test	39
5.2	Pièce meublée	41
5.3	<i>Cornell Box</i>	42
5.4	Lenna	43
5.5	Poivrons	44

5.6	Calcul de distance : projection gaussienne	51
5.7	Calcul de distance : quantification couleur en 8 niveaux	52
6.1	Choix des rectangles	54
6.2	Différences suivant la répartition des erreurs	55
6.3	Segmentation de l'image de distances	56

Liste des tableaux

1.1	<i>Cornell Box</i> : Résultats des tests.	5
3.1	Prise en compte des orientations avec les ondelettes de Haar.	23
4.1	Méthode globale : temps de calcul.	30
4.2	Sondage de l'image : Temps de calcul	31
4.3	Temps de calcul selon la condition de subdivision	35
4.4	Temps de calcul en <i>LLAB</i>	36
4.5	Transitivité : caractéristiques des images	36
5.1	Caractéristiques de la pièce meublée	40
5.2	Caractéristiques des images photographiques	40
5.3	Résultats de la première expérience sur la pièce	45
5.4	Résultats de la première expérience sur la <i>Cornell Box</i>	46
5.5	Résultats de la première expérience sur Lenna	46
5.6	Résultats de la première expérience sur les poivrons	47
5.7	Poids affecte aux images	47
5.8	Distances calculées	48
5.9	Ordonnancement des images de la pièce	48
5.10	Ordonnancement des images de la <i>Cornell Box</i>	48
5.11	Ordonnancement des images de Lenna	49
5.12	Ordonnancement des images des poivrons	49
5.13	Troisième test	50

Introduction

La formation doctorale orchestrée par l'*IUP Vision*, de l'*Université Jean Monnet de Saint-Etienne* et l'*École Nationale Supérieure des Mines de Saint-Etienne* suppose un stage d'initiation à la recherche. Cet enseignement s'effectue en collaboration avec l'*École Supérieure de Chimie, Physique et Électronique (CPE) de Lyon*.

L'École des Mines m'a donc accueilli; plus précisément, le stage s'est déroulé dans les locaux du *LISSE* (Laboratoire d'Images de Synthèse de Saint-Etienne, sous l'égide de *Bernard Péroche*), rattaché au centre *SIMADE* (Sciences de l'Information, Mathématiques, Aide à la Décision et Environnement) de l'école.

D'une durée de six mois (de mars à août 1998), le but de ce stage était de proposer une distance entre images. En effet, pouvoir quantifier l'écart séparant deux images est très utile. Imaginons, par exemple, un algorithme de tracé de rayons incrémental. Un tel outil nous permettrait de connaître les différences entre deux images calculées successivement. Ainsi, nous pourrions savoir lorsqu'il n'est plus utile de raffiner certaines zones. La synthèse d'images étant un domaine très coûteux en ressources, les calculs inutiles sont à bannir.

Ce travail n'en est pas à ses débuts. Les premières pierres ont été posées par Gilles Rougeron dans sa thèse [28]. C'est dans la continuité que les résultats présentés ici s'inscrivent.

Ce rapport se découpe en plusieurs parties.

Dans le chapitre 1, nous étudierons le modèle de départ de Gilles Rougeron.

Dans les chapitres 2 et 3, nous aborderons le système visuel humain et les solutions existantes pour le modéliser.

Les chapitres 4 et 5 seront consacrés aux solutions que nous proposons pour améliorer les résultats et aux tests effectués.

Enfin, dans le chapitre 6, nous évoquerons quelques idées pour le futur.

Chapitre 1

Travaux préliminaires

La définition d'une distance entre images est un problème fondamental. Si, en traitement et analyse d'images, des travaux ont été effectués, peu de résultats en revanche sont connus pour la synthèse d'images. Or, le problème est tout aussi crucial. De plus, quel que soit le domaine d'application, cette voie de recherche n'en est qu'à ses débuts.

Nous allons maintenant présenter les solutions proposées par Gilles Rougeron dans [28], sur lesquelles nous nous sommes appuyés.

1.1 Cadre d'étude

La particularité de la distance présentée est d'être spécifique à la synthèse d'images. L'auteur a choisi d'utiliser tous les avantages que procure la synthèse d'images, à savoir :

- la parfaite connaissance de la géométrie de la scène ;
- la possibilité de travailler avec des grandeurs physiques.

D'autre part, il faut noter que l'espace colorimétrique *LLAB*, décrit en 1.3, est utilisé. La méthode suppose plusieurs hypothèses :

1. Le point de vue doit être identique pour les deux images ;
2. les objets de la scène ne doivent pas comporter de textures ;
3. la scène ne doit contenir qu'un seul type de source lumineuse.

Ces restrictions réduisent beaucoup le champ d'application de la distance. Nous verrons, plus loin, que nous pouvons nous en passer.

1.2 Champs visuels

L'auteur propose de prendre en compte l'acuité visuelle de l'observateur. Ainsi, à chaque pixel, on associe deux masques représentant les champs visuels de 2° et 20° . Cela correspond à la zone fovéale¹ et à la zone d'arrière-fond.

La figure 1.2, tirée de [28], représente les masques associés à chaque pixel, calculés à l'aide des formules 1.1.

¹aussi appelée *zone de focus*.

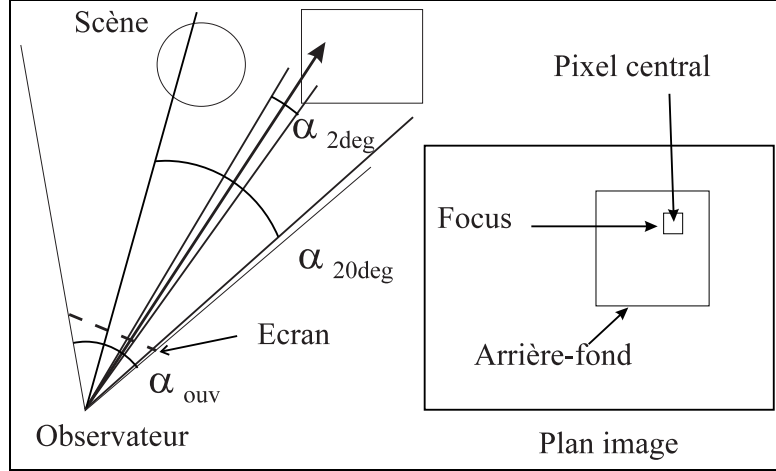


FIG. 1.1: Champs visuels associés à chaque pixel.

$$ng = i - \frac{1}{2} \frac{2i n_{ph} \tan \frac{\alpha_h}{2} - n_{ph}^2 \tan \frac{\alpha_p}{2}}{n_{ph} \tan \frac{\alpha_h}{2} + 2i \tan \frac{\alpha_h}{2} \tan \frac{\alpha_p}{2}} \quad (1.1a)$$

$$nd = \frac{1}{2} \frac{2j n_{ph} \tan \frac{\alpha_h}{2} + n_{ph}^2 \tan \frac{\alpha_p}{2}}{n_{ph} \tan \frac{\alpha_h}{2} - 2j \tan \frac{\alpha_h}{2} \tan \frac{\alpha_p}{2}} - i \quad (1.1b)$$

$$nh = j - \frac{1}{2} \frac{2i n_{pv} \tan \frac{\alpha_v}{2} - n_{pv}^2 \tan \frac{\alpha_p}{2}}{n_{pv} \tan \frac{\alpha_v}{2} + 2i \tan \frac{\alpha_v}{2} \tan \frac{\alpha_p}{2}} \quad (1.1c)$$

$$nb = \frac{1}{2} \frac{2j n_{pv} \tan \frac{\alpha_v}{2} + n_{pv}^2 \tan \frac{\alpha_p}{2}}{n_{pv} \tan \frac{\alpha_v}{2} - 2j \tan \frac{\alpha_v}{2} \tan \frac{\alpha_p}{2}} - j \quad (1.1d)$$

où

- (i, j) désigne le pixel ;
- α_p représente l'angle du champ visuel (2° ou 20°) ;
- α_h et α_v représentent les angles d'ouverture de la scène ;
- n_{ph} et n_{pv} désignent la taille de l'image.

D'autre part, l'acuité visuelle étant meilleure dans la zone fovéale, le masque est pondéré avec un poids de 1 pour le focus. La zone d'arrière-fond est pondérée linéairement de manière décroissante de 1 à 0, du centre vers les bords.

La figure 1.2 illustre cette distribution.

L'espace *LLAB* impose que la cible soit uniforme. Or, la portion d'image dans la zone fovéale peut contenir plusieurs objets. Connaissant facilement la géométrie de la scène, une segmentation est effectuée pour ne garder que les pixels appartenant à la même face du même objet que le pixel central (cf. figure 1.2). Le reste de la zone de focus est considéré comme appartenant l'arrière-fond.

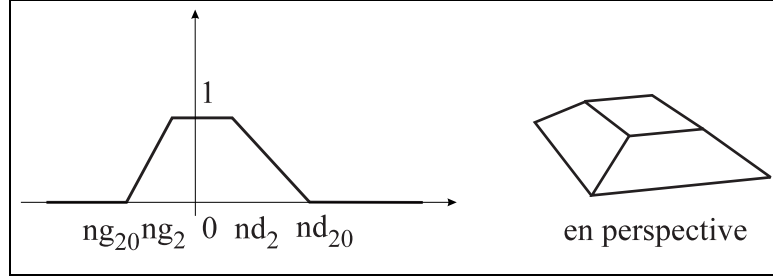


FIG. 1.2: Distribution des poids dans le champ visuel.

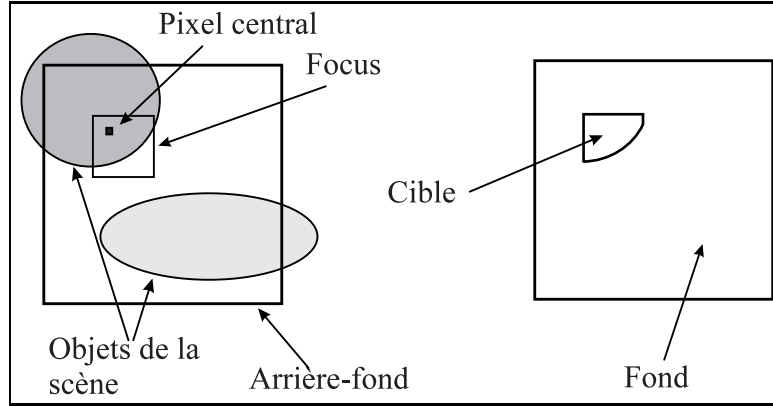


FIG. 1.3: Segmentation du champ visuel.

1.3 L'espace colorimétrique *LLAB*

Ces dernières années, de nombreux travaux sur la couleur ont abouti à des propositions. Les espaces *LLAB*, défini par Luo dans [16], *RLAB* [8] et *Hunt94* [14], [13] en sont des exemples. Se basant sur l'apparence des couleurs, ils permettent de calculer les différences entre couleurs. Gilles Rougeron a choisi d'utiliser le premier.

Le détail des calculs peut être trouvé dans [28].

La première étape est constituée d'un recalage sur une source lumineuse de type D6500 à partir d'une source quelconque, connaissant ses composantes *XYZ*. Des transformations matricielles permettent d'obtenir les composantes du pixel sous l'éclairage d'une source D6500.

La seconde étape est inspirée de l'espace *CIE Lab* 1976. Ainsi, des valeurs de clarté (L_L), opposition rouge-vert (A_L), opposition jaune-bleu (B_L), chromaticité (C_L), angle de teinte (h_L) et teinte (H_L) peuvent être calculées sur la zone de focus. La valeur médiane de chaque canal est alors affectée au pixel central. Les trois canaux utilisés sont L_L , C_L et H_L .

La conversion *LAB* repose sur des valeurs fixes. Cependant, ce n'est pas le cas ici. Les constantes dépendent de la clarté L_L au pixel courant. Cela nous empêche d'exploiter toute forme de cohérence, et donc oblige à recommencer tous les calculs pour chaque pixel.

Pour une image de dimension 512 par 512, la taille moyenne du masque est de $7,6^2$, soit 57,7 pixels. Pour chaque image, on a donc $512 \times 512 \times 7,6^2 = 1,51 \times 10^6$ conversions *LLAB*.

1.4 Calcul de distance

Soit Z_p la zone de focus segmentée affectée au pixel p . La conversion de l'espace XYZ vers $LLAB$ peut se formaliser de la façon suivante :

$$LLAB(p) = \text{median}(\mathcal{F}(p))$$

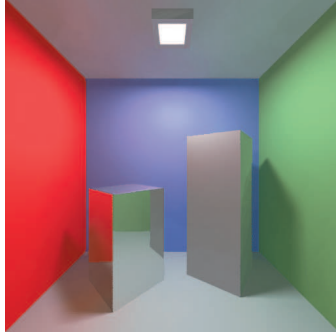
On note $\mathcal{F}(p)$, la fonction calculant les valeurs $LLAB$ en un seul pixel p de la zone de focus. L'écart couleur est ensuite calculé pour tous les pixels des images I_1 et I_2 .

$$\forall p, \quad \text{distance}(p) = |LLAB(p_{I_1}) - LLAB(p_{I_2})|.$$

1.5 Résultats

Des tests ont été effectués sur la scène standard nommée "Cornell Box". Trois algorithmes de rendu différents ont été utilisés :

- Tracé de rayons avec terme ambiant ;
- Tracé de rayons avec la méthode d'interpolation vectorielle de la composante indirecte [36] (VEI) ;
- méthode de Monte-Carlo avec 256 échantillons par hémisphère.

Monte-Carlo	Ambiant	VEI
		
14 h 5 mn 13 s	3 mn 27 s	4 mn 9 s
Distance moyenne	12, 24	4, 17

TAB. 1.1: *Cornell Box* : Résultats des tests.

Sur une station *Silicon Graphics Indigo 2*, équipée d'un microprocesseur R10000 cadencé à 250 Mhz, le temps de calcul est d'environ 32 minutes.

1.6 $LLAB$ est-il bien approprié ?

Après avoir discuté de l'intérêt d' $LLAB$, nous évoquerons quelques espaces fréquemment utilisés, parmi lesquels LAB , qui nous semble être la meilleure solution.

Si les résultats avec *LLAB* sont qualitativement encourageants, il n'est évidemment pas envisageable d'utiliser cela dans un algorithme de tracé de rayons. Les temps de calcul sont beaucoup trop importants du fait de l'emploi de *LLAB*.

Il convient donc de se demander si cet espace couleur est bien utile. Nous pouvons déjà remarquer que l'auteur ne l'exploite pas dans les conditions strictes où il a été défini. En effet, Luo, dans [16], impose que la cible soit uniforme et le fond achromatique. Cette seconde condition n'est pas remplie. Il nous est alors difficile de garantir la validité de *LLAB*.

Ensuite, Gilles Rougeron impose, à notre sens, beaucoup trop de contraintes. Les scènes que nous souhaitons synthétiser comportent souvent plusieurs sources lumineuses, de plus différentes. Ceci est encore en contradiction avec la définition de *LLAB*. Enfin, les textures ne peuvent pas être prises en compte.

Avec un champ d'application réduit, des temps de calcul prohibitifs, nous estimons que cet espace colorimétrique n'est pas utilisable à l'heure actuelle.

1.6.1 Les espaces basiques

Nous ne nous étendrons pas sur la description de ces espaces. Citons les plus répandus : *RVB* et *XYZ*. Ces espaces ne sont pas uniformes. On ne peut donc garantir la véracité des distances calculées. C'est pour cela que la *Commission Internationale de l'Eclairage*, *CIE*, tente de définir de meilleures solutions.

1.6.2 L'espace *CIECAM97*

C'est à partir des trois modèles *LLAB*, *Hunt* et *RLAB* que Luo et Hunt ont proposé, dans [17], un nouvel espace, *CIECAM97*, tentant de prendre les avantages de chacun. Il repose sur trois étapes :

- une transformation d'adaptation chromatique. C'est la phase de recalage sur un blanc de référence. La transformation utilisée est celle de Bradford.
- une fonction de réponse dynamique. Cette fonction modélise la réponse minimale de 1 pour les faibles stimulus et maximale de 41 pour les stimulus importants. La transition entre les deux approxime une fonction racine.
- une projection dans un espace couleur pour représenter les données.

On obtient alors un espace à 7 dimensions représentant : l'angle de teinte, la teinte, la luminosité, le contraste, la saturation, la chromaticité et le piqué des couleurs.

Nous ne donnerons pas le détail des opérations qui peuvent être trouvées dans [17].

Si cet espace semble, a priori, donner de meilleurs résultats que *LLAB*, il ne permet toujours pas l'emploi de sources lumineuses diverses dans une même scène. De plus, les temps occasionnés risquent d'être, là encore, prohibitifs.

1.6.3 L'espace *LAB*

Cet espace présente le double avantage d'être uniforme (pour les petites valeurs), tout en restant plus facilement calculable. S'il n'y a pas de solution miracle, celle-ci nous semble être

la moins pire comparativement aux espaces présentés ci-avant.

Chapitre 2

Le système visuel humain

Cette description est tirée de [28].

Le système visuel peut être subdivisé en trois sous-ensembles, représentant le cheminement de l'information visuelle de l'œil vers le cerveau. Le couple *œil-rétine*, d'abord, officie comme capteur des signaux lumineux. Ensuite, le *nerf optique*, le *chiasma* et le *corps genouillé latéral* permettent la transmission nerveuse de l'image au *cortex*, enfin, où elle est décodée (figure 2.1). Chaque étape influe sur le signal transmis. Des pré-traitements sont effectués mais aussi des pertes y sont occasionnées.

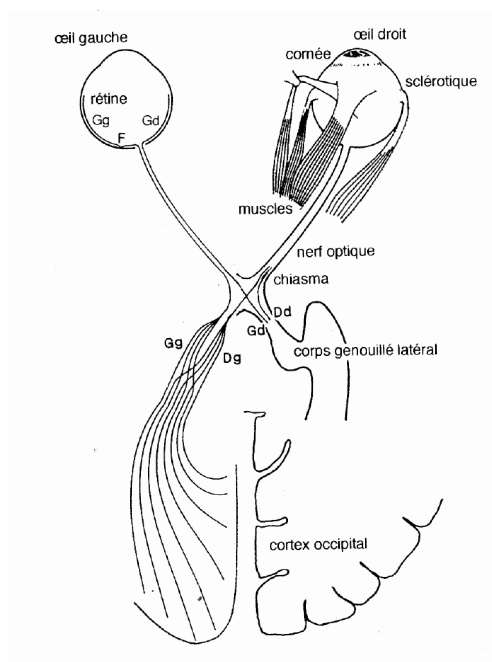


FIG. 2.1: Le système visuel humain

2.1 L'œil

L'œil est composé de plusieurs éléments qui ont chacun une influence dans le cheminement de la lumière et la compréhension du signal optique par le cerveau¹, (figure 2.2).

