Ecole Nationale Supérieure
des Mines de Saint-Etienne

Université Jean Monnet
de Saint-Etienne

N° d'ordre : 193 ID

THESE

Présentée par Jacques ZANINETTI

pour obtenir le titre de

Docteur

DE L'UNIVERSITE JEAN MONNET DE SAINT-ETIENNE ET DE
L'ECOLE NATIONALE SUPERIEURE DES MINES DE SAINT-ETIENNE

Spécialité Informatique, Synthèse d'images

Modélisation vectorielle de l'éclairement global en lancer de rayons

Soutenue à Saint-Etienne, le 19 octobre 1998

Composition du jury :

Frederik W. Jansen Claude Puech	Rapporteurs
Didier Arques Marc Fontoynont Bernard Péroche Alain Trémeau	Examinateurs

---

Remerciements

Je tiens tout particulièrement à remercier Bernard Péroche, qui a dirigé ces travaux de thèse, pour ses innombrables conseils bien sûr, mais aussi et surtout pour sa confiance et son amitié. Son aide continuelle et son expérience en ce domaine furent irremplaçables. De plus, les articles en anglais n'auraient certainement pas existé sans ses traductions.

Je remercie Messieurs Frederik Jansen et Claude Puech pour avoir accepté d'être les rapporteurs de cette thèse, et Messieurs Didier Arques, Marc Fontoynont, et Alain Trémeau pour avoir accepté d'être membres du jury. Leurs remarques et critiques constructives furent particulièrement utiles.

Je remercie tous les membres du Laboratoire d'Images de Synthèse de Saint-Etienne, pour leur accueil chaleureux, en commençant par Marie-Line Barnéoud, secrétaire pour qui les méandres sinueux de l'administration n'ont aucun secret, et dont les tartes aux pommes sont inoubliables. Dominique Michelucci, Jean Michel Moreau, Marc Roelens et Grégory Six ont su partager leur grande expérience en infographie, et aussi offrir leur amitié et être complices d'innombrables discussions animées sur bien d'autres sujets. Jean-Luc Maillot, qui m'a beaucoup aidé en stage de DEA, a aussi été un bon exemple de rigueur scientifique. Je décerne une mention particulière à Marion Personne, toujours ponctuelle à l'heure du café, qui me supporte dans le bureau depuis le début, et puis à Hélymar Balza-Gomez, Mireille Batton-Hubert et Antoine Pauze, sympathiques compagnons des repas-causettes de midi. Enfin, je souhaite un bon courage à tous ceux qui continuent leur thèse.

Je remercie Laurent Carraro et les membres de l'équipe de Mathématiques pour m'avoir aimablement aidé à résoudre certains problèmes théoriques.

Je remercie aussi les autres services de l'école, notamment Hélène Sayet, de la bibliothèque, pour avoir accepté aussi gentiment mes oublis systématiques de la date de restitution des ouvrages empruntés, et Denise Digat, de l'administration, qui ne perd pas pied parmi des monceaux de dossiers plus ou moins complets. Et puis je félicite chaleureusement Pierre Igier et Pierre Rullière, pour avoir mené à bien l'impression de ce document, malgré mes retards, les formats de fichiers incompatibles entre-eux, et les caprices des machines de reproduction.

Finalement, je remercie tous ceux, famille et amis, qui m'ont encouragé et permi de reprendre tardivement des études, et d'achever cette thèse dans de bonnes conditions, en premier lieu mes parents, ma sœur et son mari.

Chapitre 1

Introduction

1.1 Bref aperçu historique

La synthèse d'image est un domaine en constante progression depuis quelques années. Les formidables évolutions de l'informatique, tant matérielles que logicielles, permettent des réalisations de plus en plus sophistiquées. Nul ne saurait aujourd'hui se satisfaire de représentations en "fil de fer", simples visualisations des arêtes des objets. Pourtant, il y a peu encore elles étaient à l'origine de l'essor des nombreuses méthodes de la CAO (Conception Assistée par Ordinateur).

Les premières images de synthèse ne représentaient que des objets relativement simples isolés de leur contexte, des pièces mécaniques en général. La notion de *réalisme* était tout à fait secondaire, seule importait la visualisation de nouvelles pièces sans devoir réaliser des maquettes très coûteuses, en temps comme en argent. Le besoin d'observer des assemblages plus complexes est cependant rapidement apparu. Les modélisations en "fil de fer" se sont alors très vite révélées trop limitées. Les images devenaient d'inextricables enchevêtrements de segments.

Un premier progrès a été l'élimination des arêtes cachées, c'est à dire celles qui sont "derrière" les objets d'avant plan. Ceci était tout aussi complexe que peu efficace. Il était devenu nécessaire de doter les images d'une réelle notion de relief. La première idée fut de colorier les facettes délimitées par les arêtes, en attribuant une couleur différente à chaque objet. Les images étaient plus compréhensibles, mais toujours sans relief et aussi peu réalistes. Les simples effets de perspective ne sont pas suffisants pour permettre une bonne interprétation des scènes représentées.

L'introduction d'effets d'ombrage fut un pas décisif. Les premières méthodes, tout à fait empiriques, donnèrent des résultats encourageants. Les surfaces apparaissaient plus ou moins claires, suivant leur inclinaison par rapport à une source de lumière imaginaire. L'augmentation incessante des capacités de traitement des ordinateurs permit la simulation des surfaces courbes par la juxtaposition d'un grand nombre de petites facettes planes. Des algorithmes de lissage furent alors développés pour atténuer l'effet de cassure visible entre les facettes adjacentes, notamment par Gouraud [Gou71] et par Phong [Pho75]. Le réalisme était bien meilleur, mais les ombres portées restaient totalement absentes, et les matériaux avaient tous un aspect uniforme et gênant de "matière plastique".

L'apparition du *lancer de rayons* [Whi80], avec des effets de reflets et de transparences, puis celle de la *radiosité* [GTGB84], avec des pénombres douces et une approche plus globale des échanges lumineux, ouvrirent de nouveaux horizons. Pour la première fois des images de synthèse semblaient presque "vraies". Le cinéma et la publicité s'emparèrent de ces nouvelles méthodes et les rendirent très populaires.

Ces progrès étaient toutefois très insuffisants pour d'autres domaines fortement intéressés par le développement d'images de synthèse, comme l'architecture, et plus particulièrement l'éclairagisme. Il devenait alors indispensable de "voir" un bâtiment dans son environnement avant sa construction, et de pouvoir "essayer" plusieurs styles de décorations différents. Les modèles approximatifs ne sont plus utilisables, car les images produites sont souvent fort éloignées de l'effet escompté. Il ne suffit plus qu'une image soit "belle", elle doit aussi être fidèle à ce que serait la scènes si elle existait. Cette notion de réalisme, sujette à diverses interprétations, sera précisée dans la section suivante.

La dernière étape décisive fut franchie par Kajiya [Kaj86], qui introduisit l'*équation de rendu*. C'était le premier modèle mathématique et physique complet décrivant la répartition de l'énergie lumineuse dans un environnement, présenté sous la forme d'une équation unique, mais récursive. Il laissaient entrapercevoir une possible mise en œuvre, coûteuse mais enfin envisageable. Le calcul de l'*éclairement global*, où tous les paramètres sont pris en compte, devenait alors un objectif accessible, même s'il était encore lointain. Dans le même temps, la modélisation des matériaux devenait de plus en plus précise, et la plupart des phénomènes physiques pouvaient être simulés, au moins en théorie.

Il apparut alors que la complexité du problème est immense, tant du point de vue de la description des objets et des matériaux qui les composent, que de celui des interactions de la lumière avec eux. Les méthodes classiques, le lancer de rayon et la radiosité dans leurs déclinaisons initiales, ne sont pas capables d'appréhender le problème dans son ensemble. Elles sont toutes les deux basées sur des simplifications drastiques qui induisent des erreurs difficiles à évaluer.

Malgrè leurs défauts, le lancer de rayon et la radiosité donnent des images plutôt satisfaisantes, au moins visuellement. De plus, ils sont complémentaires. Le lancer de rayons privilégie les réflexions spéculaires, alors que la radiosité gère les réflexions diffuses. Des *méthodes multipasses* furent alors développées pour combiner les deux approches, inspirées par Wallace [WCG87]. Schématiquement, une première passe de radiosité calcule les interréflexions diffuses, et une seconde passe de lancer de rayons ajoute les reflets spéculaires. Avec le *lancer de rayons inverse*, Arvo [Arv86] introduisit les effets de caustiques, avec une passe supplémentaire distribuant l'énergie des sources dans la scène. L'échantillon des phénomènes visualisables devenait presque complet, mais toujours avec les simplifications des méthodes initiales qui interdisent toute validation physique.

Seules les *méthodes de Monte Carlo*, dont Kajiya [Kaj86] fut l'initiateur en synthèse d'images, semblent en mesure de relever le défi, car elles n'ont pas de limites théoriques. Par contre, les quantités de calculs sont gigantesques, et par conséquent les temps de calcul sont prohibitifs pour encore plusieurs générations d'ordinateurs.

Un grand nombre de travaux de recherche actuels en synthèse d'images tentent d'étendre la gamme des phénomènes simulés, tout en respectant plus fidèlement les lois physiques. En parallèle, des techniques d'optimisations sont développées pour réduire la durée des calculs, sans nuire à la qualité des résultats.

1.2 A propos de réalisme

Tout au long de ce document, nous emploierons souvent le terme *réalisme*, en tant qu'objectif, et aussi en tant que critère de qualité. Cette notion étant sujette à diverses interprétations [Mai96], nous allons préciser le sens que nous lui donnons.

Avant tout, nous nous plaçons dans le cadre de la simulation de phénomènes physiques. L'objectif est de calquer le plus fidèlement possible les comportements physiques de la lumière et de ses interactions avec la matière. Il s'agit donc en premier lieu de *physicoréalisme*. Cependant, ne sachant pas résoudre tous les problèmes, nous limitons notre ambition à la simulation de scènes de la vie courante, c'est à dire composées d'objets de dimensions moyennes, pour un observateur humain. La taille des détails géométriques des objets est alors très largement supérieure aux longueurs d'ondes de la lumière visible, ce qui élimine les phénomènes de diffraction, même s'ils interviennent parfois de façon implicite dans la modélisation des propriétés des matériaux. Ces hypothèses sont tout à fait valides dans le domaine de la simulation architecturale.

Dans ces conditions, notre objectif ultime serait de recréer, chez l'observateur humain, les mêmes sensations devant une image de synthèse d'une scène que celles qu'il aurait s'il se trouvait dans la scène réelle. Il est bien évident que ce but est utopique avec une image plane, quels que soient sa qualité et son support. Plus modestement, ce serait déjà un excellent résultat de créer une image de synthèse d'une scène qui soit indiscernable d'une photographie de la scène réelle. Nous nous situons donc plutôt dans le cadre du *photoréalisme*. En fait, ce dernier terme manque lui-aussi de précision. En effet, nos images sont pour la plupart calculées pour être affichées sur un écran vidéo, avec une résolution immensément plus faible que celle d'une photographie, une dynamique énergétique limitée, et avec un mode de reproduction des couleurs différent et moins nuancé. La comparaison des résultats est donc faussée. Il serait donc préférable, dans un premier temps, de rechercher le meilleur *vidéoréalisme*[1]. Il s'agit donc pour nous de créer des images de synthèse indiscernables de reproductions par une caméra vidéo de la scène. Bien que les progrès soient incontestables, cet objectif plus réaliste n'est toujours pas atteint aujourd'hui.

1.3 Objectifs de cette thèse

Durant cette thèse, nous nous sommes attachés à développer une modélisation vectorielle de la lumière, qui unifie les représentations de ses diverses composantes. Elle regroupe de façon concise des informations énergétiques et directionnelles qui sont nécessaires aux calculs de réflexions ou de réfractions. Nous l'utilisons pour calculer l'éclairement global, en intégrant des méthodes de calcul de Monte Carlo.

Afin de diminuer les temps de calculs, nous avons élaboré des procédures d'interpolation, qui exploitent les cohérences de la lumière et de la scène. Elles permettent de réutiliser des valeurs calculées ponctuellement, en leur attribuant un domaine de validité étendu. Des critères limitant les risques d'erreur ont été introduits.

De plus, nous avons établi une stratégie de subdivision adaptative des sources surfaciques. A qualité équivalente, elle réduit sensiblement le taux d'échantillonnage moyen par rapport aux méthodes habituelles.

1. Ce mot est certainement un néologisme.

Ces résultats ont été implémentés dans un logiciel de lancer de rayons [Roe93], pour valider les approches choisies.

1.4 Organisation de ce document

- Le chapitre 1 est la présente introduction, où sont décrits les principaux objectifs de ces travaux.
- Le chapitre 2 est une présentation des ondes électromagnétiques, et du cas particulier qu'est la lumière. Les grandeurs physiques et les unités utilisées dans toute la suite sont explicitées ici. Des notions psychovisuelles succintes sont abordées, pour décrire la sensation de couleur.
- Le chapitre 3 présente l'équation de rendu, qui formalise en quelques symboles l'ampleur du problème à résoudre, et les principaux modèles d'éclairement global actuels qui tentent de le faire, avec plus ou moins de réussite.
- Le chapitre 4 décrit en détail les *vecteurs lumineux*, qui sont au cœur de notre méthode de calcul de l'éclairement global. La lumière est divisée en trois composantes indépendantes, suivant le type de chemin lumineux parcouru pour atteindre un point de la scène : direct, indirect, et caustique. Trois types distincts de vecteurs lumineux sont alors définis, dont les particularités permettent des optimisations spécifiques. Des procédures d'interpolation, adaptées aux trois composantes de la lumière, exploitent les cohérences de la lumière pour éviter les opérations inutiles ou redondantes, et donc limiter la durée des calculs.
- Le chapitre 5 traite des sources lumineuses utilisées en lancer de rayons, et propose une méthode optimisée de subdivision adaptative des sources surfaciques, ainsi qu'une extension à la source naturelle qu'est le ciel.
- Le chapitre 6 décrit brièvement le cadre d'implémentation de notre méthode, et présente quelques résultats commentés.
- Le chapitre 7 est la conclusion, qui récapitule les divers éléments présentés auparavant, et propose quelques sujets de développement qui compléteraient utilement ces travaux.

Les notions théoriques importantes, les développements secondaires, et les démonstrations mathématiques un peu longues, qui auraient inutilement alourdi le texte (déjà confus parfois), sont présentés sous la forme d'annexes indépendantes.

- L'annexe A regroupe quelques rappels de probabilités et présente les méthodes de Monte Carlo, ainsi que les principales techniques de réduction de variance, notamment celles utilisées en infographie.
- L'annexe B est la démonstration mathématique un peu lourde du calcul du gradient d'éclairement indirect, que nous utilisons pour affiner les interpolations dans notre modélisation vectorielle.

- L'annexe C décrit les divers problèmes d'aliassage, hélas incontournables en synthèse d'images. De plus, nous décrivons la méthode d'antialiassage en post-traitement utilisée pour corriger la plupart des images présentées dans les autres sections.

- L'annexe D explicite la notion d'angle solide, qui apparaît dans pratiquement toutes les formulations des grandeurs physiques décrivant les comportements des ondes électromagnétiques. Nous donnons ensuite quelques formules de calculs analytiques, dans les cas simples où cela est possible.

Chapitre 2

La lumière et les matériaux

L'étude de la lumière et de ses interactions avec la matière est primordiale en synthèse d'images. Ce sujet est si vaste qu'il ne nous est pas possible de le traiter exhaustivement, aussi allons nous dans ce chapitre nous limiter à une description des phénomènes fondamentaux qui nous sont indispensables. Plusieurs ouvrages décrivent avec beaucoup plus de détails ces phénomènes, notamment [Gla95] et [PGMR98]. Nous nous attacherons autant que possible à définir sans ambiguïté les grandeurs, les unités, et les notations que nous utiliserons par la suite.

2.1 Nature de la lumière

La lumière est un rayonnement énergétique, auquel est sensible le système visuel humain. Elle possède la particularité de pouvoir s'interpréter selon deux aspects différents : l'aspect *ondulatoire* et l'aspect *corpusculaire*. Leur coexistence permet d'expliquer les comportements particuliers de cette énergie et ses interactions avec la matière.

2.1.1 Aspects ondulatoires

2.1.1.1 Ondes électromagnétiques

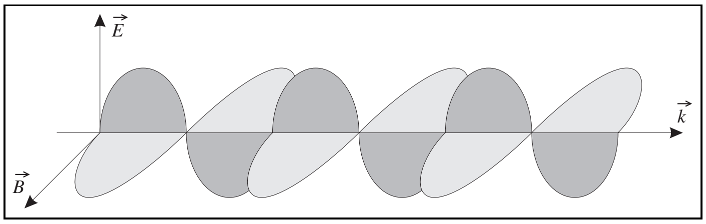


FIG. 2.1 – *Onde électromagnétique*

La lumière est composée d'ondes électromagnétiques, dont la fréquence détermine la sensation de couleur (section 2.8.3). Cet aspect ondulatoire de la lumière doit être invoqué pour interpréter les phénomènes d'interférences et de diffraction.

Les ondes lumineuses (figure 2.1) sont composées d'un champ électrique $\vec{E}$ et d'un champ magnétique $\vec{B}$. Ces deux champs se propagent dans la même direction $\vec{k}$ à la même vitesse c, d'environ 300.000 km/s dans le vide. A l'intérieur des matériaux, cette vitesse est plus faible et varie avec la longueur d'onde. Ils oscillent à la même fréquence, en phase, et restent perpendiculaires l'un de l'autre. La propagation est rectiligne dans les milieux uniformes, où l'indice de réfraction n reste constant.

L'évolution au cours du temps des directions des champs $\vec{E}$ et $\vec{B}$ confère des propriétés particulières aux ondes lumineuses. On distingue deux cas principaux :

- **Ondes polarisées rectilignement :** les champs $\vec{E}$ et $\vec{B}$ gardent un direction constante au cours de la propagation. Le plan $(\vec{E}, \vec{k})$ est par définition le *plan de polarisation.* L'*angle de polarisation* est l'angle formé par $\vec{E}$ avec un vecteur de référence fixe.

- **Ondes non polarisées :** c'est le cas général très important en synthèse d'images de la lumière naturelle, et aussi de la plupart des lumières artificielles. Une onde non polarisée est la superposition d'un très grand nombre d'ondes polarisées rectilignement, sans corrélation entre elles, sinon qu'elles se propagent dans la même direction. Elles sont de longueurs d'onde différentes lorsque la lumière est polychromatique.

La position du plan de polarisation par rapport à l'angle d'incidence de la lumière est aussi un paramètre utile lors de l'analyse de la réflexion. Le *plan d'incidence* est le plan engendré par le vecteur directeur de l'onde lumineuse et par le vecteur normal à la surface au point d'incidence. On distingue deux cas importants :

- **Polarisation parallèle :** Le plan de polarisation est parallèle au plan d'incidence.

- **Polarisation perpendiculaire :** Le plan de polarisation est perpendiculaire au plan d'incidence.

2.1.1.2 Equations de Maxwell

Le champ électrique $\vec{E}$ et le champ magnétique $\vec{B}$ sont reliés en tout point de l'espace par les quatre *équations différentielles de Maxwell.* Leur formulation dans le vide est :

$$\left[\begin{array}{ll} rot(\vec{E}) = -\dfrac{\partial \vec{B}}{\partial t} & div(\vec{E}) = 0 \\ \\ rot(\vec{B}) = \mu_0\, \varepsilon_0\, \dfrac{\partial \vec{E}}{\partial t} & div(\vec{B}) = 0 \end{array}\right]$$

avec : $\left| \begin{array}{lll} \mu_0 = 4\pi\; 10^{-7} & \text{henry/m} & : \text{permittivité électrique du vide} \\ \varepsilon_0 = 10^{-9}/36\,\pi & \text{farad/m} & : \text{perméabilité magnétique du vide} \end{array}\right.$

On en déduit les *équations de propagation* des champs $\vec{E}$ et $\vec{B}$ dans le vide :

$$\Delta\, \vec{E} = \mu_0\, \varepsilon_0\, \frac{\partial^2 \vec{E}}{\partial t^2} \qquad \text{et} \qquad \Delta\, \vec{B} = \mu_0\, \varepsilon_0\, \frac{\partial^2 \vec{B}}{\partial t^2}$$

La résolution de ces équations différentielles permettrait de connaître la répartition de l'énergie lumineuse pour tous les points de la scène, et donc de calculer des images exactes. Elle n'est malheureusement pas envisageable en pratique, hormis pour quelques configurations théoriques très simples.

2.1.1.3 Propagation de la lumière dans la matière

Les équations de Maxwell sont toujours valables dans la matière, mais les hypothèses suivantes sont nécessaires pour les résoudre :

- le milieu est globalement neutre : il n'y a accumulation de charges électriques en aucune partie du milieu ;
- le milieu suit la loi d'Ohm pour la conductivité ;
- le milieu est électriquement parfait ;
- le milieu est magnétiquement parfait.

Les trois valeurs suivantes caractérisent la propagation de la lumière dans un matériau :

- ε, la constante diélectrique relative du milieu ($\varepsilon = 1$ pour le vide) ;
- μ, la perméabilité magnétique du milieu ($\mu = 1$ pour le vide) ;
- γ, la conductivité du milieu ($\gamma = 0$ pour les matériaux non-conducteurs).

L'évolution au cours du temps en un point P du champ électromagnétique d'une onde plane monochromatique de direction $\vec{u} = \vec{k}/\|\vec{k}\|$ et de pulsation ω s'exprime par :

$$\vec{E}=\vec{E_0}\ e^{-i\omega\left(t-\frac{\vec{u}.\vec{OP}}{v}\right)} \qquad \text{et} \qquad \vec{H}=\vec{H_0}\ e^{-i\omega\left(t-\frac{\vec{u}.\vec{OP}}{v}\right)}$$

où v est la constante complexe définie par :

$$v = \frac{c}{\sqrt{\varepsilon\mu + i\dfrac{\gamma\mu}{\varepsilon_0\omega}}}$$

2.1.1.4 Indice de réfraction

L'*indice de réfraction* du matériau est défini par :

$$n = \frac{c}{v} = \sqrt{\varepsilon\mu + i\frac{\gamma\mu}{\varepsilon_0\omega}}$$

On peut noter que l'indice de réfraction dépend de la longueur d'onde. Des ondes électromagnétiques de fréquences différentes ne se propagent donc pas de façons identiques dans les matériaux.

L'indice de réfraction est complexe pour les matériaux conducteurs. Sa décomposition en une partie réelle n_r et une partie imaginaire n_i donne :

$$n_r = \sqrt{\frac{\varepsilon\mu}{2} + \frac{\mu}{2}\sqrt{\varepsilon^2 + \frac{\gamma^2}{\varepsilon_0^2\omega^2}}} \qquad \text{et} \qquad n_i = \sqrt{-\frac{\varepsilon\mu}{2} + \frac{\mu}{2}\sqrt{\varepsilon^2 + \frac{\gamma^2}{\varepsilon_0^2\omega^2}}}$$

Le quotient $\dfrac{n_i}{n_r}$ est appelé *indice d'atténuation* ou *coefficient d'extinction* du milieu.

2.1.1.5 Absorption, dispersion

Le module des champs électrique et magnétique décroit le long de l'axe de propagation dans un milieu autre que le vide. C'est le phénomène d'absorption. Il se caractérise avec le coefficient $\delta = \dfrac{c}{n_i\,\omega}$, appelé *longueur d'absorption*. Après un parcours de longueur δ dans le milieu, l'amplitude du champ électromagnétique est diminuée d'un facteur e^{-1}. δ peut être très petit pour les métaux non magnétiques, de l'ordre de $10^{-10}m$. La lumière ne se propage pratiquement pas dans ces milieux.

La vitesse de la lumière dans un milieu dépend de la partie réelle de l'indice de réfraction, et vaut $\dfrac{c}{n_r}$. Comme n_r dépend de la longueur d'onde, cette vitesse varie avec la fréquence de l'onde électromagnétique. C'est le phénomène de *dispersion*, à l'origine de la décomposition spectrale des prismes, et de la formation des arcs-en-ciel par exemple.

2.1.2 Aspects corpusculaires

Sous son aspect *corpusculaire*, la lumière est un flux de *photons*, particules sans masse qui transportent de l'énergie. L'énergie de chaque photon est infime, donc les sources de lumière en émettent des quantités extrêmement grandes. Elle vaut $q = h\nu$, où q est la constante de Planck ($q = 6,62\ 10^{-34} J.s$), et ν la fréquence de l'onde électromagnétique associée (ν varie entre $3,8\ 10^{14} Hz$ et $7,9\ 10^{14} Hz$ pour la lumière visible). Chaque photon suit sa propre trajectoire, et des propriétés statistiques définissent son comportement, notamment en cas d'interaction avec la matière. Par exemple, la réflexion de la lumière sur un matériau, pour une direction d'incidence et une direction réfléchie particulières, est définie par le rapport entre les nombres de photons incidents et réfléchis dans ces mêmes directions. De la même manière, le taux de photons absorbés par unité de longueur définit l'opacité d'un matériau. Il dépend lui aussi de la longueur d'onde.

Seule cette interprétation corpusculaire permet d'expliquer les phénomènes photoélectriques. Dans certaines conditions particulières, qui dépendent des matériaux, les photons possèdent l'énergie juste suffisante pour modifier la trajectoire des électrons, ou bien pour les arracher de l'attraction des noyaux des atomes. Il se crée alors des courants électriques.

Cette représentation permet d'utiliser la théorie du transport pour analyser la lumière et ses interactions avec la matière (cf. [Gla95], chapitre 12). Les flux de photons sont alors considérés comme des fluides qui circulent à l'intérieur d'enveloppes virtuelles, en obéissant à des lois probabilistes. Le comptage des photons présents à un moment donné en un lieu donné est une mesure de l'énergie présente en cet endroit. En pratique on ne compte pas individuellement les photons, car ils sont trop nombreux, mais on détermine le ratio de ceux qui sont présents en un point donné par rapport à ceux émis par les sources. D'un certain point de vue, la synthèse d'image consiste à estimer au mieux ce ratio, en prenant en compte les propriétés des matériaux rencontrés par les flux de photons sur leurs trajets.

2.2 Principales grandeurs et unités radiométriques

Cette section présente les grandeurs physiques indispensables pour décrire les comportements de la lumière, en nous inspirant largement des développements très complets de

Nicodemus et al. [NRH+77]. Par la suite, nous utiliserons toujours les mêmes symboles pour faire référence à ces grandeurs.

La notion d'angle solide intervient dans la plupart des définitions des grandeurs physiques que nous sommes amenés à manipuler. Nous donnons à l'annexe D des détails sur les méthodes de calcul des angles solides.

2.2.1 Energie et flux énergétique

L'*énergie* s'exprime en Joules $[J]$. Chaque photon transporte une petite quantité d'énergie (cf. section 2.1.2), qui dépend de la fréquence de l'onde électromagnétique qui lui est associée.

Le *flux énergétique* ϕ, exprimé en Watts $[W = J.s^{-1}]$, est l'énergie quittant ou atteignant une surface par unité de temps. Il est directement relié au nombre de photons en mouvement et à l'énergie qu'ils transportent.

Par la suite, nous noterons $\phi(x, \vec{\omega})$ le flux lumineux émis depuis un point x dans la direction $\vec{\omega}$. De plus, afin de ne pas surcharger les notations, nous n'indiquerons explicitement la dépendance avec la longueur d'onde que lorsque cela nous sera nécessaire.

2.2.2 Luminance

2.2.2.1 Définition

La *luminance* $L(x, \vec{\omega})$ est le flux lumineux émis ou reçu par une surface en un point x, dans une direction $\vec{\omega}$ donnée, par unité de surface dA et par unité d'angle solide projeté $\cos\theta d\omega$. Elle s'exprime en $[W.m^{-2}.sr^{-1}]$. L'angle θ est l'angle d'incidence, mesuré par convention entre la direction $\vec{\omega}$ et la normale à la surface en x (figure 2.2).

$$L(x, \vec{\omega}) = \frac{d^2\phi(x, \vec{\omega})}{dA \cos\theta \, d\omega} \tag{2.1}$$

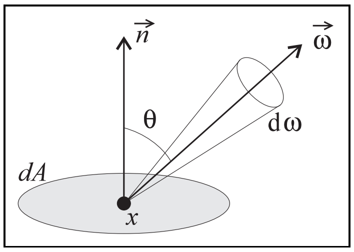


FIG. 2.2 – *Géométrie de la luminance*

La luminance est une donnée fondamentale en synthèse d'images. En effet, le système visuel humain, ainsi que la plupart des autres capteurs (appareils photographiques par exemple), sont sensibles à la luminance. La connaissance de sa répartition sur toutes les surfaces visibles des objets de la scène permet de reconstituer les images.

2.2.2.2 Conservation de la luminance

En dehors de tout phénomène d'absorption, la luminance reste constante le long des chemins lumineux. Pour cette raison, la couleur et la luminosité d'une surface ne dépendent pas de la distance à laquelle elle se situe.

Cette propriété se démontre en considérant deux surfaces élémentaires dA_1 et dA_2, qui se voient mutuellement sous les angle solides $d\omega_1$ et $d\omega_2$, et situées à une distance r l'une de l'autre (figure 2.3).

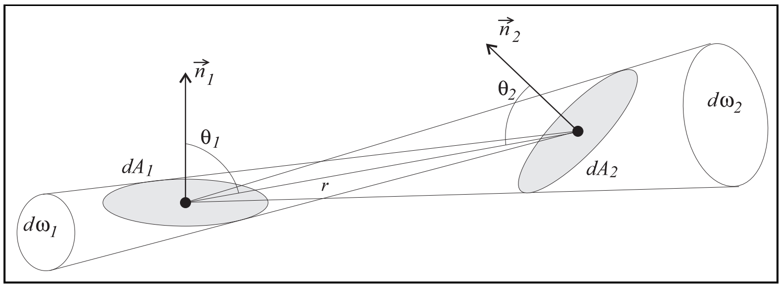


FIG. 2.3 – *Conservation de la luminance entre deux surfaces*

D'après la loi de conservation de l'énergie, le flux quittant dA_1 en direction de dA_2 est égal au flux arrivant en dA_2 depuis dA_1, soit encore $L_1 dA_1 \cos\theta_1 d\omega_2 = L_2 dA_2 \cos\theta_2 d\omega_1$.

Or, par définition des angles solides élémentaires, $d\omega_1 = \dfrac{\cos\theta_1 dA_1}{r^2}$ et $d\omega_2 = \dfrac{\cos\theta_2 dA_2}{r^2}$.

On en déduit $L_1 dA_1 \cos\theta_1 \dfrac{\cos\theta_2 dA_2}{r^2} = L_2 dA_2 \cos\theta_2 \dfrac{\cos\theta_1 dA_1}{r^2}$,

et donc $L_1 = L_2$.

2.2.3 Eclairement

L'*éclairement* E représente le flux énergétique incident total, par unité de surface. Il s'exprime en $[W.m^{-2}]$.

Il se calcule en intégrant l'expression du flux élémentaire en fonction de la luminance incidente L_i, déduit de l'équation 2.1 :

$$E \;=\; \int_{\Omega_i} \frac{d^2\phi}{dA} \;=\; \int_{\Omega_i} L_i(x,\vec{\omega}_i)\cos\theta_i d\omega_i \tag{2.2}$$

où Ω_i représente l'ensemble des directions incidentes.

2.2.4 Radiosité ou exitance

La *radiosité*, ou *exitance* M représente le flux énergétique total rayonné par une surface. Elle s'exprime en $[W.m^{-2}]$. La seule différence avec l'éclairement est que l'on considère la luminance rayonnée L_r au lieu de la luminance incidente.

$$M \;=\; \int_{\Omega_r} L_r(x,\vec{\omega})\cos\theta d\omega \tag{2.3}$$

où Ω_r représente l'ensemble des directions réfléchies.

2.2.5 Intensité

La luminance est bien adaptée pour décrire la répartition énergétique sur les sources surfaciques. Elle n'a par contre aucune signification pour les sources ponctuelles, pour lesquelles le terme dA n'existe pas. Il faut dans ce cas introduire la notion d'*intensité*, qui représente le flux énergétique par unité d'angle solide. Elle s'exprime en $[W.sr^{-1}]$.

L'intensité émise depuis le point x dans la direction $\vec{\omega}$, notée $I(x,\vec{\omega})$ est :

$$I(x,\vec{\omega}) \;=\; \frac{d\phi(x,\vec{\omega})}{d\omega} \tag{2.4}$$

2.3 Interactions de la lumière avec les matériaux

2.3.1 Théorie de Kirchhoff

Lorsqu'un champ électromagnétique $\vec{E_i}$ rencontre une surface, il exerce une force électromagnétique sur les particules chargées électriquement de la matière. Du fait de leur éloignement du noyau, les électrons des couches périphériques, et les électrons libres pour les matériaux conducteurs, sont soumis à une plus faible attraction. Ils sont donc les plus sensibles aux forces électromagnétiques induites par le champ incident, et se mettent à leur tour en mouvement, en engendrant un courant de surface.

Ce courant, comme tout autre courant électrique, crée à son tour un champ électromagnétique qui interagit alors avec le champ incident. La combinaison des deux forme un champ de surface $\vec{E_s}$, qui obéit à l'équation différentielle :

$$\Delta^2 \vec{E_s} + k^2 \vec{E_s} = 0$$

où $k = \dfrac{n\omega}{c}$ est la *constante de propagation* de l'onde.

La théorie de Kirchhoff décrit les comportements de ces champs électromagnétiques. Nous en donnons ci-après un bref aperçu.

2.3.1.1 Intégrale de Helmholtz

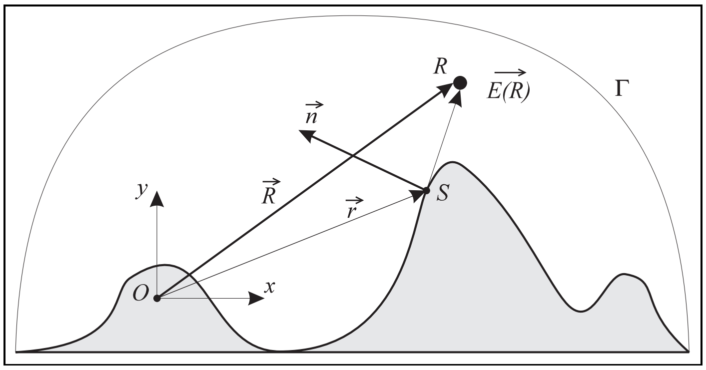


FIG. 2.4 – *Géométrie pour la théorie de Kirchhoff*

Le champ électromagnétique $\overrightarrow{E(R)}$, en tout point R situé à l'intérieur d'une surface fermée Γ (figure 2.4), peut être exprimé grâce à l'*intégrale de Helmholtz*. Sa valeur est la somme intégrale de toutes les contributions élémentaires des éléments de la surface :

$$\overrightarrow{E(R)} = \frac{1}{4\,\pi} \int\int_{\Gamma} \left[\overrightarrow{E_s(\vec{r})}\, \frac{\partial G'(\vec{R},\vec{r})}{\partial \vec{n}} - G'(\vec{R},\vec{r})\, \frac{\partial \overrightarrow{E_s(\vec{r})}}{\partial \vec{n}} \right] d\,\Gamma \tag{2.5}$$

où le terme $G'(\vec{R},\vec{r}) = \dfrac{e^{ik\|\vec{R}-\vec{r}\|}}{\|\vec{R}-\vec{r}\|}$ représente l'onde sphérique émise depuis l'élément de surface dS entourant le point S. Ce point parcourt toute la surface fermée Γ.

En général les termes $\overrightarrow{E_s(\vec{r})}$ et $\dfrac{\partial \overrightarrow{E_s(\vec{r})}}{\partial \vec{n}}$ exprimant le champ de surface ne sont pas connus. Il est cependant possible d'en déterminer une approximation en utilisant l'*hypothèse de Kirchhoff*, grâce à l'approximation par le plan tangent.

2.3.1.2 Approximation par le plan tangent

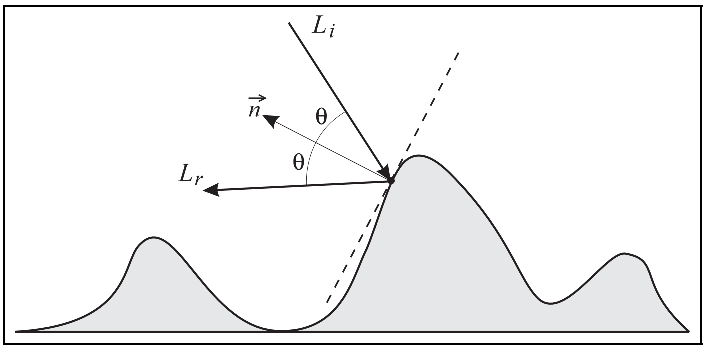


FIG. 2.5 – *Approximation par le plan tangent*

L'approximation par le plan tangent est applicable lorsque la taille des rugosités de la surface est très supérieure à la longueur d'onde de la lumière incidente, ce qui est généralement le cas. L'*hypothèse de Kirchhoff* consiste alors à considérer que le champ présent en un point quelconque d'une surface est égal au champ qui serait présent sur un plan tangent à la surface en ce point (figure 2.5). Ce plan tangent se comporte localement comme un miroir microscopique. On peut alors ne prendre en compte que l'aspect spéculaire de la réflexion et de la réfraction en chaque point.

2.3.1.3 Lois de Descartes

Lorsqu'un rayonnement électromagnétique rencontre une surface *parfaitement plane* séparant deux matériaux homogènes différents d'indices de réfraction n_1 et n_2, une partie de l'énergie est réfléchie, et le reste est transmis de l'autre côté de la surface (figure 2.6).

D'un point de vue géométrique, les directions des ondes respectent les lois de Descartes [1].

1. Ou lois de Snell pour les anglo-saxons ...

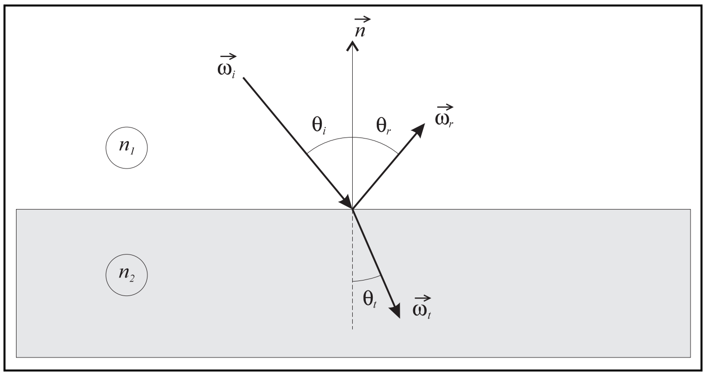


FIG. 2.6 – *Optique géométrique de Descartes*

Les ondes réfléchies et transmises sont toutes les deux dans le plan engendré par l'onde incidente et la normale à la surface, et vérifient :

- La direction de l'onde réfléchie $\vec{\omega_r}$, est symétrique par rapport à la normale, de celle de l'onde incidente $\vec{\omega_i}$, donc $\theta_r = \theta_i$.
- L'angle θ_t entre la normale et la direction de l'onde réfractée dépend à la fois de l'angle d'incidence θ_i et des indices de réfraction des matériaux : $n_1 \sin\theta_i = n_2 \sin\theta_t$.

Il y a réflexion totale lorsque $\dfrac{n_1}{n_2}\sin\theta_i \geq 1$. Dans ce cas, toute l'énergie incidente est réfléchie, et rien ne traverse la surface.

2.3.1.4 Coefficients de Fresnel

Les *coefficients de Fresnel* permettent de caractériser énergétiquement les réflexions et réfractions spéculaires sur des surfaces parfaitement planes, en exprimant l'amplitude des champs électromagnétiques réfléchi et transmis par rapport au champ incident.

Soit $\vec{E}$ le champ incident, que nous exprimons comme combinaison linéaire de deux composantes : $\vec{E_{//}}$ parallèle au plan d'incidence, et $\vec{E_\perp}$ orthogonale à ce même plan d'incidence. Les ondes réfléchies et transmises sont elles aussi exprimées en fonction de leurs composantes parallèles et orthorgonales : $\vec{R_{//}}$ et $\vec{R_\perp}$ pour l'onde réfléchie, $\vec{T_{//}}$ et $\vec{T_\perp}$ pour l'onde transmise.

En notant θ_i l'angle entre l'onde incidente et la normale et θ_t l'angle entre l'onde transmise et la normale, les *relations de Fresnel* donnent :

$$\vec{T_{//}} = \frac{2\sqrt{\frac{\varepsilon_1}{\mu_1}}\cos\theta_i}{\sqrt{\frac{\varepsilon_1}{\mu_1}}\cos\theta_t + \sqrt{\frac{\varepsilon_2}{\mu_2}}\cos\theta_i}\,\vec{E_{//}} \qquad \vec{T_\perp} = \frac{2\sqrt{\frac{\varepsilon_1}{\mu_1}}\cos\theta_i}{\sqrt{\frac{\varepsilon_1}{\mu_1}}\cos\theta_i + \sqrt{\frac{\varepsilon_2}{\mu_2}}\cos\theta_t}\,\vec{E_\perp}$$

$$\vec{R_{//}} = \frac{\sqrt{\frac{\varepsilon_2}{\mu_2}}\cos\theta_i - \sqrt{\frac{\varepsilon_1}{\mu_1}}\cos\theta_t}{\sqrt{\frac{\varepsilon_1}{\mu_1}}\cos\theta_t + \sqrt{\frac{\varepsilon_2}{\mu_2}}\cos\theta_i}\,\vec{E_{//}} \qquad \vec{R_\perp} = \frac{\sqrt{\frac{\varepsilon_1}{\mu_1}}\cos\theta_i - \sqrt{\frac{\varepsilon_2}{\mu_2}}\cos\theta_t}{\sqrt{\frac{\varepsilon_1}{\mu_1}}\cos\theta_i + \sqrt{\frac{\varepsilon_2}{\mu_2}}\cos\theta_t}\,\vec{E_\perp}$$

Soit dS une surface élémentaire appartenant au plan séparant les deux milieux. On note $d\phi_i$ l'énergie élémentaire moyenne arrivant sur dS pendant une période de temps, $d\phi_r$ l'énergie élémentaire moyenne réfléchie par dS, et $d\phi_t$ l'énergie élémentaire moyenne réfractée à travers dS. Ces énergies vérifient les relations :

$$d\phi_i = \frac{1}{2}\sqrt{\frac{\varepsilon_1\varepsilon_0}{\mu_1\mu_0}}E^2 \cos\theta_i \, dS$$
$$d\phi_r = \frac{1}{2}\sqrt{\frac{\varepsilon_1\varepsilon_0}{\mu_1\mu_0}}R^2 \cos\theta_i \, dS$$
$$d\phi_t = \frac{1}{2}\sqrt{\frac{\varepsilon_2\varepsilon_0}{\mu_2\mu_0}}T^2 \cos\theta_i \, dS$$

On définit alors les *coefficients de Fresnel*:

- la réflectance de Fresnel : $R = \dfrac{d\phi_r}{d\phi_i}$
- la transmittance de Fresnel : $T = \dfrac{d\phi_t}{d\phi_i}$

La réflectance et la transmittance de Fresnel, pour un angle de polarisation quelconque α_i de l'onde incidente, se déduisent de la réflectance et de la transmittance pour des polarisations parallèles et orthogonales, notées $R_{//}$, $R_\perp$, $T_{//}$ et $T_\perp$:

$$R = R_{//}\cos^2\alpha_i + R_\perp \sin^2\alpha_i$$
$$T = T_{//}\cos^2\alpha_i + T_\perp \sin^2\alpha_i$$

Les coefficient $R_{//}$, $R_\perp$, $T_{//}$ et $T_\perp$ vérifient :

$$\begin{aligned}
R_{//} &= \left(\frac{\sqrt{\frac{\varepsilon_2}{\mu_2}}\cos\theta_i - \sqrt{\frac{\varepsilon_1}{\mu_1}}\cos\theta_t}{\sqrt{\frac{\varepsilon_1}{\mu_1}}\cos\theta_t + \sqrt{\frac{\varepsilon_2}{\mu_2}}\cos\theta_i}\right)^2 \\
R_\perp &= \left(\frac{\sqrt{\frac{\varepsilon_2}{\mu_2}}\cos\theta_i - \sqrt{\frac{\varepsilon_1}{\mu_1}}\cos\theta_t}{\sqrt{\frac{\varepsilon_1}{\mu_1}}\cos\theta_i + \sqrt{\frac{\varepsilon_2}{\mu_2}}\cos\theta_t}\right)^2 \\
T_{//} &= \frac{\sqrt{\frac{\varepsilon_2}{\mu_2}}\cos\theta_t}{\sqrt{\frac{\varepsilon_1}{\mu_1}}\cos\theta_i}\left(\frac{2\sqrt{\frac{\varepsilon_1}{\mu_1}}\cos\theta_i}{\sqrt{\frac{\varepsilon_1}{\mu_1}}\cos\theta_t + \sqrt{\frac{\varepsilon_2}{\mu_2}}\cos\theta_i}\right)^2 \\
T_\perp &= \frac{\sqrt{\frac{\varepsilon_2}{\mu_2}}\cos\theta_t}{\sqrt{\frac{\varepsilon_1}{\mu_1}}\cos\theta_i}\left(\frac{2\sqrt{\frac{\varepsilon_1}{\mu_1}}\cos\theta_i}{\sqrt{\frac{\varepsilon_1}{\mu_1}}\cos\theta_i + \sqrt{\frac{\varepsilon_2}{\mu_2}}\cos\theta_t}\right)^2
\end{aligned}$$

Pour des ondes non-polarisées, les coefficients de Fresnel sont :

$$\begin{aligned}
R &= \frac{1}{2}\left(R_{//} + R_\perp\right) \\
T &= \frac{1}{2}\left(T_{//} + T_\perp\right)
\end{aligned}$$

La loi de conservation de l'énergie implique :

$$R \in [0,1] \qquad T \in [0,1] \qquad R + T = 1$$

Remarques :

- il n'y a pas de perte d'énergie sur la surface séparant les deux milieux, car l'absorption éventuelle se produit uniquement à l'intérieur des matériaux (section 2.1.1.5) ;
- les coefficients de Fresnel dépendent de la longueur d'onde, ce qui crée l'aspect coloré des matériaux.

2.3.2 Différents types de matériaux

Nous supposons qu'une onde lumineuse polychromatique et non polarisée, circulant dans l'air, entre en contact avec une surface plane. Les caractéristiques du milieu rencontré induisent des comportements différents, que nous allons décrire dans cette partie.

2.3.2.1 Diélectrique homogène isotrope

Le verre est un exemple de matériau diélectrique homogène isotrope. Lorsqu'un champ lumineux atteint la surface d'un diélectrique, il se forme un faisceau réfléchi et un faisceau réfracté. Leurs directions obéissent aux lois de Descartes.

L'indice de réfraction varie avec la longueur d'onde de la lumière incidente. Ainsi, la déviation du faisceau transmis dépend de la fréquence de l'onde électromagnétique. C'est le phénomène de dispersion, que nous avons vu à la section 2.1.1.5.

Le champ incident excite les charges électriques (électrons pour les solides, ions pour les liquides) des atomes situés près de la surface. Elles se mettent à vibrer à leur tour, en créant le champ électromagnétique réfléchi. Pour les longueurs d'onde du domaine visible, les vibrations sont pratiquement en phase avec le flux incident. L'onde réfléchie spéculairement oscille donc avec la même fréquence que l'onde incidente. De ce fait, une source lumineuse et son image réfléchie par un diélectrique parfait sont de même couleur.

2.3.2.2 Conducteur isotrope

Les métaux sont des matériaux conducteurs isotropes. Leur comportement est très différent de celui des diélectriques :

- la lumière pénétrant à l'intérieur est absorbée presque instantanément;
- l'énergie réfléchie est importante ;
- la couleur du faisceau réfléchi est modifiée.

Les électrons libres sont très mobiles. Excités par le champ incident, ils créent des courants de conduction superficiels, qui dissipent l'énergie sous forme de chaleur, et surtout sous forme d'ondes réfléchies spéculairement. Ils ne sont pas parfaitement en phase avec le champ incident, ce qui provoque l'aspect coloré du faisceau réfléchi. De plus, cette coloration dépend de l'angle d'incidence. Elle est maximale lorsque la direction d'incidence est perpendiculaire à la surface, et pratiquement nulle pour une incidence rasante.

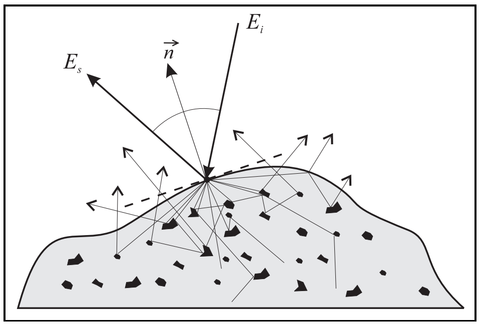


FIG. 2.7 – *Réflexions et réfractions sur un matériau diélectrique hétérogène*

2.3.2.3 Diélectrique hétérogène

Les matières plastiques colorées par des particules de pigment sont des diélectriques hétérogènes (figure 2.7). Il se produit principalement deux phénomènes :

- une réflexion spéculaire pratiquement pas colorée, dont l'intensité augmente avec l'angle d'incidence ;
- une réflexion diffuse colorée, dont l'intensité varie peu avec l'angle d'incidence.

La réflexion spéculaire est semblable à celle produite par un diélectrique homogène, car seule la surface du substrat intervient.

La réflexion diffuse est la combinaison de la réfraction sur la surface du substrat et des multiples interactions avec les grains du pigment. La trajectoire individuelle de chaque rayon est très difficile à caractériser, mais la grande quantité de photons incidents permet de mesurer des comportements statistiques moyens. La lumière est réémise dans toutes les directions, et sa couleur est modifiée par les propriétés spectrales d'absorption et de diffusion des particules du pigment.

2.3.2.4 Autres types de matériaux

Il existe une très grande diversité de matériaux, qui ne se comportent pas comme ceux vus précédemment. On peut citer par exemple :

- les cristaux, qui sont anisotropes, donnent naissance à plusieurs faisceaux réfractés, et peuvent modifier fortement la polarisation de la lumière ;
- les matériaux multicouches, constitués d'empilements de films diélectriques très fins, et dont les comportement varient avec le rapport entre la longueur d'onde incidente et les épaisseurs des couches ;
- toutes combinaisons de matériaux plus ou moins purs, dont les caractéristiques sont difficilement prévisibles.

De plus, les rugosités de la surface modifient sensiblement les aspects macroscopiques de la réflexion. Le faisceau spéculaire réfléchi est plus ou moins atténué, et une composante diffuse est réémise dans toutes les directions. Nous décrirons plus en détail ces comportements lors de l'analyse des modèles de matériaux, dans la section 2.5.

