



HAL
open science

Entre perception et images numériques : une perspective

Samuel Delepouille

► **To cite this version:**

Samuel Delepouille. Entre perception et images numériques : une perspective. Informatique [cs].
Université du Littoral - Côte d'Opale, 2022. tel-04210197

HAL Id: tel-04210197

<https://theses.hal.science/tel-04210197>

Submitted on 18 Sep 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

UNIVERSITÉ DU LITTORAL CÔTE D'OPALE

Mémoire

d'Habilitation à Diriger les Recherches

discipline : informatique

présentée et soutenue publiquement le 17 juin 2022

par

SAMUEL DELEPOULLE

Laboratoire LISIC

Entre perception et images numériques : une perspective

Rapporteurs : Pr Céline Loscos, Université de Reims Champagne Ardenne
Pr Pascal Mamassian, École Normale Supérieure
Pr Daniel Meneveau, Université de Poitiers

Examineurs : Pr Rémi Cozot, Université du Littoral – Côte – d'Opale
Dr Solène Kalenine, Université de Lille
Pr Laurent Madelain, Université de Lille

Garant : Pr Christophe Renaud, Université du Littoral – Côte – d'Opale

« *Si j'ai vu plus loin, c'est en me
tenant sur les épaules de géants* »¹

Isaac Newton, *Lettre à Robert Hooke*,
5 février 1675

à André

1. « *if I have seen further, it is by standing on the shoulders of giants* »

Table des matières

Introduction	5
I Perception numérique	7
1 Sensibilité visuelle	9
1 Acuité	9
1.1 Définition de l'acuité	9
1.2 Mesure de l'acuité	10
1.3 Acuité face à une image numérique	10
1.4 Facteurs qui modulent l'acuité	11
1.4.1 Acuité et excentricité	11
1.4.2 Acuité et luminance	12
1.4.3 Hyperacuité	14
1.5 Application à l'imagerie numérique : le rendu fovéal	14
2 Contraste	16
2.1 Définitions	16
2.1.1 Le rapport de contraste	16
2.1.2 Le contraste local	16
2.1.3 Contraste de Michelson	17
2.2 Sensibilité au contraste	17
2.2.1 Sensibilité spatiale	17
2.2.2 Sensibilité temporelle	17
2.3 Méthodes de mesure de la CSF	18
2.4 Modèles de CSF	19
2.5 Rôle de l'apprentissage dans la CSF	19
2.6 Artistes	20
2 Perception de l'espace	23
1 Perspective linéaire	25
1.1 Description	25
1.1.1 Modèle de la caméra de type sténopé	25
1.1.2 Caméra virtuelle « plan devant »	26
1.1.3 Effet d'horizon	27
1.2 Perception	29
1.3 Artistes	30

	1.3.1	Dans la peinture	30
	1.3.2	Fausse perspectives	31
2		Gradient de texture	32
	2.1	Description	32
	2.2	Perception	32
	2.3	Artistes	32
	2.4	Informatique graphique	32
3		L'occultation ou interposition	33
	3.1	Définition	33
	3.2	Perception	34
	3.3	Utilisation par les artistes	34
	3.4	Image numérique	35
	3.4.1	Élimination des faces arrières ou <i>Backface culling</i>	35
	3.4.2	Algorithme du peintre	36
	3.4.3	Z-buffer	36
	3.4.4	Lancé de rayons	37
4		Le flou de profondeur	38
	4.1	Description	38
	4.1.1	Image focalisée	39
	4.1.2	Image défocalisée	39
	4.2	Perception	41
	4.3	Artistes	41
	4.4	Informatique graphique	41
5		Distance à l'horizon	43
	5.1	Description	43
	5.2	Perception	43
	5.3	Artistes	44
6		Ombres et ombrage	44
	6.1	Description	44
	6.2	Perception	45
	6.2.1	Informations sur la géométrie de l'objet	45
	6.2.2	Informations sur la position des objets	46
	6.3	Artistes	48
	6.4	Informatique graphique	49
	6.4.1	Les volumes d'ombres (<i>shadow volumes</i>)	49
	6.4.2	Shadow map	49
	6.4.3	Méthodes perceptives	49
7		La perspective aérienne	50
	7.1	Description	50
	7.2	Perception	50
	7.3	Artistes	50
	7.4	Informatique graphique	50
8		Synthèse et discussion	51

3	Stéréo-vision et perception de la 3D	55
1	Convergence oculaire	55
1.1	Définition	55
1.2	Rôle dans la perception de la distance	55
2	Disparité rétinienne	56
2.1	Zone de stéréovision	56
2.1.1	Mesure de la stéréoacuité	56
2.2	L'horoptère théorique	57
2.3	Horoptère empirique	57
2.4	La disparité	58
2.4.1	Seuil de confort	59
2.5	La vision stéréoscopique chez les animaux	61
3	La stéréoscopie	61
3.1	Les débuts	61
3.2	Les lunettes stéréoscopiques	63
3.2.1	Lunettes anaglyphes	63
3.2.2	Lunettes polarisantes	63
3.2.3	Lunettes à occultations alternées	63
3.3	Les autres périphériques	63
3.3.1	Écrans auto-stéréoscopiques	63
3.3.2	Casques de réalité virtuelle et augmentée	64
3.4	Caméras stéréoscopiques en images de synthèse	65
4	Sur l'effet de la 3D	66
II	Image de synthèse photoréaliste et perception du bruit	69
	Introduction	71
4	Synthèse d'images photoréalistes	73
1	Introduction	73
2	Modèles des sources de lumière	73
2.1	Notion de luminance	73
2.2	Représentation des sources de lumière	74
2.2.1	Géométrie d'une source lumineuse	74
2.2.2	Distribution d'intensité	74
2.2.3	Émission spectrale	74
2.2.4	Notion de solide photométrique	75
3	Modèles des matériaux	75
3.1	Modèles de réflexion et de réfraction	75
3.1.1	Le modèle de Lambert	77
3.1.2	Le modèle spéculaire pur (ou du miroir)	77
3.1.3	Le modèle de Phong (1975)	78
3.1.4	Le modèle de Cook-Torrance	79
4	Simulation de la propagation de la lumière	80
4.1	Modèle de propagation de la lumière : l'équation du rendu	80

4.2	Méthodes de résolution de l'équation du rendu	81
4.2.1	Les méthodes stochastiques	81
5	Apports réciproques de la psychologie de la perception et de la synthèse d'image photo-réaliste	85
1	Esquisse de laboratoires virtuels	85
1.1	Motivations	85
1.2	Raisons de l'utilisation de la synthèse d'images en sciences du comportement	86
1.2.1	Méthode	86
1.2.2	Application à l'image	87
1.3	Un exemple d'utilisation de l'image 3D	88
1.3.1	Exploration du couplage perception-action	88
1.3.2	Stimulus artificiels	88
2	Étude des artefacts visuels	90
2.1	Définition de l'artefact visuel	90
2.2	Artefacts en synthèse d'images	90
3	<i>Fireflies</i>	92
3.1	Problème	92
3.2	Médiane des moyennes	92
3.3	MoN adaptif avec le coefficient de Gini	92
3.3.1	Coefficient de Gini	93
3.3.2	Choix binaire de l'estimateur	94
3.3.3	MoN adaptatif	95
3.4	Comparaison des estimateurs	95
3.4.1	Scènes utilisées	95
3.4.2	Étude de la convergence	95
3.5	Impact visuel	96
4	Perception du bruit de Monte-Carlo en synthèse d'images photoréalistes . .	96
4.1	Positionnement du problème	96
4.2	Critère d'arrêt des algorithmes stochastiques	98
4.3	Critère d'arrêt basé sur un apprentissage supervisé	99
4.3.1	Constitution d'une base d'images	99
4.3.2	Procédure de recueil des seuils perceptifs	100
4.3.3	Apprentissage du bruit par méthode supervisée	101
4.3.4	Évaluation de l'apprentissage	102
4.4	Constitution d'une nouvelle base d'images	102
4.5	Modèle de perception du bruit	103
4.5.1	Paramètres testés et résultats	105
4.6	Génération automatique de caractéristiques de bruit	105
4.7	Comparaison de modèles	107
	Projets de recherche	111
1	De l'impossible conclusion	111
2	Précisions sur la perception du bruit	112
2.1	Algorithmique des artefacts visuels	112

	2.1.1	Une meilleure prise en compte du bruit	112
	2.1.2	Le problème des séquences d'images	113
	2.2	Explorer les périphériques	114
	2.3	L'impact de la plage de luminance	114
	2.3.1	Autres périphériques de rendu	114
3		L'extension du domaine de la perception	115
	3.1	Perception du réalisme	115
	3.2	Généralisation de la méthode : une voie transdisciplinaire	115

III Annexes 129

Annexes 131

1		DEA et thèse de Fabien Montagne	131
	1.1	Contexte	131
	1.2	Résumé	131
	1.3	Soutenance	132
	1.4	Publications associées	132
	1.4.1	Montagne et al. (2003)	132
	1.4.2	Montagne et al. (2006)	133
2		DEA et Thèse de Nawel Takouachet	133
	2.1	Contexte	133
	2.2	Résumé	133
	2.3	Soutenance	133
	2.4	Publications associées	134
	2.4.1	Takouachet et al. (2007)	134
3		Mémoire bibliographique de DEA de Pierre Marie Oumoumsack	134
4		Thèse de Thuy Tram Ngo Thi	134
5		Thèse de Jérôme Buisine	135
	5.1	Contexte	135
	5.2	Résumé	135
	5.3	Soutenance	135
	5.4	Publications associées	136
	5.4.1	Buisine, Teytaud et al. (2021)	136
	5.4.2	Buisine, Delepouille et Renaud (2021b)	136
	5.4.3	Buisine, Teytaud et al. (2021)	137
	5.4.4	Buisine, Delepouille et Renaud (2021a)	137
6		ANR PrISE-3D	138
	6.1	Contexte et positionnement	138
	6.2	Objectifs	138
	6.3	Organisation et moyens	139
	6.3.1	Organisation scientifique	139
	6.3.2	Moyens	140
	6.4	Bilan et perspectives	140
	6.4.1	Bilan scientifique	140
	6.4.2	Bilan quantitatif	141

6.4.3	Publications	141
6.4.4	Thèses	141
6.5	Perspectives	142
7	Publications	142

Table des figures

1.1	Acuité normale	11
1.2	Densité des photorécepteurs	13
1.3	Répartition des photorécepteurs	13
1.4	Hyperactuité	15
1.5	CSF publiée par Campbell et Robson (1968)	18
1.6	Comparaison des différentes fonctions d'ajustement à la CSF. Figure par J. Buisine.	19
1.7	CSF pour créer des images hybrides	21
2.1	Représentation schématique des principaux indices picturaux	24
2.2	Modèle de perspective « plan derrière ».	26
2.3	Modèle de perspective « plan devant ».	27
2.4	Élévation	28
2.5	Angle d'élévation	28
2.6	Points de fuite	29
2.7	Dispositif d'acquisition de la perspective	30
2.8	Perspective accéléré du palais Spada à Rome	31
2.9	Rue de Paris.	33
2.10	Exemples de complétion visuelle.	34
2.11	Le motif de Kanizsa	35
2.12	« <i>La dernière cène</i> » de Duccio di Buoninsegna. Domaine public.	36
2.13	Hierarchie de volumes englobants	38
2.14	Principe du BSP-tree	38
2.15	Système optique focalisé.	40
2.16	Système optique défocalisé.	40
2.17	Emerald. De Philippe Barlow	42
2.18	illusion de Ponzo	44
2.19	Utilisation artistique de l'illusion de Ponzo	45
2.20	Comparaison entre ombre dure et ombre douce	46
2.21	Informations de l'ombre projetée	47
2.22	Ambiguïté de l'image sans ombre	47
2.23	L'invention de la peinture	48
2.24	Perspective aérienne en peinture	51
3.1	Champ visuel horizontalement et verticalement.	57
3.2	Horoptère théorique dans le plan horizontal et en 3D.	58

3.3	Disparité horizontale	59
3.4	Disparité croisée et homonyme	60
3.5	Le stéréoscope de Wheatstone	62
3.6	Portrait stéréoscopique de Abraham Lincoln	62
3.7	Comparaison de deux techniques auto-stéréoscopiques.	64
3.8	Trois sortes de caméras stéréoscopiques	66
4.1	Spectres d'émission	75
4.2	Solides photométriques	76
4.3	Rayon incident, rayon réfléchi	76
4.4	Matériau diffus : modèle de Lambert	77
4.5	Matériau spéculaire pur	78
4.6	Le modèle de Phong	79
4.7	Le modèle de Cook-Torrance	80
4.8	Principe du lancer de rayons.	82
4.9	Bruit de Monte-Carlo	83
5.1	Tendances sur la perception et réalité virtuelle	87
5.2	Un dispositif expérimental dans l'environnement 3D.	89
5.3	Exemples d'artefacts optiques.	90
5.4	Problème d' <i>aliasing</i> des ombres (source NVIDIA).	91
5.5	Problème des <i>Firefly</i>	93
5.6	Principe de l'estimateur MON	94
5.7	Illustration du nombre de moyennes retenues pour le calcul.	95
5.8	Images de référence	96
5.9	Convergence des estimateurs	97
5.10	Comparaison des différents estimateurs	98
5.11	Exemple d'interface de recueil des données.	100
5.12	Méthode pour l'apprentissage des seuils	101
5.13	Exemple d'image	103
5.14	Exemples d'images de la base de données.	103
5.15	L'architecture du réseau RNN	104
5.16	Résultats de la sortie du RNN	106
5.17	Architecture de modèle GGN	107
5.18	Exemple d'utilisation des seuils recueillis	108
5.19	Écart aux seuils perceptifs enregistrés sur les 16 blocs des 40 images	108
5.20	Seuils prédits par le modèle comparés au seuils observés.	109
5.21	Organisation des différentes tâches du projet.	140

Liste des tableaux

1.1	Densité optique des périphériques courants	12
1.2	Acuité par zone	14
1.3	Fonctions d'ajustement à la CSF	19
5.1	Résultats après apprentissage.	102

Remerciements

La science est un peu comme un sport qui se pratique en équipe. Si les remerciements occupent la première place de nos mémoires, c'est pour rappeler que nous sommes *des nains sur des épaules de géants*. Ceci est d'autant plus vrai lorsqu'on prétend à l'habilitation à diriger les recherches. Aussi je souhaite remercier tous ceux qui d'une façon ou d'une autre m'ont aidé dans ce long parcours, même s'ils ne sont pas cités ici.

Je veux exprimer ma gratitude à Céline Loscos, Pascal Mamassian et Daniel Menveaux qui ont accepté de rapporter ce document ainsi qu'à Rémi Cozot, Solène Kalénine et Laurent Madelain examinateurs.

Je souhaite remercier très chaleureusement Christophe Renaud qui a cru dès le départ que la perception visuelle fait partie intégrante du processus de l'image en informatique. Je le remercie de sa confiance et de son exigence scientifique. Je voudrais associer à ces remerciements les collègues chercheurs avec qui j'ai travaillé dans les équipes OASIS puis IMAP : la recherche n'empêche pas de prendre les choses en riant. Merci donc à François Rousselle, Julien Dehos, Rémi Synave, André Bigand sans oublier Michel Leblond. J'y adjoint Gaele Compiègne et Isabelle Buchard qui assurent le secrétariat du laboratoire.

Le travail des doctorants est également précieux, je sais ce que ces travaux doivent à Nawel Takouachet, à Valerie Tram et à Jérôme Buisine. Merci aussi à Thomas Boute, Richard Drouard, Victor Drouard, Quentin Huan, Antoine Sauvage et Florence Constans.

Le partage de la connaissance est indissociable du projet scientifique, c'est pourquoi je souhaite remercier mes collègues, enseignants, intervenants et personnels de l'IUT, en particulier ceux des départements informatique et GEII.

Merci à Franck Vandewièle¹ pour ses relectures attentives du texte et des formules, à mes amis, à ma famille, tout particulièrement à Nathalie, Romane, Clément et Jeanne.

1. armé de son *Lexique des règles typographiques en usage à l'Imprimerie nationale*

Introduction

Le thème de ce mémoire d'habilitation à diriger les recherches porte sur les liens très étroits qu'entretiennent perception visuelle et images numériques. Il est composé de deux grandes parties.

La première constitue un tour d'horizon des relations entre perception visuelle et informatique graphique. Ce sujet fait intervenir des méthodes, des connaissances et des techniques de champs disciplinaires qui ont peu d'intersection en apparence. Des notions seront donc empruntées à l'informatique, à la biologie, à l'optique, à la psycho-physique, à la psychologie, à la neuro-biologie et à la médecine. Cependant le choix est ici de ne pas séparer ces différents aspects dans des chapitres isolés mais bien de traiter la perception de l'image numérique dans son ensemble comme un objet d'étude à part entière. Le choix scientifique est bien de considérer l'observateur, le périphérique d'affichage et l'ensemble des algorithmes qui pilotent ces dispositifs comme un et un seul système — particulièrement complexe. Au cours du développement des connaissances scientifiques et de l'histoire des sciences les études spécialisées dans les disciplines citées précédemment ont souvent considéré des parties de ce système. Il nous paraît actuellement nécessaire de proposer une étude synthétique donc nécessairement transdisciplinaire. Car depuis que les techniques ont permis de produire des écrans graphiques et donc d'afficher des images, le même problème se pose : l'image est destinée à être perçue par un être humain — un organisme qui possède un système visuel. C'est-à-dire que l'image numérique est en réalité toujours inclus dans une interaction qui fait intervenir des éléments technologiques et un organisme biologique. L'étude de l'un décorrélée de l'autre est en partie nécessaire mais non suffisante.

Sans prétendre à l'exhaustivité, trois chapitres s'intéressent à des thèmes plus précis. Le premier envisage les relations entre la sensibilité du système visuel face aux périphériques d'affichage. Un second chapitre présente les indices monoculaires qui entrent en jeu dans la perception de la distance face à une image — numérique ou non — et un troisième chapitre explore la question de la vision binoculaire, en particulier face aux périphériques de stéréovision. Dans cette première partie, l'accent est mis sur les connaissances scientifiques solides, établies et qui font l'objet d'un consensus. Ces chapitres laissent volontairement de côté certains développements formels, mathématiques ou algorithmiques avancés. Le choix est de situer les problèmes dans une perspective d'histoire de la connaissance afin de permettre au lecteur d'envisager les grands problèmes et les réponses qui ont été adoptées, que ce soit au cours de l'évolution des connaissances ou dans les différentes spécialités. Une référence constante est donnée à l'histoire des arts visuels dans la mesure où les artistes ont souvent été les premiers à expérimenter l'interaction vision-image. Il est intéressant de souligner que la construction de l'image numérique dans la continuité de ce mouvement

a souvent utilisé, de façon formalisée, les techniques picturales.

La seconde partie de ce mémoire est davantage destinée à présenter les problèmes de recherche auxquels j'ai contribué. Après un chapitre pour présenter les méthodes informatiques de construction des images photo-réalistes, un chapitre envisage comment l'informatique graphique et la psychologie expérimentale peuvent se compléter. L'un des problèmes envisagé est celui de la perception du bruit dans des images de synthèse photo-réalistes. C'est le sujet sur lequel j'ai été le plus actif et pour lequel j'ai participé à l'encadrement de trois thèses avec Christophe Renaud.

Ce document ne reprend pas tous les sujets auxquels j'ai contribué. En particulier, j'ai choisi de ne pas exposer ce qui concerne l'apprentissage par renforcement pour lequel je renvoie à ma thèse de doctorat et à celle de Fabien Montagne co-encadrée avec Philippe Preux.

Première partie
Perception numérique

Chapitre 1

Sensibilité visuelle

« La vision est l'art de voir les choses invisibles. »

Jonathan Swift, Pensées sur divers sujets moraux et divertissants (1745)

Ce chapitre s'intéresse à deux mesures de la sensibilité visuelle : celle des détails qui correspond à l'acuité visuelle et celle des contrastes qui correspond à la CSF (*Contrast Sensitivity Function*). Dans les deux cas, leur étude présente un intérêt dans le cadre de l'imagerie numérique.

1 Acuité

1.1 Définition de l'acuité

La mesure la plus courante de l'acuité visuelle correspond au minimum séparable. C'est-à-dire qu'on présente deux stimulus nettement visibles et la distance minimale pour laquelle les points sont perçus comme doubles est enregistrée. En dessous de cette distance ils sont perçus comme un seul point. Une autre méthode pour mesurer cette acuité est de présenter des stimulus comportant des cycles de lignes qui alternent blanc et noir. A partir d'une certaine distance, ces stimulus ne peuvent plus être différenciés d'une zone grise. Cette mesure est ramenée à une valeur angulaire qui représente le pouvoir de résolution de l'œil. L'acuité est alors définie comme l'inverse du pouvoir de résolution.

$$AV = \frac{1}{\alpha} \tag{1.1}$$

Avec α la valeur angulaire exprimée en minute d'arc.

Pour un observateur, une fois corrigés les éventuels problèmes liés à la convergence oculaire, il est considéré que la normalité correspond à un pouvoir de résolution proche d'une minute d'angle (Velasco e Cruz, 1990) et l'acuité correspondante est usuellement exprimée dans le système français en dixièmes (Colenbrander, 2003) alors qu'une mesure en vingtièmes est plus utilisée aux États-Unis. Dans la pratique, des acuités supérieures à

un (10/10) ne sont pas rares, en particulier chez les sujets jeunes pour lesquels des valeurs allant jusqu'à 20/10 peuvent parfois être observées.

1.2 Mesure de l'acuité

L'acuité est en pratique mesurée à l'aide d'échelles standardisées : anneaux de Landolt, échelles Monoyer et Snellen pour la vision de loin ou Parinaud pour la vision de près. L'acuité est définie comme une fonction inverse du minimum qui peut être résolu, ce qui aboutit à une échelle non linéaire. Cela est un problème pour établir des statistiques telles que la moyenne ou la variance dans les mesures d'acuité. Pour ces raisons, il est préférable de substituer la mesure de l'acuité visuelle (AV) au $\log MAR$ (*Logarithm of the Minimum Angle of Resolution*) qui est une transformation en logarithme en base 10 (Bailey & Lovie, 1976 ; Holladay, 1997 ; Kniestedt & Stamper, 2003).

$$\log MAR = -\log(AV) \tag{1.2}$$

Et inversement, l'acuité visuelle peut alors être exprimée en décimales :

$$AV = 10^{-\log MAR} \tag{1.3}$$

Il existe également des planches qui utilisent une progression logarithmique comme l'échelle de Bailey-Lovie et l'ETDRS.

1.3 Acuité face à une image numérique

La valeur de 10/10 ou acuité de un est équivalente à une fréquence spatiale de 30 cycles par degrés (voir figure 1.1). La fréquence spatiale permet alors de définir le nombre de répétitions d'un motif visuel par unité angulaire. Le tableau 1.1 présente la fréquence spatiale associée à différents périphériques pour des distances proches de leur utilisation naturelle. Certains périphériques induisent des fréquences spatiales nettement inférieures ou supérieures à l'acuité normale.

Ce qui produit deux effets :

- lorsque la fréquence spatiale est fortement inférieure à l'acuité, les pixels peuvent-être perçus individuellement, ce qui induit une perception de crénelage ou de « pixelisation » ;
- à l'inverse lorsque la fréquence spatiale dépasse les limites de l'acuité, aucune amélioration de l'image n'est perceptible alors que le coût de traitement et de stockage des images est plus important.

En impression, la valeur de 300 DPI (*dots per inch*) correspond à environ 38 cycles par degrés pour une feuille vue à une distance 35 cm, ce qui est à nouveau de l'ordre du seuil de résolution des détails.

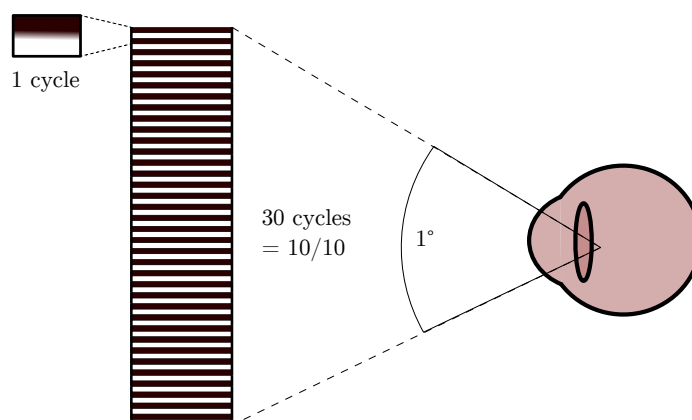


FIGURE 1.1 – Acuité normale

La vision dite normale correspond approximativement à la capacité à résoudre environ une minute d'angle soit un soixantième de degrés. Ceci correspond donc à 30 cycles par degré puisque un cycle nécessite au moins deux pixels.

1.4 Facteurs qui modulent l'acuité

1.4.1 Acuité et excentricité

L'acuité visuelle n'est réellement importante que dans une petite fraction du champ visuel. Contrairement à un capteur optique, le pouvoir de résolution de l'œil n'est pas constant puisque les récepteurs ne sont pas situés selon une grille régulière. Dans l'axe optique de l'œil, correspondant à l'axe optique, une zone d'un diamètre d'environ 0,35 mm, appelée fovéa, contient une très forte densité de cônes fins et allongés. Cet éloignement est appelé excentricité rétinienne.

Les cellules photo-sensibles ont une répartition inégale dans l'œil. Cette répartition est fonction de l'excentricité et du type de récepteur : les cônes, sensibles à la couleur et aux détails, sont responsables de l'acuité fine et sont très majoritairement regroupés au niveau de l'axe optique dans la zone de vision maximale, la fovéa. La répartition des bâtonnets décroît, elle, de façon moins drastique en fonction de l'excentricité, avec deux exceptions notable : la fovéa elle-même qui ne comporte pas de bâtonnets et la tache aveugle, zone d'insertion du nerf optique, qui est dépourvue de tout récepteur (figure 1.2). L'acuité visuelle est donc très variable suivant l'éloignement au centre de la fovéa.

Il est possible de relier l'acuité à la densité de récepteur. En effet, l'acuité étant définie comme l'inverse du plus petit angle qui peut être résolu ($AV = 1/\alpha$), l'angle α peut être retrouvé. Soient deux rayons lumineux issus de deux points A et B se projettent sur deux cellules réceptrices de la rétine, respectivement A' et B' (figure 1.3). Sous l'approximation des lentilles minces,

$$\alpha = \widehat{A'NB'} = \widehat{ANB} \quad (1.4)$$

Ecran	Résolution	diagonale	distance	cycles par degrés
Smartphone	2280 × 1080	146 mm (5,8')	20 cm	31
Montre connectée	454 × 454	35 mm (1,39')	20 cm	32
Console portable	1280 × 720	157 mm (6,2')	30 cm	24
Moniteur informatique Full-HD	1920 × 1080	686 mm (27')	50 cm	16
Moniteur informatique 4k	3840 × 2160	800 mm (31,5')	50 cm	27
TV 4k	3840 × 2160	1397 mm (55')	3 m	85
TV 8k	7680 × 4320	1626 mm (64')	3 m	144
Casque RV	Résolution	champs visuel		cycles par degrés
1 ^{er} génération	640 × 800	110°		3,5
Casque récent	1440 × 1600	110°		6,5

TABLE 1.1 – Densité optique des périphériques courants

Densité optique en cycles par degré de divers périphériques courants. Pour certains périphériques la densité est nettement inférieure à la sensibilité normale, c'est le cas des casques de réalité virtuelle, ce qui induit un effet de pixellisation perceptible. À l'inverse la résolution 8k pour un écran vu à 3m de distance produit une densité optique que la vision humaine ne peut distinguer.

Ce qui permet d'exprimer l'angle qui correspond au minimum résolu en fonction de ε l'excentricité, de ρ la densité des cônes et de NC la distance du point nodal au centre de l'œil. Le point C représente le centre de l'œil et N son point nodal¹.

$$\alpha = 2 \sin^{-1} \left[\sin \left(\frac{\sqrt{1/\rho}}{2\sqrt{(r \sin(\varepsilon))^2 + (r \cos(\varepsilon) + NC)^2}} \right) \right] \quad (1.5)$$

Il est alors possible de retrouver les données relatives à l'acuité visuelle en fonction de la zone de l'œil. Ces valeurs, consignées dans le tableau 1.2 rejoignent les mesures expérimentales de l'acuité (Wertheim, 1980).

1.4.2 Acuité et luminance

La meilleure acuité visuelle est observée pour des conditions d'illumination dites photopiques. Les récepteurs qui sont responsables de la meilleure acuité sont les cônes. Or

1. Dans la théorie, l'œil devrait plutôt être représenté par deux points nodaux. Dans la pratique, la distance très faible qui les sépare, de l'ordre de 0,3 mm, autorise à les confondre et à les assimiler au centre optique. Ce dernier se trouve proche de la partie postérieure du cristallin, à 17mm de la rétine.

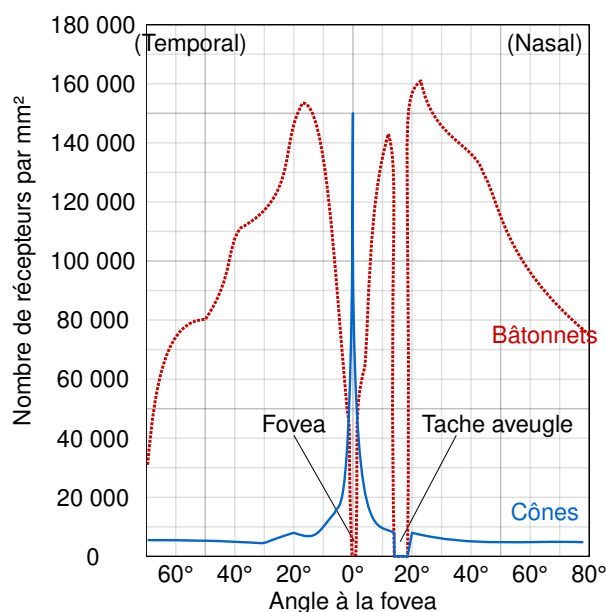


FIGURE 1.2 – Densité des photorécepteurs

La densité des photorécepteurs de type bâtonnet est figurée par la ligne pointillée rouge et celle de type cône par la ligne continue bleue. Cette densité est mesurée le long d'une ligne passant par la fovéa et la tache aveugle d'un œil humain en fonction de l'angle mesuré à partir de la fovéa. Wikimedia Cmglee CC BY-SA 3.0 d'après Wandell (1995)

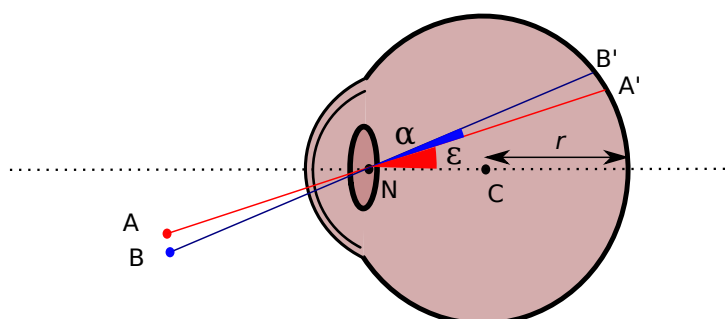


FIGURE 1.3 – Répartition des photorécepteurs

L'angle α de résolution minimale peut être assimilé à l'angle incident sous l'approximation de Gauss. L'œil est de centre C, de rayon r et son point nodal N se situe proche du cristallin. A' et B' sont les projetés de deux points A et B à la limite de la résolution et séparés d'un angle α . La valeur de α est liée à l'excentricité ϵ et à la densité des récepteurs ρ .

ceux-ci ont besoin d'une quantité importante de lumière pour répondre, ce qui implique une baisse de l'acuité visuelle lorsque la luminance décroît.

Sheedy et al. (1984) enregistrent les variations d'acuité pour une population de sujets utilisant des anneaux de Landolt et des lettres britanniques. La relation entre le loga-

Région	Excentricité ε (en °)	Distance depuis la fovéa (en mm)	Densité maximale des cônes (en unité par mm^2)	Résolution maximale (en minutes d'arc)	Meilleure acuité visuelle
fovéa	de 0 à 3,3	de 0 à 0,75	198 100	0,38	2,6
zone parafovéal	de 3,3 à 5,5	de 0,75 à 1,25	26 000	1,05	1,0
macula	de 5,5 à 12,1	de 1,25 à 2,75	19 275	1,22	0,8
periphérie	12,1 et au delà	2,75 et au delà	9 900	1,70	0,6

TABLE 1.2 – Acuité par zone

Meilleure acuité calculée en fonction de la zone de la rétine. Source : Chen et al. (2019)

arithme de l'angle minimal résolu (logMAR) et le logarithme de la luminance s'ajuste bien à une droite pour une plage de luminance photopique allant de 40 à 600 cd/m^2 . Ils notent qu'un doublement du niveau de luminance dans cette plage améliore la mesure de l'acuité d'environ une lettre sur une rangée de cinq lettres. Ce qui correspond à un facteur multiplicatif de 1,26 où 0,10 unité logarithmiques. Cependant, cet effet est moins sensible sur le test des anneaux de Landolt.

1.4.3 Hyperacuité

Dans certains cas, le système visuel dans son ensemble semble démontrer des capacités supérieures à la sensibilité de chacun des récepteurs. C'est le cas, par exemple, dans des tâches où l'on demande de juger si deux segments sont alignés (1), si trois points sont alignés (2) ou la direction d'un arc (3) comme présenté sur la figure 1.4. Dans ces tâches une sensibilité des réponses à des variations inférieures à 10 secondes d'arc peut être observée. Cette valeur est nettement supérieure aux mesures d'acuité dans des conditions standards et même à l'écart entre deux récepteurs : leur projection est de l'ordre de 30 secondes d'arc dans le meilleur des cas (McKee et al., 1990).

L'explication de ce phénomène est que de nombreux récepteurs participent à la détection de ces motifs et que les interactions de voisinage permettent d'accroître la résolution. L'hyper-acuité peut être mesurée à l'aide d'échelles dédiées.

1.5 Application à l'imagerie numérique : le rendu fovéal

Dans le cadre des périphériques de réalité virtuelle et de réalité augmentée, une évolution dans la résolution des périphériques est observée. Cependant la résolution utilisée par ces périphériques associée au fait qu'ils couvrent une grande partie du champ visuel aboutit une densité optique limitée. Dans les dispositifs actuels, la résolution angulaire dépasse rarement 10 cycles par degré. Ceci reste nettement inférieur à la capacité de discrimination de l'œil humain.

L'augmentation des résolutions pour atteindre la limite de perception humaine se heurte à un problème : le champs visuel moyen monoculaire est de l'ordre de 170° horizontalement (60° nasalement et presque 110° temporalement) et 150° verticalement (70°

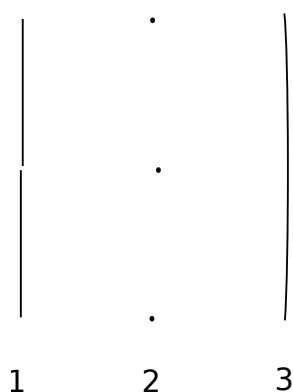


FIGURE 1.4 – Hyperacuité

Exemples de stimulus utilisés pour tester l’hyper-acuité. De très petits écarts peuvent être perçus. Pour le stimulus 1, les segments ne sont pas parfaitement alignés, de même que pour le stimulus 2, les points ne sont pas alignés et pour le stimulus 3, le segment présente une légère courbure. Dans les trois cas, des variations inférieures à l’écart théorique d’acuité peuvent être perçus.

vers le haut et 80° vers le bas). Atteindre 60 cycles par degré, nécessiterait donc une résolution de $10\,200 \times 9\,000$ pour chaque oeil — avec une forte exigence temporelle puisque l’interaction visuelle impose un taux de rafraîchissement d’au moins 60Hz. Cela reste encore nettement au dessus des capacités de traitement, même avec des matériels récents.

La solution du rendu fovéal repose sur l’utilisation de techniques d’oculométrie (*eye tracking*) afin de proposer une expérience visuelle optimale en vision centrale et une diminution importante du rendu des détails en dehors de cette zone (Mohanto et al., 2021). Pour les périphériques qui couvrent un grand champ de vision, la majorité des efforts de calcul concerne des zones dans lesquelles l’acuité est faible.

Ces techniques sont particulièrement intéressantes dans le cadre des casques de réalité virtuelle pour au moins deux raisons. D’une part le champ visuel à couvrir est important et, d’autre part, il est relativement facile d’y inclure un dispositif de capture de la position du regard. La part du champ visuel utilisée finement dans ce type de périphérique est de l’ordre de 4% (Patney et al., 2016).

L’utilisation de ces techniques avec des moniteurs « classiques » est plus problématique car l’enregistrement de la position du regard est plus difficile face à ce type de périphériques. De plus, le gain en performance semble un peu plus limité mais reste important : Wei et Sakamoto (2019) estiment que la région fovéale couvre environ 8% d’un moniteur de bureau utilisé en conditions standard.

2 Contraste

2.1 Définitions

Si l'acuité définit le seuil absolu de sensibilité aux détails du système visuel, un autre problème est d'enregistrer la sensibilité au contraste. Alors que les mesures d'acuité sont réalisées dans les conditions où le contraste est le meilleur, l'étude de la sensibilité au contraste consiste à déterminer les valeurs de contraste limite pour lesquelles un signal est perçu.

Bien qu'intuitivement simple, la notion de contraste dans des zones de gris (contraste mono-chromatique) peut se définir de plusieurs façons différentes. Trois mesures sont couramment utilisées : le rapport de contraste, le contraste local et le contraste de Michelson.

2.1.1 Le rapport de contraste

Le rapport de contraste que l'on appelle aussi contraste dynamique ou gamme dynamique mesure l'écart entre la plus grande valeur de luminance affichable rapportée à la plus petite.

$$C_s = \frac{L_{max}}{L_{min}}$$

avec L_{max} et L_{min} qui sont respectivement les valeurs de luminance maximum et minimum.

La valeur C_s peut varier entre 1 qui représente l'absence de contraste à des valeurs grandes, sans limite. Il est généralement exprimé sous la forme d'un ratio (X :1). Plus X est important plus le contraste est élevé. Cette mesure du contraste est généralement utilisée pour caractériser des périphériques d'affichage.

2.1.2 Le contraste local

La notion de contraste local a été introduite car elle permet de mieux rendre compte de la perception des détails, en particulier de la détection des contours. Le contraste local, ou contraste de Weber, est calculé comme :

$$C_w = \frac{L_{zone} - L_{fond}}{L_{fond}}$$

où L_{zone} et L_{fond} sont respectivement les valeurs de luminance d'une zone et celle du fond. Le contraste de Weber varie entre une valeur négative sans limite lorsque le fond est plus clair que la zone à une valeur positive arbitrairement grande lorsque la zone est plus claire que le fond. En cas d'absence de contraste, sa valeur est de 0.

Ernst Weber et Gustav Fechner (Ekman, 1959) ont étudié comment est perçu le contraste local. Ils aboutissent à la conclusion que la plus petite variation perceptible (ΔI) dans le contraste est liée linéairement à l'intensité (I). Cette observation ($\frac{\Delta I}{I} = k$) les amène à postuler une variation logarithmique de la sensation perçue :

$$I = k.log(S)$$

2.1.3 Contraste de Michelson

Le contraste de Michelson qui correspond au concept de modulation peut s'exprimer comme :

$$C_m = \frac{L_{max} - L_{min}}{L_{max} + L_{min}}$$

où L_{max} et L_{min} sont respectivement les valeurs de luminance minimale et maximale d'un signal. Un avantage du contraste de Michelson est que les valeurs sont dans l'intervalle $[0 : 1]$ avec la valeur 0 qui représente l'absence de contraste et 1 le contraste le plus élevé. C'est cette dernière définition qui est utilisée pour mesurer la sensibilité au contraste lorsqu'on utilise des signaux périodiques.

2.2 Sensibilité au contraste

2.2.1 Sensibilité spatiale

Le test de sensibilité au contraste s'intéresse à la mesure de la sensibilité visuelle avec des cibles plus grandes et des contrastes plus faibles. Des cibles qui comportent un signal sinusoïdal sont généralement utilisées. Le contraste de ces cibles varie afin de mesurer le plus faible contraste détectable pour une fréquence spatiale donnée. La sensibilité au contraste est l'inverse de ce contraste minimum. Dans la mesure où le signal présenté est périodique, c'est le contraste de Michelson qui est utilisé.

En faisant varier le contraste et la fréquence spatiale, Campbell et Robson (1968) mesurent la sensibilité au contraste pour une grille sinusoïdale. Ils obtiennent typiquement une courbe avec un optimum de perception du contraste autour de 8 cycles par degré. De part et d'autre de cet optimum, la sensibilité au contraste décroît. Les variations de contraste qui correspondent à la partie au dessus de la courbe de la CSF ne sont pas visibles².

On teste aussi la CSF par le biais de patches de Gabor. Le principe est de moduler le signal sinusoïdal avec une enveloppe gaussienne :

$$I(x, y) = \cos(ax + by) \exp\left(-\frac{x^2 + y^2}{2\sigma^2}\right) \quad (1.6)$$

2.2.2 Sensibilité temporelle

Il est également possible de mesurer une CSF temporelle (ou TSF pour *Temporal contrast Sensitivity Function*). Les mêmes protocoles sont utilisés mais la fréquence est alors dans le domaine temporel au lieu d'être dans le spatial.

Pour cela une cible qui clignote est présentée. L'œil semble être le plus sensible à une fréquence de 15 à 20 Hz pour des luminances élevées, ce qui correspond à la vision photopique. Une coupure de haute fréquence est enregistrée autour de 60Hz, le stimulus est alors perçu comme fixe. Pour des niveaux de lumière faibles, le maximum de sensibilité est enregistré pour des valeurs inférieures à 10Hz et le meilleur contraste est de 20%. La coupure de haute fréquence se situe alors autour de 15Hz (Kalloniatis & Luu, 2005).

2. « *it is in this region that ectoplasm, fairies and ghosts could exist in privacy.* » (Campbell, 1983)

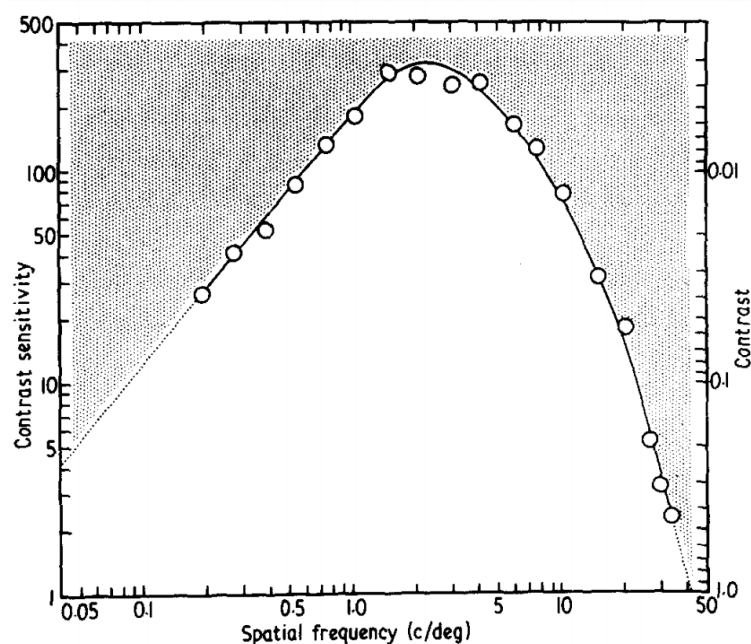


FIGURE 1.5 – CSF publiée par Campbell et Robson (1968)

2.3 Méthodes de mesure de la CSF

Au même titre que l'acuité visuelle, la CSF est intéressante à mesurer en ophtalmologie puisqu'elle rend compte des capacités du système visuel à discriminer des contrastes. La modification de la sensibilité au contraste peut être révélatrice de pathologies comme des glaucomes ou des névrites optiques. (Arden & Jacobson, 1978)

La sensibilité au contraste semble aussi très liée aux performances optiques de l'œil. Il est également possible que les modifications de la CSF chez des sujets normaux soient causées par des aberrations d'ordre supérieur comme des aberrations sphériques, la coma³ ou l'astigmatisme d'ordre supérieur (Oshika et al., 2006).

Le protocole à base de grille ou de patches de Gabor reste complexe à mettre en oeuvre dans le cadre d'une consultation ophtalmologique. Pour cette raison, les tests de Bailey-Lovie et de Pelli-Robson ont été proposés.

Le test de Pelli-Robson est un tableau qui comporte des triplets de lettres. Ces lettres deviennent progressivement moins contrastées par rapport au fond blanc à mesure que l'on descend dans le tableau. Les lettres de plus haut contraste sont en haut et celles de plus faible contraste en bas. Le tableau est présenté à distance standardisée (100 cm) et la tâche consiste à lire les lettres depuis le haut vers le bas. Le score est donné par la valeur du dernier groupe qui est lu. L'échelle est constituée de valeurs allant de 0 à 2,25. Une vision normale correspond à un score de 2,0 alors qu'un score inférieur à 1,5 est associé à une déficience visuelle.

Bien que moins utilisés que le test de Pelli-Robson, il existe d'autres tests standardisés de la mesure de la CSF : le Functional Acuity Contrast Test (FACT) et le CSV-1000E.

3. La coma désigne une aberration optique sphérique asymétrique. Elle se traduit par la présence sur les images d'une queue évoquant celle d'une comète, d'où son nom.

(Jinabhai et al., 2013)

2.4 Modèles de CSF

Plusieurs fonctions ont été proposées pour rendre compte de la valeur de la CSF en fonction de la fréquence spatiale. Les plus utilisées sont celles de (Daly, 1992; Kelly, 1975; Mannos & Sakrison, 1974). Elles sont représentées dans le tableau 1.3 et leur représentation graphique est tracée sur la figure 1.6.

Modèle	Expression avec α la fréquence spatiale
Mannos et Sakrison (1974)	$CSF(\alpha) = 2,6(0,0192 + 0,144\alpha)e^{-(0,0144\alpha)^{1,1}}$
Kelly (1975)	$CSF(\alpha) = \alpha^2 e^{-\alpha}$
Daly (1992)	$CSF(\alpha) = \left(\frac{0,008}{\alpha^3 + 1}\right)^{-0,2} 1,42\alpha e^{-0,3\alpha} \sqrt{(1 + 0,06e^{0,3\alpha})}$

TABLE 1.3 – Fonctions d’ajustement à la CSF

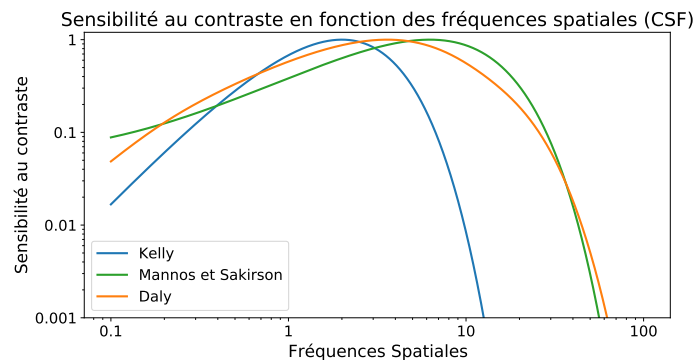


FIGURE 1.6 – Comparaison des différentes fonctions d’ajustement à la CSF. Figure par J. Buisine.

2.5 Rôle de l’apprentissage dans la CSF

Il a été rapidement détecté que la sensibilité au contraste évolue au cours de la vie.

Banks et Salapatek (1976) enregistrent la CSF auprès de nourrissons de 2 mois. La mesure de CSF à cet âge requiert une procédure adaptée. Pour cela, deux images sont présentées à des nourrissons et le temps passé à fixer l’une et l’autre est mesuré. S’ils passent un temps significativement plus élevé à regarder l’une plutôt que l’autre, il est vraisemblable qu’ils ont perçu une différence. Pour cette mesure de CSF, l’une des images comporte une grille dont le contraste et la fréquence spatiale sont contrôlés. L’autre image est un fond uni. Les valeurs mesurées diffèrent fortement de celles enregistrées chez l’adulte. Le pic de sensibilité apparaît pour des valeurs nettement plus basses (de

l'ordre de 0,3 cycle par degré) et pour des valeurs de contraste nettement plus élevées que pour les adultes.

Cette constatation amène à se poser la question du rôle de l'apprentissage dans l'évolution de la CSF. La question est de savoir si la sensibilité d'un individu peut évoluer au cours de sa vie sous l'effet de l'expérience. Pour répondre à cette question, Li et al. (2009) utilisent un protocole qui implique des jeux vidéos. Ils montrent que la pratique des jeux vidéos d'action peut entraîner une amélioration de la CSF.

Dans cette étude, les auteurs comparent l'effet d'un entraînement de 50h de jeux vidéos d'action (Unreal® Tournament et Call of Duty® 2) comparé à celui un entraînement sur un jeu vidéo qui n'est pas un jeu d'action (The Sims 2). Une amélioration significative des performances à un contraste de 6 cycles par degré est enregistrée. Ces améliorations ont été induites en vision centrale chez de jeunes adultes en bonne santé à un moment où leur capacité visuelle est censée être optimale.