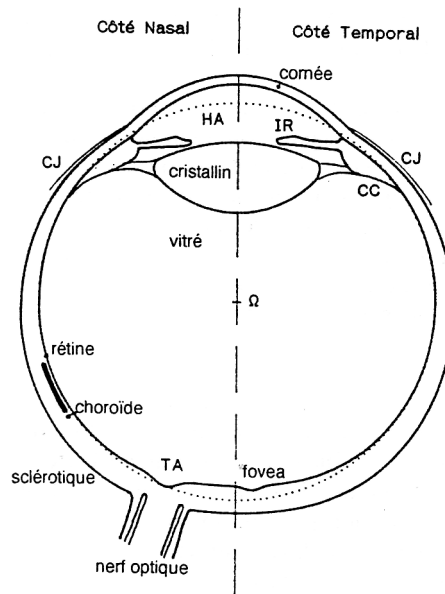


FIG. 2.2: Coupe transversale de l'œil

Parmi ces éléments, les plus importants sont :

- la *cornée* : C'est la couche externe convexe de l'œil. Elle est composée de quatre couches successives. Sa courbure dépend des individus et varie aussi avec l'âge ;
- l'*humour aqueuse* : C'est un liquide salin et alcalin sous pression qui maintient ainsi la rigidité du globe oculaire ;
- l'*iris*, un diaphragme vertical percé en son centre par la pupille. Il permet, en ajustant la taille et la forme de celle-ci, de réguler la quantité de lumière atteignant la rétine ;
- le *cristallin*, une lentille transparente placée derrière l'iris. La perte de transparence des cellules cristallines, entraînant leur mort, s'appelle la *cataracte* ;
- le *corps vitré*, qui constitue les 4/5 du volume de l'œil. Composé d'un liquide albumineux sous forme de gelée, il est parfaitement transparent. Son indice de réfraction varie suivant la longueur d'onde (d'environ 1.345 pour 400nm à 1.330 pour 700nm). Ce phénomène s'appelle *l'aberration chromatique* et il diminue l'acuité visuelle.

¹Le lecteur trouvera de plus amples explications dans [1] et [29]

2.2 La rétine

C'est la membrane la plus interne de l'œil (un prolongement du nerf optique). Elle reçoit les signaux lumineux, et assure leur transmission au cerveau par l'intermédiaire du nerf optique. La rétine possède plusieurs zones où l'acuité visuelle varie.

La **fovea** est la zone d'acuité maximale. Décalée de 4 degrés par rapport à l'axe optique, elle contient en son centre, la *foveola* où les cellules *photo-réceptrices* sont moins nombreuses. La région périphérique de la *fovea* est colorée en jaune, ce qui accroît l'acuité visuelle et compense le défaut de cellules photo-réceptrices par rapport à la *fovea*. Notons enfin que la zone où émerge le nerf optique s'appelle la tâche aveugle. Elle ne contient aucune cellules photo-réceptrices.

Outre une organisation spatiale, la rétine possède aussi plusieurs couches. La couche externe est composée de cellules photo-sensibles, les **bâtonnets** et les **cônes**. Ces photo-récepteurs reçoivent l'information optique par l'intermédiaire de pigments visuels et doivent la transmettre au cerveau par influx nerveux (figure 2.3) grâce à plusieurs types de cellules. Les cellules bipolaires, d'abord, font le lien entre les photo-récepteurs et les cellules ganglionnaires. Il en existe deux types. Les petites sont associées aux cônes et les grosses aux bâtonnets. Les cellules horizontales et amacrine, elles, sont chargées de propager l'information latéralement.

Enfin, l'impulsion visuelle est propagée via les cellules ganglionnaires, en contact avec les fibres du nerf optique.

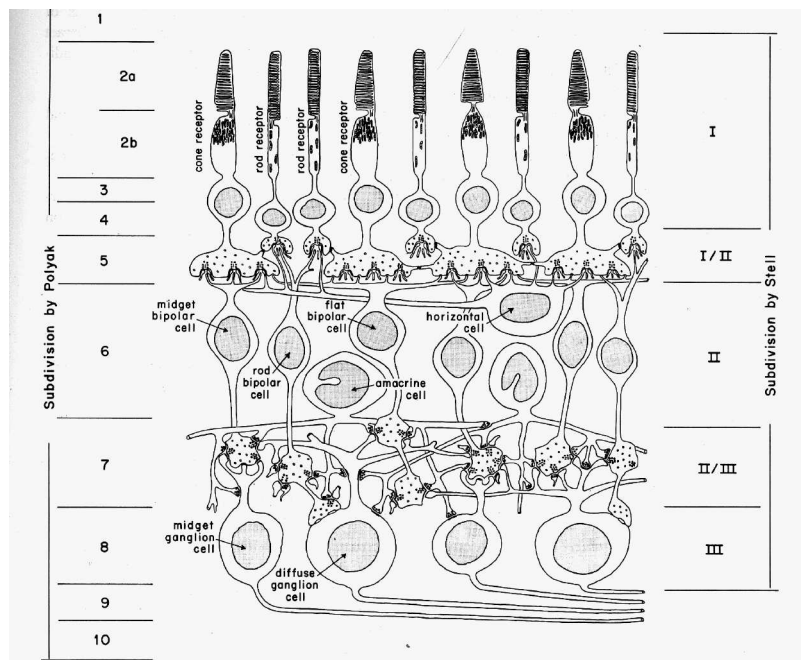


FIG. 2.3: Schéma d'une section de rétine

Il faut noter qu'il existe trois types de cônes : S, M et L (pour *short*, *medium* et *long*

wavelength). Leurs sensibilités maximales sont respectivement de : $420nm$, $530nm$ et $560nm$. Les bâtonnets, eux, ne sont sensibles que pour des longueurs d'ondes inférieures à $495nm$.

Ces photo-récepteurs fonctionnent selon trois modes :

- *photopique* le jour, seuls les cônes sont actifs,
- *scotopique* la nuit, ce sont les bâtonnets qui fonctionnent,
- *mésopique*, où tous les photo-récepteurs sont actifs.

De plus, leur répartition n'est pas égale dans la rétine. Ainsi, les cônes ne sont pratiquement présents que dans la *fovea* avec une densité maximale de $150000/mm^2$. Celle-ci décroît de manière concentrique. A 40 degrés de l'axe visuel, on considère la densité des cônes comme trop faible pour avoir une influence sur la vision.

Les bâtonnets, eux, ne sont pas présents dans la *foveola* et très peu dans la *fovea*. Leur densité maximale est d'environ $160000/mm^2$ à 20 degrés. La vision à la périphérie de la rétine est donc monochromatique et ne sert qu'à la détection des mouvements. La figure 2.4 illustre cela.

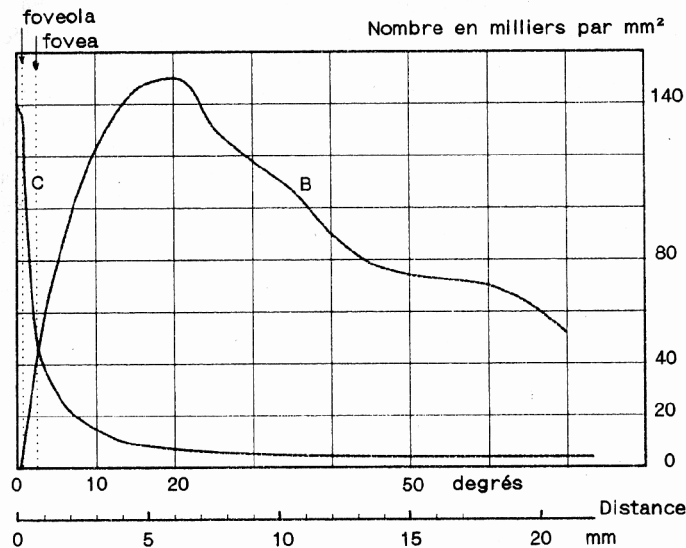


FIG. 2.4: Répartition spatiale des photo-récepteurs dans la rétine

2.3 Transmission de l'information au cerveau

Les deux nerfs optiques se rejoignent au niveau du *chiasma*. Il y a alors séparation suivant le champ observé. L'information gauche est envoyée vers l'hémisphère droit et inversement. Les reliefs peuvent ainsi être détectés par comparaison des deux demi-images. Le signal optique, transformé en signal électrique par la rétine, arrive alors au *cortex* où il est analysé. Des phénomènes complexes entrent en jeu. Ainsi, la perception visuelle dépend des

facteurs environnants comme le bruit. L'état de fatigue ou de stress joue aussi un rôle dans la compréhension de l'image fournie.

2.4 Les phénomènes importants de la vision

2.4.1 L'accomodation

L'accomodation est la faculté d'adapter l'œil en fonction de la distance aux objets observés. C'est le rôle de la cornée et du cristallin qui, en se déformant, ajustent l'image projetée sur la rétine. L'*hypermétropie* et la *myopie* sont les troubles visuels issus du dysfonctionnement de l'accomodation. Ces problèmes sont corrigés par l'ajout d'une lentille convergente ou divergente devant la cornée.

2.4.2 L'adaptation à la lumière

L'intensité lumineuse à la surface de la terre varie beaucoup. Ainsi, l'éclairage peut varier de $10^{15} W.cm^{-2}$ entre une journée ensoleillée et une nuit sans lune. L'œil doit être capable de capter les photons dans l'obscurité mais aussi de se protéger d'une lumière trop vive ; ceci grâce à un système d'adaptation à la lumière.

Ce phénomène s'illustre facilement. Lorsque nous quittons un endroit fortement éclairé pour un autre plus sombre, nous sommes temporairement aveugles. Après quelques secondes, notre vision se rétablit. C'est également le cas dans la situation inverse.

Plusieurs modes permettent cela.

La régulation de la quantité de lumière atteignant la rétine

L'iris est capable de modifier la forme et la taille de la pupille. Une pupille large laissera passer beaucoup de lumière (mode nocturne), et inversement. Ce phénomène très rapide n'est cependant que temporaire. Il permet aux autres dispositifs plus lents de se mettre en place. La régulation peut se faire aussi par des mouvements rétinien. En effet, en vision nocturne, les bâtonnets sont allongés et les cônes contractés. L'effet inverse se produit en vision diurne. Toutefois, ce mécanisme est, à l'heure actuelle, peu connu.

Le pigment visuel

Composé de *rhodopsine*² il a pour charge d'absorber les photons et d'exciter les cellules visuelles. Cette molécule possède, de plus, la propriété de blanchir avec l'intensité lumineuse. Son pouvoir absorbant est alors réduit. Il faut 7 minutes aux pigments visuels des cônes pour se régénérer et 40 minutes pour ceux contenus dans les bâtonnets. C'est pour cette raison que nous sommes moins gênés par une lumière forte soudaine que par l'entrée dans un tunnel très sombre, par exemple.

²pour les bâtonnets. La substance contenue dans les cônes dépend de leur type mais produit un effet similaire.

Le traitement neural des informations

Ce dernier phénomène est encore mal compris. Cependant, il semble que les neurones en contact avec les cellules rétinienne jouent aussi un rôle dans le contrôle du gain.

2.4.3 L'acuité et la sensibilité

L'acuité est le pouvoir de résolution, c'est à dire la faculté de déceler de très petits objets. Pour cela, il faut que l'image fournie par le système optique (la cornée et le cristallin) soit de bonne qualité. Cela dépend aussi des caractéristiques physiologiques de chaque rétine. Ainsi, on compte 7 millions de cônes et 120 millions de bâtonnets pour seulement un million de fibres dans le nerf optique. L'information est donc compressée avec pertes. Les signaux reçus par deux cellules proches seront combinés vers une seule fibre. Plus la distance entre deux photo-récepteurs telle que deux fibres nerveuses soient sollicitées est réduite, meilleure est l'acuité visuelle. Il faut noter aussi que cela dépend de l'intensité lumineuse des sources observées.

La sensibilité, elle, est l'aptitude à détecter de faibles quantités de lumière. Elle dépend de la longueur et de la surface des photo-récepteurs. Ce sont donc les bâtonnets qui offrent une meilleure sensibilité à la lumière. C'est pourquoi, il est communément admis que la région périphérique à la *fovea*, plus riche en bâtonnets, détecte l'information de luminance.

2.4.4 La transduction visuelle

La transduction est la conversion de l'énergie lumineuse en impulsions électriques. Si les études menées ont porté sur les bâtonnets, il semble que les cônes fonctionnent de manière similaire.

La perméabilité des cellules des bâtonnets permet le passage des ions et provoque une excitation électrique sur le nerf optique. L'hyperpolarisation des cellules visuelles est régie, notamment, par l'ion sodium. Cependant, ni le mécanisme en jeu, ni l'influence d'autres substances telles que l'ion calcium ou le *guanosine 3', 5'* n'ont pu être approfondis pour l'instant.

Les physiologistes ont quand même défini une loi mettant en relation l'intensité lumineuse et la réponse du récepteur (cf. figure 2.5). Cette loi prend la forme d'une sigmoïde.

2.4.5 La vision des couleurs

La couleur dépend de la discrimination des longueurs d'ondes dans le spectre. Deux facteurs permettent de qualifier une couleur : le *ton* qui est la longueur d'onde prédominante et la *saturation*, c'est à dire la quantité de cette longueur d'onde par rapport à la distribution spectrale.

Cependant, il convient de faire un distinguo entre la notion de couleur physique et la perception de celle-ci. Ainsi, certaines espèces animales ont une vision seulement monochromatique ou dichromatique³ (un seul ou deux types de cônes dans la rétine).

L'homme est considéré comme trichromate. La première théorie, développée par Young en 1802 puis Helmholtz, suggérait que la rétine contient trois types de cônes aux sensibilités

³Cette caractéristique se retrouve, très rarement, chez certains individus humains.

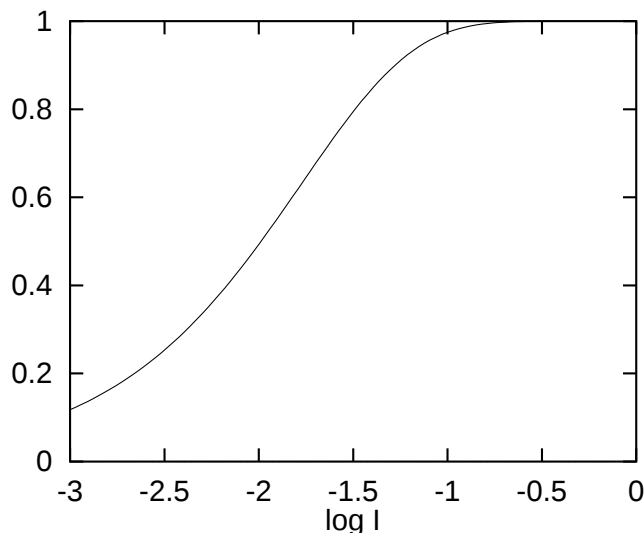


FIG. 2.5: Réponse d'un bâtonnet en fonction de l'intensité de son stimulus

spectrales différentes (rouge, vert, bleu). Certaines critiques furent émises. En 1878, Hering a proposé une approche alternative. Selon lui, l'homme serait sensible aux différences de couleurs. Ainsi, les oppositions noir/blanc, rouge/vert et bleu/jaune décriraient la vision chromatique humaine.

Des études récentes (cf Padgham & Saunders, [26]) semblent attester que les deux théories sont présentes. Les cônes seraient bien sensibles au rouge, vert et bleu comme indiqué par Young et Helmholtz. Il y aurait, de plus, un phénomène chimique dans la rétine. L'information sur les oppositions de couleur serait aussi transmise au cerveau, le jaune étant obtenu par combinaison du rouge et du vert.

2.4.6 Le rôle des fréquences et orientations

De nombreuses études ont montré l'importance de l'organisation fréquentielle de l'image dans sa compréhension. Pour le cas des orientations, la littérature est moins abondante.

Fonction de sensibilité au contraste, CSF

Watanabe et al. [32] ont effectué des mesures sur des sujets humains. Le but était de connaître leur sensibilité au contraste en fonction de la fréquence spatiale. La figure 2.6 montre les résultats.

On observe que les personnes testées sont plus particulièrement sensibles à des fréquences spatiales comprises entre 1 et 15 cycles par degré. De plus, au delà de 50 cycles par degrés, l'œil ne détecte plus rien.

Les mesures de Campbell et al. [6] (figure 2.7) sont similaires.

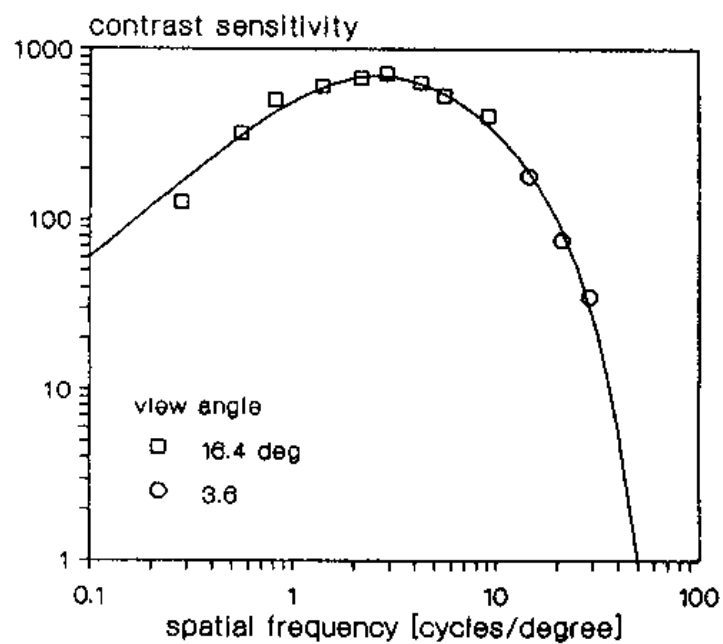


FIG. 2.6: CSF mesurée par Watanabe et al.