2.4 Modélisation des matériaux

2.4.1 Fonction de distribution de la réflectance bidirectionnelle (BRDF)

2.4.1.1 Définition de la BRDF

Les facteurs de Fresnell ne représentent que les réflexions et réfractions spéculaires pures, sur des surfaces que l'on suppose parfaitement homogènes et planes. Or, dans la réalité, les matériaux n'ont jamais ces comportements parfaits, à cause de l'inévitable présence d'impuretés et de rugosités plus ou moins marquées. Il s'ensuit qu'un faisceau lumineux incident sur une surface donne naissance à une infinité d'ondes réfléchies dans toutes les directions, avec souvent des zones privilégiées.

La *fonction de distribution de la réflectance bidirectionnelle* décrit la répartition énergétique dans toutes les directions réfléchies, et pour toutes les directions incidentes. Par la suite, nous la noterons *BRDF*, sigle dérivé du nom anglais de la fonction (bidirectional reflectance distribution function), car cette appellation est devenue la plus courante en synthèse d'images.

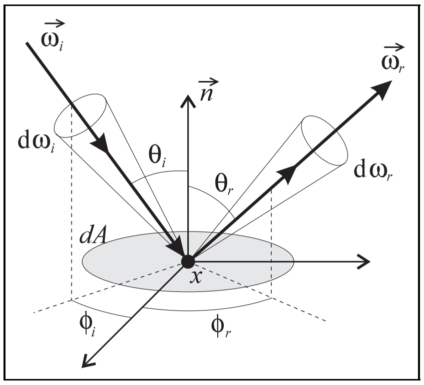


FIG. 2.8 – *Géométrie de la réflexion*

Notation : $f_r(x,\vec{\omega_i}{\to}\vec{\omega_r})$ représente la BRDF en x, pour une direction incidente $\vec{\omega_i}$ et une direction réfléchie $\vec{\omega_r}$ (figure 2.8). Encore une fois, nous ne faisons pas apparaître la dépendance de la longueur d'onde, qui est implicite pour la plupart des grandeurs photométriques.

Par définition, la BRDF en un point x d'un élément de surface dA est le quotient de la luminance réfléchie dans la direction $\vec{\omega_r}$, par l'éclairement de l'élément de surface dA provenant de la direction $\vec{\omega_i}$:

$$f_r(x,\vec{\omega_i}{\to}\vec{\omega_r}) = \frac{dL_r(x,\vec{\omega_r})}{d^2\phi(x,\vec{\omega_i})/dA}$$

La BRDF a pour unité le $[sr^{-1}]$.

On l'exprime souvent en fonction de la luminance incidente :

$$f_r(x,\vec{\omega_i}{\to}\vec{\omega_r}) = \frac{dL_r(x,\vec{\omega_r})}{L_i(x,\vec{\omega_i})\,\cos\theta_i\, d\omega_i}$$

Cette définition représente une valeur ponctuelle de la BRDF, pour une direction réfléchie fixée $\vec{\omega_r}$ et pour l'énergie incidente dans un angle solide infiniment petit $d\omega_i$ autour de la direction $\vec{\omega_i}$.

2.4.1.2 Equation de réflectance

Cette dernière notation permet d'exprimer la luminance totale réfléchie dans une direction $\vec{\omega_r}$, en fonction des luminances provenant de l'ensemble Ω_i de toutes les directions d'incidence possibles. On obtient alors l'équation de réflectance :

$$L_r(x, \vec{\omega_r}) = \int_{\Omega_i} f_r(x, \vec{\omega_i} \to \vec{\omega_r})\, L_i(x, \vec{\omega_i})\, \cos\theta_i\, d\omega_i \tag{2.6}$$

Elle est une des composantes de l'équation de rendu (équation 3.1), qui décrit l'éclairement global dans la scène (section 3).

2.4.1.3 Propriétés fondamentales de la BRDF

La définition de la BRDF suppose que la réflexion est ponctuelle : la lumière ne doit pas être réfléchie d'un autre endroit que le point d'incidence. De plus, la réflexion est sans délai, ce qui élimine la prise en compte de phénomènes tels que la phosphorescence.

La BRDF vérifie le principe de réciprocité de Helmholtz. La lumière peut suivre un trajet de façon identique dans les deux sens :

$$f_r(x, \vec{\omega_i} \to \vec{\omega_r}) \;=\; f_r(x, \vec{\omega_r} \to \vec{\omega_i}) \tag{2.7}$$

Le principe de conservation de l'énergie stipule que l'énergie réfléchie totale sur l'ensemble Ω_r des directions de réflexion ne peut pas être supérieure à l'énergie totale provenant de l'ensemble Ω_i des directions incidentes, soit :

$$\frac{\displaystyle\int_{\Omega_r}\int_{\Omega_i} f_r(x, \vec{\omega_i} \to \vec{\omega_r})\, L_i(x, \vec{\omega_i})\, \cos\theta_i\, \cos\theta_r\, d\omega_i\, d\omega_r}{\displaystyle\int_{\Omega_i} L_i(x, \vec{\omega_i})\, \cos\theta_i\, d\omega_i} \;\leq\; 1 \tag{2.8}$$

L'expression précédente est le quotient de l'énergie émise totale MdA sur un élément de surface dA (M étant l'exitance de la surface), par l'énergie reçue EdA (E étant l'éclairement total incident (équation 2.2). M se calcule en intégrant sur Ω_r l'expression de l'exitance totale (équation 2.3) due aux lumimances réfléchies provenant de Ω_i (équation 2.6).

Un modèle de réflexion est considéré comme *physiquement plausible* lorsqu'il respecte au moins ces deux lois physiques fondamentales.

2.4.1.4 Cas particuliers

2.4.1.4.1 Réflexion parfaitement diffuse Dans le cas de la réflexion parfaitement diffuse, la lumière incidente est réfléchie de façon identique dans toutes les directions (figure 2.9). La BRDF est alors constante : $f_r(x, \vec{\omega_i} \to \vec{\omega_r}) = f_{r,d}^p(x)$.

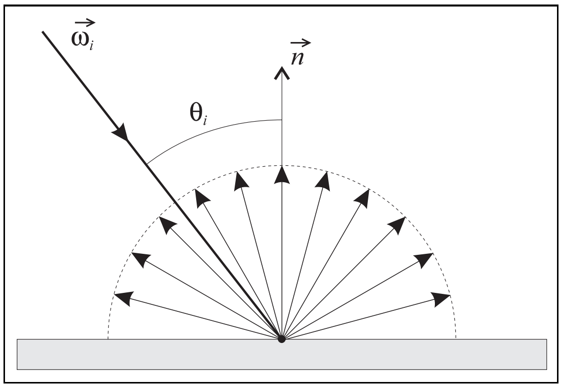


FIG. 2.9 – *Réflexion parfaitement diffuse*

On suppose en outre que l'absorption est nulle, dans ce cas théorique de réflexion parfaite, donc l'intégralité de l'énergie reçue est réémise. L'équation 2.8 devient alors :

$$\frac{\displaystyle\int_{\Omega_r}\int_{\Omega_i} f^p_{r,d}(x)\, L_i(x,\vec{\omega_i})\, \cos\theta_i\, \cos\theta_r\, d\omega_i\, d\omega_r}{\displaystyle\int_{\Omega_i} L_i(x,\vec{\omega_i})\, \cos\theta_i\, d\omega_i} \quad = \quad 1$$

que l'on peut, dans ce cas de réflexion parfaitement diffuse, mettre sous la forme :

$$\frac{\displaystyle f^p_{r,d}(x) \int_{\Omega_r} \cos\theta_r\, d\omega_r \int_{\Omega_i} L_i(x,\vec{\omega_i})\, \cos\theta_i\, d\omega_i}{\displaystyle\int_{\Omega_i} L_i(x,\vec{\omega_i})\, \cos\theta_i\, d\omega_i} \quad = \quad 1$$

on en déduit donc :

$$f^p_{r,d}(x) \;=\; \frac{1}{\displaystyle\int_{\Omega_r} \cos\theta_r\, d\omega_r}$$

L'ensemble des directions réfléchies Ω_r est l'hémisphère orienté selon la normale à la surface. Exprimé en coordonnées polaires, on a :

$$\begin{aligned}
\int_{\Omega_r} \cos\theta_r\, d\omega_r &= \int_0^{2\pi}\int_0^{\frac{\pi}{2}} \cos\theta_r\, \sin\theta_r\, d\theta_r\, d\phi_r \\
&= \int_0^{2\pi} d\phi_r \int_0^{\frac{\pi}{2}} \frac{\sin 2\theta_r}{2}\, d\theta_r \\
&= 2\pi \left[\frac{-\cos 2\theta_r}{2}\right]_0^{\frac{\pi}{2}} \\
&= \pi
\end{aligned}$$

Donc la BRDF diffuse parfaite est :

$$f^p_{r,d}(x) \;=\; \frac{1}{\pi}$$

En pratique, l'absorption n'est jamais nulle. Si $\rho_d \in [0,1[$ est le taux d'énergie réfléchie, alors la BRDF diffuse est :

$$f_{r,d}(x) = \frac{\rho_d}{\pi} \tag{2.9}$$

2.4.1.4.2 Réflexion parfaitement spéculaire La réflexion parfaitement spéculaire est la modélisation théorique d'un miroir parfait. La direction de la lumière réfléchie, donnée par les loi de Descartes (section 2.3.1.3), est dans la direction symétrique de la direction incidente par rapport à la normale. Pour ces directions, la luminance réfléchie est égale à la luminance incidente. On considère de plus qu'il n'y a pas d'absorption.

La *fonction de Dirac*, notée δ est une fonction nulle partout sur $\mathbb{R}$, sauf en 0, et qui vérifie pour toute fonction f la propriété : $\int_{-\infty}^{+\infty} \delta(x-y)f(x)dx = f(y)$. Nous l'utilisons pour modéliser la réflexion spéculaire, qui est nulle partout, sauf dans le pic spéculaire contenu dans un angle solide infiniment petit.

La BRDF spéculaire parfaite, notée $f^p_{r,s}$, a pour valeur :

$$f^p_{r,s}(x, \vec{\omega_i} \to \vec{\omega_r}) = \frac{\delta(\cos\theta_i - \cos\theta_r)}{\cos\theta_i} \cdot [\delta(\phi_i - \phi_r - \pi) + \delta(\phi_i - \phi_r + \pi)] \tag{2.10}$$

Le terme $[\delta(\phi_i - \phi_r - \pi) + \delta(\phi_i - \phi_r + \pi)]$ n'est non nul que pour la direction symétrique, c'est à dire lorsque $\phi_r = \phi_i + \pi$ ou bien $\phi_r = \phi_i - \pi$. Une seule de ces deux valeurs est atteinte pour tout $\phi_i \in [0, 2\pi[$.

2.4.1.5 Aspects qualitatifs de la BRDF

La BRDF est une fonction de dimension 7 au moins (3 pour la position, et 2 pour chaque direction), sans tenir compte de la dépendance spectrale. Il est très difficile de la mesurer avec précision sur des échantillons de matériaux, pour des raisons de précision géométrique et de mesures énergétiques. Le simple fait que le capteur de mesure ait nécessairement une surface finie induit un lissage et une perte d'information, en général à proximité de la zone de réflexion spéculaire qui est tout particulièrement significative. En outre, le stockage des nombreuses données, dues à la grande dimension de l'espace échantillonné, est problématique.

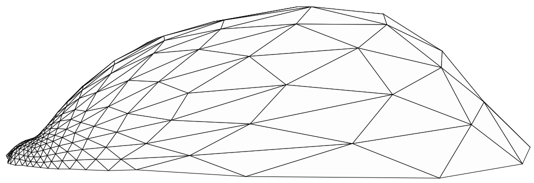


FIG. 2.10 – *BRDF mesurée : ardoise*

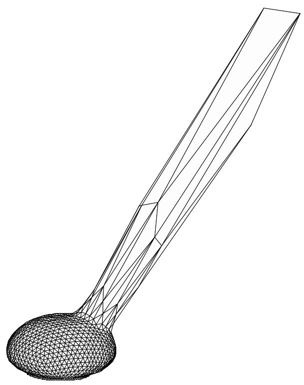


FIG. 2.11 – *BRDF mesurée : céramique*

Il n'est pas possible de représenter la BRDF graphiquement dans son ensemble. La seule solution est de fixer quelques paramètres. Il est habituel de visualiser en 3 dimensions le volume décrit pour une direction d'incidence fixée, ou bien alors la courbe plane obtenue en faisant l'intersection du volume précédent avec le plan d'incidence.

La figure 2.10 représente la BRDF mesurée d'ardoise, un matériau au comportement plutôt spéculaire. L'incidence est de 75^o, donc relativement rasante. Le pic spéculaire, assez large, est pratiquement tangent à la surface. La partie diffuse est beaucoup plus faible, et de forme sensiblement sphérique.

La figure 2.11 représente la BRDF mesurée d'une céramique à une incidence de 30^o. Cette fois la composante diffuse est beaucoup plus marquée. Le pic spéculaire est toujours important, mais il est étroit. La direction de réflexion spéculaire est plus précise.

Ces donnée ont été obtenues avec un goniophotomètre expérimental développé par le *Laboratoire Central des Ponts et Chaussées* [CB96].

2.4.2 Fonction de distribution de la transmittance bidirectionnelle (BTDF)

La *fonction de distribution de la transmittance bidirectionnelle*, que nous noterons *BTDF*, est l'équivalent de la BRDF, pour la transmission de la lumière à travers la matière. L'abréviation usuelle est $f_t(x, \vec{\omega_i} \to \vec{\omega_t})$, pour la valeur de la BTDF en un point x, pour une direction incidente $\vec{\omega_i}$ et pour une direction transmise $\vec{\omega_t}$.

La BTDF indique le rapport entre la luminance transmise dans la direction $\vec{\omega_t}$ et l'éclairement provenant de la direction $\vec{\omega_i}$:

$$f_t(x, \vec{\omega_i} \to \vec{\omega_t}) = \frac{dL_t(x, \vec{\omega_t})}{L_i(x, \vec{\omega_i}) \cos\theta_i \, d\omega_i}$$

La mesure physique de la BTDF est particulièrement difficile pour les solides. L'absorption n'est pas négligeable, comme pour des petites distances dans l'air, et perturbe les résultats. De plus, pour ne pas ajouter des réflexions parasites, il faudrait inclure des capteurs dans la matière, et ceux-ci devraient avoir le même indice de réfraction qu'elle. Ces conditions sont peu réalistes, et l'on préfère introduire des compensations numériques, hélas complexes et peu fiables si l'état de surface n'est pas parfait. Il faut donc bien souvent se satisfaire de données extrapolées de modèles théoriques qui tentent de simuler des observations plus ou moins empiriques.

2.5 Modèles analytiques de matériaux

La création d'images de synthèse "réalistes" implique de reproduire au mieux les comportements des matériaux. Les résultats de mesures physiques disponibles sont peu nombreux, à cause des difficultés matérielles de cette opération. De plus les diverses combinaisons d'états de surface, de coloris et de texturations possibles conduiraient à la création de bases de données gigantesques, et sans doute incomplètes. La simulation semble donc une solution mieux adaptée à ce problème. Plusieurs modèles analytiques ont été proposés dans la littérature, prenant en compte plus ou moins complètement les phénomènes lumineux.

Nous présentons ici quelques-uns de ceux qui sont le plus souvent utilisés en synthèse d'images, en insistant particulièrement sur ceux qui respectent les lois fondamentales de la physique.

2.5.1 Modèle de Lambert

Le modèle de matériaux de Lambert décrit les surfaces diffuses, qui renvoient l'énergie de la même façon dans toutes les directions. La différence par rapport à la réflexion parfaitement diffuse (section 2.4.1.4.1), pour laquelle la BRDF est $f^p_{r,d}(x) = \frac{1}{\pi}$, est l'ajout d'un coefficient d'absorption $C(\lambda)$ indiquant la dépendance spectrale (équation 2.9). Pour respecter la loi de conservation de l'énergie, $C(\lambda)$ est compris entre 0 et 1 pour toute longueur d'onde λ du spectre lumineux. L'ensemble des coefficients $C(\lambda)$ représente la couleur de l'objet.

La BRDF lambertienne (ou diffuse) est donc :

$$f_r(x, \vec{\omega_i} \to \vec{\omega_r}) \;=\; f_{r,d}(x) \;=\; \frac{C(\lambda)}{\pi}$$

L'équation de réflectance (équation 2.6), pour une longueur d'onde λ, devient alors :

$$L_r(x, \vec{\omega_r}, \lambda) \;=\; \frac{C(\lambda)}{\pi} \int_{\Omega_i} L_i(x, \vec{\omega_i}) \cos\theta_i \, d\omega_i \;=\; \frac{C(\lambda)}{\pi} E(\lambda)$$

où $E(\lambda)$ est l'éclairement total pour la longuer d'onde λ.

Bien qu'elle ne soit que théorique, cette modélisation simple est souvent utilisée par les algorithmes de rendu ne simulant pas les réflexions spéculaires et les réfractions.

2.5.2 Modèle de Phong

2.5.2.1 Description

Phong a été le premier, en 1975, à introduire en synthèse d'images un modèle d'éclairement augmentant notablement le réalisme [Pho75]. Il n'est pas décrit par une BRDF, notion encore inutilisée à l'époque, mais par des coefficients indiquant les proportions d'énergie réfléchie spéculaire et diffuse. Pour se rapprocher des modèles physiques actuels, on peut l'exprimer en terme de luminance :

$$L_r(x, \vec{\omega_r}, \lambda) \;=\; k_a(\lambda) L_a + \Big[d\, k_d(\lambda) \cos(\vec{N}, \vec{\omega_i}) + s\, k_s(\lambda)\, F_s(\vec{\omega_i}, \vec{\omega_r}) \Big] \; L_i(x, \vec{\omega_i}, \lambda)$$

- $\vec{N}$ est la normale à la surface en un point x ;
- $d \in [0,1]$ est la proportion de la surface se comportant comme un réflecteur diffus ;
- $s \in [0,1]$, tel que $s + d = 1$, est la proportion de la surface se comportant comme un réflecteur spéculaire ;
- L_a est une luminance ambiante arbitraire, évitant que les zones d'ombre ne soient entièrement noires ;
- $k_a(\lambda) \in [0,1]$ est un coefficient de réflexion ambiante ;
- $k_d(\lambda) \in [0,1]$ est la proportion de lumière réfléchie par le réflecteur diffus ;

- $k_s(\lambda) \in [0,1]$ est la proportion de lumière réfléchie par le réflecteur spéculaire.

Les termes $k_a(\lambda)$, $k_d(\lambda)$, et $k_s(\lambda)$, qui dépendent de la longueur d'onde, représentent la couleur du matériau.

Le terme $F_s(\vec{\omega_i}, \vec{\omega_r})$ permet de faire varier la taille du pic spéculaire, afin de modifier l'aspect de la brillance du matériau. Il est donné par la formule :

$$F_s(\vec{\omega_i}, \vec{\omega_r}) = \begin{cases} \cos^n(\vec{R}, \vec{\omega_r}) & \text{si} \quad \cos(\vec{R}, \vec{\omega_r}) > 0 \\ 0 & \text{sinon} \end{cases}$$

où $\vec{R} = 2(\vec{N}.\,\vec{\omega_i})\vec{N} - \vec{\omega_i}$ est la direction de réflexion spéculaire parfaite, et n un coefficient de réflexion spéculaire empirique. Plus n est grand, plus le reflet spéculaire est étroit.

2.5.2.2 Variante de Blinn

Blinn a proposé en 1977 [Bli77] une légère modification du modèle de Phong, en reformulant l'expression du terme $F_s(\vec{\omega_i}, \vec{\omega_r})$:

$$F_s(\vec{\omega_i}, \vec{\omega_r}) = \cos^n(\vec{N}, \vec{H})$$

où $\vec{H}$ est la direction bissectrice de $\vec{\omega_i}$ et $\vec{\omega_r}$.

Ce modèle est tout aussi empirique que celui de Phong, et les différences visuelles sont faibles. Il fut cependant beaucoup utilisé, car l'absence du terme $\vec{R}$ diminue sensiblement les temps de calculs. Il produit des images acceptables, mais sans réalisme, où les matériaux ont tous un aspect artificiel de matière plastique. Il est toujours présent dans certains logiciels de rendu simples.

2.5.2.3 Considérations physiques

Lewis a montré en 1993 [Lew93] que ni le modèle de Phong, ni la variante de Blinn ne sont physiquement plausibles (section 2.4.1.3). Ne respectant pas toujours la loi de conservation de l'énergie, il se peut sous certaines conditions que l'énergie réfléchie soit plus forte que la somme de toutes les énergies incidentes. Il propose dans le même article des solutions pour remédier à ce problème, sans toutefois éliminer tous les inconvénients de ce modèle développé empiriquement.

2.5.3 Modèle de Cook-Torrance

Cook et Torrance [CT81] ont proposé, en 1981, le premier modèle théorique de réflexion pour la synthèse d'images basé sur la physique des matériaux et de la lumière. Il s'appuie sur les travaux antérieurs de Beckmann et Spizzichino [BS63] et de Torrance et Sparrow [TS67] sur la réflexion des ondes électromagnétiques sur des surfaces rugueuses. La BRDF est donnée par la formule :

$$f_r(x, \vec{\omega_i} \rightarrow \vec{\omega_r}, \lambda) \;=\; \frac{d}{\pi} C(\lambda) \;+\; \frac{s}{\pi} \frac{F.D.G}{\cos\theta_i \cos\theta_r}$$

- θ_i et θ_r sont les angles d'incidence et de réflexion ;

- $d \in [0, 1]$ est la proportion de la surface se comportant comme un réflecteur diffus ;
- $s \in [0, 1]$, tel que $s + d = 1$, est la proportion de la surface se comportant comme un réflecteur spéculaire ;
- $C(\lambda)$ est la proportion de lumière réfléchie par le réflecteur diffus, pour la longueur d'onde λ ;
- F est le facteur de Fresnel, pour l'angle d'incidence θ_i et la longueur d'onde λ ;
- D est un terme représentatif de la distribution de micro-facettes (section 2.5.3.1) ;
- G est un facteur d'atténuation géométrique représentant les effets de masquage et d'auto-ombrage (section 2.5.3.2).

2.5.3.1 Distribution de micro-facettes

Aucun objet réel n'offre une surface parfaitement plane pouvant être représentée uniquement par les facteurs de Fresnel (section 2.3.1.4). Les rugosités jouent un rôle prédominant dans l'aspect macroscopique des matériaux, par leur taille et par leur orientation. Il n'est pas envisageable en pratique de modéliser géométriquement les milliards d'aspérités microscopiques qui recouvrent chaque centimètre-carré de surface. Par contre, on constate que les matériaux ont des propriétés macroscopiques moyennes que l'on peut reproduire d'un échantillon à l'autre. Les rugosités peuvent donc être représentées uniquement par les caractéristiques statistiques de leur forme, taille et répartition.

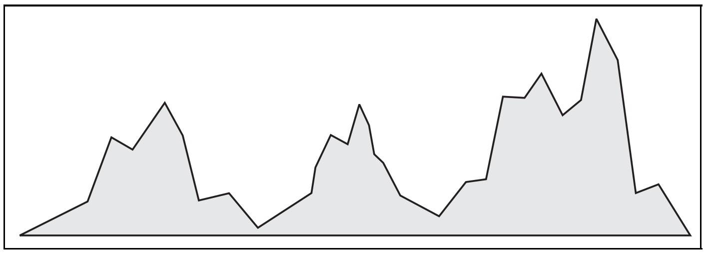


FIG. 2.12 – *Simulation des rugosité par des facettes planes*

L'objectif étant d'obtenir un modèle calculable, les rugosités sont simulées par une répartition aléatoire de micro-facettes planes à la surface des objets (figure 2.12). Les données statistiques décrivant l'état de la surface sont alors la moyenne et l'écart-type de la taille des facettes et de leur orientation. Le choix de micro-facettes planes permet de considérer qu'elles respectent individuellement les caractéristiques simples de la réflexion et de la réfraction spéculaires. Un autre avantage est que les facteurs de Fresnel peuvent être mesurés ou calculés avec une précision généralement satisfaisante pour les matériaux courants.

Dans le modèle de Cook-Torrance, le terme D représente la distribution statistique des micro-facettes, exprimé en général par la répartition des pentes par rapport à la normale. Si α est l'angle entre la normale et le vecteur $\vec{H}$, et si $m \in]0, 1[$ caractérise la rugosité, alors :

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

Cette expression est dérivée de la théorie de Beckmann. Il existe quelques autres formulations, qui dépendent des paramètres choisis pour décrire l'état de surface moyen.

2.5.3.2 Facteur d'atténuation géométrique

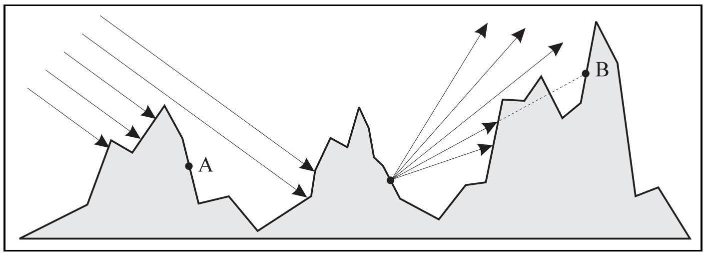


FIG. 2.13 – *Exemples d'auto-ombrage (A) et de masquage (B)*

Les rugosités de la surface peuvent arrêter une partie des rayons lumineux. Cet effet est particulièrement sensible pour les éclairages en lumière rasante.

- L'*auto-ombrage* est provoqué par les rugosités qui bloquent les rayons incidents. Par exemple, le point A de la surface n'est pas éclairé (figure 2.13).
- Le *masquage* est un phénomène similaire, mais qui concerne les rayons réfléchis. Par exemple, le point B n'est pas atteint par la lumière réfléchie sur le pic situé à sa gauche.

Dans le modèle de Cook-Torrance, l'auto-ombrage et le masquage sont simulés par le facteur d'atténuation géométrique G, indiquant le taux de lumière "perdue" :

$$G = \min \left\{ 1 \, , \, \frac{2 \cos(\vec{N}, \vec{H}) \cos \theta_r}{\cos(\vec{\omega_r}, \vec{H})} \, , \, \frac{2 \cos(\vec{N}, \vec{H}) \cos \theta_i}{\cos(\vec{\omega_r}, \vec{H})} \right\}$$

où $\vec{N}$ est la normale à la surface et $\vec{H}$ la bissectrice entre la direction incidente $\vec{\omega_i}$ et la direction réfléchie $\vec{\omega_r}$.

Cette formulation de G ne dépend que des angles d'incidence et de réflexion. Il n'introduit aucune prise en compte de la taille et de la répartition des rugosités, alors que ce sont elles qui sont à l'origine du phénomène. De plus, les réflexions éventuelles des rayons réfléchis sont ignorées, comme s'ils étaient systématiquement absorbés, ce qui n'est pas justifié. Les modèles beaucoup plus complexes, par exemple celui de Smith [Smi67], ne présentent pas ce problème.

2.5.4 Modèle HTSG

Un modèle plus complet de matériaux, entièrement basé sur la théorie de Kirchhoff (section 2.3.1), est présenté par He, Torrance, Sillion et Greenberg dans [HTSG91]. Il a été le premier à prendre en compte la polarisation des ondes électromagnétiques.

La BRDF est divisée en 3 composantes indépendantes régies par des propriétés physiques différentes (figure 2.14) :

$$f_r = f_r^{du} + f_r^{dd} + f_r^{sp} \tag{2.11}$$

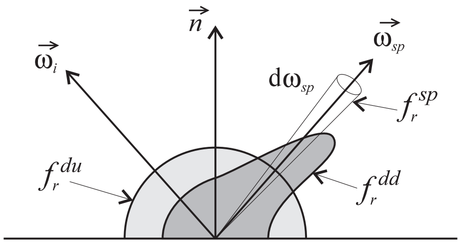


FIG. 2.14 – *Décomposition de la BRDF*

où :

- f_r^{du} est la composante diffuse uniforme, obéissant à la loi de Lambert ;
- f_r^{dd} est la composante diffuse directionnelle, orientée dans une direction proche du pic spéculaire, et englobée dans un angle solide pouvant être important ;
- f_r^{sp} est le pic spéculaire, orienté dans la direction de la réflexion spéculaire $\overrightarrow{\omega_{sp}}$, et contenu dans un angle solide infiniment petit $d\omega_{sp}$ représenté par la fonction de Dirac.

Le développement physique et mathématique du modèle est très complexe, et ne sera pas repris ici. Sa mise en œuvre informatique est tout aussi lourde, et la nécessité de calculer des séries numériques convergentes rend les temps de calcul très importants. Les résultats sont cependant à la hauteur de la complexité, et les vérifications expérimentales sont probantes. Par contre, il faut disposer des constantes électromagnétiques des matériaux et d'informations précises sur l'état de surface. Ce n'est pas toujours possible en pratique, surtout lorsque les matériaux simulés ne sont pas homogènes, ce qui est le cas le plus fréquent.

2.5.5 Modèle de Ward

Ward présente dans [War92] une réalisation simple d'un goniophotomètre, développé par le *Lawrence Berkeley Laboratory*. Classiquement, la mesure des BRDF requiert un montage mécanique de précision permettant les orientations simultanées de la sources et du capteur au dessus de l'échantillon mesuré. C'est un dispositif à la fois complexe, fragile et coûteux, lorsqu'il offre une bonne précision géométrique. A la place, Ward propose l'assemblage d'un miroir hémisphérique semi-opaque et d'un appareil photographique muni d'un objectif en "œil de poisson", disposés autour de l'échantillon. Une source lumineuse mobile éclaire à travers le miroir selon l'angle d'incidence désiré. L'image capturée indique la répartition de la lumière réfléchie sur l'échantillon, à travers sa BRDF, et sur le miroir. La BRDF est finalement reconstituée en appliquant la transformation mathématique inverse à l'image. La précision des mesures est satisfaisante au regard de la simplicité du dispositif.

Muni d'un ensemble de mesures à diverses incidences pour quelques matériaux, Ward développe dans le même article un modèle empirique de BRDF capable de reproduire approximativement les résultats physiques. Il est décliné sous deux versions : $f_{r,iso}$ pour les matériaux isotropes, et $f_{r,aniso}$ pour les matériaux anisotropes dont la BRDF varie avec l'angle d'incidence azimutal ϕ_i (figure 2.8).

Modélisation de BRDF isotrope

$$f_{r,iso}(x, \vec{\omega_i} \to \vec{\omega_r}) = \frac{\rho_d}{\pi} + \rho_s \frac{e^{-\tan^2 \frac{\delta}{\alpha^2}}}{4\pi\alpha^2\sqrt{\cos\theta_i \cos\theta_r}} \tag{2.12}$$

où :

- ρ_d est le coefficient de réflectance diffus ;
- ρ_s est le coefficient de réflectance spéculaire ;
- α est un coefficient de rugosité, semblable au terme m du modèle de Cook-Torrance ;
- δ est l'angle entre la normale et la bissectrice $\vec{H}$ des directions incidentes et réfléchies.

Modélisation de BRDF anisotrope

$$f_{r,aniso}(x, \vec{\omega_i} \to \vec{\omega_r}) = \frac{\rho_d}{\pi} + \rho_s \frac{e^{-\tan^2 \delta \left(\frac{\cos^2 \phi_i}{\alpha_x^2} + \frac{\sin^2 \phi_i}{\alpha_y^2} \right)}}{4\pi\alpha_x\alpha_y\sqrt{\cos\theta_i \cos\theta_r}} \tag{2.13}$$

où :

- α_x est un coefficient de rugosité pour une direction x sur la surface ;
- α_y est un coefficient de rugosité pour une direction y perpendiculaire à x.

Ces deux versions du modèle de Ward sont physiquement plausibles, et possèdent en outre l'avantage d'être contrôlées par un petit nombre de paramètres relativement intuitifs.

2.5.6 Modèle de Schlick

2.5.6.1 Principe

L'approche de Schlick [Sch94] est différente. Il ne propose pas vraiment un nouveau modèle, mais revoit entièrement l'écriture de ceux existant déjà. Le constat est que les modèles utilisés en pratique sont empiriques, et que de surcroit les données physiques sur les matériaux réels sont très disparates. L'idée est alors de développer un modèle qui soit physiquement plausible, ce qui est la condition minimale, et contrôlé par des paramètres intuitifs et peu nombreux, comme le modèle de Ward. Comme il s'agit d'un modèle empirique, simplifiant nécessairement les lois physiques extrêmement complexes, sa formulation mathématique doit être la plus compacte possible, pour permettre une implémentation informatique efficace.

Partant d'un modèle à distribution de facette, très similaire à celui de Cook-Torrance, Schlick remplace toutes les formules complexes par des fractions rationnelles de faible degré ayant la même allure. Les coefficients des fractions rationnelles sont choisis pour obtenir les mêmes caractérisques de valeur et de pente que la fonction à simuler, aux bornes de l'intervalle de définition et en quelques points particuliers. Le nombre de points de contrôle et les variations de la fonction influencent le degré des fractions. Des simulations numériques montrent que l'erreur moyenne est très faible lorsque le degré des fonctions et le choix des coefficients sont corrects, ce qui est bien entendu le cas pour le modèle proposé.

Il définit deux catégories de matériaux :

- les matériaux *simples* simulent les conducteurs (section 2.3.2.2), en particulier les métaux ;

- les matériaux *doubles* simulent les matériaux diélectriques hétérogènes (section 2.3.2.3) avec des matériaux simples recouverts d'une pellicule fine de vernis ou de peinture homogène.

2.5.6.2 Matériaux simples

Deux paramètres, $r \in [0,1]$ et $p \in [0,1]$, définissent l'aspect du matériau :

- r indique la proportion de la composante diffuse dans la réflexion ($r = 0$: spéculaire parfait, $r = 1$: diffus parfait) ;
- p est un coefficient d'isotropie ($p = 0$: anisotrope parfait, $p = 1$: isotrope parfait).

L'expression de la BRDF est décomposée en un terme spectral S et un terme directionnel D :

$$f_r(x, \vec{\omega_i} \to \vec{\omega_r}, \lambda) = S(x, \vec{\omega_i}, \vec{\omega_r}, \lambda)\, D(x, \vec{\omega_i}, \vec{\omega_r}) \tag{2.14}$$

avec :

$$S(x, \vec{\omega_i}, \vec{\omega_r}, \lambda) = C(\lambda) + (1 - C(\lambda))(1-u)^5 \tag{2.15}$$

et :

$$D(x, \vec{\omega_i}, \vec{\omega_r}) = \frac{G(v)G(v')}{4\pi v v'} Z(t)\, A(w) + \frac{1 - G(v)G(v')}{4\pi v v'} \tag{2.16}$$

$\vec{H}$ est le vecteur médian $(\vec{\omega_i} + \vec{\omega_r})/2$, $t = \cos(\vec{H}, \vec{n})$, $u = \cos(\vec{H}, \vec{\omega_r})$, $v = \cos\theta_r$, $v' = \cos\theta_i$, et w est le cosinus de l'angle azimutal entre $\vec{H}$ et un vecteur tangent à la surface indiquant l'axe éventuel d'anisotropie. Le terme zénithal Z, le terme azimutal D et le terme d'atténuation géométrique G valent :

$$\begin{aligned} Z(t) &= \frac{r}{(1 + rt^2 - t^2)^2} \\ A(w) &= \sqrt{\frac{p}{p^2 - p^2 w^2 + w^2}} \\ G(\xi) &= \frac{\xi}{r - r\xi + \xi} \end{aligned}$$

2.5.6.3 Matériaux doubles

Les matériaux doubles du modèle de Schlick simulent les matériaux composés de deux couches distinctes. L'énergie qui n'est pas réfléchie pas la première couche (définie par S et D) l'est par la deuxième (définie par S' et D'). Avec les mêmes notations que pour les matériaux simples, la BRDF s'exprime par :

$$f_r(x, \vec{\omega_i} \to \vec{\omega_r}, \lambda) = S(u, \lambda)\, D(t, v, v', w) + [1 - S(u, \lambda)]\, S'(u, \lambda)\, D'(t, v, v', w) \tag{2.17}$$

En général la première couche est plutôt spéculaire, comme un vernis, alors que la deuxième est plutôt diffuse comme les grains d'un pigment. Le modèle autorise cependant toutes les autres combinaisons théoriques.

2.5.6.4 Continuum diffus-spéculaire

L'expression du terme directionnel D peut être modifiée pour exprimer une variation continue entre un réflecteur lambertien et un réflecteur de Fresnel, en passant par toutes les étapes décrites par un modèle à micro-facettes :

$$D(t, v, v', w) \;=\; \frac{a}{\pi} \;+\; \frac{b}{4\pi v v'} B(t, v, v', w) \;+\; \frac{c}{d\omega_i} \Delta(d\omega_i) \tag{2.18}$$

où $B(t, v, v', w)$ est l'expression de D définie auparavant (équation 2.16) et $\Delta(d\omega_i)$ simule la fonction de Dirac en étant égal à 1 dans le pic spéculaire d'angle solide $d\omega_i$ et nul partout ailleurs.

On retrouve la décomposition en trois éléments du modèle HTSG (équation 2.11). Les coefficients a, b, et c indiquent les répartitions respectives des trois modes de réflexion ($a + b + c = 1$). Ils sont déterminés automatiquement en fonction du facteur de rugosité r :

$$\begin{cases} b = 4r(1-r); \quad a = 0; \quad c = 1 - b; & \text{si } \; r < 0,5 \\ b = 4r(1-r); \quad a = 1 - b; \quad c = 0; & \text{sinon} \end{cases}$$

2.5.6.5 Remarques

Nous avons implémenté le modèle de Schlick dans notre logiciel de rendu. Le respect des lois physiques fondamentales assure l'indispensable absence d'incohérences énergétiques, et la simplicité des opérations ne pénalise pas les temps de calcul. Il donne de bons résultats visuels pour la simulation de matériaux courants, en scènes d'intérieur notamment. Il semble que l'aspect souvent artificiel des images de synthèse provienne beaucoup plus des simplifications de la géométrie des objets et des textures que de la définition approximative des propriétés des matériaux.

2.5.7 Modèle de lobes de Cornell

Lafortune et al. proposent dans [LFTG97] une généralisation du modèle de Phong. Ce modèle simule approximativement tout type de BRDF par une somme de lobes similaires au lobe spéculaire de Phong (terme F_s, section 2.5.2), dans les directions privilégiées de réflexion. De façon simplifiée, la BRDF est de la forme :

$$f_r \;=\; \sum_{i=1}^{N} \rho_{s,i}\, k_i\, \cos^{n_i} \alpha_i$$

où :

- l'angle α_i est mesuré entre la direction réfléchie et $\vec{d_i}$, l'une des N directions privilégiées ;
- l'exposant n_i caractérise la taille du lobe dans la direction $\vec{d_i}$;
- le term $\rho_{s,i}$ est la proportion d'énergie renvoyée dans la direction $\vec{d_i}$;
- la suite de coefficients pondérateurs k_i assurent la conservation globale d'énergie.

Ce modèle est physiquement plausible, car tous les lobes respectent la loi de réciprocité. La formulation du modèle dans [LFTG97] utilise un repérage plus complet dans l'espace que celui de la version schématique présentée ici, et permet la simulation de matériaux anisotropes. Même les effets de rétro-réflexion sont possibles, grâce à la totale liberté laissée sur le nombre et la position des lobes. Dans ce cas, une partie de la lumière est réfléchie en direction de la source, en plus du traditionnel pic spéculaire. C'est un phénomène souvent relevé lors des mesures expérimentales, mais en général ignoré par les modèles de matériaux classiques.

2.6 Textures

Le principal reproche fait aux images de synthèse est leur manque de réalisme, à cause des trop nombreuses simplifications géométriques et du manque de diversité des matériaux. En effet, dans la nature, à l'échelle macroscopique, les matériaux ne sont jamais parfaitement homogènes, et les surface ne sont jamais parfaitement lisses. Bien que cela puisse sembler contradictoire, c'est le manque de défauts qui est préjudiciable. Les vitrages des images de synthèse manquent cruellement de traces de doigts, et les parquets cirés de rayures de talons. D'autre part, la description de certains détails semble une tâche insurmontable de par leur finesse et leur nombre. Les brins d'herbe d'une prairie sont tous différents, mais il est inimaginable de vouloir tous les décrire individuellement.

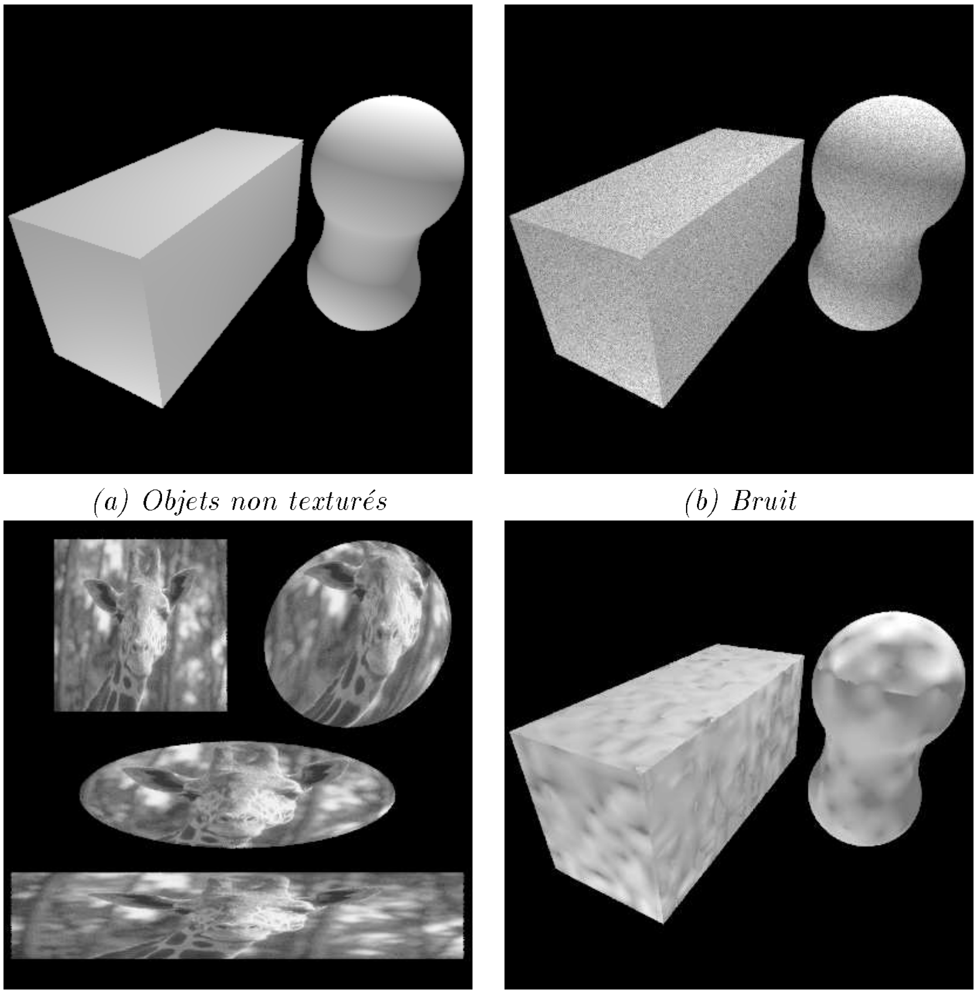
Les textures, en synthèse d'images, sont des procédés qui modifient l'aspect des surfaces, pour tenter de les rendre plus réalistes. Ells simulent des détails de couleur ou de géométrie qui ne sont pas décrits explicitement. L'idée initiale a été introduite par Catmull [Cat74] puis développée par Blinn et Newell [BN76]. De nombreux travaux en ce domaine ont été publiés depuis, aussi bien pour améliorer l'implémentation algorithmique que le réalisme. D'autres recherches sont toujours en cours, pour tenter de mieux simuler des phénomènes toujours plus complexes.

2.6.1 Textures planes

Les textures planes sont définies dans un espace à deux dimensions, puis elles sont plaquées sur les surfaces des objets pour en modifier l'apparence. L'idée semble simple, mais sa mise en œuvre pose de nombreux problèmes. La projection sur des surfaces non planes génère des déformations, et les discontinuités peuvent créer des incohérences. On peut faire l'analogie avec une tapisserie collée contre un mur. La figure 2.15 présente quelques exemples de textures planes appliqués à des objets simples.

Le cas le plus élémentaire est celui des textures de bruit (figure 2.15-b), où la couleur de chaque pixel est modifiée par une valeur aléatoire, ce qui donne un aspect sableux aux surfaces. Comme la texture n'est pas définie explicitement, il n'y a aucun problème de projection. Ce cas très particulier est parfois considéré comme une texture volumique. Des procédés de génération pseudo-aléatoires permettent une répétitivité indispensable pour les animations, afin d'éviter que la répartition des couleurs des pixels ne varie d'une image à l'autre.

Le plaquage d'images est la texture plane la plus courante, mais c'est aussi l'opération la plus sensible aux problèmes de déformation. Dans notre exemple (figure 2.15-c), la même image est plaquée sur des objets différents (une surface carrée, une sphère, un ellipsoïde, et une surface rectangulaire), par une projection parallèle. Les déformations du sujet sont évidentes, même s'il reste souriant. Il faut définir une méthode de projection adaptée aux



(a) Objets non texturés *(b) Bruit*

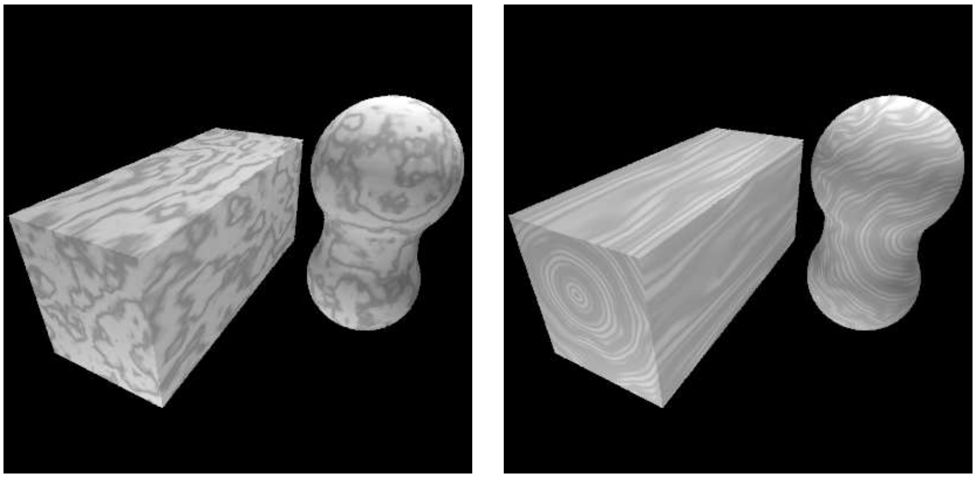
(c) Image plaquée *(d) Perturbation de normale*

FIG. 2.15 – *Exemples de textures planes*

caractéristiques de la surface et aux proportions de l'image initiale, avec des difficultés de repérage sur les objets tels que les sphères par exemple, qui ne présentent pas naturellement de direction privilégiée. L'aliassage est très important lorsque l'échantillonnage de l'image initiale est différent de celui qui sera restitué [GP91], c'est à dire presque toujours.

Certaines textures planes simulent localement des déformations géométriques, à l'instar des peintures en trompe-l'œil. Par exemple, la perturbation de la direction de la normale (figure 2.15-d), présentée par Blinn [Bli78] sous le nom de *bump mapping*, imite des surfaces bosselées. Les limites de la méthode sont cependant vite atteintes, car les incohérences sont nombreuses. En effet, tous les calculs de position étant effectués avec la surface initiale, les points sont mathématiquement tous à la même altitude, alors que visuellement ils semblent décalés. Les ombres ne correspondent pas à l'attente, et le plus gênant est que les bordures des objets restent lisses, en désaccord avec l'apparence volumique de la surface. Ces deux défauts sont visibles sur notre exemple.

2.6.2 Textures volumiques



(a) Marbre (b) Bois

FIG. 2.16 – *Exemples de textures volumiques*

Les textures volumiques sont des fonctions définissant pour tout l'espace la couleur des points des objets. Intuitivement, les objets sont sculptés dans un bloc de matière, sans problème de projection. Cette approche, présentéee par Gardner [Gar84] puis par Perlin [Per85], est particulièrement bien adaptée pour simuler des matériaux tels que le marbre (figure 2.16-a) ou le bois (figure 2.16-b).

Les diverses matières sont modélisées par des fonctions représentant la répartition des différents composants. Par exemple, le marbre est représenté par un substrat à l'intérieur duquel des veines plus foncées forment des figures aléatoires. Elle sont guidées par des séries de fonctions trigonométriques de fréquences différentes, qui décrivent des effets de vagues, et que viennent perturber des valeurs aléatoires. Le principe est similaire pour le bois, mais une corrélation supplémentaire organise les veines concentriquement autour de l'axe plus ou moins rectiligne d'un tronc imaginaire.

Ces modélisations sont empiriques, sans réelle justification géologique ou biologique, mais elles donnent des résultats visuellement intéressants sans trop pénaliser les temps d'exécution. Des méthodes plus récentes [GD95], basées sur l'analyse spectrale de modèles réels, fournissent des images beaucoup plus réalistes.

2.6.3 Hypertextures

Les textures planes ou volumiques classiques ne permettent pas la simulation correcte des phénomènes produisant une modification locale du relief, car elles introduisent des incohérences topographiques entre la géométrie simulée et l'image restituée. Les hypertextures, présentées par Perlin et Hoffert [PH89], apportent une réponse à ces problèmes. Elles sont de plus bien adaptées pour simuler des microstructures complexes très fines, telles que la fourrure par exemple.

Intuitivement, une hypertexture est une couche d'épaisseur connue plaquée sur un objet, à l'intérieur de laquelle sont placés les éléments visibles. Ces éléments ne sont pas modélisés explicitement, mais représentés par des fonctions de densité volumique qui obéissent aux lois statistiques du modèle imité. Un point est visible lorsque les densités cumulées sur le parcours d'un chemin lumineux (section 3.2) à l'intérieur de la couche de l'hypertexture dépassent un seuil donné. Le calcul de la densité totale requiert une discrétisation du chemin lumineux, très fine pour prendre en compte les détails très petits. C'est une opération très coûteuse, améliorée récemment par Dischler et Ghazanfarpour [DG95]. Des squelettes géométriques associés aux surfaces portent les informations de densité, qui sont calculées en chaque point par une interpolation pondérée en fonction de la distance aux squelettes.