L'effet provoqué est durable : il peut être mesuré à plusieurs mois voire à plusieurs années d'écart. Bien que les auteurs soulignent que les mécanismes qui permettent cette évolution soient peu connus, ils suggèrent que l'affinement, l'amélioration du gain ou le réajustement du modèle par le biais de la rétroaction et/ou des connexions latérales jouent un rôle.

2.6 Artistes

Il est possible d'utiliser le maximum de sensibilité au contraste pour encoder deux images. Le principe, détaillé par Oliva et al. (2006) consiste à séparer les contenus fréquentiels de deux images en appliquant un filtre passe-haut pour l'une et un filtre passe-bas pour l'autre. Les deux images sont alors combinées en une seule. Suivant la distance de présentation, c'est l'une ou l'autre image qui sera perçue ; c'est-à-dire celle dont la fréquence spatiale est la plus proche de 8 cycles par degré. Ce principe est utilisé pour fusionner les deux portraits de la figure 1.7.

Récemment, Straßer et al. (2020) utilisent cette technique pour une campagne au profit du WWF. C'est cependant Salvator Dali qui a exploité cet effet en premier en réalisant un portrait de « Gala nue regardant la mer qui à 18 mètres laisse apparaître le président Lincoln »⁴.

4. « *Gala desnuda mirando el mar que a 18 metros aparece el presidente Lincoln* » daté de 1975, Huile et photographie sur panneau de bois contreplaqué. 445 x 350 cm. © Salvador Dalí, Fundació Gala-Salvador Dalí, Figueres, Espagne.



FIGURE 1.7 – CSF pour créer des images hybrides

Fusion de deux portraits afin de composer une image hybride. Vue de près, elle semble représenter d'Albert Einstein alors qu'en s'éloignant c'est celui de Marilyn Monroe qui est visible. (Domaine public)

En Résumé

Le système visuel est caractérisé par sa sensibilité. La capacité à percevoir les détails correspond à l'acuité. La mesure de la sensibilité au contraste permet d'étudier les facteurs qui la déterminent : la fréquence spatiale joue un rôle important. Il est notamment possible de modéliser la sensibilité au contraste en fonction de la fréquence spatiale par le biais de la CSF (*Contrast Sensitivity Function*). Mieux connaître les limites du système visuel permet d'adapter les périphériques de rendu.

Chapitre 2

Perception de l'espace

*« qu'un peintre soit instruit, autant
que possible dans tous les arts
libéraux, mais [...] surtout qu'il
possède bien la géométrie »*

Leon Battista Alberti, *De Pictura*,
1436

Le monde dans lequel nous évoluons est en trois dimensions. Ce qui paraît une évidence pose néanmoins de nombreux problèmes au niveau de la perception. Le premier est que les surfaces réceptrices visuelles — les rétines — sont, elles, en deux dimensions. C'est-à-dire dire que notre perception du monde 3D passe par une étape 2D. C'est d'ailleurs cette étape qui permet de « tromper » le système visuel en lui présentant des images qui sont en 2D. Nous sommes néanmoins capables d'y lire une information de profondeur.

La vision binoculaire, la convergence oculaire et l'accommodation permettent d'enrichir l'information reçue et participent à la perception 3D. Ces informations seront développées ultérieurement pour ce chapitre, seules les informations monoculaires statiques seront envisagées. La perception de la distance reste possible avec uniquement ces informations bidimensionnelles. Elles sont souvent regroupées sous la désignation d'indices picturaux dans la mesure où les peintres et les artistes graphiques les ont utilisés parfois naïvement, parfois en lien avec les connaissances scientifiques disponibles à leur époque. Ces indices se retrouvent non seulement dans les dessins et peintures mais aussi dans les photographies ou images numériques. Ce sont des informations qui peuvent être perçues par un observateur statique avec un seul œil, avec le postulat que les mêmes indices sont impliqués lors de la perception d'une image en 2D et également dans la perception quotidienne d'environnements réels (Arcizet, 2007 ; Proffitt, 2006). Les mécanismes sous-jacents à la perception de la distance seraient donc les mêmes face à un tableau, face à une image, qu'elle soit numérique ou non, et face à une situation réelle.

Un second problème qui se pose très généralement au niveau de la perception est l'ambiguïté de l'image perçue. C'est-à-dire que l'image perçue à un moment donné peut être interprétée comme le résultat de nombreuses situations différentes. Il résulte de ceci que le monde tel qu'il apparaît à l'observateur est le résultat d'un processus complexe d'inférence, une sorte de pari sur ce que signifie la perception à un moment donné. Le système nerveux se comporte tel un enquêteur qui réunit un ensemble d'indices pour arriver à une conclusion qui constitue une inférence sur la situation.

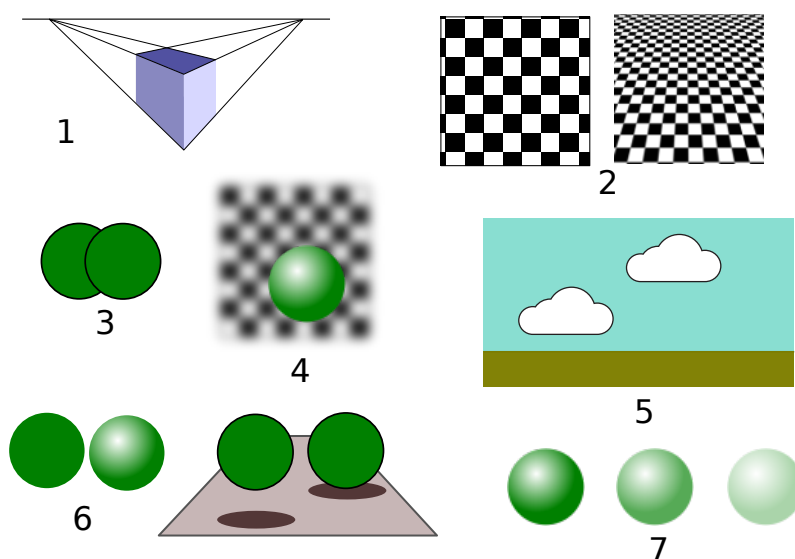


FIGURE 2.1 – Représentation schématique des principaux indices picturaux

Les principaux indices picturaux de perception de la distance sont : (1) la perspective linéaire, (2) le gradient de texture, (3) le recouvrement, (4) le flou de distance, (5) la hauteur relative à l'horizon ou élévation, (6) les ombres et ombrages et (7) la perspective aérienne

L'étude des indices perceptifs permet de renseigner les scientifiques sur l'impact de telle ou telle information. Elle permet aussi aux artistes de choisir de les représenter pour être le plus fidèle possible à la réalité ou bien au contraire de s'en jouer pour illusionner l'observateur, le spectateur. Il existe de nombreux indices qui ont fait l'objet d'études empiriques et qui sont utilisés par les arts graphiques dès lors qu'il s'agit de reproduire une impression de profondeur. Les peintres les ont souvent utilisés avec une grande maîtrise avant même qu'ils soient étudiés en laboratoire et bien avant qu'ils soient synthétisés afin de produire artificiellement des images et des environnements tridimensionnels.

Ce chapitre explore donc les sept principaux indices de perception de la profondeur en tentant de les situer dans le contexte artistique dans lequel ils ont émergé, les travaux expérimentaux qui les mettent en évidence et leur utilisation en imagerie numérique. Les indices explorés sont (Watson & Enns, 2012) :

1. la perspective linéaire
2. le gradient de texture
3. le recouvrement
4. le flou de distance
5. la hauteur relative à l'horizon ou élévation
6. les ombres et ombrages
7. la perspective aérienne

Ces indices sont définis et étudiés dans des parties spécifiques et la figure 2.1 en présente une illustration synthétique.

1 Perspective linéaire

La perspective linéaire constitue un ensemble de méthodes qui formalisent la construction d'une image en deux dimensions à partir d'un environnement en trois dimensions. Ces méthodes se sont développées initialement autour de la peinture et des arts graphiques. Les premières perspectives mathématiquement correctes sont apparues au xv^e siècle sous l'influence notamment de Leon Battista Alberti (Alberti, [1435] 1991) et de Léonard de Vinci.

1.1 Description

Les travaux relatifs à la perspective supposent généralement que l'image est formée sur une surface bidimensionnelle plate et que tous les rayons lumineux passent par un unique centre optique. Ce modèle, qui décrit de façon assez précise le fonctionnement d'un appareil photo ou d'une caméra, n'est que partiellement vrai pour le système optique humain. L'œil est en effet caractérisé par de nombreuses aberrations optiques. Par exemple, les rétines sont incurvées, le rapport focal sur ouverture est très faible (proche de 2) ce qui rend formellement invalide l'approximation des petits angles ou approximation de Gauss.

L'approximation de Gauss est un modèle simplifié qui permet de traiter un dispositif optique comme un système linéaire qu'il s'agisse d'un groupe de lentilles ou de miroirs. Pour cela, certaines conditions doivent être respectées : (1) l'angle d'incidence des rayons est petit par rapport à l'axe optique (2) les rayons sont paraxiaux, c'est à dire que leur point d'incidence est proche de l'axe optique.

Cependant, pour rendre plus facile la compréhension des mécanismes optiques, ce problème ne sera pas exploré. Pour la suite, le modèle choisi est que la perception résulte d'une projection d'un environnement en 3D sur un plan.

1.1.1 Modèle de la caméra de type sténopé

Le principe du sténopé ou caméra *pin-hole* est un dérivé de la chambre noire. Un trou de très petite taille est percé dans une plaque de faible épaisseur. Une image se forme sur la surface opposée à l'orifice. Il est même possible d'y placer un support photo-sensible afin de fixer l'image obtenue bien que ce soit un « appareil photo » de bien piètre qualité. La modélisation de ce dispositif optique élémentaire est une base pour comprendre le mécanisme de la perspective.

La figure 2.2 présente un modèle simplifié de perspective. Les points A et B , de coordonnées respectives (X_A, Y_A, Z_A) et (X_B, Y_B, Z_B) se projettent sur les points a et b du plan de projection. Les coordonnées 2D des points a et b sont respectivement (x_a, y_a) et (x_b, y_b) .

Afin de simplifier les notations, le centre de projection est confondu avec le centre du repère (O) et l'axe Oz représente l'axe optique du dispositif.

Le plan de projection \mathcal{P} , ou plan image, est parallèle au plan XY et il est séparé du centre optique de la distance f . La projection en perspective est définie comme la transformation qui associe à tout point A de l'espace 3D son image définie par l'intersection de la droite (OA) avec le plan \mathcal{P} de projection, si ce point d'intersection existe.

Ce modèle aboutit à une image renversée selon les deux axes X et Y (figure 2.2). Les points hauts se projettent en bas de l'image, les points bas en haut. Les signes de X_a et x_a sont donc opposés. Il en va de même pour Y_a et y_a . Cette propriété se retrouve d'ailleurs dans les dispositifs optiques qui reposent sur des lentilles ou des miroirs. Il est alors nécessaire de redresser optiquement les images ou d'un point de vue mathématique d'inverser les signes.

Mathématiquement, les coordonnées 2D peuvent se déduire des coordonnées en 3D :

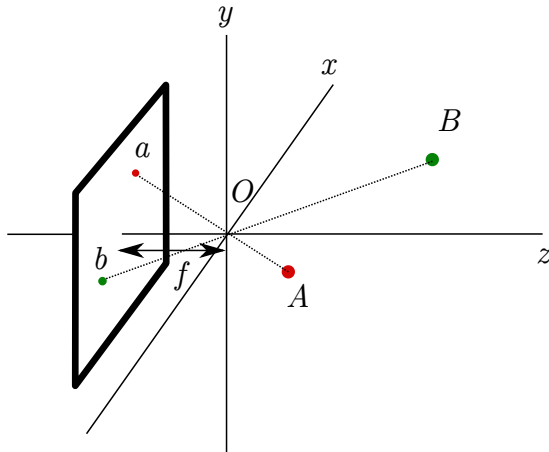


FIGURE 2.2 – Modèle de perspective « plan derrière ».

Les point a et b sont respectivement les images dans le plan de projection des point 3D A et B . Pour simplifier, le centre optique O est ici confondu avec le centre du repère.

$$\begin{cases} x = -f \frac{X}{Z} \\ y = -f \frac{Y}{Z} \end{cases} \quad (2.1)$$

On peut en déduire que :

$$\begin{cases} \Delta x = x_a - x_b = -\frac{f}{Z}(X_A - X_B) = -\frac{f}{Z}(\Delta X) \\ \Delta y = y_a - y_b = -\frac{f}{Z}(Y_A - Y_B) = -\frac{f}{Z}(\Delta Y) \end{cases} \quad (2.2)$$

Ce qui veut dire que le facteur d'agrandissement est proportionnel à la distance focale f . C'est ce même principe qui est utilisé par exemple avec un télé-objectif en photographie. Dans ce cas, l'augmentation de la longueur focale amène à une augmentation du facteur d'agrandissement. A l'inverse, si f est supposé fixe, la quantité Δx diminue si Z diminue. Ceci est donc responsable de l'effet d'échelle entre la taille apparente et la distance.

1.1.2 Caméra virtuelle « plan devant »

Pour la production des images numériques, ce sont plus généralement des représentations de caméra virtuelle « plan devant » qui sont utilisées. Elles ont l'avantage de produire une image naturellement dans le même sens (figure 2.3). Le centre de projection est à nouveau placé à l'origine mais cette fois, le plan de projection s'interpose entre les objets et l'observateur (ou la caméra).

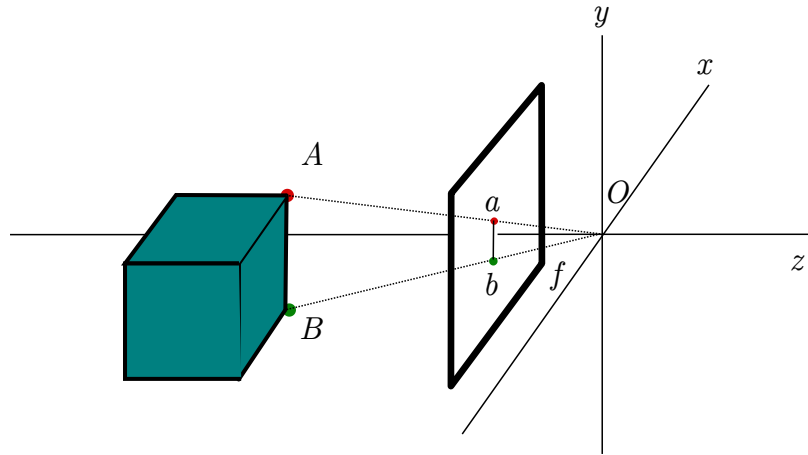


FIGURE 2.3 – Modèle de perspective « plan devant ».

Comme son nom l'indique, le plan de projection est maintenant face à l'observateur. L'un des avantages de cette projection est de conserver les orientations.

1.1.3 Effet d'horizon

L'une des propriétés les plus notables de la perspective est l'existence de points de fuite qui résultent de l'apparente convergence de droites parallèles de l'espace 3D vers l'horizon. C'est d'ailleurs cette propriété qui permet de « construire » géométriquement la perspective.

Soit un point V qui s'éloigne d'un observateur sur un plan \mathcal{S} parallèle à XZ , le plan horizontal. L'observateur est en O surélevé de la distance h par rapport au plan \mathcal{S} tel que représenté sur la figure 2.4. L'angle θ dit d'élévation — ou de dépression s'il est négatif — entre les droites Oz et OV peut s'écrire :

$$\theta(d) = \tan^{-1} \left(\frac{h}{d} \right)$$

Lorsque la distance d tend vers l'infini, l'angle θ tend vers zéro :

$$\lim_{d \rightarrow +\infty} \theta(d) = 0$$

Quelle que soit l'élévation h , l'angle de dépression θ tend vers 0 lorsque d devient très grand. Ceci explique l'effet d'horizon.

Dans le cas de droites séparées par une distance horizontale Δx , leur projection dans le plan image sera donné par l'équation :

$$\Delta x = \frac{f}{d}(\Delta X) \quad (2.3)$$

Lorsque la distance d devient à nouveau très grande, et que la focale f est supposée constante, l'écart Δx tend vers 0 :

$$\lim_{d \rightarrow +\infty} \Delta x = 0 \quad (2.4)$$

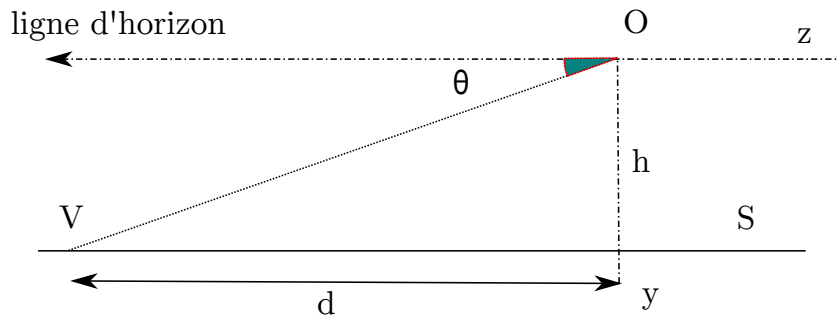


FIGURE 2.4 – Élévation

L'angle d'élévation — ou ici de dépression — θ représente la direction de visée depuis un point O à hauteur h en direction d'un point V séparés horizontalement par la distance d .

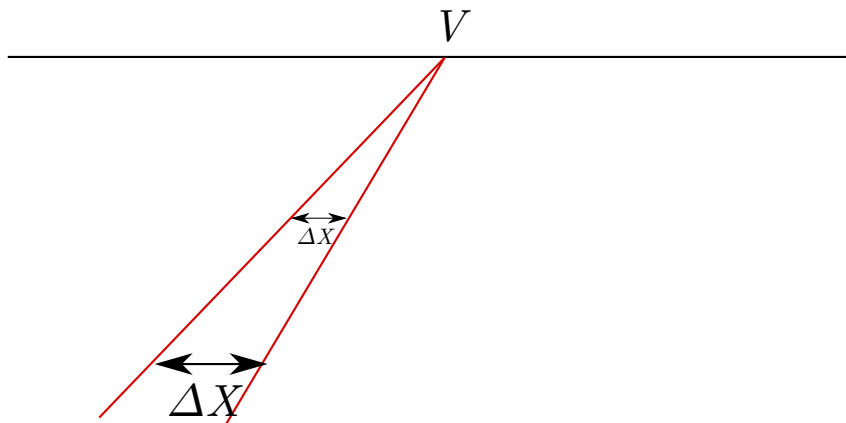


FIGURE 2.5 – Angle d'élévation

Deux droites parallèles se projettent comme deux demi-droites se joignant à l'infini. Leur point de convergence correspond à l'horizon de l'observateur.



FIGURE 2.6 – Points de fuite

Illustration des points de fuite de la perspective linéaire : la projection des droites parallèles converge à l'horizon (photo E. Delsart)

C'est-à-dire que les deux droites parallèles de l'environnement 3D se projettent comme deux demi-droites convergentes. Leur point de convergence se situe à l'horizon (figure 2.5).

Ceci rejoint les lois de la perspective linéaire (figure 2.6) qui peuvent être énoncées comme suit :

1. un ensemble de droites parallèles entre elles et parallèles au plan image a pour image un ensemble de droites parallèles ;
2. soit un ensemble de droites parallèles entre elles et non parallèles au plan image. La projection de ces droites forme un ensemble de demi-droites, d'origine commune, aussi appelé faisceau ;
3. le point de concours des demi-droites projetées sur l'horizon ne dépend que de la direction de l'ensemble de droites. Ce point est appelé point de fuite ;
4. l'ensemble des points de fuites des familles de droites parallèles entre elles qui ne sont pas parallèles au plan image forment la ligne d'horizon.

1.2 Perception

La perception des indices picturaux de façon générale est vraisemblablement le résultat d'un apprentissage. Pour la perspective linéaire, certaines observations vont dans ce sens.

Ainsi, les résultats de nombreuses études sur l'approche préférentielle suggèrent que les nourrissons réagissent pour la première fois aux signaux picturaux de profondeur entre cinq et sept mois (Kavšek et al., 2009). Pour la perspective linéaire, Yonas et al. (1986) présentent des objets, des poupées, que des nourrissons peuvent attraper. Les jeunes enfants montrent une préférence pour les objets perçus comme plus proches et tentent de les attraper. Cette perception de la distance peut être manipulée par la présence d'un fond donnant une fausse



FIGURE 2.7 – Dispositif d’acquisition de la perspective

Le « perspectographe » ou portillon de Dürer est une fenêtre quadrillée qui, associée à une contrainte sur la position du regard, permet de reproduire la perspective - Domaine public

impression de perspective. La présence du fond impacte la préférence des bébés de sept mois en vision monoculaire mais n’a aucun impact sur ceux de cinq mois.

L’interprétation de ces études reste cependant compliquée puisque de nombreux facteurs évoluent dans la même période, en particulier l’image rétinienne qui évolue également au cours du développement.

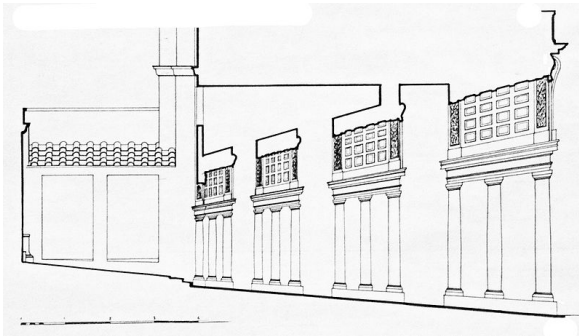
1.3 Artistes

1.3.1 Dans la peinture

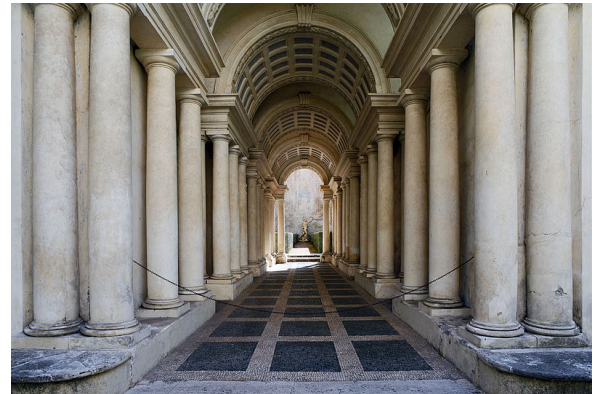
Classiquement la formalisation de la perspective remonte à la Renaissance ou un peu avant. L’invention de la perspective en tant que telle est attribuée à Brunelleschi (1377-1446). En 1415, pour représenter le baptistère Saint-Jean de Florence, il utilise un dispositif basé sur des miroirs afin de comparer son dessin à ce qu’il voit et vérifier que son dessin est bien fidèle à la perspective. Pour ce faire, il place son dessin sur une planchette dans laquelle il a percé un petit trou pour voir l’image du baptistère. En utilisant un miroir il peut comparer l’image réelle et sa production (Lynes, 1980).

Mais c’est surtout l’architecte Alberti (Alberti, [1435] 1991) qui propose une méthode géométrique et mathématique pour représenter la perspective dans son traité *Della Pittura*.

Pour l’école Flamande de peinture, il est généralement admis que Jean van Eyck ne connaissait pas la perspective. Néanmoins Simon (2021) effectue une analyse des points de fuite pour cinq tableaux que le peintre a réalisés dans les années 1430. Il note un motif qui revient de façon récurrente (« en arête de poisson ») qui semble montrer qu’une machine à perspective a été utilisée. Sur la base d’une étude statistique d’une part, et sur une modélisation 3D du célèbre « *Les époux Arnolfini* » d’autre part, l’auteur conclut que le peintre devait utiliser un outil de mesure afin de représenter au plus près la vision humaine : une « machine à perspective polyscopique » dont la forme exacte est inconnue puisqu’elle n’a malheureusement pas été conservée. Il compare ce dispositif à une sorte de réalité augmentée. Ce dispositif n’est pas sans rappeler le « perspectographe » utilisé par Albrecht Dürer pour représenter fidèlement la perspective d’une scène (figure 2.7). Comme noté par Hofmann (1990), ce dispositif et sa description s’apparentent au lancé de rayons actuel.



(a) Coupe de la galerie du Palazzo Spada à Rome. Wikimedia. CC-BY-2.5



(b) Photo : Livioandronico2013. Wikimedia. CC-BY-SA-4.0

FIGURE 2.8 – Perspective accéléré du palais Spada à Rome

Exemple de perspective accélérée dans la galerie du palais Spada (en italien *Palazzo Spada*) de Rome qui paraît occuper 30m si les colonnes étaient de même taille mais qui occupe en réalité moins de 9m. Les colonnes de l'arrière-plan sont plus petites que les premières de façon à simuler une perspective.

1.3.2 Fausses perspectives

Les artistes jouent également avec la perspective dans des domaines différents de la peinture. C'est le cas, par exemple, des anamorphoses et des fausses perspectives. L'utilisation de « trompe l'œil » se retrouve dès l'antiquité dans certaines fresques murales et dans des mosaïques.

Le temple du Parthénon à Athènes est en lui-même une forme de manipulation de la perspective. Les bâtisseurs avaient remarqué que lorsque un bâtiment est construit sur des lignes parfaitement rectilignes, celui-ci paraît déformé sous l'effet de la perspective. Les constructeurs du Parthénon ont donc utilisé une transformation géométrique pour « corriger » cette transformation apparente. Contrairement aux apparences, le bâtiment ne comporte que très peu de lignes parfaitement droites. Lorsque l'observateur voit les colonnes comme rectilignes, les architectes ont en réalité utilisé des transformations subtiles aussi appelées « raffinements optiques ». Par exemple, pour éviter une impression d'affaissement, la façade est plus haute en son centre que dans les coins. Les variations si elles peuvent paraître modestes (environ 6 cm) imposent cependant que chaque pierre de l'ensemble soit taillée de façon quasi unique avec une précision millimétrique. La technique de l'entasis est utilisée. Elle consiste en une courbure convexe du profil des colonnes, sans doute dans le but de les faire sembler plus rectilignes pour compenser l'illusion d'optique qui les ferait paraître courbées.

Plus récemment, en 1632, la galerie des colonnes du palais Spada de Rome, construite par l'architecte Francesco Borromini, est l'un des exemples les plus connus de perspective accélérée. La galerie est construite sous la forme d'une perspective dite forcée ou accélérée de façon à simuler une galerie de plus grande taille (figure 2.8b).

2 Gradient de texture

2.1 Description

Le gradient de texture est la distorsion que présentent les textures des objets du fait de leur orientation. Même si la déformation est de même type que celle induite par la perspective linéaire, il est d'usage de l'isoler tant pour des raisons graphiques que perceptives.

De plus, la perspective linéaire décrit davantage une organisation globale de la scène alors que le gradient de texture décrit plutôt à des variations locales de l'organisation des textures. Dans ces variations locales, un changement abrupte est généralement le signe d'un contour. En revanche, un changement graduel est plutôt relié à des variations dans l'orientation des objets.

2.2 Perception

Il a été établi depuis longtemps que le gradient de texture est un indice important pour la perception des distances (J. J. Gibson, 1947 ; Weinstein, 1957). Mais c'est plus récemment que Tsutsui et al. (2002) montrent qu'il existe des corrélats neuronaux de la perception de la profondeur à partir du gradient de texture dans le cortex. Dans une région précise du cortex — la partie caudale de la rive latérale du sillon intrapariétal — se trouvent des neurones sélectifs à l'orientation de la surface 3D. Leur réponse semble invariable pour différents types de motifs de texture. Ils se comportent donc comme de véritables détecteurs d'orientation, indépendamment du pattern présenté. Cependant la réponse de ces neurones n'est pas spécifique, puisque la plupart d'entre eux sont également sensibles à un gradient de disparité impliqué dans la stéréo-vision. Cela suggère qu'ils intègrent notamment les signaux de texture et de gradient de disparité pour construire une représentation généralisée de l'orientation de la surface 3D.

2.3 Artistes

Les peintres utilisent abondamment les variations de textures pour produire la sensation de distance. La capacité à déformer les textures fait même souvent partie du savoir-faire de l'artiste, de sa technique.

Par exemple, le dégradé de texture est soigneusement utilisé dans le tableau « *Rue de Paris, jour de pluie* » de Gustave Caillebotte (figure 2.9).

Ce tableau est reconnu pour ses dimensions exceptionnelles (212×276 cm) et sa géométrie. Celle-ci repose sur une construction de la perspective autour de deux points de fuite pour les bâtiments et de la ligne d'horizon au niveau des regards. Comme le souligne Varnedoe (1988), l'artiste a particulièrement travaillé la texture des pavés et a réalisé plusieurs études préparatoires sur ce point spécifique (figure 2.9b). Bien qu'apparemment vide, cet espace contribue fortement à la perception de l'espace dans ce tableau.

2.4 Informatique graphique

La technique du mappage de textures est également utilisée abondamment en informatique graphique car elle permet d'introduire des informations de détails sur un modèle 3D. Les textures informatiques sont généralement stockées dans des fichiers à part des modèles géométriques ou bien générées algorithmiquement par fonctions mathématiques : des textures procédurales.

La première application utilisant des textures a été développée par Catmull (1974). Initialement, le mappage de texture concerne uniquement la mise en correspondance des pixels avec



(a) Rue de Paris, temps de pluie Peinture à l'huile sur toile de Gustave Caillebotte, 212 × 276 cm, 1877, Chicago, Art Institute – Domaine public.



(b) Étude préliminaire de Gustave Caillebotte – Domaine public.

FIGURE 2.9 – Rue de Paris.

Tableau de Gustave Caillebotte et l'une de ses études préliminaires sur le gradient de texture des pavés.

une image purement diffuse projetée sur une surface 3D. Cependant, l'évolution des techniques de rendu s'est accompagnée d'une grande évolution des techniques de texturage. C'est le cas, par exemple, avec l'utilisation de mappage de cartes de normales, de hauteurs, de cartes de relief (*bump mapping*) et de nombreuses autres variations de ces techniques.

Enfin, l'introduction de matériels spécifiquement dédiés aux graphiques 3D a permis une utilisation massive et rapide des textures afin de les intégrer en temps réel dans les images. Le traitement des textures, leurs déformations et leurs filtrages représentent ainsi une part importante du pipeline graphique.

3 L'occultation ou interposition

3.1 Définition

Lorsqu'un objet est placé géométriquement devant un autre objet, ce dernier peut être partiellement occulté par le premier. L'occultation partielle renseigne donc sur l'ordre des objets dans une scène à partir du point de vue. Selon la géométrie des objets, il se peut que ce soit une partie de l'objet lui-même qui en occulte une autre partie.

Bien que le mécanisme optique qui provoque cette occultation soit élémentaire, il semble que ce soit un indice très puissant de perception des distances. Il s'agit d'un indice relatif puisqu'il ne porte pas d'information de distance dans l'absolu, mais permet de situer un objet relativement à un autre que ce soit devant ou derrière, lorsqu'il y a occultation partielle.

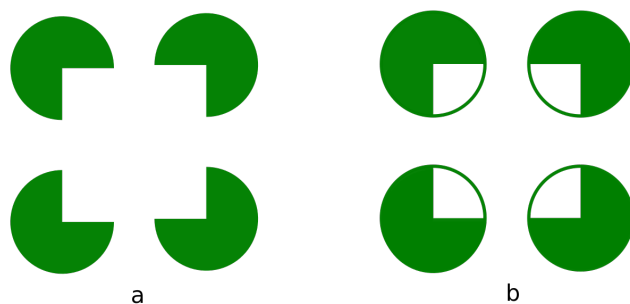


FIGURE 2.10 – Exemples de complétion visuelle.

Dans la complétion visuelle modale (a), les limites du carré semblent émerger. Par contre, dans la complétion amodale (b) les coins d'un carré semblent vus au travers de trous. On considère généralement que la situation de complétion modale est plus prégnante que la situation amodale.

3.2 Perception

Les résultats expérimentaux tendent à prouver que l'indice du recouvrement domine les autres, y compris face aux indices stéréoscopiques. Il est possible de créer une situation expérimentale artificielle dans laquelle le recouvrement et d'autres indices sont discordants. C'est le cas, par exemple, avec un stéréogramme comportant deux barres l'une verticale et l'autre horizontale. Le stéréogramme est calculé de façon à ce que l'une des barres paraisse plus loin mais qu'elle ne soit pas occultée par la barre proche. Les réponses des participants sont enregistrées pour connaître l'indice privilégié. Dans ce cas, c'est l'interposition (Hillis et al., 2004 ; Knill & Saunders, 2003) qui prévaut.

La capacité à percevoir correctement des objets partiellement occultés démontre la capacité constructive du système visuel humain. À partir d'informations partielles, le système permet de prendre des décisions avec une fiabilité remarquable (Burke, 1952). La « complétion perceptuelle » désigne l'ensemble des mécanismes qui conduisent à percevoir un objet alors qu'une partie du stimulus n'est pas perceptible. Ils constituent une forme d'interpolation visuelle (Kellman, 2003) qui permettent une représentation stable et continue, parfois en l'absence de délimitations perceptibles.

Il est d'usage dans la littérature de distinguer la complétion modale et la complétion amodale (Burke, 1952 ; Kanizsa, 1979 ; Kellman, 2003 ; Michotte & O., 1964). Même si cette distinction est pour partie critiquable (van Lier & Ekroll, 2020), la complétion modale se produit lorsqu'un objet du premier plan est « camouflé » par rapport à l'arrière plan. La complétion amodale désigne le fait que des objets sont complétés alors qu'ils sont occultés partiellement par d'autres objets. La figure 2.10 présente ces deux types de complétion.

Le motif de Kanizsa (Kanizsa, 1976), par exemple, intègre à la fois une complétion modale et une complétion amodale (voir figure 2.11).

3.3 Utilisation par les artistes

Bien avant la maîtrise et la formalisation de la perspective linéaire, les artistes utilisent l'interposition comme un moyen de figurer les positions relatives d'objets et de personnages. Par exemple dans « *La dernière cène* » de Duccio di Buoninsegna en 1308, la perception de

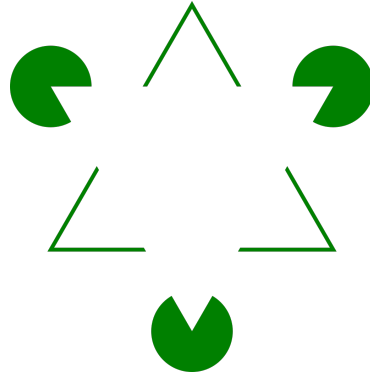


FIGURE 2.11 – Le motif de Kanizsa

Ce motif a pour particularité de faire apparaître non seulement une complétion modale : le triangle blanc semblant flotter devant. Il fait également apparaître une forme de complétion amodale : les disques au second plan.

profondeur est rendue quasi exclusivement par cet indice alors que la perspective dans son acceptation actuelle n'est pas respectée (figure 2.12).

3.4 Image numérique

Si dans le monde physique un objet proche occulte naturellement les objets plus lointains que lui, l'élimination des parties cachées d'une scène reste un problème en informatique graphique. En effet, dans des scènes complexes, le nombre de surfaces visibles représente une petite partie des surfaces du modèle (Teller & Sequin, 1991). Plusieurs problèmes se posent : certaines surfaces sont tout simplement en dehors du champ de la caméra virtuelle, certaines faces sont orientées dans la mauvaise direction et enfin des objets en occultent partiellement ou totalement d'autres.

Dans le cadre du rendu photo-réaliste, le calcul des rayons mobilise une fraction importante des ressources. Une partie du problème consiste à rechercher l'intersection la plus proche entre le rayon et les objets de la scène. Cette opération géométrique consiste à simuler le transport de lumière suivant les lois de la physique. Elle est relativement simple puisqu'on suppose à cette échelle que le trajet de la lumière est rectiligne. Mais dans la pratique ces opérations doivent être évaluées de très nombreuses fois. Dans la mesure où une scène peut être composée de millions de primitives et que les algorithmes doivent échantillonner de nombreuses fois les chemins de lumière afin de converger vers une solution stable, l'élimination des parties cachées depuis une direction est un enjeu important.

Le problème de l'élimination des faces cachées a été à l'origine de nombreux travaux et diverses techniques ont été proposées.

3.4.1 Élimination des faces arrières ou *Backface culling*

Une méthode élémentaire est d'éliminer les parties des surfaces dont la normale pointe dans la même direction que le regard de l'observateur. L'évaluation de la direction relative est une opération facile à réaliser puisqu'elle consiste à calculer le signe du produit scalaire entre la direction du regard et la direction du vecteur normal à la face. Ceux dont le produit scalaire est positif ou nul sont ainsi éliminés ce qui permet un gain d'environ 50% sur les autres calculs avec un surcoût très raisonnable.



FIGURE 2.12 – « *La dernière cène* » de Duccio di Buoninsegna. Domaine public.

Pour un objet convexe, la simple application du *Backface culling* est suffisante à éliminer toutes les parties cachées. Cette technique ne permet cependant pas d'éliminer les parties d'un objet concave ou celles qui sont cachées par d'autres objets.

3.4.2 Algorithme du peintre

L'algorithme du peintre consiste à calculer la valeur des pixels relativement aux objets du plus éloigné au plus proche par rapport à l'observateur dans l'espace 3D.

Dans le cas où la scène est constituée de polygones, ceci est réalisé en deux étapes :

1. les facettes sont triées de la plus éloignée à la plus proche ;
2. les facettes sont affichées suivant cet ordre et les parties les plus proches viennent recouvrir les parties éloignées.

Dans la pratique, l'algorithme est un peu plus compliqué car le tri lui même des facettes n'est pas forcément trivial et peut poser des problèmes de recouvrements partiels.

3.4.3 Z-buffer

L'algorithme du Z-buffer ou « tampon de profondeur » travaille sur l'écran 2D, dans l'espace de l'image, contrairement à l'algorithme du peintre, qui travaillait dans l'espace 3D ou espace objet. Pour chaque pixel, une information de profondeur est également enregistrée. Elle représente la profondeur de l'objet qui est affiché en un pixel à un moment donné du calcul. Lorsqu'un nouvel objet est projeté sur un pixel, la valeur de distance est comparée à la valeur actuelle. Si la distance est plus faible, la valeur est retenue et la valeur du Z-buffer est actualisée ainsi que celle du pixel. Dans le cas où il est plus loin, il est considéré comme non visible et n'est pas pris en compte.

Contrairement à l’algorithme du peintre, il n’est pas nécessaire de trier les objets, ce qui accélère les opérations. Historiquement, cet algorithme ne s’est pas imposé immédiatement pour des raisons de coût mémoire mais il est maintenant intégré matériellement aux cartes 3D.

3.4.4 Lancé de rayons

3.4.4.1 Principe

Le principe du lancé de rayon consiste à suivre le trajet d’un rayon depuis l’observateur — représenté par son œil ou une caméra. Ce rayon passe par les pixels qui représentent l’écran. Pour chacun, les intersections entre ce rayon et des objets de la scène sont calculés. La première surface rencontrée est visible dans le cas d’un matériau opaque mais il est également possible de prendre en compte la notion de transparence. Cet algorithme a été implémenté pour la première fois en informatique par Appel (1968b) et il constitue un algorithme efficace d’élimination des parties cachées.

Cependant, dans la mesure où l’exploration exhaustive des différentes surfaces de la scène est impossible en terme de temps, il est nécessaire d’utiliser des structures accélératrices. C’est le cas, par exemple, des hiérarchies de volumes englobants et des partition binaires de l’espace.

3.4.4.2 Hiérarchie de boîtes englobantes

Les hiérarchies de volumes englobants ou *BVH* pour *bounding volume hierarchy* se sont imposées dans de nombreuses applications (Meister et al., 2021) : dans un premier temps, les objets sont remplacés par leur boîte englobante (*bounding box*) qui sont des boîtes virtuelles qui entourent les objets dont les intersections sont recherchées comme illustré par la figure 2.13. Il est alors beaucoup plus rapide de rechercher dans un premier temps les intersections avec la boîte englobante (opérations élémentaires). Si une intersection est trouvée avec la boîte englobante, le calcul d’intersection avec le volume lui-même est alors réalisé.

Dans un second temps, les boîtes englobantes sont rangées dans une structure d’arbre binaire dans lequel les boîtes englobantes des objets constituent les feuilles. Les nœuds représentent de nouvelles boîtes englobantes qui en regroupent d’autres. Il existe différentes techniques pour construire des BVH. La méthode SAH (*Surface Area Heuristic*) est reconnue pour son efficacité dans la recherche des intersections. Dans cette méthode, la construction se fait de façon récursive à partir de la scène d’origine qui est partitionnée en utilisant une stratégie gloutonne afin de minimiser le coût de traversée (Pharr et al., 2016).

Enfin, lors de l’utilisation, c’est l’intersection avec les boîtes de plus haut niveau qui est débord recherchée et une branche n’est parcourue que s’il y a une intersection. Lorsqu’elle n’est pas parcourue, elle offre donc la possibilité d’exclure un grand nombre d’objets rapidement.

L’un des problèmes de ce type d’algorithmes est d’engendrer un surcoût qui correspond à la phase de construction de la structure accélératrice.

Pour résoudre ce problèmes, il existe des solutions pour optimiser la construction du BVH lui-même. (Gu et al., 2013) proposent d’utiliser une méthode d’agglomération basé sur des approximations successives : *Approximate Agglomerative Clustering* (AAC). L’algorithme est efficace et facilement parallélisable pour générer des hiérarchies de volumes de haute qualité. L’idée de l’AAC est de calculer une approximation de la véritable solution de regroupement. Ceci est réalisé en restreignant l’ensemble des candidats inspectés lors de l’identification de la géométrie voisine dans la scène. Le résultat est un algorithme simple qui produit généralement des hiérarchies de meilleure qualité. Cette qualité est mesurée sur le coût du ray tracing ultérieur. La méthode se révèle meilleure qu’une construction SAH de balayage complet, tout en s’exécutant en moins de temps que l’algorithme de construction SAH.

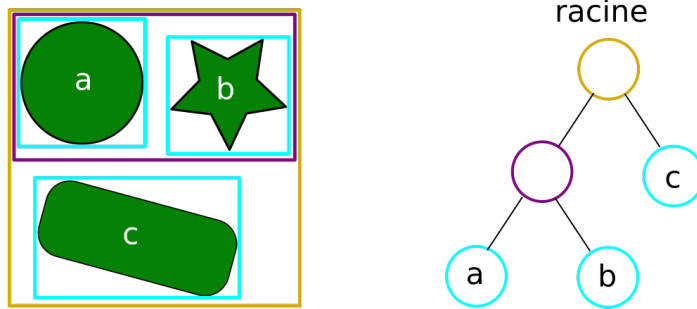


FIGURE 2.13 – Hiérarchie de volumes englobants

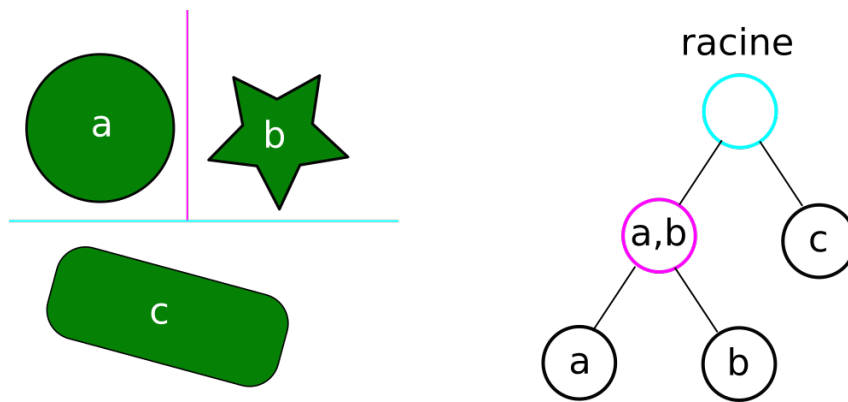


FIGURE 2.14 – Principe du BSP-tree

3.4.4.3 Partition binaire de l'espace

Une autre structure d'accélération classique est l'arbre de partitionnement binaire de l'espace : BSP pour *binary space partitioning*. Avec l'algorithme BSP, la scène est séparée en plusieurs régions disjointes : des cellules. Cela est réalisé par des plans de division à travers la scène. Chacun de ces plans sépare la scène en deux régions qui sont à leur tour divisées de manière récursive. Cette procédure est appliquée jusqu'à ce qu'un certain critère final soit atteint.

Les performances des structures accélératrices sont fortement dépendantes du type de scène, de leur implémentation logicielle et des accélérations matérielles possibles (Chung et al., 2019 ; Vinkler et al., 2015). Néanmoins, l'utilisation d'un arbre BSP est généralement plus rapide que celle d'un BVH. Ce gain lors de l'utilisation a comme revers de s'accompagner d'un temps de construction de l'arbre plus important.

4 Le flou de profondeur

4.1 Description

Dans cette partie, seuls les aspects liés au flou à proprement parler sont envisagés. La variation de longueur focale de l'œil lors de l'accommodation sera envisagée ultérieurement puisqu'il s'agit d'un indice non visuel et qu'il est en partie relié à la vision stéréoscopique.

Le flou de profondeur désigne le fait que les objets plus ou moins éloignés du plan focal sont plus ou moins flous, ce qui signale des variations de profondeur dans la scène.

Les systèmes optiques — y compris les yeux — produisent des images partiellement nettes. Le comportement d'un système optique lors de la projection d'une image peut être assimilé au modèle de la lentille mince sous l'approximation de Gauss. Pour la suite, les ensembles de lentilles ou de miroirs seront représentés par une unique lentille caractérisée par une longueur focale.

Afin de comprendre le fonctionnement du système optique en situation défocalisée, il faut d'abord décrire la géométrie d'un système focalisé.

4.1.1 Image focalisée

Pour qu'une image soit nette, il est nécessaire que l'image se forme dans un plan à une distance d_I de la lentille ou du groupe de lentilles. Pour cela, tous les rayons issus d'un même point de l'objet doivent se focaliser en un même point de l'image. Soit une lentille de centre optique O et de focale $f = OF$ et un objet AB dans le plan parallèle au plan des lentilles. A est supposé aligné avec l'axe optique et éloigné d'une distance $OA = d_O$ du centre optique (figure 2.15).

Un image focalisée se projette en $A'B'$ sous condition que :

$$\frac{1}{OF} = \frac{1}{OA} + \frac{1}{OA'}$$

Ou encore :

$$\frac{1}{f} = \frac{1}{d_O} + \frac{1}{d_I}$$

Le plan parallèle au plan de la lentille et passant par $A'B'$ est appelé plan de focalisation. Il dépend non seulement de la focale mais aussi de la distance de l'objet à la lentille. Dans le cas où le point A n'est pas dans l'axe optique le même raisonnement peut être appliqué avec la projection orthogonale de A sur l'axe optique. Par ailleurs, si l'objet n'est pas dans un plan parallèle au plan des lentilles, l'ensemble de l'image ne pourra pas être focalisé sur un plan orthogonal à l'axe optique et un flou apparaîtra.

4.1.2 Image défocalisée

Lorsque l'image est formée devant ou derrière le plan focalisé (figure 2.16), la projection d'un point n'est plus un point mais une tâche de lumière appelée *cercle de confusion* induisant ainsi un flou qu'on qualifiera de flou de défocalisation. Afin de quantifier la taille de ce cercle de confusion, il faut prendre en compte deux facteurs supplémentaires : la distance entre le plan image et le plan de focalisation d'une part et l'ouverture du système optique d'autre part.

Lorsque la distance objet s'éloigne d'une quantité Δd_O , le diamètre s de la tâche de lumière est déterminé par f la focale du dispositif optique, a l'ouverture qui va conditionner la quantité de rayons divergents, d_O la distance à l'objet focalisé et la distance et $D_O = d_O + \Delta d_O$ (Bove, 1989).

$$s(d_O, D_O, f, a) = \left| a \frac{f(D_O - d_O)}{d_O - f} \right|$$

La *profondeur de champ* désigne l'ensemble des points qui sont focalisés dans l'image.

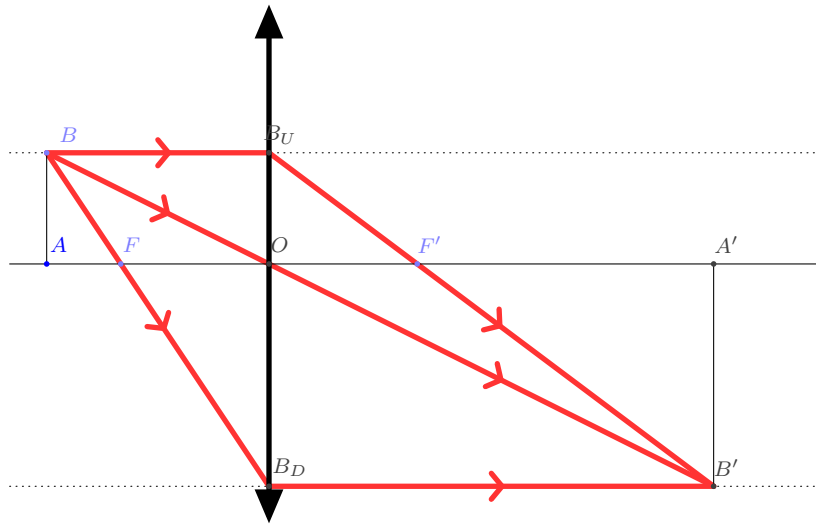


FIGURE 2.15 – Système optique focalisé.