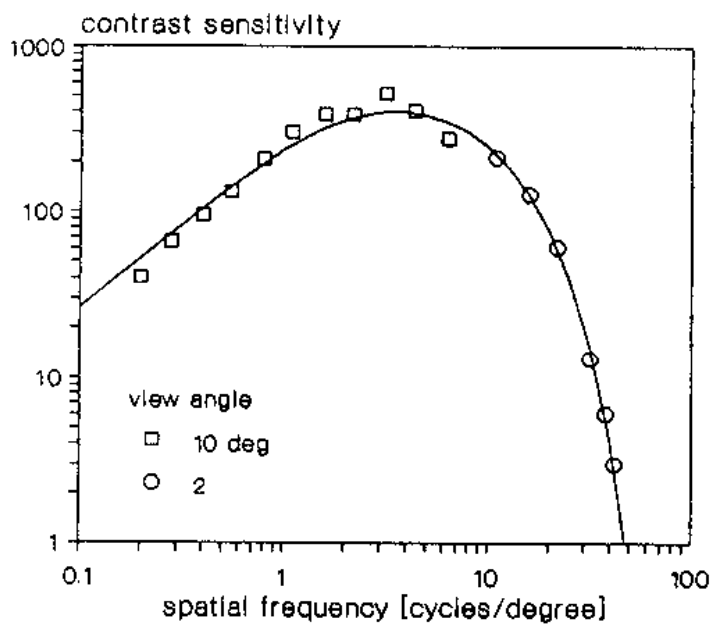


FIG. 2.7: CSF mesurée par Campbell et al.

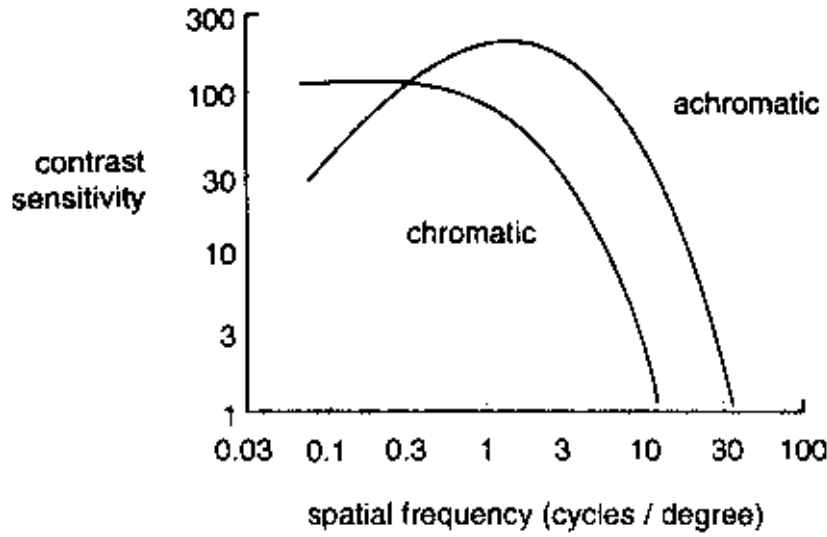


FIG. 2.8: Aberration chromatique.

Des études ont été menées pour trouver une formule analytique proche des résultats expérimentaux. Parmi les plus connues, nous pouvons citer la formule de Mannos et Sakrison [19]. Celle-ci est une des premières solutions proposées. Elle est utilisée dans bon nombre de modèles.

$$CSF(f) = 2,6(0,0192 + 0,114f)e^{-(0,114f)^{1,1}} \quad (2.1)$$

Citons aussi la solution de Ngan et al. [24], plus récente, et donnant de meilleurs résultats.

$$CSF(f) = (0,31 + 0,69)e^{-0,29f} \quad (2.2)$$

Les orientations

Il y a eu très peu d'études menées sur le sujet. Cependant, il est facile de se rendre compte intuitivement de son influence. Lorsque l'on regarde une image, nous sommes plus sensibles à l'aliassage sur des formes verticales ou horizontales qu'obliques. De la même manière, dans la vie courante, il est plus difficile d'ajuster un objet avec précision s'il est incliné à 45°.

2.4.7 L'aberration chromatique

Cette sensibilité fréquentielle diffère selon le type de cônes. Des études ont montré que les cônes *S* ne détectaient rien au delà de 4 cycles par degré, [22], [27]. Nous devons donc utiliser trois fonctions *CSF* différentes.

La figure 2.8 montre deux *CSF* achromatique et chromatique.

2.4.8 Le masquage

La détection d'une texture n'est pas seulement fonction de la fréquence spatiale locale. La présence d'autres textures dans la zone influe aussi. Considérons par exemple une image avec un fond texturé orienté horizontalement. Un objet ayant la même fréquence et la même orientation sera difficilement détectable. Si au contraire, cet objet est orienté verticalement, il sera mieux repéré. Donc, l'objet devra avoir un poids plus important en terme d'information significative (figure 2.9).

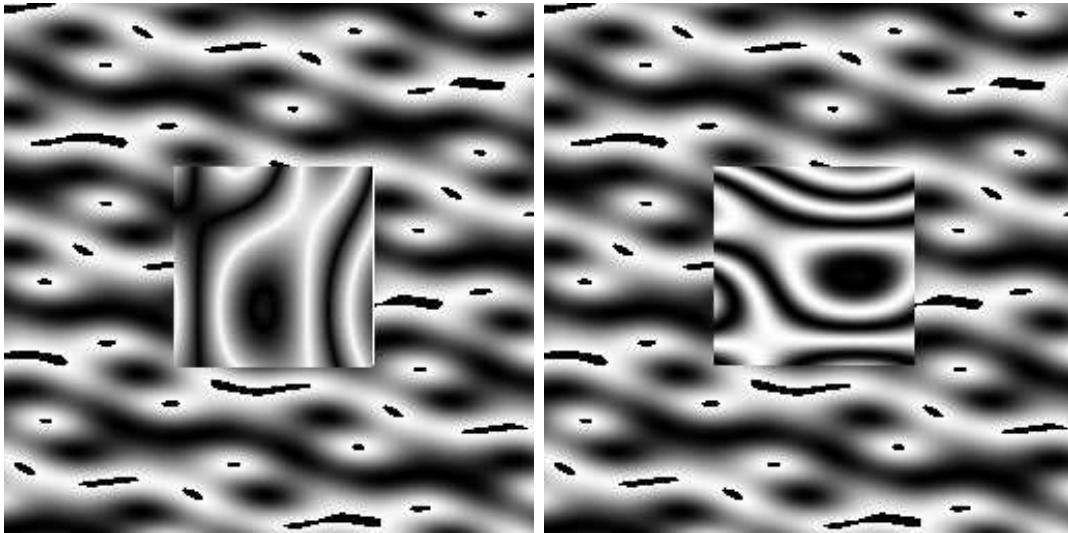


FIG. 2.9: Le phénomène de masquage

2.4.9 La fonction psychométrique

On définit le seuil de détection comme l'inverse de la fonction CSF.

$$S = \frac{1}{CSF}. \quad (2.3)$$

Si le contraste est inférieur à ce seuil, on considère que le système visuel ne détecte rien, ou tout au plus une forme grise peu précise. Au contraire, si le contraste est supérieur, la zone de l'image est considérée comme significative. Mais, ce n'est pas une alternative binaire. La transition se fait progressivement. La probabilité de détection varie donc en fonction de l'écart du contraste calculé avec le seuil. Lorsque la distance est égale à 1, la probabilité vaut 75%.

C'est le rôle de la fonction psychométrique. La figure 2.10 montre un exemple. La forme sigmoïde est une bonne simulation du phénomène. Remarquons aussi la progression linéaire pour des valeurs proches du seuil.

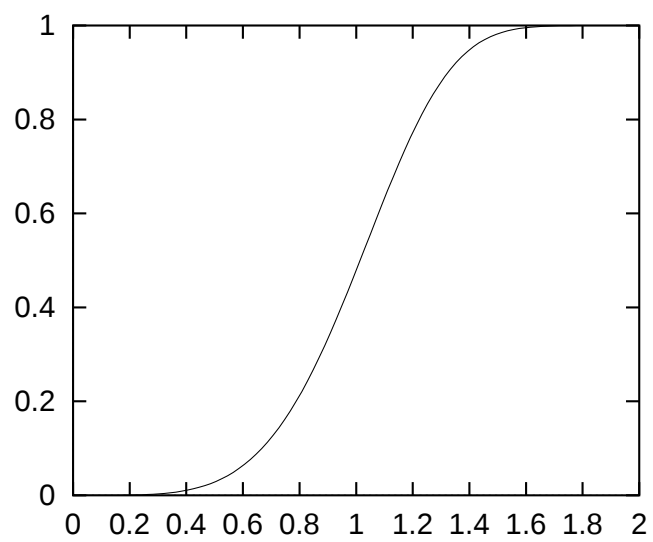


FIG. 2.10: Fonction psychométrique

Chapitre 3

Les modèles de vision

La modélisation des phénomènes régissant notre vision est un domaine capital en imagerie numérique. En effet, les temps de calcul en synthèse d'images sont très longs. Il serait alors utile d'avoir un outil permettant de ne calculer que ce que l'œil est capable de voir. Les études sur la qualité des images ont tout autant besoin de ces modèles. Comment savoir si un résultat est pertinent si l'algorithme d'analyse ne prend pas en compte les mécanismes du système visuel humain ?

A partir de deux images (dont l'une pourrait être la référence, par exemple), on applique le modèle. On obtient ainsi les deux images telles que les comprend le cerveau. Un calcul de distance est alors effectué entre celles-ci. Deux résultats sont possibles : soit avoir une image de distances complète, soit une valeur unique symbolisant l'écart entre les images. De plus, nous souhaiterions être capables d'obtenir une seule valeur à partir de la carte des distances, et inversement. La figure 3.1 représente la démarche.

En synthèse d'image, le calcul peut s'effectuer soit dans la scène, soit sur l'image affichée sur l'écran comme en analyse. Dans la seconde hypothèse, on rajoutera un dispositif permettant de connaître l'image affichée en fonction des caractéristiques précises du moniteur utilisé.

Beaucoup de solutions ont été proposées. La plupart ne prennent en compte que quelques phénomènes. C'est le cas de [10] ou encore [34]. Cependant, deux modèles sont plus complets. Ce sont les modèles de Sarnoff [15] et de Daly [7].

3.1 Le modèle de Sarnoff, un algorithme séquentiel

Ce modèle, défini au centre de recherche David Sarnoff, tente de prendre en compte les phénomènes se produisant dans le système visuel humain, de manière *séquentielle*. Toutefois, il est important de préciser que ce modèle est défini pour des images mono-chromatiques. Nous verrons plus loin que Meyer et Bolin ont proposé, dans [21], une évolution pour traiter la couleur. Mais, ce n'est pas sans difficultés.

Chaque étape repose sur des mécanismes physiologiques, permettant ainsi des résultats plausibles. La figure 3.2 représente l'architecture du modèle.

Stimuli Il s'agit des images de départ, en luminosité.

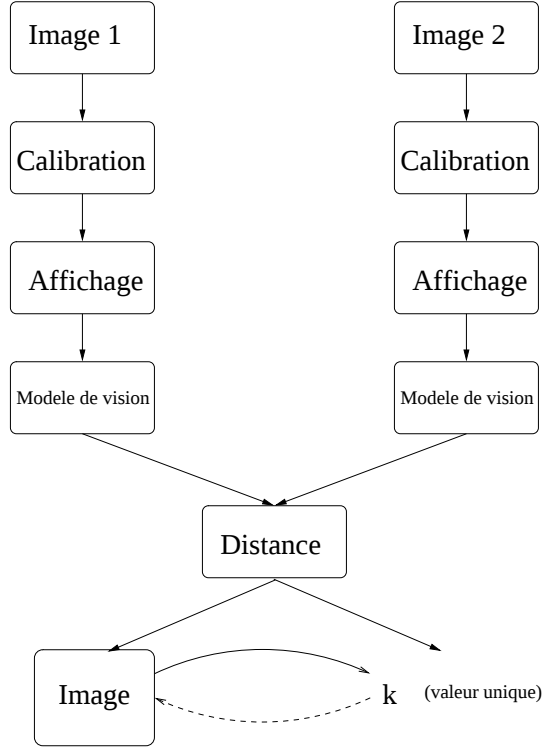


FIG. 3.1: Diagramme d'utilisation d'un modèle de vision

Système optique Le but de cette étape est de modéliser la cornée et le cristallin. On veut prendre en compte l'influence d'un point lumineux sur la vision des ses voisins. Ceci se fait par une fonction approximant le phénomène de cercle de confusion. Cette fonction, donnée par Weistheimer dans [35], est :

$$Q(\rho) = 0,952e^{-2,59|\rho|^{1,36}} + 0,048e^{-2,43|\rho|^{1,74}}, \quad (3.1)$$

avec $Q(\rho)$ l'intensité lumineuse en fonction de la distance ρ à un point de puissance unitaire.

Échantillonnage Après avoir obtenu l'image déformée par le système optique, il faut calculer l'image telle que la voit la rétine. La méthode choisie est aussi une convolution, mais par une gaussienne. Le poids affecté dépend de la région. Pour la *fovea*, la densité est de 120 pixels par degré. En dehors de la *fovea*, la densité décroît suivant l'excentricité.

$$d = \frac{120}{1 + 0,4e} \quad (3.2)$$

Malheureusement, Lubin ne donne pas de justification pour le choix des densités.

Réponse de l'œil aux bandes passantes de contraste Les intensités sont converties en contraste local. Lubin propose d'utiliser, pour cela, une pyramide laplacienne. On obtient

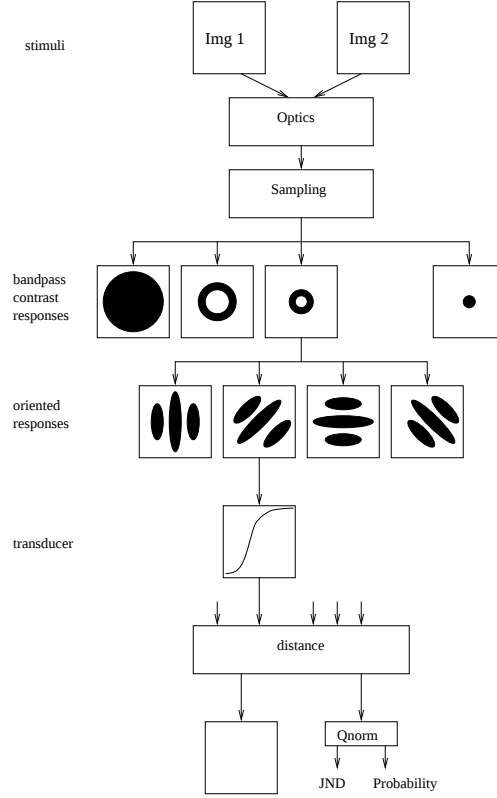


FIG. 3.2: Le modèle de Sarnoff

ainsi 7 niveaux de fréquences allant de 0,5 à 32 cycles par degré. La localité est obtenue en divisant par la valeur de la gaussienne située deux niveaux plus bas dans la pyramide. Le contraste local $\hat{c}_k(\vec{x})$ peut se formuler ainsi :

$$\hat{c}_k(\vec{x}) = \frac{I(\vec{x}) * (G_k(\vec{x}) - G_{k+1}(\vec{x}))}{I(\vec{x}) * G_{k+2}(\vec{x})}, \quad (3.3)$$

avec $\vec{x}$ un point de l'image, $I(\vec{x})$ l'intensité lumineuse après l'opération *Échantillonnage*, et G_k un noyau gaussien tel que :

$$G_k(\vec{x}) = \frac{1}{(\sqrt{2\pi}\sigma_k)^2} e^{-\frac{(x^2+y^2)}{2\sigma_k^2}} \quad \text{et } \sigma_k = 2^{k-1}\sigma_0. \quad (3.4)$$

Ce type de calcul est assez lourd. L'utilisation d'une transformation en ondelettes permet un gain en temps significatif pour une précision similaire (cf 3.1.1).

Réponse orientée Il s'agit maintenant de tenir compte de l'orientation dans l'image. Pour cela, Lubin calcule le contraste local donné plus haut suivant quatre directions : 0, 45, 90 et 135 degrés. Il semble que cela soit un bon compromis entre précision et temps de calcul.

Pour chaque direction, on a un couple d'opérateurs : la dérivée seconde d'une gaussienne orientée d'abord, et sa transformée de Hilbert ensuite. La réponse énergétique en fonction d'une certaine fréquence et d'une orientation est alors obtenue par la formule suivante :

$$e_{k,\theta}(\vec{x}) = (o_{k,\theta}(\vec{x}))^2 + (h_{k,\theta}(\vec{x}))^2, \quad (3.5)$$

où o est l'opérateur orienté et h , sa transformée de Hilbert. L'intérêt d'avoir un couple de filtres est d'être moins sensible à la position exacte des zones à forts gradients. En effet, la détection de ces zones par l'œil n'est pas au pixel près.

Transducteur C'est l'opération correspondant au phénomène dit de transduction visuelle (cf 2.4.4). De plus, on peut interpréter cela comme le seuil du contraste nécessaire à la détection.

Étalement Le résultat, fonction du nombre de cycles, est sensible jusqu'à un cycle par degré. Or, dans la *fovea*, la sensibilité maximale est à cinq cycles par degré. La solution proposée consiste à convoluer avec un disque de diamètre 5.

Distance A ce stade de l'algorithme, on a quatre pyramides de sept niveaux chacune, donnant le contraste dans l'image de départ en fonction des phénomènes psycho-visuels entrant en jeu. Il s'agit donc de comparer les résultats obtenus pour les deux images.

D'abord, les premiers niveaux de la pyramide sont "étirés" pour avoir la même taille que le niveau le plus bas. On obtient alors un vecteur de dimension 28 pour chaque pixel. Ensuite, on calcule la distance, dans l'espace L^Q , entre les pixels des deux images, ce qui nous donne une *image de distances*.

$$D(\vec{x}_1, \vec{x}_2) = \left\{ \sum_{i=1}^m [P_i(\vec{x}_1) - P_i(\vec{x}_2)]^Q \right\}^{\frac{1}{Q}} \quad (3.6)$$

Q prend généralement comme valeur 2, 4. Lubin ne donne pas d'explications à cela. Cependant, 2, 4 est une valeur aussi utilisée dans d'autres modèles.

Outre la carte de distances, une valeur unique peut-être générée. Cela permet d'avoir un descripteur de qualité d'images, et de comparer voire même de classer plusieurs images.