2.7 Milieux participants

Jusqu'à maintenant, nous avons toujours supposé que la lumière se déplace en ligne droite d'une surface à une autre, et sans modification énergétique le long de son chemin, grâce à la conservation de la luminance (section 2.2.2.2). Cette hypothèse est vraie dans le vide, et elle peut être admise sans erreur notable dans de l'air propre et sec, pour des distances raisonnables. Par contre, il n'en va plus du tout de même en présence de poussières en suspension, de fumée ou de brume. Les phénomènes d'absortion et de dispersion deviennent prépondérants.

Les milieux participants modélisent les régions où se produisent ces phénomènes, dont les frontières ne sont pas clairement définies. Les lois physiques sont difficiles à appréhender de par leur complexité. Imaginons par exemple un nuage, constitué de milliards de gouttes d'eaux minuscules. Chaque rayon lumineux qui rencontre une goutte donne naissance à un rayon réfléchi et à un rayon réfracté, qui eux-même vont rencontrer une autre goutte dans leur voisinage. Ces événements se multiplient tant que l'énergie n'est pas complètement absorbée ou tant qu'un rayon ne sort pas du milieu. Les gouttes d'eau sont des réflecteurs de Fresnel presque parfaits, et le nombre d'interactions peut devenir très grand avant que l'absortion ne soit notable. Même en supposant que les gouttelettes soient de simples sphères, la quantité de calculs nécessaires au suivi des rayon est prohibitive, d'autant plus que leur nombre croît exponentiellement avec le nombre d'intersections.

De nombreux modèles théoriques ont été développés pour décrire certains types de milieux participants [PPS97], notamment les phénomènes atmosphériques tels que les fumées, les brumes et les nuages. Certains sont empiriques, alors que d'autres sont dérivés de complexes théories physiques. Ce domaine de recherche est encore en pleine évolution, pour augmenter le réalisme des effets simulés, mais aussi pour réduire les temps de calcul qui sont très conséquents. La plupart des méthodes proposées nécessitent une discrétisation très fine du parcours des rayons lumineux.

2.8 Quelques notions photométriques

2.8.1 Radiométrie vs. photométrie

Les grandeurs radiométriques, que nous avons présentées au début de ce chapitre, permettent l'analyse des ondes électromagnétiques et de leurs propriétés physiques. Cette approche masque cependant un aspect fondamental de la lumière : par définition, elle est *visible* par les humains. Elle n'est donc pas seulement un cas particulier d'onde électromagnétique (figure 2.17), car elle implique en plus des notions de sensibilité. Les images, qui sont des répartitions spatiales de lumière, existent uniquement parce que nous sommes capables de les percevoir, et éventuellement de leur donner un sens. Le récepteur, c'est à dire le système visuel humain, joue alors un rôle primordial. Il ne réagit pas linéairement aux stimuli énergétiques, donc les données radiométriques ne sont pas suffisantes pour décrire ces ondes particulières.

La photométrie est la science qui étudie la lumière, en prenant en compte les caractéristiques de la vision humaine. Elle intègre des notions physiques (les ondes électromagnétiques), des données biologiques (la structure de l'œil), et des comportements psychophysiologiques (interprétation par le cerveau des informations reçues). Les travaux de Rougeron [Rou98] et de Callet [Cal98] nous servent de liens avec la synthèse d'images, et nous présentons sommairement quelques notion psychovisuelles, notamment la sensation de couleur.

2.8.2 La vision humaine

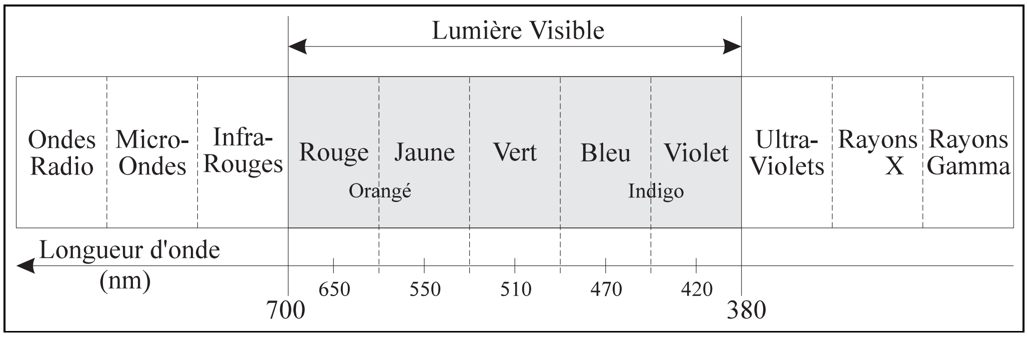


FIG. 2.17 – *Répartition spectrale des ondes électromagnétiques*

2.8.2.1 Le système visuel humain

Le système visuel humain intègre l'œil, le nerf optique et le cerveau, qui sont indissociables dès qu'il s'agit de décrire la sensation visuelle. L'œil transforme l'énergie lumineuse qu'il reçoit en influx nerveux. Le nerf optique les met en forme, et les transmet au cerveau, où ils sont analysés.

D'un point de vue physiologique simpliste (figure 2.18), l'œil peut être comparé à un appareil photographique[2]. La pupille, l'iris et le cristallin forment une lentille d'ouverture variable, qui projette une image sur la rétine. Elle est recouverte de micro-capteurs sensibles à la lumière et reliés au nerf optique : les cônes et les bâtonnets. La fovéa est la zone centrale de la rétine, là où se forme l'image la plus nette. La plupart des cônes sont répartis sur la fovéa.

Les cônes servent à la vision des couleurs (section 2.8.3) et à la vision diurne, dite *photopique*, alors que les bâtonnets, plus sensibles à la lumière mais non discriminants pour les couleurs, servent à la vision nocturne, dite *scotopique*. La vision intermédiaire, entre photopique et scotopique, est dite *mésopique*. Par la suite, nous ne considérerons que le cas de la vision photopique.

2. En fait, c'est plutôt l'appareil photographique qui peut être comparé à un œil.

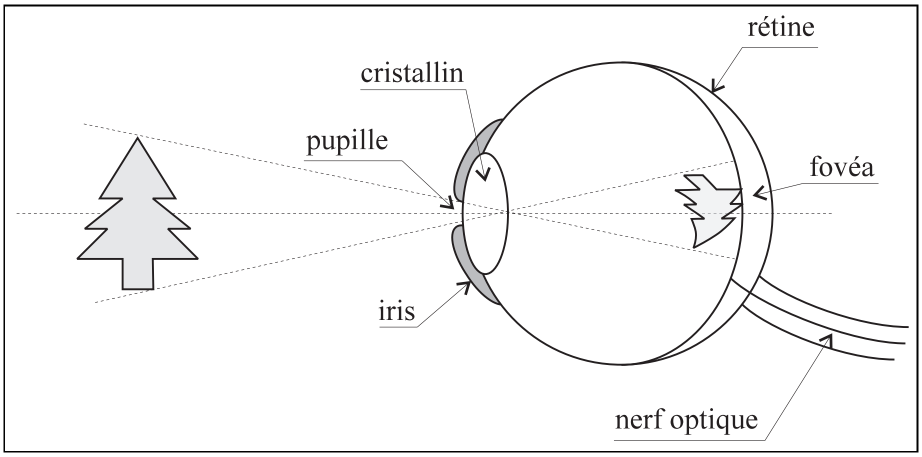


FIG. 2.18 – *Vue en coupe simplifiée d'un œil humain*

2.8.2.2 Grandeurs photométriques

Le système visuel humain peut percevoir les ondes électromagnétiques dont la longueur d'onde se situe entre 380 et 780 nm environ (figure 2.17). Ces limites varient d'un individu à l'autre, et évoluent aussi avec l'âge. Afin d'uniformiser les représentations, il a été défini par la *Commission Internationale d'Eclairage* (CIE) un observateur moyen, issu de nombreuses expériences. Toute les grandeurs photométriques, qui définissent la vision humaine, se rapportent à cet observateur moyen. Pour les différencier des grandeurs radiométriques, elles sont explicitement et systématiquement accompagnées de l'adjectif *lumineux*. Cette notation est lourde, mais évite toute ambiguïté lors de la désignation de notions homonymes presque semblables mais incompatibles entre elles.

La sensibilité visuelle n'est pas uniforme, ni même linéaire, pour les longueurs d'onde du domaine visible. La courbe de réponse en luminosité, en écartant pour l'instant toute notion de couleur, a sensiblement l'allure de la courbe L de la figure 2.19. Nous noterons $V(\lambda)$ cette valeur de la sensibilité relative pour longueur d'onde λ. $V(\lambda)$ est nulle en dehors du domaine visible.

Le **flux lumineux** $\phi(\lambda_0)$d'une source monochromatique de longueur d'onde λ_0 et de flux énergétique $\phi_e(\lambda_0)$ (section 2.2.1) est :

$$\phi(\lambda_0) \;=\; 680\, V(\lambda_0)\, \phi_e(\lambda_0)$$

Le flux lumineux s'exprime en lumen $[lm]$.

Pour une source polychromatique, dont le flux énergétique spectral pour une longueur d'onde λ est $\phi_e(\lambda)$, le flux lumineux est :

$$\phi \;=\; 680 \int_0^{+\infty} V(\lambda)\, \phi_e(\lambda)\, d\lambda$$

De la même façon que pour les grandeurs radiométriques, le flux lumineux permet de définir :

- la **luminance lumineuse** : $L(x, \vec{\omega}) \;=\; \dfrac{d^2\phi}{dA\, \cos\theta\, d\omega}$, exprimée en $[lm.m^{-2}.sr^{-1}]$;

- l'**éclairement lumineux** : $E(x) = \dfrac{d\phi}{dA}$, exprimé en lux $[lx]$;
- l'**intensité lumineuse** : $I(x, \vec{\omega}) = \dfrac{d\phi}{d\omega}$, exprimé en candela $[cd]$.

La luminance lumineuse permet de quantifier la *sensation lumineuse* B, avec la loi de Weber :

$$B = 25\, L^{1/3} - 17$$

Bodman a proposé une autre formulation, en prenant en compte le contraste de la zone observée par rapport à son voisinage, mais qui suppose que le fond soit de luminance lumineuse uniforme L_f :

$$B = 23 \log L - 5,5 \log L_f - 1,6$$

2.8.3 Sensation de couleur

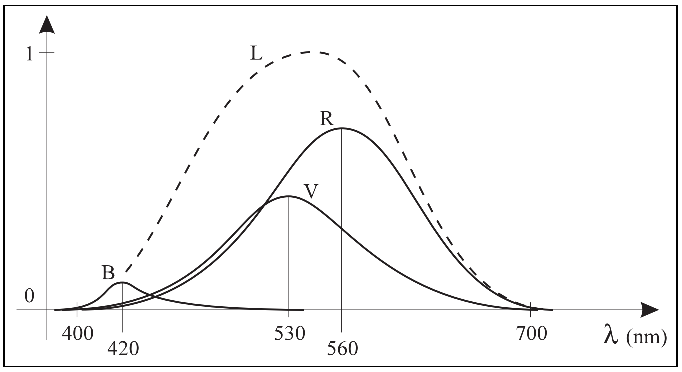


FIG. 2.19 – *Sensibilité relative des trois types de cônes et en luminosité*

Il existe trois sortes de cônes, qui réagissent à des gammes différentes de longueurs d'onde, dont les couleurs dominantes sont le rouge, le vert, et le bleu (courbes R, V et B de la figure 2.19). Les cônes rouges ont leur maximum de sensibilité pour des longueurs d'onde proches de 560 nm, les cônes verts aux environs de 530 nm, et les cônes bleus vers 420 nm. Lorsqu'ils sont soumis à un éclairement, la combinaisons des informations qu'ils transmettent au cerveau crée la sensation suggestive de couleur. La couleur n'est donc pas une entité physique clairement identifiable, mais une sensation propre à chaque individu due à la stimulation des cônes par les diverses fréquences de la lumière observée. Il est difficile de la décrire sans équivoque.

La longueur d'onde de la lumière induit la sensation de couleur. Lorsque la lumière est monochromatique, chaque fréquence correspond à une teinte saturée (figure 2.17). La lumière polychromatique crée des teintes moins pures, qui peuvent prendre toutes les nuances perceptibles par le système visuel, des teintes saturés jusqu'au blanc en passant par toutes les combinaisons et les couleurs pastels.

Le diagramme de chromaticité (figure 2.20) représente dans un plan l'ensemble des couleurs perceptibles. Les teintes saturées sont situées sur la bordure, où sont indiquées quelques longueurs d'onde correspondantes. Le centre du diagramme est la zone du blanc, qui est aussi une notion suggestive, et des diverses sources que l'on perçoit comme blanches quand elles

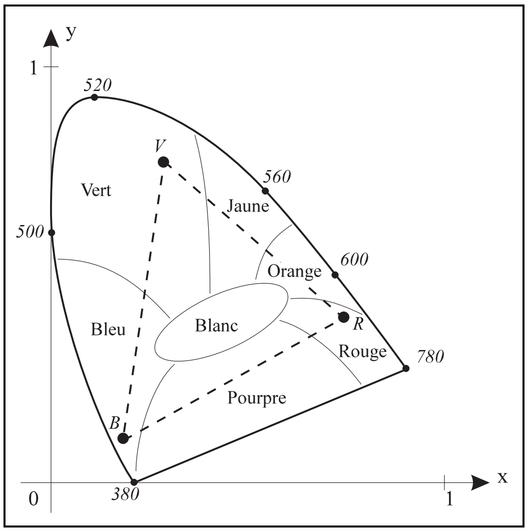


FIG. 2.20 – *Diagramme de chromaticité*

servent de référence pour comparer les autres couleurs d'une scène. Les régions intermédiaires sont les teintes pastels, de plus en plus claires lorsque l'on se dirige vers la zone blanche. Il n'y a aucune transition brusque, mais uniquement des dégradés. La position des frontière sur le graphique n'est qu'indicative et varie d'un individu à l'autre.

2.8.3.1 Trichromie

Des expériences, menées par la CIE, ont montré que l'ensemble des couleurs perceptibles peut être reconstitué en combinant selon les proportions appropriées trois ondes monochromatiques bien choisies. Pour cela, des observateurs cherchent les proportions énergétiques de trois sources monochromatiques qui créent la même sensation visuelle que la couleur d'un échantillon témoin.

La simulation est possible lorsque les trois couleurs sont primaires entre elles, c'est à dire lorsque aucune d'elles ne peut être reconstituée par une combinaison des deux autres. Elles forment alors la base d'un espace tridimensionnel de couleurs visibles. C'est le cas par exemple pour les longueurs d'onde de 645 nm (rouge), 526 nm (vert), et 444 nm (bleu).

La génération de certaines couleurs n'est pas possible par une simple addition des couleurs primaires. Il faut dans ce cas soustraire l'une d'elles du mélange, ce qui correspond en pratique à l'utiliser pour éclairer l'échantillon témoin. Cet espace des couleurs comporte donc des coordonnées négatives pour représenter toutes les nuances perceptibles. La description des couleurs avec ce modèle trichromatique n'est pas très intuitive.

2.8.3.2 Espace de couleurs XYZ

La CIE a adopté en 1931 un système trichromatique de représentation des couleurs qui supprime le recours à des coordonnées négatives, en introduisant trois couleurs primaires artificielles notées X, Y, et Z.

Les fonctions $\bar{x}(\lambda)$, $\bar{y}(\lambda)$, et $\bar{z}(\lambda)$ sont définies à partir d'un observateur moyen. Elles sont utilisées pour pondérer la distribution spectrale $P(\lambda)$ d'une couleur donnée, sur toute l'étendue du spectre visible :

$$X = \int_{380}^{780} P(\lambda)\bar{x}(\lambda)d\lambda \qquad Y = \int_{380}^{780} P(\lambda)\bar{y}(\lambda)d\lambda \qquad Z = \int_{380}^{780} P(\lambda)\bar{z}(\lambda)d\lambda$$

Par convention, la coordonnée Y représente la *luminance lumineuse*. Elle est sensiblement proportionnelle à la perception visuelle de la clarté d'une surface, du plus sombre au plus clair. Elle est utilisée pour transformer un image colorée en niveaux de gris.

La *chromaticité* définit une couleur en normalisant ses trois coordonnées entre 0 et 1 :

$$x = \frac{X}{X+Y+Z} \qquad y = \frac{Y}{X+Y+Z} \qquad z = \frac{Z}{X+Y+Z}$$

2.8.4 Autres modèles de couleur

2.8.4.1 Espace de couleurs RVB

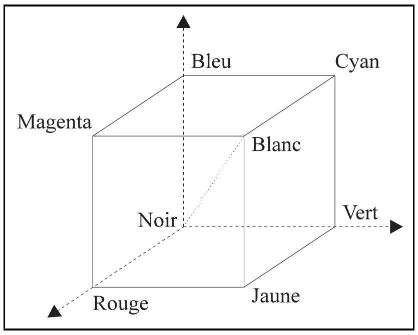


FIG. 2.21 – *Espace de couleurs RVB*

Chaque pixel d'un écran cathodique couleur est composé de trois luminophores, colorés en rouge, en vert et en bleu. La couleur d'un pixel est contrôlée en ajustant les proportions de ces trois types d'émissions, c'est la *synthèse additive*. Les luminophores n'émettent pas des ondes monochromatiques parfaites, mais des rayonnements dont la couleur est située à l'intérieur du diagramme de chromaticité, en général à proximité de la bordure. Ce sont par exemple les points R, V, et B de la figure 2.20, qui forment un triangle inclus dans le diagramme.

La synthèse additive permet d'obtenir toutes les couleurs situées à l'intérieur du triangle. Les nuances du diagramme de chromaticité ne sont pas toutes accessibles, car la zone couverte est plus petite que l'ensemble des couleurs perceptibles. Elle dépend des caractéristiques physiques des luminophores. Il n'y a donc pas unicité de représentation d'une même couleur entre deux écrans différents.

L'*espace de couleurs RVB* est la représentation, sous forme d'un cube unitaire, de l'ensemble des couleurs engendrées par les trois primaires rouge, verte et bleue (figure 2.21). Une couleur est définie par trois coordonnées (r, v, b) comprises entre 0 et 1, qui indiquent les proportions de rouge, vert, et bleu qui la composent. Bien qu'il ne couvre pas l'ensemble complet des couleurs, cet espace est très utilisé en synthèse d'images, car il est très simple et calque approximativement le fonctionnement d'un écran d'ordinateur.

La luminance lumineuse associée à un triplet (r, v, b) est approximativement :

$$L_v = 0.299\,r \; + \; 0.587\,v \; + \; 0.114\,b \tag{2.19}$$

Les valeurs numériques des coefficients sont obtenues à partir des caractéristiques standard des luminophores de téléviseurs. Cette formule permet de convertir très rapidement une image couleur en une image en niveaux de gris. La valeur L_v correspond sensiblement à la composante Y de l'espace XYZ.

2.8.4.2 Espace de couleurs CMJ

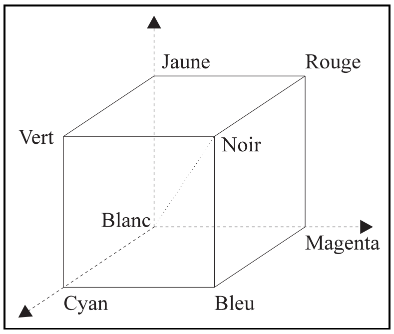


FIG. 2.22 – *Espace de couleurs CMJ*

L'*espace de couleurs CMJ* (Cyan, Jaune, Magenta) est le complément de l'espace de couleurs RVB, pour la *synthèse soustractive*. La sensation de couleur n'est plus due à l'émission de trois sources primaires, mais à l'absorption par trois éléments d'une partie du spectre de la lumière blanche. Il est utilisé pour les impressions en couleur, où les trois encres colorées absorbent sélectivement la lumière réfléchie par le papier. Le cube CMJ (figure 2.22) se déduit du cube RVB par un changement de repère.

En pratique, une quatrième composante est souvent ajoutée : le noir. C'est le principe de la quadrichromie, très utilisée en imprimerie. En effet, les encres réelles n'ont pas des propriétés d'absorption parfaites. La combinaison des trois teintes ne permet alors pratiquement jamais d'obtenir un noir parfait, mais plutôt une couleur marron plus ou moins foncée.

2.8.4.3 Espaces de couleurs uniformes

Les espaces de couleurs XYZ, RVB, et CMJ ont pour défaut majeur de ne pas être visuellement uniformes. La distance entre les points qui représentent deux couleurs n'est pas du tout proportionnelle à la différence de sensation psychovisuelle qu'elles induisent.

Différents modèles d'espaces pseudo-uniformes sont proposés par la CIE, dont les plus utilisés actuellement sont les espaces *Luv* et *Lab*, et plus récemment *LLab*. Ces différentes déclinaisons marquent l'évolution encore inachevée des connaissances sur le fonctionnement du système visuel humain. Ils sont obtenus en déformant l'espace XYZ de façon à ce qu'il épouse la forme d'un espace visuel. Ce sont des opérations non linéaires coûteuses.

On peut noter que toutes ces modélisations dissocient un canal L, représentant la luminance lumineuse, de deux canaux contenant les informations chromatiques. Le cerveau fonctionne selon le même schéma lors de l'analyse des informations visuelles.

2.8.4.4 Modèle spectral

Les modèles de couleurs trichromatiques ne permettent pas de définir complètement les propriétés de réflexion et de réfraction des matériaux. En effet, il n'y a pas bijection entre l'ensemble des couleurs perceptibles et l'ensemble des spectres énergétiques, c'est le phénomène de métamérisme. Le triplet représentant la sensation visuelle unique d'une couleur précise peut correspondre à plusieurs spectres différents, qui n'ont pas le même comportement physique lors des interactions avec la matière.

Un spectre lumineux est composé d'une infinité de longueur d'ondes situées dans le domaine visible. Il ne peut pas être physiquement assimilé à un triplet dans un espace de couleurs. La conversion en sensation de couleur n'a de sens que lors de la toute dernière étape de la création d'une image, lorsque l'on considère l'énergie atteignant effectivement l'œil. Tous les autres échanges énergétiques ne font pas intervenir la sensibilité du système visuel, et ils doivent donc être abordés uniquement d'un point de vue radiométrique.

Dans un *modèle spectral*, tous les matériaux de la scène sont décrits par leurs BRDF et BTDF spectrales, et toutes les sources le sont par leur distribution énergétique spectrale. Informatiquement, la discrétisation des spectres est nécessaire. Le pas d'échantillonnage optimal couramment admis est de $5nm$, ce qui conduit à gérer un très grand nombre de données numériques. Des travaux récents [RP97] proposent une méthode d'échantillonnage adaptatif intégrant des critères perceptuels, afin de raffiner la discrétisation uniquement lorsque l'erreur commise risque d'être visible lors de l'affichage.

Chapitre 3

Calcul de l'éclairement global

3.1 Eclairement global

L'éclairement global d'une scène est obtenu par la prise en compte de toutes les interactions de la lumière avec la matière. Dès leur mise en route, les sources éclairent les objets, qui à leur tour réfléchissent une partie de la lumière reçue pour s'éclairer les uns les autres. Une situation d'équilibre de la répartition spatiale de l'énergie est atteinte presque instantanément, car la vitesse de la lumière est très grande. La phase transitoire est beaucoup trop brève pour être perçue par un observateur humain, donc, par la suite, nous considèrerons uniquement des situations équilibrées.

Durant un intervalle de temps donné, les sources lumineuses émettent un très grand nombre de photons. Par exemple, une source monochromatique, de longueur d'onde 500 nm et d'une puissance de 1 W seulement, émet environ $2,5\ 10^{18}$ photons par seconde. Ils sont indépendants les uns des autres, mais leurs propriétés physiques, principalement la fréquence de l'onde électromagnétique associée (section 2.1.1.1) et la direction d'émission, obéissent à des lois statistiques qui dépendent uniquement des caractéristiques de la source. Les photons sont fortement affectés lors de leurs interactions avec la matière, que ce soit lorsqu'ils se déplacent à travers le milieu ambiant ou bien lorsqu'ils rencontrent des objets. Ils peuvent être déviés, ou bien absorbés, ce qui transforme l'énergie lumineuse en énergie thermique. Les photons absorbés peuvent être réémis avec d'autres caractéristiques énergétiques et après un délai temporel, pour les phénomènes de phosphorescence. Lorsqu'ils atteignent l'œil de l'observateur, ils sont convertis en influx nerveux qui déterminent les sensations visuelles. Tous ces phénomènes suivent eux-aussi des lois statistiques, qui dépendent à la fois des caractéristiques des photons et de celles de la matière rencontrée. Le même processus se répète récursivement pour les photons déviés ou réémis, jusqu'à l'absorption.

Dans ces conditions, calculer l'éclairement global dans une scène de synthèse d'images est excessivement simple. Il suffit de générer une quantité suffisante de photons, de leur attribuer les propriétés adéquates, puis de les envoyer depuis les sources vers la scène où ils subissent diverses interactions. Les caractéristiques électromagnétiques de la scène étant parfaitement connues, l'image est facilement générée, lorsque la situation d'équilibre est atteinte, en comptabilisant les photons présents en chaque point et en calculant l'énergie qu'ils émettent vers l'observateur.

La méthode précédente se heurte malheureusement à des problèmes pratiquement insurmontables. Le nombre de photons et d'interactions à gérer est gigantesque, et dépasse largement les capacités de mémoire des ordinateurs actuels. Même avec des machines très performantes, la durée des calculs serait prohibitive. De plus, la description de la scène et de ses caractéristiques géométriques et électromagnétiques complètes, en tout point, aboutirait à des bases de données de taille démesurée, en supposant que toutes ces connaissances soient disponibles.

En synthèse d'images, la recherche d'un plus grand réalisme impose de calculer le plus fidèlement possible l'éclairement global dans les scènes. La complexité du problème est telle qu'il faut élaborer des stratégies simplificatrices, en limitant l'introduction d'erreurs, pour les maintenir si possible en dessous du seuil de perception de l'observateur. Nous présentons dans ce chapitre les principales approches développées dans ce domaine.

3.2 Les chemins lumineux

Les *chemins lumineux*, introduits en synthèse d'images par Heckbert [Hec90], caractérisent sous la forme d'une grammaire les différentes interactions subies par la lumière le long d'un trajet entre deux points donnés. Nous utiliserons cette terminologie pour différencier les diverses composantes lumineuses.

Le vocabulaire comporte quatre symboles :

L	Source de lumière (Light source)
D	Réflexion ou réfraction diffuse
S	Réflexion ou réfraction spéculaire
E	Point visé (Eye)

Ainsi, un photon partant de L, pour arriver en E, après avoir subi deux réfexions diffuses et une réfraction spéculaire, suit un chemin lumineux de type $LDDSE$. Ces quatre symboles sont suffisants pour regrouper les chemins en classes distinctes représentant les composantes lumineuses que nous étudierons plus loin. Si nécessaire, on peut cependant imaginer une décomposition plus fine, en introduisant les symboles adéquats, par exemple pour distinguer les réflexions des réfractions ou les sources ponctuelles des autres [Sch92, Vea97].

Par convention dans l'écriture des expressions régulières, les symboles suivants décrivent les agencements des mots :

X	X, exactement une fois
X^-	X, une fois au plus
X^+	X, répété une fois au moins
X^*	X, répété zéro fois ou plus
$X\|Y$	X ou Y

Exemples :

LD^-S^*E	$LDSE$ ou LSE ou $LSSE$ ou $LDSSSE$ ou ...
LD^*SE	$LDSE$ ou $LDDDSE$ ou LSE ou ...
$L(D\|S)E$	LDE ou LSE
$L(D\|S)^+E$	LDE ou LSE ou $LDSSE$ ou $LDDDE$ ou $LSDS$ ou ...
$L(D\|S)^*E$	LE ou LDE ou LSE ou $LDSSE$ ou $LDDDE$ ou $LSDS$ ou ...

Les divers chemins lumineux sont des expressions régulières obtenues en combinant ces symboles. Ils partent tous de la source L, pour aboutir au point E, qui est en général l'observateur, apres avoir subi un nombre variable de réflexions ou réfractions de type D ou S.

L'expression générique d'un chemin lumineux quelconque est donc du type :

$$L\ (D|S)^*\ E$$

3.3 Equation de rendu

3.3.1 Définition

L'équation intégrale de rendu est dérivée du modèle de Kajiya [Kaj86], qui fut le premier à formaliser explicitement le problème de l'éclairement global. Elle est modifiée pour s'exprimer en terme de luminance, et non d'intensité comme dans la version originale. En tout point de la scène, l'équation de rendu est :

$$L_r(x,\vec{\omega_r}) = L_e(x,\vec{\omega_r}) + \int_{\Omega_i} f_r(x,\vec{\omega_i}\rightarrow\vec{\omega_r})\, L_i(x,\vec{\omega_i})\, \cos\theta_i\, d\omega_i \tag{3.1}$$

Le terme $L_e(x,\vec{\omega_r})$ est la luminance émise depuis le point x dans la direction $\vec{\omega_r}$, lorsque l'objet est une source, et le deuxième terme est constitué par l'équation de réflectance déjà présentée (équation 2.6). Cette formulation est récursive, car le calcul de la luminance incidente $L_i(x,\vec{\omega_i})$ fait lui-aussi intervenir l'équation de rendu, en un autre point de la scène, situé dans la direction $\vec{\omega_i}$. L'équation de rendu indique que la luminance réfléchie est égale à la luminance propre, plus la luminance réfléchie due à la lumière incidente selon toutes les directions possibles.

Cette équation est le point de départ de tous les algorithmes actuels de calcul de l'éclairement global. Même si les solution proposées divergent fortement dans leur approche de la résolution du problème, ces algorithmes cherchent tous à le résoudre le plus rapidement possible, et en minimisant les erreurs commises.

3.3.2 Problème dual

3.3.2.1 Potentiel

Le rôle des sources n'apparaît pas explicitement dans l'équation de rendu, hormis le terme L_e pour le point courant s'il appartient lui-même à une source. Elles sont pourtant fondamentales, puisque ce sont elles qui apportent l'énergie dans la scène. Pour appliquer certaines méthodes, il peut être utile de modifier la formulation, afin qu'elles interviennent directement [PB95]. Dans ce but, il est possible de développer l'équation de rendu relativement aux sources :

$$\begin{aligned} L_r(x,\vec{\omega_r}) \;=\; & \left[L_e(x,\vec{\omega_r})\right] \\ & + \int_{\Omega_{x,i}} f_r(x,\vec{\omega_y}\rightarrow\vec{\omega_r})\, \left[L_e(y,\vec{\omega_y})\right] \cos\theta_y\, d\omega_y \\ & + \int_{\Omega_{x,i}} f_r(x,\vec{\omega_y}\rightarrow\vec{\omega_r}) \cos\theta_y \int_{\Omega_{y,i}} f_r(y,\vec{\omega_z}\rightarrow\vec{\omega_y})\, \left[L_e(z,\vec{\omega_z})\right] \cos\theta_z\, d\omega_z \\ & + \ \ldots \end{aligned} \tag{3.2}$$

où y est un des points vus depuis x dans la direction ω_y, z est un des points vus depuis y dans la direction ω_z, et où $\Omega_{x,i}$ et $\Omega_{y,i}$ sont les ensembles des directions incidentes en x et y.

Le principe de conservation de l'énergie assure la convergence de ce développement infini, mais il n'est pas commode à utiliser sous cette forme. La notion de *potentiel*, introduite initialement par Pattanaik [PM93], permet une écriture plus compacte.

Soit y un point sur une source de lumière, dA_y l'aire élémentaire entourant le point y, et $\vec{\omega_y}$ une direction en y orientée d'un angle θ_y par rapport à dA_y.

Le potentiel $\mathcal{W}(y,\vec{\omega_y})$ est une grandeur représentant, par définition, la luminance différentielle en x, dans une direction $\vec{\omega_x}$, due à une unité de puissance ($1\ W$) émise depuis l'aire élémentaire dA_y, en y, dans la direction $\vec{\omega_y}$. Le potentiel exprime la contribution de l'énergie émise depuis le point y à la luminance en x dans la direction $\vec{\omega_x}$.

La puissance propre émise depuis y dans la direction $\vec{\omega_y}$ est :

$$d^2\phi(y,\vec{\omega_y}) = L_e(y,\vec{\omega_y})\cos\theta_y d\omega_y dA_y$$

donc la luminance différentielle en x, dans la direction $\vec{\omega_x}$, et due à l'élément de surface dA_y est :

$$d^2L(x,\vec{\omega_x}) = L_e(y,\vec{\omega_y})\cos\theta_y d\omega_y dA_y \mathcal{W}(y,\vec{\omega_y})$$

Ainsi, la luminance cumulée due à tous les points de la source S, dans l'ensemble Ω_y des directions autour d'elle, est :

$$L(x,\vec{\omega_x}) = \int_{y\in S}\int_{\Omega_y} L_e(y,\vec{\omega_y})\mathcal{W}(y,\vec{\omega_y})\cos\theta_y d\omega_y dA_y$$

Cette définition s'étend immédiatement à n'importe quel point de l'environnement E, qu'il soit sur une source ou non. La formulation générique est donc :

$$L(x,\vec{\omega_x}) = \int_{y\in E}\int_{\Omega_y} L_e(y,\vec{\omega_y})\mathcal{W}(y,\vec{\omega_y})\cos\theta_y d\omega_y dA_y \tag{3.3}$$

La lumière quittant y dans la direction $\vec{\omega_y}$ peut quitter x dans la direction $\vec{\omega_x}$ après zéro ou plus réflexions. Le cas zéro réflexion survient lorsque $(x,\vec{\omega_x}) = (y,\vec{\omega_y})$. Comme l'équation de rendu, l'expression générique du potentiel comporte donc deux partie : un terme direct et un terme indirect intégral. Elle s'obtient en égalisant l'équation 3.2 avec le développement de l'équation 3.3 :

$$\mathcal{W}(y,\vec{\omega_y}) = \mathcal{W}_e(y,\vec{\omega_y}) + \int_{\Omega_z} f_r(z,\vec{\omega_y}\to\vec{\omega_z})\mathcal{W}(z,\vec{\omega_z})\cos\theta_z d\omega_z \tag{3.4}$$

avec :

$$\mathcal{W}_e(y,\vec{\omega_y}) = \begin{cases} \dfrac{1}{dA_y\cos\theta_y d\omega_y} & \text{si} \quad (x,\vec{\omega_x}) = (y,\vec{\omega_y}) \\ 0 & \text{sinon} \end{cases}$$

La similitude d'écriture avec l'équation de rendu 3.1 est très importante. En fait, ces deux approches sont duales l'une de l'autre. Dans un cas, on considère en tout point la lumière incidente depuis toutes les directions, alors que dans l'autre cas on considère en tout point la lumière émise dans toutes les directions.

3.3.2.2 Liens entre l'équation de rendu et l'équation des potentiels

3.3.2.2.1 Description par opérateurs Dans sa thèse, Sillion [Sil89] propose l'introduction d'un *opérateur intégral de réflexion*, afin de formaliser des algorithmes de calcul d'images tels que le lancer de rayons ou la radiosité. Cette notion est maintenant étendue à l'équation de rendu et à l'équation des potentiels.

Opérateur pour l'équation de rendu : En utilisant un opérateur intégral T, l'équation de rendu 3.1 s'écrit :

$$L = L_e + TL$$

L'opérateur intégral T caractérise l'effet d'une réflexion de la lumière incidente. Il est défini par :

$$(Tf)(x,\vec{\omega_x}) = \int_{\Omega_x} f(y,\vec{\omega_y})\, f_r(x,\vec{\omega_y}{\rightarrow}\vec{\omega_x})\, \cos\theta_y\, d\omega_y$$

Opérateur pour l'équation des potentiels : En utilisant un opérateur intégral T^*, l'équation des potentiels 3.4 s'écrit :

$$\mathcal{W} = \mathcal{W}_e + T^*\mathcal{W}$$

L'opérateur intégral T^* est défini par :

$$(T^*f)(x,\vec{\omega_x}) = \int_{\Omega_y} f(y,\vec{\omega_y})\, f_r(x,\vec{\omega_x}{\rightarrow}\vec{\omega_y})\, \cos\theta_y\, d\omega_y$$

3.3.2.2.2 Equations adjointes Soient deux opérateurs intégraux T et T^*, et les équations correspondantes :

$$\begin{cases} f &= f_e + Tf \\ g &= g_e + T^*g \end{cases}$$

Ces deux opérateurs sont *adjoints* l'un de l'autre, pour un produit scalaire $< \cdot, \cdot >$, si et seulement si :

$$< f, g_e > = < f_e, g >$$

pour tous f_e et g_e.

Si A est l'ensemble des surfaces de la scène, alors un produit scalaire est défini par :

$$< f, g > = \int_A \int_{\Omega_x} f(x,\vec{\omega_x})\, g(x,\vec{\omega_x})\, \cos\theta_x d\omega_x dA_x$$

Avec ce produit scalaire, en utilisant l'équation de rendu et l'équation des potentiels, on remarque que le flux énergétique ϕ s'exprime de deux façons :

$$\phi = < L, W_e > = < L_e, W >$$

Ces deux équations sont donc adjointes l'une de l'autre. Ce résultats est utilisé par de nombreux auteurs ([PM95], [Laf96], [Vea97], ... entre autres), pour formaliser les algorithmes de rendu, et justifier mathématiquement les procédés choisis.

3.4 Algorithmes de rendu réaliste

L'amélioration du réalisme des images de synthèse est fortement dépendante, entre autre [1], de l'évaluation correcte de l'éclairement global dans les scènes. Il s'agit de prendre en compte toutes les interactions entre les sources et les objets, et entre les objets eux-mêmes. L'équation de rendu (équation 3.1) modélise complètement les phénomènes à simuler, mais elle ne possède pas de solution analytique exploitable.

On distingue principalement deux familles d'algorithmes, qui abordent différemment ce problème.

- Les méthodes d'éléments finis sont basées sur une décomposition complète de la scène en éléments de petite taille, assez fins pour admettre que la répartition énergétique soit homogène sur leur surface. C'est un procédé classique en physique et en analyse numérique. En synthèse d'images, la méthode d'éléments finis la plus célèbre est la *radiosité* (section 3.5.1).

- Le problème de l'éclairement global étant modélisé par une intégrale, l'équation de rendu, il peut se résoudre par une méthode de Monte Carlo (annexe A). La solution est calculée en quelques points particuliers, puis est étendue à leur voisinage. Le *lancer de rayons* (section 3.6.1) peut être considéré comme un cas particulier de méthode de Monte Carlo (section 3.6.6).

Actuellement, aucune des deux approches précédentes n'est en mesure de résoudre efficacement, et dans tous les cas de figure, le problème de l'éclairement global. Toutes les deux introduisent des simplifications qui génèrent des erreurs, plus ou moins bien contrôlées. Cependant, ce sont des procédés complémentaires, car les points faibles de l'une sont souvent les points forts de l'autre, hormis pour les temps de calculs, pratiquement aussi désastreux dans les deux cas.

Des *méthodes mixtes* (section 3.7) tentent de concilier les avantages respectifs des deux approches classiques.

Les méthodes les plus courantes sont décrites dans la suite de ce document.

3.5 Calcul par une méthode d'éléments finis

3.5.1 Radiosité classique

L'algorithme de la radiosité est très populaire en synthèse d'image. Il a donné lieu à de nombreuses publications, et quelques ouvrages lui sont entièrement consacrés [CW93, SP94]. La méthode a été initialement utilisée en physique, dans les années 1950, pour simuler les transferts thermiques. Les premiers à l'utiliser en informatique graphique furent Goral et al. [GTGB84] et Nishita et al. [NN85]. C'était la première méthode pratique prenant explicitement en compte les interréflexions diffuses.

1. Les autres problèmes concernent principalement la modélisation géométrique des objets, et la simulation spectrale de la lumière et des matériaux.

3.5.1.1 Principe général

3.5.1.1.1 Hypothèses simplificatrices La méthode de la radiosité consiste à résoudre l'équation de rendu (équation 3.1) en la simplifiant, pour se limiter à des cas plus faciles à résoudre. Les simplifications sont de deux types : la facétisation de la scène et la prise en compte partielle des BRDF des matériaux.

Discrétisation géométrique La scène complète est remplacée par des assemblages de facettes planes, en général des triangles ou des quadrilatères, suffisamment petites pour que la luminance de leur surface soit considérée comme constante. L'ensemble infini des directions incidentes Ω_i, en un point x quelconque, est alors remplacé par l'ensemble fini des directions des facettes visibles depuis x. L'expression intégrale devient une somme finie.

Simplification de la BRDF La radiosité classique suppose que toutes les surfaces sont parfaitement diffuses (ou lambertiennes) (section 2.4.1.4.1). Dans ce cas, la BRDF est constante, indépendante des directions d'incidence et de réflexion. Elle a pour valeur $\frac{\rho_d}{\pi}$ (équation 2.9), où ρ_d est une constante comprise entre 0 et 1 caractérisant la réflexion diffuse. Le terme intégral de l'équation de rendu peut alors être factorisé.

Seuls les chemins lumineux de type LD^*E sont donc pris en compte. Malgrè cette simplification drastique, les images produites sont souvent visuellement satisfaisantes, car la plupart des matériaux des scènes courantes ont un comportement plutôt diffus. Il reste cependant impossible de reproduire ainsi les reflets métalliques, et les réfractions, ce qui limite fortement le champ d'application de la radiosité, dans sa version classique. Un mirroir plan parfait peut toutefois être simulé, en générant le symétrique virtuel de la scène par rapport au plan du mirroir [Sil89]. Le nombre d'objets intervenant effectivement dans les calculs est alors doublé.

3.5.1.1.2 Expression analytique L'équation de rendu est :

$$L_r(x,\vec{\omega_r}) \;=\; L_e(x,\vec{\omega_r}) \;+\; \int_{\Omega_i} f_r(x,\vec{\omega_i}\!\rightarrow\!\vec{\omega_r})\, L_i(x,\vec{\omega_i})\, \cos\theta_i\, d\omega_i$$

L'indépendance directionnelle de la BRDF, et donc des luminances émises par les surfaces, permet d'écrire :

$$L_r(x) \;=\; L_e(x) \;+\; \frac{\rho_d(x)}{\pi} \int_{\Omega_i} L_i(x)\, \cos\theta_i\, d\omega_i$$

Cette équation s'exprime classiquement en terme de radiosité (d'où le nom de la méthode), appelée aussi exitance (section 2.2.4). La radiosité, indépendante de la direction, est bien adaptée pour décrire la répartition énergétique sur les facettes. Dans ce cas d'indépendance directionnelle, la radiosité $B(x)$ est liée linéairement à la luminance $L(x)$: $B(x) = \pi\, L(x)$.

L'équation de rendu devient alors :

$$B(x) \;=\; E(x) + \rho_d(x) \int_{y\in S} B(y)\, \frac{\cos\theta\cos\theta'}{\pi\, r^2}\, V(x,y)\, dy$$

où $B(x)$ est la radiosité en x, E(x) est l'exitance propre en x, S est l'ensemble des surfaces de la scène, et $V(x,y)$ est un terme de visibilité valant 1 lorsque les deux points x et y sont mutuellement visibles et 0 sinon. Les angles θ et θ' indiquent les angles entre les normales aux surfaces en x et en y (figure 3.1).

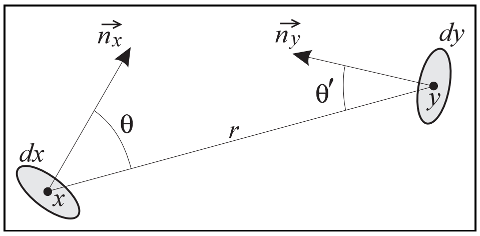


Fig. 3.1 – *Géométrie entre deux facettes*

3.5.1.1.3 Les facteurs de forme Un *facteur de forme* est une grandeur purement géométrique qui quantifie les échanges énergétiques entre deux facettes, dans le cas qui nous concerne ici des radiations purement diffuses. Le facteur de forme $F_{i,j}$ est la proportion de l'énergie émise par la facette P_i, d'aire A_i, qui est reçue par la facette P_j. Il s'exprime par :

$$F_{i,j} = \frac{1}{A_i} \int_{x \in P_i} \int_{y \in P_j} \frac{\cos\theta \cos\theta'}{\pi\, r^2} V(x,y)\, dx\, dy \tag{3.5}$$

Les facteurs de forme vérifient la loi de réciprocité : $A_i F_{i,j} = A_j F_{j,i}$.

Les facteurs de forme, associés à la discrétisation de la scène, permettent d'écrire l'*équation de radiosité*. Bien que très concise, elle décrit les échanges énergétiques globaux dans la scène, avec les hypothèses simplificatrices de la radiosité :

$$B_i = E_i + \rho_i \sum_{j=1}^{n} F_{i,j}\, B_j \tag{3.6}$$

où B_i est la radiosité de la $i^{\grave{e}me}$ facette de la scène (qui en compte n au total).

3.5.1.1.4 Système matriciel d'équations L'équation de radiosité peut s'écrire sous forme matricielle :

$$\begin{pmatrix} 1-\rho_1 F_{1,1} & -\rho_1 F_{1,2} & \cdots & -\rho_1 F_{1,n} \\ -\rho_2 F_{2,1} & 1-\rho_2 F_{2,2} & \cdots & -\rho_2 F_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ -\rho_n F_{n,1} & -\rho_n F_{n,2} & \cdots & 1-\rho_n F_{n,n} \end{pmatrix} \begin{pmatrix} B_1 \\ B_2 \\ \vdots \\ B_n \end{pmatrix} = \begin{pmatrix} E_1 \\ E_2 \\ \vdots \\ E_n \end{pmatrix} \tag{3.7}$$

Soit encore : $MB = E$. La solution du système est donc simplement : $B = M^{-1}E$.

La matrice M est de taille n^2, avec n très grand (une scène comporte souvent au moins plusieurs dizaines de milliers de facettes). Son stockage en mémoire et son inversion ne sont donc pas envisageables en pratique. Par contre, la matrice est à diagonale strictement dominante [SP94], c'est à dire que $M_{i,i} > \sum_{j=1,j\neq i}^{n} |M_{i,j}|$ pour tout i. De ce fait, les méthodes de résolution itérative de Gauss-Seidel ou de Southwell sont utilisables, et elles convergent relativement rapidement vers la solution, sans que la matrice n'ait besoin d'être stockée entièrement en mémoire.

3.5.1.2 Calcul des facteurs de forme

Il n'existe en général pas de formule analytique pour calculer le facteur de forme entre deux facettes, mais plusieurs méthodes de calculs approchés existent. C'est une tâche fastidieuse, car les facteurs de forme à calculer sont très nombreux.

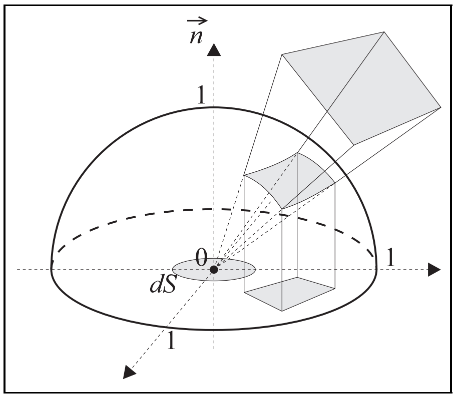


FIG. 3.2 – *Analogie de Nusselt*

L'analogie de Nusselt (1928) indique que le facteur de forme entre un élément de surface différentiel dS et un carreau de taille fini a une définition géométrique simple. Sa valeur est égale à l'aire de la projection sur le plan contenant dS de la projection, sur un hémisphère unitaire, de la facette (figure 3.2). Cette approximation reste utilisable, avec une erreur modérée, entre des facettes de taille finie, si elles sont assez petites et éloignées l'une de l'autre.

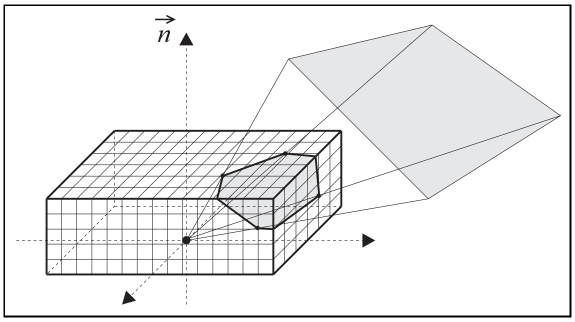


FIG. 3.3 – *Méthode de l'hémicube*

La méthode de l'hémicube [CG85] exploite l'analogie de Nusselt. Un demi-cube discrétisé est centré en x, dont la contribution de chacune des mailles est connue par la méthode de Nusselt. La facette est projetée sur les cinq faces du cube, afin de déterminer les mailles qu'elle recouvre, dont les contributions au facteur de forme sont ajoutées. Cette méthode est rapide, d'autant plus que la projection des facettes peut s'effectuer en détournant la carte d'affichage de l'ordinateur et son tampon de profondeur cablé, ce qui détermine en même temps les masquages entre facettes. Elle n'est cependant pas très précise, car le domaine de validité de l'analogie de Nusselt n'est pas toujours respecté, et la discrétisation du cube introduit un aliassage qui peut causer des artefacts visibles.

Des optimisations de cette méthode ont été proposées. Par exemple, le plan de Sillion [SP89] consiste à remplacer l'hémicube par un plan discrétisé parallèle à la surface. Les faces manquantes introduisent une erreur minime, dont un majorant qui dépend de la taille et de la position du plan est estimé. Une autre amélioration consiste à orienter la grande diagonale de l'hémicube vers la facette [BKP91]. De cette façon, la facette se projette au plus sur trois faces du cube au lieu de cinq, ce qui limite le nombre total d'opérations.

La méthode de Monte Carlo utilise le lancer de rayons pour déterminer la proportion de rayons qui touchent une facette, qui est proportionnelle à la valeur du facteur de forme. Cette méthode est plus précise et plus souple que l'hémicube, mais sa mise en œuvre est plus coûteuse.

3.5.1.3 Affichage d'une image de radiosité

La méthode de la radiosité calcule une valeur pour chaque maille de la scène discrétisée. La visualisation de la scène consiste alors à reconstituer la fonction de radiosité pour tous les points visibles. L'affichage brut des valeurs calculées rend fortement visibles les facettes, ce qui est particulièrement gênant pour les objets courbes. Il est préférable d'effectuer un lissage au moment de l'affichage, par la méthode de Gouraud [Gou71] par exemple. Il atténue les défauts visuels, mais introduit une erreur non contrôlée supplémentaire. Cette opération est bien souvent réalisée très rapidement par les cartes graphiques.

3.5.2 Améliorations de la méthode de la radiosité

3.5.2.1 Radiosité progressive

La radiosité progressive [CCWG88] est une modification de la méthode itérative de résolution du système matriciel (équation 3.7). A chaque étape, les facettes qui émettent la plus grande quantité d'énergie sont privilégiées, notamment les sources. Le résultat final est identique à celui des autres méthodes. Cependant, comme les facettes les plus influentes distribuent en priorité leur énergie, les images obtenues dès les premières étapes sont visuellement proches de l'image définitive.