Tous les rayons issus de B se projettent en B' , quel que soit le trajet optique. Selon les lois de conjugaison de Descartes, les rayons qui passent par le centre optique ne sont pas déviés, les rayons parallèles à l'axe optique avant la lentille passent par le foyer image (F'), les rayons qui passent par le foyer objet (F) ressortent parallèles après la lentille.

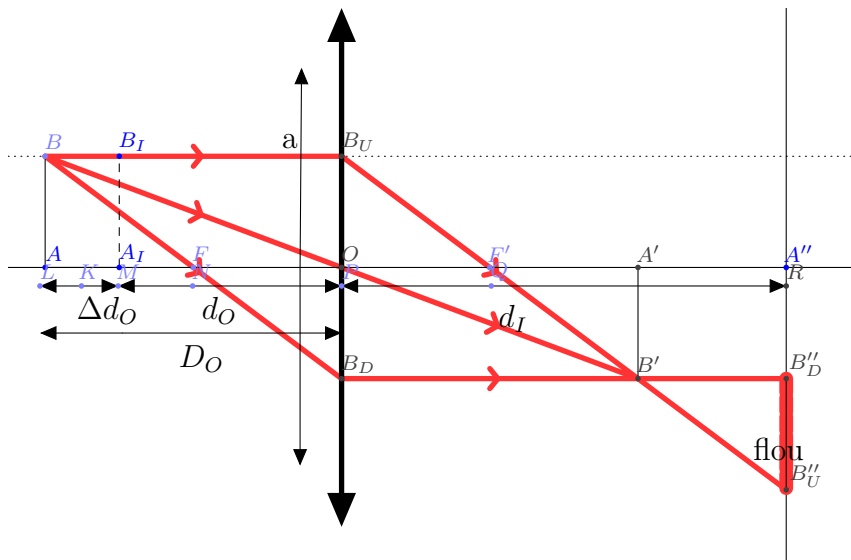


FIGURE 2.16 – Système optique défocalisé.

Les mêmes lois de construction que pour l'image focalisée sont appliquées, à la différence que l'objet est éloigné d'une quantité Δd_O , ce qui provoque la formation de l'image focalisée devant la projection. Dans l'espace image, les rayons issus de B sont maintenant dispersés.

4.2 Perception

Même s'il est perçu, le signal du flou reste compliqué à interpréter pour plusieurs raisons. D'une part, le signe de la variation de profondeur n'est pas connu puisque la quantité de flou dépend de la valeur absolue de la variation de distance. D'autre part, cette variation peut résulter de nombreux autres facteurs optiques et physiologiques comme la profondeur de champ, le contenu fréquentiel de l'image, la forme des sources lumineuses, les défauts de la vision, *etc.*

Watt et al. (2005) montrent que le flou provoqué par la plage limitée des focales de l'œil est un indice utilisé dans la perception. Mooney et al. (2019) fournissent des indications sur la façon dont le système visuel réussit ou échoue à le résoudre. En particulier dans le cas où il existe plusieurs sources de dé-focus. Une ombre douce peut être perçue comme identique à une ombre dure perçue comme floue et réciproquement. Dans ce cas, d'autres indices sont utilisés pour désambiguïser la perception.

Même si ce résultat peut paraître contre-intuitif, les mécanismes de perception du flou sont en partie indépendants de ceux qui contrôlent l'accommodation de l'œil puisque le système visuel semble capable d'utiliser l'information du flou sans forcément rechercher l'accommodation (Águila-Carrasco et al., 2017).

4.3 Artistes

La photographie et le cinéma utilisent la profondeur de champ comme un moyen d'expression à part entière. L'utilisation combinée d'un objectif très lumineux et d'une grande ouverture permet de produire un effet de flou décrit comme doux et harmonieux appelé *bokeh*.

En revanche, dans la peinture, l'utilisation du flou reste assez marginale. Dans la peinture classique, c'est sans doute l'indice perceptif qui est le moins utilisé. Les peintres préfèrent généralement utiliser la perspective aérienne et parfois un léger flou comme le *Sfumato*. C'est pendant la période impressionniste que le flou fait réellement son apparition dans la peinture. Ceci correspond à un moment où la peinture se veut moins une copie de la réalité qu'une tentative de reproduire les perceptions. Ce mouvement dans la peinture se produit également en parallèle des progrès dans la compréhension de la vision, de l'optique et de la photographie.

Gerhard Richter réalise des œuvres photo-réalistes auxquelles il ajoute un effet de flou. Dans cette lignée, Philip Barlow (figure 2.17) artiste contemporain, utilise le flou dans sa peinture. Il peint des toiles qui comportent des images artificiellement et totalement floues. Il utilise le flou comme un moyen d'expression qui n'est pas sans rappeler l'effet de *bokeh*.

4.4 Informatique graphique

Les images de synthèse sont par défaut parfaitement nettes. Afin de reproduire les effets optiques de profondeur de champ des optiques et du système visuel humain, il est nécessaire soit de simuler en détail le parcours optique soit d'ajouter artificiellement la notion de profondeur de champ. L'introduction de la profondeur de champ dans les images numériques a un double rôle. D'une part elle augmente le réalisme des images et d'autre part, elle permet un contrôle des images produites. Tout comme pour la photo, la profondeur de champ fait partie de « l'écriture » de la production et les infographistes souhaitent pouvoir la contrôler à la fois pour des contraintes artistiques et narratives.

La netteté des images produites en informatique graphique provient de l'utilisation de la caméra sténopé ou *pinhole camera*. Demers (2004) recense cinq méthodes pour recréer le flou :

Ray-Traced Depth of Field ou profondeur de champ par traçage de rayons pour simuler la lentille. Cette méthode est la plus réaliste, elle est également réputée comme la plus lente



FIGURE 2.17 – Emerald. De Philippe Barlow

Peinture « floue » à la manière d'une photo défocalisée. 120×180 cm, Huile sur toile. Avec l'aimable autorisation de l'auteur.

mais l'apparition des algorithmes de raytracing dans les processeurs graphiques permet de l'accélérer.

Accumulation-Buffer Depth of Field avec cette méthode, la scène est rendue plusieurs fois avec une légère variation de la position de la caméra alors que le point visé par la caméra reste identique. Un buffer d'accumulation permet de fusionner les différentes images obtenues et donc de produire le flou de profondeur.

Layered Depth of Field Dans la technique de profondeur de champs par couche, une approche de type 2D est utilisée. A ceci est ajouté une notion de couche. Cette méthode est parfois qualifiée de 2D $\frac{1}{2}$. Le principe est que les objets sont triés en couches qui ne se chevauchent pas et chaque couche est rendue à part. L'image finale est composée avec un filtre flou dont la quantité est paramétrée par la distance au plan de focalisation. Ce qui donne une impression de profondeur de champs.

Forward-Mapped Z-Buffer Depth of Field la technique de mappage vers l'avant repose sur l'idée de conserver une correspondance entre la valeur des couleurs calculée dans l'image avec leur distance à l'observateur (z-buffer). L'image finale est reconstruite en remplaçant les pixels par des cercles (*sprites*) dont la taille est déterminée par le cercle de confusion. A l'inverse, la valeur d'alpha (transparence) est augmentée afin de simuler l'étalement de la lumière. Ce mécanisme n'est appliqué que pour les pixels les plus éloignés afin d'éviter que des pixels flous ne viennent masquer des pixels plus nets.

Reverse-Mapped Z-Buffer Depth of Field Dans cette technique de mappage inverse, une information de couleur et de distance est également utilisée pour chaque pixel mais les deux informations sont combinées dans une texture unique qui prend en compte l'écart avec le plan de focalisation.

Demers (2004) note que les différentes méthodes possèdent des avantages et inconvénients en matière de rapidité d'exécution et de réalisme. Si un résultat réaliste est souhaité au détriment des performances, les techniques de *ray tracing* ou à base de buffer d'accumulation sont meilleures, alors que les techniques à base de couches ou de mapping du z-buffer sont plus rapides, mais induisent des artefacts qui peuvent être gênants. Ces derniers se manifestent par des crénelages des ombres.

5 Distance à l'horizon

5.1 Description

La position relative d'un objet à l'horizon peut aussi servir d'indice perceptif. Sous l'hypothèse que l'observateur se déplace sur un plan, l'information angulaire sur l'élévation des objets suffit à renseigner sur la distance absolue. Comme mentionné dans la partie 1.1.3 l'angle θ suit la relation $\theta(d) = \tan^{-1}(h/d)$ avec h la hauteur et d la distance.

5.2 Perception

Sedgwick et al. (1983) ont proposé ce qu'ils appellent la relation horizon-distance comme l'un des mécanismes permettant à un observateur debout sur une surface au sol de retrouver la distance égocentrique absolue par rapport à des emplacements environnementaux situés également sur cette surface au sol.

Dans les environnements réels, la question se pose de savoir comment l'information relative à la position de l'horizon est perçue. Rand et al. (2010) démontrent expérimentalement que lorsque l'horizon réel n'est pas visible, les informations gravitationnelles sont utilisées. Ils constatent que dans des tâches de marche et de lancer dirigées visuellement, l'augmentation de l'information de l'horizon visible entraîne une sous-estimation de la distance absolue par rapport à un horizon non modifié.

La variation dans la distance perçue peut induire également une variation de la taille perçue d'un objet : selon l'hypothèse de l'invariance de la distance, lorsque deux objets situés à des distances différentes sont perçus sous le même angle visuel, le système visuel perçoit l'objet le plus proche comme plus petit et l'objet le plus éloigné comme plus grand. Ozkan et Braunstein (2010).

À l'inverse, lorsqu'un objet de taille apparente identique est perçu avec un angle d'élévation différent, l'objet le plus proche de l'horizon est perçu comme plus grand et plus éloigné que l'objet éloigné de l'horizon. Une information très limitée de profondeur comme des lignes convergentes suffisent à produire une variation de taille perçue comme c'est le cas dans l'illusion de Ponzo (Harré et al., 2012) représentée sur la figure 2.18.

L'information de la distance à l'horizon pourrait également expliquer partiellement une illusion très célèbre : l'illusion de la lune (Holway & Boring, 1940). Pour rappel, la lune semble avoir un diamètre apparent plus important lorsqu'elle est proche de l'horizon alors que ce n'est pas le cas des mesures optiques objectives. Lors des photographies il n'est pas observé pas de variation de taille de la lune lorsqu'elle est à son zénith ou à l'horizon. L'une des raisons de l'illusion serait que la lune proche de l'horizon semble alors plus éloignée et que le mécanisme de constance de la taille aboutit à la percevoir plus grande : l'inférence d'un objet grand à grande distance semble effectuée. Alors qu'à l'inverse lorsque la lune est plus haute, elle apparaîtrait comme plus proche et donc plus petite.

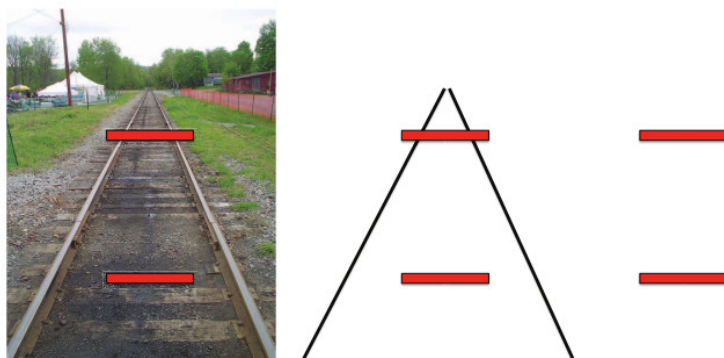


FIGURE 2.18 – Illusion de Ponzo

L'illusion est montrée successivement sur un fond de photographie et devant des lignes convergentes. Si le contexte n'est pas présent, l'illusion disparaît. Figure sous licence : Creative Commons Attribution-NonCommercial-ShareAlike 3.0 Unported. (Harré et al., 2012)

5.3 Artistes

Plusieurs artistes ont utilisé l'indice de position relative à l'horizon dans des contextes assez similaires à l'illusion de Ponzo. C'est le cas de Juan Luis Roldan (figure 2.19).

6 Ombres et ombrage

Bien que les ombres soient omniprésentes dans notre environnement, elles sont rarement remarquées.¹

Les ombres résultent de l'occultation de tout ou partie de l'éclairage des sources de lumière. Optiquement il s'agit de phénomènes relativement simples mais elles jouent un rôle important dans la perception visuelle d'une scène. Elles augmentent l'information visuelle et permettent de lever ambiguïté de l'image perçue depuis un point de vue.

6.1 Description

L'ombre peut se définir de façon relativement simple comme un masquage de tout ou partie du flux lumineux provenant des sources de lumière. Les ombres sont omniprésentes dans notre environnement puisque les objets possèdent quasiment toujours une ombre sauf dans des cas très particuliers. Au-delà de l'apparente simplicité des mécanismes optiques qui les produisent, une grande variété de formes est observée. Une partie de celles-ci peut être décrite par des termes précis présentés en figure 2.20 (Casati & Cavanagh, 2019) :

L'ombre propre est la partie d'un objet qui ne reçoit pas de lumière directement depuis une source de lumière.

L'ombre intrinsèque est une ombre créée par un objet sur lui-même dans le cas d'objets concaves.

1. « *Shadows are everywhere, but they usually go unnoticed.* » (Kersten & Mamassian, 2017)

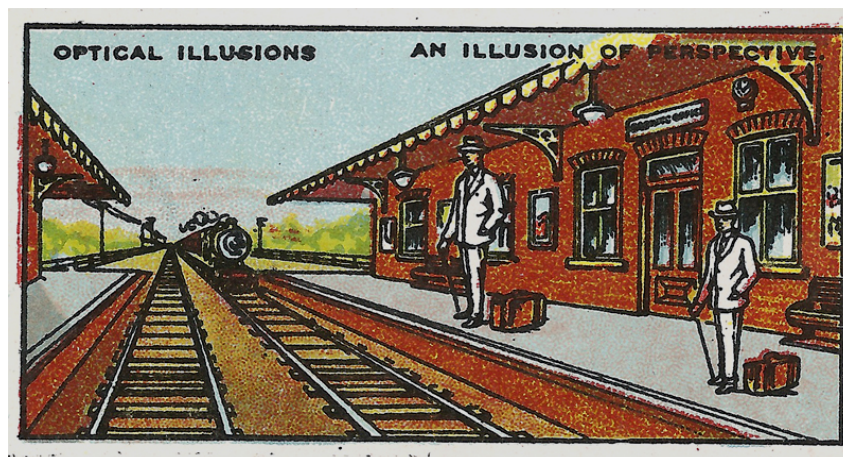


FIGURE 2.19 – Utilisation artistique de l'illusion de Ponzo

Utilisation de l'illusion de Ponzo par Juan Luis Roldan. Les deux personnages ont rigoureusement la même taille mais le contexte et l'effet de perspective semble indiquer le contraire. Reproduction sous licence Attribution-NonCommercial-ShareAlike 2.0 Generic (CC BY-NC-SA 2.0)

L'ombre portée est une zone sombre créée par l'interposition d'un objet devant une surface.

La pénombre désigne la zone de transition entre la partie éclairée et l'ombre. Dans le cas d'une source de lumière étendue, il existe une zone pour laquelle seule une partie de la source est occultée. Dans le cas où la zone de pénombre est présente, l'ombre paraît douce par opposition à une ombre dure produite par une source supposée ponctuelle.

L'ombrage désigne la variation de la teinte et de la luminance sur l'objet 3D en fonction de facteurs comme l'angle de la surface par rapport aux sources de lumière (distance et orientation) et les propriétés du matériau.

6.2 Perception

Les résultats de Knill et al. (1997) suggèrent que les ombres intrinsèques peuvent être utilisées pour déduire directement l'inclinaison de l'illuminant, la structure qualitative globale de la surface, et, aux intersections avec les plis de la surface, la concavité/convexité d'une surface.

Dans les situations d'ombrage ambiguë, Gerardin et al. (2010) montrent que les connaissances antérieures jouent un rôle. L'observateur effectue un certain nombre d'inférences pour résoudre le problème de l'ambiguïté. Ils explorent ce problème par une méthodologie qui combine indices comportementaux et mesures psycho-biologiques tels de l'IRM fonctionnel (fMRI). Ils démontrent par exemple que face à des figures ambiguës dont l'ombrage peut provenir de directions différentes, les sujets choisissent généralement un biais de réponse pour une source au dessus et à gauche des objets.

6.2.1 Informations sur la géométrie de l'objet

Les ombres apportent des informations sur la géométrie des objets en eux-mêmes (figure 2.21). Elles sont un facteur important qui contribue à enrichir la perception et donc à lever une

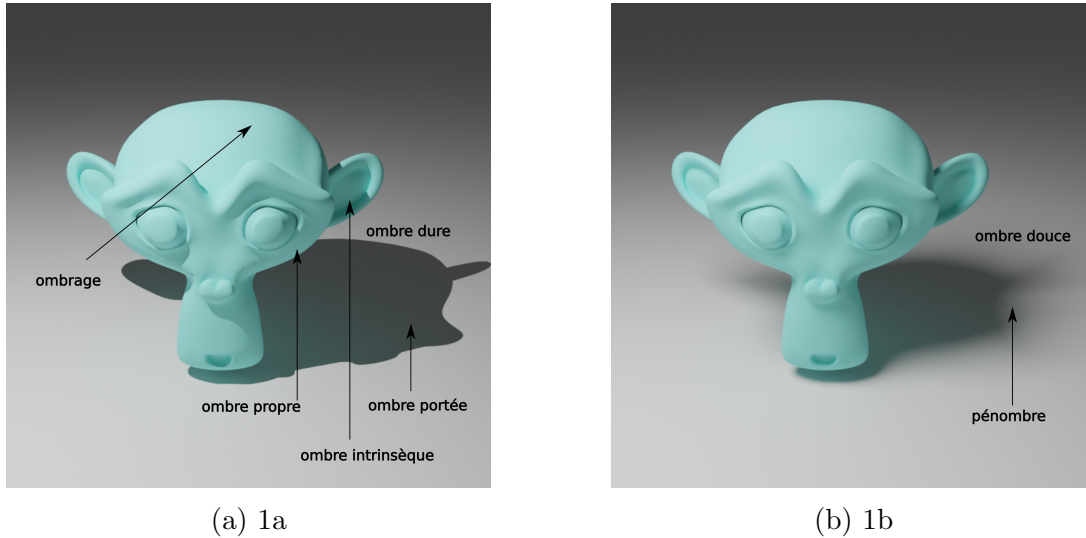


FIGURE 2.20 – Comparaison entre ombre dure et ombre douce

partie de l’ambiguïté de l’image perçue.

6.2.2 Informations sur la position des objets

D’autre part, les ombres apportent des informations importantes sur les positions des objets relativement aux surfaces sur lesquelles elles sont projetées comme l’illustre la figure 2.22. Kersten et Mamassian (2017) notent que les ombres portent une information importante sur la disposition spatiale des objets. De façon simplifiée, plus un objet est proche de la surface de l’arrière-plan, plus l’ombre est proche de l’objet projeté.

Yonas et Granrud (2006) indiquent que comme les autres indices picturaux, c’est également entre cinq et sept mois que les bébés apprennent à utiliser les indices liés à l’ombre portée. Pour cela, ils ont testé la préférence pour atteindre un objet dont l’ombre portée le fait paraître plus proche. Leurs résultats indiquent que les nourrissons de sept mois perçoivent la profondeur sur la base des ombres portées. Les enfants de cinq mois n’ont pas cherché à atteindre l’objet apparemment le plus proche et n’ont donc pas montré de sensibilité aux ombres portées comme information sur la profondeur.

Mamassian et al. (1998) montrent que quand un objet projette son ombre sur une surface d’arrière-plan, l’ombre peut informer sur la forme de l’objet, la forme de la surface du fond et la disposition spatiale de l’objet par rapport au fond. Parmi tous ces rôles, les auteurs constatent que les ombres projetées étaient perceptiblement les plus pertinentes pour l’interprétation de l’arrangement spatial — la position relative des objets —, en particulier lorsque l’ombre se trouve dans la zone de l’arrière-plan. Ce résultat semble intrigant lorsqu’on considère les ambiguïtés dans les façons possibles d’interpréter le mouvement de l’ombre. Aussi ils supposent que le système visuel doit utiliser des contraintes *a priori* pour lever ambiguïté sur le mouvement de l’ombre portée. L’une de ces contraintes est que la source de lumière est stationnaire. La variation de position d’une ombre est prioritairement interprétée comme une variation de la position de l’objet par rapport à un déplacement des sources. Bien que simple, la contrainte de la source lumineuse stationnaire permet des inférences riches et fiables sur les mouvements qualitatifs des objets en trois dimensions.

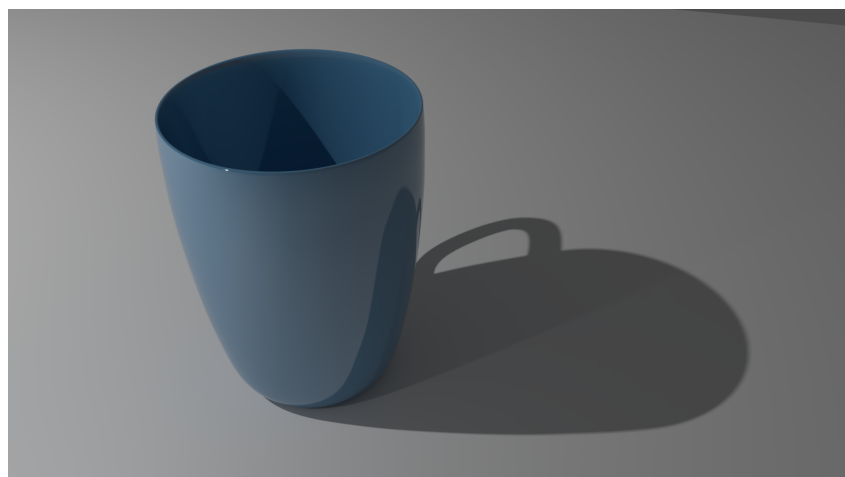
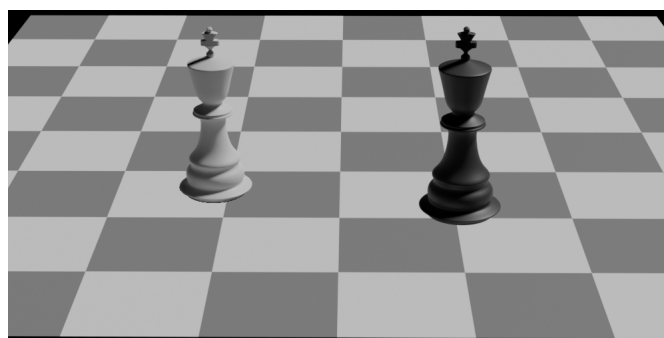
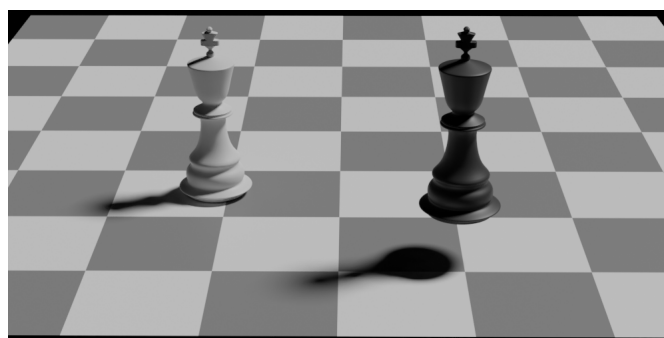


FIGURE 2.21 – Informations de l'ombre projetée

L'ombre projetée peut apporter une information qui n'est pas visible depuis un point de vue donné. Dans ce cas, l'ombre trahit la présence d'une anse qui ne serait pas visible sinon. L'ombre vient enrichir la perception d'un objet afin de mieux en comprendre la forme tridimensionnelle par une forme de second point de vue.



(a) En l'absence d'ombre projetée, la position des objets est ambiguë. Ceci produit une sensation de lévitation.



(b) La simple présence des ombres projetées suffit lever l'ambiguïté.

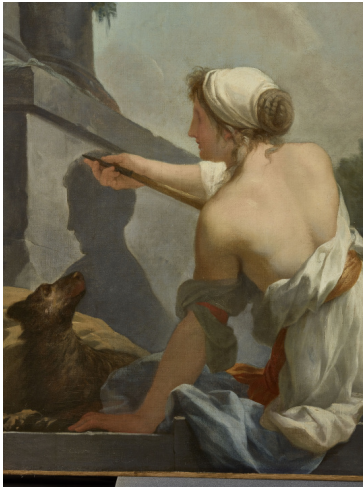
FIGURE 2.22 – Ambiguïté de l'image sans ombre

6.3 Artistes

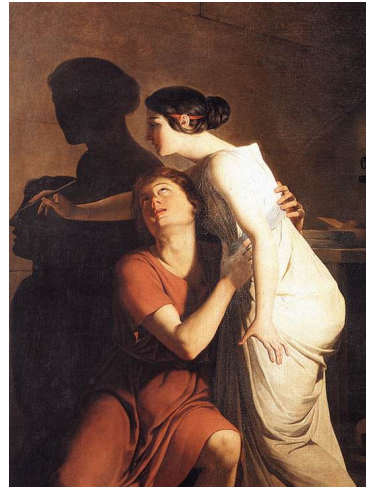
La représentation de l'ombre dans la peinture a beaucoup varié en fonction des époques et des cultures. Par exemple l'abbé Dubos note « *Les Chinois, si curieux des peintures de leur pays, ont peu de goût pour les tableaux d'Europe, où, disent-ils, on voit trop de taches noires. C'est ainsi qu'ils appellent les ombres.* » (Dubos, 1732). La peinture médiévale ne représente pas ou très peu les ombres des personnages, sans doute en partie à cause de raisons religieuses. Ce qui apparaît comme un double sombre des êtres est considérée comme impur et la peinture avant les années 1430 ne veut pas représenter la « part d'ombre », en particulier celle des personnages saints qui occupent une grande part de la peinture occidentale de cette époque.

Puis au cours des années 1425–1435, l'ombre fait son apparition dans la peinture chez plusieurs artistes : Masaccio à Florence, Giovanni di Paolo à Sienne et Van Hecke pour les peintres flamands. Au départ ténue, elle se renforce petit à petit dans la peinture occidentale. Leonardo Da Vinci (Vinci, 2012) avait noté que la distance entre l'ombre et le fond est une information importante quant à leur position relative et il encourage ses élèves à utiliser cet indice.

Au XVII^e siècle, en lien avec l'optique géométrique, les ombres se précisent. Mais, avec l'ombre, c'est surtout la lumière qui fait son apparition dans les tableau. C'est également sur l'utilisation de l'ombre qu'est bâtie un mythe fondateur de la peinture, celui de l'invention du dessin. Ce récit mythique rapporté par Pline l'ancien raconte qu'une jeune Corinthienne souhaite conserver les traits de son amant qui doit partir. Elle trace le contour de son visage, sa silhouette à même la pierre. Ce récit a donné lieu a de nombreux tableaux, notamment au XVIII^e siècle comme illustré en figure 2.23.



(a) Détail de *L'Origine de la peinture*. Jean-Baptiste Regnault (1754–1829). Domaine public.



(b) Détail de *L'Invention de l'art du dessin* — Joseph-Benoît Suvée (1743–1807). Domaine public.

FIGURE 2.23 – L'invention de la peinture

Mythe de Dibutade dans laquelle la jeune femme cherche à conserver l'image de son amant. Elle utilise la projection du soleil ou d'une lampe pour produire son contour. De nombreuses œuvres ont représenté cette scène au XVIII^e siècle. Ici les versions de Jean-Baptiste Regnault et Joseph-Benoît Suvée.

6.4 Informatique graphique

De la même façon que les peintres, l'image numérique a très vite intégré le rendu des ombres comme un aspect important de l'expérience visuelle. La détermination des ombres portées est donc un enjeu important dès le départ de l'informatique graphique.

6.4.1 Les volumes d'ombres (*shadow volumes*)

La méthode des shadow volumes a été proposée par Crow (1977). Elle consiste à calculer un volume d'ombre pour chaque objet qui produit une ombre : les « *shadow casters* ». Ce volume est obtenu par une extrusion infinie de la silhouette qui est vue depuis la source de lumière. Mais plusieurs volumes peuvent s'intersecter entre eux. Lors du rendu, il faut déterminer si le point rendu fait partie ou non du volume d'ombre. Pour ce faire, le nombre de fois où l'on entre ou sort des volumes d'ombre entre le point de vue de l'observateur et l'objet représenté par le pixel est compté.

Cette méthode est reconnue pour son exactitude puisqu'il est toujours possible de déterminer sans erreur si le point représenté appartient ou non à la zone d'ombre. De plus, la méthode est applicable avec n'importe quelle source de lumière ponctuelle (source directionnelle omnidirectionnelle ou spot).

Cependant l'implémentation de la méthode des shadow volumes est complexe : elle repose sur plusieurs étapes qui sont lourdes et de plus elle ne peut rendre compte des ombres douces.

6.4.2 Shadow map

Williams (1978) propose une méthode complètement différente qui se déroule en deux étapes.

- Dans un premier temps un rendu de la scène est effectué telle qu'elle serait vue depuis la source avec la technique du z-buffer qui permet de détecter les parties cachées. Ces parties appartiennent à l'ombre pour la source considérée. Ceci est stocké dans une *shadow map*.
- Dans un second temps, un rendu est effectué depuis la position de la caméra et pour chaque pixel qui doit être rendu, la valeur du z-buffer est récupérée dans la *shadow map*. Si elle est inférieure à la distance entre l'objet et la source, il existe au minimum un objet qui masque la source de lumière et qu'il n'est pas éclairé. Sinon le pixel est bien éclairé par la source de lumière.

Cette méthode s'avère plus rapide mais présente plusieurs problèmes. D'une part le fait d'utiliser deux cartes différentes avec des orientations différentes dans l'espace objet induit des approximations visibles sous forme de crénelage (*aliasing*) des ombres. D'autre part, il est difficile de représenter des sources omnidirectionnelles.

De nombreuses autres méthodes ont été proposées depuis et font l'objet d'optimisations (Hasenfratz et al., 2003 ; Woo et al., 1990). C'est le cas, notamment des partitions de volumes d'ombres proposées par Gerhards (2017).

6.4.3 Méthodes perceptives

Plutôt que d'essayer de produire des ombres réalistes, Sattler et al. (2005) s'intéressent davantage au niveau de complexité requis pour produire des ombres suffisamment détaillées pour être perçues comme acceptables par les utilisateurs. Leur objectif est d'utiliser des modèles simplifiés des objets de la scène afin de réduire le coût d'implémentation des ombres. Leurs résultats montrent que la génération d'ombres perçues comme satisfaisantes pour 90% des observateurs peut être effectuée avec des modèles utilisant seulement 1% de la géométrie des objets.

7 La perspective aérienne

7.1 Description

En plus des modifications liées à projection dans le plan 2D, la distance induit également des modifications chromatiques liées au milieu dans lequel se déplace la lumière.

Ainsi, la perspective atmosphérique se définit par une modification du spectre lumineux en fonction de la distance. Elle est caractérisée par une prédominance des teintes bleues avec la distance.

D'un point de vue optique, l'air qu'on peut supposer comme transparent sur des distances faibles peut également diffuser une partie de lumière lorsque la distance augmente. La diffusion de la lumière produit alors une sorte de voile lumineux qui réduit le contraste avec le ciel. La prédominance des courtes longueurs d'ondes est responsable de la dérive des teintes vers le bleu.

7.2 Perception

D'un point de vue perceptif, la perspective aérienne est étudiée sous le nom de *covariance proximité-luminance*.

Dosher et al. (1986) démontrent que cette information est utilisée lors de la perception. Leur méthode utilise l'illusion du cube de Necker qui peut être perçue de deux façons. Dans le cas où le cube est mis en rotation, l'illusion se caractérise par deux perceptions possibles. Dans un cas, un cube en rotation est réellement perçu alors que dans l'autre, c'est une déformation du cube qui est vue. Suivant le contexte, c'est l'une ou l'autre des perceptions qui domine. Les auteurs montrent que lorsqu'on ajoute une covariance luminance-proximité, cela suffit à faire disparaître l'ambiguïté perceptive.

7.3 Artistes

L'utilisation de la perspective aérienne est antérieure à la perspective linéaire. Thomas Corneille note

Outre la perspective linéale ou linéaire qui enseigne, comme nous venons de le dire, la diminution des lignes, les Peintres observent la perspective aérienne, qui consiste dans la diminution des teintes et des couleurs, selon le plus ou le moins d'éloignement des objets.

Thomas Corneille, Le dictionnaire des arts et des sciences, t. 2, 1732, p. 202.

Les maîtres flamands de peinture ont particulièrement développé cette technique au xv^e siècle. Jan Van Eyck l'utilise et plus encore, Joachim Patinier réputé pour ses peintures de paysages (figure 2.24).

7.4 Informatique graphique

Dans le cadre des applications de cartographie 3D, (Jenny & Patterson, 2020) notent que la perspective aérienne joue un rôle important pour guider l'utilisateur dans la perception des reliefs. Ces applications utilisent souvent des variations de contraste. Un fort contraste de luminance est associé à des variations importantes d'altitude. Dans ce cas, l'introduction de perspective aérienne permet d'augmenter la lisibilité et aide l'utilisateur à comprendre plus facilement la structure du paysage. Les auteurs proposent une méthode qui permet d'ajouter une perspective aérienne à un relief ombré.



(a) Paysage avec Saint Christophe, vers 1520, huile sur bois, 127×172 cm El Escorial, Monasterio de San Lorenzo, Espagne. Domaine public.



(b) Paysage avec Saint Jérôme, vers 1515, huile sur bois, 74×91 cm, Musée du Prado, Madrid, Espagne. Domaine public.

FIGURE 2.24 – Perspective aérienne en peinture

Joachim Patinier utilise la perspective atmosphérique pour représenter la distance, Cette technique est typique de la peinture flamande de la Renaissance.

8 Synthèse et discussion

Ce chapitre explore l'utilisation de différents indices statiques et monoculaires qui induisent une perception de distance. La séparation des différents indices est souvent pour partie artificielle car ils se recouvrent partiellement. Surtout dans la pratique, les indices sont perçus ensembles. La situation de laboratoire qui consiste à les isoler pour les analyser est en grande partie artificielle. Il a été également montré que chaque indice possède une plage de distances pour laquelle il est opérant.

En situation réelle, la question se pose de savoir quels indices prédominent. Pour cela, Surdick et al. (1994) testent sept indices visuels de profondeur : la luminosité relative, la taille relative, la hauteur relative, la perspective linéaire, le raccourcissement, le gradient de texture et la disparité stéréoscopique à des distances d'observation de un et deux mètres pour répondre à deux questions.

- Premièrement, quels sont les indices qui fournissent une information efficace sur la profondeur ? C'est-à-dire que seule une petite variation de l'indice de profondeur entraîne un changement notable de la profondeur perçue.
- Deuxièmement, comment l'efficacité de ces indices de profondeur change-t-elle en fonction de la distance d'observation ?

Dans une étude sur six sujets, ils testent l'effet des indices à deux distances d'observation grâce à une procédure à stimulus constants et un écran stéréoscopique.

Ils notent que la précision des réglages de correspondance perceptive pour tous les repères était très élevée (les erreurs constantes moyennes étaient proches de zéro), et aucun repère n'était significativement plus ou moins précis que les autres. L'efficacité des indices de perspective (perspective linéaire, raccourcissement et gradient de texture) était supérieure à celle des autres indices de profondeur, tandis que l'efficacité de la luminosité relative était nettement inférieure. Ces résultats ont des implications théoriques pour les modèles de perception spatiale humaine

et des implications pratiques pour la conception et le développement d'environnements virtuels en 3D.

D'autres problèmes n'ont pas été évoqués dans ce chapitre. En particulier, les facteurs d'ordre plus cognitifs ont été passés sous silence, alors qu'ils jouent un rôle important. La taille habituelle des objets joue un rôle important sur la perception de distance. La lecture de ce chapitre ne doit pas laisser penser que la perception visuelle de la distance est un processus purement « bottom-up », alors qu'il existe de nombreuses interactions avec des processus de plus haut niveau. Dans le même ordre d'idée, le rôle fondamental de l'action a été ignoré dans ce panorama des indices perceptifs.

Comme indiqué au départ, seuls les indices monoculaires statiques ont été retenus. Pourtant un autre facteur important de la perception est aussi à prendre en compte : la variation de tous ces indices. Le système visuel est extrêmement sensible aux variations et leur étude est nécessaire à une meilleure compréhension. L'étude d'un observateur placé face à un flux vidéo est un véritable enjeu non seulement pour l'étude théorique de la perception mais aussi pour la synthèse d'images animées.

En Résumé

La perception visuelle de la profondeur est reliée à des indices de perception de la distance. Les indices présentés ici sont dit monoculaires et ont fait l'objet d'une exploitation dans les arts graphiques pour suggérer des environnements en trois dimensions. Les méthodes de production d'images de synthèse exploitent également ces indices également désignés de picturaux. Les principaux indices sont la perspective linéaire, le gradient de texture, le recouvrement, le flou de distance, la hauteur relative à l'horizon ou élévation, les ombres et ombrages ainsi que la perspective aérienne.

Chapitre 3

Stéréo-vision et perception de la 3D

« Voir le relief, c'est recevoir au moyen de chaque œil l'impression simultanée de deux images dissemblables du même sujet. »

Euclide, III^e siècle avant J.-C.

En plus des indices monoculaires de distance, il existe au moins deux autres indices qui font intervenir la vision binoculaire. Le premier est la convergence oculaire qui permet d'amener le même objet en vision centrale pour les deux yeux. Le second est la disparité rétinienne, c'est à dire la variation entre les images reçues par chaque œil.

1 Convergence oculaire

1.1 Définition

La convergence oculaire — parfois dénommée aussi vergence oculaire — est définie comme le mouvement simultané des deux yeux en direction opposée pour atteindre ou maintenir la vision binoculaire. Pour cela, les deux yeux doivent tourner selon leur axe vertical de façon à ce qu'un même point de l'environnement se projette au centre des deux rétines. Contrairement aux autres mouvements oculaires qui sont conjugués, c'est-à-dire dans le même sens, la convergence oculaire entraîne une adduction de l'œil.

Le point de convergence le plus proche, ou NPC pour *near point of convergence*, peut être mesuré comme le point le plus proche qui n'entraîne pas de diplopie ou vision dédoublée. Les valeurs normales sont proches de 10 cm. Ce qui veut dire que pour un sujet normal, il n'est pas possible de faire converger les yeux pour une distance inférieure à 10 cm.

1.2 Rôle dans la perception de la distance

Dans la pratique, la convergence oculaire est très corrélée à l'accommodation, l'adaptation de la focale du cristallin. Ces deux mécanismes sont liés à la distance en vision de près et agissent toujours ensemble en situation naturelle. Pour cette raison, les deux mécanismes ont tendance à être associés en un « *réflexe d'accommodation-convergence* ». En réalité, il s'agit plutôt d'un

ensemble de réflexes en interaction les uns avec les autres et qui font intervenir de nombreux mécanismes d'une grande complexité (J. H. Jung, 2019).

La question de savoir si la convergence oculaire à elle seule est utile en tant qu'indice de perception de la distance est un problème scientifique ouvert. Il est d'ailleurs débattu depuis longtemps puisque Kepler (1604) posait déjà la question dans son traité d'optique.

Actuellement, la question reste l'objet de nombreux débats, d'une abondante littérature et de conclusions peu compatibles. Dans une série d'études, Rogers (2019) conclut par exemple à l'évidence d'un lien entre la vergence et la perception de la distance¹ alors que d'autres comme Linton (2020) remettent en cause cette liaison.

Le problème majeur, défi de la convergence, est qu'il est extrêmement complexe de séparer expérimentalement son rôle sans affecter les autres indices de perception de la distance, si bien qu'en fonction de la méthodologie choisie, les auteurs aboutissent à des conclusions opposées.

2 Disparité rétinienne

En plus de la convergence oculaire, la vision binoculaire permet surtout une perception des distances grâce à la disparité visuelle.

La disparité est liée au fait que les deux yeux ont une position légèrement différente. La séparation horizontale chez l'humain représente environ 6,5 cm. La perception tridimensionnelle est impactée par cette différence, en particulier pour les petites distances.

La sensation de profondeur provoquée par la disparité binoculaire est appelée *stéréopsie*.

2.1 Zone de stéréovision

Le champ de vision d'un seul œil est de 60° nasalement et de 100° temporalement. Verticalement le champ visuel s'étend de 60° vers le haut à 70° vers le bas. Sur le plan binoculaire, les champs de vision des deux yeux se chevauchent dans un secteur d'environ 120°. La région de stéréovision utile est, cependant, beaucoup plus petite en raison d'une mauvaise acuité loin de l'axe optique de l'œil. Par exemple, la reconnaissance des mots est limitée à quelques degrés de la ligne de vue. La reconnaissance des symboles est limitée à environ 30° et une discrimination fiable des couleurs reste possible entre 30 et 60°. (Schwiegerling, 2004)

2.1.1 Mesure de la stéréoacuité

Tout comme l'acuité visuelle, la stéréoacuité mesure la capacité à résoudre des détails en utilisant la vision binoculaire.

Le test stéréoscopique de Howard-Dolman est un dispositif physique qui a été conçu pour mesurer le seuil de stéréoacuité. Pour cela, deux cibles sont présentées — des petits bâtons — à une distance de six mètres de l'observateur. Les deux bâtonnets sont placés l'un devant l'autre de façon à ce que l'un occulte parfaitement l'autre. La distance entre les bâtonnets est variable. Du fait de la disparité, si la distance entre les deux bâtonnets augmente, les deux deviennent visibles. Le seuil de stéréovision est alors défini comme la différence minimale de l'angle de convergence entre les deux yeux qui provoque la perception de deux objets. Cette différence correspond également à la mesure de la disparité horizontale.

Ce seuil peut atteindre des valeurs de 5 à 10 secondes d'arc, ce qui représente environ 25 à 50 microradians. Ces valeurs sont nettement inférieures aux valeurs d'acuité visuelle de l'ordre

1. « *No one would deny that binocular disparities and eye vergence are sufficient to "specify perceived depth relations"* »

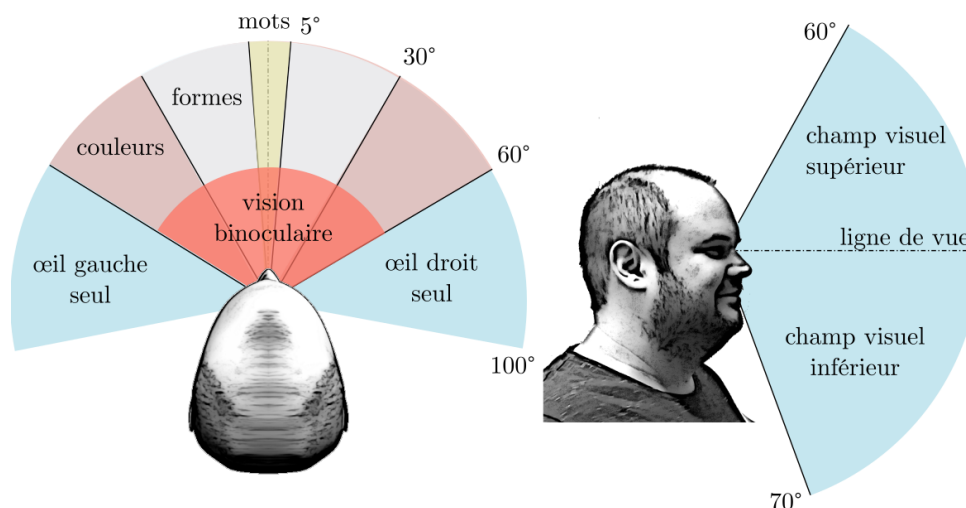


FIGURE 3.1 – Champ visuel horizontalement et verticalement.

Lorsque le champ est libre, la ligne de vue peut être considérée comme horizontale. Le champs de vision s'étend sur 130° verticalement et près de 200° horizontalement. Mais la zone de vision binoculaire utile est beaucoup plus limitée.

de 30 secondes d'arc. Ceci constitue une forme d'hyperacuité (voir section 1.4.3) dont la raison reste encore peu connue. Elle résulte d'un mécanisme de sommation interoculaire.

Notons néanmoins que les valeurs de ce seuil sont très variables. Elles sont influencées par des facteurs tels que la luminance et le contraste (Geib & Baumann, 1990) ainsi qu'une forte variabilité inter-personnelle.

2.2 L'horoptère théorique

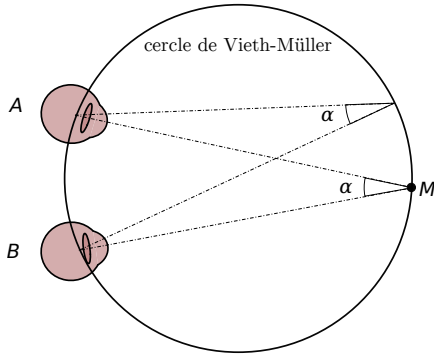
Pour une convergence oculaire déterminée, il existe un ensemble de points pour lesquels il y a une correspondance des projections sur les deux rétines. L'ensemble de ces points décrit une surface virtuelle appelée horoptère². Cette surface dont le modèle est appelé *horoptère théorique* est représentée dans un plan transversal par le cercle de Vieth-Müller qui passe par le point de convergence et les points nodaux des deux yeux comme représenté par la figure 3.2a (Howard, 2008).

Toujours d'un point de vue théorique, les points de correspondance binoculaire sont des portions de droites qui passent par le point de fixation verticalement à ce dernier. Ceci confère à l'horoptère théorique la forme d'une section de cylindre délimitée par le cercle de Vieth-Müller d'une part et tronquée suivant le champ de vision (Gibaldi & Banks, 2019) comme illustré en figure 3.2b

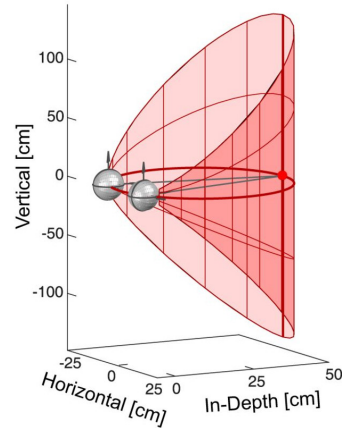
2.3 Horoptère empirique

Lorsque l'horoptère empirique est mesuré, sur la base d'expériences perceptives, un écart est enregistré avec l'horoptère théorique prédit par le cercle de Vieth-Müller. Cet écart porte le nom

2. Historiquement, le terme d'horoptère a plutôt désigné l'*horoptère oculomoteur* c'est à dire l'ensemble des points pour lesquels la convergence oculaire est de même valeur. La forme théorique de l'horoptère oculomoteur est torique. Actuellement, la notion d'horoptère désigne l'*horoptère binoculaire*.



(a) Selon la réciproque du théorème de l'angle inscrit, le lieu des points M tels que la valeur de l'angle \widehat{AMB} est égal à une constante α pour deux points A et B fixés est un arc de cercle appelé *arc capable*. Le cercle de Vieth-Müller, qui passe par les points nodaux des deux yeux et le point de focalisation, représente le modèle théorique de l'horoptère.



(b) Représentation de l'horoptère théorique en 3D.

FIGURE 3.2 – Horoptère théorique dans le plan horizontal et en 3D.

de *déviaton de Hering-Hillebrand*.

La comparaison de l'horoptère théorique et des résultats empiriques, laisse apparaître plusieurs phénomènes (Harrold & Grove, 2021 ; Schreiber et al., 2008).

1. L'horoptère empirique est caractérisé par une tolérance en profondeur, si bien que l'horoptère s'assimile davantage à une zone : la zone de fusion stéréoscopique aussi appelée aire de Panum³.
2. Horizontalement, la forme réelle de l'horoptère est plutôt à mi-chemin entre le cercle de Vieth-Müller et le plan frontal.
3. En élévation, l'horoptère n'apparaît pas vertical mais légèrement incliné vers le bas.

2.4 La disparité

Les objets dans l'espace pour lesquels il n'y a pas de correspondance rétinienne produisent des images disparates. Si la disparité dépasse un seuil, en dehors de la zone Panum, l'objet apparaît alors double : c'est le phénomène de diplopie.

La disparité horizontale pour un point A de l'espace peut-être mesurée comme la différence de l'angle de convergence pour l'objet focalisé F et la convergence pour le point A (voir figure 3.3) avec EIO l'écart interoculaire, P la distance du point de focalisation F et ΔP l'écart en profondeur entre F et A . Selon l'approximation des petits angles, $\tan(\alpha) \approx \alpha$.