3.1.1 Améliorations

En synthèse d'images, le principal reproche que nous pouvons faire à ce modèle, est son approche mono-chromatique. Meyer et Bolin ont proposé, lors du *SIGGRAPH 98*, une évolution traitant la couleur [21]. Nous verrons aussi que les auteurs utilisent une transformée en ondelettes moins coûteuse que la pyramide laplacienne associée aux filtres gaussiens.

La première étape, *Système optique*, du modèle de Sarnoff est supprimée. A la place, on trouve un passage dans l'espace colorimétrique *SML*. L'étape *Échantillonnage* est aussi abandonnée.

Le contraste en fonction d'une bande de fréquence et d'une orientation est calculé à l'aide d'une seule transformation : *les ondelettes de Haar*. Pour une bande de fréquence, trois types

de contrastes sont détectés : horizontalement, verticalement et obliques. Cette dernière classe regroupe aussi bien les orientations à 45 ° qu'à 135 °. Le schéma 3.1 illustre cette répartition.

d3	d2	d3
d1	c	d1
d3	d2	d3

TAB. 3.1: Prise en compte des orientations avec les ondelettes de Haar.

$$\begin{aligned}
c_{l-1} \left[\frac{x}{2}, \frac{x}{2} \right] &= \frac{1}{4}(c_l[x, y] + c_l[x, y + 1] + c_l[x + 1, y] + c_l[x + 1, y + 1]) \\
d_{l-1}^1 \left[\frac{x}{2}, \frac{x}{2} \right] &= \frac{1}{4}(c_l[x, y] - c_l[x, y + 1] + c_l[x + 1, y] - c_l[x + 1, y + 1]) \\
d_{l-1}^2 \left[\frac{x}{2}, \frac{x}{2} \right] &= \frac{1}{4}(c_l[x, y] + c_l[x, y + 1] - c_l[x + 1, y] - c_l[x + 1, y + 1]) \\
d_{l-1}^3 \left[\frac{x}{2}, \frac{x}{2} \right] &= \frac{1}{4}(c_l[x, y] - c_l[x, y + 1] - c_l[x + 1, y] + c_l[x + 1, y + 1])
\end{aligned} \tag{3.7}$$

La grosse différence avec le modèle de Sarnoff tient dans la gestion de la couleur. Meyer et Bolin proposent de passer dans l'espace AC_1C_2 . Cela permet de prendre en compte l'aberration chromatique. Ainsi, sur le canal achromatique, la fonction *CSF* est celle donnée par Barten dans [4] et [5]. Sur les canaux en opposition rouge/vert et bleu/jaune, elle est construite à partir des données fournies par Mullen [22].

Ce sont les principales différences apportées par Meyer et Bolin. L'architecture du modèle reste la même. Ainsi, une image de distances ou une valeur sont calculées à partir des images correspondant aux différentes sensibilités fréquentielles et angulaires.

3.2 Le modèle de Daly, *Visual Difference Predictor*

Quoique destinée à prédire la qualité des images, l'approche proposée par Daly est assez différente. Elle s'appuie sur la physiologie mais n'est pas optimisée en temps. De plus, le calcul se fait sur l'image affichée. C'est à dire en fonction d'une certaine distance à l'écran. Ce modèle, également séquentiel, est structuré en trois étapes (figure 3.3) :

1. la *non linéarité de la réponse*,
2. l'utilisation d'une *fonction de sensibilité au contraste* (CSF) pour détecter les fréquences spatiales et les orientations,
3. la prise en compte des mécanismes : *masquage* et *fonction psychométrique*.



FIG. 3.3: Le modèle de Daly

3.2.1 Amplitude non linéaire

On ne travaille qu'en luminosité. Daly propose de modéliser la réponse rétinienne par l'équation suivante :

$$\frac{R(\vec{x})}{R_{max}} = \frac{L(\vec{x})}{L(\vec{x}) + (12.6L(\vec{x}))^{0.63}}. \quad (3.8)$$

R/R_{max} est la réponse normalisée en fonction de la luminosité L . Cette phase simulant la transduction visuelle est assez approximative. Daly ne donne pas de justifications quant à la valeur des constantes. La fonction sigmoïde utilisée dans le modèle de Sarnoff est plus proche des résultats obtenus par Bader dans [2].

D'autre part, nous pouvons de suite remarquer que le modèle n'intègre pas le système optique de l'œil. Ce phénomène pourtant indispensable peut être facilement rajouté en pré-traitement, grâce à une convolution gaussienne par exemple.

3.2.2 Fonction de sensibilité au contraste

Nous avons vu en 2.4.6 que nous sommes moins sensibles aux détails dans les zones à hautes fréquences. Pour chaque fréquence, il y a donc un contraste seuil au dessus duquel nous ne détectons rien. La fonction utilisée est en deux dimensions, ce qui permet de prendre en compte à la fois les fréquences et les orientations.

Nous pouvons remarquer sur la courbe que la sensibilité à 45° est bien inférieure à celles pour les directions horizontale ou verticale.

3.2.3 Mécanismes de détection

A partir de là, plusieurs phénomènes sont gérés. Premièrement, les fréquences spatiales sont repérées. Ensuite, le masquage est intégré au modèle. Troisièmement, Daly propose l'utilisation d'une fonction psychométrique modélisant la probabilité de détection du contraste. Enfin, ces résultats doivent être combinés, donnant ainsi une idée de la réponse perceptuelle en chaque pixel.

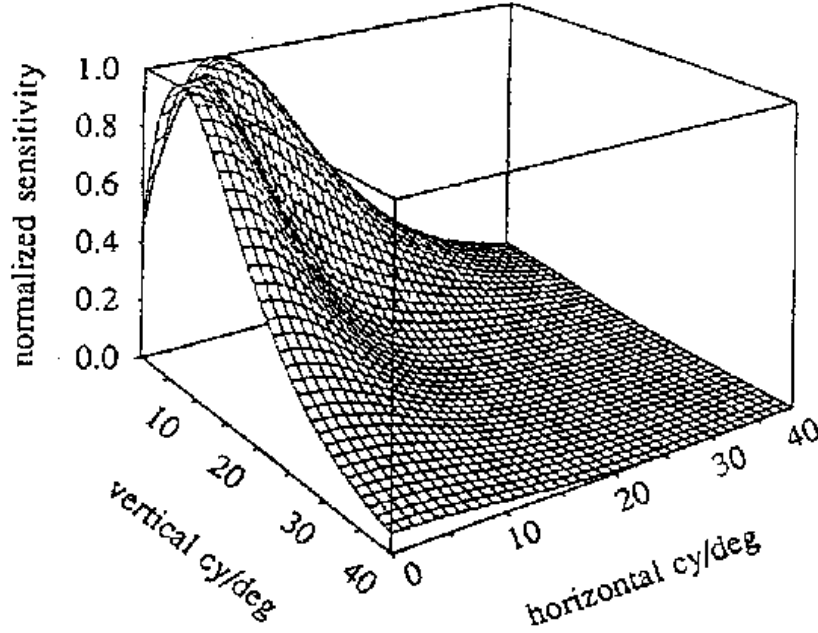


FIG. 3.4: Fonction de sensibilité au contraste en deux dimensions

Fréquences spatiales

Le contraste en fonction d'une plage de fréquence et d'une orientation est calculé grâce à une transformation nommée *Cortex Transform*, définie par Watson dans [33]. Elle se compose de deux filtres, l'un fréquentiel, *dom filter*, l'autre angulaire, *fan filter*.

– Dom filter :

$$\tilde{m}_0(u, v) = \left(\frac{\gamma}{f}\right)^2 e^{-\pi \left(\frac{r\gamma}{f}\right)^2} * \Pi\left(\frac{r}{2f}\right) \quad (3.9)$$

$$\tilde{m}_k(u, v) = \tilde{m}_0(2^k u, 2^k v) \quad (3.10)$$

Un premier filtre correspondant à une plage $[0; k]$ est d'abord calculé. La composante radiale est non linéaire. Ensuite, il suffit de faire la différence avec son suivant pour obtenir la sensibilité sur une plage $[k; k + 1]$ (cf. figure 3.6).

$$\tilde{d}_k(u, v) = \tilde{m}_k(u, v) - \tilde{m}_{k+1}(u, v) \quad (3.11)$$

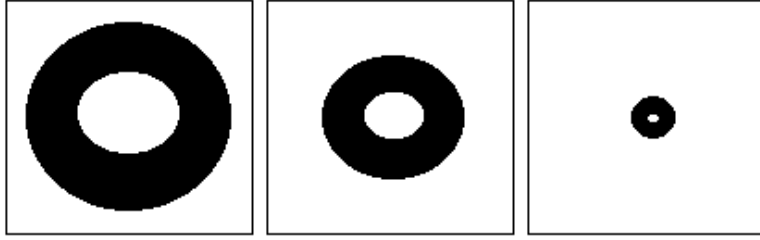


FIG. 3.5: Cortex Transform : filtre fréquentiel

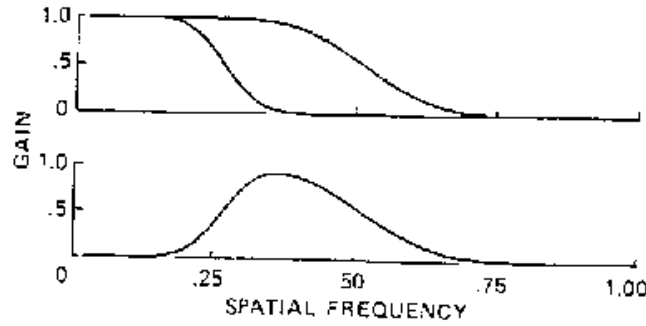


FIG. 3.6: Cortex Transform : composante radiale du filtre fréquentiel

– Fan filter : Soit N , le nombre de filtres¹.

L'écart angulaire de chaque filtre est donc de : $\theta_0 = \frac{180}{N}^\circ$.

Le filtre f peut alors se calculer suivant la formule suivante :

$$fan_f(\theta) = \frac{1}{2} \left\{ 1 + \cos \left[\frac{\pi |\theta - \theta_c(f)|}{\theta_0} \right] \right\} \quad \text{pour } |\theta - \theta_c f| \leq \theta_0 \quad (3.12a)$$

$$= 0 \quad \text{pour } |\theta - \theta_c f| > \theta_0, \quad (3.12b)$$

avec $\theta_c f$, l'orientation du sommet du filtre f , soit $\theta_c f = (f - 1)\theta_0 - 90$.

La figure 3.7 représente la projection sur un plan du filtre angulaire ainsi calculé.

Le filtre final n'est alors que la composition de ces deux opérateurs (figure 3.8) :

$$Cortex_{k,l}(\rho, \theta) = Dom_k(\rho) \bullet Fan_f(\theta). \quad (3.13)$$

¹Daly choisit $N = 6$

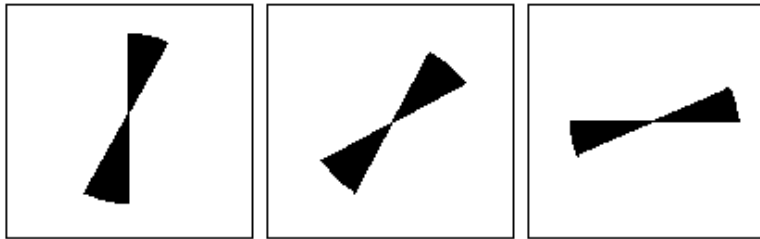


FIG. 3.7: Cortex Transform : orientations

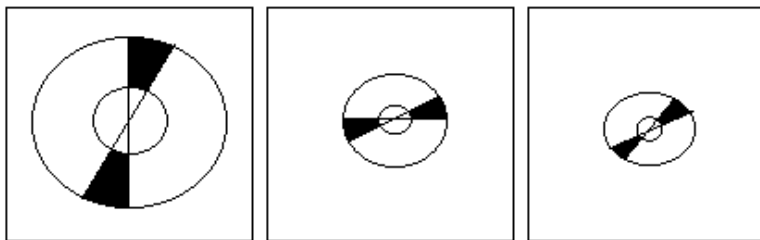


FIG. 3.8: Cortex Transform

Fonction de masquage

Daly définit l'information de masquage comme le produit de la fonction *CSF* par l'opérateur *Cortex* appliqué à l'image.

$$m(\rho, \theta) = I * csf(\rho, \theta) * cortex(\rho, \theta) \quad (3.14)$$

L'auteur introduit ensuite d'autres notions que nous ne détaillerons pas. Il s'agit de la modification du masquage suivant la phase et de l'effet d'apprentissage.

Fonction psychométrique

Le caractère aléatoire est enfin modélisé par la fonction psychométrique.

$$P(c) = 1 - e^{-(\frac{c}{\alpha})^\beta} \quad (3.15)$$

3.3 D'autres approches

3.3.1 Distance perceptuelle à base d'ondelettes

Gaddipatti et al. ont défini une distance perceptuelle entre images [10]. Elle repose sur la combinaison d'une transformation en ondelettes, celles de *Daubechies* en l'occurrence, avec la fonction *CSF* en deux dimensions.

Soit $W(m, \vec{x})$, le résultat de la transformation en ondelettes au niveau m au pixel $\vec{x}$.

Il faut d'abord établir une pondération sur chaque niveau. Ce poids représente le volume sous la surface *CSF*, pour une bande de fréquence (BF) donnée. Pour π , la fréquence maximale, on a :

$$C_m = \frac{\int_{BF_m} CSF \omega d\omega}{\int_{BF_m} \omega d\omega}, \quad \text{avec} \quad BF_m = \left(\frac{\pi}{2^m}, \frac{\pi}{2^{m-1}} \right). \quad (3.16)$$

On définit alors la sensibilité S au niveau m en $\vec{x}$ comme :

$$S(m, \vec{x}) = C_m W(m, \vec{x}). \quad (3.17)$$

La métrique perceptuelle n'est autre qu'une moyenne des différences de sensibilité.

$$M_p(A, B) = \frac{\sum_{m, \vec{x}} |S_A - S_B|^2}{N_h \times N_v} \quad (3.18)$$

3.3.2 Un tracé de rayons fréquentiel

Gary Meyer et Mark Bolin ont aussi proposé un algorithme de tracé de rayons en fonction des fréquences spatiales [20]. Il est basé sur la méthode de compression *JPEG*.

Le domaine est divisé en blocs de taille 8 par 8 pixels. L'idée principale est de trouver la représentation fréquentielle sur chaque bloc. Ceux-ci sont alors triés selon leur importance en fonction de la *CSF*. A un instant t , on tire des échantillons dans les blocs significatifs. En réitérant le processus, on obtient une image calculée selon les bonnes fréquences.

De plus amples explications peuvent être trouvées dans [20].

3.3.3 Le modèle de Watson

Nous parlerons, enfin, du modèle de vision défini par A. Watson et J. Solomon [34]. Très similaire au modèle de Sarnoff, nous retrouvons la même organisation séquentielle. Chaque image, **en luminance**, subit des transformations successives ; les résultats sont en-

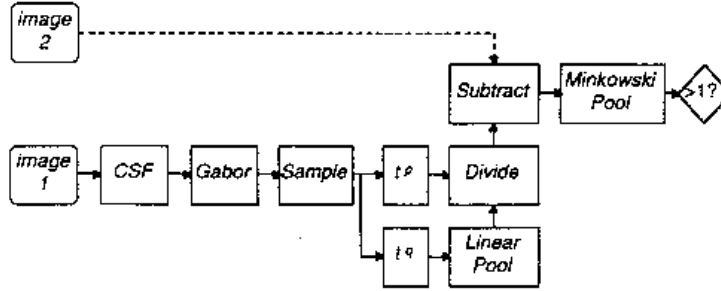


FIG. 3.9: Le modèle de Watson

suite comparés et sommés.

Remarquons l'absence de pré-traitement par un filtre passe-bas, simulant le système optique de l'œil. La première étape du modèle consiste à récupérer les informations fréquentielles suivant la fonction de sensibilité au contraste (*CSF*) combinée à une *transformation de Gabor* [9] [11]. Cette opération est un cas particulier de la transformée de Fourier fenêtrée, aussi appelée *Short-Time Fourier Transform*.

Celle-ci peut s'écrire :

$$X_g(t_x, \omega) = \kappa \int x(t)g(t - t_g)e^{-j\omega t}dt. \quad (3.19)$$

La plupart du temps, la fonction g est de la forme :

$$g(t) = 1, \quad t \in [t_1; t_2] \quad (3.20a)$$

$$= 0, \quad \text{sinon} \quad (3.20b)$$

On a alors bien la transformée de Fourier réduite à l'intervalle $[t_1; t_2]$.

Lorsque g est une fonction gaussienne sur $[t_1; t_2]$, on appelle X_g transformation de Gabor. Il est ainsi possible de connaître les fréquences localement dans l'image.

Ensuite, les auteurs construisent une pyramide, la taille de chaque niveau n étant divisée par un facteur 2^n . Comme avec la pyramide laplacienne du modèle de Sarnoff (cf. 3.1), cette opération permet d'obtenir les composantes fréquentielles de l'image.

Le signal est divisé en deux parties inhibitrice et excitatrice. Chacune suit une fonction non linéaire de la forme x^p . Différentes valeurs de p sont admises. Les auteurs se conforment au modèle de Teo et Heeger [30] [31] où p est égal à 2. De plus, le signal inhibiteur est convolué par un filtre gaussien. Il ne reste plus qu'à réunir les deux signaux pour obtenir le contraste relatif pour chaque image.

La distance finale est calculée dans l'espace L^4 .

Chapitre 4

Accélération des calculs : sonder l'image

La comparaison des images définie par Gilles Rougeron prenait en compte tout le domaine, chaque pixel étant calculé. Les opérations étant très coûteuses, cela occasionne un temps de calcul important.

Espace couleur	Temps
<i>LLAB</i>	28 mn
<i>LAB</i>	2,2 s

TAB. 4.1: Méthode globale : temps de calcul.

Cependant, nous pouvons nous demander s'il est vraiment utile de calculer tous les pixels. Lorsque l'on effectue un sondage, un panel de dix mille, voire même mille personnes permet d'avoir un idée très précise de l'opinion générale. La même technique ne pourrait-elle pas être utilisée dans ce cas précis ?

Il serait en effet intéressant d'obtenir un résultat global quasi-équivalent avec quelques milliers d'échantillons au lieu de 250000.

Cette technique a été développée dans [23]. Nous nous en inspirons.

NB : Il convient de préciser que les temps donnés ci-après correspondent à un calcul dans l'espace *LAB*.