3.5.2.2 Maillage de discontinuité

Un maillage régulier de la scène n'est pas optimal. Il est soit trop fin dans les zones régulières, soit trop grossier dans les zones perturbées de l'image. Le maillage de discontinuité [LTG92] consiste à subdiviser plus finement les régions de pénombre. Leur position est déterminée en prétraitement, par projection des facettes sources sur les autres, pour détecter les zones éclairées, partiellement éclairées, et d'ombre. C'est une opération très lourde pour les scènes importantes, mais qui améliore considérablement les résultats.

3.5.2.3 Radiosité hiérarchique

La radiosité hiérarchique [HSA91] représente les objets par une arborescence de facettes. Chaque facette est éventuellement subdivisée en éléments plus petits, lorsque cela est utile. Le principe est que le niveau de subdivision nécessaire dépend de la distance depuis laquelle un groupe d'objet est vu. Par exemple, un crayon posé sur une table n'a pas d'influence notable quant à la lumière rayonnée de la table sur le plafond. Par contre, le crayon ne doit pas être ignoré pour détecter la zone de pénombre qu'il crée sur la table. Ce procédé donne lieu à de multiples travaux d'amélioration, notamment pour tenir compte des visibilités et occlusions à l'intérieur des groupements d'objets [Sil95].

3.5.2.4 Radiosité étendue

Un des principaux défauts de la méthode de la radiosité est sa limitation aux surfaces parfaitement diffuses. La prise en compte de BRDF plus complètes est une lourde tâche, car elle remet en cause la factorisation qui simplifie l'équation de rendu. Une solution très complexe réside dans la discrétisation à la fois des surfaces de la scène et de l'espace des directions de réflexion [AH93]. La méthode zonale, avec l'introduction de facteurs de forme volumiques, permet la prise en compte de phénomènes plus complexes, tels que les milieux participants (section 2.7). Elle peut être adaptée à la prise en compte des matériaux spéculaires [AMP98].

3.6 Calcul par une méthode de Monte-Carlo

3.6.1 Lancer de rayons

Le lancer de rayons a été présenté pour la première fois par Appel [App68], en 1968. Il s'agissait à l'époque d'une méthode d'élimination des parties cachées, indiquant simplement les objets visibles depuis la position de l'observateur. Cet algorithme a été rendu célèbre par Whitted [Whi80], en 1980, en tant que méthode de rendu. Pour la première fois des images de synthèse comprenaient des reflets spéculaires et des ombres portées. De multiples améliorations ont été apportées depuis, aussi bien pour la qualité des résultats que pour la vitesse d'exécution (qui reste tout de même très lente).

Le lancer de rayons, même dans sa version élémentaire, peut être considéré comme une résolution de l'équation intégrale de rendu par une méthode de Monte Carlo (section 3.6.6).

3.6.1.1 Principe général

Le *lancer de rayons* reproduit les trajectoires des rayons lumineux, en respectant au mieux les lois de l'optique géométrique (figure 3.4). Les phénomènes de réflexion et de réfraction sont obtenus en suivant les rayons jusqu'aux sources de lumière, en calculant les déviations qu'ils subissent en rencontrant les surfaces des objets.

Les rayons sont orientés de l'œil vers la scène, en passant par les pixels de l'écran. Cette direction inverse de celle de la lumière permet de ne lancer que les rayons qui participent effectivement à l'image. Cette inversion est possible, car les lois de Descartes sont symétriques (section 2.3.1.3). Le gain est considérable, car seule une très faible fraction des rayons émis par les sources atteignent l'œil.

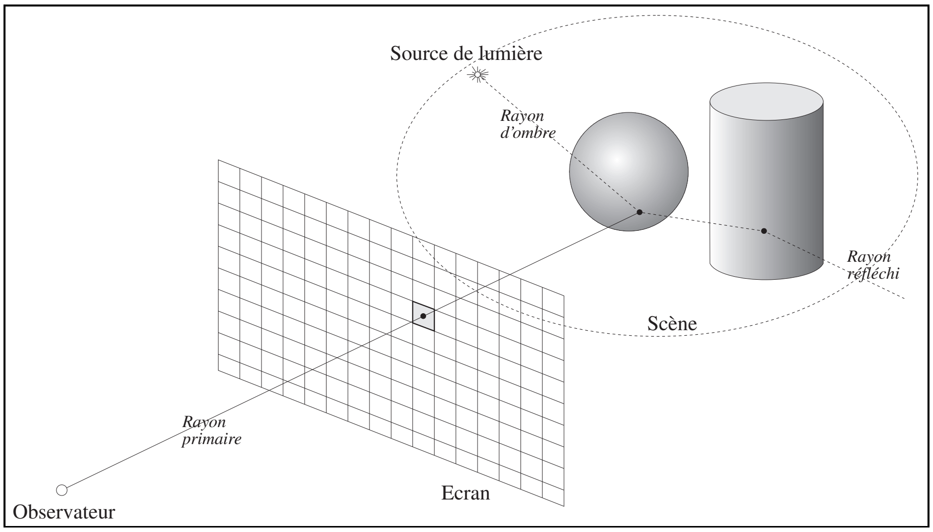


FIG. 3.4 – *Schéma de principe du lancer de rayons*

L'algorithme générique du lancer de rayons est très simple :

- **Pour** chaque pixel de l'écran **Faire**
  - Générer un rayon orienté de l'œil vers la scène passant par le pixel ;
  - Calculer l'intersection de ce rayon avec la scène ;
  - **Si** un objet au moins est rencontré **Alors**

    Déterminer le point d'intersection le plus proche ;

    Calculer son éclairement ;

    FinSi

  Fait

L'éclairement du point d'intersection, qui est calculé par la fonction de rendu, est la somme de trois énergies distinctes :

- **Energie directe** : obtenue en lançant des *rayons d'ombre*[2] vers les sources de lumière (voir le chapitre 5). Si un rayon d'ombre ne rencontre aucun objet, alors la source est visible depuis le point, et il faut ajouter sa contribution.
- **Energie réfléchie** : obtenue en générant la direction réfléchie par les lois de l'optique géométrique, puis en calculant l'énergie incidente selon cette direction, encore une fois en lançant un rayon. Cette fonction est récursive.
- **Energie réfractée** (selon la matière de l'objet) : comme pour l'énergie réfléchie, mais en générant cette fois la direction réfractée vers l'intérieur de l'objet.

2. Les *rayons d'ombre* permettent de savoir si un objet est dans l'ombre d'un autre pour une source donnée.

Dans cette version classique du lancer de rayons, avec uniquement des sources ponctuelles, pour chaque pixel, un rayon primaire est généré, qui donne naissance à des rayons d'ombre (un par source), et éventuellement à un rayon réfléchi et un rayon réfracté, qui ont eux aussi le même comportement. Les rayons sont donc organisé sous une forme arborescente : l'*arbre des rayons*. Cette structure est mise a profit par des algorithmes qui améliorent progessivement la qualité de l'image, en revenant en arrière dans l'arbre pour affiner certaines régions.

3.6.1.2 Complexité du lancer de rayons classique

Soit une scène composée de N objets et de N_s sources, et dont la profondeur moyenne des arbres des rayons est p. Le calcul d'un image de N_p pixels nécessite la génération de N_p rayons primaires, dont il faut calculer l'intersection avec la scène, en testant les N objets. Cette opération est répétée pour leurs p fils, la visibilité des sources devant être testée à chaque fois. Le nombre de calculs d'intersections rayon-objet est pN_pN_sN. La complexité théorique du lancer de rayons classique en fonction de la taille de la scène est donc optimale, en $O(N)$, sans même tenir compte de l'optimisation très efficace des boîtes englobantes (section 3.8.2).

Ce bon résultat théorique ne signifie hélas pas que la méthode est rapide. Les temps de calcul sont en pratique très longs, car le nombre d'opérations est considérable. Par exemple, pour une scène de $N = 1000$ objets (ce qui est peu aujourd'hui), une image de taille moyenne 512x512 et $p = 5$, le nombre total de calculs d'intersections est de l'ordre de $1,3$ milliard.

Des études plus approfondies d'un modèle théorique de lancer de rayons sont proposées dans [Ris96] et dans [RKJ96].

3.6.1.3 Terme ambiant

Le lancer de rayons, dans sa version initiale, ne permet pas la prise en compte des éclairements indirects. Seuls les chemins lumineux de type LD^-S^*E sont calculés. Chaque rayon incident sur une surface ne génère qu'un seul rayon réfléchi (et/ou réfracté), alors que dans le cas de réflexion diffuse chaque rayon incident génère une infinité de rayons réfléchis dans toutes les directions.

(a) Terme ambiant nul. *(b) Valeur moyenne.* *(c) Valeur trop élevée*

FIG. 3.5 – *Influence du terme ambiant*

De ce fait, les zones qui ne sont pas éclairées paraissent parfaitement noires, en contradiction avec toute attente. Pour compenser ce phénomène, un *terme ambiant* est ajouté à l'énergie de tous les points, qui simule artificiellement une pseudo-pénombre. L'amélioration visuelle est importante, mais ce procédé n'est pas justifiable physiquement, car le respect du principe de conservation de l'énergie n'est pas assuré. Le choix de la valeur du terme ambiant est totalement empirique, basé sur une estimation de l'éclairement indirect moyen (figure 3.5). De plus, sa valeur est constante pour toute la scène, que l'on soit à proximité d'un objet clair fortement éclairé ou bien à l'intérieur d'une armoire fermée.

Des solutions beaucoup plus élaborées, respectant les lois fondamentales de la physique, ont été apportées depuis à ce problème (voir les sections 3.6.3 et 4.2.3).

3.6.2 Lancer de rayons distribué

La méthode du lancer de rayons classique induit par nature un fort aliassage, à tous les niveaux de l'arbre des rayons. En effet, les rayons étant infiniment petits, ils réalisent un échantillonnage ponctuel de la répartition surfacique de l'énergie.

Le *lancer de rayons distribué* [CPC84] remédie à ce problème, par un sur-échantillonnage systématique de l'espace des rayons. D'un point de vue méthode de Monte Carlo, l'estimateur primaire (équation A.1), non biaisé mais fortement bruité, est remplacé par un estimateur secondaire (équation A.2) de plus faible variance. Chaque rayon devient un cône de même direction, qui est échantillonné par un nombre variable de rayons.

Suivant le type de rayons auquel il est appliqué, ce procédé génère des résultats différents :

- rayons primaires : diminution de l'aliassage de l'image ;
- rayons d'ombre : effets de pénombres douces ;
- rayons réfléchis : effet de polissage des matériaux, qui ne sont plus des mirroirs parfaits ;
- rayons réfractés : simulation de matériaux translucides.

De plus, en modifiant la position de l'œil ou des objets, entre deux rayons consécutifs, il est possible de simuler des flous de mouvements, comme sur les photographies d'objets mobiles.

Malheureusement, le lancer de rayons distribué est une méthode très coûteuse. Le nombre de rayons est fortement multiplié, même si des techniques d'échantillonnage adaptatif atténuent ce problème.

3.6.3 Méthode de Ward

Le logiciel de lancer de rayons *RADIANCE* [War94], développé principalement par Ward, est célèbre dans le monde de la synthèse d'images. Il intègre une méthode de calcul des éclairements indirects par une méthode de Monte Carlo, qui est à l'origine de nos propres développements (section 4.2.3).

3.6.3.1 Modèle de la sphère coupée

Le modèle de la sphère coupée *(split sphere model)*, présenté par Ward dans [WRC88] est utilisé pour déterminer une borne supérieure de la variation de l'éclairement indirect sur une surface, en se limitant à la composante purement diffuse.

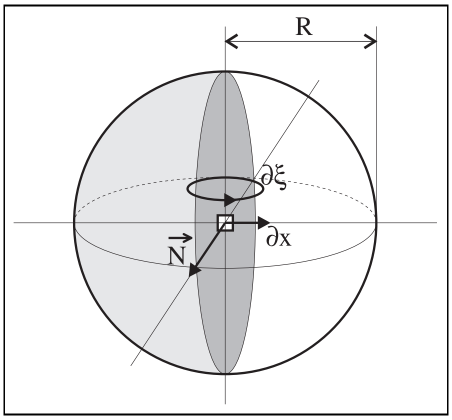


FIG. 3.6 – *Modèle de la sphère coupée*

Ce modèle consiste en une sphère divisée en deux parties : l'une est complètement sombre, alors que l'autre est complètement éclairée. Cette construction représente le plus grand contraste que l'on puisse attendre pour une surface infiniment petite placée au centre de la sphère (figure 3.6). L'orientation est telle que des petites variations géométriques induisent les plus fortes variations de l'éclairement.

Un majorant approximatif de la variation ε de l'éclairement E au centre de la sphère, pour un déplacement élémentaire de x_0 en x, et un changement d'orientation infinitésimal de ξ_0 en ξ, est obtenu par un développement de Taylor du premier ordre :

$$\varepsilon \leq \left| \frac{\partial E}{\partial x}(x - x_0) + \frac{\partial E}{\partial \xi}(\xi - \xi_0) \right|$$

La variation d'éclairement, pour une surface infiniment petite, est proportionnelle à la variation de son aire projetée. La variation avec la position est $\dfrac{2R\partial x}{\frac{1}{2}\pi R^2} = \dfrac{4\partial x}{\pi R}$, et la variation avec l'orientation est $\dfrac{\frac{1}{2}\pi R^2 \partial\xi}{\frac{1}{2}\pi R^2} = \partial\xi$. Donc la variation de l'éclairement à l'intérieur de la sphère coupée est :

$$\varepsilon \leq \frac{4\, E_0}{\pi\, R}\, |x - x_0| + E_0\, |\xi - \xi_0|$$

Cette formulation est étendue au cas général, pour exprimer un majorant de la variation de l'éclairement, pour un déplacement du point P_0, de normale $\vec{n_0}$ et d'éclairement E_0, vers un point P, de normale $\vec{n}$:

$$\varepsilon(P) \leq E_0 \left[\frac{4}{\pi} \frac{\left\| \overrightarrow{P - P_0} \right\|}{R_0} + \sqrt{2 - 2\left(\vec{n} \cdot \vec{n_0}\right)} \right]$$

Le terme R_0 est la distance moyenne harmonique (qui privilégie les faibles valeurs) entre P_0 et les autres objets de la scène, et $\sqrt{2 - 2\cos\alpha} \approx \alpha$, lorsque l'angle α est petit.

Le calcul de ce majorant comporte un grand nombre d'approximations, parfois difficiles à justifier. Plutôt que de l'utiliser comme une valeur fiable, Ward ne conserve que l'allure de la fonction, sans les coefficients qui n'ont de sens que pour la sphère coupée théorique. Il l'utilise comme une fonction de pondération inversement proportionnelle à l'erreur probable, pour l'interpolation des éclairements indirects décrite plus loin :

$$w(P) = \frac{1}{\frac{\left\|\overrightarrow{P - P_0}\right\|}{R_0} + \sqrt{1 - \left(\vec{n} \cdot \vec{n_0}\right)}} \tag{3.8}$$

3.6.3.2 Eclairement indirect diffus

3.6.3.2.1 Principe de la méthode de Ward L'éclairement indirect diffus en un point P d'une surface, tel qu'il est calculé par Ward [WRC88], regroupe toute l'énergie incidente qui a été réfléchie au moins une fois sur une surface diffuse, par des chemins lumineux de type LD^+P. Le calcul de l'éclairement indirect diffus en un point est une opération longue. Ward propose des optimisations basées sur l'hypothèse suivante:

> "*La composante diffuse de l'éclairement varie lentement à l'intérieur d'une scène.*"

Cette hypothèse intuitive est couramment admise en synthèse d'images, pour les scènes courantes comportant des matériaux majoritairement diffus. Le calcul complet de l'éclairement diffus peut alors n'être effectué qu'en un nombre restreint de points. Sa valeur en d'autres points "proches" est déduite par interpolation des valeurs connues. L'introduction d'une notion de gradient permet d'affiner la reconstruction.

Empiriquement, cette hypothèse peut se justifier ainsi. L'éclairement indirect est la somme de réflexions sur des surfaces dont la taille est non nulle. Comme des sources surfaciques, elles produisent un éclairement dont les variations sont douces, avec des zones de pénombre plus ou moins étendues. L'éclairement indirect est alors une somme de fonctions continues, et il est donc lui aussi représenté par une fonction continue.

L'éclairement (section 2.2.3) est calculé en cumulant les énergies incidentes, pondérées par l'angle solide projeté. La grandeur finale ne contient plus aucune information directionnelle, et elle ne peut être utilisée qu'avec la composante purement diffuse de la BRDF, pour déterminer la luminance réfléchie par la surface vers l'observateur. De ce fait, nous employons l'expression éclairement indirect diffus, qui précise la prise en compte seulement partielle de l'éclairement indirect.

3.6.3.2.2 Méthode de Monte Carlo L'éclairement (section 2.2.3) en un point P, noté $E(P)$, provient de toutes les directions visibles depuis P. Comme la formulation intégrale n'a pas de solution analytique générale, une méthode de Monte Carlo est utilisée pour calculer l'éclairement. Il s'agit ici uniquement de réflexions, donc l'ensemble des directions incidentes est un hémisphère centré en P.

Ward préconise une discrétisation uniforme de l'hémisphère, à angle solide projeté constant (voir la figure 3.7 pour les notations). Un rayon est lancé dans chaque cellule, à une position aléatoire. Si M est le nombre de découpages zénithaux, et N le nombre de

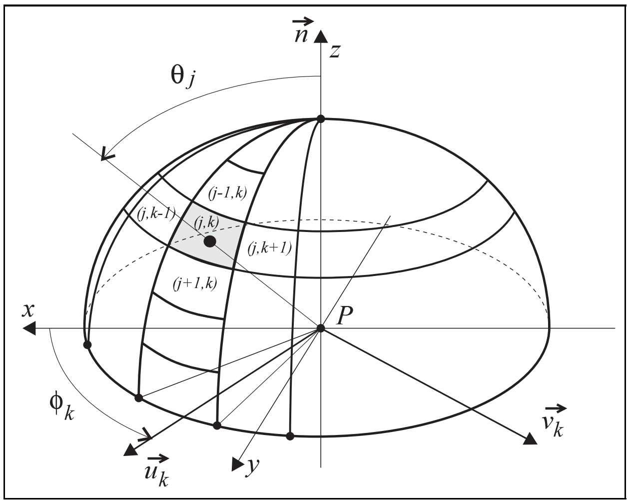


FIG. 3.7 – *Discrétisation de l'hémisphère*

découpages zénithaux, alors la direction du rayon tiré dans la cellule (j, k) est donnée par :

$$\begin{cases} \theta_j = \sin^{-1}\left(\sqrt{\dfrac{j - X_j}{M}}\right) \\ \phi_k = \pi \, \dfrac{k - Y_k}{N} \end{cases}$$

où X_j et Y_k sont des variables aléatoires uniformes comprises entre 0 et 1.

L'angle solide projeté élémentaire est le terme $\cos \theta d\omega$ de l'équation de l'éclairement (équation 2.2). L'angle solide projeté de l'hémisphère complet a pour valeur π, donc chaque cellule sous-tend un angle solide projeté de valeur $\frac{\pi}{MN}$, qui est considéré comme élémentaire lorsque le découpage est fin. L'équation discrétisée de l'éclairement est alors :

$$E(P) = \frac{\pi}{MN} \sum_{j=1}^{M} \sum_{k=1}^{N} L(P, \theta_j, \phi_k)$$

Le calcul par la méthode de Monte Carlo consiste donc à générer MN rayons, pour calculer les luminances incidentes. Comme cette procédure est récursive, un seuil arbitraire limite la profondeur maximale, car le nombre de rayons augmente alors exponentiellement. Ce calcul est une opération très longue, le nombre de rayons devant être important pour limiter le bruit (typiquement plusieurs centaines par hémisphère). La méthode de Ward permet de l'effectuer beaucoup moins souvent que lors d'une évaluation systématique en chaque pixel.

Pour limiter la profondeur de récursion, infinie en théorie, le processus est stoppé lorsqu'un seuil prédéfini est atteint. Cette méthode de Monte Carlo est biaisée, car cette troncature ne garantit pas la convergence vers la solution exacte. Cependant, l'absorption des matériaux lors de la réflexion fait que la contribution diminue avec la profondeur de récursion. Ainsi, même si une surface fortement éclairée est oubliée, seule une faible partie de son énergie manque dans le résultat global. Des méthodes de Monte Carlo plus sophistiquées et non biaisées, comme le suivi de chemins (section A.4.3), permettraient de résoudre plus élégamment ce problème.

3.6.3.2.3 Interpolation pondérée Lorsque des valeurs de l'éclairement indirect diffus $\{E(P_1), E(P_2), \ldots, E(P_n)\}$ sont déjà connues dans un voisinage du point P, plutôt que d'utiliser la coûteuse méthode de Monte Carlo, l'éclairement indirect diffus en P est estimé par une interpolation pondérée. L'hypothèse de variation lente de l'éclairement indirect autorise cette opération.

Pour qu'une valeur calculée en P_i soit acceptée dans l'interpolation, le poids $w(P_i)$ (équation 3.8) doit être supérieur à un seuil arbitraire, afin de limiter l'erreur de reconstruction. La fonction d'interpolation pondérée des éléments acceptés est classique :

$$E(P) \approx \frac{\sum_{i=1}^{n} w(P_i)\, E(P_i)}{\sum_{i=1}^{n} w(P_i)}$$

La luminance réfléchie en P par l'éclairement indirect diffus est alors calculée avec :

$$L(P, \overrightarrow{\omega_r}) = \frac{\rho_d(P)}{\pi} E(P)$$

où $\rho_d(P)$ est le coefficient de réflexion purement diffuse de la surface au point P.

3.6.3.3 Gradient d'éclairement diffus

Dans [WH92], Ward propose une amélioration de sa méthode de calcul des éclairements indirects diffus. L'estimation d'un gradient permet d'évaluer plus précisément les variations de l'éclairement, et donc d'affiner le résultat de l'interpolation pondérée. Le gradient est décomposé en deux parties : un gradient de rotation $\overrightarrow{\nabla}_r E$, pour des petites variations de la normale, et un gradient de translation $\overrightarrow{\nabla}_t E$, pour des petits déplacements du point de calcul. Ils sont tous les deux calculés en considérant les variations de l'influence des cellule de l'hémisphère sur la valeur de l'éclairement, pour ces petits mouvements. Nous ne détaillons pas plus ces calculs très lourds ici.

Le gradient est utilisé lors de l'interpolation pondérée, qui devient :

$$E(P) \approx \frac{\sum_{i=1}^{n} w(P_i) \left[E(P_i) + (\overrightarrow{n_i} \wedge \overrightarrow{n}) \cdot \overrightarrow{\nabla}_r E_i + (\overrightarrow{P - P_i}) \cdot \overrightarrow{\nabla}_t E_i \right]}{\sum_{i=1}^{n} w(P_i)}$$

L'amélioration induite par le calcul du gradient permet de réduire le nombre total de points de calcul complets, par la méthode de Monte Carlo. Le gain de temps compense largement le surcoût du calcul du gradient, qui est relativement modéré car les informations nécessaires sont disponibles lors du parcours de l'hémisphère initial.

Toujours dans le même article, la méthode est testée sur un exemple comprenant une sphère, dont une face ne reçoit que l'éclairement indirect. L'amélioration est sensible, avec une forte réduction du zonage causé par l'interpolation. Cependant Ward ne donne pratiquement aucune indication sur le taux d'échantillonnage des hémisphères et sur le nombre de points de calcul effectifs.

3.6.4 Lancer de rayons inverse

3.6.4.1 Les caustiques

Les *caustiques* sont la conséquence de concentrations de lumière créées par des objets spéculaires ou réfractants sur la composante diffuse d'autres surfaces. C'est un phénomène très fréquent. Par exemples, les taches lumineuses sous les objets en verre, une image projetée sur un écran, les reflets fluctuants des vagues sur le fond de sable, ..., sont des caustiques. Elles apparaissent rarement dans les images de synthèse, car leur simulation est très difficile et coûteuse.

Les chemins lumineux comportent dans ce cas au moins une fois le terme S^*D. Partant de l'œil, le lancer de rayons classique n'est pas capable de simuler de tels chemins. En effet, il n'est pas possible de déterminer par avance l'ensemble des chemins aboutissant à une source après avoir subi plusieurs déviations. La recherche de *chemins de Fermat* [MH92] est une solution théorique, mais elle n'est utilisable en pratique que pour des configurations géométriques extrèmement simples. La radiosité classique n'est pas capable elle non plus de simuler ces phénomènes, car elle ne manipule que les surfaces diffuses.

3.6.4.2 Prise en compte des caustiques

Une solution à ce problème est le *lancer de rayons inverse* [Arv86], présenté par Arvo en 1986, à la suite des travaux beaucoup plus anciens de Appel en 1968 [App68] (les moyens techniques de l'époque ne permettaient pas de pousser très loin les développements). Cette méthode est maintenant déclinée sous de multiples formes par de nombreux auteurs, car les évolutions de l'informatique permettent des simulations de plus en plus sophistiquées.

Le lancer de rayons inverse consiste à émettre des rayons depuis les sources, recréant ainsi le processus naturel de cheminement de la lumière. Lors d'une première passe, l'énergie émise par les sources est stockée sur les surfaces diffuses, à la suite de chemins de type LS^*D. Cette énergie est recueillie lors de la passe suivante, un lancer de rayons classique depuis l'œil, correspondant à des chemins de type DS^*E. Le regroupement de ces deux passes donne donc les chemins de type LS^*DS^*E.

Arvo propose de stocker les informations de la première passe dans une structure fixe, l'*Illumination Map*, qui est un quadrillage régulier des surfaces diffuses plaqué comme une texture. Cette structure très lourde nécessite une paramétrisation des surfaces difficilement compatible avec la représentation des scènes par *Arbre CSG*. De plus, les caractéristiques complètes de la BRDF ne sont pas prises en compte, et les caustiques n'apparaissent que sur les surfaces parfaitement lambertiennes. La *Photon Map*, introduite par Jensen [Jen96], résoud ces problèmes en utilisant une structure adaptative de stockage (un kd-arbre) indépendante de la géométrie de la scène.

Nous proposons plus loin une méthode de prise en compte des caustiques (section 4.2.2), basée sur une modélisation vectorielle de la lumière. Notre méthode de reconstruction utilise une estimation de densité avec noyau, du même type que celle proposée par Collins [Col94], mais nous tenons compte de la BRDF complète, tout comme Jensen [Jen97]. De plus, notre méthode est utilisable avec des sources étendues.

3.6.5 Suivi de chemins

3.6.5.1 Principe

Le suivi de chemins est l'évaluation de l'équation intégrale de rendu (équation 3.1) par une méthode de Monte Carlo (section A.3), si possible d'importance pour réduire la variance, et si possible non biaisée pour garantir la convergence vers une solution correcte. Kajiya [Kaj86] fut le premier à utiliser cette approche, pour résoudre la première formulation globale des éclairements, qu'il venait lui-même d'introduire dans le même article, qui fit une grande impression en informatique graphique et constitue toujours un référence incontournable.

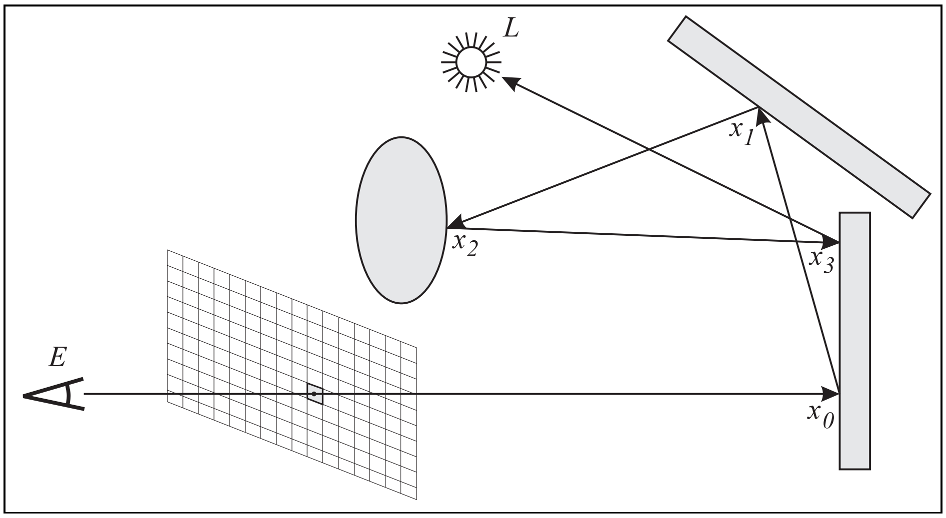


FIG. 3.8 – *Suivi de chemins*

Le principe est de tracer une marche aléatoire (section A.4.3) depuis l'œil, en passant par un pixel donné, pour résoudre l'équation de rendu. Une suite de points $x_0, x_1, x_2, \cdots$ de la scène est générée aléatoirement (figure 3.8). A chaque étape x_{i-1}, un point x_i est choisi selon une fonction de densité de probabilité p_i, qui dépend de la direction d'incidence et de la BRDF du matériau rencontré. Il s'agit de simuler le comportement le plus probable d'un photon de lumière, en suivant un chemin inverse (pour être sûr d'atteindre un pixel).

Le chemin lumineux généré ainsi est en théorie infini. Pour que les calculs se terminent, il est donc nécessaire de limiter sa longueur. Une troncature arbitraire au-delà d'un certain seuil est un procédé simple, mais qui introduit un biais dans la solution. La méthode de la roulette russe (section A.4.2) est une réponse non biaisée à ce problème

Lorsqu'une source lumineuse est atteinte, sa contribution est ajoutée au flux énergétique suivant ce chemin jusqu'à l'œil. Ce procédé suppose alors l'utilisation de sources étendues, car la probabilité d'atteindre ainsi une source ponctuelle est parfaitement nulle.

Chaque chemin généré ainsi est un estimateur primaire de l'équation de rendu pour un pixel donné. Le résultat est fortement bruité. Pour réduire la variance, comme toujours avec les méthodes de Monte Carlo, le procédé est réitéré plusieurs fois, et l'on affecte au pixel la moyenne des résultats obtenus, qui est non biaisée si les résultats individuels le sont.

Si le matériau est plutôt spéculaire, la probabilité de choisir une direction proche de la direction de réflexion spéculaire est forte, alors que si le matériau est plutôt diffus aucune direction n'est vraiment privilégiée. Tous les chemins lumineux $L(D|S)^*E$ sont en théorie accessibles, mais la convergence peut être très lente, surtout si les sources sont petites et la proportion de surfaces spéculaires importante.

3.6.5.2 Echantillonnage plus sophistiqué

La connaissance de scène, et des caractéristiques des sources, est mise à profit par Maillot [Mai96] pour accélérer la convergence, sans introduire de biais. Ces informations sont prises en compte lors du tirage des directions réfléchies, afin d'augmenter la probabilité d'atteindre les sources. Les directions dans lesquelles sont situées des sources deviennent des directions privilégiées par la fonction aléatoire, au même titre que la direction spéculaire l'est pour les matériaux brillants. Des conditions locales (la BRDF), globales (les caractéristiques des sources), et mixtes (l'énergie potentiellement apportées par les sources en un point) sont prises en compte lors du choix d'une direction réfléchie, afin de déterminer s'il est péférable de sélectionner l'éclairement direct, et donc une source, ou bien l'éclairement indirect, et donc une direction un peu plus quelconque, mais tout de même influencée par la BRDF.

D'autre méthodes d'échantillonnages d'importance optimisées ont été proposées dans la littérature pour réduire encore la variance, par exemple [Shi90a, LW95, VG95, VG97]. Ce domaine de recherche est toujours très actif.

3.6.5.3 Suivi de photons

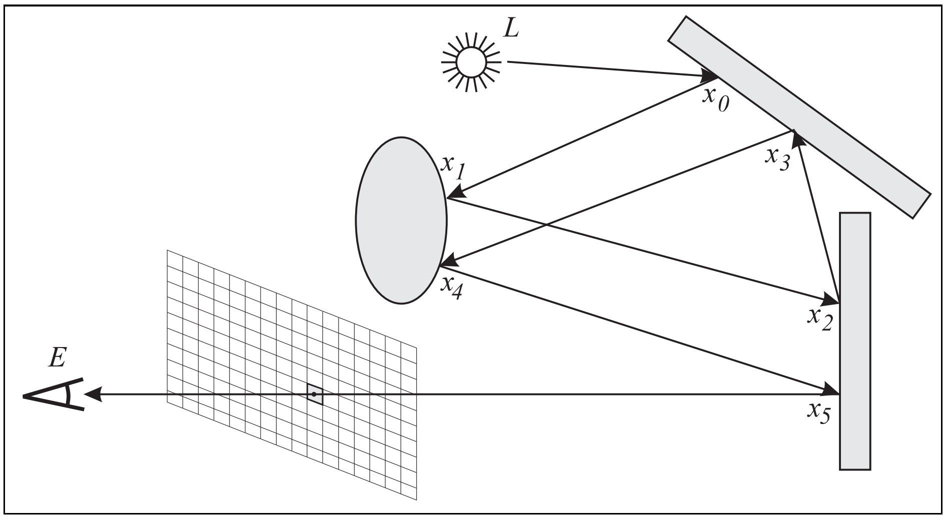


FIG. 3.9 – *Suivi de photons*

La méthode du *suivi de photons* est la résolution par une méthode de Monte Carlo du problème dual (section 3.3.2). Cette fois, les chemins aléatoires sont lancés depuis les sources (figure 3.9). Lorsqu'un pixel de l'image est atteint, la contribution énergétique du chemin lui est ajouté. Comme pour le suivi de chemins, diverses techniques de réduction de variance ont été développées [DW94, DW95, Jen95].

3.6.5.4 Suivi de chemins bidirectionnel

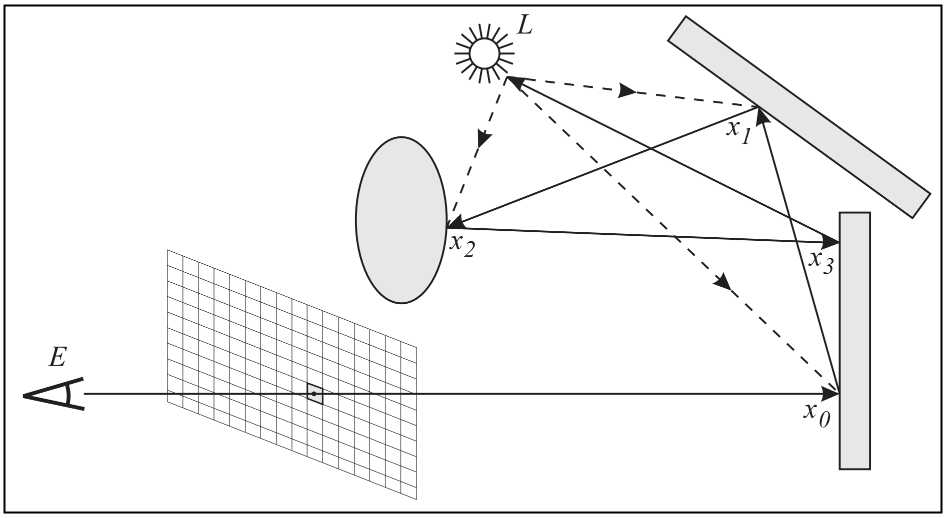


FIG. 3.10 – *Suivi de chemins bidirectionnel*

Le *suivi de chemins bidirectionnel* [LW94, Vea97, JC98] est la combinaison des deux méthodes précédentes. Des chemins aléatoires sont lancés à la fois depuis les sources et depuis l'observateur, avec l'objectif de les faire se croiser. Les rencontres peuvent être forcées en lançant systématiquement des chemins directs depuis les sources, vers les points de bifurcation du chemin tracé depuis l'œil (figure 3.10).

3.6.6 Analogies avec le lancer de rayons classique

Le lancer de rayons classique, tel qu'il est décrit initialement par Whitted [Whi80], peut s'interpréter selon la méthode de Monte Carlo du suivi de chemins bidirectionnel, bien qu'il n'ait pas été écrit dans cette optique. Il comporte cependant des simplifications drastiques qui introduisent un biais important.

Les chemins qui rencontrent une surface diffuse sont systématiquement tronqués, et les réfexions sur les matériaux brillants n'ont lieu que dans la direction spéculaire parfaite de Descartes. Pour compenser les troncatures brutales et les directions de réflexion absentes, un terme ambiant est ajouté à tous les calculs. Ce procédé est comparable à la méthode de réduction de variance du contrôle de variations (section A.3.2.3), mais la fonction de contrôle sensée approcher la fonction calculée est constante, et sa valeur est complètement arbitraire.

De la même façon, le lancer de rayons inverse présenté par Arvo [Arv86] peut être assimilé à un cas particulier de la méthode de Monte Carlo du suivi de photons, là encore avec un biais important.

3.7 Approches mixtes

Constatant la relative complémentarité des qualités et des lacunes des méthodes du lancer de rayons et de la radiosité, dans leurs versions classiques, Wallace et al. ont proposé une méthode mixte [WCG87], dite *en deux passes*, pour les réconcilier. Cette approche a été souvent reprise et améliorée, en particulier par Sillion et Puech [SP89], Shirley [Shi90b], Chen et al. [CMRT91], et par Kok et Jansen [KJ92]. Plus récemment, cette technique a été reprise par Zimmerman et Shirley [ZS95], pour des scènes comprenant de nombreuses sources.

- La première passe, de type radiosité, calcule la répartition des éclairements diffus dans la scène, avec les simplifications de la méthode.
- La deuxième passe, de type lancer de rayons, ajoute les reflets spéculaires. Les valeurs de radiosité de la passe précédente remplacent le terme ambiant, avec une erreur moindre.

Bien que certains calculs soient communs aux deux passes, la durée d'exécution globale est très longue. Il est cependant à noter que, dans le cas où la position de l'observateur est modifiée, seule la deuxième passe doit être reprise. C'est intéressant pour la réalisation d'animations, car cette passe de lancer de rayons est la moins lente.

3.8 Optimisations habituelles

3.8.1 Scènes comportant de nombreuses sources

Des optimisations spécifiques ont été développées pour les scènes comportant de nombreuses sources.

La méthode de Kok et Jansen [KJ91] permet de réduire le nombre de rayons d'ombre lancés vers les sources de lumière. Elle limite le nombre de sources à considérer pour l'éclairement direct, l'énergie fournie par les sources les moins significatives étant ajoutée à l'éclairement indirect, qui a été calculé lors d'une première passe de radiosité. Pour chaque facette, les n plus importantes sources (facettes) sont sélectionnées selon la radiosité qu'elles émettent. La composante directe est alors évaluée en lançant des rayons vers des points disposés aléatoirement sur la surface des sources. L'image est calculée à une basse résolution, puis les zones perturbées sont raffinées, selon des critères d'homogénéité géométrique et de variance des composantes directes et indirectes de l'éclairement.

L'esprit de la méthode proposée par Ward [War91] est proche. Pour chaque point, les sources sont triées selon l'énergie qu'elles sont succeptibles de fournir, puis la visibilité des sources les plus significatives est prises en compte en premier. Lorsque le taux de l'énergie prise en compte atteint un seuil donné, l'énergie restante est simplement ajoutée au calcul en cours, en la pondérant par une estimation statistique de visibilité.

3.8.2 Boîtes englobantes

Le système des *boîtes englobantes* est utilisé pratiquement depuis les débuts de l'essor de l'informatique graphique [Whi80]. L'idée est d'insérer tous les objets de la scène dans des boîtes virtuelles de forme simple, en général des sphères ou des parallélépipèdes rectangles, orientés ou non selon des axes privilégiés [Roe93]. Lorsqu'une condition géométrique doit être vérifiée sur un objet, le test est d'abord effectué sur la boîte englobante de l'objet. S'il est négatif, il est inutile de tester plus précisément l'objet lui même.

Cette méthode est systématiquement utilisée en lancer de rayons, où l'opération la plus coûteuse, et aussi la plus fréquente, est le calcul de l'intersection d'un rayon avec la scène, pour déterminer les objets visibles. Il est clair que lorsque le rayon ne rencontre pas la boîte englobante, il ne peut pas rencontrer l'objet qu'elle contient. De nombreuses opérations sont ainsi évitées.

En radiosité, le regroupement d'objets *(clustering)* [Sil95] obéit à la même démarche. Lorsqu'un groupe d'objets est assez loin, il est inutile de détailler la contribution de chacun de ses composants.

Pour que la méthode soit efficace, les boîtes englobantes doivent être aussi proches que possible des objets, tout en étant de forme simple pour être testées rapidement. Une optimisation complémentaire est apportée par une structure hiérarchique des boîtes englobantes. Certaines boîtes englobent alors d'autres boîtes, regroupant des objets géométriquement proches. En effet, il est inutile de tester la visibilité d'un barreau de chaise, si la boîte englobante de la chaise complète n'est pas visible.

3.8.3 Sous-échantillonnage

La technique du sous-échantillonnage cherche à diminuer le nombre de rayons, en utilisant autant que possible les cohérences de l'image, des objets, et de la lumière. Lorsqu'un point de la scène apparaît d'une couleur donnée, la probabilité qu'un point "voisin" soit d'une couleur identique (ou très similaire) n'est pas négligeable. Par exemple, la méthode de Ward (section 3.6.3) utilise la cohérence de l'éclairement indirect, qui varie lentement, pour ne pas le calculer partout.

C'est aussi une méthode utilisée pour obtenir rapidement une prévisualisation des images [JvW83]. L'écran est divisé en carreaux de taille prédéfinie. Un seul pixel est calculé pour chaque carreau, et il donne sa couleur au carreau complet. L'image devient de plus en plus précise, au fur et à mesure que la taille des carreaux diminue.

Le lancer de rayons progressif [MCP92] exploite la constatation précédente, dans l'espace de l'image. Celle-ci est décomposée en zones, à l'intérieur desquelles une proportion réduite de pixels sont calculés. Si les pixels sont de couleurs proches, selon des critères élaborés, la zone est considérée comme uniforme, et les couleurs des pixels manquants sont calculées par interpolation. Sinon, la zone est hétérogène, et elle est subdivisée en zones plus petites pour lesquelles le processus est réitéré récursivement. La qualité de l'image s'améliore incrémentalement, ce qui permet d'obtenir rapidement une première ébauche acceptable.

Chapitre 4

Les Vecteurs Lumineux

Ce chapitre présente la notion de *vecteur lumineux*, qui est au cœur de notre méthode de calcul de l'éclairement global. Cette notion homogénéise la représentation des diverses énergies présentes en un point de la scène, et permet d'optimiser le calcul des images. Elle est tout d'abord introduite dans un cadre général, puis des versions adaptées à des cas particuliers de chemins lumineux (section 3.2) sont proposées.

4.1 Cas général

4.1.1 Objectifs

Le calcul de l'éclairement global est une opération complexe et très coûteuse, quelle que soit la méthode de calcul d'image choisie. L'équation de rendu (équation 3.1), plus ou moins modifiée, doit être évaluée pour tous les points de la scène. Cependant, il apparaît bien souvent en pratique que l'éclairement incident varie très peu entre des points proches, en dehors des frontières entre objets et des zones de transition ombre-lumière. Des calculs complexes sont alors effectués à de nombreuses reprises, pour finalement obtenir des résultats presque semblables.

L'observation précédente est utilisée pour développer des méthodes de réduction des temps de calcul. En effet, la détection des zones où l'éclairement est stable permet de réutiliser à moindre frais les valeurs déjà calculées, lorsque des conditions de validité sont réunies. Nous proposons une modélisation de la lumière incidente en un point, le *vecteur lumineux*, qui nous permet d'appliquer des algorithmes rapides de reconstruction des éclairements inconnus, par interpolation ou estimation de densité à partir de valeurs connues.

Ces travaux ont été publiés dans [ZP97] et [ZSP98]. Ils sont repris avec plus de détails ici.

4.1.2 Le vecteur d'éclairement

Le *vecteur d'éclairement* $\overrightarrow{\Phi(x)}$ représente la quantité et la direction de l'énergie lumineuse totale présente en un point quelconque x de l'espace. Il est obtenu en intégrant, sur la sphère unitaire S^2, centrée en x, l'ensemble des directions incidentes, pondérées par les luminances incidentes. La mesure canonique sur la sphère unitaire correspond en pratique à la valeur de l'angle solide associé à la direction incidente $\overrightarrow{\omega_i}$. Elle est notée $\sigma(\overrightarrow{\omega_i})$. Dans le cas où le point x appartient à la surface d'un matériau opaque, le domaine d'intégration est limité à l'hémisphère orienté vers l'extérieur. Le vecteur d'éclairement vaut :

$$\overrightarrow{\Phi(x)} = \int_{S^2} L_i(x, \vec{\omega_i}) \; \vec{\omega_i} \; d\sigma(\vec{\omega_i}) \tag{4.1}$$

Lorsqu'il est défini sur une surface, il permet de calculer l'éclairement $E(x)$ (section 2.2.3) de n'importe quel point x de la surface de normale $\vec{n}$:

$$E(x) = -\overrightarrow{\Phi(x)} \cdot \vec{n}$$

L'ensemble des vecteurs d'éclairement forme un champ vectoriel auquel s'appliquent de nombreuses propriétés mathématiques. Arvo [Arv94] l'utilise pour modéliser l'émission de sources surfaciques planes uniformes partiellement obstruées. Dans ce cas, le jacobien du champ vectoriel est calculable. Il permet une simplification des calculs d'éclairement, en quantifiant ses variations sur la surface. Dans [Arv95], Arvo étend les applications de cette méthode à des sources directionnelles et des BRDF non lambertiennes, en exploitant la notion de tenseur d'éclairement.

4.1.3 Le vecteur lumineux

4.1.3.1 Description

La définition de *vecteur lumineux* est inspirée par celle de vecteur d'éclairement (section 4.1.2). La principale différence est que nous tenons compte des caractéristiques du matériau au point considéré, et de la direction d'observation. En effet, il s'agit de représenter localement l'énergie incidente, mais en prenant en compte uniquement ses effets dans une direction réfléchie donnée. Notre ambition est de limiter ainsi les erreurs de reconstruction dans la direction réfléchie, et donc en particulier pour la lumière perçue par l'observateur.

Le vecteur lumineux calculé en un point x est noté $\overrightarrow{VL(x)}$. Dans sa version générique, il comporte deux champs : une direction $\vec{D}$, et un terme énergétique P. Leur signification sera précisée par la suite.

$$\overrightarrow{VL(x)} = \begin{pmatrix} \vec{D} \\ P \end{pmatrix}$$

Ce vecteur est de dimension $n+3$, où n dépend du mode de représentation des couleurs ($n = 3$ en *rvb*, beaucoup plus avec un modèle spectral, cf. section 4.1.3.2.2). Intuitivement, il se conçoit comme une source de lumière virtuelle simulant seule les effets cumulés des énergies incidentes au point de calcul (figure 4.1). Il contient les informations minimales nécessaires au calcul de la luminance réfléchie, et aussi aux interpolations éventuelles.

Afin d'optimiser les procédures d'interpolation, nous distinguerons par la suite des cas particuliers pour les diverses composantes de la lumière (section 4.1.5). Des données spécifiques supplémentaires seront alors attachées à ces variantes de vecteurs lumineux.

4.1.3.2 Composante directionnelle

4.1.3.2.1 Méthode de calcul La direction $\vec{D}$ du vecteur lumineux est la direction d'incidence moyenne de la lumière au point considéré. Cette direction est identique à celle du

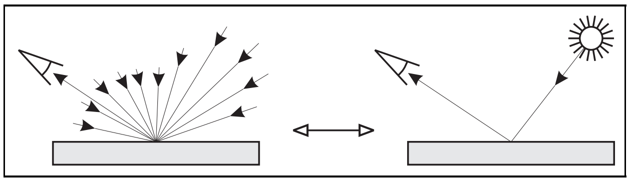


FIG. 4.1 – *Le vecteur lumineux simule une source ponctuelle virtuelle.*

vecteur d'éclairement (équation 4.1), sauf pour l'ensemble des directions incidentes que nous noterons Ω_i. En général, Ω_i n'est pas la sphère complète centrée en x ; il est restreint à certaines directions dépendant de la composante lumineuse traitée (équation 4.7).

$$\vec{D} = \int_{\Omega_i} L_i(x, \vec{\omega_i})\ \vec{\omega_i}\ d\sigma(\vec{\omega_i}) \tag{4.2}$$

En général, cette direction indique la région de la scène où se trouvent les sources les plus significatives, ou bien les objets clairs fortement éclairés qui deviennent à leur tour des sources secondaires prédominantes. Afin de faciliter les comparaisons et les interpolations ultérieures, le vecteur $\vec{D}$ est normé avant d'être mémorisé.

Comme le calcul est effectué sur la surface d'un objet, l'ensemble des directions incidentes prises en compte appartient à l'hémisphère orienté selon la normale extérieure. Dans le cas de matériaux translucides, un calcul équivalent supplémentaire est effectué sur un hémisphère orienté vers l'intérieur de l'objet. On distingue donc deux cas particuliers, pour la face externe et la face interne, qui engendrent des ensembles de vecteurs lumineux indépendants. En effet, il paraît improbable, en pratique, de relever une cohérence exploitable entre des éclairements provenant de deux régions opposées de la scène, surtout après que la lumière ait été modifiée pour une seule de ces régions par les propriétés de transmission de la matière.

4.1.3.2.2 Direction moyenne et espace de couleurs La présentation précédente du calcul de la direction moyenne d'incidence occulte les problèmes liés au modèle de couleurs. Pour obtenir une seule direction représentative, le terme $L_i(x, \vec{\omega_i})$ (équation 4.2) indique l'énergie totale incidente, intégrée sur tout le spectre lumineux. Diverses interprétations peuvent être données à cette notion.

Dans notre implémentation, nous utilisons le modèle de couleurs *rvb* classique. Celui-ci ne permet pas de simuler les phénomènes de décomposition spectrale lors des réfractions. Il est donc inutile d'associer des propriétés directionnelles différentes aux trois composantes. Comme de plus nous nous limitons au domaine visible des ondes électromagnétiques, la valeur de pondération $L_i(x, \vec{\omega_i})$ est la luminance lumineuse, que nous calculons avec l'équation 2.19.

Dans le cas d'un modèle spectral complet, la perte d'informations induite par le calcul d'une direction moyenne risque d'engendrer des erreurs conséquentes, principalement lors de la simulation des réfractions. La prise en compte des petites variations des directions réfractées liées à la longueur d'onde est primordiale. Il faudra probablement dans ce cas envisager de calculer une direction d'incidence moyenne par composante spectrale, ou par intervalle pour les méthodes adaptatives [RP97], ...et prévoir une très grande occupation de mémoire.