Les angles de convergence α et β respectivement pour F et A sont estimés par :

$$\alpha = 2 \times \frac{\alpha}{2} \approx 2 \times \frac{\frac{EIO}{2}}{P} = \frac{EIO}{P}$$

$$\text{et } \beta \approx \frac{EIO}{P + \delta P}$$

3. Les aires de Panum, nommées en référence à Peter Ludvig Panum (19 décembre 1820 – 2 mai 1885) médecin et physiologiste danois qui en a fait la découverte (Roper-Hall & Jørgensen, 2008)

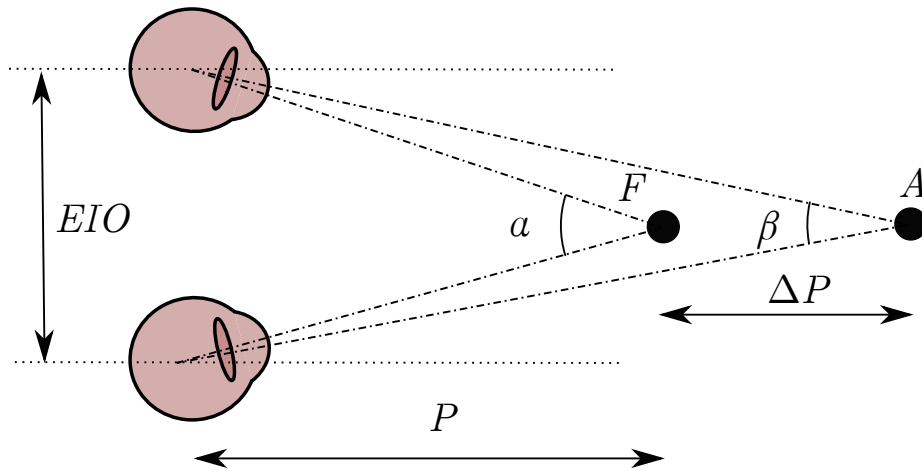


FIGURE 3.3 – Disparité horizontale

La disparité horizontale dépend de l'écart interoculaire (EIO), des angles α et β respectivement vers le point de fixation F et A un point quelconque. Ceux-ci dépendent de la distance en profondeur P au point F ainsi que de l'écart ΔP entre F et A .

Ce qui permet de connaître la valeur δ de la disparité horizontale.

$$\delta = \alpha - \beta \approx EIO \left(\frac{1}{P} - \frac{1}{P + \Delta P} \right) = EIO \left(\frac{\Delta P}{P^2 + P \times \Delta P} \right)$$

Lorsque $\Delta P \ll P$, il est possible d'écrire (Ritter, 1977)

$$\delta \approx EIO \left(\frac{\Delta P}{P^2} \right)$$

Deux cas peuvent se produire selon que A est plus loin ou plus proche que F , c'est-à-dire selon le signe de ΔP .

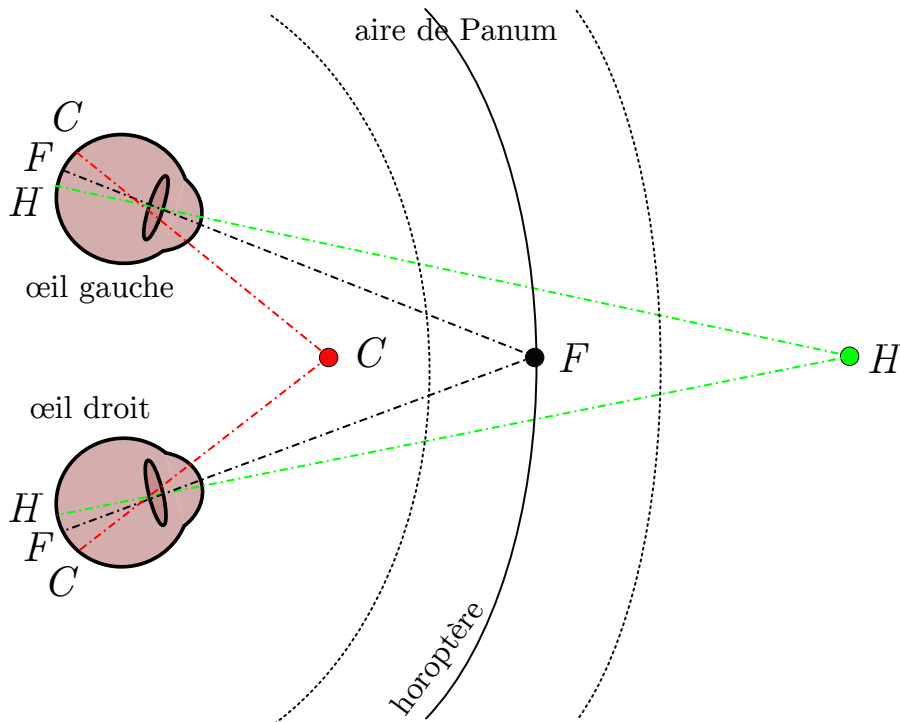
La disparité est positive lorsque A est plus loin que F .

La disparité est négative lorsque A est plus proche que F .

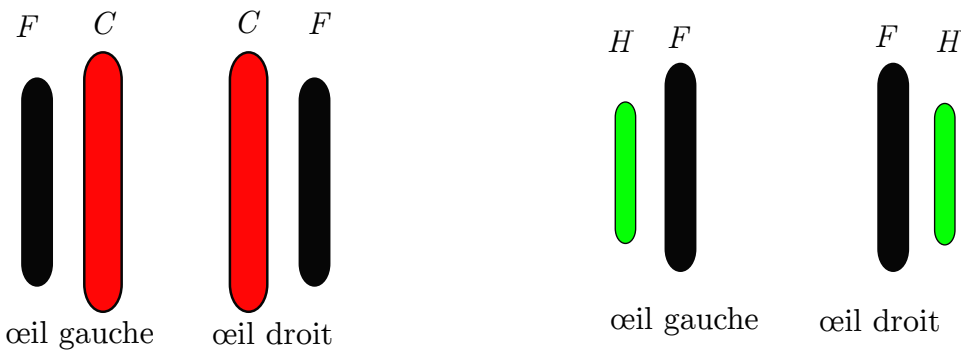
Lorsque la disparité est positive, ceci produit une *disparité homonyme* (Devisme, 2008). L'œil droit perçoit l'objet plus éloigné à droite et l'œil gauche le perçoit à gauche du point de fixation. Dans le cas où la disparité est négative, le phénomène inverse se produit : une *disparité croisée* comme le montre la figure 3.4.

2.4.1 Seuil de confort

Si une disparité faible ne provoque pas de perception de la profondeur, une disparité trop importante réduit le confort de visualisation de la scène. Wöpking (1995) enregistre un large éventail de conditions de vision binoculaire. Pour cela, il utilise des angles de disparité qui vont de 0 à 140 minutes. Sur la base d'images stéréoscopiques, il établit à 70 minutes d'angle le seuil de confort. Cependant, dans certaines conditions, en particulier en présence de flou en dehors de la zone de focalisation, des disparités plus importantes peuvent être tolérées.



(a) Disparité pour deux stimulus C et F situés respectivement devant et derrière l'aire de Panum d'un stimulus fixé F .



(b) Le stimulus C est perçu à gauche de F pour l'œil droit et à droite de F pour l'œil gauche : la disparité est croisée.

(c) Le stimulus H est perçu à droite de F pour l'œil droit et à gauche de F pour l'œil gauche : la disparité est homonyme.

FIGURE 3.4 – Disparité croisée et homonyme

F est le point de fixation, les points pour lesquels la correspondance existe entre les deux yeux s'appelle l'aire de Panum. Elle s'étend de part et d'autre de l'horoptère. En dehors de cette zone de tolérance, il existe une disparité. Celle-ci est croisée pour les stimulus devant la frontière de l'aire de Panum alors qu'elle est homonyme pour les stimulus situés derrière.

2.5 La vision stéréoscopique chez les animaux

La vision binoculaire est un phénomène fondamental de la vision humaine. Cette capacité apparaît dans les premiers mois de la vie. La stéréopsie a été démontrée chez de nombreuses espèces, y compris des mammifères, des oiseaux, des amphibiens et des invertébrés (Nityananda & Read, 2017). Cependant, il s'agit d'un mécanisme apparu tardivement dans l'évolution des espèces car il nécessite notamment que les yeux soient en position frontale. Ce phénomène permet au système nerveux central de détecter et de fusionner simultanément à partir de chaque œil.

Orssaud (2006) introduit une classification des capacités binoculaires : la perception simultanée et la fusion constituent le premier et second degré de vision binoculaire. Ce phénomène donne également au système nerveux central la possibilité de percevoir la distance grâce à l'écart inter-oculaire. La vision stéréoscopique est alors appelée le troisième degré de vision binoculaire.

Historiquement, les premières espèces à avoir démontré des capacités stéréoscopiques ont été des prédateurs. Le fait que cette capacité soit fréquemment rencontrée dans les espèces carnivores a été à l'origine de l'hypothèse que la vision stéréoscopique aurait un lien avec la prédation (Cartmill, 1974). Cependant, des données plus récentes ont contredit en partie cette hypothèse. Des formes de stéréopsie ont été découvertes chez des espèces non-prédatrices comme le cheval, la chèvre, le mouton ou le cochon d'Inde (Murphy et al., 2009)

3 La stéréoscopie

L'utilisation de la disparité a été exploitée dans de multiples techniques et périphériques regroupés sous le nom de stéréoscopie (Michel, 2012).

3.1 Les débuts

En 1838, Charles Wheatstone fait breveter la première forme de stéréoscope. Il s'agit d'un dispositif dans lequel deux images sont placées face à face (Wade & Ono, 1985). Deux miroirs placés à 45° permettent d'envoyer les deux images aux deux yeux comme représenté par la figure 3.5. À l'origine, les images étaient dessinées mais ont été ultérieurement remplacées par des photographies. Le dispositif reste relativement encombrant et peu utilisé. La stéréoscopie prend réellement une place plus importante lorsqu'en 1849 David Brewster eut l'idée d'utiliser des lentilles pour réduire l'encombrement du dispositif. De plus, il devient possible d'associer les deux vues sur un même support, ce qui facilite la distribution des images stéréoscopiques. Le dispositif fut présenté à la reine Victoria lors de l'exposition universelle de 1851. C'est ainsi qu'est né le stéréoscope « moderne », lointain ancêtre des casques de réalité virtuelle.

Lors de la deuxième partie du XIX^e siècle, la photographie stéréographique se démocratise peu à peu, jusqu'à être utilisée pour le portrait d'Abraham Lincoln de 1865 (figure 3.6). En 1893, un industriel français produit en série un appareil photo stéréographique sur plaque de verre, rechargeable en plein jour, le *Vérscope*.

Ensuite, au début du XX^e siècle, la stéréoscopie est retombée peu à peu dans l'oubli avec quelques exceptions comme le View-Master, dispositif qui permettait de visionner des séquences d'images à partir de disques cartonnés de sept paires d'images.

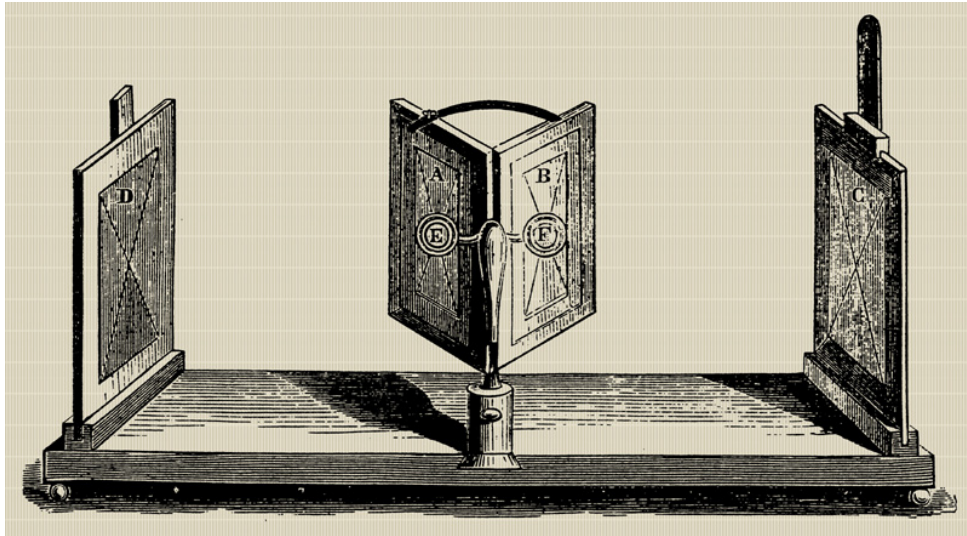


FIGURE 3.5 – Le stéréoscope de Wheatstone
Les images sont placées face à face et renvoyées aux deux yeux par un jeu de miroirs
(Domaine public).



FIGURE 3.6 – Portrait stéréoscopique de Abraham Lincoln
Abraham Lincoln, le 5 février 1865. Photographie par Alexander Gardner. Domaine public.

3.2 Les lunettes stéréoscopiques

3.2.1 Lunettes anaglyphes

La première technique est celle des *lunettes anaglyphes* qui utilisent un codage en couleur des images et un filtre coloré associé pour les visualiser. Plusieurs paires de couleurs complémentaires peuvent être utilisées mais le codage le plus fréquent est le rouge pour la vue gauche et le cyan pour la vue droite.

Le principal avantage de cette technique est la simplicité des lunettes, qui peuvent être produites à un coût modique. Son inconvénient majeur est de rendre quasiment impossible la perception des couleurs et aussi de s'accompagner d'artefacts visuels désagréables, tels des images fantômes suivant la qualité de l'ajustement entre l'encodage coloré et le filtrage.

3.2.2 Lunettes polarisantes

Le principe des lunettes polarisantes est le même que celui des lunettes anaglyphes à la différence que c'est une lumière polarisée qui est employée. L'utilisation de deux filtres polarisants orientés à 90° l'un de l'autre permet de n'envoyer qu'un seul signal à chaque œil.

L'avantage de cette technique est de permettre une « vraie » vision stéréoscopique avec restitution des couleurs. Il est également possible de produire des lunettes peu onéreuses, les filtres polarisants étant bon marché. Cependant, la technique est essentiellement utilisable par projection de deux images sur un écran avec deux projecteurs : un filtre polarisant est interposé entre chaque projecteur et l'écran. De plus, il faut soigneusement contrôler la concordance. Cela limite essentiellement cette méthode à des projections dans des salles dédiées même si son utilisation est possible sur d'autres supports : Polaroid a utilisé cette technique pour des tirages de photographies en relief sous le nom de *Vectograph*.

3.2.3 Lunettes à occultations alternées

Une autre solution consiste à utiliser des lunettes qui filtrent les images temporellement. La source, que ce soit un écran ou un vidéo-projecteur, envoie alternativement les deux images. Les lunettes sont synchronisées avec la source de façon à ce que chaque œil ne perçoive que l'une des deux images. La fréquence d'occultation doit être le double de la fréquence de rafraîchissement d'un écran « classique » et la synchronisation doit être effectuée de façon filaire ou au moyen d'un émetteur-récepteur infra-rouge.

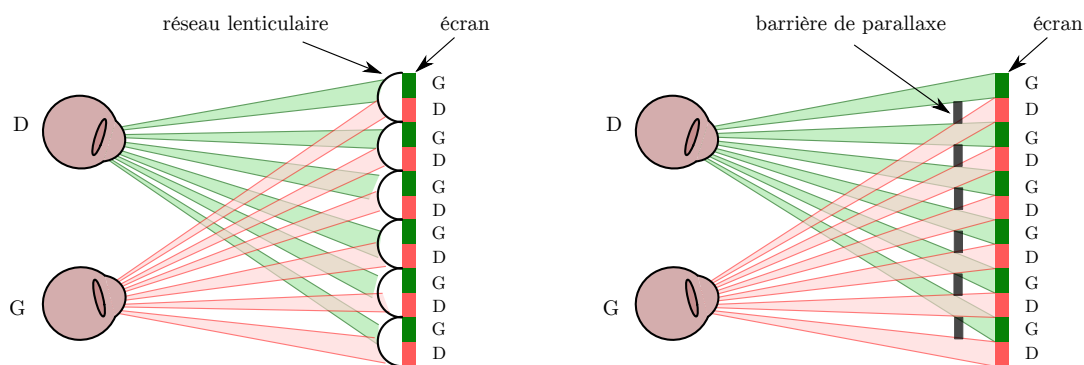
L'avantage est de permettre une vision stéréoscopique colorée sur tous les supports de type écran. En revanche le coût du dispositif est plus important et peut provoquer des « images fantômes » liées à l'imprécision de la synchronisation. De plus, le fait que l'un des yeux soit privé de la source de lumière au moins la moitié du temps aboutit à diminuer la luminosité perçue. Il est alors nécessaire d'augmenter la puissance des périphériques.

3.3 Les autres périphériques

3.3.1 Écrans auto-stéréoscopiques

Les écrans auto-stéréoscopiques sont des périphériques qui permettent la visualisation stéréoscopique « sans lunettes » sur certains périphériques d'affichage.

Le principe de l'auto-stéréoscopie reste assez simple, puisqu'il consiste à produire des images légèrement différentes suivant le point de vue. Ce principe a été utilisé en photographie depuis le début du XX^e siècle.



(a) Principe du réseau lenticulaire.

(b) Principe de la barrière de parallaxe.

FIGURE 3.7 – Comparaison de deux techniques auto-stéréoscopiques.

Dans les deux cas, l'image est découpée en bandes verticales. Chaque bande n'est visible que d'un seul œil en utilisant soit la déviation optique avec un réseau lenticulaire, soit le masquage d'une partie de l'image avec une barrière de parallaxe. Pour simplifier, ce schéma illustre deux directions de visualisation mais la même méthode permet de produire de multiples points de vue.

L'un des avantages de ce type d'écran, outre de ne pas nécessiter d'équipement supplémentaire, est de produire à la fois une disparité et une convergence visuelle associée. Les principales techniques sont les réseaux lenticulaires et les barrières de parallaxe comme illustré par la figure 3.7.

3.3.1.1 Auto-stéréoscopie à base de réseaux lenticulaires.

Pour produire des points de vue différents, un réseau lenticulaire constitué de microlentilles est placé à la surface d'une image constituée d'images imbriquées : l'image alterne des bandes de l'une et l'autre des images face aux microlentilles.

Cette technique a été utilisée par exemple dans des supports publicitaires ou des cartes postales. La société Alioscopy[®] a appliqué le même principe pour commercialiser des écrans 3D avec huit points de vue.

3.3.1.2 Auto-stéréoscopie à base de barrière de parallaxe.

Dans le cas de la barrière de parallaxe, qui est une technique plus ancienne, la répartition des parties de l'image est réalisée par un filtrage directionnel qui masque des parties de l'image. Tout comme les réseaux lenticulaires, il est possible de l'adapter à plus de deux points de vue. Plusieurs écrans de ce type ont été commercialisés mais l'application la plus connue reste l'écran auto-stéréoscopique de la Nintendo[®] 3DS.

3.3.2 Casques de réalité virtuelle et augmentée

Les casques de réalité virtuelle et de réalité augmentée sont des casques immersifs ou HMD pour *Head Mounted Display*. Ces périphériques associent généralement :

- des afficheurs à proximité de l'œil ou NEDs pour *near-eye displays* ;

- des capteurs pour mesurer l'orientation et la position de la tête ;
- un système audio qui permet d'augmenter l'immersion et de communiquer ;
- des manettes pour agir sur l'environnement avec éventuellement retour de force ;
- une ou plusieurs caméras qui enregistrent l'environnement pour la réalité augmentée.

L'ensemble doit d'une part couvrir la plus grande part du champ visuel et d'autre part avoir un encombrement et un poids réduits. Dans les systèmes optiques de réalité augmentée et de réalité virtuelle, les images virtuelles sont projetées sur les yeux par le biais de lentilles dans un plan apparent éloigné des yeux, généralement autour de 3 m.

Comme l'image virtuelle projetée se trouve dans un plan fixe, les utilisateurs portant un NED souffrent généralement d'un *conflit de vergence-accommodation*. En situation naturelle, lorsqu'un objet réel est présent, la convergence oculaire correspond à leur accommodation : l'adaptation de la longueur focale du cristallin. Mais en présence d'un affichage stéréoscopique, la convergence des yeux correspond à différentes profondeurs liées aux images affichées, tandis que la profondeur d'accommodation est fixe. Il a été noté que le conflit vergence-accommodation des périphériques 3D provoque de l'inconfort et ce d'autant plus que les variations sont rapides. (Kim et al., 2014 ; Reinhardt-Rutland, 1996 ; Vienne et al., 2015).

Ceci est confirmé par une étude récente : Yoon et al. (2021) utilisent des smartphones montés en tant que périphériques de réalité virtuelle. Ils mesurent les performances perceptives de participants qui ont utilisé ce dispositif pendant deux heures, comparé à un groupe contrôle qui utilise un smartphone pendant la même durée. Dans les comparaisons avant-après, les auteurs enregistrent des différences significatives notamment dans la convergence proche et l'accommodation. De telles variations ne sont pas enregistrées lorsque le smartphone est utilisé seul.

Il existe plusieurs techniques pour réduire le conflit vergence-accommodation. Elles reposent sur une modification de la longueur focale en fonction de la distance d'affichage (Y.-J. Wang & Lin, 2021).

3.4 Caméras stéréoscopiques en images de synthèse

La génération d'images pour les périphériques de rendu 3D repose sur la définition de caméras stéréoscopiques afin de produire deux images différentes de la scène. La génération de ces deux images utilise deux paramètres stéréo : la séparation et la profondeur.

La séparation est définie comme la distance entre les axes des caméras virtuelles

La profondeur correspond à la distance du plan de focalisation

Plusieurs modèles de caméras stéréoscopiques virtuelles seraient possibles (Trapp et al., 2011). Ces types de caméra sont illustrés par la figure 3.8 :

les caméras parallèles n'ont qu'une partie du plan de focalisation en commun ;

les caméra *toe in* ou convergentes pour lesquelles les plans images ne sont pas confondus ;

la caméra *off-axis* présente un recouvrement des images dans la zone de focalisation et des plans images confondus. Ce type de caméra est proche des caméras parallèles mais utilise une caméra non symétrique (Trapp, 2013).

Il a été noté par Jones et al., 2001 que les caméras parallèles et convergentes sont source d'inconfort visuel. Pour cette raison, le modèle de caméra *off-axis* est préférable.

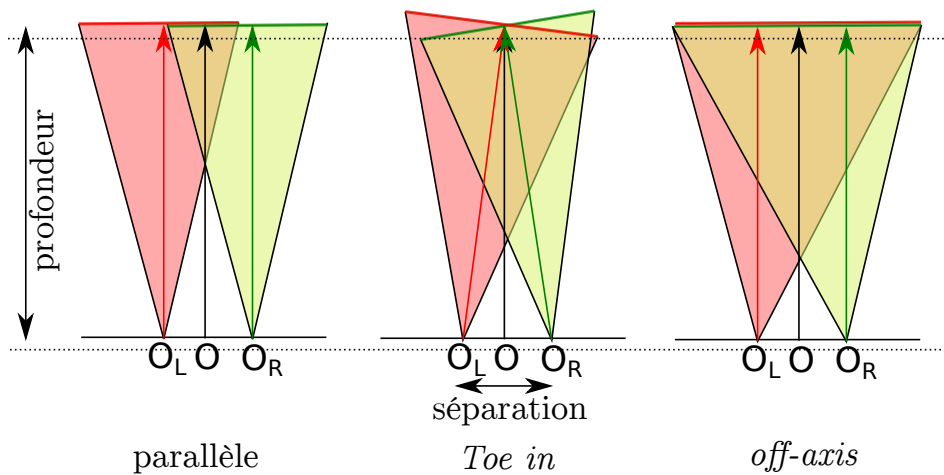


FIGURE 3.8 – Trois sortes de caméras stéréoscopiques

Les caméras parallèles pour lesquelles les axes sont parallèles, les caméras *toe in* qui convergent vers le point de focalisation et les caméras *off-axis* qui sont des tronçons non symétriques de la pyramide de vue.

4 Sur l'effet de la 3D

McIntire et al. (2014) réalisent une méta-analyse effectuée sur plus de 180 expériences couvrant plus de 50 années de recherche dans le domaine de la stéréovision. Ces études couvrent les disciplines de la psychologie et l'ingénierie des facteurs humains, l'interaction homme-machine, la science de la vision, la visualisation et la médecine. Les publications qui sont retenues décrivent au moins une tâche avec évaluation expérimentale d'une comparaison entre un périphérique de stéréo 3D et un affichage 2D dans des conditions comparables.

Les études sont classées en cinq catégories en fonction de la nature de la tâche sur laquelle les performances ont été enregistrées :

- des jugements sur la position ou la distance à proprement parler ;
- des recherches, des classifications ou des identifications d'objets ;
- des manipulations dans l'espace d'objets réels ou virtuels ;
- des tâches de navigation dans un environnement 3D ;
- des tâches de compréhension, de mémorisation et de rappel spatial ;
- des tâches de planification, d'apprentissage ou d'entraînement.

Les auteurs constatent que l'affichage stéréo 3D amène une amélioration des performances dans 75% des expériences considérées : un avantage important dans 60% des cas, marginal dans 15%. Dans 25% des cas, la stéréoscopie n'apporte aucun avantage et parfois aboutit à nuire aux performances. Les performances semblent réellement meilleures dans les cas de manipulation, de localisation et de jugement sur la position, en particulier quand ces tâches font intervenir des configurations complexes. En revanche, c'est dans les cas de mémorisation et d'apprentissage que l'intérêt semble le plus faible.

En Résumé

Les indices picturaux du chapitre précédent sont complétés par des indices binoculaires : la convergence oculaire et la disparité. La disparité est utilisée dans les périphériques stéréoscopiques tels les écrans auto-stéréoscopiques et les casques immersifs de réalité virtuelle et augmentée.

Deuxième partie

Image de synthèse photoréaliste et perception du bruit

Introduction à la deuxième partie

L'objectif de la première partie était de situer l'image numérique dans le contexte de la perception visuelle. Il paraît intéressant de montrer que ce problème peut être envisagé sous plusieurs aspects : psychologie de la perception, physiologie du système visuel, optique géométrique et production d'image. Ces approches sont en outre complémentaires. La production d'une image — numérique ou non — s'adresse à un observateur. C'est pourquoi il apparaît nécessaire de connaître ses limitations perceptives ainsi que les mécanismes impliqués dans la perception d'images, que ces images soient présentées sur un écran conventionnel, dans un contexte stéréoscopique ou celui d'un casque immersif. Une grande partie des références bibliographiques traite de ces sujets dans le contexte d'images « analogiques ». La plupart des résultats restent valides dans le contexte de l'image numérique mais il est parfois nécessaire de prendre en compte ce paramètre. Les images numériques offrent des possibilités quasi illimitées mais ajoutent elles aussi de nouvelles contraintes.

Cette seconde partie est essentiellement destinée à présenter les contributions réalisées dans ce contexte. Cependant la problématique générale de la perception des images numériques est très vaste. Aussi le problème qui a été exploré concerne essentiellement la perception du bruit dans les images photoréalistes. Afin de pouvoir exposer les recherches qui ont été menées, il est nécessaire de préciser comment de telles images peuvent être produites.

C'est pourquoi le chapitre 4 s'attache à décrire les modèles et méthodes utilisés dans le cadre de la synthèse d'images photoréaliste et en particulier le modèle du transport de la lumière dans une scène caractérisée par des sources lumineuses et par des surfaces géométriques. Dans cette scène la propagation de la lumière est simulée en prenant en compte les propriétés d'émission des sources et l'interaction entre la lumière et les surfaces jusqu'à un observateur. Afin de comprendre pourquoi ces techniques génèrent des artefacts visuels, tel le bruit haute fréquence, il est nécessaire de préciser les méthodes qui permettent le calcul des images.

Après cet aperçu des techniques de production des images réalistes, c'est finalement le chapitre 5 qui regroupera les travaux qui ont été réalisés.

Chapitre 4

Synthèse d'images photoréalistes

« *J'appelle rayon de lumière, ou simplement rayon, la ligne droite tirée de l'objet lumineux, ou éclairé, par laquelle se transmet l'action qui le rend visible.* »

Nicolas Malebranche, De la recherche de la vérité, 1674

1 Introduction

Le rendu photo-réaliste est le domaine dans lequel on cherche à produire des images qui soient indiscernables de photographies de la même scène. La perception du réalisme est liée de nombreux facteurs. Certains facteurs sont purement géométriques et dépendent de la représentation géométrique des objets et de leur relations spatiales. Cet aspect ne sera pas évoqué dans ce chapitre qui couvre essentiellement la simulation de l'éclairage depuis la source jusqu'au système visuel de l'observateur ou à une caméra. Ceci revient à simuler le plus exactement possible le comportement de la lumière : sa propagation dans le milieu et ses interactions vis à vis des matériaux de la scène. Cette simulation nécessite donc de décrire les sources lumineuses et les matériaux présents dans la scène, ce qui sera présenté respectivement dans les sections deux et trois de ce chapitre. Un modèle, si possible physiquement correct, permet alors de simuler la propagation de la lumière dans la scène. Ce modèle qui sera présenté dans la section quatre repose sur une équation intégrale, appelée *équation du rendu*.

2 Modèles des sources de lumière

2.1 Notion de luminance

La création d'images réalistes a pour objectif de donner à notre système visuel l'illusion que ce qu'il voit pourrait être réel. Or notre perception du monde est liée à l'énergie lumineuse que reçoivent les cellules photo-réceptrices de nos yeux. La principale grandeur énergétique à laquelle

ces cellules sensibles est appelée *luminance énergétique*. Elle correspond à l'énergie, exprimée en Watts, émise par une surface mesurée en mètres carrés dans une portion d'angle solide en stéradians. Elle correspond donc à des Watts par mètre carré et par stéradian ($W.m^{-2}.sr^{-1}$).

2.2 Représentation des sources de lumière

La simulation de l'éclairage physiquement réaliste d'une scène nécessite la connaissance des sources de lumière et des matériaux constituant la scène. Cette partie s'intéresse aux sources de lumière. En général, un luminaire est composé d'un ou plusieurs émetteurs (filament, néon, LED, *etc.*), de réflecteurs permettant de diriger l'émission de lumière et d'objets perturbant l'émission tels que les fixations d'un filament ou une grille de protection. Il est courant de considérer qu'une source de lumière peut être décrite par trois caractéristiques : sa géométrie, sa distribution d'intensité et son émission spectrale.

2.2.1 Géométrie d'une source lumineuse

Les sources de lumière en informatique graphique peuvent être ponctuelles, linéaires ou surfaciques¹. Le modèle le plus simple et très utilisé est la source ponctuelle. C'est un modèle théorique puisque les sources ponctuelles n'existent pas dans la réalité. Ce type de source produit des ombres aux contours très nets, ce qui nuit au réalisme des images. Pour obtenir des ombres douces, il est nécessaire d'utiliser des sources émettant depuis une ligne ou une surface. Lorsque la source est proche d'un objet, sa géométrie est importante du fait de la variation d'éclairage que cette proximité est susceptible de générer. En revanche, sa géométrie peut être négligée quand l'émetteur est suffisamment éloigné de l'objet. Dans ce cas, il peut être acceptable d'utiliser une source ponctuelle. Ceci conduit aux notions de source en champ proche ou *near-field* et de source en champ lointain ou *far-field*. Cette différenciation n'obéit pas à une loi précise mais la règle d'Asdown peut être adoptée (Asdown, 1993) : une source est en champ lointain si la distance à la source est supérieure à cinq fois la plus grande dimension du luminaire.

2.2.2 Distribution d'intensité

Les sources lumineuses réelles n'émettent pas la lumière uniformément dans toutes les directions : par exemple, les luminaires utilisés en architecture ou les phares de voiture sont caractérisés par une forme de distribution spécifique de l'intensité. Il est important de tenir compte de cette distribution, lorsqu'elle est disponible, afin d'améliorer la qualité du calcul d'éclairage. La plupart des constructeurs de luminaires fournissent d'ailleurs cette information sous forme de *solides photométriques* qui représentent des échantillonnages de l'intensité lumineuse émise par le dispositif d'éclairage dans toutes les directions. Parfois cette information est donnée comme une information tridimensionnelle mais plus souvent sous la forme de courbes en deux dimensions².

2.2.3 Émission spectrale

Une source peut être caractérisée par son spectre d'émission, c'est-à-dire l'intensité de son émission à différentes longueurs d'onde. Quelques exemples sont montrés en figure 4.1.

1. Des sources comme les LEDs ou les lampes au xénon ne possèdent pas en fait de géométrie bien définie.

2. voir <https://leclairage.fr/courbe-photometrique/>

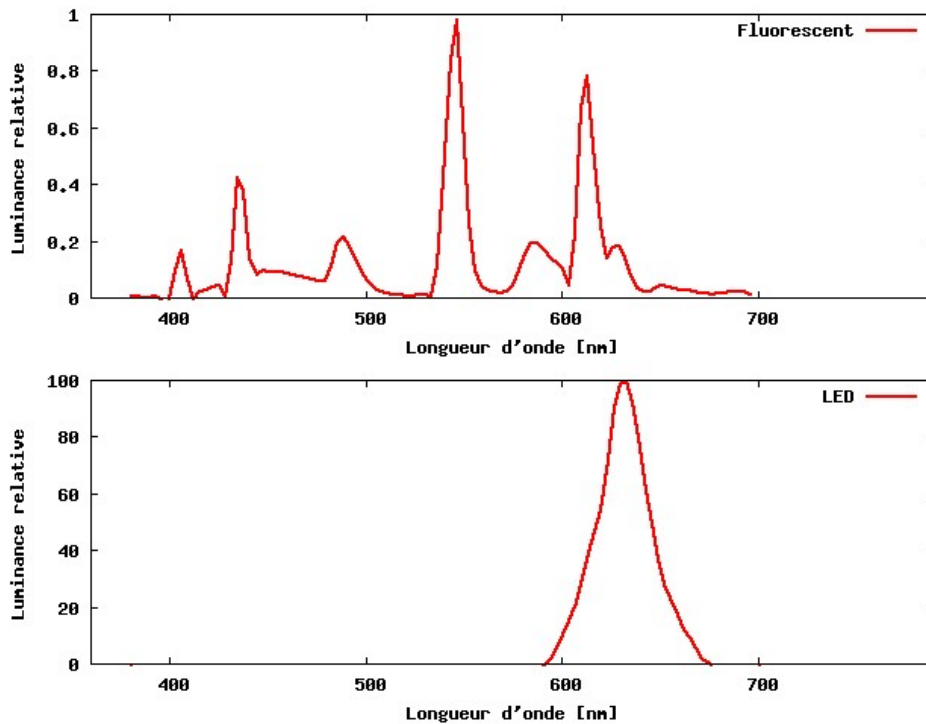


FIGURE 4.1 – Spectres d'émission

Pour une source fluorescente — en haut — et pour une LED — en bas.

2.2.4 Notion de solide photométrique

Les trois propriétés précédentes permettant de caractériser une source de lumière peuvent se formaliser de la manière suivante : une source est représentée par une surface munie en tout point d'une luminance $L_\lambda(x, \vec{\omega})$ (voir section 2.1), fonction de :

- x : position paramétrique (u, v) du point sur la surface émettrice ;
- $\vec{\omega}$: direction d'émission en coordonnées sphériques (θ, ϕ) depuis le point x ;
- λ : longueur d'onde considérée.

Cette représentation de la luminance est généralement appelée *solide photométrique étendu* (voir figure 4.2). Ce solide dépend de cinq paramètres et il est de ce fait difficile à mesurer, à stocker et à reconstruire. Quand on réduit la dimension spatiale à un point, on obtient un modèle de *solide photométrique ponctuel* (voir également la figure 4.2), modèle qui n'a plus que deux paramètres si on considère la luminance monochromatique.

3 Modèles des matériaux

3.1 Modèles de réflexion et de réfraction

La représentation des propriétés optiques des matériaux est elle aussi de très grande importance dès lors que l'on souhaite effectuer une simulation réaliste d'éclairage. En effet, nous percevons visuellement l'aspect final d'un objet uniquement par l'intermédiaire des photons qui atteignent nos yeux après avoir interagi avec le matériau de l'objet : réflexion sur la surface de

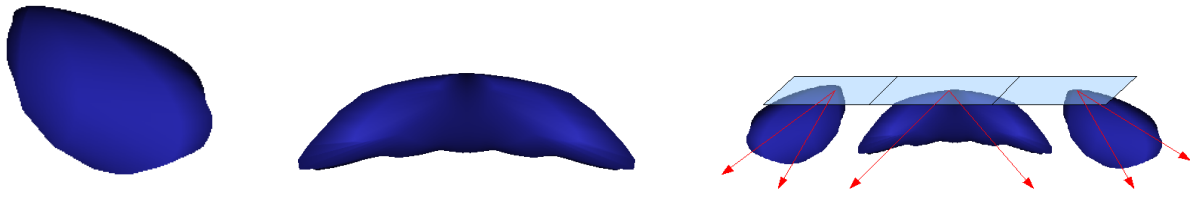


FIGURE 4.2 – Solides photométriques

Les deux figures de gauche représentent chacune un solide photométrique ponctuel ; la figure de droite représente un solide photométrique étendu, chaque point de la source surfacique possédant sa propre distribution de luminance.

l'objet, transmission au travers de ce matériau, réémission dans le cas de la phosphorescence, etc. Cette interaction lumière-matière est un problème extrêmement complexe, qui peut être abordé à différentes échelles : au niveau microscopique (mécanique quantique), au niveau mésoscopique (électromagnétisme et optique géométrique) ou au niveau macroscopique (fonctions de dimensions variables). Dans ce paragraphe, seul le niveau macroscopique est envisagé. C'est le niveau le plus utilisé en informatique graphique au travers de la notion de *fonction de distribution de la réflectance bidirectionnelle* (BRDF). Introduite par Nicodemus et al. (1977), la BRDF est le rapport entre la luminance réfléchi dans la direction de réflexion $\vec{\omega}_r$ en un point x d'une surface infinitésimale dA et l'éclairement incident à celle-ci dans la direction d'incidence $\vec{\omega}_i$ (voir figure 4.3)³.

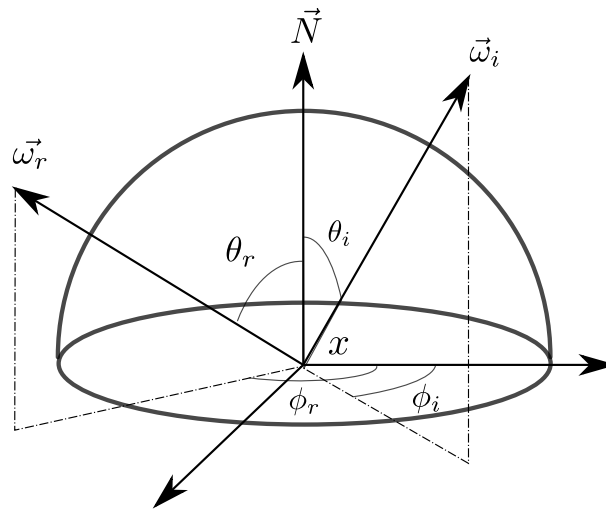


FIGURE 4.3 – Rayon incident, rayon réfléchi

Le rayon incident $\vec{\omega}_i$ est caractérisé par sa norme et par les angles θ_i (*colatitude*) et ϕ_i (*longitude*). De même le rayon réfléchi est défini par la norme de $\vec{\omega}_r$ et par les angles θ_r et ϕ_r .

Elle s'exprime sous la forme :

$$f_r(x, \vec{\omega}_i, \vec{\omega}_r, \lambda) = f_r(x, \theta_i, \phi_i, \theta_r, \phi_r, \lambda) = \frac{dL_r(x, \theta_r, \phi_r, \lambda)}{L_i(x, \theta_i, \phi_i, d\omega_i, \lambda)}$$

3. Il faut noter que la BRDF est une quantité sans dimension.

La BRDF caractérise la réflexion des surfaces⁴. Cependant, on peut définir de manière similaire une notion de BTDF permettant de caractériser les matériaux transparents.

S'il existe des modèles basés sur la compréhension des processus physiques mis en jeu dans la réflexion lumineuse, par exemple, le modèle de Kubelka et Munk (1931), la plupart des modèles utilisés en informatique graphique sont phénoménologiques. Nous allons maintenant présenter les plus classiques de ces modèles.

3.1.1 Le modèle de Lambert

Ce modèle suppose la surface parfaitement diffuse, c'est-à-dire réfléchissant la lumière de façon équiprobable dans toutes les directions :

$$f_r(x, \vec{\omega}_i, \vec{\omega}_r, \lambda) = \frac{\rho_\lambda}{\pi}$$

où ρ_λ est la réflectance (ou albédo) de la surface, qui représente le rapport entre l'énergie totale incidente à la surface et l'énergie totale réfléchi. Ce modèle est illustré en figure 4.4.

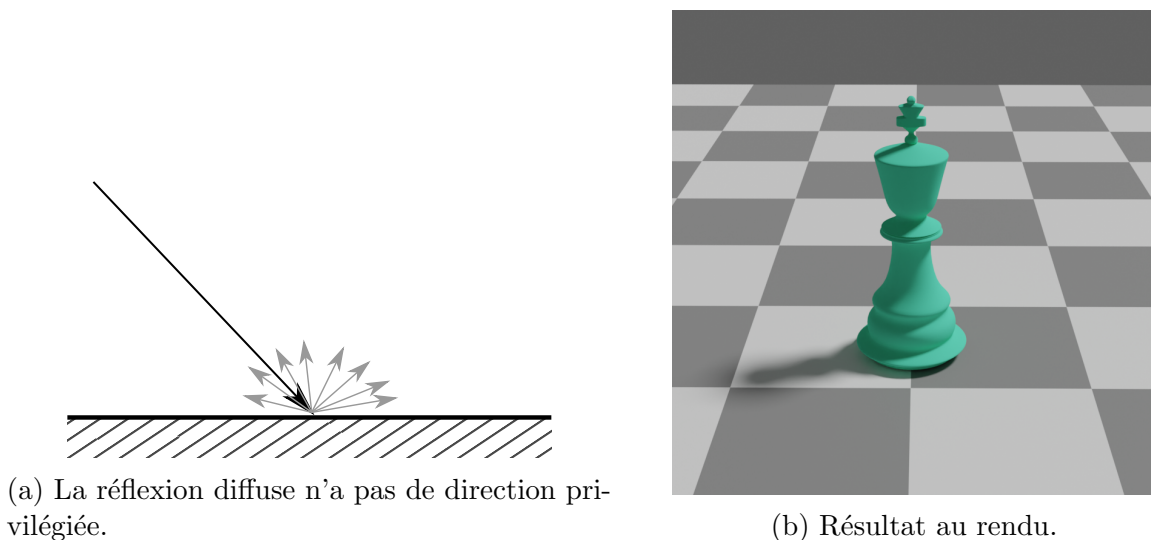


FIGURE 4.4 – Matériau diffus : modèle de Lambert

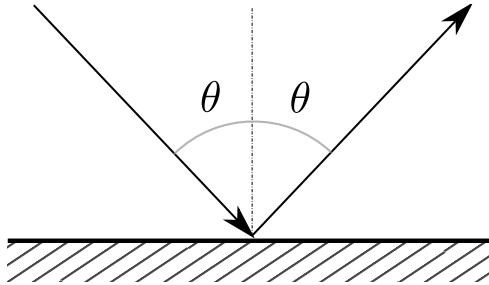
3.1.2 Le modèle spéculaire pur (ou du miroir)

Ce modèle, illustré en figure 4.5, suppose la surface parfaitement spéculaire, c'est-à-dire réfléchissant la lumière dans la direction symétrique à la direction d'incidence par rapport à la normale à la surface :

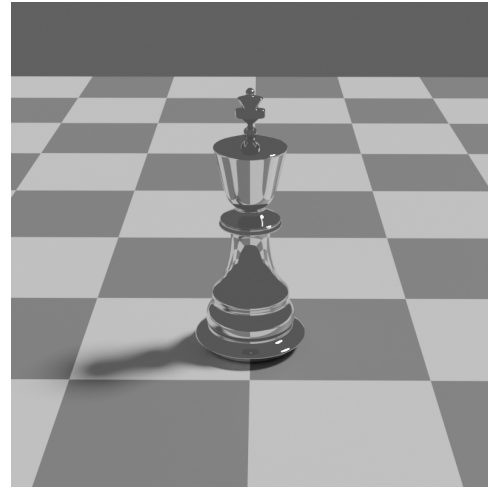
$$f_r(x, \theta_i, \phi_i, \theta_r, \phi_r) = \frac{\delta[\cos\theta_i - \cos\theta_r]}{\cos\theta_i} \delta(\phi_i - (\phi_r \pm \pi))$$

4. En première approximation, la BRDF suppose que l'énergie lumineuse incidente en un point x d'une surface est réfléchi depuis ce même point. Ceci n'est pas toujours vrai (matériaux translucides, par exemple) et il faut dans ce cas introduire la notion de *fonction de distribution de la réflectance sous-surfacique bidirectionnelle*, BSSRDF (Nicodemus et al., 1977).

δ étant la distribution de Dirac⁵



(a) La réflexion spéculaire pure est celle d'un miroir.



(b) Résultat au rendu.

FIGURE 4.5 – Matériau spéculaire pur

3.1.3 Le modèle de Phong (1975)

Ce modèle repose sur le principe de la combinaison linéaire d'un comportement diffus et d'un comportement spéculaire comme illustré par la figure 4.6. Si \vec{L} est la direction d'observation, \vec{V} la direction de réflexion et \vec{N} la normale à la surface, alors la BRDF est donnée par :

$$f_r(\vec{L}, \vec{V}) = k_d(\vec{N} \cdot \vec{L}) + k_s F_s(\vec{L}, \vec{V})$$

Les paramètres k_d et k_s représentent respectivement la couleur diffuse et la couleur spéculaire de l'objet, c'est à dire la part de la lumière qui est réfléchiée de manière purement diffuse et celle qui est réfléchiée de manière spéculaire. On doit normalement avoir $k_d + k_s \leq 1$. Le terme $F_s(\vec{L}, \vec{V})$ est appelé *lobe spéculaire* et il traduit le fait que la réflexion spéculaire peut se faire de manière non parfaite, en autorisant une dispersion plus ou moins importante autour de la direction de réflexion spéculaire pure. Ce terme peut être exprimé de plusieurs façons.

Dans le modèle de Phong, on prend

$$F_s(\vec{L}, \vec{V}) = (\vec{R} \cdot \vec{V})^n$$

où \vec{R} représente la direction de réflexion spéculaire et n appelé brillance, caractérise la rugosité de la surface : plus n est grand, moins la dispersion autour de la direction de réflexion spéculaire est forte et plus la surface apparaît lisse.⁶

C'est un modèle très utilisé en informatique graphique, avec quelques adaptations, c'est le modèle proposé par les API OpenGL et DirectX⁷

5. Les distributions généralisent la notion de fonction. La distribution de Dirac en est un cas particulier, apparenté à une fonction qui prendrait une valeur infinie en zéro et serait nulle partout ailleurs, et dont l'intégrale sur \mathbb{R} vaudrait un.

6. Avec cette formulation, le modèle de Phong est réciproque, mais il ne respecte pas le principe de conservation de l'énergie. Afin de rendre le modèle physiquement plausible, Lewis (2003) a proposé un ensemble de contraintes sur les paramètres qui doivent être respectées.

7. Dans la pratique, c'est la formulation proposée par Blinn (1977) qui est alors utilisée : $F_s(\vec{L}, \vec{V}) = (\vec{N} \cdot \vec{H})^n$ où \vec{H} est le vecteur bissecteur des vecteurs \vec{L} et \vec{V} .

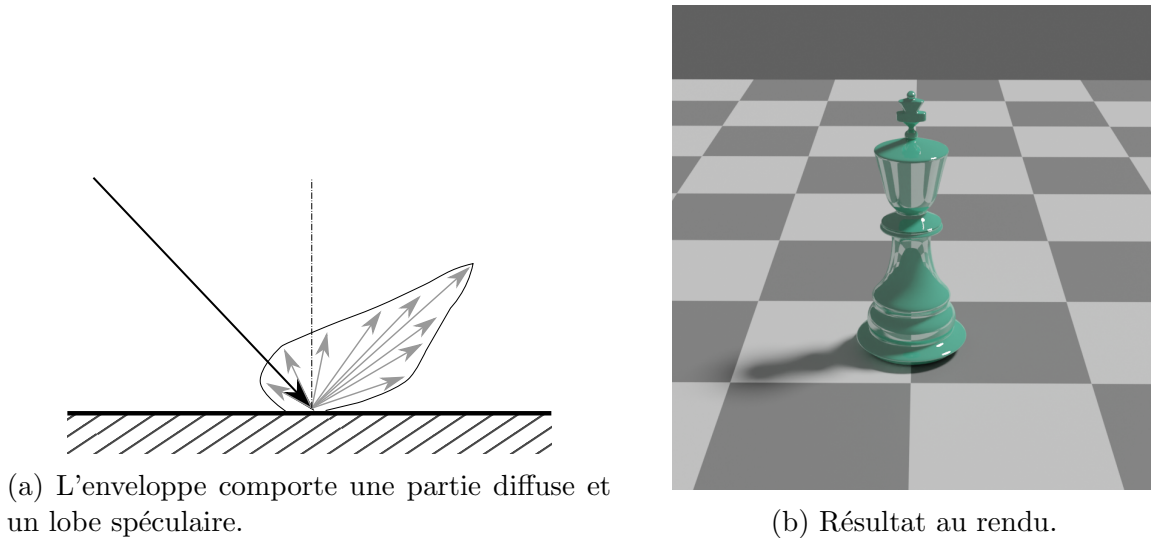


FIGURE 4.6 – Le modèle de Phong

Il repose sur l'addition d'une réflexion diffuse et d'un lobe autour de la direction de réflexion spéculaire.