4.1 Échantillonnage

4.1.1 Précision

Pour cela, nous utilisons la méthode de *Monte-Carlo* [11]. Soient deux images contenant T pixels de distance moyenne $\mathcal{M}$ et M défini comme l'écart

moyen sur N pixels choisis au hasard :

$$M = \frac{1}{N} \int_0^N \text{distance}(p) dp, \quad (4.1)$$

p étant la position dans l'image. Lorsque N est égal à T , on a donc : $M = \mathcal{M}$.

Le but est alors de déterminer le nombre de pixels N tel que M soit suffisamment proche de la valeur idéale $\mathcal{M}$. Pour cela, nous avons effectué plusieurs tests suivant le nombre d'échantillons tirés. La figure 4.1 montre l'écart en pourcentage suivant le nombre d'individus, et cela pour deux tailles d'images ¹ (200×200) et (512×512). D'autre part, ces tests ont été effectués sur cinq séries de tirages aléatoires. Nous montrons donc deux courbes représentant l'erreur maximale et l'erreur moyenne.

Les courbes obtenues montrent que quelle que soit la taille de l'image, nous avons besoin de 2000 individus pour obtenir une erreur inférieure à 2% et 5000 individus pour se ramener à moins de 1%. Nous pouvons remarquer que cela est en accord avec la théorie des sondages où le nombre de personnes interrogées ne dépend pas de la population totale, pour peu que celle-ci soit suffisamment grande.

Bien sûr, faire une moyenne de cinq séries n'est pas suffisant pour tirer des conclusions. Il faudrait, de plus, utiliser des images beaucoup plus variées. Nous pouvons cependant considérer que ces résultats donnent une bonne indication du nombre d'échantillons à tirer.

4.1.2 Temps de calcul

Le tableau 4.2 montre quelques exemples.

Taille	Méthode globale Temps de calcul	Sondage			
		Individus	Temps de calcul	Rapport	Précision
200×200	0,33s	2000	0,20s	60%	0,86%
200×200	0,33s	5000	0,60s	181%	0,68%
400×400	1,30s	2000	0,20s	15%	0,44%
400×400	1,30s	5000	0,60s	46%	0,89%
512×512	2,19s	2000	0,40s	19%	2,07%
512×512	2,19s	5000	0,70s	32%	0,46%

TAB. 4.2: Sondage de l'image : Temps de calcul

Le gain de temps est bien sûr meilleur pour les images de grande taille. Nous pouvons même noter que faire un sondage avec 5000 individus sur de petites images est plus long. Le sur-coût est dû aux calculs statistiques.

¹toujours la "Cornell Box"

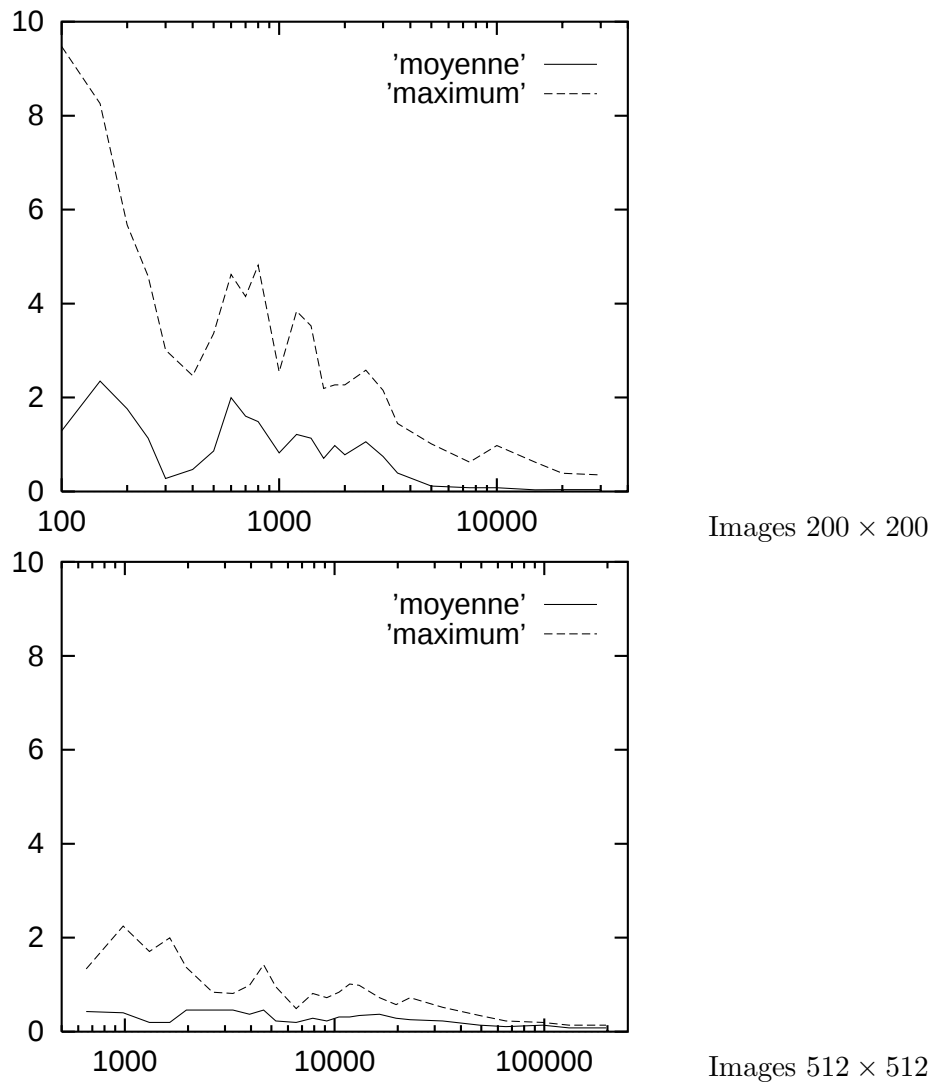


FIG. 4.1: Convergence du sondage d'une image.

4.2 Reconstruction de l'image de distances

Il est possible de calculer une valeur unique à partir d'une carte de distances.² Inversement, il serait très utile de pouvoir connaître les zones où la différence est importante (*resp.* insignifiante), sans pour autant être obligés de calculer tous les pixels. Pour cela, nous avons défini une méthode adaptative permettant de reconstruire une carte de distances en un temps meilleur.

L'algorithme est basé sur un découpage de l'image en *arbre quaternaire* (quad-tree). Nous avons un processus récursif en deux étapes :

1. Le calcul de l'erreur, par Monte-Carlo, sur la zone concernée, d'abord ;
2. le découpage de cette zone, lorsqu'une condition n'est pas remplie, ensuite.

Après un certain nombre de subdivisions, on obtient donc une image plus ou moins raffinée. La figure 4.2 illustre la progression de l'algorithme suivant les zones découpées.

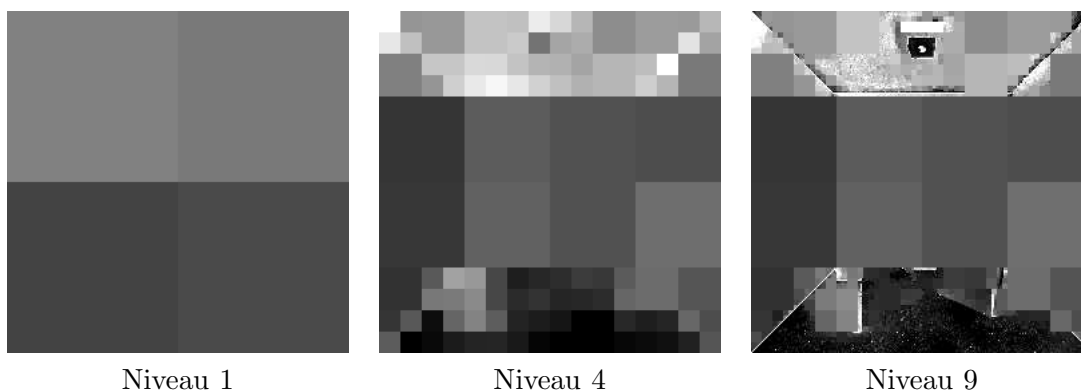


FIG. 4.2: Image de distances : Méthode adaptative

4.2.1 Nombre d'échantillons

Une fois définie la méthode de découpage, il nous faut connaître le nombre d'éléments qu'il va falloir tirer dans une zone. Nous avons choisi de déterminer ce nombre selon la taille de la zone à analyser.

Lorsque le domaine contient un nombre important de pixels, nous avons vu en 4.1 que 2000 individus était un bon choix. Ainsi, nous adoptons la règle suivante :

Pour T , la taille de la zone :

- Si $T > 10000$, on prend $N = 2000$;
- si $1000 < T \leq 10000$, on prend $N = 500$;
- si $T \leq 1000$, on prend $N = \lceil \frac{T}{3} \rceil$

Il peut y avoir des zones de deux pixels sur deux, voire même un seul pixel. Pour les petites surfaces, nous sommes donc obligés d'avoir un nombre d'échantillons proportionnel à leur taille.

²La moyenne en est bon exemple.

4.2.2 Quelle condition ?

La descente dans l'arbre, c'est à dire le raffinement du calcul, est sujette à un test. Deux approches sont possibles.

- La première solution consiste à répertorier les différentes zones de l'image en classes d'équivalence. D'après les tests que nous avons effectués (cf. chapitre 5), il semble que les observateurs discernent les différences entre deux images selon trois classes. Elles correspondent aux différences non perceptibles, perceptibles mais peu gênantes et enfin inacceptables.

Ainsi, pour une zone donnée, si une proportion suffisante de pixels appartient à la même classe, nous considérons cette zone comme homogène. Il est donc inutile de la subdiviser. Nous fixons arbitrairement le taux d'homogénéité à 95%. Dans l'image de distances, la moyenne calculée est affectée à tous les pixels de la zone considérée.

Le seuil départageant les deux premières classes est fixé à 5. Cette valeur a été déterminée d'après les tests effectués par des observateurs neutres (cf. 5.3.1). Le second seuil est fixé arbitrairement à 12. Nous obtenons ainsi une image de distances segmentée en trois composantes. Les zones non différenciables prennent la valeur 0, tandis que les zones aux différences inacceptables reçoivent le niveau de gris maximum, soit 255. Dans les zones intermédiaires, nous effectuons une interpolation linéaire, étendant ainsi l'intervalle $[5; 12]$ à $[1; 254]$.

- La deuxième approche est tirée de [18]. Nous voulons connaître les zones où la distance est à peu près constante. Il n'y a donc pas besoin de subdiviser ces zones là. Ainsi, nous considérons une zone comme constante lorsque les valeurs calculées sont toutes proches de la moyenne. Comme précédemment, nous nous donnons une tolérance de 5%. Soit X l'ensemble des pixels calculés.

$$Si \quad \frac{Card\{x \in X, x \in [x - \epsilon; x + \epsilon]\}}{|X|} < 0,95 \Rightarrow subdivision$$

Reste alors à fixer ϵ . Il semble logique de prendre la moitié de l'écart maximal entre deux couleurs imperceptibles, que nous avons fixé à 5, soit $\epsilon = 2,5$.

Nous avons implémenté les deux méthodes sur plusieurs scènes. Le tableau 4.3 montre les temps de calcul et le nombre de subdivisions. La figure 4.3, quant à elle, montre les images produites.

Si la deuxième condition semble permettre de meilleurs temps de calcul, nous préférons quand même la première qui possède l'avantage de segmenter le domaine en trois classes. Or, comme nous le verrons au chapitre 5, les différences de couleurs peuvent être *imperceptibles*, *visibles mais acceptables* ou *inacceptables*.

Il faut préciser que nous essayons d'améliorer des temps de calculs déjà très bas. Il est donc normal que, dans certains cas, les temps obtenus soient supérieurs à ceux avec la méthode globale. Dans l'optique future où un modèle de vision ainsi qu'un espace colorimétrique performant seraient utilisés, nous pensons que notre approche donnera de bien meilleurs résultats.

Nous avons aussi effectué quelques tests avec l'espace *LLAB* en utilisant la première condition de subdivision. Les temps obtenues (tableau 4.4) sont là très encourageants.

	Méthode	Temps	Nombre de zones subdivisées
Image 1 512×512	Méthode globale	2,2 s	—
	Première condition	2,4 s	4929
	Deuxième condition	0,7 s ³	440
Image 2 512×512	Méthode globale	2,7 s	—
	Première condition	1,1 s	2251
	Deuxième condition	1,7 s	2782
Image 3 400×400	Méthode globale	1,3 s	—
	Première condition	1,8 s	4092
	Deuxième condition	1,7 s	3757

TAB. 4.3: Temps de calcul selon la condition de subdivision

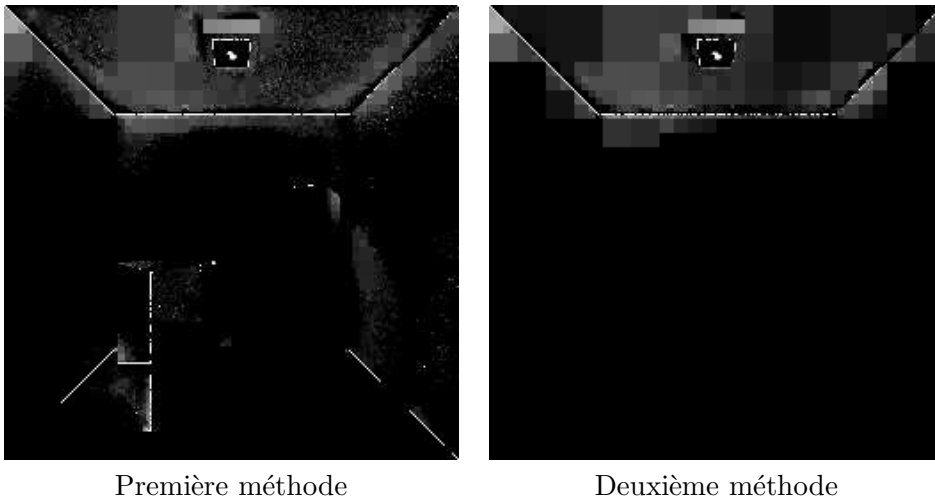


FIG. 4.3: Condition de subdivision

4.3 Une pseudo-distance

La distance que nous calculons, si elle n'en est pas une au sens mathématique, comporte quand même quelques similitudes avec la définition théorique.

1. *Réflexivité* : Il va de soi que la distance pixels à pixels entre deux images identiques est nulle. Notre distance est bien réflexive.
2. *Symétrie* : Nous calculons des écarts entre pixels en valeur absolue. Là encore, l'ordre des images n'a aucune influence sur le résultat.
3. *Transitivité* : Nous sommes, à l'heure actuelle, incapables de prouver la transitivité ou non de notre solution. Cependant, nous avons effectué quelques tests sur une série d'images de synthèse où nous modifions progressivement la scène.

Nous voyons, dans le tableau 4.5, les distances calculées entre les différentes images.

Tous les pixels		28 mn
Méthode adaptative	Scène 1	7 mn 47 s
	Scène 2	6 mn 10 s
	Scène 3	11 mn 17 s

TAB. 4.4: Temps de calcul en *LLAB*

N °	Différences avec la n ° 1	Distances avec l'image n °		
		1	2	3
1	—	—	—	—
2	sans anti-aliasage	0,72	—	—
3	sans anti-aliasage, source ponctuelle	1,20	0,54	—
4	sans anti-aliasage, source ponctuelle, texture de la table	1,36	0,71	0,17
5	sans anti-aliasage, source ponctuelle, texture de la table	1,33	0,68	0,14

TAB. 4.5: Transitivité : caractéristiques des images

Nous remarquons que l'absence d'anti-aliasage produit une distance 0,72 entre nos deux images. De même, la présence d'une source ponctuelle entre les images 2 et 3 induit une distance de 0,54. La distance entre les images 1 et 3 est de 1,20. Et, $0,72 + 0,54 = 1,26 \sim 1,20$. Le même phénomène se produit si on modifie la texture de la table :

$$1,36 \sim 1,20 + 0,17 \sim 0,72 + 0,54 + 0,17.$$

Bien sûr, cela ne prouve rien. Il semble cependant que les modifications de la scène se répercutent transitivement dans le calcul de la distance.

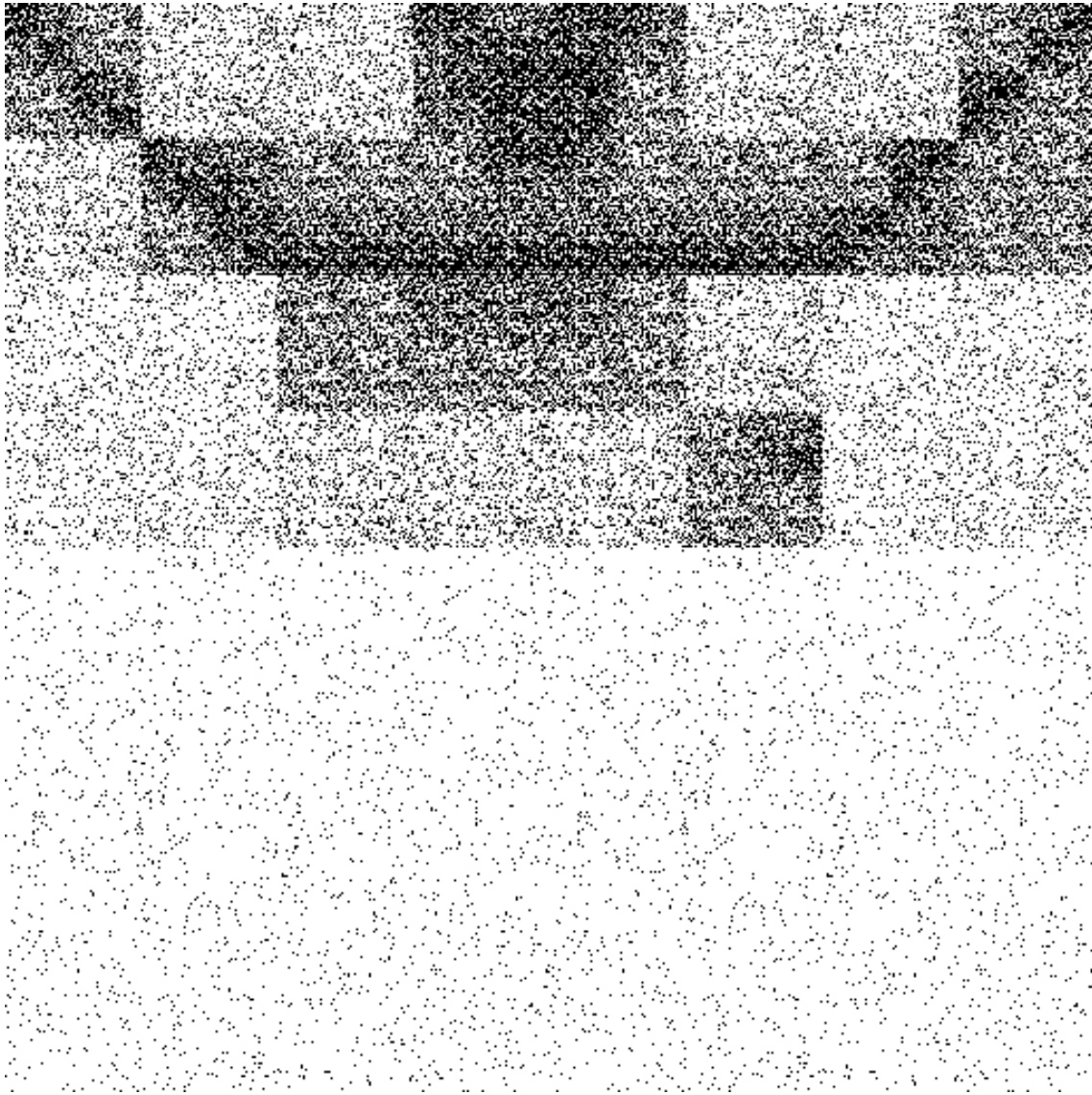


FIG. 4.4: Points effectivement calculés

Chapitre 5

Expérimentation

Après avoir défini les bases d’une distance entre images, il nous est apparu indispensable de mettre en place une procédure de test pour connaître la robustesse de notre modèle d’une part, et pour nous permettre d’évaluer les constantes que nous utilisons d’autre part.