4.1.3.3 Composante énergétique

Le calcul de la luminance réfléchie dans une direction donnée est une évaluation de l'équation de réflectance (équation 2.6). Nous ne tenons pas compte ici de l'émission propre, car elle est traitée à part.

$$L_r(x, \vec{\omega_r}) = \int_{\Omega_i} f_r(x, \vec{\omega_i} \rightarrow \vec{\omega_r})\, L_i(x, \vec{\omega_i}) \cos\theta_i \, d\omega_i$$

On peut admettre, sous certaines conditions qui seront précisées plus loin, que la luminance incidente depuis une direction donnée est sensiblement la même en un point x', "proche" du point x, que celle calculée en x. La connaissance de l'ensemble des luminances incidentes en x permet donc d'évaluer une valeur approchée de la luminance réfléchie en x' dans une direction $\vec{\omega'_r}$:

$$L_r(x', \vec{\omega'_r}) \approx \int_{\Omega_i} f_r(x', \vec{\omega_i} \rightarrow \vec{\omega'_r})\, L_i(x, \vec{\omega_i}) \cos\theta_i \, d\omega_i$$

Cette observation conduit naturellement à une méthode d'optimisation très efficace, en remplaçant les nombreux et couteux calculs de luminances par la réutilisation de valeurs déjà connues. Elle n'est malheureusement pas envisageable en pratique. En effet, la quantité de données à stocker serait prohibitive, car l'ensemble Ω_i des directions incidentes est très grand (infini en théorie). De plus, la formulation récursive de l'équation de réflectance induit une croissance exponentielle du nombre de directions traitées. En lancer de rayons, il faudrait stocker l'arbre complet des rayons, qui est de taille démesurée, surtout si Ω_i est échantillonné pour une méthode de Monte Carlo.

Le vecteur lumineux limite autant que possible les informations conservées, en assimilant l'ensemble Ω_i à la seule direction $\vec{D}$. Le terme P est alors déterminé de façon à introduire une erreur parfaitement nulle, lorsque le vecteur lumineux est utilisé là où il a été calculé, pour la direction de réflexion $\vec{\omega_r}$. La valeur de P est :

$$P = \frac{\int_{\Omega_i} f_r(x, \vec{\omega_i} \rightarrow \vec{\omega_r})\, L_i(x, \vec{\omega_i}) \cos\theta_i \, d\omega_i}{f_r(x, \vec{D} \rightarrow \vec{\omega_r}) \cos(\vec{n}, \vec{D})} \tag{4.3}$$

Le terme énergétique P est exprimé en $[W.m^{-2}]$, afin que la formulation soit homogène. Il n'est cependant pas directement relié à la notion physique d'éclairement, bien qu'elle s'exprime dans la même unité.

Le calcul numérique de l'intégrale, au numérateur de l'expression, requiert en général une discrétisation de l'ensemble Ω_i, puis le parcours dans une boucle des cellules ainsi créées. C'est la stratégie classique de toutes les méthodes par éléments finis. Le calcul de la direction $\vec{D}$, indispensable pour déterminer la valeur de la BRDF, au dénominateur, est effectué durant le même parcours de boucle, avec très peu de calculs supplémentaires, puisque les valeurs de luminances et d'angles solides nécessaires sont identiques.

4.1.4 Utilisation des vecteurs lumineux

4.1.4.1 Calcul de luminance

Lorsque le vecteur lumineux est connu exactement, en un point x et pour une direction de réflexion $\vec{\omega_r}$, la luminance réfléchie se calcule simplement avec la formule :

$$L_r(x, \vec{\omega_r}) \ = \ P\, f_r(x, \vec{D} \rightarrow \vec{\omega_r}) \cos(\vec{n}, \vec{D}) \tag{4.4}$$

Dans ce cas, par définition du vecteur lumineux, l'erreur est nulle.

Pour un point x' "proche" de x, de normale $\vec{n'}$, et pour une direction de réflexion $\vec{\omega'_r}$, une approximation de la luminance réfléchie est donnée par :

$$L_r(x', \vec{\omega'_r}) \ \approx \ P\, f_r(x', \vec{D} \rightarrow \vec{\omega'_r}) \cos(\vec{n'}, \vec{D}) \tag{4.5}$$

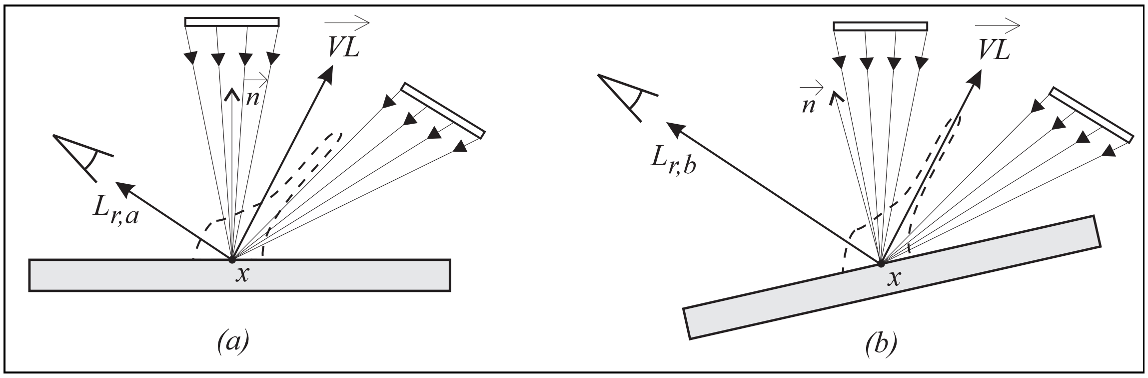


FIG. 4.2 – *Changement d'orientation de la surface.*

Ce calcul peut introduire deux types d'erreurs :

- **Une erreur due à la position :** le point x' doit être dans un voisinage suffisamment petit du point x, pour pouvoir admettre que les luminances incidentes ne varient pas de façon significative.

- **Une erreur due à l'orientation :** lorsque la BRDF est d'un aspect tourmenté, et que $\vec{\omega'_r}$ est éloignée de $\vec{\omega_r}$, la plus grande partie de l'énergie incidente doit provenir d'une direction proche de $\vec{D}$. En effet, puisque l'on ne conserve qu'une seule direction d'incidence, certains pics de la BRDF risquent d'être ignorés. Il faut donc éviter qu'une trop grande quantité d'énergie ne transite par ces zones. Ce risque d'erreur est nul pour les matériaux lambertiens, de BRDF constante. Il augmente avec la complexité de la BRDF, lorsque des pics autres que le pic spéculaire apparaissent.

  La figure 4.2 donne un exemple de cette erreur. La lumière incidente en un point x provient uniquement de deux zones, qui peuvent être deux sources surfaciques, et la BRDF présente un pic spéculaire fortement marqué. Le vecteur lumineux $\vec{VL}$ est calculé dans le cas (a), où aucune énergie ne transite par le pic de la BRDF. Il indique une direction $\vec{D}$, qui elle aussi évite ce pic. En (b), la surface a subi une légère rotation autour du point x, entraînant avec elle la BRDF et son pic, qui devient aligné avec $\vec{D}$. De ce fait, la luminance réfléchie $L_{r,b}$, calculée dans le cas (b) avec le vecteur lumineux, devient beaucoup plus imporante que dans le cas (a), alors que le pic n'est vu par aucune des deux sources. De surcroît, les incidences s'éloignant dans l'ensemble de la normale à la surface, $L_{r,b}$ devrait même être inférieure à $L_{r,a}$.

Afin d'exercer un contrôle minimal sur ces erreurs, hélas difficiles à quantifier, nous décomposerons par la suite la lumière en composantes indépendantes (section 4.1.5). Par exemple, le cas d'erreur présenté sur la figure 4.2 est éliminé en associant un vecteur lumineux indépendant à chacune des deux zones fortement lumineuses. Nous utilisons sytématiquement ce procédé avec les sources lumineuses, pour lesquelles ce risque d'erreur est particulièrement important, car elles concentrent une grande quantité d'énergie sur des zones relativement petites en général.

4.1.4.2 Interpolation

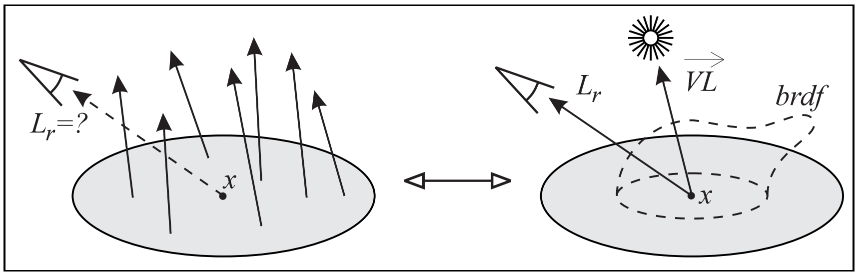


FIG. 4.3 – *Calcul d'un vecteur lumineux par interpolation.*

Lorsque le vecteur lumineux $\overrightarrow{VL(x)}$, en un point x de la scène, n'est pas connu, mais que des valeurs ont déjà été calculées dans la scène, il est parfois possible d'obtenir une valeur approchée, par une rapide interpolation vectorielle (figure 4.3). Celle-ci est valide sous certaines conditions, liées à la continuité des paramètres intervenant dans le calcul du vecteur lumineux. Ces différents critères seront détaillés plus loin.

En supposant que n vecteurs lumineux, calculés en des points $\{x_1, x_2, \ldots, x_n\}$, satisfont aux critères de validité en x, le vecteur lumineux interpolé en x s'obtient par la formule :

$$\overrightarrow{VL(x)} = \frac{\sum_{i=1}^{n} p_i \overrightarrow{VL(x_i)}}{\sum_{i=1}^{n} p_i} \tag{4.6}$$

Les coefficients de pondération $\{p_1, p_2, \ldots, p_n\}$ sont calculés de façon à minimiser l'erreur globale, en privilégiant les vecteurs lumineux risquant d'introduire l'erreur estimée la plus faible.

Les vecteurs lumineux obtenus par interpolation ne sont pas mémorisés pour une utilisation ultérieure dans d'autres interpolations. Ils risqueraient en effet d'introduire des accumulations d'erreurs non contrôlables.

4.1.4.3 Stockage en kd-arbre

Les vecteurs lumineux sont associés à des points de la scène, dans l'espace à trois dimensions. L'un des critères de validité primordiaux, lors de l'interpolation, est l'appartenance à un voisinage du point de calcul. Pour que notre méthode soit avantageuse, en durée d'exécution, nous devons être capable de retrouver le plus rapidement possible, parmi les vecteurs lumineux déjà calculés, ceux qui sont à l'intérieur d'une zone donnée. Le choix de la structure de stockage est donc capital pour obtenir de bonnes performances.

Dans notre implémentation du lancer de rayons [Roe93], la représentation interne de la scène est un graphe CSG (section 6.1.3), sans discrétisation des surfaces. Il ne nous est donc pas possible de relier les points de calcul à un quelconque maillage des objets. Le *kd-arbre* [Ben75, Ben90] est une solution particulièrement élégante au stockage spatial de données dans ce cas, et les algorithmes de recherche sont efficaces. Puisque nous travaillons en trois dimensions, nous utilisons des 3d-arbres.

Nous avons aussi testé le stockage en octree. La structure est un peu plus simple, et permet de retrouver très rapidement un élément isolé, mais les recherches dans des voisinages de tailles variables sont lents. En effet, toute notion d'adjacence à un niveau de l'arborescence est perdue au niveau suivant, à moins de tenir constamment à jour un complexe et peu efficace système de pointeurs croisés.

4.1.5 Décomposition de l'équation de rendu

L'ensemble des directions incidentes Ω_i varie avec la composante de la lumière prise en compte. Il s'agit d'un hémisphère complet dans le cas général, avec une infinité de directions possibles. Lorsque seule l'énergie provenant d'une source particulière est considérée, il se limite à l'angle solide qu'elle sous-tend, et il se réduit à une direction isolée dans le cas extrême d'une source ponctuelle. De plus, les interactions de la lumière avec les matériaux modifient sa distribution spectrale, mais aussi son comportement spatial. Par exemple, la lumière ayant subi des réflexions diffuses varie en général de façon douce, alors que les réflexions spéculaires peuvent produire des discontinuités.

Ces considérations nous ont conduits à décomposer l'équation de rendu (équation 3.1), évaluée récursivement en chaque point de la scène, en éléments plus simples. Ils prenent chacun en compte une composante différente de la lumière incidente au point de calcul. Cette distinction permet d'affiner les critères de validité pour l'interpolation et le calcul des coefficients de pondération. L'indépendance des chemins lumineux propres à chaque composante élimine le risque de comptabiliser plusieurs fois la même énergie.

$$L_r(x,\vec{\omega_r}) = L_e(x,\vec{\omega_r}) + L_{spec}(x,\vec{\omega_r}) + L_{dir}(x,\vec{\omega_r}) + L_{caust}(x,\vec{\omega_r}) + L_{ind}(x,\vec{\omega_r}) \tag{4.7}$$

- $L_e(x,\vec{\omega_r})$ est la luminance propre, émise intrinsèquement par l'objet depuis le point x dans la direction $\vec{\omega_r}$, s'il est lui même une source.

- $L_{spec}(x,\vec{\omega_r})$ est la luminance réfléchie due à l'énergie incidente selon la direction spéculaire, symétrique de $\vec{\omega_r}$ par rapport à la normale. Ce terme inclut aussi l'énergie provenant de la direction réfractée, pour les matériaux transparents obéissant aux lois de Descartes.

- $L_{dir}(x, \vec{\omega_r})$ est la luminance réfléchie due à l'énergie directe, provenant des sources sans avoir subi d'interaction avec les autres objets.

- $L_{caust}(x, \vec{\omega_r})$ est la luminance réfléchie due à la lumière n'ayant subi que des réflexions ou réfractions spéculaires entre les sources et le point x.

- $L_{ind}(x, \vec{\omega_r})$ est la luminance réfléchie due à l'énergie indirecte, qui a subi des interactions avec des objets non spéculaires de la scène.

Le terme $L_e(x, \vec{\omega_r})$, lorsqu'il existe, est défini par les propriétés de l'objet, sans méthode générique de calcul. Le terme $L_{spec}(x, \vec{\omega_r})$ est évalué par la méthode classique en lancer de rayons, en générant un rayon dans la direction spéculaire ou réfractée, ou bien en échantillonnant un cône englobant le lobe spéculaire s'il ne s'agit pas d'un pic de Dirac. Les vecteurs lumineux ne sont pas adaptés à sa prise en compte, car ses variations peuvent être très importantes sans qu'il soit possible de le prévoir.

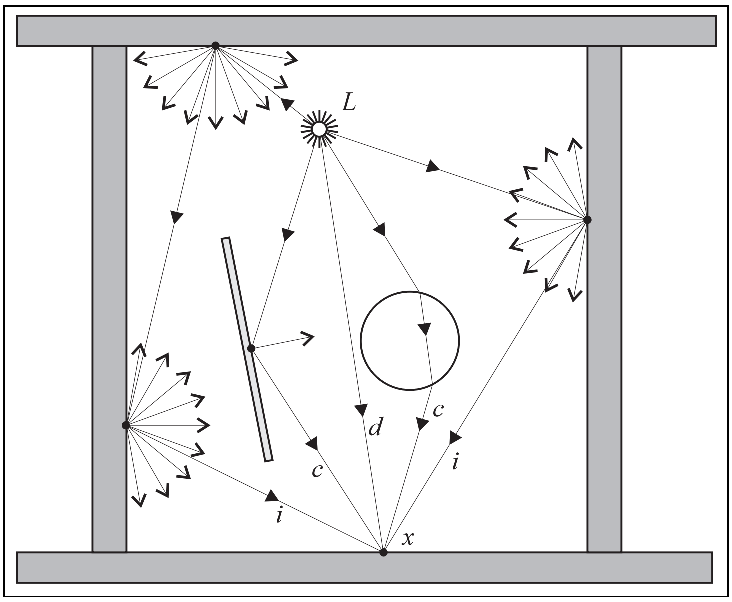


FIG. 4.4 – *Différentes composantes de la lumière en un point x de la scène*

Il reste donc trois composantes lumineuses, qui donnent naissance à trois variantes des vecteurs lumineux :

- la composante directe $L_{dir}(x, \vec{\omega_r})$, chemin de type d sur la figure 4.4, engendre les *vecteurs lumineux directs*, notés $(\overrightarrow{VLD})$;

- la composante caustique $L_{caust}(x, \vec{\omega_r})$, chemins de type c sur la figure 4.4, engendre les *vecteurs lumineux caustiques*, notés $(\overrightarrow{VLC})$;

- la composante indirecte $L_{ind}(x, \vec{\omega_r})$, chemins de type i sur la figure 4.4, engendre les *vecteurs lumineux indirects*, notés $(\overrightarrow{VLI})$.

Les trois types de vecteurs lumineux sont stockés au fur et à mesure de leur calcul dans des 3d-arbres distincts : un pour tous les $(\overrightarrow{VLC})$, un pour tous les $(\overrightarrow{VLI})$, et un pour les $(\overrightarrow{VLD})$ de chaque source.

De plus, afin de garantir une prise en compte unique des diverses énergies, nous supposerons que la BRDF peut être séparée en un terme spéculaire pur $f_{r,s}$, et un terme diffus $f_{r,d}$, qui n'est pas uniforme en général :

$$f_r(x,\vec{\omega_i}\to\vec{\omega_r}) = f_{r,s}(x,\vec{\omega_i}\to\vec{\omega_r}) + f_{r,d}(x,\vec{\omega_i}\to\vec{\omega_r}) \tag{4.8}$$

Cette décomposition est généralement naturelle avec les modèles analytiques de matériaux. Avec les BRDF mesurées expérimentalement, il faut détecter et extraire les éventuels pics de réflexion. S'ils indiquent des directions autres que la direction spéculaire, leur prise en compte se fera en générant des rayons explicites.

4.1.6 Calcul d'une image avec les vecteurs lumineux

4.1.6.1 Principe général

Lors du calcul d'une image, l'intégrale de rendu doit être évaluée à de très nombreuses reprises. Elle est décomposée selon l'équation 4.7, puis les différents termes sont traités séparément. La seule différence par rapport aux algorithmes de rendu habituels concerne les trois composantes donnant lieu au calcul de vecteurs lumineux.

A chaque fois que la composante directe, caustique, ou indirecte est requise, au lieu de lancer systématiquement un calcul classique, des vecteurs lumineux déjà calculés sont recherchés dans le 3d-arbre correspondant, au voisinage du point courant. S'ils sont trouvés en nombre suffisant, et si leurs paramètres correspondent à des critères de validité convenus, alors un vecteur lumineux approximatif est évalué par interpolation, et il est utilisé avec l'équation 4.5 pour calculer une valeur approchée de la luminance réfléchie. Dans le cas contraire, un calcul complet est effectué, et il est à son tour stocké sous forme de vecteur lumineux, pour un éventuel usage futur.

Ce principe est identique pour les $(\overrightarrow{VLD})$, $(\overrightarrow{VLC})$, et $(\overrightarrow{VLI})$. Seuls les critères de validité et les calculs de pondération sont différents. Ces trois méthodes sont présentées à la section 4.2.

4.1.6.2 Création d'un germe

La progression classique, ligne par ligne, du calcul de l'image risque d'introduire un déséquilibre entre les zones où les vecteurs lumineux sont calculés et celles où ils sont obtenus par interpolation. En effet, au début du calcul, les vecteurs lumineux sont tous calculés explicitement. Puis, lorsque leur quantité est suffisante, les interpolations deviennent de plus en plus fréquentes. Mais le processus est faussé, car la plupart des vecteurs connus se trouvent dans la même région de l'image, au dessus de la ligne courante. Tout le reste est ignoré, et des discontinuités risquent de ne pas être détectées.

Pour remédier à ce problème, un germe de vecteurs lumineux est initialisé avant le calcul de l'image. Des points répartis aléatoirement dans la scène sont générés, en lançant des rayons de directions aléatoires depuis l'œil. Les vecteurs lumineux correspondants sont calculés et mémorisés, sans tenter d'interpolation durant cette phase. Ils seront utilisés lors du calcul final de l'image.

La répartition des points de calcul n'est pas complètement uniforme, car la zone de la scène visible sur l'image est privilégiée lors du choix initial de directions. En effet, les éclairements seront très souvent calculés pour cette région, au moins une fois par pixel, et il est logique de favoriser ici les interpolations. Le reste de la scène est sollicité par la génération des rayons réfléchis et réfractés, ce qui favorise les zones intervenant de façon significative dans l'éclairement de la région observée. Ainsi, la densité de répartition des vecteurs lumineux dans la scène, à l'issue de la création du germe, est sensiblement proportionnelle aux besoins qui seront exprimés lors du calcul définitif de l'image. Dans le cas où une méthode de Monte Carlo est utilisée pour calculer l'éclairement indirect, cette répartition tend naturellement à devenir uniforme, car toutes les zones de la scène interviennent en chaque point.

4.2 Différents types de vecteurs lumineux

4.2.1 Vecteur lumineux direct

L'utilisation des vecteurs lumineux, dans le cas des éclairements directs, s'apparente à celle de la *photon map* de Jensen [Jen96], adaptée à notre modélisation vectorielle. Elle est particulièrement avantageuse lors de la simulation de sources complexes nécessitant un échantillonnage coûteux.

4.2.1.1 Définition

La prise en compte d'une source, en lancer de rayons, se fait en générant un *rayon d'ombre* du point courant vers un point de la source. Si celui-ci ne rencontre aucun objet sur son parcours, la source est visible, et sa contribution énergétique peut être calculée. Les sources étendues, surfaciques ou volumiques, sont susceptibles d'être partiellement visibles, ce qui crée les zones de pénombre. La stratégie dans ce cas consiste à les subdiviser en éléments de petite taille, vers lesquels sont lançés des rayons d'ombre indépendants (figure 4.5). Seule l'énergie des échantillons visibles est comptabilisée. Le chapitre 5 propose une méthode de subdivision adaptative des sources planes.

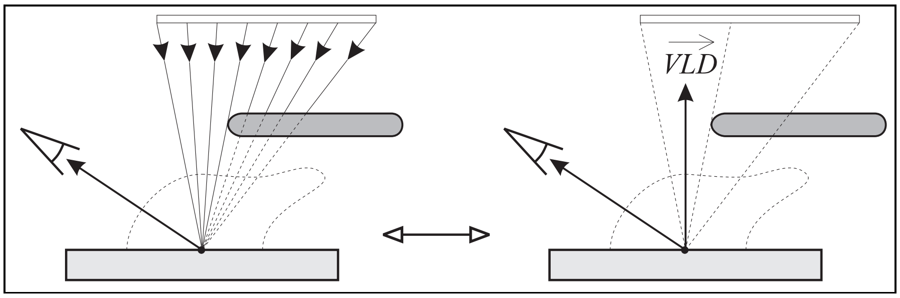


FIG. 4.5 – *Vecteur lumineux direct avec une source partiellement obstruée*

Le *vecteur lumineux direct* représente, en un point quelconque de la scène, l'énergie apportée par une source donnée. Il se compose de deux parties, comme les vecteurs lumineux génériques :

- Une direction d'incidence moyenne $\vec{D}$, obtenue en pondérant les n directions d'échantillonnage[1] de la source $\vec{\omega_i}$ par la luminance incidente L_i et l'angle solide associé Ω_i : $\vec{D'} = \sum_{i=1}^{n} L_i \Omega_i \vec{\omega_i} \;\Rightarrow\; \vec{D} = \vec{D'}/\|\vec{D'}\|$. Ce calcul est réduit à sa plus simple expression pour les sources ponctuelles.

- Une grandeur énergétique P vérifiant l'équation 4.4.

Lorsque la source est complètement occultée, le $\overrightarrow{VLD}$ est quand même mémorisé, afin de réduire le nombre de rayons ultérieurs. Dans ce cas, par convention, $\vec{D}$ indique la direction du centre de gravité de la source, et P est nul.

Lorsque la scène comporte plusieurs sources, des vecteur lumineux directs sont calculés pour chacune d'elles, et ils sont stockés dans des 3d-arbres indépendants. En effet, l'utilisation d'un seul vecteur lumineux direct global, qui regrouperait les contributions de toutes les sources, ne semble pas une solution correcte. Chaque source crée des zones de pénombres différentes, avec des risques de variations brusques qui lui sont propres.

Par contre, dans le cas de scènes comportant de très nombreuses sources, le regroupement de celles qui ont des caractéristiques spectrales similaires, et qui sont proches géographiquement les unes des autres, est un compromis permettant de limiter l'encombrement en mémoire. L'évolution de la synthèse d'images vers la simulation de scènes toujours plus complexes [War91, KJ91, SWZ96], avec tous les problèmes liés à la gestion simultanée de nombreux objets, ouvre des perspectives de recherche en ce sens.

4.2.1.2 Interpolation directe

4.2.1.2.1 Recherche des candidats Un ensemble de $\overrightarrow{VLD}$ candidats est sélectionné avant de lancer l'interpolation. Il est constitué des $nmax$ plus proches voisins du point courant, en se limitant à un voisinage de rayon $rmax$. Si le nombre de $\overrightarrow{VLD}$ trouvés est inférieur à un seuil $nmin$, cet ensemble est rejeté et un calcul complet est lancé. Le choix de ces trois paramètres dépend principalement de la taille de la scène. En pratique, pour des simulations d'intérieurs, les valeurs $nmin = 5$, $nmax = 20$ et $rmax \approx 0.1\ m$ donnent de bons résultats, avec un germe initial de 5000 $\overrightarrow{VLD}$.

4.2.1.2.2 Test de validité Même avec des sources étendues, l'éclairement direct est susceptible de varier brusquement dans les zones de pénombre. Il est indispensable de détecter ces zones où l'interpolation risque d'introduire un lissage préjudiciable à la qualité de l'image. Nous utilisons un critère basé sur les variations de luminance lumineuse, estimée avec la formule 2.19, à l'intérieur de l'ensemble des candidats. Avec des espaces colorimétriques plus sophistiqués que le simpliste modèle *rvb*, ce test pourrait être contrôlé par une distance psychovisuelle entre couleurs.

La dispersion en luminance lumineuse ne doit pas dépasser un seuil prédéterminé S_{dir}. Elle est évaluée en calculant l'écart-type entre les échantillons sélectionnés. En pratique, un seuil de 10 % à 20 % est assez faible pour détecter correctement les variations rapides de l'éclairement direct. Cette valeur empirique dépend cependant de la taille de la zone de recherche (section 6.2.2).

1. L'échantillonnage de la source peut être quelconque en pratique : ponctuel, surfacique, ou même pas d'échantillonnage du tout.

4.2.1.2.3 Fonction de pondération Le poids p_i de l'échantillon i, pour l'interpolation, est une fonction décroissante $f(d_i)$ de la distance d_i entre le point courant, où doit être évalué un nouveau $\overrightarrow{VLD}$, et le point où a été calculé le candidat i. C'est une fonction polynomiale rapide à calculer, d'allure gaussienne pour favoriser les échantillons les plus proches :

$$f(d_i) = \left[2(\frac{d_i}{rmax})^3 - 3(\frac{d_i}{rmax})^2 + 1\right]^2 \tag{4.9}$$

Finalement, l'interpolation pondérée des n candidats valides donne les valeurs des composantes du nouveau $\overrightarrow{VLD}$:

$$\vec{D} = \frac{\sum_{i=1}^{n} f(d_i)\vec{D}_i}{\sum_{i=1}^{n} f(d_i)} \qquad \text{et} \qquad P = \frac{\sum_{i=1}^{n} f(d_i)P_i}{\sum_{i=1}^{n} f(d_i)}$$

Puis, ce $\overrightarrow{VLD}$ interpolé est utilisé avec l'équation 4.5 pour calculer la luminance réfléchie due à cette source. L'opération est réitérée pour toutes les sources de la scène, pour obtenir la composante directe globale.

4.2.2 Vecteur lumineux caustique

Cette section présente l'intégration du calcul de caustiques (section 3.6.4.1) dans notre modélisation vectorielle. Ces travaux ont été réalisés conjointement avec Xavier Serpaggi, lors de son stage de DEA Images à Saint-Etienne, de mars à août 1997. Les développements très complets de la thèse de Collins [Col97] nous ont largement inspirés.

4.2.2.1 Définition vectorielle

Le *Vecteur Lumineux Caustique* $\overrightarrow{VLC}$ est semblable à la notion de *photon* de Jensen. Nous conservons cependant notre appellation, car elle met en avant l'homogénéité de la représentation vectorielle des diverses composantes lumineuses.

- La direction $\vec{D}$ est celle d'incidence de la dernière portion du chemin ayant conduit de la source au point courant.
- La grandeur énergétique P est égale à la puissance de la source divisée par le nombre total de rayons lancés. Si le rayon subit des réflexions ou des réfractions, P est atténuée par la composante spéculaire $f_{r,s}$ des matériaux rencontrés.

Les $\overrightarrow{VLC}$ sont stockés dans un 3d-arbre.

4.2.2.2 Distribution d'énergie caustique

Lors de la première passe, l'énergie des sources est distribuée dans la scène, pour être stockée sur la composante diffuse des matériaux lorsqu'elle a subi des réflexions ou réfractions spéculaires. La représentation des scènes par arbre CSG masquant toute connaissance globale des objets et de leurs caractéristiques, la recherche des surfaces spéculaires est un processus stochastique. Un grand nombre de rayons sont générés depuis la source dans des directions aléatoires. Chacun d'eux transporte une fraction de l'énergie totale de la source. Cette opération s'effectue en deux étapes, avec un échantillonnage adaptatif.

- Un premier passage parcourt toutes les directions visibles depuis la source avec un échantillonnage relativement grossier, dont le pas est guidé par la taille des objets que l'on accepte de ne pas toujours détecter. Tout en distribuant une partie de l'énergie, il permet aussi de détecter les régions où se trouvent des objets spéculaires susceptibles de créer des caustiques.
- Ensuite, des rayons supplémentaires sont générés vers les zones détectées précédemment et leurs voisinages immédiats. Ainsi, la densité de rayons est plus forte sur les objets qui créent des caustiques.

4.2.2.3 Occupation de mémoire

Les caustiques sont un phénomène essentiellement de hautes fréquences. L'échantillonnage des sources doit donc être très fin, au moins plusieurs milliers de rayons. Il se pose alors un sérieux problème d'occupation de mémoire, surtout lorsque la scène comporte des objets transparents. En effet chaque intersection génère alors un rayon réfléchi et un rayon réfracté (hormis lors des plus rares cas de réflexion totale), qui donneront eux-même naissance en cascade à d'autres rayons et d'autres intersections. Le processus étant fortement récursif, le nombre total de rayons, et donc de vecteurs lumineux caustiques à mémoriser, peut devenir considérable.

Il existe une solution envisageable à ce problème, mais elle semble complexe à mettre en œuvre. Il faudrait, à chaque fois que le 3d-arbre de stockage des $\overrightarrow{VLC}$ dépasse une taille raisonnable, le parcourir pour regrouper en un seul les vecteurs "proches" les uns des autres, puis reconstruire un nouvel arbre équilibré. Le critère de proximité reste à définir, pour ne pas introduire un lissage excessif.

Une autre solution, un peu simpliste mais efficace, consiste à abuser sans modération de l'augmentation continuelle des performances des ordinateurs. Nous avons opté pour cette méthode drastique.

4.2.2.4 Calcul de la composante caustique

La reconstitution de la composante caustique n'est pas une interpolation, mais une estimation de densité. Il faut recueillir toute l'énergie incidente dans un voisinage du point courant. Le choix de la taille de la zone de recherche détermine l'ampleur de l'effet de lissage qui élimine le bruit dû à la discrétisation. Si cette zone est trop petite le bruit persiste, mais si elle est trop grande les hautes fréquences de la caustique disparaissent. Sa taille doit être au moins égale à celle de la projection d'un pixel sur la surface. Une valeur de l'ordre de quelques millimètres est utilisée pour des simulations d'intérieurs courants.

Un *noyau d'estimation* K permet d'assurer la continuité de la reconstruction de la fonction de densité. La composante caustique de la luminance réfléchie s'exprime comme :

$$L_{caust}(x, \vec{\omega_r}) = \sum_{i=1}^{n} K(x, x_i)\, f_r(x, \vec{\omega_i} \to \vec{\omega_r})\, P(x_i)\, \cos(\vec{n}, \overrightarrow{D(x_i)})$$

Le choix du noyau optimal est toujours une question ouverte. A l'instar de Collins, nous utilisons un noyau d'allure gaussienne, mais sous la forme d'une approximation polynomiale (équation 4.9).

4.2.2.5 Caustiques et sources surfaciques

Les sources étendues créent, elles-aussi, des caustiques. Leur prise en compte est plus délicate, car elles ne sont pas seulement visibles ou masquées comme les sources ponctuelles, mais elles peuvent aussi être partiellement obstruées.

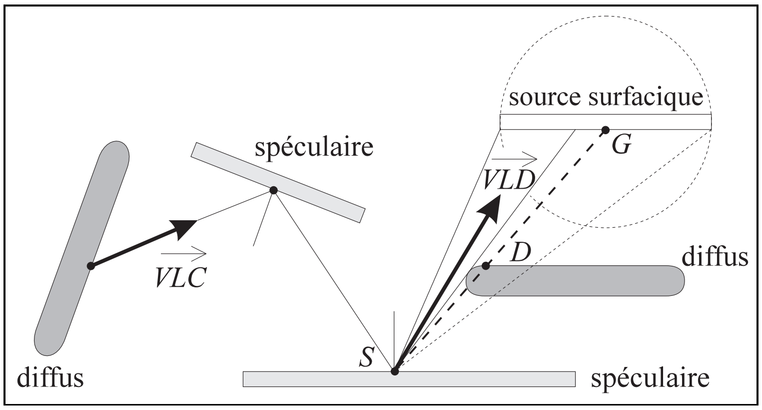


Fig. 4.6 – *Calcul de caustique avec une source surfacique*

La méthode brutale de prise en compte consiste à générer des rayons dans toutes les directions, depuis tous les points de la surface. Ce n'est pas une solution envisageable en pratique. Nous proposons une méthode plus efficace, mais tout de même assez coûteuse, mettant en œuvre les vecteurs lumineux directs et caustiques.

Comme pour les sources ponctuelles, la première phase est une étape de distribution d'énergie. Une sphère virtuelle englobant la source complète est centrée en un point G de sa surface (figure 4.6). De façon arbitraire, nous choisissons le centre de gravité. La sphère est échantillonnée dans toutes les directions, et sert de guide pour générer des rayons uniformément distribués depuis G.

Toutes les intersections des rayons avec la scène sont calculées, et non pas la première uniquement, même lorsqu'un objet opaque est rencontré. En effet, les objets situés derrière lui sont peut-être partiellement éclairés. C'est le cas du point S, situé en arrière du point D (figure 4.6). Pour tous les objets spéculaires de la liste d'intersections, le $\overrightarrow{VLD}$ est calculé pour déterminer la contribution de la source. Si celle-ci n'est pas nulle, un chemin spéculaire est généré jusqu'à ce qu'un objet diffus soit rencontré, sur lequel est mémorisé un $\overrightarrow{VLC}$, ou bien jusqu'à ce que l'énergie soit totalement absorbée.

Bien que plus simple que la méthode brutale, ce processus est coûteux, car de nombreuses intersections doivent être testées. Cependant, il n'est effectué qu'une seule fois par calcul d'image. Le principal problème réside plutôt dans l'occupation de mémoire, car la densité de rayons est importante.

Remarques :

- Lorsque la source n'émet que sur une seule de ses deux faces, la sphère est remplacée par un hémisphère.

- Le fait d'assimiler la partie visible de la source à l'unique direction du $\overrightarrow{VLC}$ introduit une erreur. Tout se passe comme si plusieurs directions incidentes donnaient la même direction réfléchie. Pour en limiter les conséquences, il est possible de scinder la source en éléments de plus petite taille. En pratique, cette erreur compense le fait que les matériaux ne sont jamais des réflecteurs spéculaires parfaits. Le cône d'incertitude simule la taille non nulle du pic spéculaire. Il ne semble malheureusement pas possible de quantifier exactement cette assertion.

- La même méthode s'applique sans difficulté aux sources volumiques classiques modélisées par des amas de sources ponctuelles.

- Lorsqu'un objet diffus est rencontré, le $\overrightarrow{VLD}$ est quand même calculé et mémorisé. Il servira éventuellement lors des interpolations directes.

4.2.3 Vecteur lumineux indirect

4.2.3.1 Principe général

La composante indirecte de la luminance réfléchie est évaluée par une méthode de Monte Carlo. Celle-ci échantillonne un hémisphère centré sur le point courant et lance des rayons vers les cellules créées pour déterminer la luminance incidente, en ne tenant compte que de la lumière ayant été réfléchie au moins une fois sur une surface diffuse, donc en excluant les composantes directes et caustiques. Cette opération est extrêmement coûteuse, à cause du très grand nombre de rayons générés et de la récursivité de l'équation de rendu (équation 3.1).

Une hypothèse est couramment admise en synthèse d'images [WRC88] : *L'éclairement indirect varie "lentement" à l'intérieur de la scène.* Elle permet de réduire très sensiblement la quantité de calculs. La connaissance de l'éclairement indirect en un nombre réduit de points est alors suffisante pour le reconstruire dans toute la scène. Nous proposons ici une méthode permettant de mener à bien cette reconstruction en limitant les erreurs commises, aidée par le calcul du gradient d'éclairement indirect.

Notre approche est inspirée par celle Ward, qui est à l'origine du célèbre logiciel de lancer de rayons *Radiance* [War94]. La principale différence concerne la prise en compte plus complète de la BRDF. Les informations directionnelles des vecteurs lumineux nous permettent de ne pas nous limiter à la seule composante parfaitement diffuse des éclairements indirects.

4.2.3.2 Méthode de Monte Carlo

4.2.3.2.1 Discrétisation La composante indirecte n'a pas de direction d'incidence privilégiée. Il faut donc, pour l'évaluer, analyser la lumière provenant de toutes les directions possibles. Le lancer de rayon procède à un échantillonnage ponctuel de la scène vue depuis le point de calcul. Nous avons implémenté une méthode de Monte Carlo stratifiée (annexe A) pour reconstituer l'information globale. Elle consiste à décomposer l'ensemble des directions incidentes en un grand nombre de cellules, suffisamment petites pour considérer qu'elles sont de luminance constante. Cette dernière est alors évaluée en générant un rayon dans chaque cellule pour déterminer sa contribution.

Comme le processus est récursif, cette opération est réitérée depuis les points de la scène vus dans les directions des cellules. La croissance du nombre de rayons à générer est exponentielle et devient vite prohibitive. Pour limiter le processus, aucun rayon supplémentaire n'est généré lorsqu'un niveau de récursivité prédéfini est atteint. Cette troncature arbitraire est très simple à mettre en œuvre, et elle réduit efficacement la durée des calculs, mais elle introduit un biais dans le calcul, tout comme celle utilisée par Ward (section 3.6.3.2.2). En effet, à un niveau de récursivité donné, rien ne peut garantir par avance que la contribution du niveau suivant ne soit pas prépondérante dans le résultat global. La convergence vers une solution exacte n'est pas toujours assurée, et aucune estimation d'erreur n'est possible.

Cependant, en pratique, nous avons remarqué que l'augmentation de la profondeur de récursivité n'apporte très rapidement plus d'amélioration visuelle, pour des scènes courantes dont la plupart des objets ont un comportement plutôt diffus. Le premier niveau de calcul est bien sûr capital, mais déjà le deuxième n'apporte en général que peu de différences visibles. Pousser plus avant les calculs ne semble nécessaire que pour les scènes complexes, où la lumière indirecte suit des chemins sinueux, et où l'éclairement direct de la zone visible est faible. Toutefois, notre modélisation vectorielle n'est pas tributaire de la méthode de Monte Carlo choisie. Il est tout à fait possible d'implémenter des méthodes non biaisées (section A), plus lourdes à mettre en œuvre, le *suivi de chemins* (section A.4.3) par exemple.

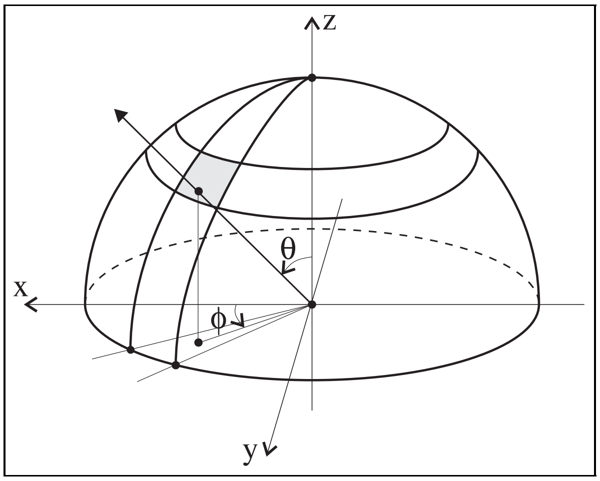


FIG. 4.7 – *Maillage de l'hémisphère*

Pour la composante indirecte, nous nous intéressons ici uniquement à la réflexion. L'ensemble des directions incidentes se limite donc à l'hémisphère orienté selon la normale (figure 4.7). Pour la discrétisation, nous proposons un découpage à angle solide constant, qui nous permettra d'accéder directement aux informations directionnelles utiles au calcul du gradient. Nous reprenons une notation similaire à celle de Ward pour repérer les cellules de l'hémisphère échantillonné, bien que le découpage ne soit pas le même, puisque le sien est à angle solide projeté constant.

L'angle zénithal $\theta \in [0, \pi/2]$ est mesuré par rapport à la normale. Il est indicé par la lettre $j \in [0, M-1]$ pour les M découpages zénithaux. L'angle azimutal $\phi \in [0, 2\pi]$ est mesuré par rapport à un repère du plan tangent. Il est indicé par la lettre $k \in [0, N-1]$ pour les N découpages azimutaux. L'hémisphère est donc échantillonné en MN cellules, repérées par leurs indices (j, k). Une direction est générée aléatoirement à l'intérieur de chaque cellule, repérée par ses coordonnées polaires (θ_j, ϕ_k).

L'hémisphère complet sous-tend un angle solide de $2\pi\, sr$. Le découpage étant à angle solide constant, chaque cellule sous-tend un angle solide $\omega_{j,k} = \frac{2\pi}{MN}\, sr$. La position explicite des bordures d'une cellule (j, k) n'est pas calculée, car seule la direction (θ_j, ϕ_k), à l'intérieur de la cellule, nous est utile. Si X and Y sont des variables aléatoires uniformes dans $[0, 1]$, nous avons :

$$\theta_j = cos^{-1}\left(\frac{M-j-X}{M}\right) \qquad \text{et} \qquad \phi_k = \frac{2\pi}{N}\,(k+Y)$$

4.2.3.2.2 Calcul de la composante indirecte

Après discrétisation, l'équation de réflectance (équation 2.6) devient :

$$L_{ind}(x, \vec{\omega_r}) \approx \frac{2\pi}{MN} \sum_{j=0}^{M-1} \sum_{k=0}^{N-1} f_{r,d}(x, \vec{\omega_{j,k}} \rightarrow \vec{\omega_r})\, L_i(x, \theta_j, \phi_k) \cos\theta_j \tag{4.10}$$

Le terme $f_{r,d}(x, \vec{\omega_{j,k}} \rightarrow \vec{\omega_r})$ est la composante diffuse, pas nécessairement uniforme, de la BRDF (équation 4.8). La composante spéculaire est traitée à part, de façon classique en lancer de rayons, en générant un ou plusieurs rayons réfléchis spéculaires. Cette décomposition de la BRDF élimine le risque de prendre en compte plusieurs fois la même énergie, notamment pour les cellules qui sont proches de la direction de réflexion spéculaire.

4.2.3.3 Direction d'incidence moyenne

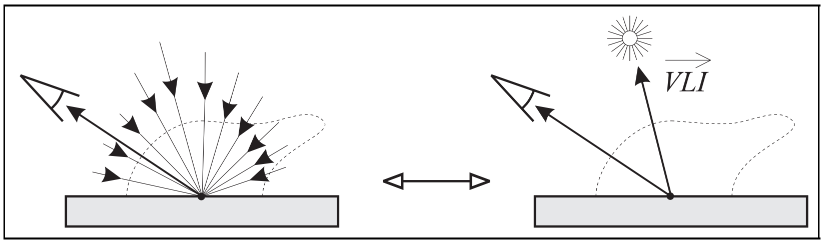


Fig. 4.8 – *Vecteur lumineux indirect*

La direction d'incidence moyenne $\vec{D}$ se calcule en discrétisant sur l'hémisphère la formulation générique (équation 4.2). Elle est donnée par l'expression :

$$\vec{D} \approx \sum_{j=0}^{M-1} \sum_{k=0}^{N-1} L_i(x, \theta_j, \phi_k)\, \vec{\omega_i} \tag{4.11}$$

Comme la mesure de l'angle solide $\sigma(\vec{\omega_i})$ est constante, elle n'intervient pas dans le calcul de direction.

La direction $\vec{\omega_i}$ de la cellule (j,k), dans un repère lié au plan tangent en x, est :

$$\vec{\omega_i} = \begin{pmatrix} \sin\theta_j \ \cos\phi_k \\ \sin\theta_j \ \sin\phi_k \\ \cos\theta_j \end{pmatrix}$$

Les coordonnées du vecteur $\vec{D}$ se calculent durant le même parcours de boucle que celui d'évaluation de la composante indirecte (équation 4.10). Ce vecteur est normé avant d'être mémorisé. Dans le cas où l'énergie indirecte est nulle, par convention, $\vec{D}$ indique la normale à la surface.

4.2.3.4 Gradient d'éclairement indirect

4.2.3.4.1 Description Un calcul de gradient est effectué lors de l'évaluation de la composante indirecte (équation 4.10), par dérivation de l'équation de l'éclairement incident (équation 2.2), pour des déplacements infinitésimaux du centre de l'hémisphère. Le détail des calculs est donné à l'annexe B. Le gradient d'éclairement est le vecteur $\vec{\Delta E}$:

$$\vec{\Delta E} = \frac{\pi^2}{MN} \begin{pmatrix} -\sum_{j=0}^{M-1}\sum_{k=0}^{N-1} \cos 2\theta_j \cdot \cos\theta_j \cdot \cos\phi_k \cdot L(\theta_j,\phi_k) \\ -\sum_{j=0}^{M-1}\sum_{k=0}^{N-1} \cos 2\theta_j \cdot \cos\theta_j \cdot \sin(\phi_k) \cdot L(\theta_j,\phi_k) \\ \sum_{j=0}^{M-1}\sum_{k=0}^{N-1} \cos 2\theta_j \cdot \sin\theta_j \cdot L(\theta_j,\phi_k) \end{pmatrix} \tag{4.12}$$

Le gradient indique les variations de l'éclairement dans un voisinage infiniment petit du point de calcul. A cause des erreurs de discrétisation et des simplifications de calcul, il n'est pas exploitable en tant que gradient mathématique. Par contre, il fournit une bonne indication du niveau de perturbation de la zone considérée. Nous l'utilisons pour déterminer la taille de la zone de validité associée au vecteur lumineux indirect (section 4.2.4.2), et pour le calcul des coefficients de pondération. En pratique, nous calculons un gradient de luminance lumineuse (équation 2.19).

4.2.4 Fuites de lumière

Le problème des *fuites de lumière*, soulevé par Vedel dans [Ved93], est partiellement résolu en radiosité par l'utilisation de maillages adaptatifs, au prix d'une grande complexité. Il se traduit par des taches trop sombres ou trop lumineuses à proximité de certains petits objets opaques, comme si une faible quantité de lumière ou d'ombre parvenait sournoisement à se glisser sous eux. Le système visuel humain est particulièrement sensible à ce type de défauts, même lorsqu'ils sont rares et peu marqués.

La zone de validité d'un $\overrightarrow{VLI}$ peut dans certains cas déborder au delà d'un objet proche, s'il est trop petit pour intervenir de façon sensible dans le calcul de la moyenne des distances. Ce $\overrightarrow{VLI}$ sera alors pris en compte dans l'interpolation de l'autre côté du petit obstacle, dont il va fausser le résultat. C'est par exemple le cas pour le point B (figure 4.9) qui se situe dans la zone de validité du point A. Cette situation est sans conséquence avec un éclairage

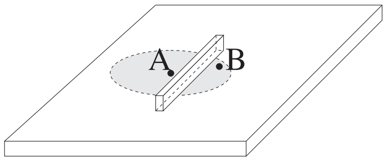


FIG. 4.9 – *Exemple de fuite de lumière*

perpendiculaire à la surface, mais risque de produire des défauts visibles si l'éclairage indirect moyen est oblique ou rasant.

Il est difficile de détecter ce risque de fuites de lumière dans un environnement *CSG*, car on n'a pas de connaissance globale de la scène. Nous proposons ici deux méthodes pour résoudre ce problème. La première est pratiquement radicale, mais coûteuse, alors que la seconde est un compromis raisonnable.

- Il est possible de limiter arbitrairement la taille de la zone de validité à la distance minimale rencontrée lors du parcours de l'hémisphère. On est ainsi certain que la zone de validité ne dépassera pas l'objet le plus proche, sauf dans le cas extrême et desespéré où il est trop petit et est oublié lors de l'échantillonnage. Cette méthode va cependant réduire les zones de validité de tous les $\overrightarrow{VLI}$, et obliger à en calculer un plus grand nombre, même dans les régions où cela n'est pas indispensable.

- L'autre possibilité, moins coûteuse mais légèrement moins fiable, est inspirée par le principe des maillages adaptatifs. Il s'agit d'adapter le poids des $\overrightarrow{VLI}$ au risque de fuites de lumière. La probabilité d'apparition de ce phénomène augmente lorsque la distance minimale détectée est très inférieure à la moyenne des distances entre le point de calcul et les autres objets. On peut donc minimiser ce risque en multipliant le poids du $\overrightarrow{VLI}$ par un coefficient inversement proportionnel à la différence entre la distance moyenne et la distance minimale.

Les calculs présentés ici correspondent à la deuxième solution. Une fonction quadratique (portion de parabole inversée) évite que le poids du $\overrightarrow{VLI}$ ne décroisse trop rapidement, lorsque la différence entre la distance moyenne $dmoy$ et la distance minimale $dmin$ augmente. Le calcul du coefficient de validité $k_{valid} \in [0, 1]$ est :

$$k_{valid} = 1 - \left(1 - \frac{dmin}{dmoy}\right)^2 \qquad (4.13)$$

4.2.4.1 Description du vecteur lumineux indirect

Le *Vecteur Lumineux Indirect* $\overrightarrow{VLI}$ mémorise les informations nécessaires à la reconstruction de la composante indirecte de la luminance réfléchie. Il possède quelques champs de plus que le $\overrightarrow{VLD}$ car la procédure d'interpolation est un peu plus sophistiquée :

- le gradient de luminance lumineuse ;

- le repère local dans lequel le gradient a été calculé ;

- la distance arithmétique moyenne entre le point de calcul et les autres objets de la scène ;
- le coefficient de validité k_{valid} (équation 4.13).