3.1.4 Le modèle de Cook-Torrance

Le modèle de Cook-Torrance (figure 4.7 a été le premier ayant des bases physiques qui a été utilisé en informatique graphique (Blinn, 1977 ; Cook & Torrance, 1982 ; Torrance & Sparrow, 1967). Ce modèle suppose la surface constituée de micro-facettes planes, lisses et parfaitement spéculaires et la source de lumière à l'infini. Si da est l'aire d'une facette et dA l'aire de la surface étudiée, le modèle est valide si $\lambda^2 \ll da \ll dA$. La BRDF est alors donnée par :

$$f_r(x, \vec{\omega}_i, \vec{\omega}_r, \lambda) = d \frac{\rho(\lambda)}{\pi} + s \frac{F(\beta, \alpha) D(\alpha) G(\theta_i, \theta_r, \alpha, \beta)}{4}$$

où :

- D est la fonction de distribution de l'orientation des facettes, qui modélise la rugosité de la surface :

$$D(\alpha) = \frac{1}{m^2 \cos^4 \alpha} e^{-\left(\frac{\tan \alpha}{m}\right)^2}$$

où le paramètre m correspond à la rugosité de la surface (plus la valeur de m est faible, plus la réflexion sera directionnelle, plus elle est forte et plus la réflexion sera diffuse).

- G est un facteur d'atténuation géométrique ($0 \leq G \leq 1$) qui module la BRDF pour prendre en compte le phénomène d'auto-ombrage entre facettes.
- F est le terme de Fresnel qui correspond au phénomène de réflexion-réfraction des ondes électromagnétiques à l'interface entre deux milieux dont les indices de réfraction sont différents.
- α représente l'angle entre la normale \vec{N} et \vec{H} le vecteur bissecteur des vecteurs \vec{L} et \vec{V} et β l'angle formé par le vecteur incident \vec{L} et le vecteur normal \vec{N} .
- les facteurs d et s qui, comme dans le modèle de Phong, permettent de pondérer respectivement la réflexion diffuse lambertienne et le lobe spéculaire.

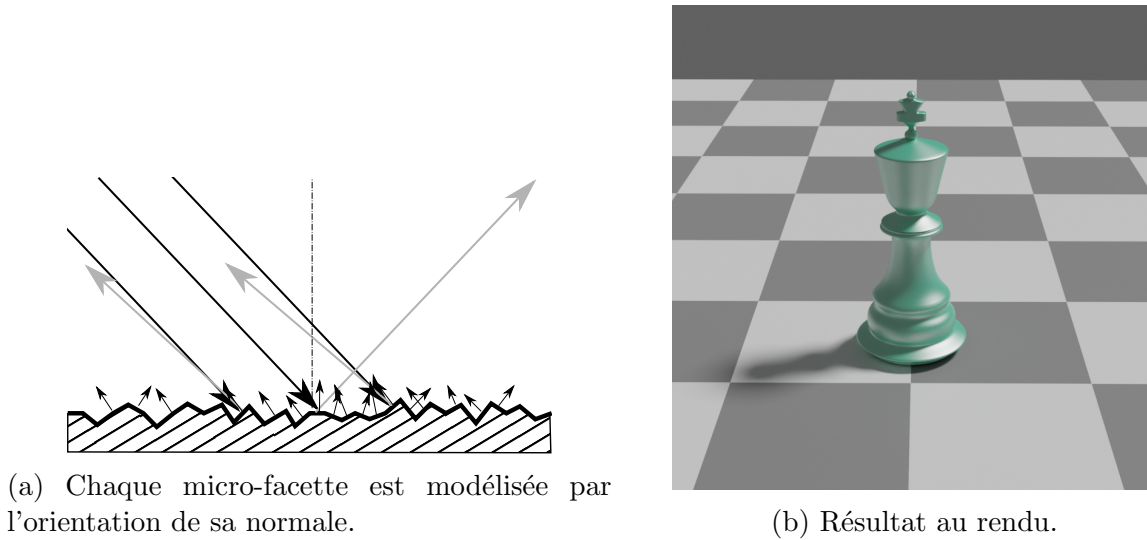


FIGURE 4.7 – Le modèle de Cook-Torrance

Ici, le matériau est modélisé par un grand nombre de micro-facettes spéculaires.

4 Simulation de la propagation de la lumière

4.1 Modèle de propagation de la lumière : l'équation du rendu

Nous pouvons maintenant formuler l'expression de la luminance monochromatique $L(x \rightarrow \vec{\omega}_r)$ émise en un point x d'une surface dans la direction $\vec{\omega}_r$. À l'équilibre, cette luminance est la somme de la luminance propre L_e émise par la surface (cas d'une source de lumière par exemple) et de la réflexion L_r de toute la lumière qui arrive sur cette surface.

$$L(x, \vec{\omega}_r) = L_e(x, \vec{\omega}_r) + L_r(x, \vec{\omega}_r) \quad (4.1)$$

L'énergie qui arrive sur une portion de surface dans une portion d'angle solide $d\vec{\omega}_i$ vaut

$$L(x, \vec{\omega}_i) \cdot \cos(\vec{N}_x, \vec{\omega}_i) d\vec{\omega}_i \quad (4.2)$$

Avec $L(x, \vec{\omega}_i)$, qui désigne la luminance incidente et \vec{N}_x la normale à la surface au point x .

La réflexion est contrôlée par la BRDF (cf. paragraphe 2.2 de ce chapitre). Ainsi, l'équilibre lumineux pour une longueur d'onde est défini par l'équation du rendu de Kajiya (1986)⁸ :

$$L_r(x, \vec{\omega}_r) = L_e(x, \vec{\omega}_r) + \int_{\Omega} f_r(x, \vec{\omega}, \vec{\omega}_i) \cdot L_i(x, \vec{\omega}_i) \cdot \cos(\vec{N}_x, \vec{\omega}_i) d\vec{\omega}_i \quad (4.3)$$

Dans cette équation, la luminance incidente $L_i(x, \vec{\omega}_i)$ est aussi inconnue. Elle peut être calculée par lancer de rayons : si un rayon émis à partir du point x rencontre la surface la plus proche au point y , alors $L_i(x, \vec{\omega}_i) = L_r(y, -\vec{\omega}_i)$ (dans un milieu non participatif, c'est-à-dire un milieu qui ne dévie pas les rayons lumineux et n'absorbe pas leur énergie). L'angle solide sous lequel on voit la portion de surface dy est alors : $d\vec{\omega}_i = \frac{\cos(\vec{N}_x, \vec{\omega}_i)}{d_{xy}^2} dy$. En introduisant la fonction $V(x, y)$ indiquant si x et y sont mutuellement visibles, l'équation 4.3 peut s'écrire :

8. Cette équation ne prend pas en compte les phénomènes de fluorescence, de phosphorescence ou les effets de polarisation. Elle est cependant largement suffisante pour prendre en compte la majorité des effets lumineux nécessaires en synthèse d'images réalistes.

$$L_r(x, \vec{\omega}_r) = L_e(x, \vec{\omega}_r) + \int_{\Omega} f_r(x, \vec{\omega}_r, \vec{\omega}_i) \cdot L_i(y, -\vec{\omega}_i) \cdot V(x, y) \cdot G(x, y) dy \quad (4.4)$$

où $G(x, y)$ est un terme géométrique :

$$G(x, y) = \frac{|\cos(\vec{N}_x, \vec{\omega}_i)| \cdot |\cos(\vec{N}_y, -\vec{\omega}_i)|}{d_{xy}^2}$$

Remarque : l'énergie réfléchie par une surface peut être divisée en deux composantes : l'énergie réfléchie provenant directement des sources de lumière, appelée *éclairage direct*, et l'énergie réfléchie provenant de la réflexion de la lumière sur les autres surfaces de la scène, appelée *éclairage indirect*. Elles sont exprimées par les équations suivantes, avec S l'ensemble des surfaces émettrices et S' l'ensemble des surfaces réfléchissantes :

$$L_{direct}(x, \vec{\omega}) = \int_S f_r(x, \vec{\omega}, \vec{\omega}_i) L_e(y, -\vec{\omega}_i) V(x, y) G(x, y) dy$$

$$L_{indirect}(x, \vec{\omega}) = \int_{S'} f_r(x, \vec{\omega}, \vec{\omega}_i) L_r(y, -\vec{\omega}_i) V(x, y) G(x, y) dy$$

$$L_r(x, \vec{\omega}) = L_{direct}(x, \vec{\omega}) + L_{indirect}(x, \vec{\omega})$$

Cette séparation peut être utilisée pour faciliter et/ou améliorer le processus de résolution de l'équation du rendu.

4.2 Méthodes de résolution de l'équation du rendu

4.2.1 Les méthodes stochastiques

Ces méthodes reposent sur des techniques d'intégration de Monte Carlo pour le calcul d'une intégrale. Soit $I = \int_0^1 f(x) dx$ l'intégrale à calculer. En effectuant N tirages aléatoires uniformes θ_i sur $[0, 1]$ et indépendants, on définit une variable aléatoire $\hat{I} = \frac{1}{N} \sum_{i=1}^N f(\theta_i)$ que l'on appelle *estimateur de I*. Cet estimateur est sans biais c'est-à-dire que l'espérance de la variable aléatoire \hat{I} est égale à I . L'erreur commise en prenant \hat{I} à la place de I est de l'ordre de $\frac{1}{\sqrt{N}}$. L'augmentation du nombre d'échantillon fait converger la somme vers la valeur de solution.

4.2.1.1 Le tracé de rayons (ray tracing)

L'algorithme le plus simple est le tracé de rayons, introduit par Appel (1968a) et popularisé par Whitted (1980). Le principe de l'algorithme (illustré par la figure 4.8) est le suivant : pour chaque pixel de l'écran, on lance un rayon provenant de l'œil. On cherche l'intersection de ce rayon avec tous les objets de la scène. Si au moins une intersection est trouvée, on garde la plus proche de l'œil. A partir de ce point d'intersection, on peut lancer des rayons d'ombre en direction de chaque source de lumière (supposée ponctuelle) pour savoir si le point est éclairé ou pas. Si le matériau rencontré est spéculaire, on peut lancer un rayon secondaire réfléchi et dans le cas où le matériau est transparent, on peut également lancer un rayon réfracté. Cette méthode revient donc à construire, pour chaque pixel, un arbre de rayons dont l'évaluation fournit l'énergie atteignant le pixel dans la direction de l'œil, donc la couleur de ce pixel.

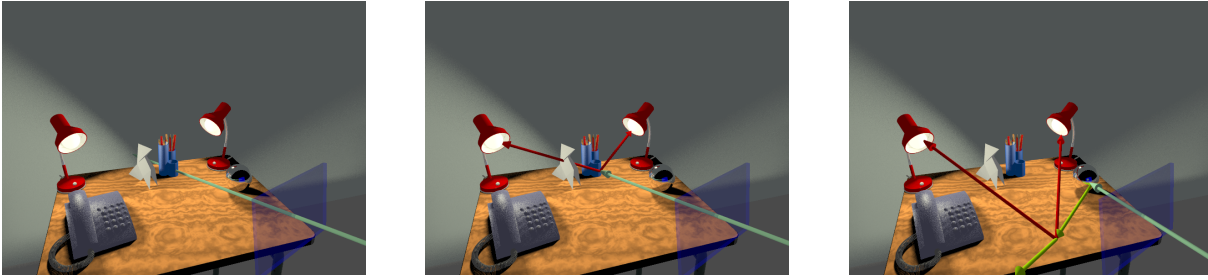


FIGURE 4.8 – Principe du lancer de rayons.

À gauche, un rayon (dit primaire) traverse un pixel de l'écran et rencontre un objet. L'éclairage direct du point trouvé est calculé en lançant un rayon d'ombrage vers chaque source de la scène (au centre) puis l'éclairage par réflexion est calculé en lançant un rayon réfléchi (à droite). Le processus est répété récursivement.

4.2.1.2 Le tracé de chemins (path tracing)

Le cas le plus classique d'application de la méthode stochastique générale est le tracé de chemins (Kajiya, 1986) : pour chaque pixel de l'image, on trace une marche aléatoire x_0, x_1, \dots, x_i partant de l'œil de la manière suivante : en chaque point x_{i-1} , on lance un rayon dont la direction est donnée par une densité de probabilité $p_i(\theta_i)$ et le premier point touché par le rayon définit le point x_i . La densité de probabilité p_i est associée à la BRDF du matériau au point x_{i-1} , ce qui permet de prendre en compte tous les types de matériaux. Quand une source lumineuse est atteinte au cours d'une marche aléatoire, elle contribue à l'estimation du flux lumineux atteignant l'œil à travers le pixel concerné. Pour que cette méthode fournisse des résultats assez précis, il faut tracer N chemins pour chaque pixel. En effet, la probabilité de parvenir sur une source lumineuse en cours de construction du chemin est relativement faible, en particulier lorsque la taille des sources est petite en regard de la taille de la scène simulée ou que l'éclairage est fortement indirect. De ce fait, cette méthode est coûteuse en temps de calcul, car N peut prendre des valeurs importantes. Suivant le contexte, c'est souvent plusieurs milliers d'échantillons qu'il faut utiliser. De plus, elle a tendance à fournir des images bruitées (voir figure 4.9).

4.2.1.3 Le tracé de photons (light tracing)

Cette méthode (Jensen, 1996a, 1996b, 2001) consiste à lancer des rayons possédant de l'énergie, appelés photons, depuis les sources de lumière, qui vont rebondir sur les surfaces rencontrées. À chaque fois qu'un tel chemin passera à travers la surface d'un pixel de l'écran, son énergie, atténuée par les rebonds sur les surfaces, sera ajoutée à celle du pixel. Cette méthode souffre des mêmes inconvénients que ceux décrits pour le tracé de chemins, puisque la probabilité qu'un photon émis par une source passe par un pixel de l'écran est généralement très faible.

4.2.1.4 Méthodes bidirectionnelles

Comme nous l'avons déjà signalé, les méthodes du tracé de chemins ou du tracé de photons fournissent des images très bruitées, car la probabilité qu'un chemin lumineux atteigne une source de lumière ou l'œil est très faible (on obtient un pixel noir si aucune source n'a été atteinte ou un pixel très lumineux dans le cas où un chemin de probabilité très faible a été trouvé (à cause de la division par cette probabilité dans l'équation (4.5)). Pour essayer de résoudre ce problème, les méthodes bidirectionnelles ont été introduites indépendamment par Lafortune et



FIGURE 4.9 – Bruit de Monte-Carlo

Illustration du phénomène de bruit présent dans les images générées par la méthode de tracé de chemins. L'image de gauche est obtenue en utilisant un seul chemin par pixel; l'image du centre utilise 25 chemins par pixel un bruit perceptible reste présent; l'image de droite, utilise 500 chemins par pixel.

Willems, 1993 et Veach et Guibas, 1995 : des rayons sont envoyés à la fois à partir des sources de lumière et à partir de l'œil, puis combinés, augmentant ainsi la corrélation entre les sources lumineuses et l'œil. La combinaison des chemins partant de l'œil ou des sources de lumière est réalisée par un *échantillonnage d'importance multiple* (MIS, Multiple Importance Sampling) qui peut être défini de la manière suivante : soient $\omega_1(x), \dots, \omega_N(x)$ des nombres réels tels que $\omega_i(x) \geq 0, \forall i, \sum_{i=1}^N \omega_i(x) = 1$. Soient $X_{i,j}$ des variables aléatoires indépendantes de densité p_i . Alors

$$\tilde{I} = \sum_{i=1}^N \frac{1}{N_i} \sum_{j=1}^{N_i} \omega_{i,j}(X_{i,j}) \frac{f(X_{i,j})}{p_i(X_{i,j})} \quad (4.5)$$

est un estimateur sans biais de I .

4.2.1.5 Autres méthodes

Les méthodes stochastiques constituent un domaine de recherche très actif depuis quelques années. Il existe donc un grand nombre d'algorithmes permettant de résoudre l'équation du rendu. Il n'est pas possible de développer ces méthodes dans le contexte de ce chapitre. Nous ne pouvons que les citer en renvoyant à la bibliographie les lecteurs intéressés : *Instant radiosity* Keller, 1998, *Metropolis* Veach, 1997, *Metropolis Instant Radiosity* Segovia et al., 2007, etc. et enfin une méthode très populaire dans le domaine de l'informatique graphique, mais qui est biaisée : les *cartes de photons* (photon mapping) introduites par Jensen (2001).

En Résumé

Les méthodes stochastiques basées sur l'estimation de Monte Carlo permettent de calculer des images photo-réalistes en tirant partie d'une modélisation des sources de lumière, des matériaux, de la géométrie et du transfert de la lumière dans une scène virtuelle.

Ces méthodes permettent un rendu physiquement réaliste au prix d'un temps de rendu important. Dans le cas où trop peu d'échantillons ont été utilisés, les images paraissent bruitées.

Chapitre 5

Apports réciproques de la psychologie de la perception et de la synthèse d'image photo-réaliste

« Celui qui trouve sans chercher est celui qui a longtemps cherché sans trouver. »

Gaston Bachelard

L'objectif de ce chapitre est de proposer une illustration de quelques approches qui associent synthèse d'image et étude du système perceptif. Bien que les travaux présentés ici couvrent des domaines et des problématiques assez différentes, ils possèdent une démarche commune : le travail transdisciplinaire qui associe la méthode de la psychologie expérimentale et la synthèse d'image.

Dans ce cadre général, les sections suivantes vont présenter plusieurs études qui ont été menées. La première section montre, à travers une illustration, comment l'image de synthèse et plus généralement la réalité virtuelle constitue un enjeu pour l'étude du comportement visuel. La seconde partie s'intéresse à un phénomène : les artefacts visuels qui sont à l'interface entre les techniques de production d'images et le système visuel. Et enfin, la troisième section couvre le problème de l'étude de la perception du bruit dans l'image de synthèse photoréaliste.

Ce chapitre est aussi une occasion de présenter une démarche originale de recherche qui, sur la base d'une collaboration entre le LISIC et SCALab, a amené à travailler dans le cadre d'un projet financé par l'Agence Nationale de la Recherche (ANR), le projet PrISE-3D (voir la partie 6 des Annexes, page 138).

1 Esquisse de laboratoires virtuels

1.1 Motivations

Depuis la première « image de synthèse », la bien connue théière de Martin Newell en 1975 (Torrence, 2006), la synthèse d'images numériques s'est imposée dans de très nombreux secteurs. Les industries du divertissement comme le cinéma, la télévision et les jeux vidéos

utilisent massivement l'image de synthèse. Au cours de la dernière décennie, des périphériques de réalité virtuelle se sont démocratisés (voir partie 3.3.2 page 64). Ils utilisent notamment la synthèse d'image, au départ essentiellement avec des méthodes de rendu temps réel. Cette évolution s'est également accompagnée de nombreuses études qui visent d'une part à comprendre les mécanismes de perception d'une image tout en utilisant abondamment l'image dans les dispositifs expérimentaux.

La quantité encore plus importante de pixels à traiter, ces périphériques peuvent sembler un nouvel obstacle à l'utilisation de méthodes photo-réalistes. La réalité virtuelle est un nouveau défi pour la synthèse d'image réaliste mais paradoxalement, elle apporte peut-être également des possibilités nouvelles dans la mesure où les dispositifs intègrent de nombreux capteurs visuels et non visuels. C'est le cas en particulier de la direction du regard. Ils permettent ainsi d'envisager un « rendu fovéal » des images qui n'était pas réellement possible sur des périphériques plus classiques (Patney et al., 2016). La partie pour laquelle la sensibilité est maximale peut alors être calculée avec le plus de précision alors que dans les zones périphériques il est possible d'utiliser de nombreuses optimisations et approximations. L'utilisation de cette méthode en conjonction avec des méthodes d'estimation du bruit pourrait concourir à intégrer un rendu photo-réaliste dans un temps s'approchant du temps réel dans la portion la plus visible de l'image. Tout ceci concourt à faire de l'image de synthèse un objet incontournable pour la perception.

La synthèse d'image connaît également une utilisation croissante dans les laboratoires de psychologie expérimentale notamment par le biais de la réalité virtuelle, dans le cadre de nombreuses applications. Le cabinet Goldman Sachs (2018), estime que le nombre d'utilisateurs de la réalité virtuelle devrait connaître une forte augmentation. En particulier, les usages liés à l'apprentissage pourraient atteindre les 15 millions d'utilisateurs à l'horizon 2025. On note également une augmentation du nombre de recherches sur la perception qui utilisent la réalité virtuelle, surtout depuis 2016 comme illustré par la figure 5.1.

La réalité virtuelle permet de nombreuses applications, y compris dans le domaine de la psychologie clinique. En particulier, pour le traitement des phobies, la possibilité de placer progressivement des personnes dans des environnements variés permet une rééducation efficace, par exemple dans le cas de stress post-traumatique.

Des méta-analyses récentes (Riva et al., 2016 ; Riva et al., 2019) évaluent l'utilisation de la réalité virtuelle dans de nombreux domaines tels les troubles anxieux, les troubles alimentaires et la gestion de la douleur. Les études notent des effets à long terme qui se généralisent au monde réel. D'autres études montrent également que ces dispositifs peuvent avoir une application à d'autres domaines comme le traitement de troubles psychotiques et la dépendance.

1.2 Raisons de l'utilisation de la synthèse d'images en sciences du comportement

1.2.1 Méthode

Un élément important de la démarche scientifique est de pouvoir tirer des conclusions de causalité entre une ou plusieurs variables.

La première étape de la découverte d'un tel lien est souvent l'observation d'une corrélation entre deux facteurs. Par une observation provoquée ou fortuite, ces deux facteurs semblent reliés. Cette observation n'est cependant pas suffisante à établir un lien causal d'un facteur sur l'autre. De nombreux autres cas de déterminisme peuvent être observés. En particulier, les deux facteurs peuvent être co-déterminés par un troisième facteur, appelé facteur caché. Il n'est donc pas possible de déduire légitimement une relation de cause à effet entre deux événements ou

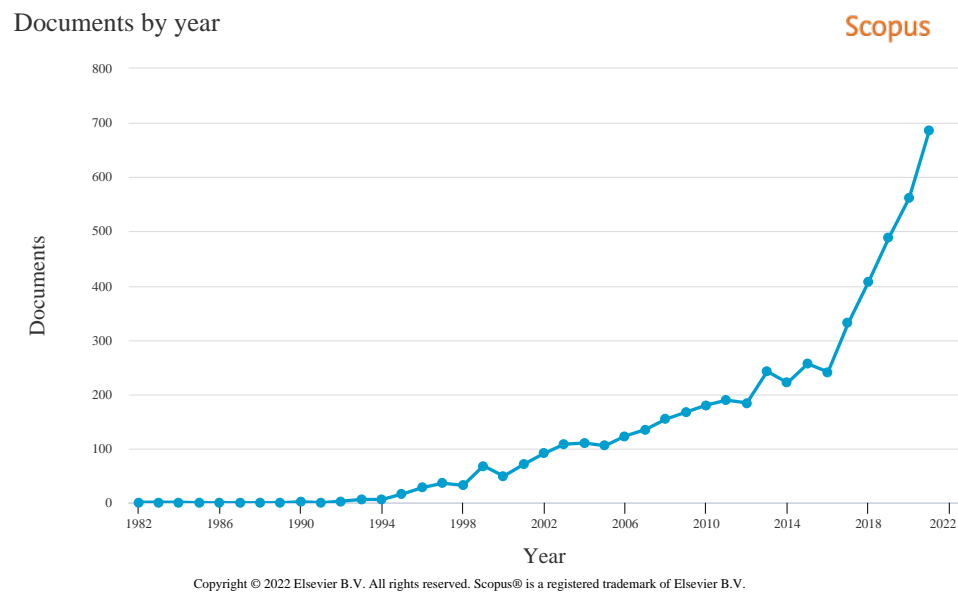


FIGURE 5.1 – Tendances sur la perception et réalité virtuelle

Évolution du nombre de documents qui traitent de perception et de réalité virtuelle référencés par Scopus. Requête (TITLE-ABS-KEY ("virtual reality" OR vr) AND TITLE-ABS-KEY (percepti*)) AND DOCTYPE (ar OR re) AND PUBYEAR > 1981 AND PUBYEAR < 2022

variables sur la seule base d'une association ou d'une corrélation observée entre eux¹.

Les critères d'administration de la preuve d'un lien causal sont à la base de la méthode expérimentale. Les variables sont identifiées et manipulées à travers un test d'hypothèse. L'hypothèse n'est considérée comme vraisemblable comme hypothèse candidate tant que les données expérimentales ne viennent pas la contredire (Popper, 2006).

Afin de mettre à l'épreuve ces hypothèses, il est alors nécessaire d'établir des plans expérimentaux qui visent à manipuler les facteurs en contrôlant la variation de tous les autres. On enregistre l'effet de ces facteurs contrôlés, appelés *variables indépendantes* sur les *variables dépendantes*. À l'aide de tests de décision statistiques, il est alors possible de généraliser ou non les résultats expérimentaux observés sur un échantillon d'une population.

1.2.2 Application à l'image

L'étude de la perception visuelle a profité des développements liés à toutes les techniques de l'image. Il est même intéressant de noter que l'évolution des connaissances relatives à la perception suit celle des techniques de production des images. On peut même considérer qu'il y a une co-évolution entre connaissance du système perceptif, savoirs relatifs à la capture et reproduction d'images réelles. Par exemple, les travaux de Maxwell relatifs à la physique de la

1. La confusion entre corrélation et causalité est appelée *effet cigogne*. Une observation fortuite de la fréquence des nids de cigogne dans certains villages d'Alsace et la natalité avait conduit à une croyance populaire d'un lien de causalité entre cigognes et naissances. Cette erreur d'attribution, comme le lien entre le déclin de la population des pirates et le réchauffement climatique mis en avant avec humour par les pastafaristes, est très courant, même dans la littérature scientifique. Le site <http://tylervigen.com/spurious-correlations> en recense un grand nombre.

lumière lui permettent de conforter la théorie trichromatique de la perception de la couleur. Ces connaissances vont l'amener à produire la première photographie couleur en 1861 (Timby, 2005). De façon naturelle, toutes les techniques d'imagerie ont alimenté les laboratoires d'étude de la perception et leurs travaux ont guidé les développements technologiques.

De manière assez logique, l'image de synthèse a remplacé pour partie la photographie pour la conception de stimulus dont il est facile de contrôler de nombreux paramètres. Alors que dans une photographie, il faut user de techniques de traitement d'image, la synthèse d'image présente l'intérêt de pouvoir manipuler toutes les variables perceptives avec finesse. Il est par exemple possible de décider quel objet aura ou non une ombre ou de sa forme très précisément.

1.3 Un exemple d'utilisation de l'image 3D

Cette partie décrit un travail qui a été mené dans le cadre d'une collaboration entre SCALab et le LISIC. Elle s'inscrit dans la problématique générale de la liaison entre perception et action, plus précisément sur l'impact des fonctionnalités des objets sur le jugement à les atteindre.

1.3.1 Exploration du couplage perception-action

Schématiquement, si on suit le modèle informatique, il a été d'usage de considérer que la perception constitue les mécanismes d'entrée (« *input* ») alors que les actions motrices sont considérées comme des sorties (« *output* ») du système nerveux. Ce découpage, issu de la cybernétique, pose néanmoins de nombreux problèmes face aux observations. Les données expérimentales vont beaucoup plus dans le sens d'une intégration forte de la perception à des schémas d'action.

Si en situation de laboratoire, il est possible d'étudier l'action ou la perception de façon plus ou moins indépendante, il convient de noter que dans une approche écologique, perception et action sont fortement couplées (J. Gibson, 2015; Gordon, 2004). Cette influence du contrôle moteur sur la capacité à reconnaître a fait l'objet de nombreuses observations qui ne se limitent pas à la perception visuelle. Par exemple, dans le cas de la reconnaissance de la parole, il a été montré que dès les premiers mois de la vie, la perception et la production de la parole sont étroitement liées. Les mécanismes neuronaux à l'origine de la perception de certains sons et ceux responsables de leur production sont dépendants l'un de l'autre. (McGettigan & Tremblay, 2018).

Ce lien s'observe aussi au niveau visuel. Par exemple la perception du flux optique est en étroite interaction avec le contrôle de la locomotion (Warren, 1990). Il a été également montré que les représentations motrices dynamiques utilisées pour déterminer ce qui est atteignable font également partie du processus perceptif menant à la représentation distincte des espaces péri-personnels et extra-personnels (Bartolo et al., 2014).

La réalité virtuelle s'avère donc un outil d'exploration de la perception, de l'action et de leurs interactions.

1.3.2 Stimulus artificiels

Il existe de nombreux indices que les représentations motrices associées à des objets manipulables sont utilisées même en l'absence de planification en direction de l'objet (Buxbaum & Kalénine, 2010). Ceci suggère qu'il existe d'étroites relations entre l'action, la perception et la représentation des objets.

Dans une étude sur l'estimation de la capacité à atteindre des objets, Kalénine et al. (2016) ont montré que chez les adultes, le traitement perceptif est affecté par la compétition entre

les représentations d'action. Lorsqu'on demande un jugement sur l'atteignabilité d'objets, la fonction de l'objet joue un rôle.

Relativement à cette tâche, on peut séparer les objets en deux catégories.

- Pour certains des objets, les mouvements pour les prendre et les utiliser sont cohérents. C'est le cas, par exemple d'un verre ou d'une balle. Ces objets sont considérés comme « non conflictuels » puisque la réponse motrice sera très proche — sinon identique — dans les deux cas. Le geste qui correspond au fait de saisir la balle et de l'utiliser sont dans ce cas quasi-similaires.
- Pour d'autres objets, les réponses motrices requises pour les prendre ou les utiliser sont différentes. C'est ce qui s'observe dans l'exemple d'une calculatrice ou d'un distributeur à savon. Dans la mesure où les mouvements sont différents, les objets sont considérés comme « conflictuels ».

La conflictualité se traduit par une augmentation du temps de décision sur un jugement d'atteignabilité. Tout se passe comme s'il y avait une étape de préparation du mouvement : une étape de planification. Lorsque l'objet est « non conflictuel » cette étape serait plus rapide que lorsqu'elle doit inclure une décision sur la nature du geste qui doit être planifié.

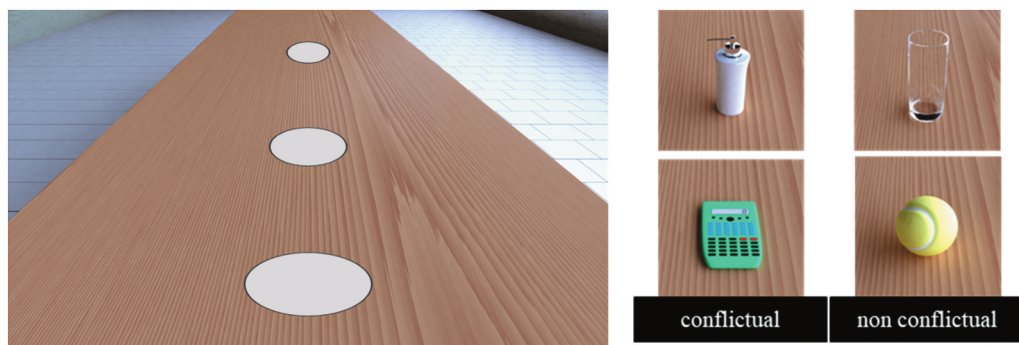


FIGURE 5.2 – Un dispositif expérimental dans l'environnement 3D.

À gauche, sont représentées les différentes distances (par exemple, proche, limite d'accessibilité, éloignée) auxquelles les stimulus sont présentés dans la scène virtuelle. À droite, sont affichés des exemples d'objets conflictuels et non conflictuels. La tâche des participants consistait à juger la catégorie de l'objet (« est-ce un objet de cuisine ? ») ou l'accessibilité de l'objet (« est-il accessible ? »), en répondant à l'aide de pédales.

Dans une étude récente portant sur le développement Godard et al. (2022) mesurent cet effet en utilisant des images de synthèse. Pour ceci, il a été nécessaire de réaliser de nombreux objets (40 objets sélectionnés sur la base d'un pré-test parmi 70 modèles). Quatre de ces objets sont présentés en figure 5.2.

Cinq groupes d'âge sont inclus : 8, 10, 12, 14 ans ainsi qu'un groupe de jeunes adultes. Cette étude montre que la compétition provoquée par l'activation de représentations d'actions conflictuelles pendant le traitement des objets induit un coût. Ceci s'observe chez les enfants dès l'âge de 8 ans. De façon critique, le développement du coût du conflit montre une trajectoire en forme de U entre l'âge de 8 ans et l'âge adulte. L'effet pénalisant de la conflictualité semble s'estomper entre 10 et 14 ans. Cette évolution non linéaire met en évidence des périodes distinctes de haute sensibilité à la compétition entre les représentations d'action et suggère des changements complexes dans l'activation des représentations d'action au cours des premières années de l'adolescence.

Dans l'ensemble, les résultats actuels fournissent des informations nouvelles sur la façon dont l'âge peut modifier l'impact des représentations d'action sur le traitement des objets. Ceci peut contribuer à enrichir les vues théoriques sur la sélection d'action et l'incarnation d'objet.

2 Étude des artefacts visuels

L'étude des artefacts visuels représente typiquement un problème situé à l'interface entre l'humain et la technologie. Ils sont non seulement inhérents à la méthode de production, de diffusion ou de stockage des images, mais ils impactent aussi notablement la perception de ces images.

2.1 Définition de l'artefact visuel

Les artefacts visuels, parfois appelés simplement artefacts², sont des anomalies qui apparaissent lors de la représentation visuelle. De nombreux domaines de l'imagerie sont touchés par des artefacts visuels. Qu'il s'agisse de déformations liées à l'optique, à l'électronique, à la transmission ou à tout traitement effectué sur les images, ils ont généralement une « signature » visuelle très caractéristique et des conditions d'apparition bien déterminées. Souvent, leur simple présence trahit la méthode utilisée pour produire l'image ou la transmettre mais de façon générale, il est difficile de les éviter. La figure 5.3 présente par exemple des artefacts optiques produits par une lentille : des *Flare* et des *Starburst*.



(a) Effet de *Flare* produit par une lentille de 18mm avec une ouverture 4.0f



(b) Effet de *Flare* et de *Starburst* produits avec la technique de Hullin et al. (2011) (CC BY-SA 3.0)

FIGURE 5.3 – Exemples d'artefacts optiques.

2.2 Artefacts en synthèse d'images

En informatique, on parle d'*artefacts computationnels* puisque ce sont généralement des processus de calculs qui induisent des distorsions dans l'image. En synthèse d'images, la valeur d'un pixel résulte d'un calcul qui peut comporter des approximations, des erreurs ou des

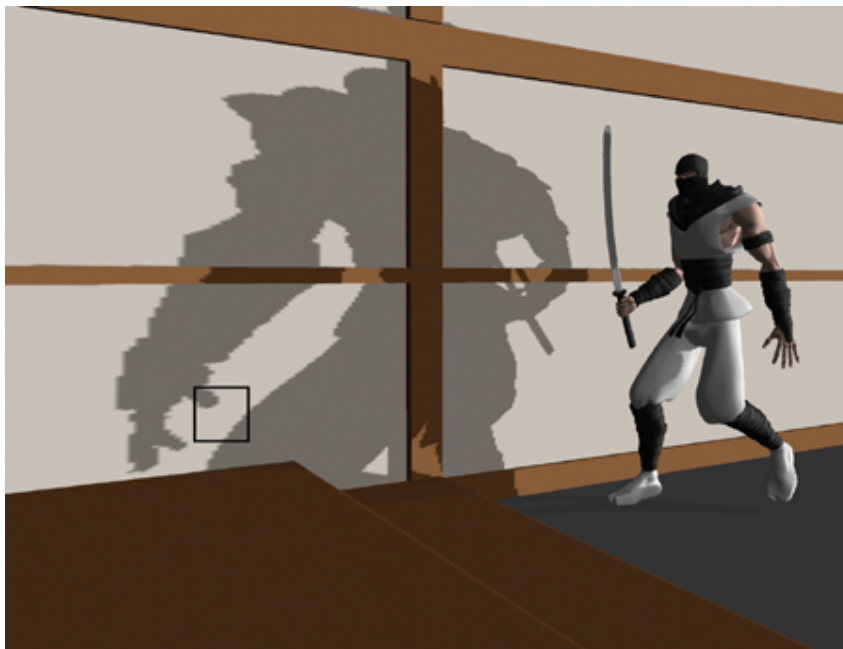
2. De façon générale un artefact — du latin *arte factus* « fait avec art » — désigne tout phénomène créé artificiellement par les conditions expérimentales, par les instruments, les périphériques ou les traitements utilisés.

arrondis numériques qui peuvent altérer le contenu de l'image produite. Par artefact computationnel, on entend toute altération qui résulte de l'algorithme de production de l'image ou de son implémentation sur une architecture donnée. Par exemple, suivant la méthode de calcul, le découpage même de l'image en une grille de pixels peut être perçu. Ceci aboutit à la perception des pixels individuellement, ce qui est généralement dénommé par un effet de crénelage ou *aliasing*. Cet effet de « pixellisation » conduit à voir des motifs tels des marches d'escalier le long de courbes ou de droites qui devraient paraître continues. C'est ici la structure de grille de pixels imposée par l'affichage sur un écran qui est à l'origine d'un phénomène perceptif. Les phénomènes d'hyper-acuité (voir partie 1.4.3 du chapitre III, page 14) ont tendance à accentuer la perception de ce problème.

Le rendu d'images photoréalistes reste particulièrement coûteux en temps d'exécution. En particulier, l'utilisation des algorithmes d'illumination globale comme présentée dans le chapitre précédent permet le rendu d'une image à haute résolution avec un éclairage physiquement réaliste. Cependant, le temps de rendu peut représenter plusieurs minutes voire plusieurs heures.

C'est pourquoi une autre solution consiste à utiliser des algorithmes plus rapides mais qui reposent sur des approximations. Ces techniques d'illumination — biaisées — vont permettre de produire une image en un temps nettement plus acceptable, parfois en temps réel. Elles vont cependant introduire d'autres artefacts visuels. L'image générée semble de bonne qualité dans sa plus grande partie mais dans certaines zones spécifiques des erreurs importantes subsistent.

Herzog et al. (2012) notent, par exemple, que la Radiosité instantanée avec *Virtual Spherical Light* (Keller, 1997) produit un bruit de basse fréquence. La *rasterisation* Open-GL avec des *shadow maps* (Reeves et al., 1987) induit un *aliasing* des ombres (figure 5.4). L'algorithme du Lightcut pour l'illumination directe (Walter et al., 2005) a tendance à assombrir les coins de l'image, ce qui provoque une modification semblable au vignettage observé en photographie³.



(a) L'ombre projetée comporte un crénelage.



(b) Vue détaillée de l'ombre dans la zone encadrée.

FIGURE 5.4 – Problème d'*aliasing* des ombres (source NVIDIA).

Les deux sections qui suivent s'intéressent plus précisément à deux artefacts visuels qui

3. Le vignettage est l'assombrissement de la périphérie (coins sombres) d'une photographie.

caractérisent l'image de synthèse photo-réaliste : les *fireflies* et le bruit visuel.

3 *Fireflies*

Note: Cette section présente succinctement les travaux à propos des fireflies. Le lecteur trouvera plus de précision sur la démarche et les résultats dans le chapitre 6 de la thèse de Jérôme Buisine (2021) ainsi que dans Buisine, Delepoulle et Renaud (2021a).

3.1 Problème

En analysant le comportement du bruit produit par les méthodes de Monte-Carlo, il apparaît que les variations dans l'estimation peuvent aboutir à des cas extrêmes 5.5. C'est le cas lorsque le calcul fait apparaître une contribution très intense associée à une faible probabilité d'apparition. Dans ce cas, cette valeur, qui peut être considérée plus ou moins comme une donnée aberrante, provoque une surestimation de la valeur de un ou quelques pixels.

Cet artefact très visible est connu dans la littérature sous le nom de *firefly*⁴ (DeCoro et al., 2010; J. W. Jung et al., 2015; Zirr et al., 2018). Il est souvent difficile de l'éliminer automatiquement.

3.2 Médiane des moyennes

Une part du problème réside dans l'utilisation de l'estimateur de la moyenne qui est très sensible aux valeurs extrêmes. Il apparaît donc plus intéressant de privilégier des estimateurs plus robustes tels que la médiane des moyennes (MoN pour *Median of meaNs*) qui est une alternative intéressante à cet estimateur classique⁵. La médiane des moyennes consiste à répartir les échantillons en groupes pour lesquels autant de moyennes sont calculées. On sélectionne comme estimateur final la médiane de ces groupes comme montré en figure 5.6.

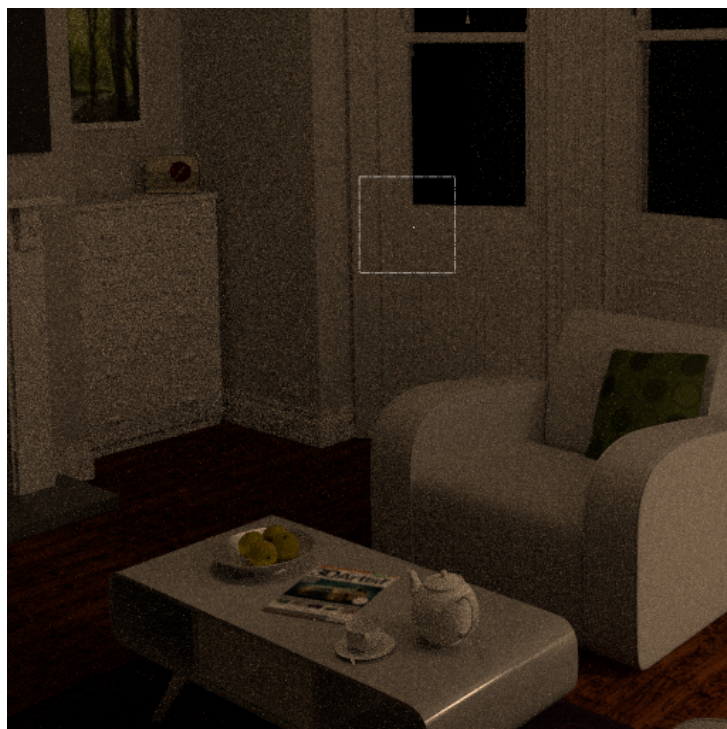
$$\hat{\mu}_{MoN} = \text{mediane} \left(\frac{1}{k} \sum_{i=1}^k x_i, \dots, \frac{1}{k} \sum_{i=n-k+1}^n x_i \right) \quad (5.1)$$

3.3 MoN adaptif avec le coefficient de Gini

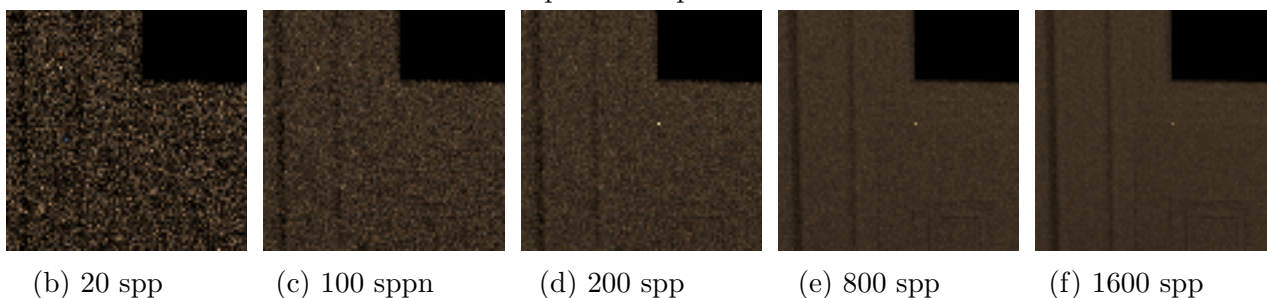
Si l'estimateur MoN s'avère particulièrement intéressant pour éviter l'apparition des *fireflies*, il semble néanmoins converger plus lentement que la moyenne dans le cas général. Nous nous sommes donc intéressés, dans le cadre de la thèse de Jérôme Buisine, à étudier la possibilité d'accélérer la convergence du MoN, en particulier via l'utilisation du coefficient de Gini.

4. Les lucioles ou *Lampyridae* sont des coléoptères qui produisent souvent de la lumière. Ici, le fait qu'un point très brillant puisse apparaître momentanément dans une image puis semble s'estomper rappelle le comportement des lucioles.

5. La médiane elle-même pourrait constituer un estimateur robuste, cependant son utilisation nécessite d'une part de stocker tous les échantillons et d'autre part de les classer ensuite, ce qui représente un surcoût important dans les traitements.



(a) Le *firefly* est un artefact très visible dans l'image, même s'il ne concerne qu'un seul pixel.



(b) 20 spp

(c) 100 spp

(d) 200 spp

(e) 800 spp

(f) 1600 spp

FIGURE 5.5 – Problème des *Firefly*

La convergence des valeurs d'éclairage se traduit par une diminution régulière du bruit dans l'image à mesure qu'on ajoute des échantillons pour chaque pixel ($spp = \text{sample per pixel}$). Cependant, la présence d'un chemin à très haute contribution mais peu probable se traduit par l'apparition d'un *firefly* pendant le calcul, ici entre le 100^e et le 200^e échantillon. Celui-ci reste visible longtemps, même lorsque le bruit s'estompe par ailleurs.

3.3.1 Coefficient de Gini

Le coefficient de Gini est introduit (Dorfman, 1979), afin d'étudier la présence de valeurs aberrantes localement pour chaque pixel. À partir de cet indicateur statistique, deux nouveaux estimateurs sont proposés en le combinant avec le MoN.

Le principe du coefficient de Gini est de proposer une mesure de l'inégalité. Il est issu des travaux d'économétrie afin de mesurer par exemple la répartition des richesses dans un pays à travers un indice. Il a été introduit par Corrado Gini (Gini, 1936) et représente la moitié de la différence absolue moyenne entre toutes les observations (E) rapportées à la moyenne des

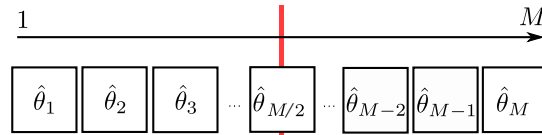


FIGURE 5.6 – Principe de l'estimateur MON

Tous les échantillons sont répartis en groupes d'effectif équivalents, par une affectation cyclique par exemple. Pour chaque groupe, la moyenne est calculée et on retient comme estimateur la médiane de ces dernières. Ici, c'est la moyenne du groupe central qui est retenue. Si le nombre de groupes est pair, on retiendra la moyenne des deux groupes centraux.

valeurs (*Moy*) :

$$G = \frac{1}{2} \times \frac{E}{Moy} \quad (5.2)$$

Avec

$$E = \frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n |x_i - x_j| \quad (5.3)$$

et

$$Moy = \frac{1}{n} \sum_{k=1}^n x_k \quad (5.4)$$

Il varie entre 0, qui correspond à une répartition parfaitement égalitaire et 1 qui correspond à la répartition totalement inéquitable. L'un des avantages de cet indice est qu'il peut être facilement approximé lorsque les données sont regroupées par classes. Dans ce cas, on peut le représenter par :

$$G = 1 - \sum_{k=0}^{n-1} (X_{k+1} - X_k)(Y_{k+1} - Y_k) \quad (5.5)$$

Avec X la part cumulée de la population et Y la part cumulée de la variable observée ($Y_{k+1} \leq Y_k$)

3.3.2 Choix binaire de l'estimateur

Une première solution consiste à calculer le coefficient de Gini et à l'utiliser comme heuristique de choix entre deux estimateurs. L'idée est que lorsque le coefficient de Gini est faible, la moyenne est un bon estimateur de la valeur d'éclairage. A l'inverse, si le coefficient dépasse une borne (B), la moyenne sera affectée par les valeurs extrêmes et c'est l'estimateur MoN qui représente mieux la valeur d'éclairage. Pour les expérimentations, une valeur de $B = 0,25$ a été retenue. Cet estimateur qui retient moyenne ou MoN selon un critère binaire sera noté $G-MoN_b$:

$$G-MoN_b = \begin{cases} \bar{\theta} & \text{si } G \leq B \\ \hat{\mu}_{MoN} & \text{sinon.} \end{cases} \quad (5.6)$$

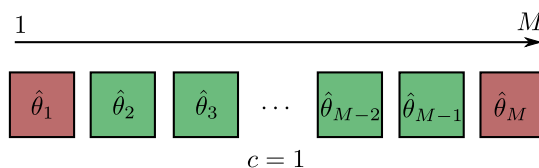


FIGURE 5.7 – Illustration du nombre de moyennes retenues pour le calcul.