Pour cela, nous avons fait appel à des observateurs neutres au sein de l’école. Sur un total de vingt personnes, il y avait dix spécialistes en synthèse d’images. Les dix autres personnes n’avaient pas de connaissances particulières dans le domaine.

Il convient cependant de préciser que ces tests ne sont là qu’à titre indicatif. Nous devons d’abord admettre que nous ne sommes pas spécialistes en la matière et la théorie de la détection est une discipline à part entière qui demande des bases solides. Il nous était donc difficile d’organiser une véritable procédure de test à grande échelle dans le temps qui nous était imparti.

Nous pensons cependant que ces résultats peuvent être une bonne indication de la validité de la méthode.

5.1 Trois tests

Nous avons soumis les observateurs à trois expériences successives.

Dans la première, nous affichions différents couples d’images pendant **quatre secondes** et cela répété cinquante fois. Pour chaque couple, nous posions la question : “*Y a-t-il des différences dans l’image ?*” La personne avait alors le choix entre trois réponses :

1. Les images sont *semblables*;
2. les images sont *peu différentes*;
3. les images sont *différentes*.

La distinction entre *peu différentes* et *différentes* n’était pas définie ; à chacun de déterminer ce qu’il entendait par différences non négligeables. En outre, les couples d’images étaient choisis aléatoirement parmi quatre séries de neuf.

Ce test comporte deux avantages. Il recueille d’abord la première impression de l’observateur, celui-ci n’ayant pas le temps d’analyser les images. Ensuite, nous pensons ainsi connaître la distance maximale entre deux images indifférenciables.

Pour la deuxième expérience, nous avons repris un test déjà effectué à l'Institut d'Ingénierie de la Vision de Saint-Etienne. Il consistait à affecter des classes d'équivalence. Nous affichions 9 images (une image traitée huit fois différemment plus la référence). L'observateur devait alors donner des numéros aux images suivant l'écart avec l'image de base. Celles n'ayant pas de différences visibles recevaient la note 1, puis 2 pour les images peu distinctes, etc. Les observateurs avaient, de plus, le choix de constituer autant de classes qu'il le désirait, pouvant aussi s'arrêter à deux.

Là encore, le plus important pour nous était de connaître toutes les images qui recevraient la note 1.

En ce qui concerne le troisième test, il s'agissait de choisir entre plusieurs images de distances calculées avec différents seuils (figure 5.1). A partir de deux images affichées, l'ob-



FIG. 5.1: Troisième test

servateur devait choisir l'image de distances qui correspondait le mieux aux différences qu'il voyait.

5.2 Les images

Nous avons choisi comme support deux scènes synthétisées (une pièce meublée et la *Cornell Box*) et deux images classiques en traitement et analyse (Lenna et les poivrons).

NB : Les images que nous montrons ne sont là qu'à titre indicatif. La qualité d'impression n'est pas suffisante pour permettre une quelconque preuve.

La pièce a été synthétisée grâce au logiciel de tracé de rayons *YART*, développé au sein du laboratoire. Chaque image comportait quelques modifications tant sur la méthode de rendu

que sur la géométrie de la scène.

	Méthode de Rendu	Différences
1	Monte-Carlo	Anti-aliasage
2	VEI	
3	VEI	sans 3 ^e chaise
4	VEI	autre tableau sur le mur
5	Terme ambiant + BRDF de Schlick	
6	Terme ambiant + BRDF de Schlick	source ponctuelle
7	VEI	autre texture de marbre
8	Terme ambiant + BRDF de Schlick	point de vue rehaussé de 5cm
9	Terme ambiant + BRDF de Schlick	angle de roulis de 2 °

TAB. 5.1: Caractéristiques de la pièce meublée

Pour la *Cornell Box*, nous avons utilisé le tracé de rayons spectral en cours de développement au laboratoire. Nous avons seulement fait varier les paramètres contrôlant l'erreur commise. Enfin, les images de Lenna et des poivrons ont subi les mêmes traitements.

	Traitement
1	Image de référence
2	Quantification couleur avec 20 niveaux
3	Quantification couleur avec 8 niveaux
4	Légère augmentation de contraste
5	Pixelisation (3×3)
6	Flou
7	Forte augmentation de contraste
8	Plaquage de l'image sur une surface gaussienne
9	Bruit aléatoire

TAB. 5.2: Caractéristiques des images photographiques

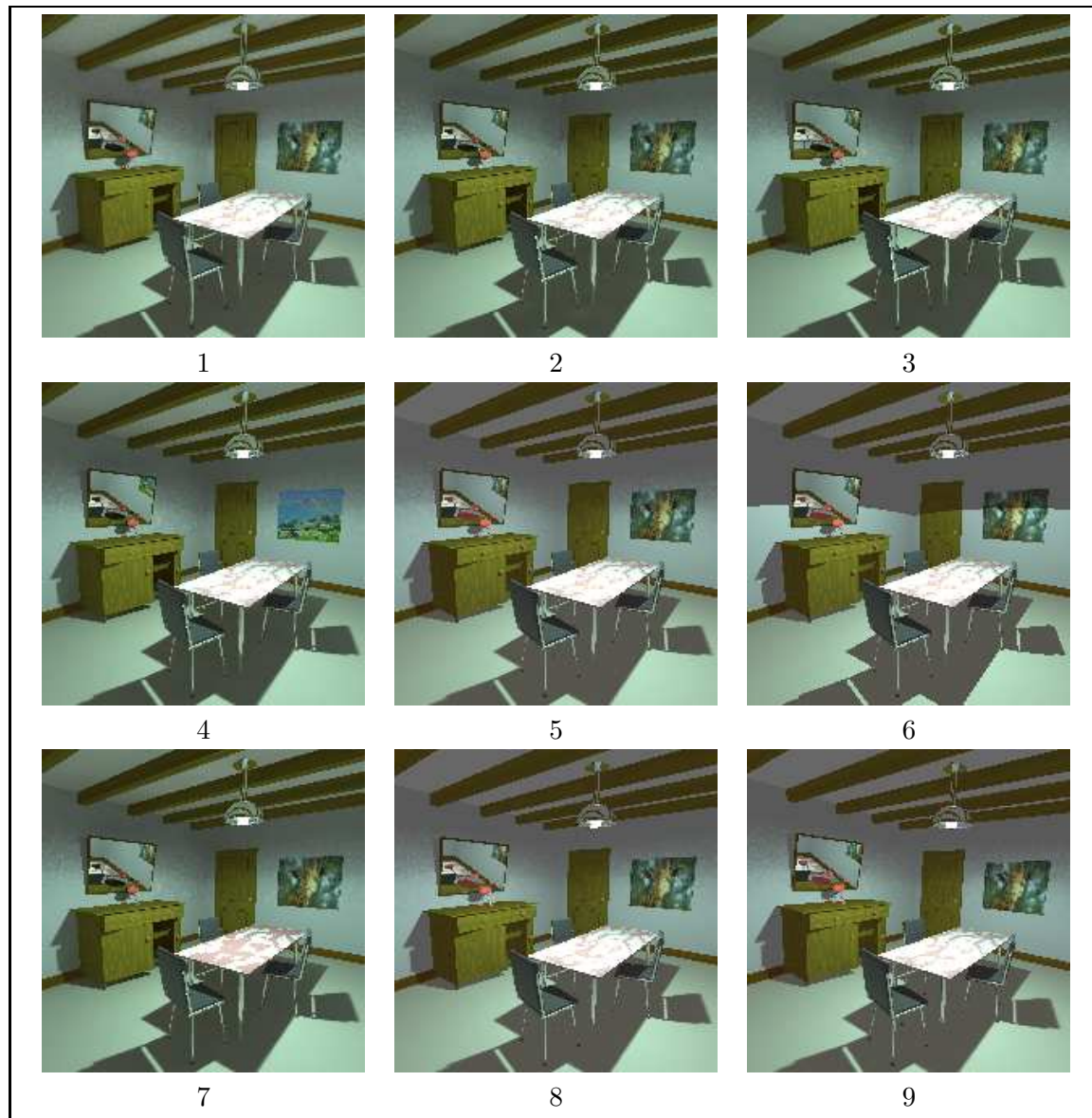
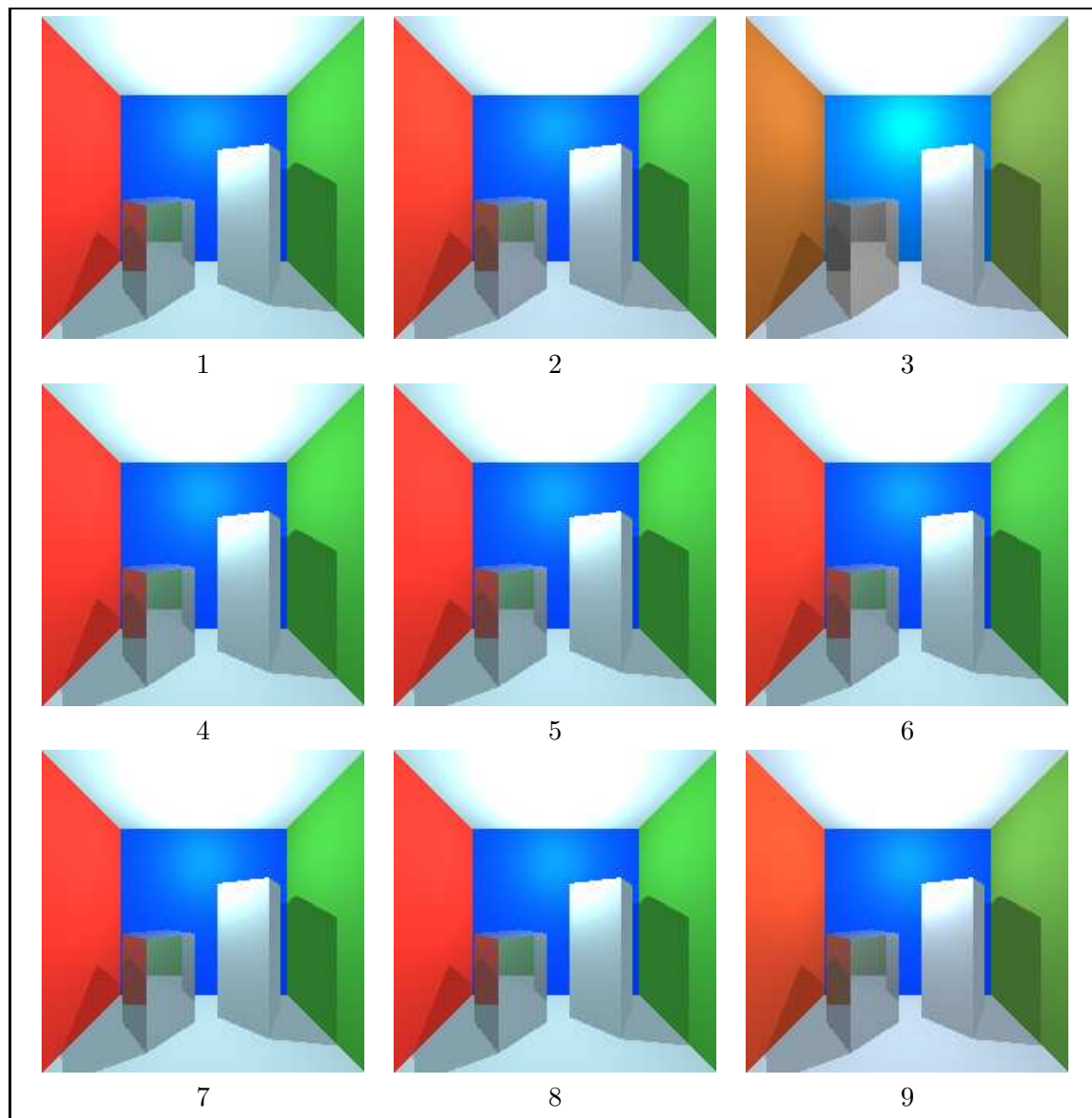