Ces données supplémentaires sont toutes calculées lors de l'unique parcours des cellules de l'hémisphère, pour un coût très faible. Elles permettent de mieux contrôler les erreurs de reconstruction, et il est ainsi possible de limiter le nombre total de $\overrightarrow{VLI}$ calculés par la méthode de Monte Carlo.

4.2.4.2 Zone de validité

Un $\overrightarrow{VLI(A)}$, calculé en un point A, doit répondre à plusieurs critères pour être admis comme valide en un point B :

- Les points A et B doivent appartenir à des surfaces de même matériau, afin de limiter les variations de BRDF. L'idéal serait de se munir d'une hypothétique notion de distance entre BRDF, pour définir des classes de matériaux proches les uns des autres, et être moins strict sur la validité des $\overrightarrow{VLI}$.
- La distance $\|\overrightarrow{AB}\|$ doit être inférieure à la distance moyenne entre A et le reste de la scène. Ce critère rend plus difficiles les interpolations à proximités des autres objets, là où les interactions indirectes sont prédominantes. Il limite, sans l'éliminer complètement cependant, le risque de fuite de lumière (section 4.2.4), en adaptant la taille de la zone de validité à l'encombrement local de la scène. Le coefficient de validité k_{valid} (équation 4.13), intervenant dans le calcul du poids du $\overrightarrow{VLI}$, vient renforcer cette condition.
- La variation relative d'éclairement due au gradient doit être inférieure à un seuil fixé par l'utilisateur : $\frac{\overrightarrow{AB}.\overrightarrow{\Delta E}}{E_A} \leq S_E$, ce qui contrôle l'effet de lissage introduit par l'interpolation, en réduisant la zone de validité lorsque l'éclairement est perturbé.
- La courbure de la surface doit être inférieure à un seuil prédéfini : $\cos(\vec{N_A}, \vec{N_B}) \geq S_{courb}$, $\vec{N_A}$ et $\vec{N_B}$ étant les normales en A et B. Ce seuil est nécessaire car seule la moitié de la scène située en avant de la surface intervient dans le calcul du $\overrightarrow{VLI}$. L'ignorance du comportement de l'autre moitié de la scène est une source d'erreur importante, car elle peut contenir des objets très lumineux.

4.2.4.3 Interpolation indirecte

Le nombre n de $\overrightarrow{VLI}$ admis pour l'interpolation doit être compris entre deux bornes prédéfinies, $nmin$ et $nmax$. Le choix de ces bornes influence le niveau de filtrage induit par l'interpolation. En pratique, $nmin = 3$ et $nmax = 16$ donnent de bons résultats. Ces faibles valeurs permettent une interpolation rapide.

Tout comme lors de l'interpolation du terme direct (section 4.2.1.2.2), la dispersion en luminance lumineuse entre les n candidats ne doit pas dépasser un seuil prédéterminé. En pratique, un seuil de 20 % à 40 % sur l'écart-type donne de bons résultats dans ce cas indirect.

Lorsque ce seuil est franchi, l'ensemble est rejeté, et un nouveau $\overrightarrow{VLI}$ est calculé explicitement, puis mémorisé.

La fonction de pondération donne moins d'importance aux échantillons risquant d'introduire la plus grande erreur, en fonction des critères définis plus haut. Un poids p_i est associé à chaque $\overrightarrow{VLI}$, de valeur 1 lorsqu'un critère est parfaitement rempli, de valeur plus faible sinon :

$$p_i = k_{valid,i} \cdot \left(1 - \frac{\|\overrightarrow{AB}\|}{dmoy_i}\right)^2 \cdot \frac{1}{1 + \frac{\overrightarrow{AB}.\overrightarrow{\Delta E}}{E_A}} \cdot \frac{\cos(\vec{N_A}, \vec{N_B}) - S_{courb}}{1 - S_{courb}} \tag{4.14}$$

L'interpolation pondérée est une interpolation classique, utilisant les poids calculés avec l'équation (4.14) :

$$\overrightarrow{VLI}(B) = \frac{\sum_{i=1}^{n} p_i \ \overrightarrow{VLI}(A_i)}{\sum_{i=1}^{n} p_i}$$

L'équation 4.5 permet alors de calculer très simplement la composante indirecte de la luminance réfléchie dans la direction $\vec{\omega_r}$.

4.3 Récapitulatif

Dans ce chapitre, nous avons défini la notion de *vecteur lumineux*, qui représente la lumière incidente en un point d'une surface. Cette entité contient les informations minimales nécessaires pour calculer la luminance réfléchie dans une direction donnée. Par définition, il n'introduit pas d'erreur lorsqu'il est utilisé là où il a été calculé, pour la même direction de réflexion. Dans un voisinage du point de calcul, et pour une direction réfléchie différente, il permet de déterminer une approximation de la luminance réfléchie. Des conditions de validité, basées principalement sur les cohérences de la lumière et de la scène, permettent de définir un domaine de validité autour de chaque vecteur lumineux.

Lorsqu'un vecteur lumineux n'est pas connu en un point, une valeur approchée peut être avantageusement estimée par une interpolation pondérée de vecteurs lumineux déjà calculés à proximité. Tous les candidats à l'interpolation doivent satisfaire des critères de validité pour le voisinage du point de calcul courant. Ce procédé réduit les temps de calcul, car l'interpolation est en général beaucoup plus rapide que le calcul explicite de l'énergie incidente. Il introduit cependant une erreur, contrôlée par les calculs de validité.

La lumière est divisée en trois composantes indépendantes, suivant le type de chemin lumineux parcouru pour atteindre un point de la scène : direct, indirect, et caustique. Trois types distincts de vecteurs lumineux sont alors définis, les $\overrightarrow{VLD}$, $\overrightarrow{VLI}$ et $\overrightarrow{VLC}$. Des critères de validité et des procédures d'interpolation[2] spécifiques sont adaptées à ces trois composantes de la lumière.

2. Dans le cas des caustiques, il s'agit d'une estimation de densité.

Chapitre 5

Les sources lumineuses en lancer de rayons

5.1 Objectifs

La simulation des sources lumineuses est un aspect primordial de la recherche d'un plus grand réalisme des images de synthèse. La simulation la plus précise possible de leurs propriétés géométriques (forme, position) et énergétiques (distribution spectrale) est nécessaire.

Le lancer de rayons réalise un échantillonnage ponctuel, car chaque rayon englobe un angle solide infiniment petit. De ce fait, il n'est pas adapté naturellement à la prise en compte de sources étendues. Ce problème est résolu classiquement par une discrétisation de la source, en zones suffisamment petites pour être considérées comme homogènes.

Nous présentons dans ce chapitre les principaux types de sources utilisées en synthèse d'images, ainsi qu'une méthode de discrétisation des sources rectangulaires. Ce modèle est ensuite étendu à des sources planes de forme quelconque, puis est modifié pour simuler l'éclairement du ciel dans son ensemble.

5.2 Sources ponctuelles

Les sources ponctuelles sont les plus simples à modéliser, ce qui leur a valu un grand succès en synthèse d'images, surtout en lancer de rayons. Elles sont aussi une des principales causes de l'aspect artificiel des images. En effet, elles ne créent pas de zones de pénombre. Une source ponctuelle est soit visible, soit masquée. Un seul rayon étant nécessaire pour le déterminer, leur prise en compte est très rapide.

Les sources ponctuelles n'existent pas dans la nature, car tout objet lumineux occupe nécessairement un volume non nul. Le soleil, par exemple, même s'il est vu de très loin, a un rayon apparent non négligeable, et génère donc des zones de pénombre, petites mais visibles. Approximativement, son rayon mesure $700\ 10^6\ m$, et il se situe à une distance de $150\ 10^9\ m$ en moyenne de la terre. De ce fait, il est vu sous un angle solide d'environ $68\ 10^{-6}\ sr$. Le seul cas où l'approximation par une source ponctuelle serait acceptable est celui des étoiles vues depuis la terre, mais il est peu probable qu'elles apportent une énergie significative par rapport aux autres sources de lumière.

Comme elle n'a pas de surface, il n'est pas possible de définir physiquement la luminance d'une source ponctuelle. Par contre, une luminance équivalente doit être évaluée, pour l'intégrer dans l'équation de rendu (équation 3.1).

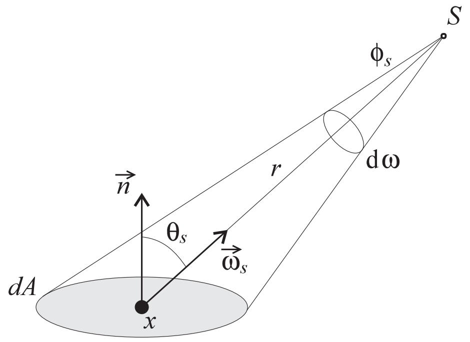


FIG. 5.1 – *Eclairement par une source ponctuelle*

Par définition (section 2.2.2), la luminance à pour valeur :

$$L(x,\vec{\omega}) = \frac{d^2\phi(x,\vec{\omega})}{dA\,\cos\theta\,d\omega}$$

En terme de flux énergétique, l'équation de rendu s'exprime :

$$\begin{aligned} L_r(x,\vec{\omega_r}) &= \int_{\Omega_i} f_r(x,\vec{\omega_i}\rightarrow\vec{\omega_r})\,\frac{d^2\phi_i(x,\vec{\omega_i})}{dA\,\cos\theta_i\,d\omega_i}\,\cos\theta_i\,d\omega_i \\ &= \int_{\Omega_i} f_r(x,\vec{\omega_i}\rightarrow\vec{\omega_r})\,\frac{d^2\phi_i(x,\vec{\omega_i})}{dA} \end{aligned}$$

Or, l'énergie provient uniquement de la direction $\vec{\omega_s}$, dans laquelle se situe la source S. L'ensemble des directions incidentes étant réduit à $\{\vec{\omega_s}\}$:

$$L_r(x,\vec{\omega_r}) = f_r(x,\vec{\omega_s}\rightarrow\vec{\omega_r})\,\frac{d\phi_i(x,\vec{\omega_s})}{dA}$$

Par définition, l'angle solide élémentaire a pour valeur : $d\omega_s = \dfrac{dA\cos\theta_s}{r^2}$, d'où l'on déduit : $dA = \dfrac{d\omega_s r^2}{\cos\theta_s}$. Donc :

$$L_r(x,\vec{\omega_r}) = f_r(x,\vec{\omega_s}\rightarrow\vec{\omega_r})\,\frac{d\phi_i(x,\vec{\omega_s})}{d\omega_s}\,\frac{\cos\theta_s}{r^2}$$

Le terme $\dfrac{d\phi_i(x,\vec{\omega_s})}{d\omega_s}$ est l'intensité $I(S,\vec{\omega_s})$ émise par la source ponctuelle dans la direction $\vec{\omega_s}$. La luminance réfléchie dans la direction $\vec{\omega_r}$ due à la source ponctuelle S a donc pour valeur :

$$L_r(x,\vec{\omega_r}) = f_r(x,\vec{\omega_s}\rightarrow\vec{\omega_r})\,I(S,\vec{\omega_s})\,\frac{\cos\theta_s}{r^2} \tag{5.1}$$

5.3 Sources sphériques

Les sources sphériques sont une modélisation très approximative des ampoules électriques ou du soleil. Comme elles ont un volume défini, elles créent des zones de pénombre, et améliorent ainsi le réalisme des images.

Les sources sphériques uniformes, par convention, émettent la même luminance L dans toutes les directions, pour tous les points de leur surface. Leur contribution en un point de la scène est simple à déterminer. Il suffit de multiplier L par l'angle solide Ω sous lequel elles sont vues (équation D.7), et par un facteur d'atténuation t, exprimant la fraction de surface masquée par d'éventuels obstacles. En lancer de rayons, il est évalué par un échantillonnage de la sphère.

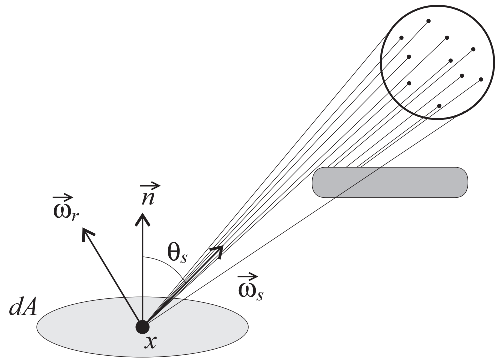


Fig. 5.2 – *Echantillonnage ponctuel d'une source sphérique*

Dans notre implémentation, chaque source sphérique est modélisée par une sphère lumineuse, renvoyant la luminance L lorsqu'elle est vue directement, et à l'intérieur de laquelle sont générés aléatoirement N points (figure 5.2). Géométriquement, la source sphérique peut alors être vue comme un amas de N sources ponctuelles, mais, énergétiquement, il s'agit bien d'une sphère d'émission uniforme qui sous-tend un angle solide mesurable. Le calcul de t consiste alors à lancer N rayons vers ces N points, afin de déterminer les N_v points visibles. On a donc simplement $t = \frac{N_v}{N}$. L'angle solide associé à chaque rayon, pour le calcul de l'équation de rendu, a pour valeur $\frac{\Omega}{N}$. Lorsqu'un $\overrightarrow{VLD}$ est calculé, sa direction $\vec{D}$ est la direction moyenne de ces N_v points visibles.

Cette méthode est équivalente à un échantillonnage volumique de la sphère. D'autres auteurs proposent un échantillonnage de sa surface [SWZ96], ou bien de la surface du disque apparent visible [Mai96].

Le nombre total de rayons N peut être déterminé en fonction de l'angle solide sous lequel est vue la source. Il est inutile de beaucoup discrétiser une source très éloignée. Ce critère est lié à la taille des petits objets que l'on accepte de ne pas détecter lors des calculs de pénombre.

L'échantillonnage produit de l'aliassage (annexe C), visible sous la forme d'un zonage des pénombres. Pour diminuer ses effets visuels, en le remplaçant par du bruit, la position aléatoire des N points est modifiée à chaque calcul. Il suffit de trois tirages aléatoires par point : deux pour choisir une direction à partir du centre de la sphère, et un pour la position dans cette direction. Cette opération ne pénalise que faiblement la durée des calculs, mais l'amélioration visuelle des résultats est très sensible. Il est ainsi possible de réduire le taux d'échantillonnage sans altérer la qualité perceptuelle de l'image.

Les exemples de la figure 5.3, page 97, représentent la même scène, comportant une chaise éclairée par une seule source sphérique, pour divers taux d'échantillonnage de la source. Dans ce cas, on constate que 32 rayons par pixel (h) perturbés produisent un résultat visuellement indiscernable de l'image de référence obtenue avec 128 rayons par pixel (b). Sans la perturbation, le zonage est toujours visible pour le même niveau de discrétisation (g). Avec 16 rayons perturbés par pixels (f), le bruit est trop important dans les zones de pénombre larges, alors qu'il est insignifiant pour les zones de pénombre étroites. Quant à la source ponctuelle (a), il est évident qu'elle ne produit pas un contour réaliste de la zone d'ombre.

5.4 Sources surfaciques planes uniformes

Les sources surfaciques sont rarement utilisées en lancer de rayons, car leur prise en compte est très coûteuse. Les sources surfaciques peuvent être seulement partiellement obstruées, ce qui a l'avantage de créer des zones de pénombre qui augmentent le réalisme des images. En contrepartie, elle demandent un échantillonnage fin pour détecter les obstructions éventuelles. Plusieurs méthodes d'échantillonnage adaptatif ont été proposées [CPC84, KJ92, War94, Col95, SWZ96], pour réduire le nombre total de subdivisions. Elles se basent sur une décomposition plus ou moins régulière de la surface de la source, en général en quadtree (section 5.4.3.1), et tentent de ne raffiner que les zones de la surface créant des pénombres. Nous proposons une méthode proche, mais dont les critères de subdivision concernent la pyramide engendrée par la source, et l'angle solide qu'elle sous-tend. Ce critère volumique ne tient pas uniquement compte de la surface de la source, mais aussi du point depuis lequel elle est vue. Il permet de réduire encore le nombre global de rayons, pour une précision équivalente aux autres solutions.

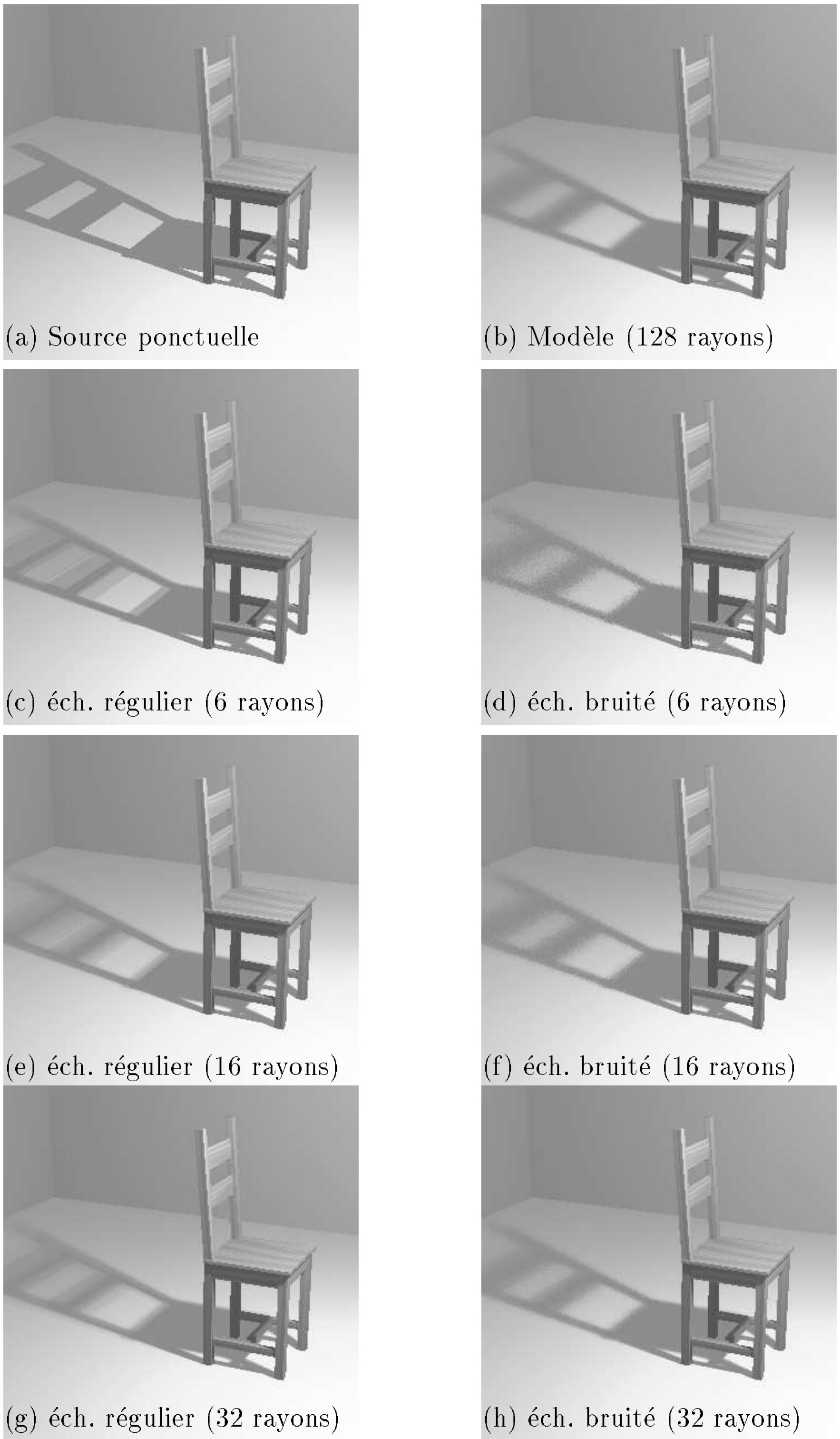
5.4.1 Principe

Dans le cas des sources planes polygonales uniformes, Arvo [Arv94] présente une définition analytique du vecteur d'éclairement (section 4.1.2). Soit une source plane d'émission constante M $[W.m^{-2}]$, dont la géométrie est un polygone de N sommets $\{p_1, \ldots, p_N\}$ (figure 5.4). Lorsqu'elle est complètement visible, elle crée en un point x quelconque de l'espace un vecteur d'éclairement $\overrightarrow{\Phi(x)}$, qui vaut :

$$\overrightarrow{\Phi(x)} = \frac{M}{2\pi} \sum_{i=1}^{N} \Theta_i(x) \overrightarrow{\Gamma_i(x)} \tag{5.2}$$

avec :

$$\begin{cases} \Theta_i(x) = \cos^{-1}\left(\dfrac{\vec{u_i}}{\| \vec{u_i} \|} \cdot \dfrac{\vec{u_{i+1}}}{\| \vec{u_{i+1}} \|} \right) \\ \overrightarrow{\Gamma_i(x)} = \dfrac{\vec{u_i} \wedge \vec{u_{i+1}}}{\| \vec{u_i} \wedge \vec{u_{i+1}} \|} \end{cases}$$



(a) Source ponctuelle
(b) Modèle (128 rayons)
(c) éch. régulier (6 rayons)
(d) éch. bruité (6 rayons)
(e) éch. régulier (16 rayons)
(f) éch. bruité (16 rayons)
(g) éch. régulier (32 rayons)
(h) éch. bruité (32 rayons)

FIG. 5.3 – *Divers taux d'échantillonnage d'une source sphérique*

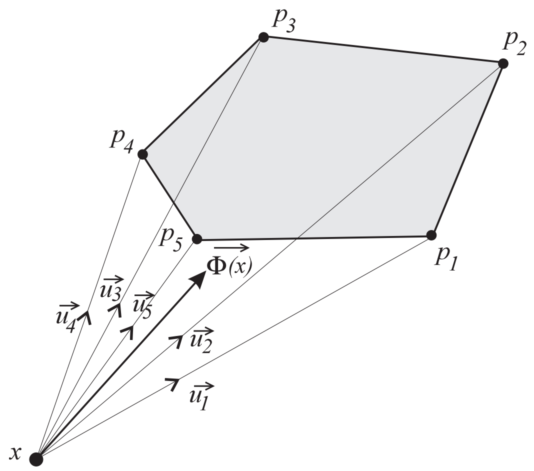


FIG. 5.4 – *Vecteur d'éclairement créé par une source polygonale*

L'émission de la source étant uniforme et lambertienne, la luminance est constante pour tous les points de la surface et pour toutes les directions. Les points émettent tous selon un hémisphère orienté vers l'extérieur de la surface, d'angle solide $2\pi\, sr$. La luminance a donc pour valeur :

$$L = \frac{M}{2\pi} \quad [W.m^{-2}.sr^{-1}]$$

L'angle solide ω_S, sous-tendu par la source, se calcule avec la formule (équation D.9) :

$$\omega_s = \sum_{i=1}^{N} \cos^{-1}\left(\frac{(\vec{u_i} \wedge \overrightarrow{u_{i+1}}) \cdot (\vec{u_i} \wedge \overrightarrow{u_{i-1}})}{\| \vec{u_i} \wedge \overrightarrow{u_{i+1}}\| \ \| \vec{u_i} \wedge \overrightarrow{u_{i-1}}\|}\right) - (N-2)\,\pi$$

Les termes intervenant dans la définition de l'angle solide se retrouvent naturellement dans celle du vecteur d'éclairement. Le calcul de l'un d'eux est donc très rapide, dès lors que l'on calcule l'autre.

Nous supposons maintenant que le point x appartienne à une surface, de normale $\vec{n}$ formant un angle θ_S avec le vecteur d'éclairement $\overrightarrow{\Phi(x)}$. Alors, le vecteur lumineux direct en x engendré par cette source se déduit du calcul du vecteur d'éclairement.

- La **direction** $\vec{D}$ du vecteur lumineux direct est égale à la direction du vecteur d'éclairement :

$$\vec{D} = \frac{\overrightarrow{\Phi(x)}}{\|\overrightarrow{\Phi(x)}\|}$$

- Le **terme énergétique** P est obtenu en calculant la luminance réfléchie avec l'équation de reflectance (équation 2.6) :

$$L_r(x, \vec{\omega_r}) = f_r(x, \vec{D} \to \vec{\omega_r})\, L \cos\theta_S\, \omega_S$$

Afin de vérifier l'équation 4.4, la valeur de P est :

$$P = L \cos\theta_S\, \omega_S \tag{5.3}$$

5.4.2 Subdivision adaptative

La formulation précédente n'est valable que sous deux conditions :

- l'angle solide ω_S est suffisamment petit pour admettre que la BRDF soit constante sur l'ensemble des directions engendrées ;
- la source n'est pas partiellement occultée.

Lorsque ces deux critères ne sont pas tous les deux satisfaits, la source est subdivisée récursivement en éléments plus petits. Leurs contributions sont ensuite cumulées pour calculer un $\overrightarrow{VLD}$, selon la méthode décrite à la section 4.2.1. La subdivision a lieu en deux phases, selon un critère géométrique d'abord, puis ensuite selon un critère de visibilité. Il s'agit d'une méthode adaptative classique, adaptée aux calculs de $\overrightarrow{VLD}$.

- **Phase géométrique.** Tant que la source est vue sous un angle solide "trop grand", elle est subdivisée en cellules jointives de plus petite taille. L'opération est réitérée avec chaque cellule.

  Le seuil d'acceptabilité est fixé arbitrairement dans notre implémentation. Il dépend d'une part de la présence de pics significatifs dans la BRDF, et d'autre part de la taille maximale des objets que l'on tolère de ne pas toujours détecter lors des calculs de pénombres.

- **Phase de visibilité.** En lancer de rayons, la visibilité d'une cellule est estimée en générant plusieurs rayons d'ombre depuis le point x, afin de détecter d'éventuelles obstructions. On distingue alors quatre cas :
  - Lorsque la cellule est entièrement visible, son apport énergétique est calculé avec l'équation 5.3.
  - Lorsqu'elle est entièrement masquée, sa contribution est nulle.
  - Lorsqu'elle est partiellement visible, et lorsque la profondeur maximale de récursion n'est pas atteinte, la cellule est divisée en éléments de plus petites tailles, auxquels sont appliqués récursivement la même procédure.
  - Lorsque la profondeur maximale de récursion est atteinte, la contribution de la cellule est estimée en calculant le taux de rayons d'ombre parvenant à la cellule par rapport à leur nombre total.

Nous proposons ci-après une méthode de subdivision optimisée des sources rectangulaires. Nous verrons ensuite qu'elle peut s'adapter relativement facilement à des sources planes de forme quelconque.

5.4.3 Sources rectangulaires uniformes

5.4.3.1 Bintree et quadtree

La méthode proposée ici est basée sur une subdivision hiérarchique adaptative en *2d-arbre* de la surface de la source [Ben75, Sam84]. Cette structure de données est souvent appelée *bintree*.

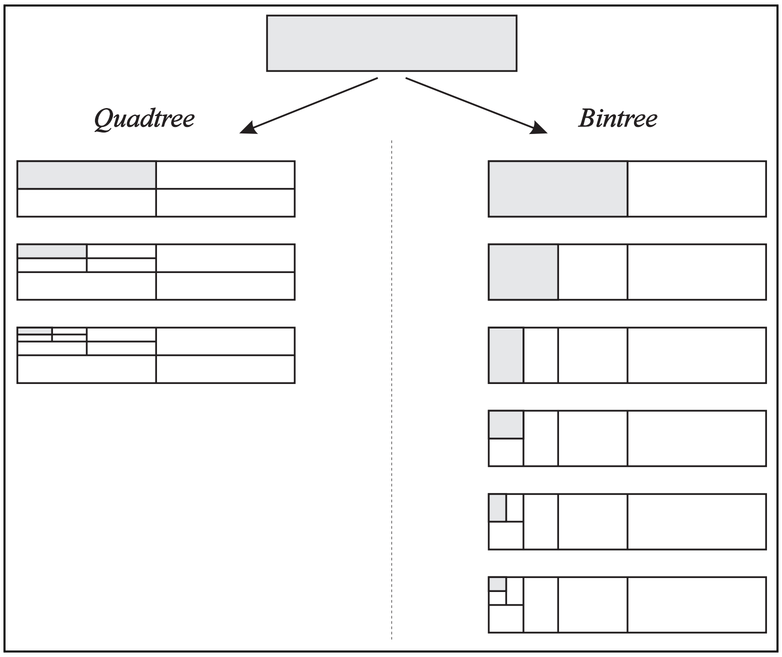


FIG. 5.5 – *Subdivision en quadtree et en bintree d'un rectangle*

Il s'agit de subdiviser récursivement la surface de la source, en la découpant à chaque étape en deux zones de même importance, selon l'axe de plus forte variance (cette notion dépend des critères que l'on désire raffiner). Avec cette méthode, le nombre total de pavés est plus faible qu'avec l'habituelle subdivision en *quadtree*, qui divise systématiquement en quatre parties, et l'on évite de générer des zones trop allongées, même lorsque le rectangle initial est lui-même très allongé.

L'avantage est manifeste sur la figure 5.5, où un rectangle est subdivisé de façon à détacher une petite zone dans le coin supérieur gauche. Le découpage en bintree, selon le plus grand côté du rectangle, crée des zones beaucoup plus équilibrées (plus proches d'un carré dans ce cas), pour une même surface finale des cellules. Le nombre d'étapes semble plus important, mais le nombre de cellules intermédiaires est plus faible (7 au lieu de 10), et donc le nombre total de points et segments calculés est réduit lui aussi.

Ces résultats sont à mettre en parallèle avec des travaux de géométrie algorithmique [LM97], où une subdivision hiérarchique binaire est utilisée pour améliorer des techniques de recherche dans un plan.

5.4.3.2 Subdivision en bintree des sources rectangulaires

Le critère de subdivision d'une cellule, lors de la phase géométrique, est la valeur de l'angle solide qu'elle sous-tend, vue depuis le point x.

Dans l'espace à trois dimensions, lorsqu'un rectangle est vu depuis un point x quelconque, il apparaît sous la forme d'un quadrilatère irrégulier, à cause des déformations de projection. De ce fait, un découpage par le milieu des segments n'est pas optimal. Il crée bien deux zones de surfaces égales sur le rectangle, mais elles n'engendrent pas le même angle solide.

Ce problème est d'autant plus crucial que x s'éloigne de la normale au rectangle passant par son centre de gravité.

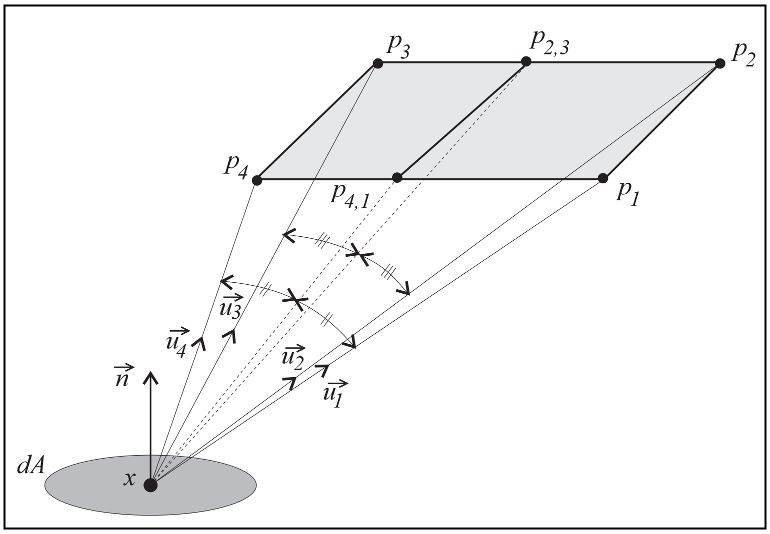


FIG. 5.6 – *Echantillonnage en bintree d'une source rectangulaire*

Notre méthode consiste à ne pas diviser le rectangle lui-même, mais le cône à base polygonale de sommet x qu'il définit (figure 5.6). Ce critère est similaire à la notion d'angle solide, et il génère donc des zones mieux équilibrées de ce point de vue. Par ce procédé, les bases des pyramides engendrées par les cellules ont une allure se rapprochant de carrés, après projection sur le plan perpendiculaire à l'axe du cône.

- La direction du découpage, à chaque étape, est celle pour laquelle la projection du segment correspondant apparaît la plus grande. C'est aussi celle de la face de la pyramide dont l'angle d'ouverture est maximal, donc de cosinus minimal. Ainsi, le segment $[p_i, p_{i+1}]$ sera divisé si :

  $$\cos(p_i, p_{i+1}) \;=\; \min\left(\cos(\vec{u_1}, \vec{u_2}),\, \cos(\vec{u_2}, \vec{u_3}),\, \cos(\vec{u_3}, \vec{u_4}),\, \cos(\vec{u_4}, \vec{u_1})\right)$$

  Dans l'exemple de la figure 5.6, c'est le segment $[p_4, p_1]$ qui est divisé, en calculant le point de séparation $p_{4,1}$.

  L'autre segment à subdiviser est celui situé “en face”, c'est à dire le segment $[p_2, p_3]$ de notre exemple.

- La subdivision du segment $[p_i, p_{i+1}]$, pour déterminer la position du point de subdivision $p_{i,i+1}$, a lieu selon la bissectrice de l'angle $(\vec{u_i}, \vec{u_{i+1}})$:

  $$\overrightarrow{u_{i,i+1}} \;=\; \frac{\vec{u_i} + \vec{u_{i+1}}}{\|\, \vec{u_i} + \vec{u_{i+1}} \,\|}$$

  Il se forme ainsi deux zones de tailles comparables, dans l'espace de la pyramide.

- La position du point sur le segment est légèrement perturbée, pour limiter les effets de l'aliassage.

Cette phase de la subdivision a lieu uniquement selon des critères géométriques. Elle est relativement rapide, car aucune intersection de rayon avec la scène n'est calculée.

5.4.3.3 Calculs de visibilité

Dans cette étape, toutes les cellules créées par la subdivision géométrique sont visitées, afin de déterminer leur contribution énergétique. Pour tester la visibilité d'une cellule, cinq rayons d'ombre sont générés : un vers chacun des quatre sommets, et un vers un point aléatoire proche de l'axe de la pyramide sous-tendue par la cellule. En pratique, ce dernier point est choisi à proximité de la bissectrice de la diagonale semblant la plus grande, vue depuis le point de calcul. Trois cas sont alors possibles :

- **Les cinq points sont visibles :** la cellule est considérée comme étant complètement visible, et sa contribution est calculée avec l'équation 5.3.
- **Les cinq points sont masqués :** la contribution énergétique de la cellule est nulle.
- **Les cinq points ne sont pas tous visibles :** lorsque N_s sommets sont visibles, avec $0 < N_s < 5$, alors :
  - **Si** le niveau maximal de récursivité est atteint, la contribution de la cellule est calculée avec l'équation 5.3, puis est pondérée par le ratio $\frac{N_s}{5}$.
  - **Sinon** la cellule est subdivisée en quatre pavés adjacents, qui ont le point central déjà déterminé comme sommet commun. Les autres sommets sont obtenus en calculant les bissectrices des angles sous-tendus par les arêtes de la cellule actuelle. Le processus est réitéré récursivement pour chacun de ces quatre pavés.

    Comme la première étape est une division en bintree des cellules (section 5.4.3.1), on suppose pour cette deuxième phase qu'elles sont suffisamment équilibrées pour appliquer une subdivision en quadtree. Ceci permet de réduire le nombre total de rayons d'ombre, grâce au plus grand nombre de sommets communs. A chaque niveau, les informations de visibilité calculées au niveau précédent sont réutilisées pour les sommets communs. Le point central étant proche de l'axe de la pyramide, les quatre zones ont une importance comparable, car elles définissent des angles solides proches.

    Comme lors de l'étape de subdivision géométrique, une légère pertubation de la position des sommets sur les arêtes atténue les effets de l'aliassage.

Cette opération de calcul de visibilité est la plus coûteuse, car il faut calculer les intersections des rayons d'ombre avec la scène. La méthode proposée ici limite le nombre total de rayons, en cherchant à tout moment à créer des zones homogènes, dans un repère lié au point de vue, et non pas à la source. La subdivision est remise à jour pour chaque point de calcul, puisque le point de vue varie. Dans l'ensemble, la subdivision est fine seulement aux endroits où cela est nécessaire.

5.4.4 Exemples de subdivisions adaptatives

La figure 5.7 représente une chaise éclairée par une source surfacique rectangulaire placée au dessus-d'elle, légèrement en avant, qui génère des pénombres complexes sous les barreaux.

FIG. 5.7 – *Echantillonnage ponctuel (196 rayons par pixel, durée 9 min)*

FIG. 5.8 – *Echantillonnage adaptatif (durée 5 min)*

Afin de servir de référence, elle est calculée avec un échantillonnage constant de la source (196 rayons par pixel). Pour une taille d'image de 400x400 pixels, avec un germe initial de 5000 $\overrightarrow{DLV}$ (section 4.2.1), et sur une machine équipée d'un processeur MIPS R10000 à 200 Mhz, le temps de calcul est d'environ 9 minutes.

La figure 5.8 représente la même scène, calculée avec notre méthode d'échantillonnage adaptatif de la source, en environ 5 minutes. Pour la phase de subdivision géométrique, la taille maximale d'angle solide est de $5\ 10^{-3}\ sr$, et pour les calculs de visibilité la profondeur de récursivité est limitée à 2 niveaux. Visuellement, les images 5.7 et 5.8 sont indistinguables. La distance moyenne pixel à pixel est d'environ $0,3$ dans l'espace de couleur Luv (section 2.8.4.3), pour lequel le seuil de perception est supérieur à 1.

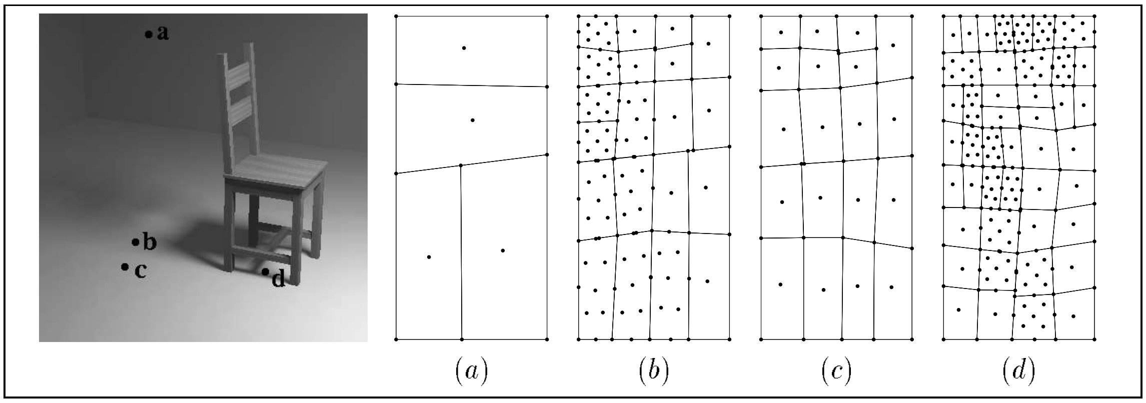


FIG. 5.9 – *Subdivision adaptative d'une source rectangulaire*

Sur la figure 5.9, la même chaise est vue sous un angle différent, où les détails de la pénombre se distinguent mieux. Les subdivisions de la source rectangulaire sont représentées sous une forme schématique, pour quatre points a, b, c, et d de la scène, sur les figures 5.9(a), 5.9(b), 5.9(c), et 5.9(d). Les traits symbolisent les segments créés lors de la phase de subdivision géométrique, et les points correspondent aux rayons effectivement lancés.

- Le point a est relativement éloigné de la source, et presque dans son plan tangent. La source entière est donc vue sous un angle solide petit. De plus il n'y a pas d'occlusion. La subdivision grossière en quatre pavés est suffisante.
- Le point b est situé dans la pénombre du montant vertical du dossier de la chaise.
- Le point c est proche du point b, mais hors de la pénombre. Le découpage géométrique est presque le même, mais le nombre de rayons est très différent. Dans le cas b la subdivision est plus forte sur la partie gauche de la source, qui est partiellement masquée par le montant du dossier.
- Le point d est situé sous la chaise, et il est en partie masqué par plusieurs barreaux, qui produisent une subdivision complexe.

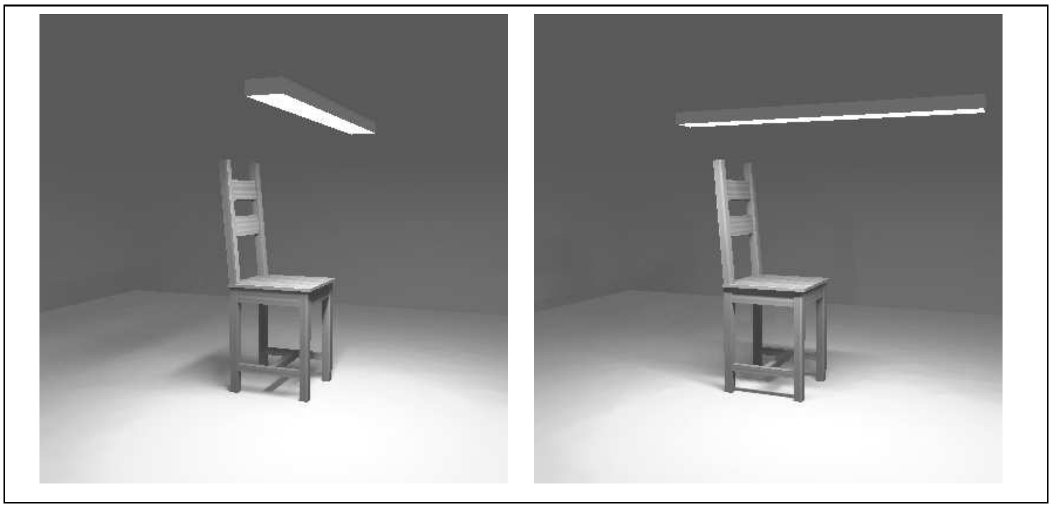


FIG. 5.10 – *Influence de l'orientation d'une source allongée*

La figure 5.10 montre l'influence de l'orientation d'une source allongée sur l'allure de la pénombre. La longueur de la source est dix fois plus grande que sa largeur, et la rotation a lieu autour du centre de gravité. L'apparente différence de taille de la source entre les deux images est due à la distortion de la projection. Cet exemple simule approximativement l'éclairement d'un luminaire contenant un tube fluorescent.

5.4.5 Sources surfaciques uniformes non rectangulaires

5.4.5.1 Principe

Les sources planes non rectangulaires sont prises en compte en les incluant dans une source rectangulaire englobante. Les zones n'appartenant pas à l'intersection des deux sont ignorées lors des calculs d'éclairement. C'est le principe de la peinture au pochoir, avec un masque reproduisant en négatif la forme de la source.

La phase de subdivision géométrique, dont le seul critère est la valeur de l'angle solide, n'est pas modifiée. L'enveloppe rectangulaire sert de guide aux divisions, car les sources surfaciques de forme quelconque n'induisent pas en général un schéma de découpage naturel.

Par contre, un critère supplémentaire est introduit lors des calculs de visibilité : le point est-il à l'intérieur ou à l'extérieur de la source ? Tout point situé à l'extérieur est considéré comme masqué, comme s'il subissait l'obstruction d'un objet de la scène, ce qui influence la décision de subdiviser ou non la cellule. Ce critère uniquement géométrique est testé sans générer de rayons supplémentaires.

5.4.5.2 Exemple de simulation d'une source circulaire

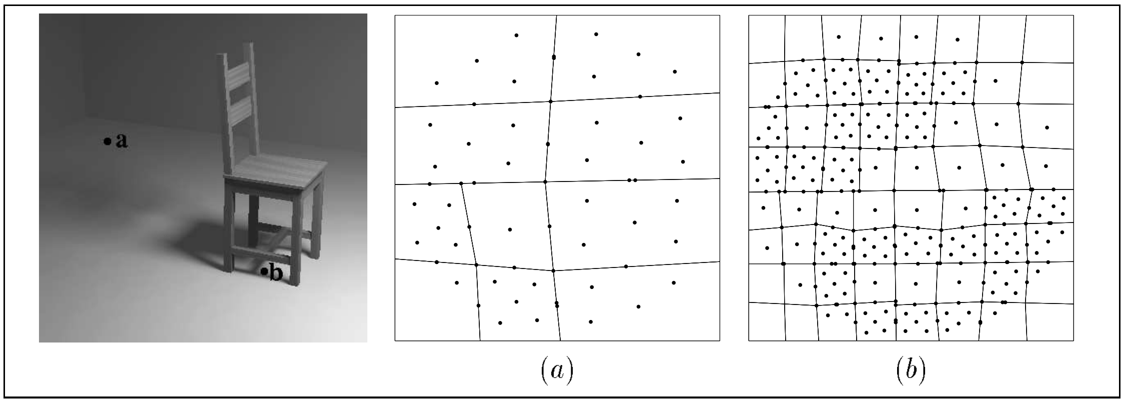


FIG. 5.11 – *Subdivision adaptative d'une source circulaire*

Simuler avec cette méthode une source circulaire est très simple. Le critère d'appartenance est la distance entre le point et le centre de la source, qui doit être inférieure au rayon de cette dernière. Des exemples de subdivisions sont reproduits sur la figure 5.11, selon le même principe que pour la figure 5.9. Les rayons ne sont explicitement lancés que pour les points situés à l'intérieur du disque, avec une densité plus forte dans les zones les plus influentes. Le point a, relativement éloigné de la source et ne subissant aucune obstruction, produit une décomposition grossière. Le point b, partiellement occulté par les barreaux de la chaise, induit une subdivision beaucoup plus complexe, avec un plus grand raffinement des frontières entre les zones masquées et visibles.

La figure 5.12 représente toujours la même chaise que dans la section précédente, mais cette fois elle est éclairée par un source circulaire, de même puissance et de même aire que la source rectangulaire de la figure 5.13. Cette dernière figure est identique à la figure 5.8. Elle est reproduite ici pour faciliter la comparaison.

Des différences sont perceptibles dans l'allure de la pénombre, mais elles sont très faibles, pratiquement insignifiantes. Elles deviennent significatives uniquement lorsque la source est très proche des objets éclairés. Ce phénomène est étudié par Ashdown dans [Ash93]. Il montre que la géométrie d'une source n'a pas vraiment d'influence sur l'éclairement qu'elle produit au delà d'une certaine distance, qui dépend de la forme et des dimensions de la source. Il

FIG. 5.12 – *Source circulaire*

FIG. 5.13 – *Source rectangulaire*

utilise ces résultats pour fixer les limites de validité de la simulation des sources de formes complexes par des agencements de sources ponctuelles.

5.5 Sources surfaciques non uniformes

Dans cette section, nous présentons brièvement une extension de la méthode de prise en compte des sources surfaciques planes uniformes à des sources plus complexes. Il s'agit de modéliser des sources non uniformes, pour s'approcher des caractéristiques mesurées de luminaires réels, ou non planes pour simuler les effets de la lumière naturelle. Ces méthodes sont actuellement en cours d'implémentation.

5.5.1 Principe

La méthode de subdivision adaptative présentée ci-dessus peut être étendue aux sources surfaciques non uniformes. Un critère supplémentaire vient alors piloter la phase de subdivision, basé sur l'homogénéité énergétique des zones créées.

La luminance émise par un point de la surface de la source peut varier avec sa position, et aussi avec la direction éclairée. Il se pose alors un sérieux problème de représentation des données, puisque l'on doit pouvoir déterminer en chaque point la luminance émise dans chaque direction. Une solution possible consiste à utiliser, pour représenter la répartition énergétique, des méthodes de représentation de BRDF mesurées. En effet, il s'agit dans ces deux cas de stocker des données qui dépendent à la fois d'une position et d'une direction. Des représentations simplifiées par harmoniques sphériques [SAWG91] ou par un modèle à lobes [LFTG97] sont envisageables, mais elles supposent tout de même la gestion d'un très grand nombre de paramètres.

5.5.2 Lumière naturelle

5.5.2.1 Eclairement naturel

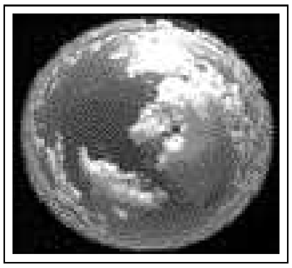


FIG. 5.14 – *Photographie d'un ciel peu nuageux avec un objectif "œil de poisson"*

Les sources traditionnelles ne permettent pas une simulation correcte de l'éclairement naturel. En effet, la modélisation du soleil par une source sphérique est d'une précision suffisante, mais la lumière parvenant au sol est fortement perturbée par sa traversée de l'atmosphère, même en l'absence de brumes ou de nuages. Ainsi, suite aux phénomènes de diffusion, une quantité importante de lumière provient de directions autres que celle où est situé le soleil. De plus, la distribution n'est pas uniforme, et varie fortement avec les conditions météorologiques.

Dans le cadre de l'architecture, et plus particulièrement de l'éclairagisme [Mit96], il est important de modéliser correctement l'éclairement provenant du ciel. Les simulations de scènes extérieures, ou de scènes intérieures éclairées à travers une fenêtre, nécessitent des modèles de ciel. On distingue des modèles analytiques, empiriques, et des modèles obtenus à partir de mesures physiques. Ces derniers sont plus difficiles à mettre en œuvre, car ils engendrent un très grand nombre de données numériques.

5.5.2.2 Modèles analytiques de ciel

La CIE[1] fournit quelques formules théoriques pour évaluer la luminance émise depuis un point P de la voûte céleste vers un point x du sol (figure 5.15). Ces modèles empiriques ne sont pas aptes à simuler l'entière diversité des phénomènes naturels. Ils reproduisent approximativement la variation de la luminance en fonction de la position sur le ciel, assimilé à un hémisphère de rayon infini.

Nous donnons ces résultats dans deux cas extrêmes diurnes : un ciel uniformément couvert, et un ciel clair où le soleil est visible.

Ciel couvert Le modèle analytique de ciel couvert, dit de Moon et Spencer, a été normalisé par la CIE en 1955. La luminance dépend uniquement de l'angle zénithal θ, selon la formule très simple :

$$L(\theta) = L_z \frac{1 + 2\cos\theta}{3}$$

où

$$L_z = 90 + 9630 \sin h_0 \qquad W.m^{-2}.sr^{-1}$$

1. Commission Internationale de l'Eclairage

est la luminance du zénith, h_0 étant la distance angulaire du soleil.

Ce modèle décrit assez fidèlement un ciel diurne, avec une couverture nuageuse très dense. La luminosité est maximale au zénith, et elle décroît progressivement en se rapprochant de l'horizon.