3.3.3 MoN adaptatif

Pour rappel, dans le calcul de la médiane des moyennes, uniquement les données appartenant au groupe médian sont utilisées pour le calcul de l'estimateur. Bien que les autres données y contribuent indirectement par le biais du classement induit par la médiane, il y a une sous-exploitation des données disponibles (Orenstein, 2019) ce qui se traduit par une convergence plus robuste mais aussi plus lente.

Le principe du *G-MoN* adaptatif est de « réintégrer » une partie des groupes qui ne sont pas utilisés pour le calcul de la moyenne afin d'accélérer la convergence.

Le second estimateur, nommé *MoN adaptatif* sera donc calculé comme la moyenne arithmétique des groupes centraux :

$$\hat{\mu}_{G-MoN} = \frac{\sum_{j=1+c}^{M-c} \hat{\theta}_j}{M - 2c} \quad (5.7)$$

avec $\hat{\theta}_j$, la j^{e} moyenne. Les moyennes des groupes sont ordonnées de la même manière que pour le calcul de la Médiane des Moyennes. La valeur $2c$ correspond au nombre de moyennes qui ne sont pas incluses dans le calcul comme illustré en figure 5.7 : on retire c moyennes des deux côtés. La valeur c est choisie de façon à utiliser une plus grande part des moyennes lorsque le coefficient de Gini est faible. Il est alors possible d'accorder une plus grande confiance à la moyenne. En pratique la valeur c est calculée comme $c = \lfloor G \times k \rfloor$ avec $k = \lfloor \frac{M}{2} \rfloor$.

L'un de ces deux estimateurs, le MoN adaptatif exploitant le coefficient de Gini (G-MoN) montre une convergence rapide vers la solution et supprime les *fireflies*. Toutefois, il reste moins bon que l'estimateur proposé par (Zirr et al., 2018) qui traite des informations des pixels voisins contrairement à G-MoN qui est purement local.

3.4 Comparaison des estimateurs

3.4.1 Scènes utilisées

Afin d'étudier la convergence des estimateurs, quatre images ont été calculées avec le moteur de rendu PBRT version 4⁶ (voir figure 5.8).

Elles seront utilisées pour comparer les estimateurs entre-eux, mais également avec celui proposé par Jung qui s'apparente au MON dans la pratique.

3.4.2 Étude de la convergence

La figure 5.9 représente l'évolution de la SSIM (pour *Structural SIMilarity* : voir Z. Wang et al. (2004)) entre l'image en cours de calcul et l'image de référence. On constate que l'évolution des cinq estimateurs est relativement proche en l'absence de firefly (sous figure 5.9a). On retrouve la convergence relativement lente de MON. En présence de firefly (5.9b), MON est très efficace

6. <https://github.com/mmp/pbrt-v4>



FIGURE 5.8 – Images de référence

Images utilisées pour les comparaisons des estimateurs : Veach et Villa sont calculées avec 100 000 échantillons par pixel en utilisant BDPT, Bathroom et Crown avec 1 million d'échantillons en utilisant le tracé de chemins classique.

pour ne pas prendre en compte les grandes valeurs alors que la moyenne est fortement affectée par quelques valeurs. Cependant, dans tous les cas G - MoN semble présenter un comportement intéressant sur les différentes scènes.

3.5 Impact visuel

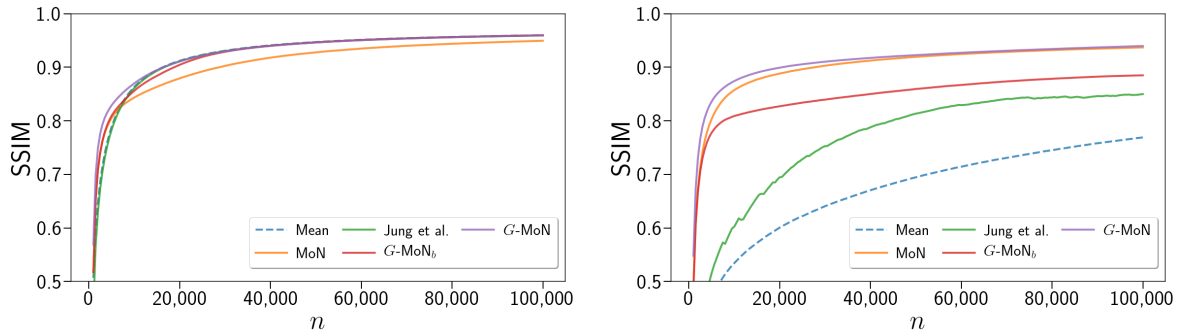
La figure 5.10 montre la comparaison des différents estimateurs pour 10 000 échantillons. Les méthodes basées sur l'estimation robuste (MoN et autre méthodes dérivées) produisent des images sans firefly. La méthode semble meilleure que celle de J. W. Jung et al., 2015. En revanche, d'autres méthodes arrivent à des résultats meilleurs. C'est le cas de la méthode de Zirr et al. (2018) à laquelle il est toutefois difficile de se comparer puisqu'elle exploite également des régularités inter-pixels.

4 Perception du bruit de Monte-Carlo en synthèse d'images photoréalistes

Un autre aspect de la démarche transdisciplinaire consiste à utiliser des données perceptives et à les modéliser afin d'optimiser les algorithmes de rendu d'images (Weier et al., 2017). C'est cette démarche qui est présentée ici, en s'appuyant sur les thèses de Nawel Takouachet et de Jérôme Buisine.

4.1 Positionnement du problème

Les méthodes stochastiques de résolution de l'équation du rendu permettent de calculer une solution physiquement correcte. Cependant, leur convergence est très lente et l'erreur liée à



(a) Comparaisons de cinq estimateurs pendant le rendu sur l'image de la scène *Bathroom*.

(b) Comparaisons de cinq estimateurs pendant le rendu de l'image de la scène de *Veach*.

FIGURE 5.9 – Convergence des estimateurs

Étude de convergence des estimateurs basés sur G avec $M = 21$ en utilisant l'indicateur SSIM jusqu'à 100 000 échantillons par pixel. Les estimateurs sont comparés à ceux de J. W. Jung et al. (2015), au MoN classique (avec $M = 21$) et à la moyenne sur deux scènes en utilisant les images de référence disponibles dans la figure 5.8 : Bathroom, sans firefly (à gauche) et Veach avec une grande quantité de fireflies (à droite).

l'estimation se traduit d'un point de vue visuel par la présence de « bruit ».

Ce bruit présente plusieurs caractéristiques perceptives.

1. Il est hautement visible.
2. Il est perçu comme une information de haute fréquence sur l'image.
3. Il s'apparente à un bruit « poivre et sel » ou *bruit impulsif* observé lors de transmissions de données ou de perturbations de capteurs.
4. Sa répartition n'est pas uniforme sur l'image mais il semble se concentrer dans certaines zones. La localisation des zones bruitées est affectée par de nombreux facteurs : quantité de lumière reçue, forme des ombres, matériaux, géométrie, etc.
5. La quantité de bruit semble diminuer avec le nombre d'échantillons.

Ces facteurs sont à l'origine d'un problème qui est de choisir le nombre d'échantillons nécessaire pour produire une image de bonne qualité. La qualité d'une image dans ce contexte pourrait se définir selon deux aspects. Il serait possible de mesurer l'erreur par rapport à une référence (qualité numérique) ou selon un observateur (qualité perceptive). C'est cette qualité perceptive qui est présentée ici.

Il s'agit donc de déterminer le seuil de perception du bruit induit par l'estimation de Monte-Carlo. Ce seuil, qui correspond à un nombre d'échantillons par pixel doit être soigneusement déterminé, avec deux conséquences pratiques.

- Si le nombre d'échantillons est inférieur au seuil, l'image paraît de mauvaise qualité. Il faudra alors ajouter de nouveaux échantillons, lorsqu'il est possible de reprendre le calcul en cours.
- Une fois le seuil dépassé, les calculs n'améliorent pas la qualité perçue. Une part du temps de calcul a été consommée sans gain visuel.

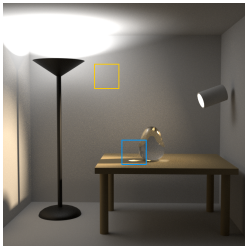


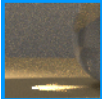


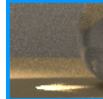
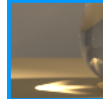
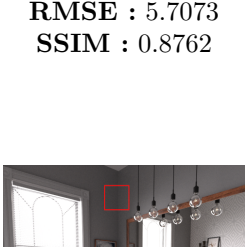
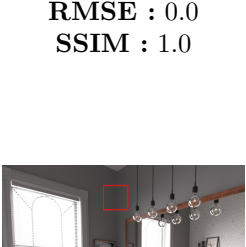






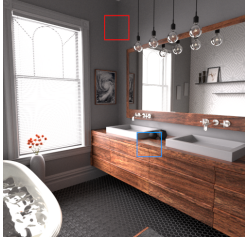






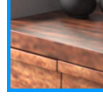
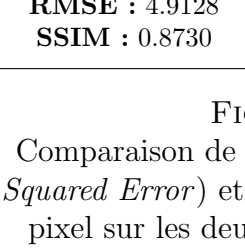
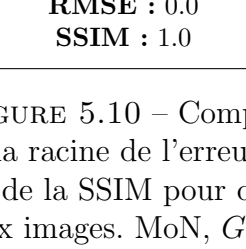






G -MoN	Reference	Mean	MoN	Jung et al.	G -MoN _b	G -MoN	Reference
							
RMSE : 5.7073 SSIM : 0.8762	RMSE : 0.0 SSIM : 1.0	RMSE : 7.3924 SSIM : 0.4637	RMSE : 8.8608 SSIM : 0.6622	RMSE : 7.3580 SSIM : 0.5123	RMSE : 8.4038 SSIM : 0.6567	RMSE : 8.1029 SSIM : 0.7032	RMSE : 0.0 SSIM : 1.0
							
RMSE : 4.9128 SSIM : 0.8730	RMSE : 0.0 SSIM : 1.0	RMSE : 6.2369 SSIM : 0.5397	RMSE : 4.0659 SSIM : 0.8391	RMSE : 6.2866 SSIM : 0.6478	RMSE : 6.0058 SSIM : 0.6927	RMSE : 3.7751 SSIM : 0.8670	RMSE : 0.0 SSIM : 1.0
							
RMSE : 4.9128 SSIM : 0.8730	RMSE : 0.0 SSIM : 1.0	RMSE : 5.7721 SSIM : 0.9101	RMSE : 5.6142 SSIM : 0.9009	RMSE : 5.7602 SSIM : 0.9054	RMSE : 5.9734 SSIM : 0.9070	RMSE : 5.2760 SSIM : 0.9206	RMSE : 0.0 SSIM : 1.0
							
RMSE : 4.9128 SSIM : 0.8730	RMSE : 0.0 SSIM : 1.0	RMSE : 5.8071 SSIM : 0.6621	RMSE : 7.0671 SSIM : 0.6942	RMSE : 5.7703 SSIM : 0.6658	RMSE : 6.8819 SSIM : 0.6921	RMSE : 5.8747 SSIM : 0.6923	RMSE : 0.0 SSIM : 1.0

FIGURE 5.10 – Comparaison des différents estimateurs

Comparaison de la racine de l'erreur quadratique moyenne (RMSE pour *Root Mean Squared Error*) et de la SSIM pour différents estimateurs avec 10 000 échantillons par pixel sur les deux images. MoN, G -MoN_b and G -MoN sont calculés avec $M = 21$.

4.2 Critère d'arrêt des algorithmes stochastiques

Note: La section suivante constitue un résumé du travail exposé dans le chapitre 4 de la thèse de Nawel Takouachet (2009) et publié dans (Takouachet et al., 2007).

Afin de rendre les méthodes de simulation d'éclairage utilisables automatiquement, il est nécessaire de posséder un critère qui permet de déterminer leur convergence perceptive. L'idée est ainsi de stopper les calculs dès que l'image atteint une qualité perceptive identique à la solution totalement convergée. Le problème majeur est qu'au moment du calcul, cette solution n'est pas disponible.

Dans ce cadre, la thèse de Nawel Takouachet (Takouachet, 2009) s'intéresse à différentes méthodes de calcul de l'écart colorimétrique qui peuvent être utilisées pour comparer la qualité des images durant le processus de rendu et donc décider de stopper la convergence à un certain seuil.

Toutefois, du fait de leurs approximations, ces méthodes ne prennent pas correctement en

compte les différents aspects de la sensibilité du système visuel humain. Une solution pour augmenter la précision des mesures de distance est de reproduire le comportement global du système visuel de l'observateur via un modèle de vision complet comme le VDP : *Visible Difference Predictor* (Daly, 1992) ou le VDM : *Visual Difference Metric* (Bolin & Meyer, 1999).

Deux approches basées sur le VDP ont été proposées pour définir un critère d'arrêt perceptif : la *comparaison successive* et la *comparaison inverse*. Dans le cadre de la *comparaison successive*, on confronte l'image obtenue avec une version précédente réalisée avec moins d'échantillons par pixel. Cette comparaison représente en quelque sorte l'amélioration produite à un moment donné du calcul. L'idée de la *comparaison inverse* est de confronter l'image à la toute première image produite : celle qui comporte le moins d'échantillons par pixels. Cette deuxième comparaison mesure l'amélioration depuis la première image qui peut-être considérée comme une estimation — de mauvaise qualité — de l'image finale. Cette dernière donne les meilleurs seuils perceptifs en matière de qualité de l'image finale et elle est plus précise que l'approche de Myszkowski (Myszkowski, 1998 ; Volevich et al., 2000). Tous les résultats obtenus ont été validés par comparaison avec des résultats expérimentaux.

Néanmoins, le VDP est un modèle complexe qui tente de reproduire le comportement général de comparaison et de sélection du système visuel humain. Son objectif est de détecter tous types d'erreurs visibles comme le bruit, la modification de la géométrie, de l'intensité ou de la couleur. Dans le cas du rendu, l'erreur perceptive est plus spécifiquement liée à un artefact visuel bien spécifique. L'erreur visible dans les méthodes d'illumination globale de type Monte-Carlo comme le *Path Tracing* est provoquée par un bruit haute fréquence. Il serait donc plus simple et plus efficace d'utiliser un modèle spécifique au type de l'erreur perceptive, afin d'améliorer le critère d'arrêt des algorithmes stochastiques.

4.3 Critère d'arrêt basé sur un apprentissage supervisé

Note: Le lecteur souhaitant davantage de précisions sur cette section se reportera aux chapitres 5 et 6 de la thèse de Nawel Takouachet (2009) et à (Takouachet et al., 2017).

Une approche proposée dans la thèse de Nawel Takouachet est d'exploiter des méthodes d'apprentissage supervisés pour décider d'un critère d'arrêt lors du rendu. Ce genre de méthode nécessite une connaissance, une expertise, qu'il est nécessaire d'obtenir pour qu'un tel modèle puisse apprendre.

4.3.1 Constitution d'une base d'images

Une première base d'images a été constituée ; elle comportait 16 scènes calculées en résolution 512×512^7 en utilisant un algorithme de *Path Tracing With Next Event Estimation*⁸. L'image de référence était celle calculée avec 10 000 échantillons par pixel et chaque image calculée a été stockée, par pas de 100 échantillons.

Afin de pouvoir mieux explorer l'effet perceptif, chaque image a elle même été découpée en zones. L'objectif était double :

7. la résolution de 512×512 a été choisie initialement pour permettre d'afficher côte à côte deux images (référence et cible) dans la résolution minimale XGA soit 1024×768

8. Le rendu a été calculé en utilisant le moteur IGLOO développé par Christophe Renaud et Christophe Cassagnabère (Cassagnabère, 2007)



FIGURE 5.11 – Exemple d'interface de recueil des données.

1. permettre le recueil de la qualité perceptive dans des zones relativement fines de l'image tout en conservant un temps d'expérimentation acceptable. Plus les zones seront petites, plus le recueil des données sera coûteux en temps.
2. préparer les données en vue d'un apprentissage automatique en limitant la taille des échantillons d'images, sachant que la dimension de l'image est un facteur fortement limitant pour la plupart des algorithmes de classification. On parle de « *malédiction de la grande dimension* ».

Un compromis a été de découper les images en 16 zones de 128×128 pixels.

4.3.2 Procédure de recueil des seuils perceptifs

Les données de 33 sujets ont été enregistrées — essentiellement des étudiants de première et deuxième année de DUT Informatique.

Afin de rester au plus proche du processus de calcul des images, la procédure suivante est proposée.

1. Une image de test est constituée avec au départ, le plus faible nombre d'échantillons. Cette image est présentée avec en référence l'image calculée avec le plus grand nombre d'échantillons.
2. L'utilisateur peut choisir d'allouer un nombre plus important d'échantillons par pixel dans chaque zone de l'image de test en choisissant l'endroit où il perçoit du bruit. Il pointe ce lieu par un clic souris. Dans ce cas, le nombre d'échantillons est augmenté dans l'ensemble de la zone correspondante (voir figure 5.11).
3. L'utilisateur réalise cette tâche jusqu'à ce qu'il considère qu'il a égalisé la qualité de l'image de test avec l'image de référence. Il valide ce choix et passe à l'image suivante.

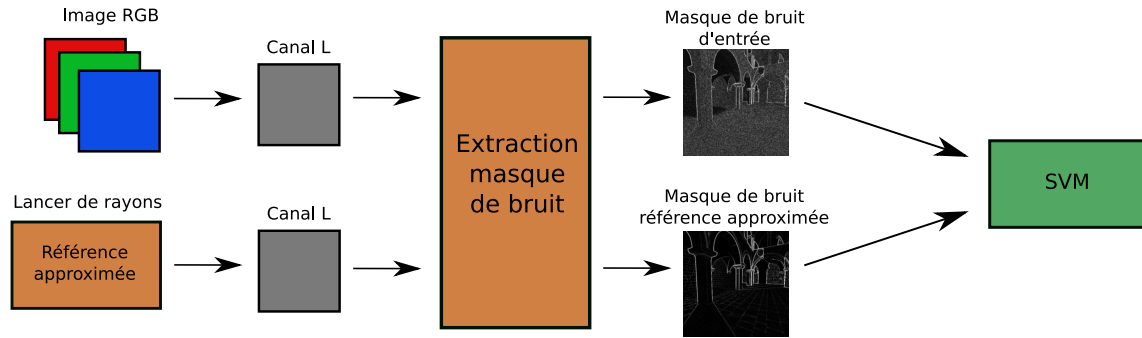


FIGURE 5.12 – Méthode pour l'apprentissage des seuils

La luminance est extraite de l'image calculée d'une part et de l'estimation de l'image de référence, d'autre part. Dans les deux cas, le masque de bruit est extrait afin d'obtenir les données d'entrée. Ce sont ces données qui servent d'exemples d'apprentissage.

On enregistre alors la valeurs S_z^a qui correspond au seuil choisi pour le sujet a dans la zone z . Afin de garantir que la zone sera perçue comme non bruitée pour la plus grande partie des observateurs, on retient une valeur qui intègre la moyenne des utilisateur pour une zone ainsi que l'écart-type de la distribution.

La valeur empirique de $S_z = \bar{m}_z + 1,1645 \times \sigma_z$ a été retenue pour chaque sous-image z , avec \bar{m}_z et σ_z respectivement la moyenne et l'écart-type du nombre d'échantillons utilisés dans la zone z . Cette valeur correspond à la précision nécessaire pour que 95% des sujets ne détectent plus le bruit.

4.3.3 Apprentissage du bruit par méthode supervisée

L'objectif initial était de proposer l'ensemble des données à un algorithme d'apprentissage supervisé (de type SVM) en utilisant les recueils expérimentaux.

Pour chaque image et sous image, on l'attribue à la classe :

- **bruitée** : si la valeur d'échantillonnage est inférieure au seuil S de la zone considérée ;
- **non bruitée**, dans le cas contraire.

L'apprentissage des données est réalisé avec la bibliothèque SVM^{light}⁹.

Une première tentative d'apprentissage a été réalisée sur l'ensemble des sous-images de la base mais ne produit pas les résultats souhaités. Après analyse, il apparaît que le modèle a produit une classification des images qui repose davantage sur des facteurs de luminance ou de géométrie des images.

Pour résoudre ce problème, on fournit d'une part au modèle une estimation de l'image finale, calculée avec une méthode peu coûteuse, et, d'autre part, il est nécessaire d'extraire l'information haute fréquence de l'image plus spécifiquement. Ceci est réalisé par le calcul d'un masque de bruit (Russ, 2016). Alors que l'utilisation générale d'un masque de bruit consiste à le soustraire de l'image pour la débruiter, dans le cas présent, il s'agit de l'utiliser pour extraire l'information visuelle haute fréquence susceptible d'être perçue comme du bruit. Les données expérimentales vont alors permettre à l'algorithme de faire la différence entre ce qui relève effectivement du bruit de Monte-Carlo du reste des informations de haute fréquence telles les variations dans la géométrie, l'éclairage ou les textures (figure 5.12).

9. <https://www.cs.cornell.edu/people/tj/svm.light/>

Apprentissage			Test		
scène	exp.	modèle	scène	exp.	modèle
OCULISTE	3278	3287	BUREAU	3025	3012
CUBE	2344	2300	BUREAU	2450	2481
BAR-V1	3234	3181	BAR-V2	2850	2893
BOULANGER	2215	2187	CLASSE	2255	2300
QUINCAILLIER	2385	2381	DRAPIER	2767	2737
SPONZA	2900	2862	EPICIER	3246	3162

TABLE 5.1 – Résultats après apprentissage.

Résultat moyen de l'estimation du nombre d'échantillons nécessaire pour obtenir une image de qualité indiscernable de l'image de référence pour 95% des utilisateurs. La colonne de gauche concerne les données d'apprentissage et la colonne de droite les données de test.

4.3.4 Évaluation de l'apprentissage

Pour cette partie, l'apprentissage est réalisé sur un sous-ensemble de six images (et leurs 16 blocs respectifs) jusqu'à obtenir un bon ajustement du modèle aux données. Le modèle est alors confronté aux autres images qui n'ont pas été utilisées pour l'apprentissage. L'ajustement du modèle aux données reste très bon comme le montre le tableau 5.1.

Les résultats qui sont montrés ici sont une présentation synthétique, c'est pourquoi la moyenne globale sur l'image est représentée. Dans la mesure où les prédictions sont réalisées par zone, on peut également vérifier que l'ajustement est également de bonne qualité à l'intérieur de l'image comme montré par exemple sur la scène sponza en figure 5.13.

Enfin, une phase de validation perceptive est réalisée en proposant les images qui utilisent les seuils par zones confrontés à l'image de référence.

Bien qu'encourageante, cette approche possède deux limites principales : la première réside dans la nécessité de disposer d'une approximation de l'image finale, qui peut être difficile à obtenir lorsque des effets lumineux complexes sont présents. La seconde est liée aux limites des scènes présentes dans la base de données constituée. Ces dernières sont à la fois géométriquement et photométriquement trop simples.

4.4 Constitution d'une nouvelle base d'images

Afin d'actualiser la base, de nouvelles images ont été calculées dans le cadre de la thèse de Jérôme Buisine.

L'ensemble de données proposé est composé de 40 points de vue d'images photoréalistes avec différents niveaux d'échantillons pour chacune. Chaque image a une taille de 800×800 pixels¹⁰ avec un échantillonnage de 20 à 10 000 échantillons. Le moteur de rendu *pbrt* (version 3) a été utilisé pour générer ces images (Pharr et al., 2016).

Les images sont divisées en 16 zones de 200×200 pixels de la même manière que pour l'approche de Nawel Takouachet ; elles sont produites à partir de scènes variées afin de contenir

10. La résolution de 800×800 pixels a été cette fois choisie dans la mesure où les périphériques en résolution minimum Full HD (1920×1080 pixels) se sont généralisés, ce qui permet l'affichage côte à côte de deux images



2671 / 2600	2006 / 1900	3068 / 3100	3467 / 3400
3554 / 3500	2102 / 2100	2239 / 2100	3104 / 3100
3537 / 3500	2734 / 2700	2829 / 2800	2720 / 2700
4058 / 4100	3535 / 3500	2324 / 2300	2462 / 2400

FIGURE 5.13 – Exemple d'image

Exemple d'une scène utilisée dans la première base d'image. Les données représentent d'une part les valeurs observées pour lesquelles 95% des observateurs ne perçoivent pas de différence avec l'image de référence et d'autre part la sortie du modèle après apprentissage.

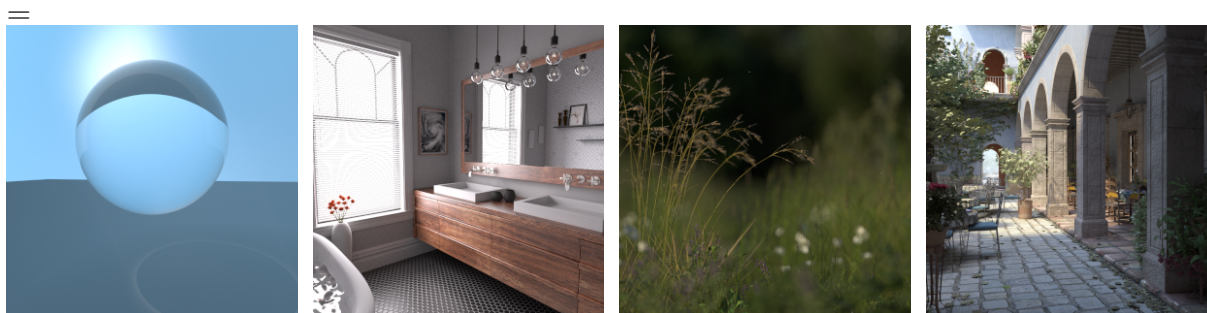


FIGURE 5.14 – Exemples d'images de la base de données.

des scènes d'intérieur et d'extérieur couvrant une large panel de matériaux et de géométries. (figure 5.14).

La base d'images ainsi que les seuils mesurés ont été mis à disposition. Jérôme Buisine a choisi la licence Creative Common pour publier ses bases : <https://zenodo.org/record/4964303>.

4.5 Modèle de perception du bruit

Note: Cette section résume la démarche proposée dans le chapitre 7 de la thèse de Buisine (2021) et publiée par Buisine, Bigand, Synave, Delepouille et Renaud, 2021.

Dans le contexte de la thèse de Jérôme Buisine nous avons été confrontés au problème de la dimensionnalité et celui lié à la grande quantité de données à traiter par les méthodes d'apprentissage qui s'appuient sur SVM.

L'un des problèmes rencontrés concerne la dimension des données : le passage de blocs images 128×128 à 200×200 n'a pas permis d'appliquer directement la méthode présentée dans la section 4.2. De plus, cette méthode nécessite d'utiliser une image de référence qui n'est pas forcément disponible. C'est pourquoi il a été nécessaire de développer une nouvelle méthode capable de prendre une décision sur la perceptibilité du bruit uniquement à partir des images disponibles.

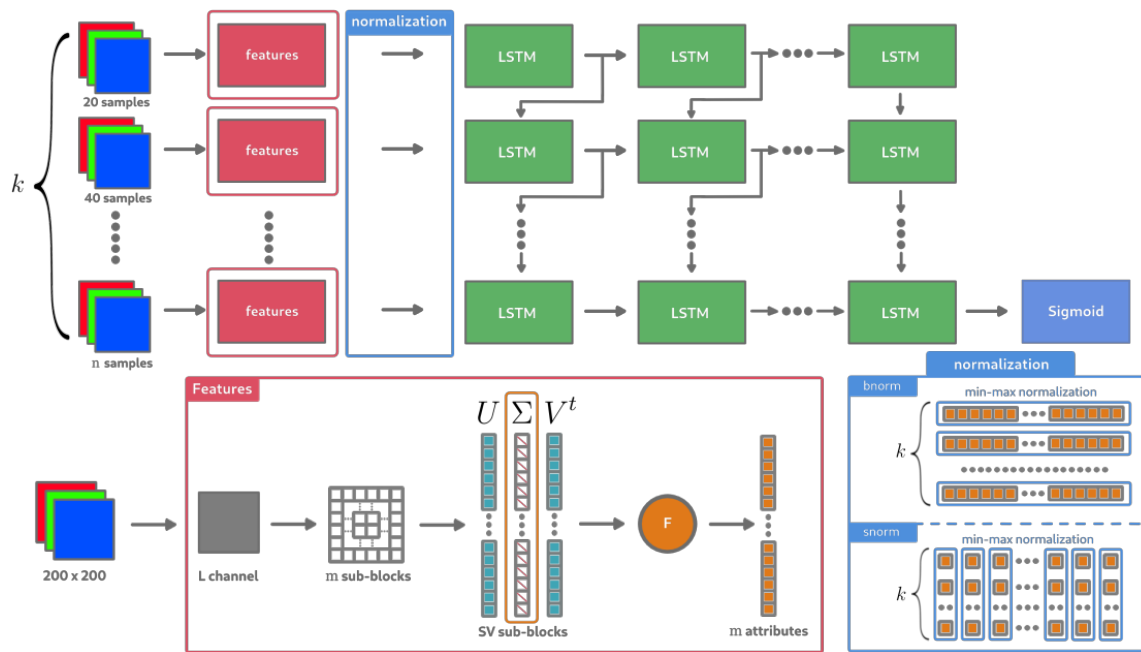


FIGURE 5.15 – L'architecture du réseau RNN

Elle comprend trois parties : le réseau RNN lui-même figuré par les cellules vertes LSTM (*Long Short-Term Memory*). A l'origine, sur la séquence des $k - 1$ images précédentes et l'image courante, une phase d'extraction des caractéristiques (*features* encadrée en rouge) permet de réduire la dimension du vecteur d'entrée. Les valeurs d'entrée doivent également subir une étape de normalisation encadrée en bleu sur la figure qui peut-être effectuée de plusieurs façon : bnorm ou snorm.

Le principe est de développer un modèle de perception du bruit de Monte Carlo basé sur la méthode d'apprentissage profond. Le modèle proposé exploite un réseau de neurones récurrents et une fenêtre glissante d'images.

Les attributs locaux qui caractérisent le bruit sont cette fois extraits par le biais de la *SVD-Entropy* car elle permet de collecter efficacement les caractéristiques de bruit lors de la génération d'images (Alter et al., 2000).

Pour effectuer l'apprentissage puis la décision de visibilité du bruit, un réseau de neurones récurrents (RNN pour *Recurrent Neural Network*) est introduit. L'utilisation de ce RNN permet de prendre en compte la temporalité¹¹ du bruit généré dans les images — au travers d'une fenêtre glissante d'images — et d'en interpréter l'évolution de ses différences.

L'architecture RNN repose sur un *pipeline* dont voici le principe (également illustré par la figure 5.15).

1. Les images produites par le moteur de rendu sont utilisées selon une fenêtre glissante qui inclut les k dernières images.
2. Une première phase d'extraction des caractéristiques est réalisée (partie « Features » encadrée en rouge dans le diagramme). Elle consiste à utiliser le canal de luminance des blocs 200×200 et à les décomposer en m sous-blocs. Sur la base de la SVD-entropie, un

11. Dans ce contexte, la « temporalité » renvoie à la variation au cours du calcul d'images intermédiaires avec un nombre croissant d'échantillons par pixel. Elle ne doit pas être confondue avec le temps dans le calcul d'une séquence.

vecteur de caractéristiques est obtenu. Il est constitué par les M décompositions SVD des différents sous-blocs. Sur ce vecteur, on calcule l'entropie (représentée ici par l'opérateur F). On obtient donc en sortie de ce processus un vecteur de m attributs.

3. Suit une phase de normalisation (encadré bleu de la figure). Elle peut prendre plusieurs formes suivant les données qui ont servies à la normalisation.
4. Enfin ces valeurs permettent l'apprentissage d'un RNN (figuré en vert par les cellules LSTM pour *Long Short-Term Memory*).

4.5.1 Paramètres testés et résultats

Une étude empirique de l'efficacité de l'approche a été menée, en particulier pour les paramètres :

- la taille k de la séquence d'images prises en compte, variant de 3 à 10 ;
- le nombre m de sous-blocs, avec $m \in [4, 25, 100, 400]$. Ce qui veut dire que la taille des sous-blocs est respectivement de taille $100 \times 100 = 10\,000$ pixels, $40 \times 40 = 1\,600$ pixels, $20 \times 20 = 400$ et $10 \times 10 = 100$ pixels.

Il en ressort que les meilleurs résultats sont observés pour fenêtre relativement grande d'images successives (5 à 10) et pour des tailles de sous-blocs plutôt petites (20×20).

Le modèle de perception de bruit fournit des prédictions intéressantes aussi bien sur les blocs appris que non appris (figure 5.16). Ainsi, les critères d'arrêt permettent d'obtenir des images où le bruit est fortement réduit. Toutefois, le modèle fournit un seuil d'arrêt trop précoce dans quelques rares cas, ce qui peut amener à une présence de bruit encore perceptible.

4.6 Génération automatique de caractéristiques de bruit

Note: Cette section résume succinctement la démarche proposée dans le chapitre 8 de la thèse de Jérôme Buisine (2021) et publiée par Buisine, Teytaud, Delepoulle et Renaud, 2021.

Une deuxième approche à base d'apprentissage profond a été proposée afin d'identifier les caractéristiques représentatives du bruit dans les méthodes de Monte-Carlo. Celle-ci repose sur une extension de la notion de masque de bruit.

En utilisant également l'apprentissage profond, une architecture composée de trois modèles d'apprentissage est proposée sous la forme de réseau GGN (*Guided-Generative Network*) inspiré du GAN (*Generative Adversarial Network*). Chaque réseau est responsable d'un aspect spécifique comme illustré en figure 5.17.

- Le premier permet de débruiter une fenêtre glissante d'images.
- Le second permet de localiser le bruit par l'intermédiaire d'une carte de caractéristiques générée à partir de la fenêtre glissante.
- Le troisième interprète la différence de bruit entre une fenêtre glissante d'images bruitées, et une autre fenêtre débruitée par le premier modèle, en analysant les cartes de caractéristiques générées respectivement pour chacune des fenêtres glissantes.

Même s'ils doivent être confirmés, les résultats du modèle GGN sont encourageants. Leur ajustement aux données semble bon. L'un des avantages de l'approche est qu'il n'exige aucune connaissance explicite sur la nature de bruit, mais qu'il l'apprend au fur et à mesure sur les données. Ce type d'architecture proposée pourrait être utilisé pour tous types de bruit à traiter.

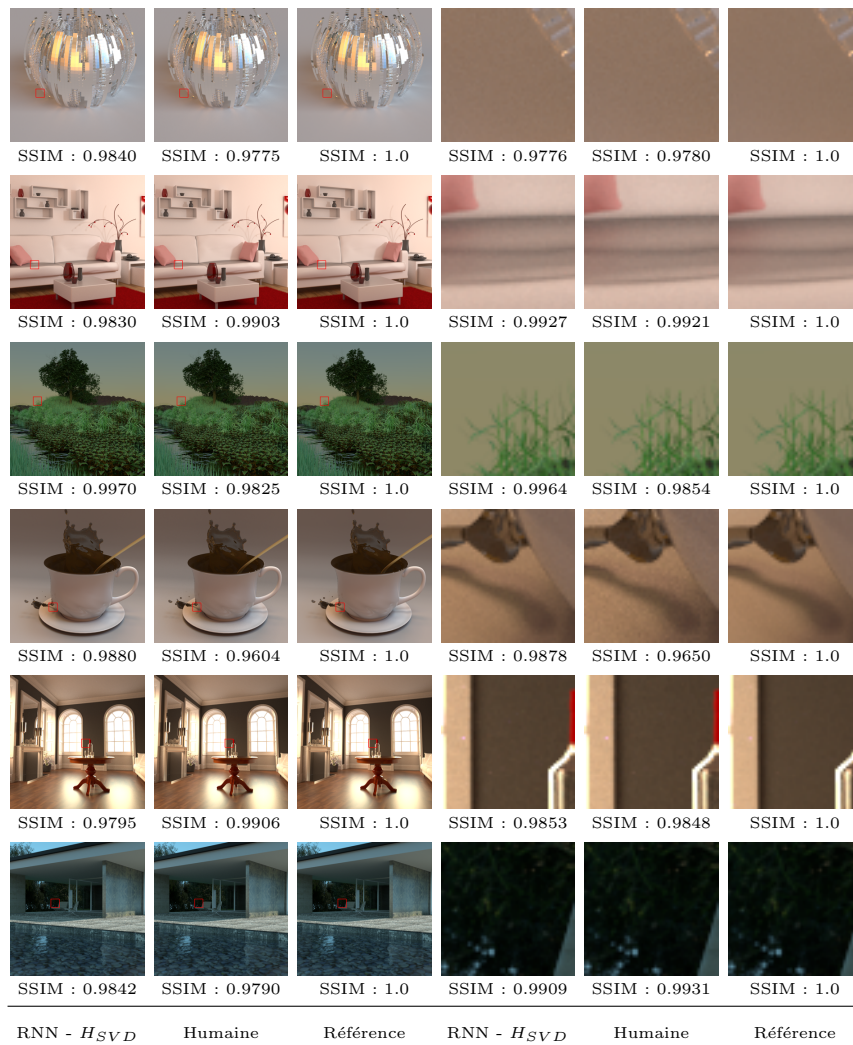


FIGURE 5.16 – Résultats de la sortie du RNN

Elle est ici comparée aux images reconstruites à partir des seuils humains d'une part et de l'image de référence d'autre part

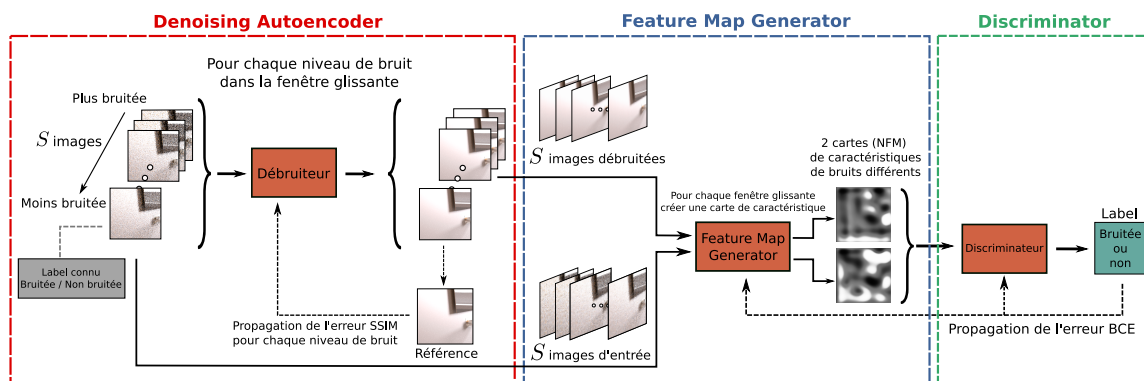


FIGURE 5.17 – Architecture de modèle GGN

Une fenêtre glissante d'images de différents niveaux de bruit et de taille S est donnée en entrée au modèle « *Denoising Autoencoder* » pour le débruitage afin d'obtenir une fenêtre glissante d'images de référence approximées de taille S . L'une après l'autre, les deux fenêtres coulissantes d'images, celle d'entrée et celle de référence approximée, sont envoyées au modèle « *Feature Map Generator* » pour obtenir pour chacune une carte de caractéristiques de bruit (NFM). Enfin, les deux NFM obtenues sont transmises au « *Discriminator* » pour évaluer si la dernière image de la fenêtre glissante d'entrée est considérée comme étant toujours bruitée ou non. La propagation de l'erreur est alors possible grâce à l'étiquette connue (bruitée / non bruitée) de la dernière image de la fenêtre glissante d'entrée.

4.7 Comparaison de modèles

Note: Cette section résume succinctement la démarche et les résultats détaillés dans le chapitre 10 de la thèse de Jérôme Buisine (2021) et soumis à pour publication dans le livre « *Deep Learning Applications Volume 4* »

Cette section présente le problème de la perception du bruit dans le cadre de l'utilisation des méthodes de Monte-Carlo pour la synthèse d'images photoréalistes. Il ressort de ces études que ce bruit peut être localisé en utilisant des protocoles expérimentaux. Le recueil des seuils permet de fournir des exemples d'apprentissage utilisables par plusieurs approches. Dans la mesure où plusieurs méthodes ont été développées dans le cadre de la thèse de Jérôme Buisine, il devient possible de les comparer entre elles. Avec les mêmes données, il est possible de comparer la capacité des méthodes à prédire le nombre d'échantillons nécessaires pour produire une image indiscernable d'une image de référence.

Il est alors possible d'utiliser les données en sortie des modèles pour produire des images reconstruites. Cette démarche est présentée par la figure 5.18. A partir du découpage en blocs, les seuils humain sont mesurés. A partir de ces seuils, une image reconstruite peut être composée et comparée à l'image de référence. L'image peut être recomposée à partir des méthodes présentées dans les sections précédentes comme la méthode basée sur la SVD-Entropie.

La figure 5.19 montre la répartition des erreurs pour les deux modèles. On note une dispersion des erreurs plus faible pour le modèle GGN et cela est confirmé par le meilleur ajustement observé entre ce modèle et les seuils observés : la corrélation linéaire dans le cas du GGN est de 0,902

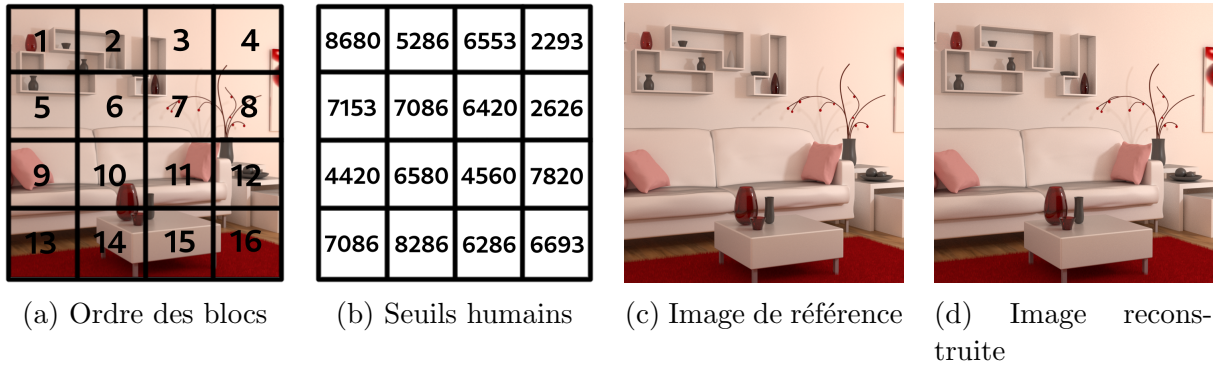


FIGURE 5.18 – Exemple d'utilisation des seuils recueillis

Les seuils sont enregistrés sur des blocs d'image (voir 4.4, page 102). (a) Les 16 blocs de l'image, (b) le seuil humain pour chaque bloc (en nombre d'échantillons par pixel) (c) l'image de référence utilisant 10 000 échantillons par pixels et (d) l'image reconstruite à partir des seuils observés.

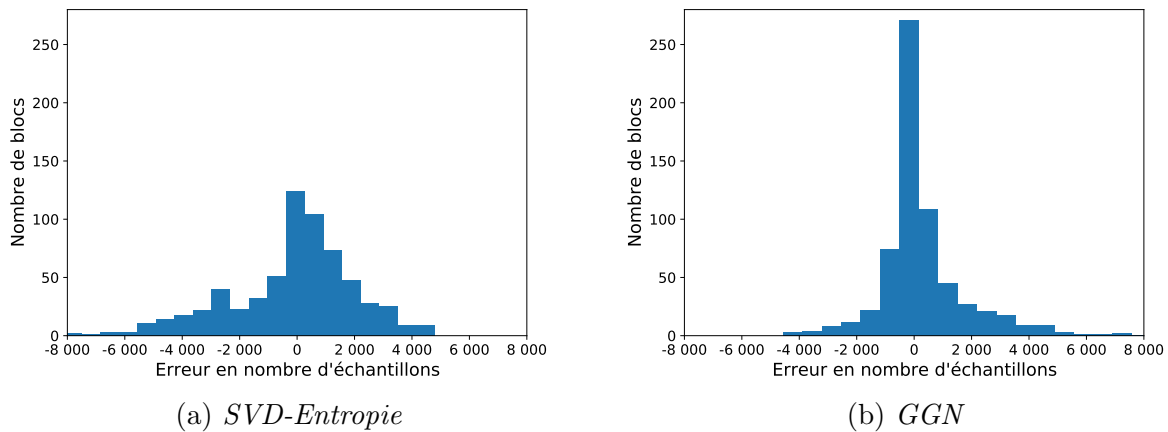


FIGURE 5.19 – Écart aux seuils perceptifs enregistrés sur les 16 blocs des 40 images

contre 0.718 pour le GGN comme présenté en figure 5.20.

Contrairement à l'approche qui repose sur la SVD-Entropie, le modèle GGN automatise la partie de recherche des caractéristiques extraites pour réaliser l'apprentissage. Les résultats du modèle observés suggèrent que cette extraction produit un ajustement de meilleur qualité. De plus, dans la mesure où aucune hypothèse n'est nécessaire quand au type de bruit présent dans l'image, on peut penser que cette méthode est également plus robuste.

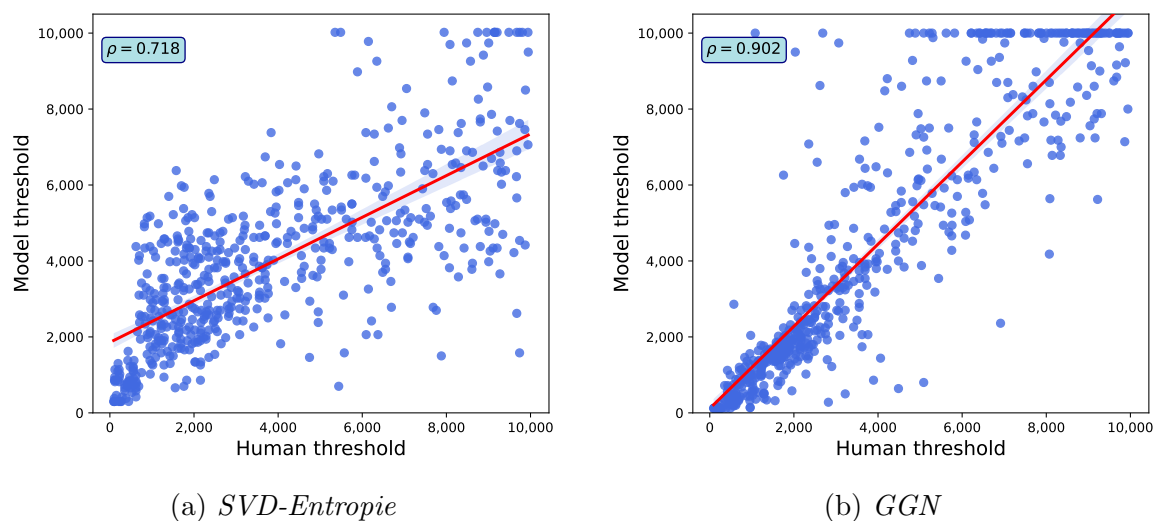


FIGURE 5.20 – Seuils prédits par le modèle comparés au seuils observés.

En Résumé

Les méthodes de Monte-Carlo pour la production d'images sont efficaces pour produire des images photo-réalistes qui seront perçues comme des scènes réelles. Leur production s'accompagne d'un bruit de haute fréquence, si le nombre d'échantillons n'est pas suffisant. Ce bruit peut-être mesuré par des méthodes psychophysiques adaptées et permet de fournir des exemples pour qu'un modèle d'apprentissage supervisé puisse prédire s'il est perceptible ou non pour un observateur. Ceci permet de fournir un critère d'arrêt des calculs.

Conclusion provisoire, perspectives et futurs travaux

1 De l'impossible conclusion

L'ambition de ce document est de proposer un tour d'horizon de l'état des connaissances et des techniques relatives à la perception des images numériques au début de la décennie 2020. Ce projet est par essence transdisciplinaire puisqu'il mobilise des connaissances issues de nombreux domaines scientifiques et technologiques. Comme la première partie a tenté de le montrer, la perception de l'image fait intervenir une large partie de la palette des sciences : l'optique, la psycho-physiologie, la médecine, la psycho-physique et plus récemment les sciences du numérique. La fabrication de l'image par les artistes n'y est pas non plus étrangère car ils ont souvent été les premiers à expérimenter, à utiliser ou à manipuler les liens si particuliers entre l'œil et l'image. C'est logiquement que la science neuve de l'imagerie numérique a puisé dans ce capital de connaissances de la relation observateur-image. L'étude de cette relation, à des niveaux divers, a traversé des siècles de recherches artistiques, scientifiques et techniques.