FIG. 5.2: Pièce meublée

FIG. 5.3: *Cornell Box*

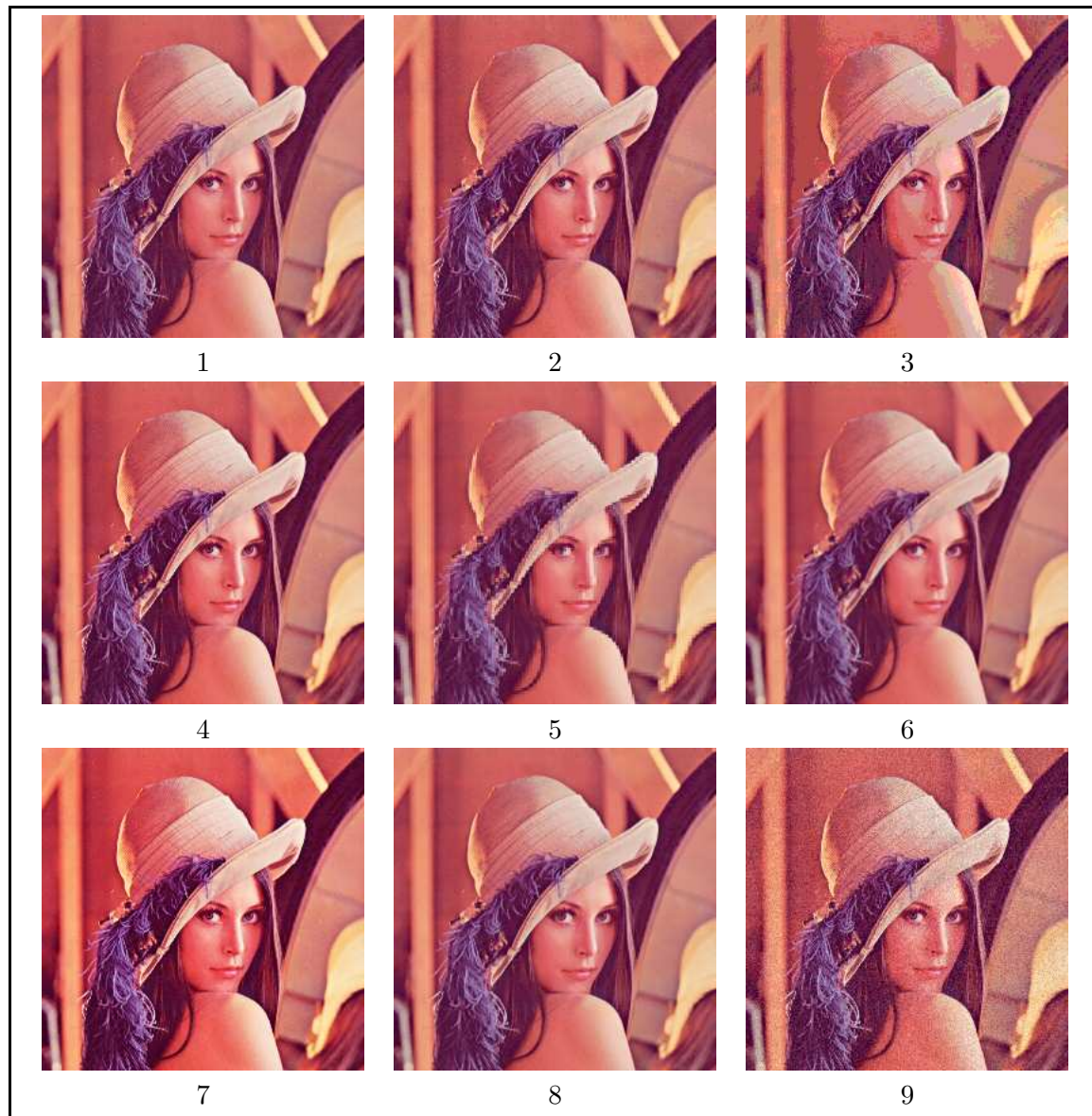


FIG. 5.4: Lenna

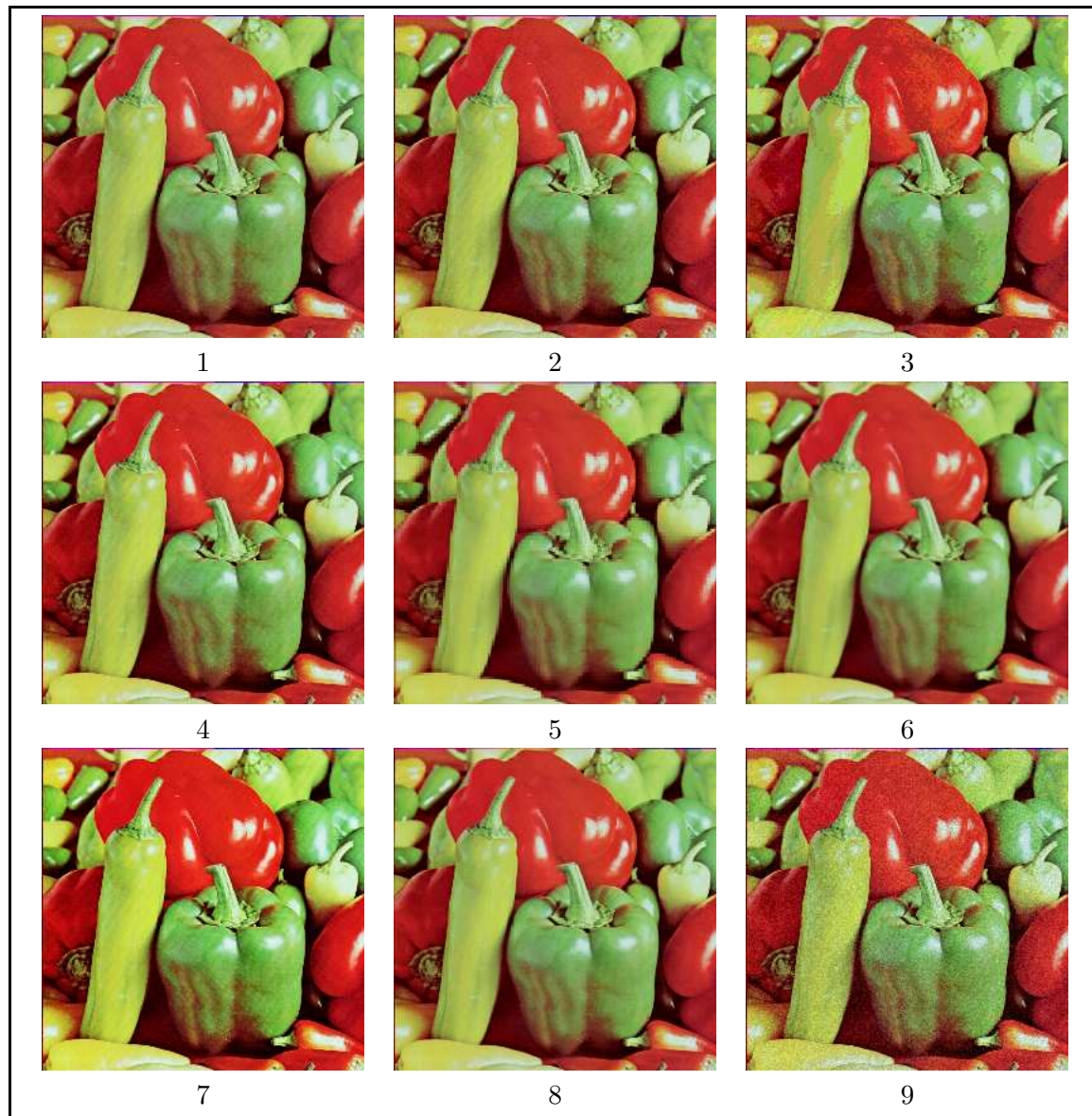


FIG. 5.5: Poivrons

5.3 Résultats

5.3.1 Première expérience

Nous demandons de comparer une image toujours par rapport à la référence. Celles-ci sont la numéro 2 pour la pièce, la 8 pour la *Cornell Box* et la 1 pour Lenna et les poivrons.

NB : Le choix des images étant aléatoire, certaines n'ont jamais été affichées. Au contraire, d'autres l'ont été plus souvent.

Les tableaux 5.3 à 5.6 montrent les réponses données par les observateurs en pourcentage. En gras se trouvent les images dont une réponse a été choisie à plus de 50%, avec les distances à l'image de référence correspondantes. Nous considérons que les images dont les réponses sont partagées, ne sont pas pertinentes pour notre étude.

Pièce meublée

N°	Semblables	Peu différentes	Différentes	Distance
1	51,67	31,67	16,67	4,42
3	62,5	17,5	20	1,53
5	20	42,5	37,5	6,19
6	50	20	30	7,03
7	50	30	20	1,45
8	30	42,5	27,5	8,96
9	15	30	55	13,33

TAB. 5.3: Résultats de la première expérience sur la pièce

Nous pouvons noter que les images 3 et 7 sont vues semblables à la référence. Cela est logique puisque l'absence d'une chaise ou le changement de texture n'affecte qu'une petite partie de la surface. Les distances calculées sont faibles, ce qui est correct.

En ce qui concerne l'image 11, les observateurs la voient comme distincte de la numéro 2. Le changement de méthode de rendu et l'angle de roulis génère bien une image différente. Là encore, la valeur calculée semble en corrélation avec les observations.

Cependant, il semble que le changement brusque de contraste provoqué par la source ponctuelle (image 6) soit moins détecté. En quatre secondes, la personne n'a pas le temps de voir cela. Nous verrons pour le prochain test (cf. 5.3.2) que ce défaut est bien diagnostiqué lorsque l'observateur dispose du temps nécessaire.

Pour l'image 1, les différences occasionnées avec la numéro 2 proviennent essentiellement de l'anti-aliasage. Tout en restant voisine, elle comporte quand même de légères différences. Nous disons qu'elle est proche de l'écart maximal entre deux images semblables.

Cornell Box

Les images spectrales, dont les seules variations portent sur la couleur, sont bien diagnostiquées. La numéro 3, dont la distance est très grande, est largement classée comme différente.

N°	Semblables	Peu différentes	Différentes	Distance
1	75	20	5	0,15
2	65	35	0	4,30
3	5	0	95	29,36
4	72,5	17,5	10	0
5	68,33	30	1,67	0,49
6	85	15	0	2,93
7	82,5	17,5	0	0,08
9	20	45	35	11,37

TAB. 5.4: Résultats de la première expérience sur la *Cornell Box*

Nous pouvons noter que 5 semble être encore le seuil de détection (image 2). Remarquons aussi une anomalie pour l'image 4 : certains observateurs ont vu des différences alors qu'il n'y en avait pas. Nous pensons que ceux-ci, étant prévenus que chaque image était générée différemment, ont cherché à forcer leur réponse.

Images photographiques

N°	Semblables	Peu différentes	Différentes	Distance
3	5	5	90	5,20
4	52,5	45	2,5	1,87
5	5	7,5	87,5	2,39
6	30	35	35	1,97
7	10	45	45	5,46
8	50	50	0	3,12
9	5	10	85	2,33

TAB. 5.5: Résultats de la première expérience sur Lenna

Les résultats sur les images non synthétisées sont moins probants. L'effet de pixelisation, très désagréable à l'œil, n'est pas reconnu comme gênant par notre méthode. Les valeurs calculées sont, en effet, assez faibles (image 5).

La transformation numéro 8 (projection gaussienne) semble ne pas gêner les personnes testées. Cela est plus marqué pour les poivrons. Or, d'après nos calculs, c'est avec Lenna que cette transformation semble générer le moins d'erreurs.

Le bruit aléatoire (image 9) dégrade beaucoup la compréhension de Lenna selon les observateurs. Néanmoins, notre méthode nous donne une distance assez faible. Cela montre que la solution n'est pas encore parfaite.

Cependant, il y a quand même des motifs de satisfaction avec les images photographiques. Les résultats avec la quantification couleur sont corrects. Quelle que soit l'image, la réduction en vingt couleurs est moins coûteuse que celle ne gardant que huit niveaux (images 2 et 3). Les calculs sur le rehaussement de contraste donnent aussi de bons résultats. Si une faible

N°	Semblables	Peu différentes	Différentes	Distance
2	81,67	18,33	0	2,18
3	5	15	80	5,80
4	47,5	37,5	15	2,76
5	5	17,5	77,5	3,51
6	45	33,33	21,67	2,95
7	10	50	40	8,22
8	72,22	27,78	0	4,09

TAB. 5.6: Résultats de la première expérience sur les poivrons

augmentation n'est pas gênante à l'œil (image 4), une forte élévation rend l'image différente de sa référence (aussi bien pour les observateurs que pour le calcul, image 7).

5.3.2 Deuxième expérience

Les observateurs n'ont pas de limite de temps. En affectant des notes à chaque image, ils établissent un ordre de distance. Nous cherchons alors à connaître l'ordre moyen sur chaque série d'images. Cependant, le nombre de classes n'est pas fixe. Par exemple, voici deux séquences de réponses données pour Lenna :

$$\{1, 2, 4, 8\} < 7 < 6 < 5 < \{3, 9\}$$

$$\{1, 2, 4, 7, 8\} < \{3, 5, 6\} < 9$$

Les chiffres entre accolades représentent les images que les observateurs n'ont pas pu départager. Nous décidons d'affecter des poids à chaque image. Ainsi, pour la première réponse, les images 1,2,4 et 8 reçoivent le poids 1. L'image 7 reçoit le poids 5. Le poids 6 est donné à l'image 6. Enfin, les images 3, 5 et 9 reçoivent respectivement 7, 8 et 8.

Pour la seconde réponse, les poids respectifs pour les images de 1 à 9 sont :

$$1, 1, 6, 1, 6, 6, 1, 1, 9.$$

Le tableau 5.7 montre les poids moyens reçus par chaque image.

Série	Images								
	1	2	3	4	5	6	7	8	9
Pièce	1	1,77	3,33	5,88	4	6,77	3,33	3,88	4,33
<i>Cornell Box</i>	1,72	2,16	8,72	1,61	1,55	2,66	1,11	1	7,33
Lenna	1	1,41	6,70	1,17	7,23	5,70	4,58	1,76	6,52
Poivrons	1	1,64	7,05	1,29	7,47	5,47	4,35	1,82	5,76

TAB. 5.7: Poids affectés aux images

Il nous est alors facile d'obtenir l'ordre moyen en classant les images suivant leurs poids. Par exemple, cela donne pour la pièce : $1 < 2 < \{3, 7\} < 8 < 5 < 9 < 4 < 6$.

Série	Images								
	1	2	3	4	5	6	7	8	9
Pièce	–	4,45	4,72	5,19	7,22	8,12	4,88	10,07	13,21
<i>Cornell Box</i>	0,15	4,30	29,36	0	0,49	2,93	0,08	–	11,37
Lenna	–	1,89	5,20	1,87	2,39	1,97	5,46	3,12	2,33
Poivrons	–	2,18	5,80	2,76	3,51	2,95	8,22	4,09	3,11

TAB. 5.8: Distances calculées

Calculons maintenant les distances.

De la même manière, nous obtenons un ordre, cette fois-ci calculé. D'autre part, nous regroupons dans un même ensemble les valeurs proches.

Pièce meublée

Test	$1 < 2 < \{3, 7\} < 8 < 5 < 9 < 4 < 6$
Calcul	$1 < 2 < \{3, 7\} < 4 < 5 < 6 < 8 < 9$

TAB. 5.9: Ordonnancement des images de la pièce

Si la distance donne des résultats en corrélation avec les tests pour les images proches, il n'en est pas de même pour les images plus éloignées. Les images 4 et 6 sont très distinctes de la référence pour les observateurs ; cela a cause des forts changements de contraste autour du tableau pour la 4 et au niveau de l'ombre pour la 6. Cependant, notre modèle ne prend pas en compte la sensibilité au contraste. Au contraire, les images 8 et 9 qui ont une géométrie décalée, provoque une erreur importante. Or, l'œil n'est pas sensible à la position exacte des objets. Les solutions vues au chapitre 3 sont donc indispensables pour avoir une meilleure précision.

Cornell Box

Test	$8 < 7 < \{1, 4, 5\} < \{2, 6\} < 9 < 3$
Calcul	$8 < \{1, 4, 5, 7\} < 6 < 2 < 9 < 3$

TAB. 5.10: Ordonnancement des images de la *Cornell Box*

Pour les images calculées selon le modèle spectral, les résultats sont bons (tableau 5.10). Nous obtenons quasiment le même ordre que celui donné par les observateurs.

Images photographiques

Quelle que soit la scène, les ordres sur les distances calculées sont les mêmes. Idem, pour les ordres donnés par les observateurs. Nous remarquons que les images 5 et 9, correspondant respectivement à une pixelisation et à un bruit aléatoire, sont classées parmi les plus

Test	$1 < \{2, 4\} < 8 < 7 < 6 < \{9, 3\} < 5$
Calcul	$1 < \{2, 4, 6\} < \{5, 9\} < 8 < \{3, 7\}$

TAB. 5.11: Ordonnancement des images de Lenna

Test	$1 < 4 < \{2, 8\} < 7 < \{6, 9\} < \{3, 5\}$
Calcul	$1 < 2 < 4 < \{6, 9\} < 5 < 8 < 3 < 7$

TAB. 5.12: Ordonnancement des images des poivrons

différentes. Or, les distances calculées sont assez faibles.

Au contraire, l'image 8, correspondant à un plaquage de l'image sur une surface gaussienne est mal interprétée. La distance calculée est en effet beaucoup trop forte par rapport à la vision de nos observateurs. C'est aussi le cas, dans une moindre mesure de l'image numéro 7.

Si les premières transformations affectent la qualité de l'image, les secondes ne provoquent pas de changement brusque de contraste dans l'image obtenue. C'est, à notre sens, la raison pour laquelle ces images semblent moins dégradées pour les observateurs.

Il est, une fois de plus, clair qu'une solution modélisant les changements de contraste est nécessaire.

5.3.3 Troisième expérience

Ce test a été effectué sur huit couples d'images successifs. Les trois premiers concernaient la pièce. A l'image numéro 3, étaient comparées les 5, 6 et 7, c'est à dire un changement de méthode de rendu, une source ponctuelle et un changement de texture. Ensuite, deux images de la *Cornell Box*, l'une très proche, l'autre distante devaient être comparées à la référence. Pour Lenna, nous avons choisi la quantification couleur à vingt niveaux et la projection gaussienne. Enfin, pour les poivrons, c'était la deuxième quantification à huit niveaux.

Quinze personnes ont participé à l'expérience. Outre le choix d'une image de distances, elles avaient la possibilité de ne pas répondre si aucune image ne leur convenait, ou de répondre que l'image de distances devait être plus blanche ou plus noire.

Les cartes de distances ont été calculées selon la méthode globale (pour tous les pixels) en *LAB*, la plage de valeurs [*imperceptible*; *inacceptable*] étant ramenée sur [0; 255]. Nous avons pris 12 comme écart minimum inacceptable entre deux couleurs, et 2, 3, 3,5, 4,5 et 7 pour l'écart couleur maximal entre deux couleurs imperceptibles, construisant ainsi cinq images de distances différentes.

Le tableau 5.13 montrent les résultats. La colonne *Mode* représente la réponse la plus fréquente, c'est à dire le numéro de l'image de distances choisie (5 étant l'image la plus noire). Le signe + représente la réponse "*L'image devrait être plus noire*". La réponse "*Aucune image ne correspond*" est indiquée par le signe $\emptyset$. La colonne *Nombre* indique le nombre d'occurrences du mode. Enfin, la dernière colonne donne le nombre d'observateurs qui n'ont pas pu répondre.

Les réponses les plus fréquentes sont 4 (seuil : 4,5), 5 (seuil : 7) et +. Les observateurs trouvent une plus grande corrélation entre les cartes de distances foncées et les différences qu'ils discernent dans les images à analyser. Conformément aux résultats obtenus lors des deux

	Mode	Nombre	Sans réponse
1	4	7	1
2	4	5	
3	5	8	
4	+	4	2
5	4	5	
6	5	6	1
7	+	10	5
8	$\emptyset$	5	5

TAB. 5.13: Troisième test

premiers expériences, 5 semble être une valeur raisonnable comme seuil d'imperceptibilité entre deux couleurs dans l'espace *LAB*.

Les tests 7 et 8 mettent en évidence les limites de notre méthode. La projection gaussienne sur Lenna génère une distance importante dans la plume comme on peut le voir sur l'image de distances (figure 5.6). Or, aucun observateur n'a détecté cette différence, la plupart qualifiant même l'image comme identique à la référence. Cela s'explique simplement par le fait que la plume est une zone très texturée, avec, de plus, une fréquence spatiale élevée. Comme nous l'avons vu en 2.4.6, page 14, l'œil n'est pas sensible à toutes les fréquences de la même manière, qui plus est les fréquences élevées. Nous voyons bien là l'intérêt d'un modèle de vision.

Les résultats pour la quantification couleur en huit niveaux sur les poivrons sont médiocres. En effet, à la grande majorité, les observateurs n'ont pas trouvé d'image significative. De plus, les différences de couleurs qu'ils voyaient ne se situaient pas aux mêmes endroits que ceux donnés par le programme. Ainsi, les personnes testées ont été sensibles à la dégradation des couleurs sur les poivrons du premier plan. Or, les écarts couleurs les plus importants se situent sur la gauche, au second plan (cf. figure 5.7).

Il nous est cependant difficile de dire si l'œil humain effectue une segmentation spatiale de l'image pour donner un poids plus important aux objets de premier plan. Il se pourrait tout aussi bien que nous examinions avec plus de précision le centre de l'image par rapport aux extrémités. Dans l'incertitude, des tests complémentaires nous semblent donc utiles.

5.3.4 Conclusion

Ces expériences nous ont apporté deux enseignements importants. Primo, en ce qui concerne l'écart *LAB* maximal entre deux couleurs imperceptibles, 5 semble être une bonne valeur. Il nous reste alors à fixer le deuxième seuil. Secundo, la prise en compte des phénomènes visuels sont indispensables.