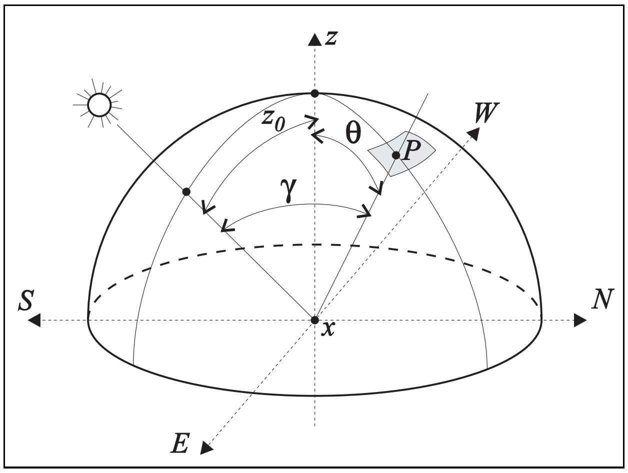


FIG. 5.15 – *Géométrie pour l'éclairement du ciel*

Ciel clair Le modèle analytique de ciel clair est plus complexe que celui de ciel couvert, car la position du soleil intervient directement. Il a été normalisé par la CIE en 1973. La luminance incidente émise depuis le point P varie avec l'angle zénithal θ, l'écart angulaire γ entre P et la position du soleil (figure 5.15), et l'angle zénithal du soleil z_0 :

$$L(\theta, \gamma, z_0) \;=\; L_z \,\frac{(0.91 + 10e^{-3\gamma} + 0.45\cos^2\gamma)(1 - e^{\frac{-0.32}{\cos\theta}})}{0.274\,(0.91 + 10e^{-3z_0} + 0.45\cos^2 z_0)}$$

Une valeur empirique de la luminance au zénith est donnée par :

$$L_z \;=\; 100 + 63h_0 + h_0(h_0 - 30)e^{0.0346(h_0-68)} \qquad W.m^{-2}.sr^{-1}$$

où h_0 est la distance angulaire du soleil, exprimée en degrés.

Energie solaire L'éclairement direct produit par le soleil sur une surface horizontale est empiriquement donné par :

$$E_{soleil} \;=\; 1000\,(1 - e^{-0.055h_0})\,\sin h_0 \qquad W.m^{-2}$$

où h_0 est la distance angulaire du soleil, exprimée en degrés.

5.5.2.3 Modèles mesurés de ciel

Les modèles analytiques vus précédemment sont trop simples pour simuler toutes les configurations possibles de ciel. Par exemple, ils ne simulent pas les ciels partiellement nuageux ou brumeux laissant traverser directement une partie des rayons solaires (figure 5.14). De plus, ils ne tiennent pas compte des conditions météorologiques particulières au lieu considéré. Pour obtenir une bonne adéquation entre les simulations et les réalisations, il est nécessaire de disposer de données réelles, en divers lieux géographiques, à diverses heures de la journée, et sous différentes conditions climatiques représentatives.

Le Laboratoire des Sciences de l'Habitat (LASH), de l'Ecole Nationale des Travaux Publics de l'Etat (ENTPE) de Vaulx-en-Velin, réalise depuis plusieurs années des mesures régulières de l'éclairement naturel, en divers sites mondiaux. Il dispose ainsi d'une base de données obtenues dans des conditions réelles [FLM96], décrivant quantitativement et qualitativement l'énergie provenant des différentes régions du ciel. Notre méthode de calcul d'images sera prochainement adaptée à la prise en compte de ces données physiques.

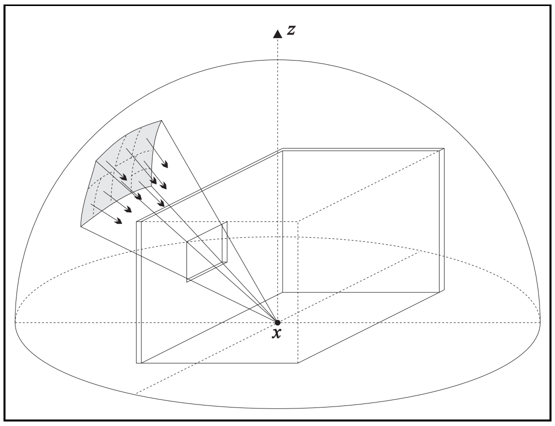


FIG. 5.16 – *Eclairement du ciel à travers une fenêtre*

5.5.2.4 Subdivision adaptative de la voûte céleste

Une fenêtre éclairant l'intérieur d'une pièce peut être simulée par une source plane rectangulaire (figure 5.16). La méthode de subdivision adaptative développée ci-dessus permet de la découper en éléments d'angles solides approximativement identiques. La luminance qu'ils émettent est obtenue en les projetant sur la voûte céleste, puis en faisant appel à un modèle de ciel, analytique ou mesuré. La variation de luminance sur leur surface est alors un critère de subdivision supplémentaire, jusqu'à atteindre un seuil prédéterminé d'uniformité. Les éventuelles obstructions par les autres objets sont détectées lors de la dernière phase, en lançant des rayons d'ombre vers les éléments de surface.

Dans le cas d'une scène extérieure, l'ensemble du ciel doit être pris en compte. En nous inspirant de la technique de l'*hémicube* [CG85], utilisée en radiosité pour déterminer les facteurs de forme, et de Tadamura [TNK$^+$93], nous proposons d'englober le point courant dans une boîte composée de cinq fenêtres rectangulaires adjacentes (figure 5.17). Elles n'interviennent pas directement dans le résultat global, mais elles servent de support initial lors de la phase de subdivision adaptative. La voûte céleste est décomposée en cinq zones indépendantes, qui sont subdivisées adaptativement comme les fenêtres simples. Leurs contributions sont finalement cumulées pour obtenir l'éclairement total du ciel.

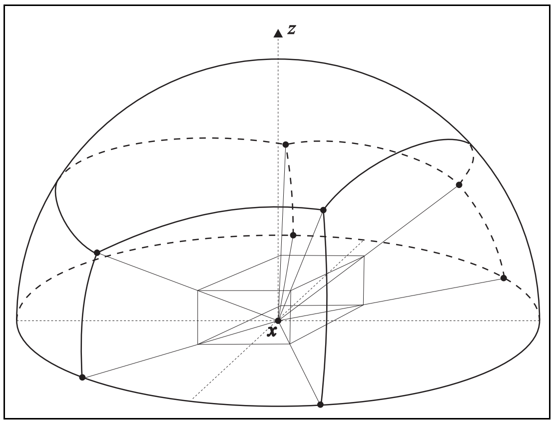


FIG. 5.17 – *Subdivision initiale du ciel complet*

Chapitre 6

Implémentation et résultats

6.1 Implémentation

6.1.1 Lancer de rayons YART

Le logiciel de lancer de rayons *YART (Yet Another Ray-Tracer)* a été développé par Marc Roelens [Roe93] à l'Ecole des Mines de Saint-Etienne. Il est écrit en langage C, pour l'efficacité, mais sous forme de modules indépendants dont la philosophie est proche de celle des langages orientés objets. *YART* est utilisé en base de développement et de recherche pour le laboratoire, aussi bien dans le domaine du rendu que de la géométrie. La modularité totale permet d'implémenter et valider les nouvelles méthodes de façon complètement transparente pour le reste du programme.

Le module de rendu, qui manipule les énergies, a servi de support à nos développements. Il est complètement séparé du module géométrique, qui contient la description de la scène et calcule les intersections avec les rayons. De même, les représentations des sources de lumière, des matériaux et des textures sont, elles aussi, implémentées sous forme de modules indépendants. Le respect de conventions prédéfinies permet les communications et les échanges de résultats entre les modules.

6.1.2 Occupation de mémoire

Les structures de données utilisées ont la taille suivante :

- $\overrightarrow{VLI} = 336$ octets ;
- $\overrightarrow{VLD} = 96$ octets ;
- $\overrightarrow{VLC} = 64$ octets.

Chaque vecteur étant encapsulé dans un nœud de kd-arbre, il faut ajouter 40 octets aux valeurs précédentes. De plus, ils comportent un pointeur vers une donnée de type énergie, formée de 3 nombres réels (24 octets) avec un modèle de couleurs RVB.

Par exemple, l'occupation totale de mémoire due aux vecteurs lumineux est de l'ordre de 10 Mo pour l'image 6.13.

6.1.3 Représentation des scènes par graphe CSG

Cette modélisation décompose les scènes en un ensemble d'objets élémentaires simples, déformés et assemblés selon les critères désirés, le tout mémorisé avec une structure d'arbre. Un arbre CSG (*Constructive Solid Geometry*) est une arborescence classique (figure 6.1) dont :

- La racine symbolise la scène à reproduire.
- Les feuilles représentent des objets élémentaires, qui sont déformés puis réunis pour former la scène complète. Ces primitives sont souvent des corps géométriques simples (cube, sphère, cylindre, tore, cône, ...), mais elles peuvent aussi être plus complexes, des surfaces de Bézier ou des surfaces implicites par exemple.
- Les nœuds peuvent être unaires ou binaires :
  - Un nœud unaire représente une transformations appliquée à toute sa descendance (homothétie, affinité, rotation, translation, ...).
  - Un nœud binaire représente une opération de combinaison logique entre ses deux fils (union, intersection, soustraction d'un objet à l'autre ...).

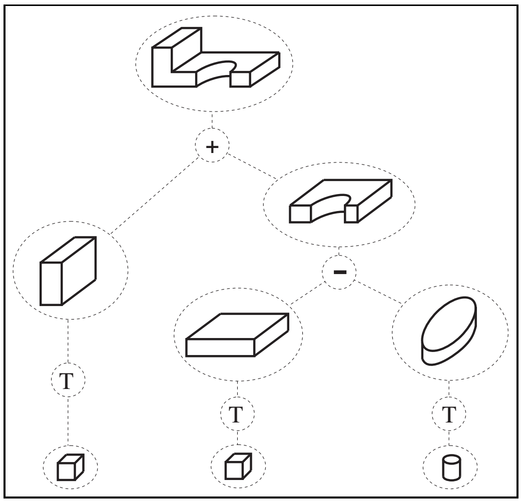


FIG. 6.1 – *Modélisation par arbre CSG*

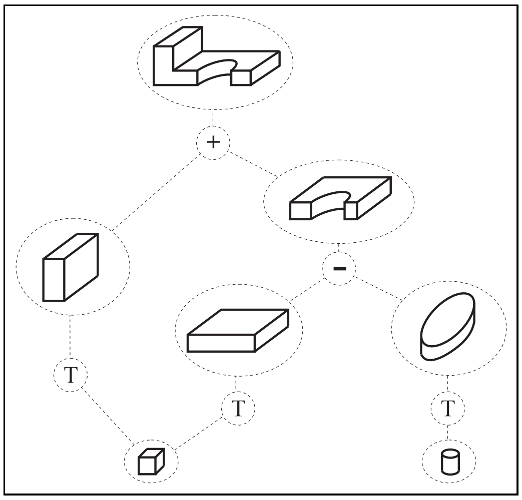


FIG. 6.2 – *Modélisation par graphe CSG*

La mise en œuvre de cette modélisation est plus complexe que celle des ensembles de facettes, mais la représentation en mémoire est compacte et les objets décrits ne souffrent d'aucune ambiguïté géométrique. De plus, cette méthode possède une meilleure tolérance envers les imprécisions numériques. Après de multiples transformations, un objet risque d'être très légèrement plus petit ou plus grand que ce qu'il devrait être, mais il sera toujours un objet parfaitement défini. Il est en pratique fort improbable qu'un tel défaut causé par les seules imprécisions numériques soit perceptible.

Des améliorations ont été apportées par Marc Roelens [Roe93] pour transformer cette structure arborescente en graphe planaire sans circuits. La mise en place de chemins transversaux entre les nœuds permet d'éviter l'inutile duplication de sous-objets identiques. Par exemple, la figure 6.2 représente sous forme d'un graphe le même objet que la figure 6.1. Les deux cubes élémentaires sont alors réunis en un seul et même objet, qui est partagé par deux sous-graphes distincts. De la même façon, des objets plus élaborés peuvent être partagés. Ainsi, si plusieurs chaises identiques sont représentées dans une scène, une seule description de chaise existe en mémoire, avec des chemins d'accés différents qui indiquent les positions, et éventuellement des transformations. Le gain en occupation de mémoire est très important pour les scènes composées de nombreux objets.

6.1.4 Description des scènes

Le langage *CASTOR*, développé à l'Ecole des Mines de Saint-Etienne [Bei88, Roe93], est un langage de description de scènes adapté à la représentation par arbre CSG. Il permet la définition géométrique d'objets quelconques, en leur associant des caractéristiques physiques (couleur, texture, indice de réfraction, ...), ainsi que des propriétés plus générales de la scène (position du point de vue, paramètres de projection, ...). Ce langage gère les inclusions multiples de fichiers. La réutilisation d'objets déjà créés pour d'autres images est aisée.

Des paramètres supplémentaires peuvent être associés aux objets. Il est ainsi relativement simple d'intégrer de nouveaux modules de rendu nécessitant ces paramètres. Le module gérant les informations géométriques n'a pas à être modifié, car il sait ne prendre en compte que les données le concernant, tout en laissant les autres données accessibles depuis les autres modules.

Des modules de développement (Bibliothèques *castorC*) permettent de définir des objets directement en langage *C*. Chacun d'eux est défini comme une fonction qui représente une partie de l'arbre CSG de la scène complète. Des modules spécifiques gèrent les diverses opérations applicables aux objets. Il est possible de créer des objets avec des paramètres variables, ce qui est tout particulièrement utile pour réaliser des animations. Il suffit de décrire les trajectoires et les modifications des objets au cours du temps. La génération des fichiers en langage *CASTOR* correspondant aux différentes images est ensuite complètement automatique.

6.2 Quelques résultats

Nous présentons dans cette section des images réalisées en utilisant les vecteurs lumineux, ainsi que des comparaisons avec d'autres méthodes. Les temps de calcul sont donnés à titre indicatif. Comme les images sont calculées sur la même machine, une station de travail Silicon Graphics Indigo 2 équipée d'un processeur MIPS R10000 cadencé à 200 Mhz, et avec la même implémentation logicielle, les durées d'exécution sont approximativement proportionnelles à la complexité des algorithmes. Elles ne sont pas des preuves irréfutables de validité, surtout sur des images de test qui ne sont que des cas particuliers, mais elles constituent néanmoins un élément d'appréciation fort utile.

6.2.1 Comparaison de méthodes

Une scène a été calculée avec des algorithmes de rendu différents, pour l'estimation du terme indirect (figure 6.6, page 123). Elle se compose d'une table, de sphères en acier, en or, et en matière plastique, et d'une unique source ponctuelle. Les matériaux sont simulés avec le modèle de Schlick (section 2.5.6). Une série de sphères est posée sur la table, où elles reçoivent directement la lumière de la source, alors qu'une autre série est placée sous la table. Ces dernières sphères sont éclairées seulement par l'éclairement indirect.

Une image de référence de taille 400x400 a été obtenue en environ 3 heures, avec une méthode de Monte Carlo brutale (128 rayons par hémisphère), pour l'estimation de la composante indirecte.

Elle est comparée (figure 6.6, page 123) à des images obtenues par trois autres méthodes de calcul du terme indirect : un terme ambiant constant, notre méthode des vecteurs lumineux, et une méthode ne calculant que la composante indirecte diffuse pure, similaire à la méthode de Ward (section 3.6.3).

La colonne de droite contient les images de différence dans l'espace de couleurs Luv. Les zones claires indiquent les plus grands écarts. Pour chaque cas, nous indiquons la différence moyenne pixel à pixel e et l'écart-type correspondant σ.

Il apparaît clairement que la méthode du terme ambiant constant introduit des différences notables, mais sa grande rapidité permet des prévisualisations lors de la conception des scènes. La méthode des vecteurs lumineux est la plus proche de l'image de référence, avec une erreur moyenne assez modérée, et une vitesse assez rapide (18 fois plus rapide que le calcul complet). Ne pas tenir compte des aspects directionnels du l'éclairement indirect génère de plus grandes erreurs, surtout pour les objets dont la composante spéculaire est importante.

6.2.2 Influence des paramètres (cas direct)

Une série d'images de la même scène, une chaise éclairée par une source plane, a été réalisée, avec des paramètres d'interpolation différents, uniquement pour la composante directe. Les images ne sont pas antialiassées, afin de ne pas fausser les comparaisons en ajoutant des variations autres que celles dues à l'interpolation des $\overrightarrow{VLD}$. Pour la même raison, un terme ambiant constant simule les éclairements indirects.

La source est de taille relativement grande (1x0.5 m^2), et elle est plaçée à une faible hauteur (1.5 m), légèrement en avant de la chaise afin d'avoir une pénombre complexe visible sous les barreaux de la chaise. L'angle solide maximal est de $0,004\,sr$, ce qui génère en moyenne 85 rayons par calcul d'éclairement direct, avec notre méthode de subdivision adaptative.

Une image de référence a été calculée sans interpolation, en une durée de 12' 39", pour une taille de 256x256=65536 pixels.

Le tableau 6.3 récapitule les résultats numériques de cette série de tests. La colonne "Germe" indique le nombre initial de $\overrightarrow{VLD}$ calculés en des point aléatoires de la scène, avant le parcours complet des pixels. R est le rayon de la zone de recherche des $\overrightarrow{VLD}$ valides pour l'interpolation, et V est le seuil de rejet sur la variance des candidats. La distance moyenne par rapport à l'image de référence, dans l'espace perceptuel Luv, est indiquée, ainsi que l'écart-type correspondant. Ces données sont commentées plus loin.

L'image de référence et les images caractéristiques des lignes (a) à (d) (figure 6.4, page 116) permettent de visualiser l'influence du seuil sur la variance tolérée. La colonne centrale contient les images de différence dans l'espace de couleurs Luv. Les zones claires indiquent les plus grands écarts. Les images de colonne de droite représentent en noir les pixels pour lesquels les $\overrightarrow{VLD}$ ont effectivement été calculés. Elles confirment que le choix de ce seuil est important pour la détection des zones de pénombre.

	Germe	R	V	Durée	Nombre de $\overrightarrow{VLD}$	Distance Luv	Ecart-type
	1000	0,05	10 %	3' 26"	11942	0,020	0,001
	1000	0,05	20 %	2' 56"	10780	0.021	0.002
	1000	0,05	50 %	2' 23"	9551	0.025	0.005
	1000	0,05	80 %	2' 02"	8771	0.029	0.011
	1000	0,1	10 %	3' 58"	10869	0.030	0.002
	1000	0,1	20 %	2' 48"	7517	0.033	0.002
	1000	0,1	50 %	2' 02"	5569	0.041	0.010
	1000	0,1	80 %	1' 16"	3918	0.055	0.036
(a)	1000	0,2	10 %	5' 52"	18744	0.035	0.003
(b)	1000	0,2	20 %	3' 35"	9090	0.042	0.004
(c)	1000	0,2	50 %	2' 05"	4863	0.063	0.056
(d)	1000	0,2	80 %	1' 08"	2790	0.079	0.086
	2000	0,05	10 %	3' 31"	11989	0,020	0,001
	2000	0,05	20 %	3' 01"	10809	0.022	0.002
	2000	0,05	50 %	2' 26"	9569	0.025	0.004
	2000	0,05	80 %	2' 09"	8937	0.028	0.008
	2000	0,1	10 %	3' 56"	10795	0.026	0.001
	2000	0,1	20 %	2' 59"	7858	0.028	0.002
	2000	0,1	50 %	2' 11"	6055	0.037	0.015
	2000	0,1	80 %	1' 26"	4552	0.051	0.037
	3000	0,03	10 %	4' 37"	21141	0.015	0.001
	3000	0,03	20 %	4' 16"	20373	0.016	0.001
	3000	0,05	10 %	3' 31"	12059	0.020	0.001
	3000	0,05	20 %	3' 03"	10930	0.022	0.002
	3000	0,1	10 %	4' 04"	11229	0.025	0.001
	3000	0,1	20 %	3' 02"	8301	0.027	0.001
	3000	0,2	10 %	4' 58"	13080	0.032	0.002
	3000	0,2	20 %	3' 31"	8479	0.035	0.003

FIG. 6.3 – *Tableau récapitulatif de l'influence des paramètres (cas direct)*

Les valeurs numériques des distances par rapport à l'image de référence sont très faibles, bien en deçà du seuil de sensibilité dans l'espace Luv, qui est de l'ordre de 1. En fait, il s'agit d'une distance moyenne, pixel à pixel. Pour notre méthode, les zones délicates de l'image sont les zones de pénombre, dont l'aire est relativement petite. La distance moyenne est donc peu affectée par les erreurs locales. Cette donnée n'est donc pas suffisante pour comparer deux images. L'écart-type est un élément supplémentaire de jugement à ne pas négliger. Sur cet exemple, des artefacts visibles apparaissent lorsque la valeur de l'écart-type devient du même ordre que celle de la distance moyenne.

Il nous manque une réelle notion de distance entre images, ce qui est toujours un sujet de recherche ouvert [Rou98].

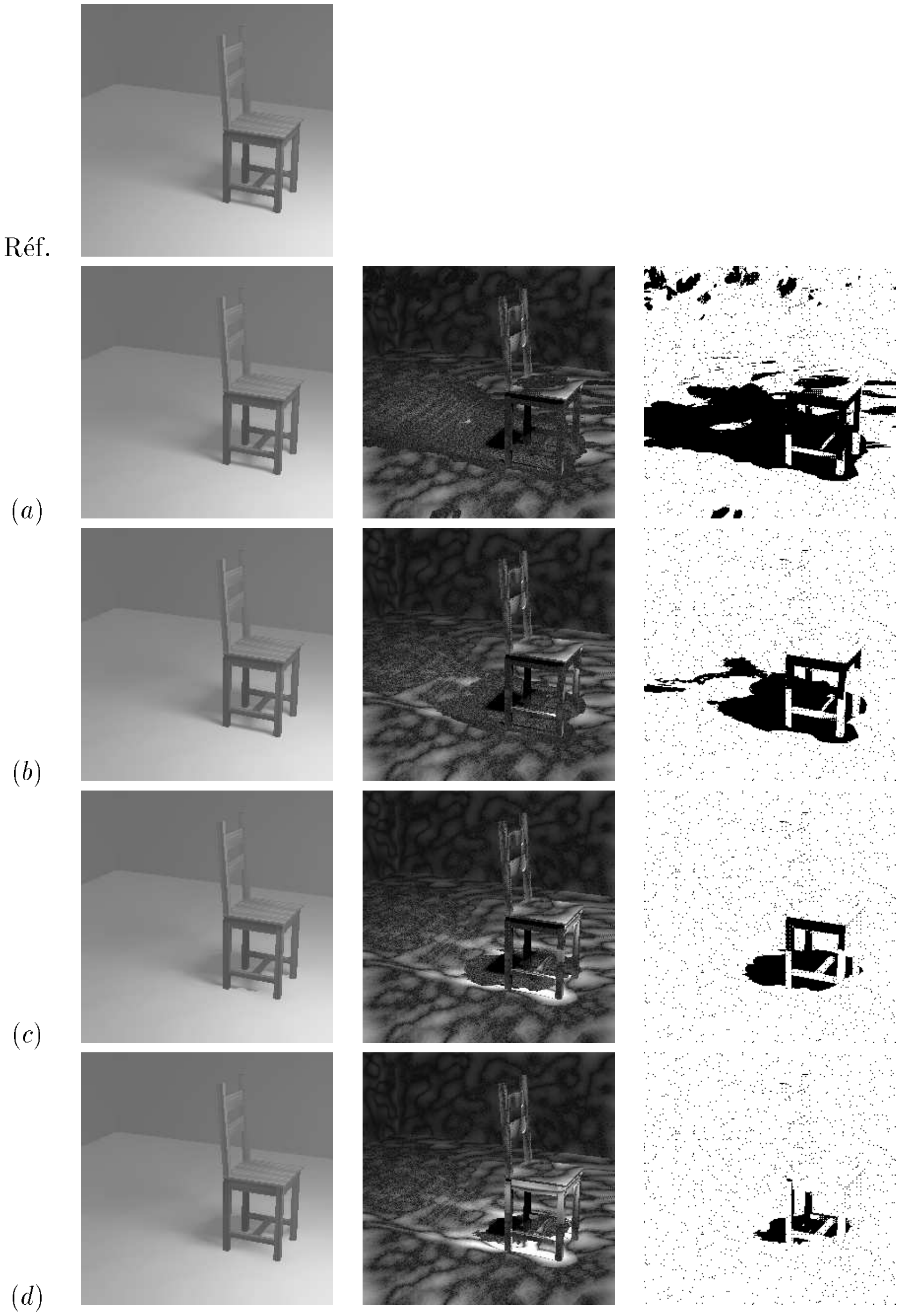


Fig. 6.4 – *Influence des paramètres (cas direct)*

Ces résultats ne sont pas d'une interprétation aisée, car le choix des paramètres optimaux obéit en fait à des contraintes contradictoires. Si la zone de recherche est trop petite, il arrivera souvent que l'on ne trouve pas assez de $\overrightarrow{VLD}$ pour interpoler. Par contre, si elle est trop grande, la variance maximale des échantillons sera souvent dépassée. Les paramètres sont liés, et leur choix dépend fortement de la scène. Actuellement, ils sont fixés empiriquement et arbitrairement. L'idéal serait qu'ils puissent varier adaptativement, en fonction de la complexité de la scène et de la précision désirée.

6.2.3 Influence des paramètres (cas indirect)

La scène de test (figure 6.7, page 125) représente une pièce complètement fermée. Elle contient une chaise en bois, et une sphère de verre légèrement teinté, soutenue à quelques centimètres du sol par un support métallique. L'énergie est fournie uniquement par une source plane orientée vers le plafond, maintenue par un bras articulé fixé au mur. Elle n'est pas centrée dans la pièce, afin d'éviter une symétrie des pénombres. Tous les matériaux sont diffus, à l'exception de la sphère et de son support.

Seuls le plafond et une petite frange supérieure des murs sont éclairés directement. Les autres objets ne reçoivent que l'éclairement indirect, provenant principalement du plafond, qui se comporte comme une source surfacique non uniforme de grande taille. Il apparaît alors plusieurs phénomènes relativement difficiles à reproduire correctement :

- Le support de la source crée une zone sombre fortement marquée vers le pied, s'élargissant et s'atténuant vers le bas.
- La chaise et ses barreaux génèrent une pénombre de forme complexe.
- La sphère crée une *caustique indirecte*, en concentrant la lumière provenant de chemins lumineux de type $LDS(D|S)^{*}DE$. Il s'agit d'une image inversée et déformée du plafond, où l'on distingue une tache rectangulaire correspondant à la face sombre de la source, dans la partie inférieure droite de la zone claire.
- Les murs, et en partie le sol, sont de grandes surfaces planes, où l'éclairement indirect engendre un dégradé très régulier.

Il s'agit d'un cas particulier, qui n'a donc pas valeur de démonstration. Par contre, il regroupe les principales difficultés auxquelles notre méthode est sensible. Comme les points délicats sont tous dans la région de la scène ne recevant que l'éclairement indirect, les défauts dus à l'interpolation sont très visibles. Nous nous servirons des résultats pour faciliter un choix raisonnable des paramètres de calcul, en fonction des résultats désirés, tout en regrettant de ne pas pouvoir fournir une méthode plus générale.

L'image de référence a été obtenue après 9 h 19 min de calculs, en évaluant en chaque point le terme indirect par la méthode de Monte Carlo complète, sans interpolation. Chaque hémisphère est échantillonné avec 512 rayons, ce qui engendre un faible niveau de bruit. Afin de ne pas perturber les comparaisons, l'image n'est pas antialiassée.

$N_{MtCarlo}$	Germe	n_{min}	n_{max}	S_E	V_E	NbTotal	Durée	e_{Luv}	σ_{Luv}	Aspect
complet							9 h 19'	0	0	4 (a)
ambiant							1' 10"	1.2	0.97	0 (b)
128	100	1	10	0.8	0.8	877	1' 59"	0.49	0.20	0 (c)
128	100	2	10	0.8	0.8	1541	2' 28"	0.41	0.18	0
128	400	1	10	0.8	0.8	1077	2' 16"	0.40	0.13	0
128	400	2	10	0.8	0.8	1661	2' 44"	0.39	0.13	1 (d)
128	400	3	10	0.8	0.8	2226	3' 15"	0.38	0.13	1
128	400	3	16	0.8	0.8	2229	3' 16"	0.38	0.13	1
128	400	3	16	0.8	0.4	2709	4' 20"	0.37	0.10	1
128	3600	3	16	0.8	0.4	5059	6' 48"	0.26	0.09	2
128	3600	3	16	0.8	0.8	4390	5' 39"	0.26	0.10	2
128	3600	3	16	0.8	0.2	12920	22' 38"	0.26	0.10	1
128	6400	3	16	0.8	0.2	16127	24' 25"	0.26	0.10	2
128	6400	3	16	0.8	0.3	8269	8' 32"	0.25	0.09	2
128	6400	3	16	0.8	0.4	7631	7' 06"	0.25	0.09	2
256	1600	3	16	0.1	0.8	3617	8' 17"	0.28	0.15	0
256	1600	3	16	0.4	0.8	3027	7' 06"	0.27	0.14	1
256	1600	3	16	0.8	0.8	2983	6' 58"	0.27	0.14	1
256	1600	3	16	1.6	0.8	2956	6' 53"	0.27	0.14	1
256	1600	3	16	0.8	0.4	3332	8' 10"	0.28	0.12	1
256	3600	3	16	0.1	0.8	5275	11' 42"	0.28	0.14	0
256	3600	3	16	0.4	0.8	4772	10' 34"	0.27	0.11	1
256	3600	3	16	0.8	0.8	4739	10' 32"	0.27	0.11	1
256	3600	3	16	1.6	0.8	4721	10' 23"	0.27	0.11	1
256	3600	3	16	0.8	0.4	5030	11' 33"	0.27	0.10	2 (e)
256	3600	3	16	0.8	0.2	7715	21' 20"	0.27	0.11	4 (f)
256	10000	3	16	0.1	0.8	11223	23' 26"	0.26	0.10	1
256	10000	3	16	0.4	0.8	10838	22' 13"	0.26	0.09	2
256	10000	3	16	0.8	0.8	10811	22' 21"	0.26	0.09	2
256	10000	3	16	1.6	0.8	10791	22' 13"	0.26	0.09	2
256	10000	3	16	0.4	0.4	11073	23' 42"	0.26	0.09	3
256	10000	3	16	0.8	0.4	11019	23' 02"	0.26	0.09	3
256	10000	3	16	0.8	0.1	> 30000	$> 120'$			
256	10000	3	16	0.8	0.2	14289	36' 58"	0.27	0.10	4
512	6400	3	16	0.8	0.4	7606	30' 40"	0.25	0.08	3
512	6400	3	16	0.8	0.2	9925	47' 30"	0.26	0.08	4
512	10000	3	16	0.8	0.4	11099	43' 45"	0.26	0.08	3

FIG. 6.5 – *Tableau récapitulatif de l'influence des paramètres (cas indirect)*

La colonne *Aspect* est une appréciation subjective de la qualité de l'image. Elle vient compléter les mesures d'écart-type et de variance, qui ne sont que des valeurs moyennes peu représentatives des aspects visuels. En effet, un petit nombre de pixels erronnés affecteront peu ces valeurs, alors qu'ils peuvent avoir un impact psycho-visuel très fort. Les images sont arbitrairement regroupées en cinq classes :

Aspect	Description
0	Image mauvaise, peu ressemblante et/ou fortement bruitée
1	Qualité médiocre, avec un zonage marqué et peu de détails dans les pénombres
2	Qualité acceptable, mais un léger zonage est perceptible, et certains détails sont absents
3	Bonne qualité, mais certains détails fins sont absents
4	Image visuellement très proche du modèle

Remarques :

- La distance moyenne pixel à pixel est un bien piètre évaluateur de différences entre images. Les zones posant des problèmes étant très petites, elles n'influencent pas de façon assez significative la valeur moyenne.

- Le gradient n'est pas suffisant pour détecter toutes les zones difficiles, il doit être complété par la condition sur la variance.

- Lorsque les contraintes sont trop rigoureuses, l'effet obtenu est contraire aux attentes. Le mélange des régions interpolées avec les pixels calculés ajoute un bruit important. La distance moyenne reste faible, mais l'écart-type augmente, et les images sont visuellement inacceptables. De plus, les temps de calculs augmentent de façon importante.

A partir d'un taux d'échantillonnage suffisant (256 rayons par hémisphère pour cette image), on remarque qu'il est préférable d'augmenter le nombre de $\overrightarrow{VLI}$ plutôt que d'augmenter le nombre de rayons par hémisphère. La contrainte sur le gradient est peu sélective, alors que celle sur la variance permet de détecter les détails fins, notamment sous la sphère, à l'intérieur de la caustique indirecte. Elle augmente cependant les temps de calcul, puisqu'elle provoque le calcul d'un plus grand nombre de $\overrightarrow{VLI}$.

Quelques images caractéristiques (figure 6.7, page 125) sont présentées, afin de visualiser divers niveaux de qualité. Elle sont repérées par une lettre dans la colonne *Aspect* du tableau 6.5.

- L'image (*a*) est la référence, calculée avec une méthode de Monte Carlo complète.

- L'image (*b*) est calculée avec un terme ambiant constant. Il est évident ici que ce procédé n'est pas du tout adapté à ce type de scène.

- L'image (*c*) est calculée avec des paramètres beaucoup trop tolérants, ce qui provoque un bruit excessif. Les zones d'influence des trop rares $\overrightarrow{VLI}$ apparaissent nettement.

- L'image (*d*) est assez grossière, mais elle indique dans ses grandes lignes la répartition de l'éclairement indirect. Elle est obtenue en moins de 3 minutes.

- L'image (e) semble satisfaisante d'un premier abord, mais certain détails de la pénombre sont absents, principalement sous la chaise et à proximité de la fixation du luminaire, là où l'hypothèse de variation lente de l'éclairement indirect n'est plus vérifiée.
- L'image (f), obtenue en moins de 22 minutes, est presque indiscernable en pratique de l'image de référence, visuellement. Pour cet exemple, le gain de temps est supérieur à 26.

6.2.4 Images plus complexes

Nous présentons dans cette section des images de scènes plus complètes, avec une prise en compte variée de phénomènes.

La figure 6.8, page 127, est une *boite de Cornell*, qui est un modèle célèbre en synthèse d'image. Elle se compose d'une pièce fermée, dont les murs sont de couleurs vives, contenant un cube mat et un cube brillant. Le plafond est éclairé uniquement de façon indirecte, et de ce fait sa couleur est modifiée par celle des murs. La source surfacique crée des pénombres douces. La durée des calculs est de 5 minutes, avec environ 5000 $\overrightarrow{VLD}$ et 2500 $\overrightarrow{VLI}$.

La figure 6.9, page 127, est une autre *boite de Cornell*, contenant en plus une sphère de verre, flottant par magie à quelques centimètres du sol. Elle est éclairée par une source sphérique. Contrairement à l'image précédente, les caustiques sont calculées. Elles se manifestent par le reflet du sommet du cube brillant dans l'angle supérieur gauche, par une tâche claire sous la sphère, visible par réflexion sur le cube brillant, et par une autre tâche en avant de la sphère, due à la lumière réfléchie par la face avant du cube et traversant la sphère. La durée des calculs est de 13 minutes, avec environ 15000 $\overrightarrow{VLD}$, 2800 $\overrightarrow{VLI}$ et 20000 $\overrightarrow{VLC}$. On peut alors remarquer que l'image précédente (figure 6.8) est incomplète, car les réflexions sur le cube brillant ne sont pas négligeables.

La figure 6.10, page 129, représente un anneau cannelé en verre, créant un enchevêtrement complexe de caustiques. La durée des calculs est de 90 minutes, avec environ 30000 $\overrightarrow{VLC}$. Cette durée assez importante est due au très grand nombre de réflexions complètes à l'intérieur de l'anneau et aux réfractions qui génèrent de nombreux rayons secondaires. Elle a été réalisée par Xavier Serpaggi, lors de son stage de DEA à l'Ecole des Mines de Saint-Etienne, en 1997.

La figure 6.11, page 129, est une scène d'intérieur, avec une grande source surfacique qui simule l'éclairement d'une fenêtre. Le paysage extérieur est une image plaquée sur la fenêtre, et il n'intervient pas dans les calculs d'éclairement. Les pieds de la table et des chaises provoquent une grande subdivision de la source, pour obtenir les nombreuses pénombres. La durée des calculs est de 220 minutes, avec environ 25000 $\overrightarrow{VLD}$, et 5000 $\overrightarrow{VLI}$.

La figure 6.12, page 131, est elle-aussi une scène d'intérieur, avec un plus grand nombre d'objets, des textures volumiques variées (bois et marbre), une texture plaquées (image plaquée de la girafe), et combinant des matériaux diffus (murs, sol et plafond) et spéculaires (miroir, bouteilles). La durée des calculs est de 180 minutes, avec environ 15000 $\overrightarrow{VLD}$, et 11000 $\overrightarrow{VLI}$.

La figure 6.13, page 131, est très similaire à la figure précédente, mais avec le calcul des caustiques en plus. L'énergie est fournie par une source surfacique qui crée des pénombres douces. La prise en compte des caustiques occasionne l'éclairement de l'arrière de la statue d'époque indéfinie posée sur la commode, visible par reflet dans le miroir. Elle crée aussi des zones plus lumineuses sous la table, dues elles-aussi à la réflexion de la lumière sur le miroir. De façon moins visible, des petits reflets supplémentaires apparaissent sur les objets en verre. La durée des calculs est de 300 minutes, avec environ 20000 $\overrightarrow{VLD}$, 10000 $\overrightarrow{VLI}$ et 24000 $\overrightarrow{VLC}$.

6.2.5 Vers un plus grand photoréalisme

Dans le cadre d'une collaboration avec le laboratoire LASH de l'Ecole Nationale des Travaux Publics de l'Etat (ENTPE) de Vaulx-en-Velin, une salle de cours a été modélisée, pour comparer des images de synthèse à des photographies (figure 6.14, page 133). L'objectif est d'intégrer, dans un avenir proche, des mesures physiques de l'éclairement naturel du ciel. Les premiers résultats présentés ici sont encourageants, bien que de nombreux défauts persistent.

Dans cette version préliminaire, le ciel est simulé par une source surfacique de grande taille, ce qui fausse la répartitions des ombres. De plus, la modélisation géométrique et les BRDF des matériaux sont fortement simplifiés. La durée des calculs de l'image est actuellement supérieure à 6 heures, mais la simulation par une source surfacique classique n'est pas efficace, car elle induit une subdivision excessive (plus de 300 rayons d'ombres, en moyenne, par calcul d'éclairement direct). La prochaine réalisation sera d'implémenter un modèle de ciel qui soit à la fois réaliste et optimisé.

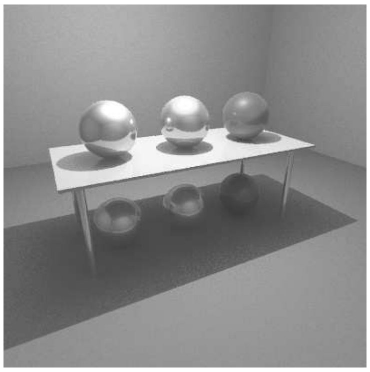
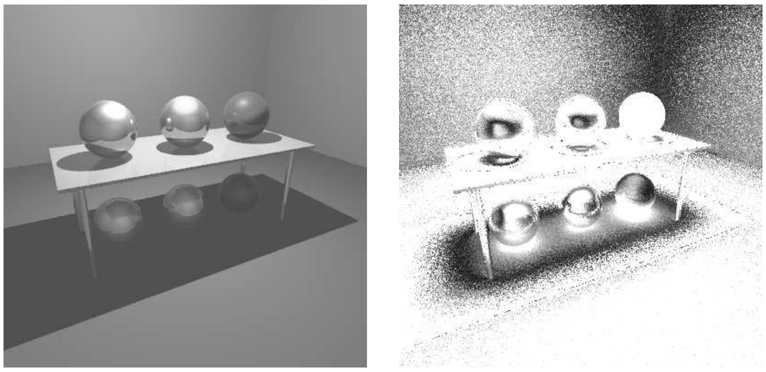
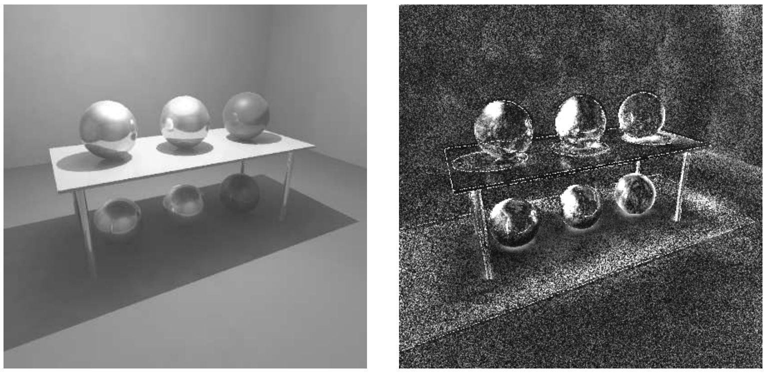


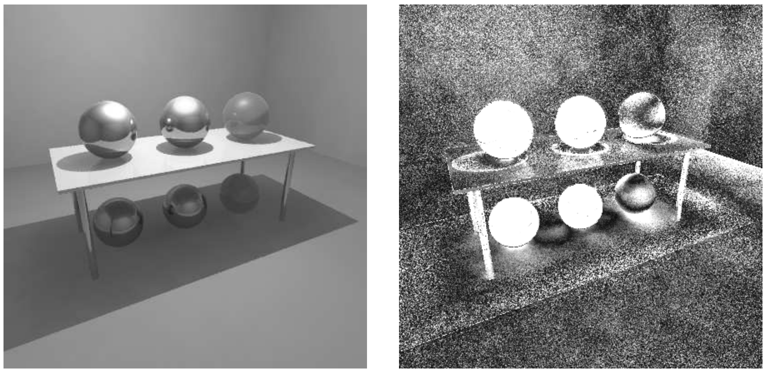
Image de référence
Monte Carlo
(3 heures)



Terme ambiant
(2 minutes)
$e = 1,23 \quad \sigma = 0,93$

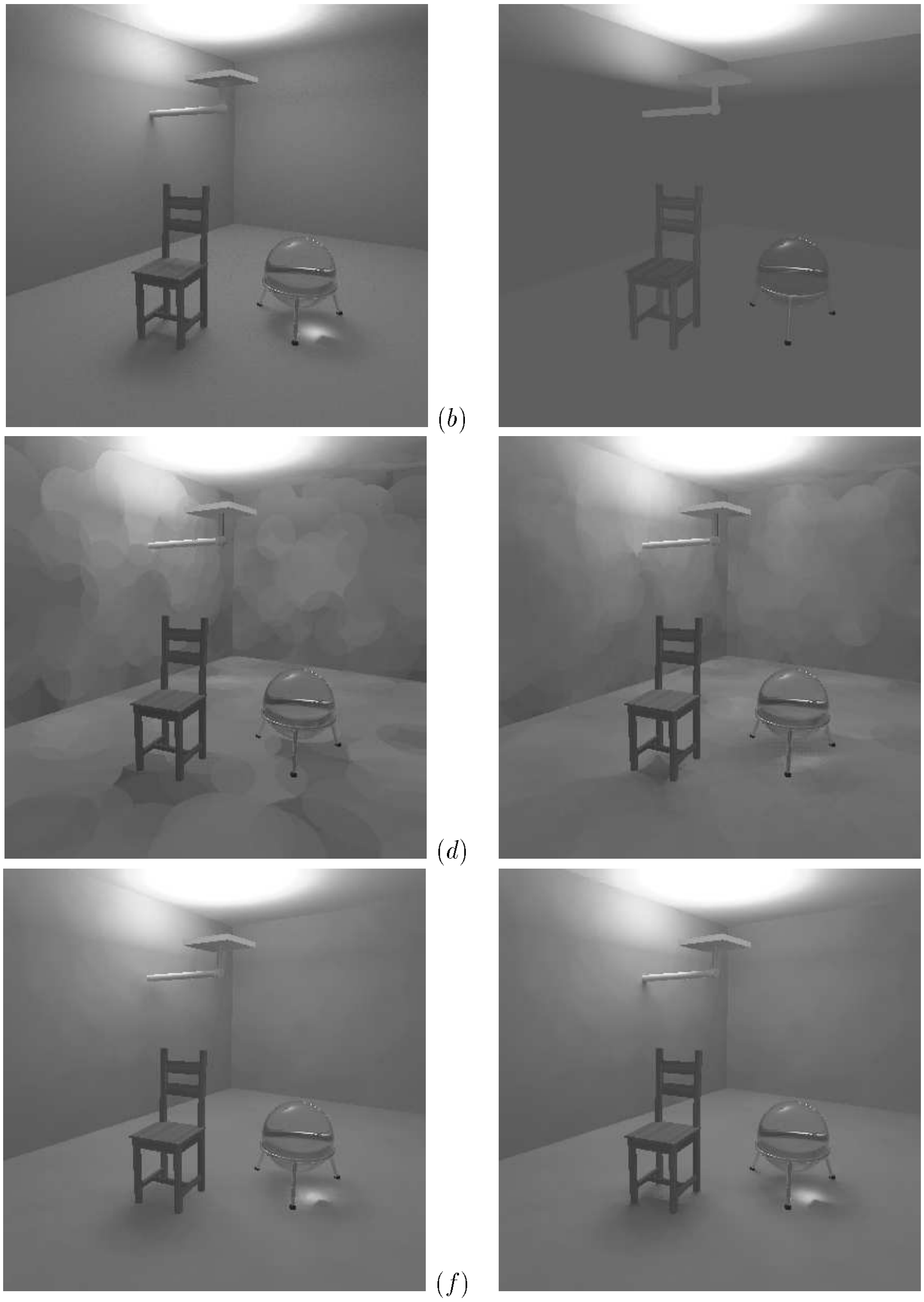


Vecteurs lumineux
4200 $\overrightarrow{VLI}$
(10 minutes)
$e = 0,33 \quad \sigma = 0,24$



Indirect diffus pur,
sans terme directionnel
4200 $\overrightarrow{VLI_{dif}}$
(10 minutes)
$e = 0,64 \quad \sigma = 0,55$

FIG. 6.6 – *Différentes méthodes de calcul du terme indirect*



(a) (b) (c) (d) (e) (f)

FIG. 6.7 – *Influence des paramètres (cas indirect)*

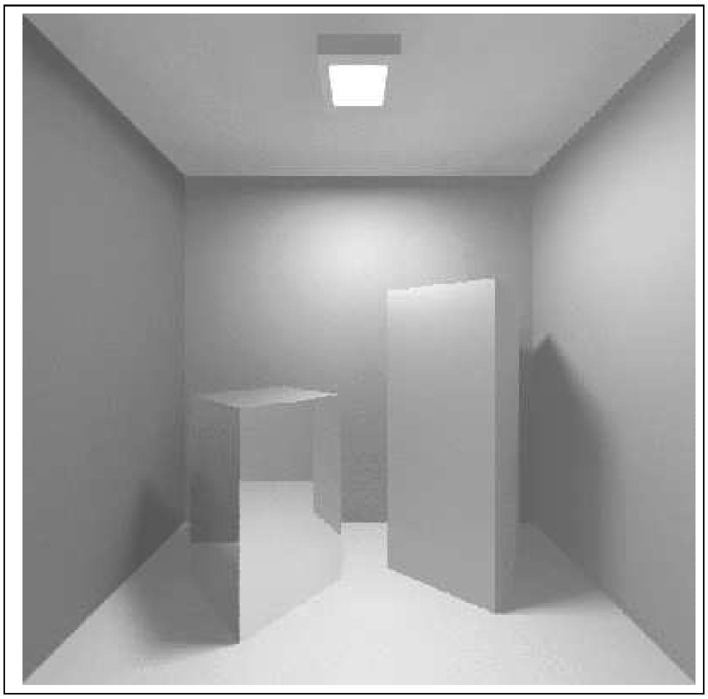


FIG. 6.8 – *"Boîte de Cornell" (5000 VLD, 2500 VLI, 5')*

FIG. 6.9 – *"Boîte de Cornell" (15000 VLD, 2800 VLI, 20000 VLC, 13')*

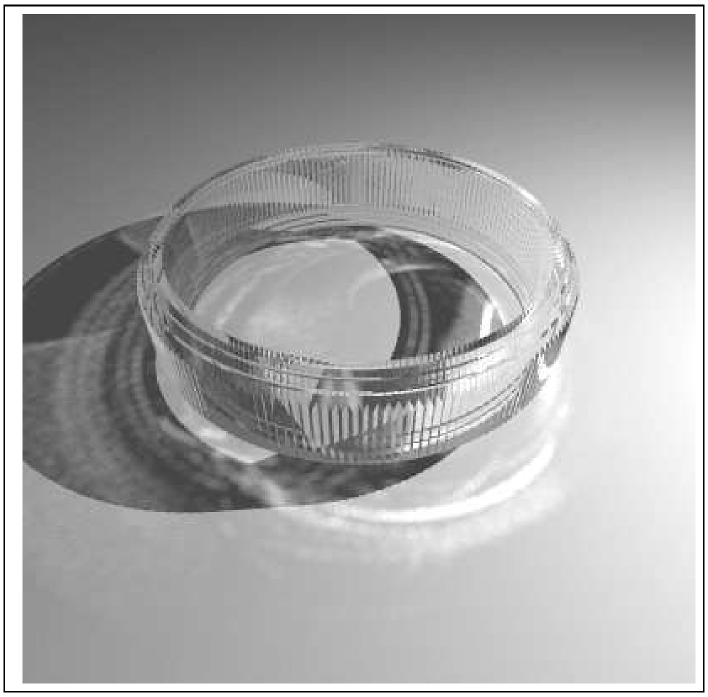


FIG. 6.10 – *Anneau cannelé et caustiques (30000 VLC, 90')*

FIG. 6.11 – *Scène d'intérieur éclairée par une fenêtre (25000 VLD, 5000 VLI, 220')*

FIG. 6.12 – *Scène d'intérieur, sans caustiques (15000 VLD, 11000 VLI, 180')*

FIG. 6.13 – *Scène d'intérieur, avec caustiques (20000 VLD, 10000 VLI, 24000 VLC, 300')*

(*a*) Photographie

(*b*) Image de synthèse

FIG. 6.14 – *Comparaison entre une photographie et une image de synthèse*

Chapitre 7

Conclusion et perspectives

7.1 Conclusion

Les images de synthèse actuelles ne sont pas très réalistes, visuellement comme physiquement, car elles souffrent de trop nombreuses simplifications. En particulier, les descriptions des scènes ne décrivent pas l'ensemble des détails caractéristiques des objets, et les modèles de calcul d'éclairement ne simulent pas toutes les propriétés physiques des interactions lumière-matière. Durant cette thèse, nous nous sommes attachés à améliorer ce deuxième point, pour prendre en compte l'éclairement global en limitant le nombre de calculs.