Dans ce tour d'horizon, plusieurs chapitres supplémentaires auraient mérité leur place : le problème de la perception des matériaux, celui de la perception des couleurs ou de la dynamique de restitution des luminances des périphériques auraient, par exemple, pu faire l'objet de développements supplémentaires.

Dans la seconde partie, nous avons tenté de montrer la validité de cette approche transdisciplinaire sur un problème précis : la réalisation d'images photoréalistes basées sur la résolution de l'équation du rendu par des méthodes de Monte-Carlo. Nous montrons comment la perception visuelle peut être prise en compte dans la production de ces images. Mais la thèse centrale de ce mémoire est qu'il existe une véritable intrication de la perception visuelle et de l'image. C'est l'idée qui en constitue en quelque sorte le fil rouge. C'est également cette idée qui a permis de définir un projet scientifique qui a été esquissé au travers des collaborations entre le LISIC, laboratoire d'informatique et SCALab, laboratoire de psychologie. Plus particulièrement, le problème de la perception du bruit visuel produit par des méthodes de synthèse d'images photo-réalistes a pu faire l'objet d'une étude réunissant des chercheurs de ces deux communautés impliqués dans un projet ANR commun.

Comme nous avons tenté de le démontrer, l'image numérique est forcément un domaine à l'interface de la technologie et des sciences humaines. Même s'il est souvent difficile à appréhender dans son ensemble, l'image, les mécanismes de sa production, de son affichage et l'observateur qui la regarde forment un système, ou pour être plus précis, un ensemble de sous-systèmes qui sont en permanentes interactions les uns avec les autres.

C'est cette démarche qui vise à une étude intégrée des algorithmes de production d'image, des périphériques d'affichage et de la perception humaine qu'il nous semble nécessaire de poursuivre.

Ce projet est ambitieux et en tirer une conclusion définitive est utopique tant les ques-

tions et problèmes restent importants. C'est plutôt un ensemble de pistes de travail qui ont été entrouvertes et quelques-unes de ces pistes sont proposées au lecteur pour terminer.

2 Précisions sur la perception du bruit

Une partie du travail à venir consiste à mieux qualifier le bruit perceptible. Les travaux à propos de la perception du bruit dans le cadre de l'utilisation de méthodes de Monte Carlo doivent être approfondis et généralisés. La base d'image doit notamment être complétée pour offrir davantage de données sur de nouvelles scènes afin de préciser les mesures et améliorer la généralisation des méthodes d'apprentissage.

Il apparaît également que de nombreux facteurs sont importants à prendre en considération. Parmi ceux-là, deux axes méritent une investigation complémentaire car ils constituent des verrous à la fois scientifiques et techniques pour les méthodes que nous mettons en place : une partie concerne la prise en compte du bruit dans le processus de calcul et une seconde partie concerne les périphériques de rendu qui constituent l'interface entre le système perceptif et l'image numérique.

2.1 Algorithmique des artefacts visuels

2.1.1 Une meilleure prise en compte du bruit

Les résultats que nous avons obtenus sur la détection, la représentation, la modélisation et l'impact du bruit perceptif montrent qu'il s'agit d'un artefact bien spécifique qui nécessite de développer des méthodes et des techniques adaptées pour le traiter de façon efficace.

Une étape nécessaire sera de poursuivre cette recherche au travers de plusieurs études.

2.1.1.1 Les procédures de recueil des données

Le recueil de données complémentaires, avec des scènes nouvelles, dans des conditions différentes est un travail important. Il consistera à mieux comprendre les facteurs perceptifs impliqués dans la perception du bruit. Ceci implique de réaliser non seulement des mesures en laboratoire, en utilisant les dernières méthodes de la psycho-physique dans des situations contrôlées, mais aussi de les compléter par des études plus proches des situations de visionnage des images. Les résultats de la thèse de Myrodia (2021) sont encourageants quant à l'utilisation de cette méthode. Il a en effet été constaté que les données mesurées avec des participants qui répondent à des études en ligne dans un cadre peu contrôlé sont globalement cohérentes avec celles réalisées dans le cadre d'un laboratoire. Aussi, les études en ligne peuvent aider à compléter des études en environnement contrôlé car elles permettent principalement d'accroître le nombre de participants à celles-ci.

2.1.1.2 Mieux quantifier pour mieux estimer

Les travaux sur les estimateurs robustes et efficaces dans le cadre de l'estimation de Monte-Carlo ont permis de montrer leur intérêt, notamment par l'utilisation de la Médiane des Moyennes (Buisine, Delepoulle & Renaud, 2021a). Ces travaux mériteraient d'être approfondis, comparés par exemple à des méthodes de rejet statistique ou enrichis par d'autres méthodes d'estimations robustes comme des estimateurs MoN bayésiens, l'estimateur de la médiane des moyennes invariant par permutation ou l'estimateur MoN par tournoi.

2.1.1.3 Intégrer le débruitage

Depuis plusieurs années, l'utilisation d'algorithmes de débruitage tend à se généraliser lors de l'utilisation de méthodes de tracé de chemins. Naïvement, le problème du bruit dans l'image pourrait paraître résolu par l'emploi de ces méthodes. Il est clair que leur utilisation élimine une bonne partie du bruit visuel de haute fréquence et permettent effectivement une accélération de la convergence visuelle. Cependant la problématique que nous avons étudiée reste pertinente pour deux raisons. D'une part ces méthodes apportent également de nouveaux artefacts visuels qui mériteraient d'être étudiés. Les méthodes de débruitage reposent généralement sur une forme de filtrage de l'image qui est très efficace pour réduire le bruit de haute fréquence. Cependant l'erreur globale sur l'image reste importante au début du calcul même si elle est moins perceptible. Il faudra donc étudier comment cette forme de « dilution de l'erreur » se traduit. A quel point les déformations induites sont perçues, dans quelles conditions et si elles perturbent ou non la perception. D'autre part, même si la convergence est plus rapide, cela ne donne pas davantage de critères d'arrêt des calculs. Ceci pourrait donc être à l'origine de nouvelles études.

2.1.2 Le problème des séquences d'images

Pour des raisons pratiques et méthodologiques, beaucoup d'études s'intéressent à la production d'images fixes ; ceci inclut également nos études actuellement. Passer à des séquences d'images représente pourtant un enjeu majeur pour plusieurs raisons.

Quelque soit le problème étudié, le calcul d'une séquence d'images représente en pratique un coût sans commune mesure avec celui d'une image seule. L'étude de la perception du bruit, comme d'autres recherches, nécessite la constitution de base de données qu'il faudra constituer en tenant compte de ce problème de facteur d'échelle. La production de séquences animées requiert de calculer chacune des images. Ce calcul est souvent réalisé de façon indépendante. Pour donner une idée du facteur, un minute d'animation représente 1500 images pour une fréquence d'affichage de 25 hertz. Ceci accentue d'autant tous les problèmes liés à la synthèse d'images réalistes. L'image animée est donc un verrou supplémentaire pour ces méthodes. Du fait de la complexité des mécanismes du transfert de la lumière, les méthodes photo-réalistes sont nécessairement gourmandes en ressources, ce qui les a longtemps cantonné à la production d'images fixes et limite leur utilisation pour produire des séquences.

Une seconde raison est liée à la cohérence des méthodes de calcul entre images. Dans la mesure où chaque image est traitée de façon indépendante, toute étape du calcul risque de faire apparaître des discontinuités temporelles. Dès lors qu'une partie de ce calcul utilise des algorithmes stochastiques, il n'est pas garanti que la solution produite à un moment donné soit temporellement stable. Par exemple, l'algorithme de *Metropolis light transport* utilise des variations aléatoires de chemins de lumière (Veach & Guibas, 1997). Ceci permet généralement une convergence plus rapide cependant, il est possible que deux images de la même scène soit légèrement différentes à une étape donnée du rendu. Ceci se traduit inévitablement par des variations lors de la réalisation d'une séquence. Dans un autre domaine, l'algorithme du *Lightcut* permet d'estimer l'éclairage direct (Walter et al., 2005). Cet algorithme a tendance à sous estimer l'éclairage dans certaines conditions. Ceci est très rarement pénalisant pour une image donnée. Cependant, dans le cas d'une séquence, ceci se traduit par un clignotement de l'image qui peut la rendre inutilisable. (Yuksel, 2019).

De plus ces discontinuité ont une implication perceptive très liée à notre système visuel qui présente une sensibilité forte aux variations. À de nombreux niveaux, les changements perceptifs jouent un rôle important. Cette sensibilité aux changements de l'image peut rendre particulièrement saillant le problème de discontinuité temporelle. Ces mécanismes, qui jouent par exemple un rôle important dans la détection du mouvement, rendent d'autant plus saillantes

de petites approximations de calcul. Par exemple, un *firefly* va se manifester dans une image sous la forme d'un clignotement. Or on sait qu'il existe dans les voies visuelles des cellules spécialisées dans l'apparition d'un signal lumineux ou cellules « ON », d'autres cellules les cellules « OFF » réagissent à une disparition. Une seule erreur dans l'estimation va entraîner une cascade de réactions qui augmentent fortement sa probabilité de détection. Il faut donc s'attendre à enregistrer des valeurs différentes pour les seuils dans ce cadre. Ce mécanisme s'applique vraisemblablement au bruit dans son ensemble. Un bruit non perceptible dans chacune des images individuelles d'une séquence risque donc de devenir perceptible dans la séquence du fait de ses variations temporelles.

2.2 Explorer les périphériques

2.3 L'impact de la plage de luminance

L'œil humain est sensible à une gamme de luminance bien plus étendue que celle des écrans. Le rapport entre la plus forte intensité qui peut-être perçue sans dommage oculaire et la plus faible enregistrée après adaptation à l'obscurité est de l'ordre de 10^{14} . Cette gamme de luminance est sans commune mesure avec celle des moniteurs standards dont la capacité de restitution est proche de 10^3 . De nouveaux périphériques désignés comme HDRI (*High Dynamique Range Imaging*) permettent de représenter des gammes de luminances suivant un rapport de 10^6 . La représentation de grandes gammes de luminances induit néanmoins des difficultés. En particulier, afin de préserver les écarts entre les échelles de luminances, il est nécessaire d'utiliser un codage adéquat. Le codage standard des images dont les couleurs sont codées sur 8 bits, s'avère trop limitée et il est donc nécessaire d'utiliser des codages dans lesquels chaque canal de couleur est représenté par au minimum 10 bits (HDR10). Ceci implique également que ce codage soit respecté dans l'ensemble de la chaîne qui va du calcul à l'affichage de l'image.

Dans le cadre du rendu d'image photoréaliste, la luminance est un facteur clé : des indices permettent de formuler l'hypothèse d'un lien entre la perceptibilité du bruit et la plage de luminance utilisée par le périphérique utilisé. Nous avons noté, par exemple que lorsqu'on transforme l'image dans un espace colorimétrique qui sépare la luminance, comme avec l'espace $L^*a^*b^*$ CIE 1976, il semble que ce soit essentiellement le canal L qui est affecté par la présence de bruit. Dans un premier temps, cette observation mériterait d'être confirmée en utilisant la base de données des images dont nous disposons. Par ailleurs, de nouvelles mesures peuvent être envisagées afin de mesurer spécifiquement l'impact de la luminance sur la perception du bruit.

Les périphériques HDRI disponibles au sein de notre équipe seront particulièrement intéressants pour étudier ce problème.

2.3.1 Autres périphériques de rendu

Des résultats préliminaires ont été enregistrés sur des périphériques de rendu stéréoscopiques et auto-stéréoscopiques. La perception du bruit dans ce cas semble déterminée par la méthode de calcul des images :

- dans une version de base, les deux images de vue stéréoscopique sont calculées de façon totalement indépendante ;
- dans une autre version, une partie des calculs sont réutilisés pour le calcul des deux points de vue qui correspondent aux deux yeux.

La différence perceptive est que dans le cas du calcul indépendant, le bruit semble « flotter » devant l'image à la manière d'un rideau alors que dans le cas du calcul cohérent, le bruit

semble davantage plaqué sur les objets à la manière d'une texture.

Ces résultats devront être confortés par des observations plus systématiques, par exemple en utilisant des casques de réalité virtuels qui s'avèrent intéressants pour recueillir des données (Huan, 2021) et mesurer l'effet de plusieurs variables. Il est par exemple envisageable de modifier la quantité globale d'échantillons utilisée pour les deux images afin d'étudier le rôle de la disparité sur la perception du bruit. De plus, l'utilisation de l'oculométrie dans le casque permet d'utiliser des niveaux de rendu différents en fonction des zones de projection sur la rétine et donc d'analyser le rôle de l'excentricité.

3 L'extension du domaine de la perception

3.1 Perception du réalisme

Un autre axe de recherche serait de conduire des expérimentations sur la perception du réalisme lui-même afin de mesurer à quel point le rendu modifie la perception d'une scène, sa mémorisation et les capacités à interagir. La possibilité de représenter la même scène avec des méthodes de rendu plus ou moins réalistes permettrait d'une part de mieux comprendre les processus impliqués dans la perception d'une scène réelle. Ceci pourrait conduire également à sélectionner la qualité nécessaire afin de produire une image de bonne qualité.

3.2 Généralisation de la méthode : une voie transdisciplinaire

La méthode utilisée ici peut également être transposée à de nombreux domaines de la production d'images numériques : pour des méthodes de production d'images, il est souvent crucial de prendre en compte les facteurs perceptifs. Ceci implique généralement de recueillir des données expérimentales, de les modéliser afin de pouvoir les utiliser dans la partie production de l'image. La démarche pourrait être appliquée à d'autres types d'artefacts visuels résultant d'autres méthodes. Par exemple, l'utilisation de cartes de photons ou d'images débruitées produisent des artefacts visuels très différents qui nécessitent sans doute d'autres méthodes de traitement ou une adaptation des méthodes proposées.

L'un des points importants que nous avons soulignés est l'intérêt d'une démarche transdisciplinaire. Cette démarche a toute sa place dans le domaine de l'image qui est par essence une production technologique mais aussi un objet perceptif et social. Si la production d'images met en jeu de plus en plus de traitements numériques, elle restera destinée à être vue par un observateur. Elle est également réalisée dans une intention spécifique afin de produire un effet perceptif, à évoquer des actions, à induire une émotion. De plus, cette image est produite dans un contexte particulier, à un moment donné de l'histoire. Cette démarche, comme nous l'avons montré, rejoint également le travail des artistes peintres qui ont appris à maîtriser ces techniques au cours des siècles d'évolution de la représentation picturale.

Bibliographie

- Águila-Carrasco, A. J. D., Marin-Franch, I., Bernal-Molina, P., Esteve-Taboada, J. J., Kruger, P. B., Montés-Micó, R. & López-Gil, N. (2017). Accommodation Responds to Optical Vergence and Not Defocus Blur Alone. *Investigative Ophthalmology & Visual Science*, 58(3), 1758. <https://doi.org/10.1167/iovs.16-21280>
- Alberti, L. ([1435] 1991). *On painting*. Penguin.
- Alter, O., Brown, P. O. & Botstein, D. (2000). Singular value decomposition for genome-wide expression data processing and modeling. *Proceedings of the National Academy of Sciences*, 97(18), 10101-10106. <https://doi.org/10.1073/pnas.97.18.10101>
- Appel, A. (1968a). Some techniques for shading machine renderings of solids. *AFIPS 1968 Spring Joint Computer Conference*, 32, 37-45.
- Appel, A. (1968b). Some techniques for shading machine renderings of solids. *Proceedings of the April 30–May 2, 1968, spring joint computer conference on - AFIPS '68 (Spring)*. <https://doi.org/10.1145/1468075.1468082>
- Arcizet, F. (2007). *Les indices monoculaires de la perception tridimensionnelle étude électrophysiologique de l'aire V4 du macaque vigile* (Theses). Université Paul Sabatier - Toulouse III. <https://tel.archives-ouvertes.fr/tel-00332039>
- Arden, G. B. & Jacobson, J. J. (1978). A simple grating test for contrast sensitivity : preliminary results indicate value in screening for glaucoma. *Investigative ophthalmology & visual science*, 17(1), 23-32.
- Asdown, I. (1993). Near-field photometry : a new approach. *Journal of the Illuminating Society*, 22, 163-180.
- Bailey, I. L. & Lovie, J. E. (1976). New Design Principles for Visual Acuity Letter Charts. *Optometry and Vision Science*, 53(11), 740-745. <https://doi.org/10.1097/00006324-197611000-00006>
- Banks, M. S. & Salapatek, P. (1976). Contrast sensitivity function of the infant visual system. *Vision Research*, 16(8), 867-III. [https://doi.org/10.1016/0042-6989\(76\)90147-4](https://doi.org/10.1016/0042-6989(76)90147-4)
- Bartolo, A., Coello, Y., Delepouille, S., Edwards, M., Endo, S. & Wing, A. (2009). Neurobiological basis of reachability judgement : an fMRI study. *NeuroImage*, 47, S171. [https://doi.org/10.1016/s1053-8119\(09\)71841-5](https://doi.org/10.1016/s1053-8119(09)71841-5)
- Bartolo, A., Coello, Y., Edwards, M. G., Delepouille, S., Endo, S. & Wing, A. M. (2014). Contribution of the motor system to the perception of reachable space : an fMRI study. *European Journal of Neuroscience*, 40(12), 3807-3817. <https://doi.org/10.1111/ejn.12742>
- Blinn, J. F. (1977). Models of light reflection for computer synthesized pictures. *SIG-GRAPH '77 Proceedings*, 11(2), 192-198.

- Bolin, M. & Meyer, G. (1999). Visual difference metric for realistic image synthesis [Proceedings of the 1999 Human Vision and Electronic Imaging IV ; Conference date : 25-01-1999 Through 28-01-1999]. *Proceedings of SPIE - The International Society for Optical Engineering*, 3644, 106-120.
- Bove, V. M. (1989). Discrete fourier transform based depth-from-focus. *Image Understanding and Machine Vision* *Image Understanding and Machine Vision*, 14, 118-121.
- Buisine, J. (2021). *Méthodes d'apprentissage automatique pour la prise en compte du bruit dans les images de synthèse* (thèse de doct.) [Thèse de doctorat dirigée par Renaud, Christophe et Delepouille, Samuel Informatique Littoral 2021]. Université du Littoral Côte d'Opale.
- Buisine, J., Bigand, A., Synave, R., Delepouille, S. & Renaud, C. (2021). Stopping Criterion during Rendering of Computer-Generated Images Based on SVD-Entropy. *Entropy*. <https://hal.archives-ouvertes.fr/hal-03100109>
- Buisine, J., Delepouille, S. & Renaud, C. (2021a). Fireflies removing in Monte Carlo rendering with adaptive Median of meanS. *Eurographics Symposium on Rendering (2021)*. <https://doi.org/10.2312/sr.20211296>
- Buisine, J., Delepouille, S. & Renaud, C. (2021b). Minimalist And Customisable Optimisation Package. *Journal of Open Source Software*, 6. <https://doi.org/10.21105/joss.02812>
- Buisine, J., Delepouille, S., Synave, R. & Renaud, C. (2021). Subjective human thresholds over computer generated images. <https://doi.org/10.5281/ZENODO.4531460>
- Buisine, J., Teytaud, F., Delepouille, S. & Renaud, C. (2021). Guided-Generative Network for noise detection in Monte-Carlo rendering. *International Conference On Machine Learning And Applications*. <https://hal.archives-ouvertes.fr/hal-03374214>
- Burke, L. (1952). On the tunnel effect. *Quarterly Journal of Experimental Psychology*, 4(3), 121-138.
- Buxbaum, L. J. & Kalénine, S. (2010). Action knowledge, visuomotor activation, and embodiment in the two action systems. *Annals of the New York Academy of Sciences*, 1191(1), 201-218. <https://doi.org/10.1111/j.1749-6632.2010.05447.x>
- Campbell, F. W. & Robson, J. G. (1968). Application of fourier analysis to the visibility of gratings. *The Journal of Physiology*, 197(3), 551-566. <https://doi.org/10.1113/jphysiol.1968.sp008574>
- Campbell, F. (1983). Why do we measure contrast sensitivity? *Behavioural Brain Research*, 10(1), 87-97. [https://doi.org/10.1016/0166-4328\(83\)90154-7](https://doi.org/10.1016/0166-4328(83)90154-7)
- Cartmill, M. (1974). Rethinking Primate Origins. *Science*, 184(4135), 436-443. <https://doi.org/10.1126/science.184.4135.436>
- Casati, R. & Cavanagh, P. (2019). *The Visual World of Shadows*. The MIT Press. <https://doi.org/10.7551/mitpress/11007.001.0001>
- Cassagnabère, C. (2007). *Etude et implémentation d'algorithmes d'illumination globale stochastiques accélérés par le matériel* (thèse de doct.) [Thèse de doctorat dirigée par Renaud, Christophe et Rousselle, François Informatique Littoral 2007]. Université du Littoral – Côte d'Opale. <http://www.theses.fr/2007DUNK0167>
- Catmull, E. E. (1974). A subdivision algorithm for computer display of curved surfaces.
- Chelle, D. M., Renaud, C., Delepouille, S. & Combes, D. (2009). Light phylloclimate within growth chambers : a modeling approach. *Comparative Biochemistry and Physiology*

- Part A : Molecular & Integrative Physiology*, 153(2), S229. <https://doi.org/10.1016/j.cbpa.2009.04.635>
- Chen, J., Mi, L., Chen, C. P., Liu, H., Jiang, J. & Zhang, W. (2019). Design of foveated contact lens display for augmented reality. *Optics Express*, 27(26), 38204. <https://doi.org/10.1364/oe.381200>
- Chung, S., Choi, M., Youn, D. & Kim, S. (2019). Comparison of BVH and KD-Tree for the GPGPU Acceleration on Real Mobile Devices. *Lecture Notes in Electrical Engineering* (p. 535-540). Springer Singapore. https://doi.org/10.1007/978-981-13-3648-5_62
- Colenbrander, A. (2003). Aspects of vision loss – visual functions and functional vision. *Visual Impairment Research*, 5(3), 115-136. <https://doi.org/10.1080/1388235039048919>
- Constantin, J., Delepoulle, S., Bigand, A. & Renaud, C. (2013). Reduced-reference quality assessment of computer-generated images based on RVM. *2013 Third International Conference on Communications and Information Technology (ICCIT)*. <https://doi.org/10.1109/iccitechnology.2013.6579572>
- Cook, R. L. & Torrance, K. E. (1982). A reflectance model for computer graphics. *ACM Transactions on Graphics*, 1(1), 7-24.
- Crow, F. C. (1977). Shadow algorithms for computer graphics. *ACM SIGGRAPH Computer Graphics*, 11(2), 242-248. <https://doi.org/10.1145/965141.563901>
- Daly, S. J. (1992). Visible differences predictor : an algorithm for the assessment of image fidelity. In B. E. Rogowitz (Éd.), *Human Vision, Visual Processing, and Digital Display III*. SPIE. <https://doi.org/10.1117/12.135952>
- DeCoro, C., Weyrich, T. & Rusinkiewicz, S. (2010). Density-based Outlier Rejection in Monte Carlo Rendering. *Computer Graphics Forum*, 29(7), 2119-2125. <https://doi.org/10.1111/j.1467-8659.2010.01799.x>
- Delepoulle, S., Bigand, A. & Renaud, C. (2012). A no-reference computer-generated images quality metric and its application to denoising. *2012 6th IEEE INTERNATIONAL CONFERENCE INTELLIGENT SYSTEMS*. <https://doi.org/10.1109/is.2012.6335116>
- Delepoulle, S., Bigand, A., Renaud, C. & Colot, O. (2016). Low-Level Image Processing Based on Interval-Valued Fuzzy Sets and Scale-Space Smoothing. *Recent Contributions in Intelligent Systems* (p. 1-22). Springer International Publishing. https://doi.org/10.1007/978-3-319-41438-6_1
- Delepoulle, S., Preux, P. & Darcheville, J.-C. (2001). Selection of Behavior in Social Situations Application to the Development of Coordinated Movements. *Lecture Notes in Computer Science* (p. 384-393). Springer Berlin Heidelberg. https://doi.org/10.1007/3-540-45365-2_40
- Delepoulle, S., Preux, P. & Darcheville, J.-C. (2002). Learning as a Consequence of Selection. *Lecture Notes in Computer Science* (p. 350-361). Springer Berlin Heidelberg. https://doi.org/10.1007/3-540-46033-0_28
- Delepoulle, S., Renaud, C. & Chelle, M. (2009). Improving Light Position in a Growth Chamber through the Use of a Genetic Algorithm. *Artificial Intelligence Techniques for Computer Graphics* (p. 67-82). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-540-85128-8_5

- Delepoulle, S., Renaud, C. & Preux, P. (2009). Light Source Storage and Interpolation for Global Illumination : A Neural Solution. *Studies in Computational Intelligence* (p. 87-104). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-03452-7_5
- Demers, J. (2004). Depth of Field : A Survey of Techniques. In R. Fernando (Éd.), *GPU Gems* (p. 375-390). Addison-Wesley.
- Devisme, C. (2008). *Etude de l'influence des disparités horizontales et verticales sur la perception de la profondeur en champ visuel périphérique* (thèse de doct.). Université Pierre et Marie Curie -Paris VI.
- Dorfman, R. (1979). A Formula for the Gini Coefficient. *The Review of Economics and Statistics*, 61(1), 146-49. <https://EconPapers.repec.org/RePEc:tpr:restat:v:61:y:1979:i:1:p:146-49>
- Dosher, B. A., Sperling, G. & Wurst, S. A. (1986). Tradeoffs between stereopsis and proximity luminance covariance as determinants of perceived 3D structure. *Vision Research*, 26, 973-990.
- Dubos, J.-B. (1732). *Reflexions critiques sur la poesie et sur la peinture*. E. Neaulme.
- Ekman, G. (1959). Weber's Law and Related Functions. *The Journal of Psychology*, 47(2), 343-352. <https://doi.org/10.1080/00223980.1959.9916336>
- Geib, T. & Baumann, C. (1990). Effect of luminance and contrast on stereoscopic acuity. *Graefe's Archive for Clinical and Experimental Ophthalmology*, 228(4), 310-315. <https://doi.org/10.1007/bf00920053>
- Gerardin, P., Kourtzi, Z. & Mamassian, P. (2010). Prior knowledge of illumination for 3D perception in the human brain. *Proceedings of the National Academy of Sciences*, 107(37), 16309-16314. <https://doi.org/10.1073/pnas.1006285107>
- Gerhards, J. (2017). *Partition de volumes d'ombres une alternative pour le rendu d'ombres en temps réel* (thèse de doct.). Université de Limoges.
- Gibaldi, A. & Banks, M. S. (2019). Binocular Eye Movements Are Adapted to the Natural Environment. *The Journal of Neuroscience*, 39(15), 2877-2888. <https://doi.org/10.1523/jneurosci.2591-18.2018>
- Gibson, J. (2015). *The ecological approach to visual perception*. Psychology Press.
- Gibson, J. J. (1947). *Motion picture testing and research*. (T. 46). U.S. Government Printing Office. <https://doi.org/10.1037/h0050640>
- Gini, C. (1936). On the measure of concentration with special reference to income and statistics. *Colorado College Publication, General Series*, 208(1), 73-79.
- Godard, M., Wamain, Y., Ott, L., Delepoulle, S. & Kalénine, S. (2022). How Competition between Action Representations Affects Object Perception during Development. *Journal of Cognition and Development*, 1-25. <https://doi.org/10.1080/15248372.2022.2025808>
- Goldman Sachs. (2018). Profiles in innovation report [[Online ; accessed 08-janv-2022]].
- Gordon, I. (2004). *Theories of visual perception*. Psychology Press.
- Gu, Y., He, Y., Fatahalian, K. & Blelloch, G. (2013). Efficient BVH construction via approximate agglomerative clustering. *Proceedings of the 5th High-Performance Graphics Conference on - HPG '13*. <https://doi.org/10.1145/2492045.2492054>
- Harré, M., Bossomaier, T. & Snyder, A. (2012). The Perceptual Cues that Reshape Expert Reasoning. *Scientific Reports*, 2(1). <https://doi.org/10.1038/srep00502>

- Harrold, A. L. & Grove, P. M. (2021). The vergence horopter. *Vision Research*, 180, 63-79. <https://doi.org/10.1016/j.visres.2020.12.003>
- Hasenfratz, J.-M., Lapierre, M., Holzschuch, N. & Sillion, F. X. (2003). A survey of real-time soft shadows algorithms. *Computer Graphics Forum*, 22(4), 753-774.
- Herzog, R., Čadík, M., Aydın, T. O., Kim, K. I., Myszkowski, K. & Seidel, H.-P. (2012). NoRM : No-Reference Image Quality Metric for Realistic Image Synthesis. *Computer Graphics Forum*, 31(2), 545-554. <https://doi.org/10.1111/j.1467-8659.2012.03055.x>
- Hillis, J. M., Watt, S. J., Landy, M. S. & Banks, M. S. (2004). Slant from texture and disparity cues : Optimal cue combination. *Journal of Vision*, 4(12), 1. <https://doi.org/10.1167/4.12.1>
- Hofmann, G. R. (1990). Who invented ray tracing? *The Visual Computer*, 6(3), 120-124. <https://doi.org/10.1007/bf01911003>
- Holladay, J. T. (1997). Proper Method for Calculating Average Visual Acuity. *Journal of Refractive Surgery*, 13(4), 388-391. <https://doi.org/10.3928/1081-597x-19970701-16>
- Holway, A. H. & Boring, E. G. (1940). The Moon Illusion and the Angle of Regard. *The American Journal of Psychology*, 53(1), 109. <https://doi.org/10.2307/1415964>
- Howard, I. (2008). *Seeing in depth*. Oxford University Press.
- Huan, Q. (2021). *Perception du bruit visuel en vision stéréoscopique* (mém. de mast.). ENSEM.
- Hullin, M. B., Eisemann, E., Seidel, H.-P. & Lee, S. (2011). Physically-Based Real-Time Lens Flare Rendering. *ACM Trans. Graph. (Proc. SIGGRAPH 2011)*, 30(4), 108 :1-108 :9.
- Jenny, B. & Patterson, T. (2020). Aerial perspective for shaded relief. *Cartography and Geographic Information Science*, 48(1), 21-28. <https://doi.org/10.1080/15230406.2020.1813052>
- Jensen, H. W. (1996a). Global Illumination using Photon Maps. *Eurographics* (p. 21-30). Springer Vienna. https://doi.org/10.1007/978-3-7091-7484-5_3
- Jensen, H. W. (1996b). Rendering Techniques '96. In X. Pueyo & P. Schröder (Éd.). Springer-Verlag.
- Jensen, H. W. (2001). *Realistic image synthesis using photon mapping*. AK Peters/crc Press.
- Jinabhai, A., Young, G., Hall, L. & Wolffsohn, J. (2013). Clinical techniques to assess the visual and optical performance of intraocular lenses : A review.
- Jones, G. R., Lee, D., Holliman, N. S. & Ezra, D. (2001). Controlling perceived depth in stereoscopic images. In A. J. Woods, M. T. Bolas, J. O. Merritt & S. A. Benton (Éd.), *Stereoscopic Displays and Virtual Reality Systems VIII*. SPIE. <https://doi.org/10.1117/12.430855>
- Jung, J. H. (2019). Accommodation and Convergence. *Primary Eye Examination* (p. 31-36). Springer Singapore. https://doi.org/10.1007/978-981-10-6940-6_3
- Jung, J. W., Meyer, G. & DeLong, R. (2015). Robust Statistical Pixel Estimation. *Computer Graphics Forum*, 34(2), 585-596. <https://doi.org/10.1111/cgf.12586>
- Kajiya, J. T. (1986). The rendering equation. *SIGGRAPH '86 Proceedings of the 13th annual conference on Computer graphics and interactive techniques*, 143-150.

- Kalénine, S., Wamain, Y., Decroix, J. & Coello, Y. (2016). Conflict between object structural and functional affordances in peripersonal space. *Cognition*, 155, 1-7. <https://doi.org/10.1016/j.cognition.2016.06.006>
- Kalloniatis, M. & Luu, C. (2005). Temporal Resolution. In H. K. Al (Éd.). University of Utah Health Sciences Center.
- Kanizsa, G. (1976). Subjective Contours. *Scientific American*, 234(4), 48-52. <https://doi.org/10.1038/scientificamerican0476-48>
- Kanizsa, G. (1979). *Organization in Vision*. Praeger. https://www.ebook.de/de/product/3694754/gaetano_kanizsa_organization_in_vision.html
- Kavšek, M., Granrud, C. E. & Yonas, A. (2009). Infants' responsiveness to pictorial depth cues in preferential-reaching studies : A meta-analysis. *Infant Behavior and Development*, 32(3), 245-253. <https://doi.org/10.1016/j.infbeh.2009.02.001>
- Keller, A. (1998). *Quasi-Monte Carlo Methods for Photorealistic Image Synthesis* (thèse de doct.). University of Kaiserslautern. Kaiserslautern, Germany.
- Keller, A. (1997). Instant radiosity. *Proceedings of the 24th annual conference on Computer graphics and interactive techniques - SIGGRAPH '97*. <https://doi.org/10.1145/258734.258769>
- Kellman, P. J. (2003). Interpolation processes in the visual perception of objects. *Neural Networks*, 16(5-6), 915-923.
- Kelly, D. (1975). Spatial frequency selectivity in the retina. *Vision Research*, 15(6), 665-672. [https://doi.org/10.1016/0042-6989\(75\)90282-5](https://doi.org/10.1016/0042-6989(75)90282-5)
- Kepler, J. (1604). *Optics Paralipomena to Witelo & Optical Part of Astronomy*. Green Lion Press, 2000. https://www.ebook.de/de/product/13766886/johannes_kepler_optics_paralipomena_to_witelo_optical_part_of_astronomy.html
- Kersten, D. & Mamassian, P. (2017). *Cast Shadow Illusions*. Oxford University Press. <https://doi.org/10.1093/acprof:oso/9780199794607.003.0020>
- Kim, J., Kane, D. & Banks, M. S. (2014). The rate of change of vergence–accommodation conflict affects visual discomfort. *Vision Research*, 105, 159-165. <https://doi.org/10.1016/j.visres.2014.10.021>
- Kniestedt, C. & Stamper, R. (2003). Visual acuity and its measurement. *Ophthalmology clinics of North America*, 16(2), 155-170. [https://doi.org/10.1016/s0896-1549\(03\)00013-0](https://doi.org/10.1016/s0896-1549(03)00013-0)
- Knill, D. C., Mamassian, P. & Kersten, D. (1997). Geometry of shadows. *J. Opt. Soc. Am. A*, 14(12), 3216-3232. <https://doi.org/10.1364/JOSAA.14.003216>
- Knill, D. C. & Saunders, J. A. (2003). Do humans optimally integrate stereo and texture information for judgments of surface slant? *Vision Research*, 43(24), 2539-2558. [https://doi.org/10.1016/s0042-6989\(03\)00458-9](https://doi.org/10.1016/s0042-6989(03)00458-9)
- Kubelka, P. & Munk, F. (1931). Ein Beitrag zur Optik der Farbanstriche. *Zeitschrift für Technische Physik*, 12, 593-601.
- Lafortune, E. P. & Willems, Y. D. (1993). Bi-Directional Path Tracing. *Proceedings of third international conference on computational graphics and visualisation techniques (COMPUGRAPHICS'93)*, 145-153.
- Lewis, R. (2003). Making Shaders More Physically Plausible. *Fourth Eurographics Workshop on rendering*, 47-62.

- Li, R., Polat, U., Makous, W. & Bavelier, D. (2009). Enhancing the contrast sensitivity function through action video game training. *Nature Neuroscience*, *12*(5), 549-551. <https://doi.org/10.1038/nm.2296>
- Linton, P. (2020). Does vision extract absolute distance from vergence? *Attention, Perception, & Psychophysics*, *82*(6), 3176-3195. <https://doi.org/10.3758/s13414-020-02006-1>
- Loth, M., Preux, P., Delepouille, S. & Renaud, C. (2009). ECON : A Kernel Basis Pursuit Algorithm with Automatic Feature Parameter Tuning, and its Application to Photometric Solids Approximation. *2009 International Conference on Machine Learning and Applications*. <https://doi.org/10.1109/icmla.2009.61>
- Lynes, J. A. (1980). Brunelleschi's Perspectives Reconsidered. *Perception*, *9*(1), 87-99. <https://doi.org/10.1068/p090087>
- Mamassian, P., Knill, D. C. & Kersten, D. (1998). The perception of cast shadows. *Trends in Cognitive Sciences*, *2*(8), 288-295. [https://doi.org/10.1016/S1364-6613\(98\)01204-2](https://doi.org/10.1016/S1364-6613(98)01204-2)
- Mannos, J. & Sakrison, D. (1974). The effects of a visual fidelity criterion of the encoding of images. *IEEE Transactions on Information Theory*, *20*(4), 525-536. <https://doi.org/10.1109/tit.1974.1055250>
- McGettigan, C. & Tremblay, P. (2018). Links Between Perception and Production (S.-A. Rueschemeyer & M. G. Gaskell, Éd.). <https://doi.org/10.1093/oxfordhb/9780198786825.013.14>
- McIntire, J. P., Havig, P. R. & Geiselman, E. E. (2014). Stereoscopic 3D displays and human performance : A comprehensive review. *Displays*, *35*(1), 18-26. <https://doi.org/10.1016/j.displa.2013.10.004>
- McKee, S. P., Welch, L., Taylor, D. G. & Bowne, S. F. (1990). Finding the common bond : Stereoacuity and the other hyperacuties. *Vision Research*, *30*(6), 879-891. [https://doi.org/10.1016/0042-6989\(90\)90056-q](https://doi.org/10.1016/0042-6989(90)90056-q)
- Meister, D., Ogaki, S., Benthin, C., Doyle, M. J., Guthe, M. & Bittner, J. (2021). A Survey on Bounding Volume Hierarchies for Ray Tracing. *Computer Graphics Forum*, *40*(2), 683-712. <https://doi.org/10.1111/cgf.142662>
- Michel, B. (2012). *La stéréoscopie numérique tourner, éditer, diffuser, imprimer, projeter*. Eyrolles.
- Michotte, A. & O., T. G. (1964). *Les compléments amodaux des structures perceptives* (I. de Psychologie de l'Université de Louvain, Éd.).
- Mohanto, B., Islam, A. T., Gobbetti, E. & Staadt, O. (2021). An integrative view of foveated rendering. *Computers & Graphics*. <https://doi.org/10.1016/j.cag.2021.10.010>
- Montagne, F., Delepouille, S. & Preux, P. (2003). A critic-critic architecture to combine reinforcement and supervised learnings. *European Workshop on reinforcement Learning*.
- Montagne, F., Preux, P. & Delepouille, S. (2006). Introducing interactive help for reinforcement learners, Workshop on planning, learning and monitoring with uncertainty and dynamic worlds. *ECAI Workshop*.
- Mooney, S. W., Marlow, P. J. & Anderson, B. L. (2019). The perception and misperception of optical defocus, shading, and shape. *Elife*, *8*. <https://doi.org/10.7554/elife.48214>

- Murphy, J., Hall, C. & Arkins, S. (2009). What Horses and Humans See : A Comparative Review. *International Journal of Zoology*, 2009, 1-14. <https://doi.org/10.1155/2009/721798>
- Myrodia, V. (2021). *Études psychophysiques sur la perception visuelle du bruit de rendu de Monte Carlo* (thèse de doct.) [Thèse de doctorat dirigée par Madelain, Laurent, Psychologie 2021]. Université de Lille.
- Myrodia, V., Buisine, J. & Madelain, L. (2021). Comparison of threshold measurements in laboratory and online studies using a Quest+ algorithm. *Journal of Vision*, 21(9), 1959. <https://doi.org/10.1167/jov.21.9.1959>
- Myrodia, V., Delepouille, S. & Madelain, L. (2020). Foveal and peripheral vision for assessing the quality of computer-generated images. *Journal of Vision*, 20(11), 355. <https://doi.org/10.1167/jov.20.11.355>
- Myszkowski, K. (1998). The Visible Differences Predictor : Applications to Global Illumination Problems. *Rendering Techniques*.
- Nicodemus, F., Richmond, J., Hsia, J., Ginsberg, I. & Limperis, T. (1977). *Geometric Considerations and Nomenclature for Reflectance* (rapp. tech. Monograph 161). National Bureau of Standards (US).
- Nityananda, V. & Read, J. C. A. (2017). Stereopsis in animals : evolution, function and mechanisms. *Journal of Experimental Biology*, 220(14), 2502-2512. <https://doi.org/10.1242/jeb.143883>
- Oliva, A., Torralba, A. & Schyns, P. G. (2006). Hybrid images. *ACM Transactions on Graphics*, 25(3), 527-532. <https://doi.org/10.1145/1141911.1141919>
- Orenstein, P. (2019). Robust Mean Estimation with the Bayesian Median of Means.
- Orssaud, C. (2006). Vision binoculaire. *EMC - Ophtalmologie*, 3(3), 1-10. [https://doi.org/10.1016/s0246-0343\(06\)45657-4](https://doi.org/10.1016/s0246-0343(06)45657-4)
- Oshika, T., Okamoto, C., Samejima, T., Tokunaga, T. & Miyata, K. (2006). Contrast Sensitivity Function and Ocular Higher-Order Wavefront Aberrations in Normal Human Eyes. *Ophthalmology*, 113(10), 1807-1812. <https://doi.org/10.1016/j.ophtha.2006.03.061>
- Ozkan, K. & Braunstein, M. L. (2010). Background surface and horizon effects in the perception of relative size and distance. *18(2)*, 229-254. <https://doi.org/10.1080/13506280802674101>
- Patney, A., Salvi, M., Kim, J., Kaplanyan, A., Wyman, C., Benty, N., Luebke, D. & Lefohn, A. (2016). Towards foveated rendering for gaze-tracked virtual reality. *ACM Transactions on Graphics (TOG)*, 35(6), 1-12.
- Péroche, B., Delepouille, S. & Renaud, C. (2013). Rendering and Computer Graphics. *Digital Color* (p. 93-118). John Wiley & Sons, Inc. <https://doi.org/10.1002/9781118562680.ch4>
- Pharr, M., Humphreys, G. & Jakob, W. (2016). *Physically Based Rendering*. Elsevier LTD, Oxford. https://www.ebook.de/de/product/25867811/matt_pharr_greg_humphreys_wenzel_jakob_physically_based_rendering.html
- Phong, B. T. (1975). Illumination for Computer Generated Pictures. *Communication of ACM*, 18(6), 311-317.
- Popper, K. (2006). *Conjectures et réfutations la croissance du savoir scientifique*. Payot.
- Proffitt, D. R. (2006). Distance Perception. *Current Directions in Psychological Science*, 15(3), 131-135. <https://doi.org/10.1111/j.0963-7214.2006.00422.x>

- Rand, K., Tarampi, M., Creem-Regehr, S. & Thompson, W. (2010). The importance of a visual horizon for distance judgments under severely degraded vision. *Journal of Vision*, 10(7), 61-61. <https://doi.org/10.1167/10.7.61>
- Reeves, W. T., Salesin, D. H. & Cook, R. L. (1987). Rendering antialiased shadows with depth maps. *Proceedings of the 14th annual conference on Computer graphics and interactive techniques - SIGGRAPH '87*. <https://doi.org/10.1145/37401.37435>
- Reinhardt-Rutland, A. H. (1996). Remote Operation : A Selective Review of Research into Visual Depth Perception. *The Journal of General Psychology*, 123(3), 237-248. <https://doi.org/10.1080/00221309.1996.9921276>
- Renaud, C., Delepoulle, S. & Takouachet, N. (2012). Detecting Visual Convergence for Stochastic Global Illumination. *Intelligent Computer Graphics 2011* (p. 1-17). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-22907-7_1
- Ritter, M. (1977). Effect of disparity and viewing distance on perceived depth. *Perception & Psychophysics*, 22(4), 400-407. <https://doi.org/10.3758/bf03199707>
- Riva, G., Baños, R. M., Botella, C., Mantovani, F. & Gaggioli, A. (2016). Transforming experience : the potential of augmented reality and virtual reality for enhancing personal and clinical change. *Frontiers in psychiatry*, 7, 164.
- Riva, G., Wiederhold, B. K. & Mantovani, F. (2019). Neuroscience of virtual reality : from virtual exposure to embodied medicine. *Cyberpsychology, Behavior, and Social Networking*, 22(1), 82-96.
- Rogers, B. (2019). Toward a new theory of stereopsis A critique of Vishwanath 2014. *Psychological Review*, 126(1), 162-169. <https://doi.org/10.1037/rev0000131>
- Roper-Hall, G. & Jørgensen, H. (2008). Peter Ludvig Panum (1820–1885), Danish Physician and Physiologist. *American Orthoptic Journal*, 58(1), 99-107. <https://doi.org/10.3368/aoj.58.1.99>
- Russ, J. (2016). *The Image processing handbook*. CRC Press LLC.
- Sattler, M., Sarlette, R., Mücken, T. & Klein, R. (2005). Exploitation of human shadow perception for fast shadow rendering. *Proceedings of the 2nd symposium on Applied perception in graphics and visualization - APGV '05*. <https://doi.org/10.1145/1080402.1080426>
- Schreiber, K. M., Hillis, J. M., Filippini, H. R., Schor, C. M. & Banks, M. S. (2008). The surface of the empirical horopter. *Journal of Vision*, 8(3), 7. <https://doi.org/10.1167/8.3.7>
- Schwiegerling, J. (2004). *Field Guide to Visual and Ophthalmic Optics*. SPIE. <https://doi.org/10.1117/3.592975>
- Sedgwick, H., Beck, J., Hope, B. & Rosenfeld, A. (1983). Human and machine vision.
- Segovia, B., Iehl, J.-C. & Péroche, B. (2007). Metropolis Instant Radiosity. *Computer Graphics Forum*, 26(3), 425-434. <http://liris.cnrs.fr/publis/?id=2821>
- Sheedy, J. E., Bailey, I. L. & Raash, T. W. (1984). Visual Acuity and Chart Luminance. *Optometry and Vision Science*, 61(9), 595-600. <https://doi.org/10.1097/00006324-198409000-00010>
- Simon, G. (2021). Jan van Eyck's Perspectival System Elucidated Through Computer Vision. *Proceedings of the ACM on Computer Graphics and Interactive Techniques*, 4(2), 1-8. <https://doi.org/10.1145/3465623>
- Straßer, T., Kurtenbach, A., Langrová, H., Kuehlewein, L. & Zrenner, E. (2020). The perception threshold of the panda illusion, a particular form of 2D pulse-width-

- modulated halftone, correlates with visual acuity. *Scientific Reports*, 10(1). <https://doi.org/10.1038/s41598-020-69952-6>
- Surdick, R. T., Davis, E. T., King, R. A., Corso, G. M., Shapiro, A., Hodges, L. & Elliot, K. (1994). Relevant Cues for the Visual Perception of Depth : Is Where You See it Where it is ? *Proceedings of the Human Factors and Ergonomics Society Annual Meeting.*, 38(19), 1305-1309. <https://doi.org/10.1177/154193129403801912>
- Takouachet, N. (2009). *Utilisation de critères perceptifs pour la détermination d'une condition d'arrêt dans les méthodes d'illumination globale* (thèse de doct.) [Thèse de doctorat dirigée par Renaud, Christophe et Delepouille, Samuel Informatique Littoral 2009]. Université du Littoral – Côte d'Opale. <http://www.theses.fr/2009DUNK0229>
- Takouachet, N., Delepouille, S. & Renaud, C. (2007). A perceptual stopping condition for global illumination computations. *Spring Conference on Computer Graphics*. <https://hal.archives-ouvertes.fr/hal-03479491>
- Takouachet, N., Delepouille, S., Renaud, C., Zoghlami, N. & Tavares, J. M. R. (2017). Perception of noise and global illumination : Toward an automatic stopping criterion based on SVM. *Computers and Graphics*, 69, 49-58. <https://doi.org/10.1016/j.cag.2017.09.008>
- Teller, S. J. & Sequin, C. H. (1991). Visibility Preprocessing For Interactive Walkthroughs. *COMPUTER GRAPHICS (SIGGRAPH 91 PROCEEDINGS)*, 61-69.
- Timby, K. (2005). Colour photography and stereoscopy : Parallel histories. *History of Photography*, 29(2), 183-196. <https://doi.org/10.1080/03087298.2005.10441370>
- Torrance, K. & Sparrow, E. (1967). Theory for Off-Specular Reflection from Roughened Surfaces. *Journal of the Optical Society of America*, 57(9), 1105-1114.
- Torrence, A. (2006). Martin Newell's original teapot. *ACM SIGGRAPH 2006 Teapot on - SIGGRAPH '06*. <https://doi.org/10.1145/1180098.1180128>
- Trapp, M. (2013). *Interactive Rendering Techniques for Focus+Context Visualization of 3D Geovirtual Environments* (thèse de doct.).
- Trapp, M., Lorenz, H., Jobst, M. & Döllner, J. (2011). Enhancing Interactive NonPlanar Projections of 3D Geovirtual Environments with Stereoscopic Imaging. *Lecture Notes in Geoinformation and Cartography* (p. 297-312). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-12272-9_20
- Tsutsui, K.-I., Sakata, H., Naganuma, T. & Taira, M. (2002). Neural Correlates for Perception of 3D Surface Orientation from Texture Gradient. *Science*, 298(5592), 409-412. <https://doi.org/10.1126/science.1074128>
- van Lier, R. & Ekroll, V. (2020). A Conceptual Playground Between Perception and Cognition : Introduction to the Special Issue on Amodal Completion. *i-Perception*, 11(4), 204166952093910. <https://doi.org/10.1177/2041669520939108>
- Varnedoe, K. (1988). *Gustave Caillebotte*. Adam Biro.
- Veach, E. (1997). *Robust Monte Carlo Methods for Light Transport Simulation* (thèse de doct.). Stanford University. Stanford, California.
- Veach, E. & Guibas, L. (1995). Bidirectional Estimators for Light Transport. In G. Sakas, S. Müller & P. Shirley (Éd.), *Photorealistic Rendering Techniques* (p. 145-167). Springer Berlin Heidelberg.
- Veach, E. & Guibas, L. J. (1997). Metropolis light transport. *Proceedings of the 24th annual conference on Computer graphics and interactive techniques*, 65-76.