FIG. 5.6: Calcul de distance : projection gaussienne

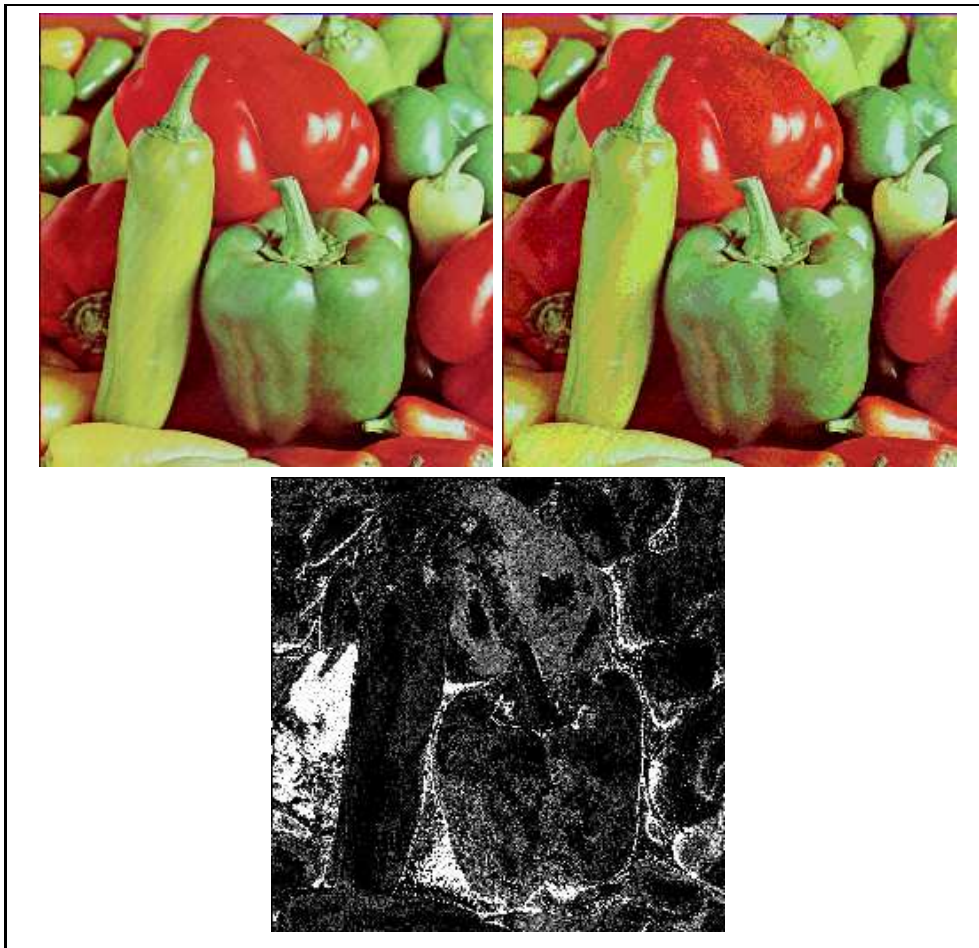


FIG. 5.7: Calcul de distance : quantification couleur en 8 niveaux

Chapitre 6

Perspectives

6.1 Un modèle de vision

Nous avons vu, lors des tests, que la prise en compte de la perception visuelle humaine était indispensable pour le calcul d'une distance entre images correcte.

Le but de cette section n'est pas de proposer un n -ième modèle, mais plutôt de garder les avantages de chacune des méthodes citées au chapitre 3. Nous devrions ainsi obtenir une solution à la fois cohérente physiologiquement et peu coûteuse.

6.1.1 Système optique

La quasi-totalité des méthodes vues modélisent l'effet de la cornée et du cristallin par un filtre passe-bas. La fonction de Weistheimer [35] utilisée dans le modèle de Sarnoff (cf. 3.1) semble être une bonne solution, à la fois rapide et précise.

$$Q(\rho) = 0,952e^{-2,59|\rho|^{1,36}} + 0,048e^{-2,43|\rho|^{1,74}} \quad (6.1)$$

6.1.2 Fréquences et orientations

Le principal phénomène que notre modèle doit prendre en compte est la sensibilité en fonction des fréquences et des orientations spatiales de l'image.

Nous reprenons l'idée donnée par Purgathofer dans [23]. Elle consiste à poser des rectangles aléatoirement dans l'image. La distance obtenue est la moyenne de toutes les moyennes des rectangles. De plus, leurs tailles ne sont pas choisies au hasard. Elles dépendent de la fonction de sensibilité au contraste (*CSF*). On utilise la technique de Monte-Carlo selon un échantillonnage d'importance sur la *CSF*. Ainsi, il est possible de pondérer la moyenne obtenue suivant les zones aux fréquences les plus détectables.

Purgathofer utilise la fonction de Mannos et Sakkrison. Par contre, le rapport entre la hauteur et la largeur est choisi aléatoirement. Là, nous pensons que la solution proposée par Daly [7], c'est à dire prendre une *CSF* en deux dimensions, est meilleure. Une fois choisie la taille du rectangle, son orientation permet de connaître la fréquence spatiale dans cette direction (cf. figure 6.1). Sachant que nous sommes plus sensibles aux objets horizontaux

et verticaux, il nous suffit de pondérer le choix de l'orientation des rectangles suivant cette sensibilité.

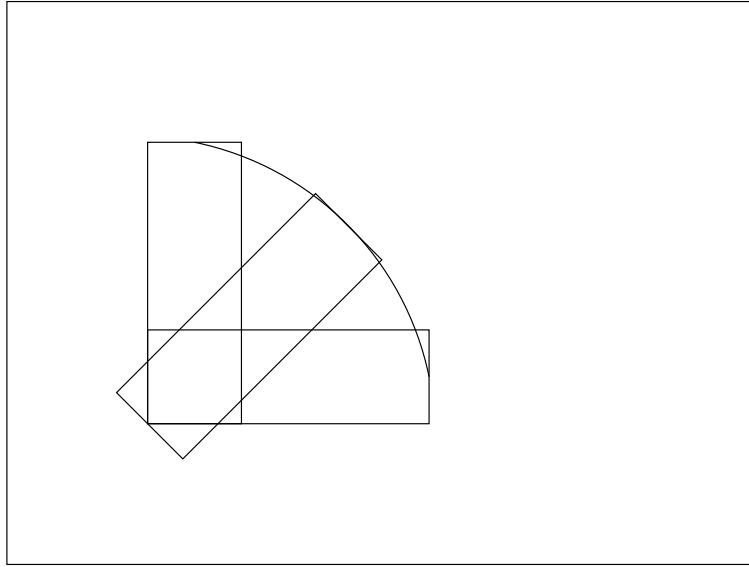


FIG. 6.1: Choix des rectangles

6.1.3 Fonction psychométrique

Le caractère aléatoire de la réponse peut être facilement reproduit par la fonction psychométrique. Celle donnée par Graham dans [12] semble correcte.

$$P(R) = 1 - 2^{-(1+kR)^3} \quad (6.2)$$

6.2 La dispersion des erreurs

Nous avons vu au chapitre 5 que l'image de la pièce meublée où seul changeait le tableau sur le mur était considérée comme très différente, au contraire d'autres où des erreurs minimales affectaient toute l'image. Quoiqu'ayant une valeur de distance proches, ces scènes ne sont pas vues de la même façon. Dans l'exemple 6.2, les deux images de distances ont la même valeur moyenne. Cependant, celle de gauche devra avoir un poids plus important que l'autre.

Notre idée est de nous inspirer de l'algorithme de quantification couleur de Bouman et Orchard [25] et de son optimisation proposée par Balasubramanian et al. [3].

La méthode repose sur la partition binaire de l'espace couleur en trois dimensions. A chaque étape, le but est de partitionner une zone selon la perpendiculaire à l'axe de plus grande variance. En réitérant, nous obtenons un certain nombre de zones où l'activité spatiale est pratiquement constante.

En remplaçant l'espace couleur par notre carte de distances, nous pensons obtenir ainsi une segmentation de notre image.

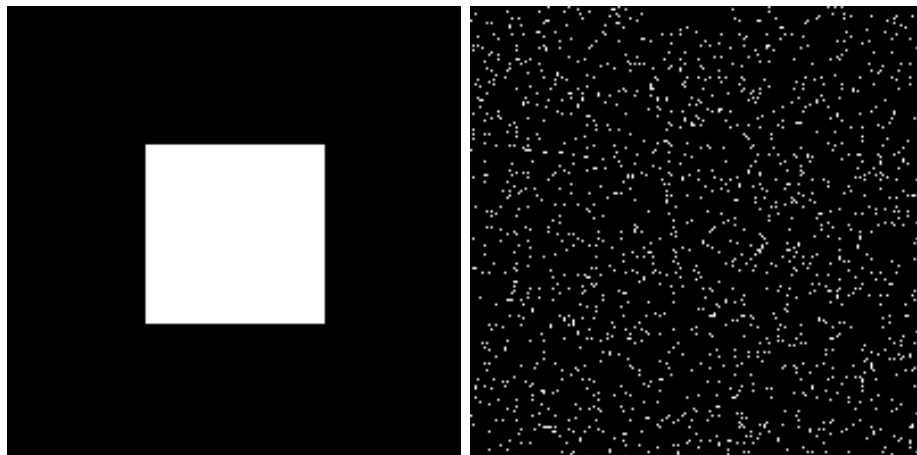


FIG. 6.2: Différences suivant la répartition des erreurs

Il convient maintenant d'analyser chaque zone. L'influence qu'une surface aura dans la distance résultante dépend de :

- sa valeur moyenne,
- son aire,
- son élongation (les zones circulaires sont plus prépondérantes que les zones allongées),
- peut-être sa position — nous avons vu en 5.3.3 (figure 5.7) que le placement des objets pouvait influencer —,
- peut-être aussi l'orientation.

Nous ne savons pas, à l'heure actuelle, comment ces facteurs doivent être pondérés entre eux, mais à chaque zone devrait être affectée un coefficient. La réunion de ceux-ci pourrait nous permettre de calculer une valeur globale (cf. figure 6.3).

L'optimisation de Balasubramanian était de travailler, non pas sur l'espace couleur, mais sur l'histogramme de chaque composante. Nous avons des valeurs de distances réelles, mais il serait intéressant de faire une première discrétisation en ne gardant que deux ou trois chiffres significatifs, pour obtenir un histogramme des distances et y appliquer la méthode.

6.3 Un tracé de rayons incrémental

L'objectif final du calcul d'une distance entre images est de l'intégrer dans un algorithme de tracé de rayons incrémental. Ainsi, la prise en compte de la vision humaine en synthèse d'images nous permettra de ne calculer que ce qui est nécessaire et, notamment dans les zones à fortes fréquences spatiales.

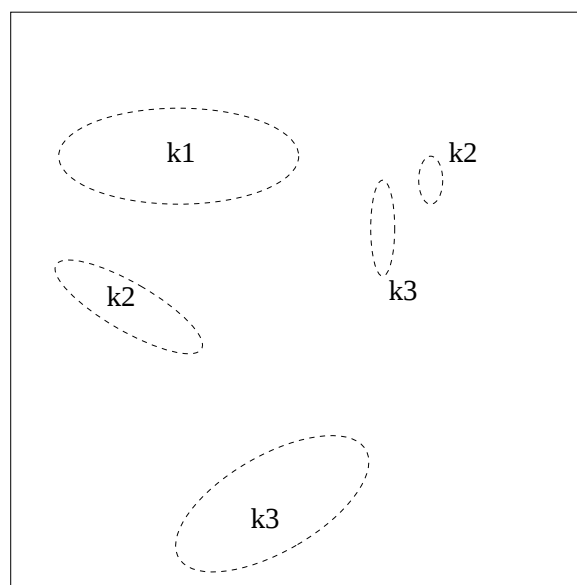


FIG. 6.3: Segmentation de l'image de distances

Conclusion

La distance définie pendant ce stage permet de connaître avec une bonne précision l'écart séparant deux images. Lors des tests effectués, nous avons vu que nos résultats étaient, en général, proches des diagnostics de nos observateurs. En outre, d'un calcul qui prenait plus de trente minutes auparavant, nous sommes passés à moins de dix secondes pour un résultat équivalent.

Nous avons dû abandonner, faute de temps, l'élaboration d'un modèle de vision. Cependant, il semble clair que cela constitue la prochaine étape pour une distance encore plus précise. Bien sûr, des tests complémentaires de plus grande ampleur devront aussi être mis en place.

Ce n'est qu'alors que le vaste domaine de la qualité des images pourra être traité.

Il faut préciser que ce travail ajouté à la partie distance de la thèse de Gilles Rougeron a fait l'objet d'un article en collaboration avec l'Institut d'Ingénierie de la Vision de Saint-Etienne. Ce papier sera présenté au mois d'octobre 1998 lors de l'*Inter-Society Color Council* (ISCC) à Baltimore, États-Unis.

Sur un plan personnel, ce stage m'a permis de découvrir le monde de la recherche scientifique auquel j'espère appartenir dans les années à venir.

Bibliographie

- [1] M.A. Ali and M.A. Klyne. *La vision chez les vertébrés*. Décarie Masson, Québec, 2^e édition, 1986.
- [2] C.R. Bader, P.R. MacLeish, and E.A. Schwartz. A voltage-clamp study of the light response in solidary rods of the tiger salamander. *J. Physiology*, 296(1), 1979.
- [3] R. Balasubramanian, J. Allebach, and C. Bouman. Color-image quantization with use of a fast binary splitting technique. *J. Opt. Soc. Am. A*, 11(11) :2777–2786, November 1994.
- [4] P.G. Barten. Evaluation of the effect of noise on subjective image quality. In *Human Vision, Visual Processing and Digital Display*, volume 1453, pages 2–15. SPIE, 1991.
- [5] P.G. Barten. Physical model for the contrast sensitivity of the human eye. In *Human Vision, Visual Processing and Digital Display*, volume 1666, pages 57–72. SPIE, 1992.
- [6] F.W. Campbell and J.G. Robson. Application of fourier analysis to the visibility of gratings. *J. Physiology*, 197 :551–566, 1968.
- [7] S. Daly. The visual difference predictor : an algorithm for the assessment of image fidelity. In A. B. Watson, editor, *Digital Images and Human Vision*, pages 179–206, Cambridge, 1993. MIT Press.
- [8] M. Fairchild. Refinement of the rlab color space. *Color Research and Applications*, 21 :338–346, 1996.
- [9] D. Gabor. Theory of communication. *Journal of IEEE*, 93 :429–457, 1946.
- [10] A. Gaddipatti, R. Machiraju, and R. Yagel. Steering image generation with wavelet based perceptual metric. *Eurographics '97*, 16(3) :241–251, 1997.
- [11] A. Glassner. *Principles of digital image synthesis*. Morgan Kaufmann publishers inc., San Francisco, California, 1995.
- [12] N. Graham. *Visual pattern Analyzers*. Oxford University Press, New York, 1989.
- [13] R. Hunt. Evaluation of a model of colour vision by magnitude scalings : discussion of collected results. *Color Research and Applications*, 19 :27–33, 1994.
- [14] R. Hunt. An improved predictor of colourfulness in a model of colour vision. *Color Research and Applications*, 19 :23–26, 1994.
- [15] Lubin J. A visual discrimination model for imaging system design and evaluation. *Vision Models for Target Detection and Recognition*, pages 245–283, 1995.

- [16] R. Luo. The llab model for colour appearance and colour difference evaluation. volume 2658, pages 261–269. SPIE, 1996.
- [17] R. Luo and R. Hunt. The structure of the cie 1997 colour appearance model (ciecam97s). *Color Research and Applications*, 23(3) :138–146, June 1998.
- [18] J-L. Mailliot. *Pseudo-réalisme et progressivité pour le tracé de rayons*. PhD thesis, Ecole Nationale Supérieure des Mines de Saint-Etienne, 1996.
- [19] J. Mannos and D. Sakrison. The effects of a visual fidelity criterion on the encoding of images. In *IEEE Transactions on Information Theory*, volume IT-20, pages 525–536, 1974.
- [20] G. Meyer and M. Bolin. A frequency based ray tracer. In *Computer Graphics, Annual Conference Series*, pages 409–418. ACM Siggraph, 1995.
- [21] G. Meyer and M. Bolin. A perceptually based adaptative sampling algorithm. In *Computer Graphics, Annual Conference Series*. ACM Siggraph, 1998. à paraître.
- [22] K. Mullen. The contrast sensitivity of human colour vision to red-green and blue-yellow chromatic gratings. *J. Physiology*, 359 :381–400, 1985.
- [23] L. Neumann, K. Matkovic, and W. Purgathofer. Perception based color image difference. Technical report, Insitut fur Computergraphik, Vienna, Austria, December 1997.
- [24] K. Ngan, K. Rao, and H. Singh. Cosine transform coding incorporating human visual system model. presented at SPIE fiber’ 86, 1986.
- [25] M. Orchard and C. Bouman. Color quantization of images. *Transactions on Signal Processing*, 39(12) :2677–2690, December 1991.
- [26] Padgham and Saunders. 1975.
- [27] A. Poirson and B. Wandell. A pattern-color separable pathways predict sensivity to simple colored patterns. *Vision Research*, 36(4) :515–526, 1996.
- [28] G. Rougeron. *Problèmes liés à la couleur en synthèse d’images*. PhD thesis, Ecole Nationale Supérieure des Mines de Saint-Etienne, 1998.
- [29] H. Saraux, B. Biais, and C. Rossazza. *Ophthalmologie*. Masson, Paris, 1988.
- [30] P. Teo and D. Heeger. Perceptual image distorsion. In *Human Vision, Visual Processing and Digital Display*, volume 2179, pages 127–139. SPIE, 1994.
- [31] P. Teo and D. Heeger. Perceptual image distorsion. In *First International Conference on Image Processing*, volume 2, pages 982–986, Austin, November 1994.
- [32] A. Watanabe, T. Mori, S. Nagata, and K. Hiwatashi. Spatial sine-wave responses of the human visual system. *Vision Research*, 8(9) :1245–1263, Sept. 1968.
- [33] A. Watson. The cortex transform : Rapid computation of simulated neural images. *Computer Vision, Graphics, and Image Processing*, 39 :311–327, 1987.
- [34] A. Watson and J. Solomon. Model of visual contrast gain control and pattern masking. *J. Opt. Soc. Am. A*, 14(9) :2379–2391, September 1997.
- [35] G. Weistheimer. The eye as an optical instrument. In K. Boff, L. Kauffman, and J. Thomas, editors, *Handbook of Perception and Human Performance*, NY, 1986. J. Wiley and sons.

- [36] J. Zaninetti and B. Péroche. A vector model for global illumination in ray tracing. *Proceedings of WSCG'98*, 3 :448–455, february 1998.

Manuel d'utilisation

Notre programme doit connaître certains paramètres. Il fonctionne de la sorte :

```
distis -1 <image1> -2 <image2> [-o sortie] [-p] [-s] {-t | -v | -a}
```

<image1> et <image2> représentent les deux images à comparer. Celles-ci doivent être au format LUM si l'option **-s** est spécifiée, au format PPM sinon.

<sortie> représente le nom du fichier de distance auquel sera rajouté l'extension **.dist**.

-p permet d'afficher les résultats sur l'écran. En outre, le programme ne renvoie pas de code d'erreur mais la valeur de la distance. Il peut ainsi être inséré dans un autre.

La méthode de calcul doit être obligatoirement spécifiée par :

- **-t** : tous les pixels ;
- **-v** : une valeur unique ;
- **-a** : méthode adaptative.