- Velasco e Cruz, A. A. (1990). Historical Roots of 20/20 as a (Wrong) Standard Value of Normal Visual Acuity. *Optometry and Vision Science*, 67(8), 661. <https://doi.org/10.1097/00006324-199008000-00022>
- Vienne, C., Blondé, L. & Mamassian, P. (2015). Depth-of-Focus Affects 3D Perception in Stereoscopic Displays. *Perception*, 44(6), 613-627. <https://doi.org/10.1177/0301006615594261>
- Vinci, L. D. (2012). *The Notebooks of Leonardo da Vinci, Vol. 1*. Apa Publications. https://www.ebook.de/de/product/22388772/leonardo_da_vinci_the_notebooks_of_leonardo_da_vinci_vol_1.html
- Vinkler, M., Havran, V. & Bittner, J. (2015). Performance Comparison of Bounding Volume Hierarchies and Kd-Trees for GPU Ray Tracing. *Computer Graphics Forum*, 35(8), 68-79. <https://doi.org/10.1111/cgf.12776>
- Volevich, V., Myszkowski, K., Khodulev, A. & Kopylov, E. A. (2000). Using the Visual Differences Predictor to Improve Performance of Progressive Global Illumination Computation. *ACM Trans. Graph.*, 19(2), 122-161. <https://doi.org/10.1145/343593.343611>
- Wade, N. J. & Ono, H. (1985). The stereoscopic views of Wheatstone and Brewster. *Psychological Research*, 47(3), 125-133. <https://doi.org/10.1007/bf00309263>
- Walter, B., Fernandez, S., Arbree, A., Bala, K., Donikian, M. & Greenberg, D. P. (2005). Lightcuts : A Scalable Approach to Illumination. *ACM Trans. Graph.*, 24(3), 1098-1107. <https://doi.org/10.1145/1073204.1073318>
- Wandell, B. (1995). *Foundations of vision*. Sinauer Associates.
- Wang, Y.-J. & Lin, Y.-H. (2021). Liquid crystal technology for vergence-accommodation conflicts in augmented reality and virtual reality systems : a review. *Liquid Crystals Reviews*, 9(1), 35-64. <https://doi.org/10.1080/21680396.2021.1948927>
- Wang, Z., Bovik, A., Sheikh, H. & Simoncelli, E. (2004). Image quality assessment : from error visibility to structural similarity. *IEEE Transactions on Image Processing*, 13(4), 600-612. <https://doi.org/10.1109/TIP.2003.819861>
- Warren, W. H. (1990). The perception-action coupling. *Sensory-motor organizations and development in infancy and early childhood* (p. 23-37). Springer.
- Watson, M. & Enns, J. (2012). Depth Perception. *Encyclopedia of Human Behavior* (p. 690-696). Elsevier. <https://doi.org/10.1016/b978-0-12-375000-6.00130-0>
- Watt, S. J., Akeley, K., Ernst, M. O. & Banks, M. S. (2005). Focus cues affect perceived depth. *Journal of Vision*, 5(10), 7. <https://doi.org/10.1167/5.10.7>
- Wei, L. & Sakamoto, Y. (2019). Fast calculation method with foveated rendering for computer-generated holograms using an angle-changeable ray-tracing method. *Applied Optics*, 58(5), A258. <https://doi.org/10.1364/ao.58.00a258>
- Weier, M., Stengel, M., Roth, T., Didyk, P., Eisemann, E., Eisemann, M., Grogorick, S., Hinkenjann, A., Kruijff, E., Magnor, M., Myszkowski, K. & Slusallek, P. (2017). Perception-driven Accelerated Rendering. *Computer Graphics Forum (Proc. of Eurographics)*, 36. <https://doi.org/10.1111/cgf.13150>
- Weinstein, S. (1957). The Perception of Depth in the Absence of Texture-Gradient. *The American Journal of Psychology*, 70(4), 611. <https://doi.org/10.2307/1419453>
- Wertheim, T. (1980). Peripheral visual acuity : Th. Wertheim. *American journal of optometry and physiological optics*, 57, 915-924.

- Whitted, T. (1980). An improved illumination model for shaded display. *Communications of the ACM*, 23(6), 343-349.
- Woo, A., Poulin, P. & Fournier, A. (1990). A survey of shadow algorithms. *IEEE Computer Graphics and Applications*, 10(6), 13-32. <https://doi.org/10.1109/38.62693>
- Wöpking, M. (1995). Viewing comfort with stereoscopic pictures : An experimental study on the subjective effects of disparity magnitude and depth of focus. *Journal of the Society for Information Display*, 3(3), 101. <https://doi.org/10.1889/1.1984948>
- Yonas, A. & Granrud, C. E. (2006). Infants' perception of depth from cast shadows. *Perception & Psychophysics*, 68(1), 154-160. <https://doi.org/10.3758/bf03193665>
- Yonas, A., Granrud, C. E., Arterberry, M. E. & Hanson, B. L. (1986). Infants' distance perception from linear perspective and texture gradients. *Infant Behavior and Development*, 9(3), 247-256. [https://doi.org/10.1016/0163-6383\(86\)90001-9](https://doi.org/10.1016/0163-6383(86)90001-9)
- Yoon, H. J., Moon, H. S., Sung, M. S., Park, S. W. & Heo, H. (2021). Effects of prolonged use of virtual reality smartphone-based head-mounted display on visual parameters : a randomised controlled trial. *Scientific Reports*, 11(1). <https://doi.org/10.1038/s41598-021-94680-w>
- Yuksel, C. (2019). Stochastic Lightcuts. *High-Performance Graphics (HPG 2019)*, 27-32. <https://doi.org/10.2312/hpg.20191192>
- Zirr, T., Hanika, J. & Dachsbacher, C. (2018). Re-Weighting Firefly Samples for Improved Finite-Sample Monte Carlo Estimates. *Computer Graphics Forum*, 37.

Troisième partie

Annexes

Annexes

Depuis ma prise de fonction en tant que Maître de Conférences en septembre 2001, j'ai participé à l'encadrement de quatre doctorants.

Tout d'abord, dans la continuité de mes travaux de thèse sur l'apprentissage par renforcement, j'ai co-encadré le travail de Fabien Montagne avec Philippe Preux. Le but de cette thèse était de trouver des méthodes pour accélérer l'apprentissage non-supervisé en proposant des méthodes de guidage lors de la phase d'exploration initiale. Fabien Montagne a soutenu sa thèse en 2008.

Ensuite, j'ai co-encadré trois doctorants avec Christophe Renaud. La première, celle de Nawel Takouachet a marqué un premier pas dans l'utilisation de méthodes d'apprentissage automatique pour déterminer un seuil d'arrêt des calculs dans le cadre de l'illumination globale. La deuxième consistait à étendre les recherches aux périphériques 3D n'a pas été soutenue à cause de problèmes personnels rencontrés par Thuy Tram Ngo Thi. Enfin, de 2018 à 2021, j'ai participé à l'encadrement du travail de Jérôme Buisine soutenu en décembre 2021, dans le cadre de l'ANR PrISE-3D.

1 DEA et thèse de Fabien Montagne

1.1 Contexte

1.2 Résumé

Cette thèse s'inscrit dans le cadre de l'apprentissage par renforcement. L'un des principaux avantages est qu'il ne nécessite pas de connaître explicitement le comportement attendu. Durant son apprentissage, l'agent perçoit des états, reçoit un ensemble de retours et sélectionne des actions. Il adapte son comportement en optimisant la quantité de retour. Néanmoins, le temps de calcul nécessaire peut vite être prohibitif. La nécessité d'explorer son environnement en est la principale raison. Notre approche consiste à utiliser des connaissances externes pour « guider » l'agent dans son exploration. Ces connaissances constituent une aide pouvant s'exprimer sous forme de trajectoires formant une base de connaissances. Elles limitent l'exploration de l'environnement tout en permettant d'acquérir un comportement de bonne qualité. Aider un agent n'implique pas de connaître une politique, même partielle, ni d'avoir la même perception que l'agent. L'architecture critique-critique a été conçue pour répondre à cette problématique. Elle combine un algorithme d'apprentissage par renforcement standard avec une aide exprimée sous forme de potentiels, lesquels associent une valeur à chaque transition des trajectoires. L'estimation de la valeur par l'agent et le potentiel de l'aide sont combinés au cours de l'apprentissage. La variation de cette combinaison permet de remettre en cause l'aide tout en garantissant une politique optimale ou proche rapidement. Il est montré que l'algorithme proposé converge dans certaines conditions. De plus, des travaux empiriques montrent que l'agent est capable de tirer

profit d'une aide même en dehors de ces conditions.

1.3 Soutenance

co-direction du DEA et de la thèse de Fabien Montagne (titre : « Une architecture logicielle pour aider un agent apprenant par renforcement », thèse soutenue en janvier 2008.

Rapporteurs :

- Mohamed Slimane Université de Tours
- Abdel-illah Mouaddib Université de Caen

Examineurs :

- Pr. Henri Basson
- Dr. Dominique Leclét

Directeur et co-directeur :

- Philippe Preux (LIL-Calais)
- Co-directeur : Dr. Samuel Delepouille (LIL-Calais)

1.4 Publications associées

1. Montagne, F., Delepouille, S. & Preux, P. (2003). A critic-critic architecture to combine reinforcement and supervised learnings. *European Workshop on reinforcement Learning*
2. Montagne, F., Preux, P. & Delepouille, S. (2006). Introducing interactive help for reinforcement learners, Workshop on planning, learning and monitoring with uncertainty and dynamic worlds. *ECAI Workshop*

1.4.1 Montagne et al. (2003)

Dans la vie réelle, l'apprentissage est grandement accéléré par l'intervention d'un enseignant qui donne des exemples ou montre comment effectuer une certaine tâche. Nous laissons de côté les simplifications structurelles du problème par le concepteur qui ne traitent pas explicitement de l'apprentissage. L'intervention de l'enseignant peut être réalisée de différentes manières : explication verbale, démonstration, guidage, modelage du comportement, ...

Conceptuellement, cela revient à combiner l'apprentissage par renforcement avec l'apprentissage supervisé. Dans ce travail, nous nous concentrons sur la technique de guidage dans laquelle nous prenons virtuellement la main de l'apprenant et lui faisons exécuter la tâche, ou une partie de celle-ci. Les exemples sont des trajectoires dans l'espace d'état. Ayant un ensemble d'exemples, de nombreuses façons d'effectuer une telle combinaison peuvent être envisagées. Ici, nous souhaitons fournir des conseils à un apprenant par renforcement et tirer parti d'un certain nombre de ses propriétés : utiliser les conseils comme une aide plutôt que comme un ordre strict ; améliorer les exemples qui ont été donnés comme conseils ; s'adapter lorsque l'environnement change ; généraliser autant que possible à partir de l'ensemble d'exemples. Cette dernière caractéristique est liée à l'utilisation d'une architecture pertinente pour stocker les estimations actuelles de la valeur des états. Les trois autres caractéristiques sont liées au compromis entre l'exploration et l'exploitation.

Montagne, F., Delepouille, S. & Preux, P. (2003). A critic-critic architecture to combine reinforcement and supervised learnings. *European Workshop on reinforcement Learning*

1.4.2 Montagne et al. (2006)

Le problème de l'apprentissage par renforcement est un problème très difficile lorsqu'on considère des applications de grande taille (taille réelle). Pour le résoudre, nous pensons que de nombreuses questions doivent être étudiées en même temps. Pour réaliser un tel effort, nous pensons également qu'il est assez courant qu'un débutant humain puisse fournir de l'aide à la volée à l'apprenant par renforcement, c'est-à-dire lorsqu'il/elle voit comment l'apprenant se comporte, ou pourrait faire mieux que ce qu'il/elle fait actuellement. Ce travail vise précisément à concevoir une méthode de principe pour aider un agent qui apprend par renforcement dans un tel contexte. Nous mettons l'accent sur le fait qu'il n'est pas nécessaire à ce que l'aide fournie soit parfaite (quel que soit le sens du mot "parfait" ici...).

Montagne, F., Preux, P. & Delepouille, S. (2006). Introducing interactive help for reinforcement learners, Workshop on planning, learning and monitoring with uncertainty and dynamic worlds. *ECAI Workshop*

2 DEA et Thèse de Nawel Takouachet

2.1 Contexte

2.2 Résumé

La thèse s'intéresse aux modèles de rendu réaliste en synthèse d'images, en particulier aux algorithmes non-biaisés d'illumination globale. Leur intérêt est de calculer précisément une solution d'illumination qui permet notamment de produire des images réalistes. Cependant ces modèles sont soumis à la présence de bruit visuel, du fait de la nature stochastique des méthodes sur lesquelles elles reposent. Ce bruit s'atténue avec l'augmentation du nombre d'échantillons utilisés, ce qui conduit généralement à des temps de calculs très importants. Dans le cadre de cette thèse, nous nous sommes intéressés à la recherche d'un critère d'arrêt automatique de ces algorithmes, sur la base de critères perceptifs permettant de déterminer la présence ou non de bruit visible. Après un tour d'horizon des différentes méthodes employées pour le rendu d'image, le problème de l'intégration des modèles perceptifs est envisagé. Il s'agit d'utiliser les connaissances établies sur le système visuel humain afin de guider des algorithmes de rendu d'images. Dans un second temps, deux méthodologies sont proposées, reposant respectivement sur l'utilisation d'un modèle de perception et d'un modèle d'apprentissage supervisé. Ces deux approches sont calibrées à partir de données expérimentales obtenues auprès d'observateurs humains. Une comparaison entre ces deux méthodes montre que celle reposant sur l'apprentissage est plus avantageuse en terme de coût (calcul, mémoire) et permet une répartition différenciée des efforts de calcul au travers des images, en se focalisant sur les zones qui laissent apparaître du bruit.

2.3 Soutenance

La thèse a été soutenue le 29 janvier 2009.

Rapporteurs :

- Pr. Kadi Bouatouch (IRISA-Rennes)
- Pr. Bernard Peroche (LIRIS-Lyon)

Examineurs :

- Pr. Denis Hamad (LASL-Calais)

- Dr. Alan Chauvin (UPMF-Grenoble)
Directeur et co-directeur :
- Pr. Christophe Renaud (LIL-Calais)
- Co-directeur : Dr. Samuel Delepouille (LIL-Calais)

2.4 Publications associées

1. hal-03479491, v1
2. Takouachet, N., Delepouille, S., Renaud, C., Zoghلامي, N. & Tavares, J. M. R. (2017). Perception of noise and global illumination : Toward an automatic stopping criterion based on SVM. *Computers and Graphics*, 69, 49-58. <https://doi.org/10.1016/j.cag.2017.09.008>

2.4.1 Takouachet et al. (2007)

Takouachet, N., Delepouille, S. & Renaud, C. (2007). A perceptual stopping condition for global illumination computations. *Spring Conference on Computer Graphics*. <https://hal.archives-ouvertes.fr/hal-03479491>

Résumé

L'objectif de la synthèse d'images réalistes est de produire des images de haute fidélité qui représentent authentiquement des scènes réelles. Comme ces images sont produites pour des observateurs humains, nous pouvons exploiter le fait que tout n'est pas perçu lorsque l'on regarde une scène. Ainsi, il est clair que tirer parti de la capacité limitée du système visuel humain (HVS) peut contribuer de manière significative à l'optimisation des logiciels de rendu. Les méthodes d'illumination globale sont utilisées pour simuler un éclairage réaliste dans les scènes 3D. Elles fournissent généralement une convergence progressive vers une solution de haute qualité. L'un des problèmes de ces algorithmes est de déterminer une condition d'arrêt, pour décider si les calculs ont atteint une convergence satisfaisante permettant au processus de se terminer. Dans cet article, nous proposons et discutons différentes solutions à ce problème. Nous montrons différentes techniques basées sur le Visual Difference Predictor (VDP) proposé par Daly [Daly 1993] pour définir une condition d'arrêt perceptive pour les calculs de rendu. Nous utilisons le VDP pour mesurer les différences perçues entre les images rendues et pour guider le rendu du Path Tracing afin de satisfaire une qualité perceptuelle. De plus, dans un cadre expérimental contrôlé avec des sujets réels, nous validons nos résultats.

3 Mémoire bibliographique de DEA de Pierre Marie Oumoumsack

Direction du mémoire bibliographique de DEA de Pierre Marie Oumoumsack à propos du MARL (Multi-Agent Reinforcement Learning).

4 Thèse de Thuy Tram Ngo Thi

Codirection de la thèse de Thuy Tram Ngo Thi, Thèse de Doctorat en Informatique, Université du Littoral Côte d'Opale. L'encadrement a été assuré avec Christophe Renaud et Remi

Synave. Suite à des problèmes personnels rencontrés par la candidate, la thèse n'a pas pu être soutenue.

5 Thèse de Jérôme Buisine

J'ai co-encadré avec Christophe Renaud le travail doctoral de Jérôme Buisine dans le cadre du financement par l'ANR (contrat PrISE-3d). (titre : « Méthodes d'apprentissage automatique pour la prise en compte du bruit dans les images de synthèse »), thèse soutenue le 8 décembre 2021.

5.1 Contexte

5.2 Résumé

Les méthodes de simulation de l'éclairage, utilisées en synthèse d'images, permettent d'obtenir des vues dites photo-réalistes d'environnements virtuels 3D. Pour ce faire, elles utilisent des méthodes stochastiques, s'appuyant sur la théorie des grands nombres, qui explorent l'espace des chemins lumineux et se caractérisent par une convergence progressive de l'image vers la solution. Cette progressivité se traduit visuellement par la présence de bruit, qui se résorbe progressivement au fur et à mesure de l'avancée des calculs. Ce bruit doit être identifié et quantifié, afin de disposer de critères perceptifs permettant d'arrêter les algorithmes dans les différentes zones de l'image. Ceci est d'autant plus important que les temps de calcul d'une image se comptent en heures, voire en dizaines d'heures de calcul. Disposer de critères fiables pour arrêter les calculs en différents points d'une image permettrait donc de réaliser des gains de temps importants. Dans cette thèse, nous proposons d'utiliser des méthodes statistiques et d'apprentissage automatique pour la réduction et détection de ce bruit généré. Les contributions réalisées dans le cadre de cette thèse sont : (i) la constitution d'une base d'images de synthèse avec recueil de seuils subjectifs humains du bruit résiduel, (ii) l'étude et la gestion d'un bruit local hautement perceptible, (iii) la création de modèles d'apprentissage profond sur cette base d'images étiquetées et (iv) une phase de validation des images reconstruites obtenues (appries ou non) à partir des modèles de perception à partir d'évaluations subjectives. Des travaux connexes à la thématique de la thèse, notamment relatifs à la gestion d'un bruit spécifique dans les images nommé "firefly", ont été proposés, tout comme l'application d'une méthode permettant de cibler les caractéristiques de bruit étudiés.

5.3 Soutenance

La thèse a été soutenue le 8 décembre 2021.

Rapporteurs :

- Pr. Laëtitia JOURDAN, Université de Lille (Rapporteuse)
- Pr. Daniel MENEVEAUX, Université de Poitiers (Rapporteur)

Examineurs :

- Pr. Philippe PREUX, Université de Lille (Examineur)
- Pr. Matéu SBERT, Université de Gironne (Examineur)
- Pr. Sébastien VEREL, Université du Littoral Côte d'Opale (Examineur)

Directeur et co-encadrant :

— Pr. Christophe Renaud (LIL-Calais)

— Dr. Samuel DELEPOULLE, Université du Littoral Côte d’Opale (Encadrant)

Résumé :

5.4 Publications associées

5.4.1 Buisine, Teytaud et al. (2021)

Buisine, J., Delepouille, S. & Renaud, C. (2021b). Minimalist And Customisable Optimisation Package. *Journal of Open Source Software*, 6. <https://doi.org/10.21105/joss.02812>

Résumé

Les problèmes d’optimisation sont fréquemment rencontrés dans les domaines scientifiques et industriels. Étant donné une fonction f à valeurs réelles définie sur un ensemble appelé espace de recherche X , optimiser la fonction f consiste à trouver un point $x \in X$ qui a la valeur optimale $f(x)$, ou au moins à construire une séquence $(x_t)_{t \in \mathbb{N}} \in X^{\mathbb{N}}$ qui est proche de l’optimum. En fonction de l’espace de recherche X , les problèmes d’optimisation peuvent être globalement classés en problèmes discrets (par exemple $X = \{0, 1\}^n$) ou en problèmes continus (par exemple $X = \mathbb{R}^n$). Des outils de modélisation et de résolution de problèmes discrets (Soni, 2017) et continus (Agarwal et al., 2020) ont été proposés dans la littérature.

5.4.2 Buisine, Delepouille et Renaud (2021b)

Buisine, J., Bigand, A., Synave, R., Delepouille, S. & Renaud, C. (2021). Stopping Criterion during Rendering of Computer-Generated Images Based on SVD-Entropy. *Entropy*. <https://hal.archives-ouvertes.fr/hal-03100109>

Résumé

L’estimation de la qualité des images et de la perception du bruit reste un problème important dans diverses applications de traitement d’images. Elle est également devenue un sujet d’actualité dans le domaine de l’infographie photo-réaliste où le bruit est inhérent au processus de calcul. Contrairement aux images de scènes naturelles, il n’existe pas d’image de référence pour les images générées par ordinateur. Ainsi, les méthodes classiques pour évaluer la quantité de bruit et le critère d’arrêt pendant le processus de rendu ne sont pas utilisables. Ceci est particulièrement important dans le cas des méthodes d’illumination globale basées sur des techniques stochastiques : Elles fournissent des images photo-réalistes qui sont toutefois corrompues par le bruit stochastique. Ce bruit peut être réduit en augmentant le nombre de chemins, comme le prouve la théorie de Monte Carlo, mais le problème de trouver le bon nombre de chemins nécessaires pour que les observateurs humains ne puissent percevoir aucun bruit reste entier. Jusqu’à présent, les caractéristiques participant à l’évaluation humaine de la qualité de l’image et le bruit restant perçu ne sont pas connus précisément. La génération d’images de synthèse a tendance à être très coûteuse et les ensembles de données produits sont des ensembles de données à haute dimension. Dans ce cas, trouver un critère d’arrêt en utilisant un cadre d’apprentissage est une tâche difficile. Dans cet article, une nouvelle méthode de caractérisation du bruit résiduel Monte-Carlo pour les images générées par ordinateur est présentée. Le bruit est représenté par l’entropie de la décomposition en valeurs singulières de chaque bloc composant une image. Ces valeurs d’entropie de la décomposition en valeurs singulières (SVD) sont ensuite

utilisées comme entrée d'un modèle d'architecture de réseau neuronal récurrent afin d'extraire le bruit de l'image et de prédire un seuil de convergence visuelle de différentes parties d'une image. Ainsi, une nouvelle évaluation de la qualité des images sans référence est proposée en utilisant la relation entre la SVD-Entropie et la qualité perceptive, basée sur une séquence d'images bruitées. Les expériences montrent que la méthode proposée, comparée à des scores psycho-visuels expérimentaux, démontre une bonne cohérence entre ces scores et les mesures de critères d'arrêt que nous obtenons.

5.4.3 Buisine, Teytaud et al. (2021)

Buisine, J., Teytaud, F., Delepouille, S. & Renaud, C. (2021). Guided-Generative Network for noise detection in Monte-Carlo rendering. *International Conference On Machine Learning And Applications*. <https://hal.archives-ouvertes.fr/hal-03374214>

Résumé

L'estimation des caractéristiques à extraire d'une image pour des tâches de classification est parfois difficile, surtout si les images sont liées à un type de bruit particulier. L'objectif de cet article est de proposer une architecture de réseau de neurones nommée Guided-Generative Network (GGN) pour extraire des informations précises permettant de quantifier correctement le bruit présent dans une fenêtre glissante d'images. Le GGN tend à trouver les caractéristiques souhaitées pour aborder un tel problème afin de sélectionner un critère de détection de ce bruit. Le GGN proposé est appliqué sur des images photoréalistes qui sont rendues par des méthodes Monte-Carlo en évaluant un grand nombre d'échantillons par pixel. Un nombre insuffisant d'échantillons par pixel tend à produire un bruit résiduel qui est très perceptible par un observateur humain. Ce bruit peut être réduit en augmentant le nombre d'échantillons, comme le prouve la théorie de Monte-Carlo, mais cela implique un temps de calcul considérable. Trouver le bon nombre d'échantillons nécessaires pour que les observateurs humains ne perçoivent aucun bruit reste un problème ouvert. Les résultats obtenus montrent que le GGN peut résoudre correctement le problème sans connaissance préalable du bruit tout en étant compétitif avec les méthodes existantes.

5.4.4 Buisine, Delepouille et Renaud (2021a)

Buisine, J., Delepouille, S. & Renaud, C. (2021a). Fireflies removing in Monte Carlo rendering with adaptive Median of meaNs. *Eurographics Symposium on Rendering (2021)*. <https://doi.org/10.2312/sr.20211296>

Résumé

L'estimation de l'équation de rendu à l'aide des méthodes de Monte-Carlo produit des images photoréalistes en évaluant un grand nombre d'échantillons de l'équation de rendu par pixel. La valeur finale pour chaque pixel est alors calculée comme la moyenne de la contribution de chaque échantillon. La moyenne est un bon estimateur, mais pas nécessairement robuste, ce qui explique l'apparition de certains artefacts visuels tels que pixels *lucioles* (*firefly* en anglais), dus à une surestimation de la valeur de la moyenne. L'estimateur MoN (Median of meaNs) est un estimateur plus robuste que la moyenne qui permet de réduire l'impact des valeurs aberrantes qui sont la cause de ces pixels lucioles. Cependant, cette méthode converge plus lentement que la moyenne, ce qui réduit son intérêt pour les pixels dont la distribution ne contient pas de valeurs aberrantes. Pour surmonter ce problème, nous proposons une extension du MoN basée sur le

coefficient de Gini afin d'exploiter le meilleur des deux estimateurs lors du calcul. Cette approche est simple à mettre en œuvre quel que soit l'intégrateur et ne nécessite pas de paramétrage complexe. Enfin, elle présente un surcoût de calcul réduit et conduit à la disparition des pixels lucioles.

6 ANR PrISE-3D

Également dans le cadre de cette collaboration avec SCALab et plus précisément d'une délégation au CNRS en 2016, il a été possible de définir un projet plus ambitieux : le projet PrISE-3D (Perception Interaction et Simulation d'Éclairage 3D »).

6.1 Contexte et positionnement

Ce projet s'inscrit dans le cadre d'une meilleure intégration des spécificités perceptives humaines dans le pilotage de systèmes artificiels. Les outils de production informatiques ont fait d'importants progrès au cours des dernières décennies, en grande partie grâce à l'évolution de la puissance de calcul. Cependant, l'utilisation rationnelle des ressources reste un enjeu important. La simulation d'éclairage est un domaine dans lequel cette problématique est particulièrement flagrante. Si elle permet désormais de produire des images de très haute qualité (dites "photoréalistes", car impossibles à différencier de photographies de scènes réelles), son utilisation en production reste freinée par la puissance de calcul requise par les algorithmes sous-jacents. À titre d'exemple, la production d'une seule image destinée à un film peut représenter plusieurs heures de calculs. Ce temps doit être multiplié par le nombre d'images composant le film et par le nombre de points de vue, dès lors que celui-ci doit être perçu en relief. Ceci aboutit à des temps de production considérables, de l'ordre de plusieurs années de temps CPU ¹²

À l'inverse, les techniques de rendu temps réel, popularisées par les jeux vidéo, reposent sur une utilisation de matériels spécialisés (GPU) et des approximations importantes (modèles 3D, éclairage). Elles ont bénéficié de progrès importants dans des domaines majoritairement liés aux applications interactives (jeux vidéos, simulateurs, réalité virtuelle et augmentée). Elles restent cependant loin de pouvoir prétendre au photoréalisme et le recours à des algorithmes de simulation d'éclairage reste indispensable lorsqu'on souhaite un rendu exact et fidèle d'un point de vue physique et perceptif.

L'obtention de résultats physiquement exacts passe par l'utilisation d'algorithmes stochastiques qui permettent d'explorer l'espace (potentiellement infini) des chemins lumineux transitant, au sein de la scène, entre la caméra et les sources de lumières. L'exploration progressive de cet espace permet d'assurer une convergence visuelle des algorithmes vers l'image finale, les étapes intermédiaires étant porteuses d'informations incomplètes qui se traduisent par du bruit visuel comme présenté dans les deux chapitres précédents, réparti de manière hétérogène sur la surface de de l'image.

6.2 Objectifs

Le projet PrISE-3D s'est fixé pour objectifs principaux :

12. Le temps de rendu du long métrage « *Les nouveaux héros* » de Disney dépasse le million d'heure de temps CPU pour une machine et le stockage représente cinq péta-octets.

1. rechercher à automatiser le processus de détection des seuils de convergence visuelle des algorithmes de simulation d'éclairage stochastique dans le cadre d'images perçues en relief sur une grande variété de périphériques dédiés ;
2. intégrer les données issues de mesures perceptives, en vue de guider les algorithmes de synthèse d'images photoréalistes dans le cadre d'applications interactives (ordonnement des zones de l'images et/ou des objets de la scène à rendre en priorité).

Ces deux objectifs visent alors à réduire les temps de calcul d'une image avec, pour le premier, l'intégration de données perceptives humaines, permettant de ne réaliser que les calculs nécessaires en chaque zone de l'image et, pour le second point, de viser un rendu photoréaliste interactif, en répartissant les calculs en fonction de la tâche courante de l'utilisateur, afin de lui assurer un rendu de qualité sur les zones principales d'attention, tout en étudiant la perception de la dégradation des autres zones de l'image.

Pour atteindre ces objectifs, notre approche consiste à utiliser des systèmes d'apprentissage supervisés de type classifieur, afin d'apprendre la notion de bruit perceptif sur des ensembles d'images de synthèse représentatifs des effets lumineux qui peuvent être présents dans une scène. L'utilisation des dispositifs de restitution en relief (casques de réalité virtuelle, lunettes 3D, écrans auto-stéréoscopiques, salles immersives, etc.), semble induire une restitution du bruit différente entre eux. Cette variabilité est encore accrue avec la possibilité d'utiliser des algorithmes qui réutilisent une partie des calculs entre images composant une vue (auto)stéréoscopique, en vue de réduire leur temps de calcul. Ces techniques provoquent cependant une corrélation du bruit entre images et, par conséquent, une perception très différente de ce dernier. Les modèles perceptifs obtenus pourraient ensuite être intégrés dans le cadre d'applications de rendu interactif 3D, de manière à en déterminer leur niveau d'ergonomie avec des utilisateurs.

Deux laboratoires étaient impliqués dans cette ANR :

- LISIC : l'équipe IMage et APprentissage de ce laboratoire développe des recherches dans le domaine de la simulation d'éclairage depuis plusieurs années. Elle s'intéresse en particulier à la problématique de l'apprentissage de la détection du bruit visuel dans les images de synthèse 3D ;
- SCALab : l'équipe AVA (Action, Vision et Apprentissage) possède une expertise relative aux méthodes psycho-physiques pour les mesures perceptives et l'étude de l'interaction perception-action. En particulier, plusieurs thèses ont été soutenues concernant l'étude psychophysique de la vision pour la perception et l'action, l'étude du liage sensoriel dans le contexte de l'action et l'étude de l'apprentissage perceptif par renforcement ;

6.3 Organisation et moyens

6.3.1 Organisation scientifique

Le travail a été découpé en cinq tâches en plus de la coordination du projet comme illustré en figure 5.21 :

- tâche 1** dont l'objectif principal était de trouver des méthodes efficaces pour le recueil des seuils perceptifs du bruit ;
- tâche 2** qui consistait développer et à comparer des modèles capables de décider de la présence de bruit perceptible ;
- tâche 3** avait essentiellement pour visée de développer des outils d'interaction avec le moteur de rendu ;

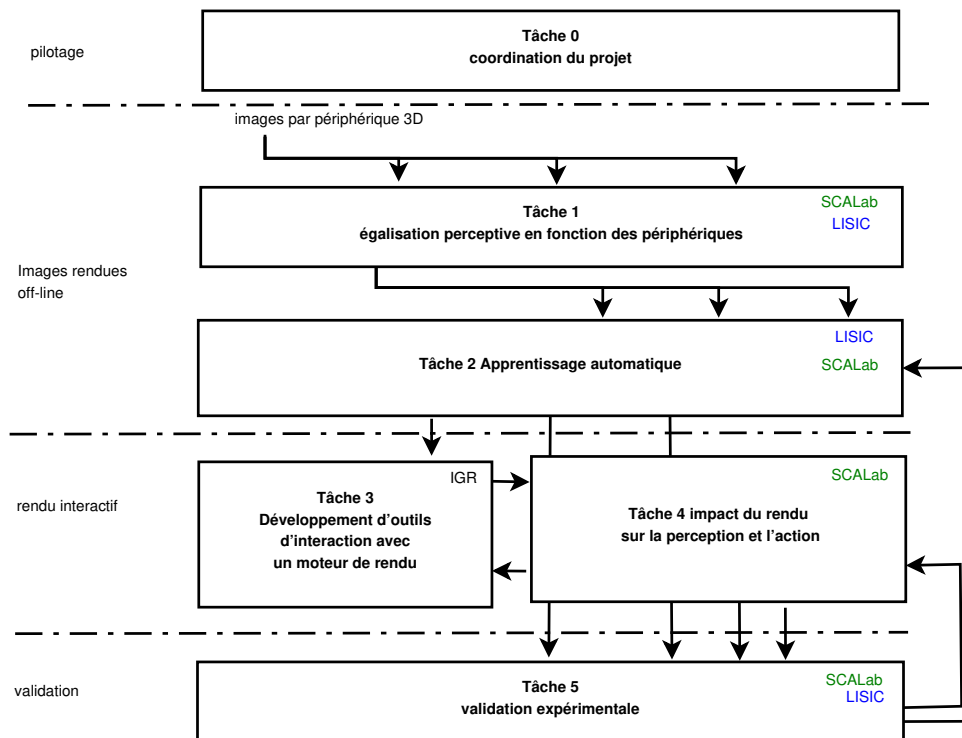


FIGURE 5.21 – Organisation des différentes tâches du projet.

tâche 4 visait à enregistrer l'impact des méthodes de rendu sur la perception et l'action ;

tâche 5 qui était essentiellement un étape de comparaison et de validation des méthodes développés afin de les comparer entre elles et à d'autres méthodes.

6.3.2 Moyens

Le projet a été sélectionné par le CES 38 : *La Révolution numérique : rapports aux savoirs et à la culture* de l'ANR au cours de l'appel à projet générique 2017.

Il a été contractualisé sous la référence ANR-17-CE38-0009. La subvention de l'ANR se montait à 224 640€ pour le LISIC et à 203 040€ pour SCALab soit un montant total de 427 680€. Ceci a permis essentiellement de financer les thèses de Jérôme Buisine et de Vasiliki Myrodiia ainsi que l'embauche de Thomas Boute et Richard Drouard en tant qu'ingénieurs de recherche.

Il a concerné une quinzaine de chercheurs (majoritairement des enseignants-chercheurs). J'ai pris en charge la constitution du dossier puis la coordination en tant que responsable scientifique.

6.4 Bilan et perspectives

6.4.1 Bilan scientifique

Les deux premières tâches qui correspondaient à la méthodologie de recueil de données psychométrique d'une part et à l'apprentissage automatique appliqué au problème ont finalement occupé la plus grande part de la réalisation. Les résultats obtenus ont d'ailleurs fait l'objet de plusieurs publications et de mise à la disposition d'outils communs et de partage de données.

Une relative difficulté a été rencontrée sur la tâche 3 liée au départ à une difficulté de recrutement d'un ingénieur de recherche. Son travail a néanmoins permis d'aboutir à une préversion d'un démonstrateur qui mériterait d'être finalisée. Enfin la problématique sanitaire a rendu délicate son évaluation par des tests subjectifs. Pour ces raisons, la tâche 4 n'a pas réellement abouti. Nous disposons de de suffisamment d'éléments pour poursuivre les développements dans ce domaine.

Enfin, il a été possible au fur et à mesure des développements de valider expérimentalement les résultats comme cela avait été prévu dans la tâche 5.

6.4.2 Bilan quantitatif

En terme de production scientifique, sept publications ont été réalisées et deux thèses ont été soutenues.

6.4.3 Publications

1. Myrodia, V., Delepouille, S. & Madelain, L. (2020). Foveal and peripheral vision for assessing the quality of computer-generated images. *Journal of Vision*, 20(11), 355. <https://doi.org/10.1167/jov.20.11.355>
2. Buisine, J., Delepouille, S. & Renaud, C. (2021a). Fireflies removing in Monte Carlo rendering with adaptive Median of meaNs. *Eurographics Symposium on Rendering (2021)*. <https://doi.org/10.2312/sr.20211296>
3. Buisine, J., Teytaud, F., Delepouille, S. & Renaud, C. (2021). Guided-Generative Network for noise detection in Monte-Carlo rendering. *International Conference On Machine Learning And Applications*. <https://hal.archives-ouvertes.fr/hal-03374214>
4. Buisine, J., Bigand, A., Synave, R., Delepouille, S. & Renaud, C. (2021). Stopping Criterion during Rendering of Computer-Generated Images Based on SVD-Entropy. *Entropy*. <https://hal.archives-ouvertes.fr/hal-03100109>
5. Buisine, J., Delepouille, S. & Renaud, C. (2021b). Minimalist And Customisable Optimisation Package. *Journal of Open Source Software*, 6. <https://doi.org/10.21105/joss.02812>
6. Buisine, J., Delepouille, S., Synave, R. & Renaud, C. (2021). Subjective human thresholds over computer generated images. <https://doi.org/10.5281/ZENODO.4531460>
7. Myrodia, V., Buisine, J. & Madelain, L. (2021). Comparison of threshold measurements in laboratory and online studies using a Quest+ algorithm. *Journal of Vision*, 21(9), 1959. <https://doi.org/10.1167/jov.21.9.1959>

6.4.4 Thèses

1. Buisine, J. (2021). *Méthodes d'apprentissage automatique pour la prise en compte du bruit dans les images de synthèse* (thèse de doct.) [Thèse de doctorat dirigée par Renaud, Christophe et Delepouille, Samuel Informatique Littoral 2021]. Université du Littoral Côte d'Opale
2. Myrodia, V. (2021). *Études psychophysiques sur la perception visuelle du bruit de rendu de Monte Carlo* (thèse de doct.) [Thèse de doctorat dirigée par Madelain, Laurent, Psychologie 2021]. Université de Lille

6.5 Perspectives

La réalisation de ce projet a amené à des développements qu'il nous paraît intéressant de poursuivre. Il s'agira dans un premier temps de finaliser les travaux d'optimisation et d'intégration des différentes méthodes de rendu.

Dans un second temps, il nous paraît important d'explorer certaines questions que nous avons identifiées. C'est le cas notamment de l'application aux séquences d'images de synthèse et à l'affichage HDR.

7 Publications

Thématique :

IA Intelligence artificielle ;

SI : Synthèse d'images ;

PV : Perception visuelle.

Nature :

R Revue ;

C : Conférence ;

A : Autre.

publication	IA	SI	PV	Nature
Delepoulle, S., Preux, P. & Darcheville, J.-C. (2001). Selection of Behavior in Social Situations Application to the Development of Coordinated Movements. <i>Lecture Notes in Computer Science</i> (p. 384-393). Springer Berlin Heidelberg. https://doi.org/10.1007/3-540-45365-2_40	•			R
Delepoulle, S., Preux, P. & Darcheville, J.-C. (2002). Learning as a Consequence of Selection. <i>Lecture Notes in Computer Science</i> (p. 350-361). Springer Berlin Heidelberg. https://doi.org/10.1007/3-540-46033-0_28	•			R
Montagne, F., Delepoulle, S. & Preux, P. (2003). A critic-critic architecture to combine reinforcement and supervised learnings. <i>European Workshop on reinforcement Learning</i>	•			C
Montagne, F., Preux, P. & Delepoulle, S. (2006). Introducing interactive help for reinforcement learners, Workshop on planning, learning and monitoring with uncertainty and dynamic worlds. <i>ECAI Workshop</i>	•			C
Takouachet, N., Delepoulle, S. & Renaud, C. (2007). A perceptual stopping condition for global illumination computations. <i>Spring Conference on Computer Graphics</i> . https://hal.archives-ouvertes.fr/hal-03479491		•	•	C
Delepoulle, S., Renaud, C. & Preux, P. (2009). Light Source Storage and Interpolation for Global Illumination : A Neural Solution. <i>Studies in Computational Intelligence</i> (p. 87-104). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-03452-7_5	•	•		R
Loth, M., Preux, P., Delepoulle, S. & Renaud, C. (2009). ECON : A Kernel Basis Pursuit Algorithm with Automatic Feature Parameter Tuning, and its Application to Photometric Solids Approximation. <i>2009 International Conference on Machine Learning and Applications</i> . https://doi.org/10.1109/icmla.2009.61	•	•		C

publication	IA	SI	PV	Nature
Bartolo, A., Coello, Y., Delepouille, S., Edwards, M., Endo, S. & Wing, A. (2009). Neurobiological basis of reachability judgement : an fMRI study. <i>NeuroImage</i> , 47, S171. https://doi.org/10.1016/s1053-8119(09)71841-5		•	•	C
Chelle, D. M., Renaud, C., Delepouille, S. & Combes, D. (2009). Light phylloclimate within growth chambers : a modeling approach. <i>Comparative Biochemistry and Physiology Part A : Molecular & Integrative Physiology</i> , 153(2), S229. https://doi.org/10.1016/j.cbpa.2009.04.635	•	•		R
Delepouille, S., Renaud, C. & Chelle, M. (2009). Improving Light Position in a Growth Chamber through the Use of a Genetic Algorithm. <i>Artificial Intelligence Techniques for Computer Graphics</i> (p. 67-82). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-540-85128-8_5	•	•		R
Renaud, C., Delepouille, S. & Takouachet, N. (2012). Detecting Visual Convergence for Stochastic Global Illumination. <i>Intelligent Computer Graphics 2011</i> (p. 1-17). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-22907-7_1	•	•	•	R
Delepouille, S., Bigand, A. & Renaud, C. (2012). A no-reference computer-generated images quality metric and its application to denoising. <i>2012 6th IEEE INTERNATIONAL CONFERENCE INTELLIGENT SYSTEMS</i> . https://doi.org/10.1109/is.2012.6335116	•	•		C
Péroche, B., Delepouille, S. & Renaud, C. (2013). Rendering and Computer Graphics. <i>Digital Color</i> (p. 93-118). John Wiley & Sons, Inc. https://doi.org/10.1002/9781118562680.ch4		•	•	A
Constantin, J., Delepouille, S., Bigand, A. & Renaud, C. (2013). Reduced-reference quality assessment of computer-generated images based on RVM. <i>2013 Third International Conference on Communications and Information Technology (ICCIT)</i> . https://doi.org/10.1109/iccitechnology.2013.6579572	•	•		C
Bartolo, A., Coello, Y., Edwards, M. G., Delepouille, S., Endo, S. & Wing, A. M. (2014). Contribution of the motor system to the perception of reachable space : an fMRI study. <i>European Journal of Neuroscience</i> , 40(12), 3807-3817. https://doi.org/10.1111/ejn.12742			•	R
Delepouille, S., Bigand, A., Renaud, C. & Colot, O. (2016). Low-Level Image Processing Based on Interval-Valued Fuzzy Sets and Scale-Space Smoothing. <i>Recent Contributions in Intelligent Systems</i> (p. 1-22). Springer International Publishing. https://doi.org/10.1007/978-3-319-41438-6_1	•	•		R
Takouachet, N., Delepouille, S., Renaud, C., Zoghliami, N. & Tavares, J. M. R. (2017). Perception of noise and global illumination : Toward an automatic stopping criterion based on SVM. <i>Computers and Graphics</i> , 69, 49-58. https://doi.org/10.1016/j.cag.2017.09.008		•	•	R
Myrodiia, V., Delepouille, S. & Madelain, L. (2020). Foveal and peripheral vision for assessing the quality of computer-generated images. <i>Journal of Vision</i> , 20(11), 355. https://doi.org/10.1167/jov.20.11.355		•	•	R
Buisine, J., Delepouille, S. & Renaud, C. (2021a). Fireflies removing in Monte Carlo rendering with adaptive Median of meaNs. <i>Eurographics Symposium on Rendering (2021)</i> . https://doi.org/10.2312/sr.20211296		•		C

publication	IA	SI	PV	Nature
Buisine, J., Teytaud, F., Delepoulle, S. & Renaud, C. (2021). Guided-Generative Network for noise detection in Monte-Carlo rendering. <i>International Conference On Machine Learning And Applications</i> . https://hal.archives-ouvertes.fr/hal-03374214	•	•		C
Buisine, J., Bigand, A., Synave, R., Delepoulle, S. & Renaud, C. (2021). Stopping Criterion during Rendering of Computer-Generated Images Based on SVD-Entropy. <i>Entropy</i> . https://hal.archives-ouvertes.fr/hal-03100109	•	•		C
Buisine, J., Delepoulle, S. & Renaud, C. (2021b). Minimalist And Customisable Optimisation Package. <i>Journal of Open Source Software</i> , 6. https://doi.org/10.21105/joss.02812	•			R
Buisine, J., Delepoulle, S., Synave, R. & Renaud, C. (2021). Subjective human thresholds over computer generated images. https://doi.org/10.5281/ZENODO.4531460		•	•	A
Myrodi, V., Buisine, J. & Madelain, L. (2021). Comparison of threshold measurements in laboratory and online studies using a Quest+ algorithm. <i>Journal of Vision</i> , 21(9), 1959. https://doi.org/10.1167/jov.21.9.1959		•	•	R
Godard, M., Wamain, Y., Ott, L., Delepoulle, S. & Kalénine, S. (2022). How Competition between Action Representations Affects Object Perception during Development. <i>Journal of Cognition and Development</i> , 1-25. https://doi.org/10.1080/15248372.2022.2025808		•	•	R

c1 : 16 8, c2 : 41 30, c3 : 56 10, c4 : 81 15, c5 : 103 4

Résumé

Le contexte scientifique des travaux présentés dans ce manuscrit concerne les apports des données perceptives à la génération d'images de synthèse photo-réalistes.

Tout d'abord un tour d'horizon des problématiques liées à la perception est effectué, en s'attachant à identifier les variables impliquées lors de la perception d'une image, numérique ou non. Certains paramètres sont à prendre en compte, au premier lieu desquels l'acuité visuelle et la sensibilité au contraste. La question de la perception 3D pour sa part soulève des problèmes théoriques et pratiques qui sont explorés par le biais des indices de perception de la profondeur : les indices monoculaires qui sont connus des peintres et artistes graphiques ainsi que les indices binoculaires impliqués dans la vision stéréoscopique. Nous nous focalisons ensuite sur la problématique générale de la simulation d'éclairage, pour laquelle l'utilisation des algorithmes de Monte-Carlo permet de produire des images photo-réalistes. Ces méthodes se basent sur une modélisation des sources de lumière, des matériaux, de la géométrie et du transfert de la lumière dans une scène virtuelle. Bien qu'elles permettent d'obtenir des images de très grande qualité, elles se caractérisent par une convergence relativement lente et induisent l'apparition d'artefacts visuels. Un bruit de haute fréquence spatiale est notamment visible dans les premières phases du calcul. Celui-ci s'estompe avec le nombre d'échantillons mais tend à rester très perceptible. De plus, il est difficile de déterminer un critère d'arrêt précis du calcul.

Nos contributions se focalisent sur ces artefacts, en utilisant des données réelles issues d'expériences utilisateurs afin de déterminer où ils sont visibles dans une image et jusqu'à quel niveau les calculs sont nécessaires pour les rendre imperceptibles. Ces travaux ont été menés dans un cadre transdisciplinaire associant les méthodes et résultats de la psychologie expérimentale aux techniques de calculs d'images numériques.

Laboratoire d'Informatique Signal et Image de la Côte d'Opale
Maison de la Recherche Blaise Pascal
50, rue Ferdinand Buisson
B.P. 719 62228 Calais Cedex

Tél. : 03.21.46.36.53

<http://www-lisic.univ-littoral.fr>