Nous avons présenté un modèle vectoriel général, représentant la lumière en un point d'une surface : le vecteur lumineux. Il prend en compte à la fois l'ensemble de l'énergie incidente, et les propriétés physiques locales du matériau. La scène complète est plongée dans un champ vectoriel. Lorsque sa valeur en un point n'est pas connue, un vecteur lumineux approché peut se déduire par interpolation de valeurs déjà calculées dans son voisinage, en exploitant des propriétés de continuité, après avoir établi des critères de validité.

Pour faciliter les interpolations, la lumière incidente est décomposée en trois termes : direct, indirect et caustique. Les variations de ces trois composantes sur une surface sont différentes. Les critères de validité sont alors optimisés pour ces trois cas particulier. D'autres décompositions sont envisageables, pour détailler plus finement les phénomènes simulés. Le modèle du vecteur lumineux unifie les représentations des diverses composantes lumineuses.

Parallèlement, nous avons développé une procédure de subdivision des sources surfaciques planes rectangulaires, organisée selon une structure hiérarchique binaire, qui réduit le nombre total d'échantillons. Nous avons montré que le procédé s'adapte facilement à des sources planes de forme quelconque. Il est ensuite étendu à des sources hémisphériques, dans le but de simuler des ciels naturels.

Actuellement, il apparaît que seules les méthodes de Monte Carlo soient capables de simuler l'ensemble des phénomènes physiques sans simplifications outrancières, au moins d'un point de vue théorique. En effet, elles sont capables de reproduire les comportements stochastiques des photons de lumière lorsqu'ils rencontrent les molécules de la matière. Cependant, il ne nous est pas possible de modéliser avec une telle finesse, et de toute façon les capacités des ordinateurs actuels seraient beaucoup trop limitées pour traiter de telles quantité de données.

Les méthode par éléments finis sont beaucoup plus difficiles à étendre au delà de leur cadre initial. Par contre, elle fournissent relativement rapidement une solution approchée satisfaisante au problème de l'éclairement indirect. Celles-ci peuvent être exploitées par des méthodes de Monte Carlo, en tant que solution initiale, pour réduire la variance, et accélérer la convergence vers un résultat plus précis.

Le modèle des vecteurs lumineux a été implémenté au sein d'un logiciel de lancer de rayons. La méthode a été testée dans différentes conditions, par rapport à des images de référence calculées sans interpolation. La réduction des temps de calcul est importante, pour des différences en général faibles, souvent visuellement insignifiantes. Avec ces optimisations, il est donc possible de prendre en compte un plus grand nombre de phénomènes physiques, sans pénaliser exagérément la durée des calculs.

Nous espérons que cette contribution permette une avancée vers un plus grand réalisme des images de synthèse.

7.2 Perspectives

De nombreux sujets n'ont été qu'à peine effleurés durant ces travaux, et quelques autres n'ont même pas été abordés. Il demeure de nombreuses questions ouvertes, qui mériteraient d'être traitées en détail. Nous donnons ici une liste des principaux thèmes connexes à cette thèse, que nous souhaiterions voir développés dans un proche avenir. Elle n'est malheureusement pas exhaustive, tellement le domaine de la synthèse d'images réaliste est vaste, et encore peu exploré.

- La prise en compte de BRDF mesurées sur des matériaux réels, avec leurs inévitables imperfections de surface, serait un atout certain pour améliorer le réalisme des images. Il semble indispensable d'aborder cet aspect de la modélisation pour les simulations d'éclairagisme.

- Modéliser correctement la lumière naturelle est indispensable pour une utilisation en architecture de la synthèse d'images. Les quelques modèles théoriques proposés par la CIE sont trop simplistes pour simuler la diversité des éclairements naturels, qui dépendent à la fois de paramètres géographiques et climatiques.

- La modélisation de sources non-lambertiennes permettrait de simuler des luminaires directionnels. Il peut s'agir, par exemple, de phares d'automobiles, ou de fenêtres équipées de stores sophistiqués.

- La prise en compte de milieux participants augmenterait encore la gamme des phénomènes simulés. La très grande complexité de leur modélisation réaliste constitue cependant un frein à leur mise en œuvre pratique. Des travaux résolvent partiellement ce problème, qui reste ouvert actuellement.

- Nous utilisons un modèle de couleurs classique à trois composantes (rouge, vert, bleu). Cette représentation n'est pas capable de simuler correctement la réflexion de la lumière, car elle ne travaille que sur des couleurs métamères dont le mélange n'a pas de signification. L'utilisation d'un modèle de couleurs spectral, plus lourd à manipuler, résoud ce problème. Le développement récent de représentations adaptatives de spectres semble une solution élégante.

- Nous n'avons pas du tout abordé le problème des grandes scènes, avec de nombreux objets et de nombreuses sources, ni celui des scènes extérieures qui cumulent ces deux aspects. Notre méthode est en théorie applicable dans ce cas, avec des ressources importantes. Il sera sans doute indispensable de développer des optimisations plus spécifiques.
- Pour le calcul de la composante indirecte des éclairements, nous avons implémenté une méthode de Monte Carlo simple, mais biaisée. Il serait intéressant de la remplacer par une méthode plus fiable, le suivi de chemins par exemple, avec laquelle un contrôle d'erreur serait envisageable. Il serait alors possible d'améliorer incrémentalement les résultats.
- Les interpolations de vecteurs lumineux utilisent des paramètres arbitraires et empiriques comme seuils de validité. En connaissant mieux les erreurs introduites, des solutions adaptatives automatiques pourraient calculer ces paramètres durant l'exécution du programme, en fonction de la précision souhaitée et des caractéristiques de l'image.
- Une estimation globale de l'erreur, pour toutes les composantes lumineuses, en cours de calcul, permettrait un contrôle adaptatif de la précision des images produites.
- L'introduction de notions psycho-visuelles permettrait d'adapter la précision des calculs à la sensibilité du système visuel humain. Lorsqu'ils n'influencent pas d'autres résultats, il est inutile de calculer des détails que l'observateur ne peut percevoir.

Annexe A

Méthodes de Monte-Carlo

Cette section décrit les fondements mathématiques des principales méthodes de Monte Carlo, notamment celles utilisées en synthèse d'images. Les ouvrages de Hammersley et Handscomb [HH64] et Rubinstein [Rub81] présentent toutes les définitions et démonstrations indispensables. De plus, les travaux de thèse de Dutré [Dut96], Lafortune [Laf96], Maillot [Mai96] et Veach [Vea97] ont été largement exploités, pour extraire les notions minimales utiles en synthèse d'images.

A.1 Présentation

L'expression *méthodes de Monte Carlo* désigne une famille de procédés mathématiques utilisés pour le calcul numérique d'intégrales. Leur principale caractéristique commune est de faire appel à des variables aléatoires pour améliorer statistiquement les vitesses de convergence, par rapport aux techniques d'intégration numériques classiques. De plus leur complexité reste constante quelle que soit la dimension de l'espace considéré. L'augmentation des capacités de calcul des ordinateurs a déclenché récemment leur développement pour résoudre des problèmes scientifiques de tous horizons. Elles sont de plus particulièrement bien adaptées à la simulation de phénomènes physiques ou interviennent des lois probabilistes, tels que l'étude des mouvements de particules.

Le nom de Monte Carlo est emprunté à une célèbre petite cité balnéaire, qui doit elle aussi son succès à la maîtrise des variables aléatoires, dans le domaine des jeux de hasard.

A.2 Quelques rappels de probabilités

A.2.1 Définitions

Une *variable aléatoire réelle* X est une application de Ω, l'ensemble des événements réalisables, vers $\mathbb{R}$. Elle décrit les résultats possibles d'une expérience.

La *fonction de répartition* F_X indique la probabilité que la valeur de la variable aléatoire X soit inférieure ou égale à une valeur x donnée :

$$\forall x, F_X(x) = P(X \leq x)$$

On lui associe une *fonction de densité de probabilité* f, qui vérifie les conditions suivantes:

1. $\forall x \in \mathbb{R}, \quad f(x) \geq 0$
2. $\displaystyle\int_{-\infty}^{+\infty} f(x)\,dx = 1$
3. $\displaystyle F_X(x) = P(X \leq x) = \int_{-\infty}^{x} f(y)\,dy$

La densité de probabilité est la dérivée de la fonction de répartition : $f(x) = \frac{dF_X(x)}{dx}$. Pour une valeur x donnée, $f(x)dx$ représente la probabilité que la valeur de la variable aléatoire soit égale à x. La probabilité que cette valeur soit comprise entre deux bornes a et b est : $P(a < x \leq b) = \int_a^b f(x)dx = F_X(b) - F_X(a)$. La fonction gaussienne, en forme de cloche, est un exemple classique de densité de probabilité.

L'*espérance* d'une variable aléatoire X est égale à la valeur moyenne obtenue en réalisant une infinité (ou tout du moins un très grand nombre) d'expériences, soit encore :

$$E(X) = \int_{-\infty}^{+\infty} x f(x)\,dx$$

La *variance* indique la façon dont sont réparties les valeurs de x autour de l'espérance :

$$\begin{aligned} V(X) &= \int_{-\infty}^{+\infty} [x - E(X)]^2 f(x) dx \\ &= E(X^2) - [E(X)]^2 \end{aligned}$$

Deux relations usuelles se déduisent des définitions de l'espérance et de la variance :

$$\forall a \in \mathbb{R} \qquad E(aX) = aE(X) \qquad V(aX) = a^2 V(X)$$

L'*écart-type*, ou *erreur standard*, est égal à la racine carrée de la variance : $\sigma(X) = \sqrt{V(X)}$
.

Plus généralement, si g est une fonction de la variable aléatoire X, alors :

$$\begin{aligned} E(g(X)) &= \int_{-\infty}^{+\infty} g(x)\, f(x)\,dx \\ \text{et} \quad V(g(X)) &= \int_{-\infty}^{+\infty} [g(x) - E(g(X))]^2 f(x) dx \\ &= E(g(X)^2) - [E(g(X))]^2 \end{aligned}$$

A.2.2 Tirage aléatoire selon une distribution donnée

Les générateurs aléatoires informatiques courants produisent des séries de nombres uniformément distribués sur un intervalle donné. La bonne qualité de la distribution obtenue est suffisante pour les applications usuelles en synthèse d'images. Il est possible de les utiliser pour générer des distributions respectant une fonction de répartition donnée. La méthode à utiliser dépend alors des données disponibles.

Fonction de répartition réciproque Le plus simple, lorsque la formulation analytique le permet, est d'utiliser la réciproque de la fonction de répartition. Si un nombre u est choisi aléatoirement de façon uniforme sur l'intervalle $[0,1]$, alors $F_X^{-1}(u)$ respecte la distribution de la variable aléatoire X.

Lorsque les données ont été obtenues expérimentalement, des tables numériques permettent d'obtenir rapidement une approximation de F_X^{-1}.

Tirage avec rejet La méthode du tirage avec rejet peut être utilisée lorsque seule la densité de probabilité f est connue. Elle consiste en tirage aléatoire uniforme de deux valeurs x et y. x appartient à l'intervalle de définition de f, alors que y est choisi dans l'intervalle $[0, M]$, où M est la valeur maximale prise par la fonction f. Si $y \leq f(x)$, la valeur de x obéit à la distribution souhaitée et elle est conservée, sinon la paire est rejetée et un nouveau tirage est effectué.

A.3 Calcul numérique d'intégrales

A.3.1 Méthodes de Monte Carlo naïves

A.3.1.1 Méthode touché-manqué

La méthode touché-manqué de calcul numérique d'intégrales est la plus simple des méthodes de Monte Carlo. L'idée intuitive est que la probabilité qu'un point généré aléatoirement soit à l'intérieur d'une surface est proportionnelle à l'aire de cette surface, relativement à l'aire totale de la boîte englobante où sont générés les points.

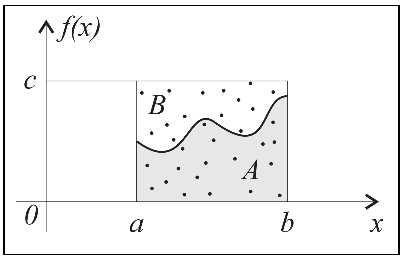


FIG. A.1 – *Calcul d'une intégrale par la méthode touché-manqué*

Soit à calculer l'intégrale : $I = \displaystyle\int_a^b f(x)dx$ (figure A.1). On suppose pour simplifier que $f(x) \in [0, c]$ pour tout $x \in [a, b]$.

Soit Ω le rectangle englobant de la fonction : $\Omega = \{(x, y), a \leq x \leq b, 0 \leq y \leq c\}$. La fonction de densité de probabilité d'un point (X, Y), généré de façon aléatoire uniforme sur Ω, est constante : $f_{XY}(x, y) = \dfrac{1}{c(b-a)}$.

La probabilité p que le point soit en dessous de la courbe de $f(x)$ (zone A de la figure A.1) est :

$$p = \frac{\text{aire } A}{\text{aire } \Omega} = \frac{\int_a^b f(x)dx}{c(b-a)} = \frac{I}{c(b-a)}$$

Une estimation $\hat{p}$ de la valeur de p peut être obtenue en générant aléatoirement N points indépendants, et en comptant les N_A points qui sont en dessous de la courbe de f (zone A de la figure A.1) : $\hat{p} = \dfrac{N_A}{N}$. On en déduit une estimation $\widehat{I}_1$ de l'intégrale I :

$$\widehat{I}_1 = c(b-a)\frac{N_A}{N}$$

$\widehat{I}_1$ est un estimateur sans biais de l'intégrale I, c'est à dire que la convergence vers la valeur exacte est garantie lorsque N devient assez grand. On le vérifie en calculant l'espérance de $\widehat{I}_1$, après avoir constaté que le tirage des N points suit une loi de Bernoulli de probabilité p :

$$E(\widehat{I}_1) = c(b-a)\frac{E(N_A)}{N} = c(b-a)\frac{Np}{N} = c(b-a)p = I$$

La variance de $\widehat{I}_1$ précise la vitesse de convergence :

$$V(\widehat{I}_1) = [c(b-a)]^2 V(\hat{p}) = [c(b-a)]^2 \frac{1}{N} p(1-p) = \frac{1}{N} I[c(b-a)-I]$$

L'écart-type de $\widehat{I}_1$ vaut donc :

$$\sigma(\widehat{I}_1) = \frac{1}{\sqrt{N}}\sqrt{I[c(b-a)-I]}$$

Pour une intégrale donnée, l'écart-type est inversement proportionnel à la racine carrée de N. Il faut donc quadrupler le nombre de points générés pour seulement doubler la précision du résultat. Cette vitesse de convergence est relativement modeste, mais la mise en œuvre de la méthode est très simple, même lorsque la dimension du problème augmente.

La connaissance de la variance permet de borner l'erreur moyenne commise, en utilisant l'inégalité de Chebyshev :

$$P(|\widehat{I}_1 - I| < \varepsilon) \geq 1 - \frac{V(\widehat{I}_1)}{\varepsilon^2}$$

Cette formule permet de déterminer le nombre minimal de points à générer pour obtenir statistiquement une tolérance donnée. Elle ne donne cependant aucune indication à propos de l'erreur commise sur un calcul particulier.

A.3.1.2 Estimateur simple d'une intégrale

La méthode précédente s'adapte tout naturellement au calcul numérique d'intégrales. Soit f une fonction de $L^2(0,1)$ (c'est à dire que $\int_0^1 f^2(x)dx$ existe), et soit I l'intégrale à calculer :

$$I = \int_0^1 f(x)\,dx$$

L'intégrale I peut être estimée en choisissant un nombre aléatoire uniforme ξ sur $[0,1]$. La valeur $f(\xi)$ est un estimateur primaire de l'intégrale :

$$\langle I \rangle_{prim} = f(\xi) \tag{A.1}$$

L'espérance de cet estimateur est égale à I, la valeur recherchée. En effet :

$$E(\langle I\rangle_{prim}) = \int_0^1 f(x)u(x)dx$$

où $u(x)$ est la densité de probabilité de la variable aléatoire. Or la distribution est uniforme sur $[0,1]$, soit encore $\forall x \in [0,1]$, $u(x)=1$. Donc :

$$E(\langle I\rangle_{prim}) = \int_0^1 f(x)dx = I$$

La variance de cet estimateur est :

$$\begin{aligned} V(\langle I\rangle_{prim}) &= \int_0^1 (f(x)-I)^2 dx \\ &= \int_0^1 f^2(x)dx - I^2 \end{aligned}$$

A.3.1.3 Méthode brutale d'intégration d'une fonction

La variance de l'estimateur primaire est en général très importante, et cette méthode de mise en œuvre très simple n'est donc pas utilisable en pratique. Afin de réduire l'incertitude, l'expérience est renouvelée N fois, pour calculer une valeur moyenne des résultats obtenus. L'intégrale de départ est alors transformée en une somme de N intégrales :

$$I \;=\; \int_0^1 f(x)dx \;=\; \sum_{i=1}^{N}\int_0^1 \frac{f(x)}{N}dx \;=\; \sum_{i=1}^{N} I_i$$

L'estimateur secondaire est la somme des estimateurs primaires $\langle I_i\rangle$:

$$\langle I\rangle_{sec} \;=\; \sum_{i=1}^{N}\langle I_i\rangle \;=\; \frac{1}{N}\sum_{i=1}^{N} f(\xi_i) \tag{A.2}$$

Cet estimateur secondaire est lui aussi non biaisé :

$$E(\langle I\rangle_{sec}) \;=\; \int_0^1 \frac{1}{N}\sum_{i=1}^{N} f(x)\,dx \;=\; \frac{1}{N}\int_0^1 Nf(x)\,dx \;=\; I$$

La variance de l'estimateur secondaire est

$$\begin{aligned} V(\langle I\rangle_{prim}) &= \frac{1}{N}\int_0^1 (f(x)-I)^2 dx \\ &= \frac{V(\langle I\rangle_{prim})}{N} \end{aligned}$$

L'écart-type étant la racine carrée de la variance, l'erreur standard diminue proportionnellement à $\frac{1}{\sqrt{N}}$. Il faut donc générer quatre fois plus d'échantillons pour diminuer par deux l'erreur moyenne, ce qui est un résultat plutôt médiocre.

A.3.2 Techniques de réduction de variance

La réduction de la variance des méthodes de Monte Carlo permet, pour une précision donnée, de limiter le nombre total d'échantillons, et donc de réduire la durée des calculs. Les principales méthodes sont présentées ici. Elles peuvent dans certains cas être combinées entre elles, pour améliorer encore les résultats.

A.3.2.1 Echantillonnage stratifié

La stratification consiste à découper le domaine d'intégration en strates, dont les intégrales sont calculées séparément par la méthode brutale. L'estimateur $\langle I\rangle_{strat}$ de l'intégrale est de la forme :

$$\langle I\rangle_{strat} = \sum_{j=1}^{k}\sum_{i=1}^{N_j} \frac{\alpha_j - \alpha_{j-1}}{N_j} f(\alpha_{j-1} + (\alpha_j - \alpha_{j-1})\xi_{ij})$$

où $\{\alpha_0, \alpha_1, \ldots, \alpha_k\}$ sont les bornes des k strates, telles que $\alpha_0 = 0 < \alpha_1 < \ldots < \alpha_k = 1$, et où N_j est le nombre d'échantillons tirés dans la strate $[\alpha_{j-1}, \alpha_j]$. De plus, ξ_{ij} est le tirage aléatoire uniforme sur $[0,1]$ de l'échantillon i de la strate j.

Cet estimateur n'est pas biaisé, et sa variance est :

$$V(\langle I\rangle_{strat}) = \sum_{j=1}^{k} \frac{\alpha_j - \alpha_{j-1}}{N_j} \int_{\alpha_{j-1}}^{\alpha_j} f(x)^2\,dx - \sum_{j=1}^{k} \frac{1}{N_j} \left(\int_{\alpha_{j-1}}^{\alpha_j} f(x)\,dx\right)^2$$

On peut montrer que cette variance est meilleure que celle de la méthode brutale [HH64], lorsque les différences entre les valeurs moyennes de f dans les strates sont plus grandes que les variations de f à l'intérieur des strates.

L'échantillonnage stratifié est dit uniforme lorsque les strates ont toutes la même taille, ce qui n'est pas toujours judicieux au vu du résultat précédent. En fait, les strates doivent être plus fines dans les régions du domaine de définition où la fonction varie le plus fortement, ce qui réduit la variance à l'intérieur de chacune des strates.

A.3.2.2 Echantillonnage d'importance

Il arrive parfois, notamment lors de la simulation de phénomènes physiques, que l'allure générale de la fonction à intégrer soit connue, au moins approximativement. Cette connaissance est alors mise à profit avec la méthode d'échantillonnage d'importance, qui concentre les échantillons dans les zones les plus "importantes" du domaine d'intégration.

Si g est une fonction "proche" de f, l'intégrale peut s'écrire :

$$I = \int_0^1 f(x)\,dx = \int_0^1 \frac{f(x)}{g(x)} g(x)\,dx = \int_0^1 \frac{f(x)}{g(x)}\,dG(x)$$

où la fonction de distribution G satisfait :

$$G(x) = \int_0^x g(y)\,dy \quad \text{et} \quad G(1) = 1$$

Alors,

$$\langle I\rangle_{imp} = \frac{1}{N}\sum_{i=1}^{N} \frac{f(\eta_i)}{g(\eta_i)}$$

est un estimateur sans biais de l'intégrale I, à la condition que les N échantillons η_i soient choisis de façon à respecter la distribution de probabilités G. Si α_i est tiré uniformément sur l'intervalle $[0,1]$, alors $\eta_i = G^{-1}(\alpha_i)$.

La variance de l'estimateur d'importance est :

$$V(\langle I\rangle_{imp}) = \int_0^1 \left(\frac{f(x)}{g(x)} - I\right)^2 dG(x)$$

Cette variance est nulle lorsque la fonction g est proportionnelle à f, et elle reste relativement faible lorsque g est seulement une approximation de f, même grossière.

A.3.2.3 Contrôle de variations

La méthode du contrôle de variation s'utilise lorqu'une approximation ϕ de la fonction f est connue, et que ϕ est suffisamment simple pour être intégrée analytiquement. L'intégrale est alors décomposée en deux éléments :

$$I = \int_0^1 f(x) - \phi(x) + \phi(x)\,dx = \int_0^1 \phi(x)\,dx + \int_0^1 [f(x) - \phi(x)]\,dx$$

La première partie est calculée analytiquement, et la seconde est évaluée par la méthode de Monte Carlo brutale. L'estimateur sans biais est donc :

$$\langle I\rangle_{var} = \int_0^1 \phi(x)\,dx + \frac{1}{N}\sum_{i=1}^{N}[f(\xi_i) - \phi(\xi_i)]$$

Intuitivement, si l'on commet une erreur, ce sera uniquement sur la différence entre les valeurs des fonctions f et ϕ, qui est sensée être petite par rapport à f, si ϕ est convenablement choisie.

A.4 Application à l'équation de rendu

L'équation intégrale de rendu (équation 3.1), qui décrit l'illumination globale dans un environnement, n'a pas de solution analytique en général. La résolution par une méthode Monte Carlo est une issue à ce problème crucial en synthèse d'image, pratiquement incontournable pour le lancer de rayons et les méthodes dérivées.

A.4.1 Résolution des équations de Fredholm du second type

Les équations de Fredholm du second type sont des expressions récursives de la forme :

$$f(x) = g(x) + \int_0^1 K(x,y)f(y)dy$$

où f est la fonction inconnue, g est une fonction connue et K est le noyau de l'opérateur intégral. La convergence de l'intégrale est assurée lorsque la norme du noyau est inférieure à 1 :

$$\|K\| = \sup_x \int |K(x,y)|\,dy < 1$$

L'équation de rendu est de ce type, et son noyau vérifie la condition de convergence, grâce au principe de conservation de l'énergie. L'état des surfaces n'étant jamais parfait, à chaque réflexion une partie de l'énergie est absorbée par les matériaux, pour être transformée en chaleur. La quantité d'énergie lumineuse transportée diminue donc à chaque rebond.

La solution peut être calculée par une décomposition en éléments finis, ce que fait la méthode de la radiosité (section 3.5.1), après que le problème ait été simplifié. Une autre approche, qui nous intéresse ici, est d'utiliser une méthode Monte Carlo adaptée à ce problème.

Le terme intégral de l'équation est évaluée récursivement par une méthode Monte Carlo avec échantillonnage d'importance. A chaque étape, un échantillon ξ_i est choisi dans $[0, 1]$, conformément à une fonction de densité de probabilité p_i (qui est éventuellemnt la loi uniforme si l'on ne dispose pas de suffisamment d'information sur la fonction), pour calculer un estimateur primaire $\langle f(x)\rangle_{prim,rec}$ de l'intégrale récursive :

$$\begin{aligned}
\langle f(x)\rangle_{prim,rec} &= g(x) + \frac{K(x,\xi_1)}{p_1(\xi_1)} \langle f(\xi_1)\rangle_{prim,rec} \\
&= g(x) + \frac{K(x,\xi_1)}{p_1(\xi_1)} \left[g(\xi_1) + \frac{K(\xi_1,\xi_2)}{p_2(\xi_2)} \langle f(\xi_2)\rangle_{prim,rec} \right] \\
&= g(x) + \frac{K(x,\xi_1)}{p_1(\xi_1)} g(\xi_1) + \frac{K(x,\xi_1)}{p_1(\xi_1)} \frac{K(\xi_1,\xi_2)}{p_2(\xi_2)} g(\xi_2) + \cdots \\
&= \sum_{i=0}^{\infty} \left(\prod_{j=1}^{i} \frac{K(\xi_{j-1},\xi_j)}{p_j(\xi_j)} g(\xi_i) \right) \qquad \text{avec} \quad \xi_0 = x
\end{aligned}$$

La suite de points $\xi_0, \xi_1, \cdots$ est une marche aléatoire, que nous appellerons en synthèse d'image *chemin*, puisqu'elle symbolise pour nous un chemin lumineux. On peut aussi noter qu'il s'agit de plus d'une *chaîne de Markov*, car chaque position ξ_i dépend uniquement de la position ξ_{i-1} et de la loi de probabilité p_i, quel que soit le chemin emprunté pour parvenir en ξ_{i-1} depuis ξ_0. Ceci lui confère des propriétés mathématiques particulières.

$\langle f(x)\rangle_{prim,rec}$ est un estimateur sans biais, puisqu'il est lui-même la somme d'estimateurs sans biais. On en déduit un estimateur de moindre variance, sans biais lui aussi, en réitérant N fois l'opération :

$$\langle f(x)\rangle_{rec} = \frac{1}{N} \sum_{k=1}^{N} \left[\sum_{i=0}^{\infty} \left(\prod_{j=1}^{i} \frac{K(\xi_{j-1},\xi_j)}{p_j(\xi_j)} g(\xi_i) \right) \right]$$

A.4.2 Troncature par la roulette russe

La méthode précédente de calcul d'intégrales récursive souffre d'un inconvénient majeur : il s'agit d'une somme infinie. Il n'est donc pas possible de l'exploiter directement sous cette forme. Tronquer arbitrairement l'évaluation à un niveau donné introduit un biais incontrôlé, car rien ne peut garantir que l'étape suivante ne serait pas prédominante. C'est pourtant un procédé très fréquent en synthèse d'image, à cause de sa simplicité.

La méthode de la roulette russe est une solution à ce problème. La troncature de la chaîne de Markov est alors décidée de manière probabiliste, après avoir modifié l'expression de l'intégrale pour n'introduire aucun biais.

Soit $P \in]0,1[$ un nombre fixé, et u une fonction nulle pour tout x supérieur à P telle que :

$$\forall x \in \mathbb{R}, \begin{cases} u(x) = 1 & \text{si } x \leq P \\ u(x) = 0 & \text{sinon} \end{cases}$$

L'intégrale à évaluer est modifiée ainsi :

$$\begin{aligned} I &= \int_0^1 f(x)\,dx \\ &= \int_0^P \frac{1}{P} f(\frac{x'}{P})\,dx' \\ &= \int_0^1 \frac{1}{P} f(\frac{x'}{P})\,u(\frac{x'}{P})\,dx' \end{aligned}$$

Cette modification s'interprète comme une déformation de la fonction à intégrer. Le domaine d'intégration est réduit d'un facteur P, alors que dans le même temps la valeur de l'intégrande augmente d'un facteur $\frac{1}{P}$. La valeur de l'intégrale n'est donc pas modifiée. En appliquant la méthode de Monte Carlo à l'intégrale modifiée, on obtient l'estimateur :

$$\langle I \rangle_{russe} = \begin{cases} \frac{1}{P} f(\frac{\xi}{P}) & \text{si } \xi \leq P \\ 0 & \text{sinon} \end{cases}$$

Cet estimateur est sans biais, car l'intégrale modifiée a la même valeur que l'intégrale originale. En contrepartie, sa variance est plus élevée que celle de l'estimateur primaire :

$$\begin{aligned} V(\langle I \rangle_{russe}) &= \int_0^1 \left[\frac{1}{P} f(\frac{\xi}{P})\,u(\frac{\xi}{P})\right]^2 d\xi - I^2 \\ &= \int_0^P \frac{1}{P^2} f^2(\frac{\xi}{P})\,d\xi - I^2 \\ &= \frac{1}{P}\int_0^1 f^2(\xi)\,d\xi - I^2 > V(\langle I \rangle_{prim}) \end{aligned}$$

A.4.3 Suivi aléatoire de chemins non biaisé

La combinaison de la méthode de la roulette russe avec l'estimateur récursif $\langle f(x) \rangle_{prim,rec}$, fournit une méthode non biaisée d'estimation d'un chemin. La marche aléatoire devient finie, car il existe un point ξ_t à partir duquel le reste de l'intégrale sera tenu pour nul. L'estimateur devient donc :

$$\langle f(x) \rangle_{prim,russe,rec} = \sum_{i=0}^{t} \left(\prod_{j=1}^{i} \frac{K(\xi_{j-1}, \xi_j)}{P_j\, p_j(\xi_j)}\, g(\xi_i) \right)$$

P_j est la probabilité que l'intégrale soit évaluée au niveau j, en poursuivant le chemin, alors que $1 - P_j$ est la probabilité de tronquer l'évaluation, avec la roulette russe.

Comme pour les autres estimateurs, la variance globale est réduite en réitérant N fois l'opération :

$$\langle f(x) \rangle_{russe,rec} = \frac{1}{N} \sum_{k=1}^{N} \left[\sum_{i=0}^{t} \left(\prod_{j=1}^{i} \frac{K(\xi_{j-1}, \xi_j)}{P_j\, p_j(\xi_j)}\, g(\xi_i) \right) \right]$$

Annexe B

Calcul du gradient d'éclairement indirect

B.1 Variations de la direction d'incidence

B.1.1 Analyse du problème

Les valeurs recueillies lors de l'échantillonnage d'un hémisphère pour calculer un éclairement permettent le calcul d'une approximation du "gradient" d'éclairement. Il est nécessaire pour cela de connaître les variations de la direction d'incidence des échantillons pour des petits déplacements du centre de l'hémisphère.

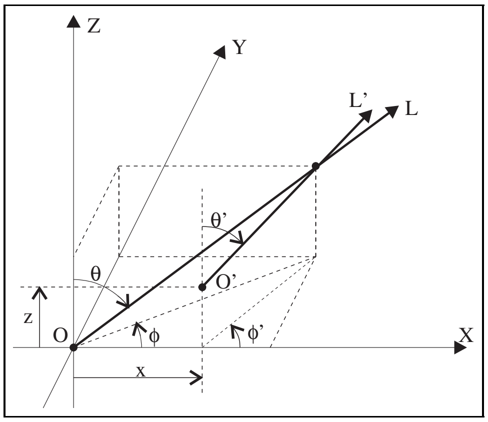


FIG. B.1 – *Déplacement de O en O' du centre de l'hémisphère*

On se place dans un repère orthonormé (O, X, Y, Z) (figure B.1) centré sur l'hémisphère. La direction d'incidence de chaque échantillon est repérée par deux coordonnées angulaires, l'angle zénithal θ et l'angle azimutal ϕ.

O	$=$	Centre de l'hémisphère
(O,X,Y)	$=$	Plan tangent à la surface
(O,Z)	$=$	Normale à la surface
R	$=$	Rayon de l'hémisphère

Il s'agit maintenant de déterminer de quelle manière sont modifiés les angles θ et ϕ pour des déplacements infiniment petits du centre de l'hémisphère dans le demi-espace ($z \geq 0$). On va donc calculer les six dérivées partielles suivantes :

$$\frac{\partial\theta}{\partial x} \quad \frac{\partial\phi}{\partial x} \quad \frac{\partial\theta}{\partial y} \quad \frac{\partial\phi}{\partial y} \quad \frac{\partial\theta}{\partial z} \quad \frac{\partial\phi}{\partial z}$$

B.1.2 Déplacement selon l'axe (O,X)

Il faut dans un premier temps exprimer θ' et ϕ', les nouvelles coordonnées angulaires, en fonction de θ , ϕ et x, pour ensuite calculer les dérivées selon x (figure B.2).

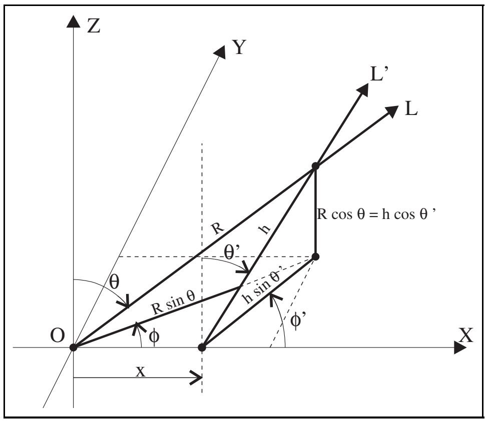


FIG. B.2 – *Déplacement selon l'axe* (O,X)

Les 3 relations suivantes sont immédiates :

$$R\cos\theta = h\cos\theta' \tag{B.1}$$
$$R\sin\theta\cos\phi = x + h\sin\theta'\cos\phi' \tag{B.2}$$
$$R\sin\theta\sin\phi = h\sin\theta'\sin\phi' \tag{B.3}$$

- **Variation de l'angle ϕ**

$$\text{B.1} \Rightarrow h = \frac{R\cos\theta}{\cos\theta'} \tag{B.4}$$
$$\text{B.2, B.4} \Rightarrow R\sin\theta\cos\phi = x + R\cos\theta\tan\theta'\cos\phi' \tag{B.5}$$
$$\text{B.3, B.4} \Rightarrow R\sin\theta\sin\phi = R\cos\theta\tan\theta'\sin\phi' \tag{B.6}$$

$$\text{B.6} \quad \Rightarrow \quad \tan\theta' = \tan\theta \frac{\sin\phi}{\sin\phi'} \tag{B.7}$$

$$\begin{aligned}
\text{B.5, B.7} \quad &\Rightarrow \quad R\,\sin\theta\,\cos\phi = x + R\,\cos\theta\,\tan\theta \frac{\sin\phi}{\tan\phi'} \\
&\Rightarrow \quad \tan\phi' = \frac{R\,\sin\theta\,\sin\phi}{R\,\sin\theta\,\cos\phi - x}
\end{aligned}$$

On obtient finalement :

$$\phi' = \arctan\left(\frac{R\,\sin\theta\,\sin\phi}{R\,\sin\theta\,\cos\phi - x}\right) \tag{B.8}$$

Il faut maintenant dériver cette expression de ϕ' en fonction de x :

$$\frac{\partial\phi'}{\partial x} = \frac{\frac{\partial}{\partial x}\left(\frac{R\,\sin\theta\,\sin\phi}{R\,\sin\theta\,\cos\phi - x}\right)}{1 + \left(\frac{R\,\sin\theta\,\sin\phi}{R\,\sin\theta\,\cos\phi - x}\right)^2} = \frac{\frac{R\,\sin\theta\,\sin\phi}{(R\,\sin\theta\,\cos\phi - x)^2}}{1 + \left(\frac{R\,\sin\theta\,\sin\phi}{R\,\sin\theta\,\cos\phi - x}\right)^2}$$

On considère des déplacements infiniment petits, d'où :

$$\lim_{x\to 0}\left(\frac{\partial\phi}{\partial x}\right) = \frac{\frac{R\,\sin\theta\,\sin\phi}{(R\,\sin\theta\,\cos\phi)^2}}{1 + \left(\frac{R\,\sin\theta\,\sin\phi}{R\,\sin\theta\,\cos\phi}\right)^2}$$

Qui se simplifie en :

$$\frac{\partial\phi}{\partial x} = \frac{\sin\phi}{R\,\sin\theta} \tag{B.9}$$

- **Variation de l'angle θ**

$$\text{B.6} \quad \Rightarrow \quad \theta' = \arctan\left(\tan\theta \frac{\sin\phi}{\sin\phi'}\right) \tag{B.10}$$

Il faut maintenant dériver cette expression de θ' en fonction de x :

$$\frac{\partial\theta'}{\partial x} = \frac{\frac{\partial}{\partial x}\left(\tan\theta \frac{\sin\phi}{\sin\phi'}\right)}{1 + \left(\tan\theta \frac{\sin\phi}{\sin\phi'}\right)^2} = \frac{\tan\theta\,\sin\phi \frac{\partial}{\partial x}\left(\frac{1}{\sin\phi'}\right)}{1 + \left(\tan\theta \frac{\sin\phi}{\sin\phi'}\right)^2}$$

Il est plus simple de dériver directement $\frac{\partial}{\partial x}\left(\frac{1}{\sin\phi'}\right)$ plutôt que de reprendre l'expression de ϕ' obtenue en (équation B.8) :

$$\frac{\partial}{\partial x}\left(\frac{1}{\sin\phi'}\right) = \frac{-\cos\phi'}{\sin^2\phi'} \cdot \frac{\partial\phi'}{\partial x}$$

En réutilisant l'expression de $\lim_{x\to 0}\left(\frac{\partial\phi}{\partial x}\right)$ obtenue en (équation B.9) :

$$\lim_{x\to 0}\left(\frac{\partial}{\partial x}\left(\frac{1}{\sin\phi'}\right)\right) = \frac{-\cos\phi\,\sin\phi}{\sin^2\phi\,R\,\sin\theta} = \frac{-1}{R\,\tan\phi\,sin\theta}$$

En utilisant $\lim_{x\to 0}\phi' = \phi$:

$$\lim_{x\to 0}\left(\frac{\partial\theta}{\partial x}\right) = \frac{-\tan\theta\,\sin\phi}{(1+\tan^2\theta)(R\,\tan\phi\,sin\theta)}$$

On obtient finalement, en simplifiant l'expression :

$$\frac{\partial\theta}{\partial x} = \frac{-\cos\theta\,\cos\phi}{R} \tag{B.11}$$

B.1.3 Déplacement selon l'axe (O, Y)

Le principe est exactement le même que pour un déplacement selon l'axe (O, X), la seule différence étant que l'angle ϕ devient $\phi - \frac{\pi}{2}$.

On obtient :

$$\frac{\partial \theta}{\partial y} = \frac{-\cos\theta \sin\phi}{R} \tag{B.12}$$

$$\frac{\partial \phi}{\partial y} = \frac{-\cos\phi}{R \sin\theta} \tag{B.13}$$

B.1.4 Déplacement selon l'axe (O, Z)

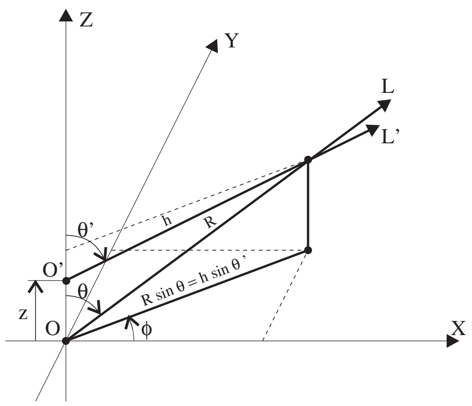


FIG. B.3 – *Déplacement selon l'axe* (O, Z)

- **Variation de l'angle ϕ**

L'angle azimutal ϕ n'est pas affecté par un déplacement selon l'axe (O, Z). On a donc simplement :

$$\frac{\partial \phi}{\partial z} = 0 \tag{B.14}$$

- **Variation de l'angle θ**

Les deux relations suivantes sont immédiates :

$$R \sin\theta = h \sin\theta' \qquad \text{et} \qquad R \cos\theta = z + h \cos\theta'$$

On en déduit facilement :

$$h = \frac{R \sin\theta}{\sin\theta'} \qquad \text{et} \qquad R \cos\theta = z + \frac{R \sin\theta}{\tan\theta'}$$

d'où :

$$\theta' = \arctan\left(\frac{R \sin\theta}{R \cos\theta - z}\right) \tag{B.15}$$

Il ne reste plus qu'à dériver cette expression en fonction de z :

$$\frac{\partial\theta}{\partial z} = \frac{\frac{\partial}{\partial z}\left(\frac{R\sin\theta}{R\cos\theta - z}\right)}{1+\left(\frac{R\sin\theta}{R\cos\theta-z}\right)^2} = \frac{\frac{R\sin\theta}{(R\cos\theta-z)^2}}{1+\left(\frac{R\sin\theta}{R\cos\theta-z}\right)^2}$$

On considère des déplacements infiniment petits :

$$\lim_{z\to 0}\left(\frac{\partial\theta}{\partial z}\right) = \frac{\frac{\tan\theta}{R\cos\theta}}{1+\tan^2\theta}$$

Et l'on obtient finalement :

$$\frac{\partial\theta}{\partial z} = \frac{\sin\theta}{R} \tag{B.16}$$

B.1.5 Récapitulation

On a obtenu les 6 relations suivantes :

$$\left\{\begin{array}{lcl} \dfrac{\partial\theta}{\partial x} & = & \dfrac{-\cos\theta\cos\phi}{R} \\ \dfrac{\partial\theta}{\partial y} & = & \dfrac{-\cos\theta\sin\phi}{R} \\ \dfrac{\partial\theta}{\partial z} & = & \dfrac{\sin\theta}{R} \end{array}\right. \quad \text{et} \quad \left\{\begin{array}{lcl} \dfrac{\partial\phi}{\partial x} & = & \dfrac{\sin\phi}{R\sin\theta} \\ \dfrac{\partial\phi}{\partial y} & = & \dfrac{-\cos\phi}{R\sin\theta} \\ \dfrac{\partial\phi}{\partial z} & = & 0 \end{array}\right. \tag{B.17}$$

B.2 Gradient d'éclairement indirect

Les “gradients” d'éclairement sont obtenus par dérivation de l'équation de calcul de l'éclairement discrétisée :

$$E = \frac{\pi^2}{2MN}\sum_{j=0}^{M-1}\sum_{k=0}^{N-1}\sin 2\theta_j \cdot L(\theta_j,\phi_k)$$

De plus, les déplacements du centre de l'hémisphère doivent être assez petits pour considérer que la luminance incidente selon une direction a donnée reste constante : $\dfrac{\partial L}{\partial a} = 0$.

• **Déplacement selon une direction $\boldsymbol{x}$** dans le plan tangent à la surface :

$$\frac{\partial E}{\partial x} = \frac{\pi^2}{2MN}\sum_{j=0}^{M-1}\sum_{k=0}^{N-1} 2\cos 2\theta_j \cdot \frac{\partial\theta_j}{\partial x}\cdot L(\theta_j,\phi_k)$$

Les résultats obtenus en (équation B.17) nous permettent de déduire :

$$\frac{\partial E}{\partial x} = \frac{-\pi^2}{MN}\sum_{j=0}^{M-1}\sum_{k=0}^{N-1}\cos 2\theta_j\cdot\cos\theta_j\cdot\cos\phi_k\cdot L(\theta_j,\phi_k)$$

• **Déplacement selon une direction $\boldsymbol{y}$** perpendiculaire à x, elle aussi dans le plan tangent à la surface :

$$\frac{\partial E}{\partial y} = \frac{-\pi^2}{MN} \sum_{j=0}^{M-1} \sum_{k=0}^{N-1} \cos 2\theta_j \cdot \cos\theta_j \cdot \cos(\phi_k - \frac{\pi}{2}) \cdot L(\theta_j, \phi_k)$$

d'où

$$\frac{\partial E}{\partial y} = \frac{-\pi^2}{MN} \sum_{j=0}^{M-1} \sum_{k=0}^{N-1} \cos 2\theta_j \cdot \cos\theta_j \cdot \sin(\phi_k) \cdot L(\theta_j, \phi_k)$$

• **Déplacement selon une direction *z*** orientée selon la normale à la surface :

$$\frac{\partial \theta_j}{\partial z} = \frac{\sin\theta_j}{R}$$

On en déduit :

$$\frac{\partial E}{\partial z} = \frac{\pi^2}{MN} \sum_{j=0}^{M-1} \sum_{k=0}^{N-1} \cos 2\theta_j \cdot \sin\theta_j \cdot L(\theta_j, \phi_k)$$

• **Le "gradient" global absolu**, utilisé comme indicateur de perturbation de l'éclairement indirect, est la somme des valeurs absolues des gradients selon les trois directions du repère :

$$GradAbs = \left|\frac{\partial E}{\partial x}\right| + \left|\frac{\partial E}{\partial y}\right| + \left|\frac{\partial E}{\partial z}\right| \tag{B.18}$$

Les expressions semblent complexes, mais la plupart des termes sont communs à plusieurs formules, et certains ont déjà été calculés pour déterminer l'éclairement. Il suffit de les mémoriser, afin de les réutiliser sans calculs supplémentaires, pour que le coût de calcul total ne soit pas excessif, d'autant plus qu'un seul passage dans la boucle de parcours des cellules de l'hémisphère est suffisant.

B.3 Scène de test du calcul de gradient

Nous avons développé un programme de vérification de notre méthode de calcul de gradient, dans le cas d'une scène très simple (figure B.4). Elle est constituée d'un plafond lumineux, d'une sphère opaque flottante et d'un plancher. Nous analysons la répartition des éclairements indirects et du gradient au niveau du sol, avec un seul niveau de récursivité. Le plafond comporte plusieurs dalles de luminosités différentes. On simule ainsi une scène où le plancher ne serait éclairé que par la lumière réfléchie sur le plafond. La sphère flottante et les murs sont des obstacles complètement absorbants.

La figure B.5 représente une image du champ de vecteurs lumineux indirects au niveau du plancher de la scène. Chaque vecteur est symbolisé par un cylindre orienté selon la direction d'incidence moyenne $\vec{D}$, et dont la longueur est proportionnelle au terme énergétique P. Ils sont pour la plupart dirigés vers un point intermédiaire entre la dalle lumineuse la plus proche et l'autre, sauf dans la pénombre de la sphère, où les directions s'inversent. Seule la lumière parvenant par les côtés crée la pénombre sous la sphère.

La figure B.6 est l'image de l'éclairement calculé au niveau du sol, avec un échantillonnage de l'hémisphère en 512 cellules. L'absence de terme ambiant amplifie l'impression de

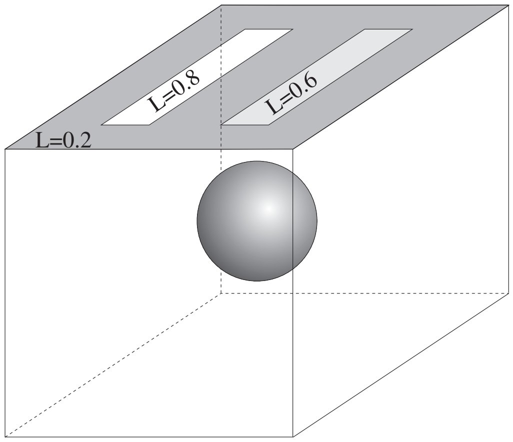


FIG. B.4 – *Scène de test*

FIG. B.5 – *Scène de test : champ de vecteurs lumineux indirects*

contraste entre les zones sombres et claires. La figure B.7 représente le gradient absolu de l'eclairement (équation B.18), calculé avec notre méthode dans le plan du sol. Les zones claires indiquent les fortes variations. Elles correspondent bien aux régions où la pénombre varie le plus rapidement, dans l'alignement des sources et dans l'ombre de la sphère.

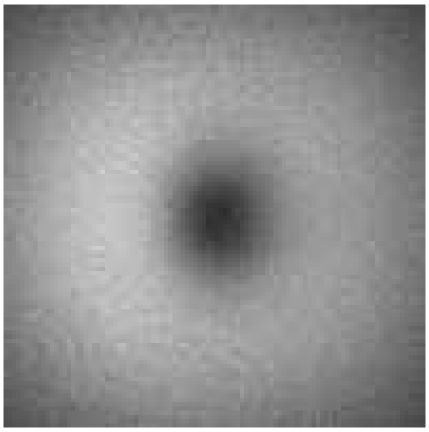


FIG. B.6 – *Eclairement du plancher*

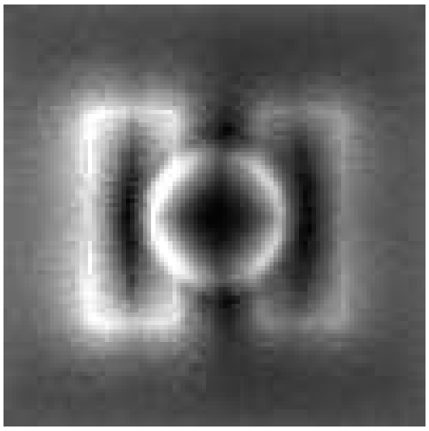


FIG. B.7 – *Gradient d'éclairement*

Annexe C

Antialiassage

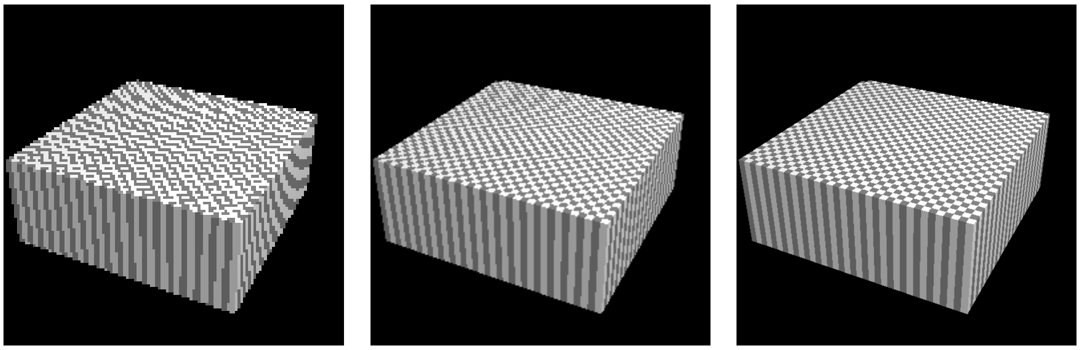
La méthode d'antialiassage décrite ici a été développée par Marlène Bougard, Nathalie Farenc et moi-même, dans le cadre du DEA Images, lors de la réalisation d'un projet, à Saint-Etienne en février 1995. Ces travaux s'inspirent directement des travaux de Witted [Whi80], qui est l'un des précurseurs de la méthode du lancer de rayons.

C.1 L'aliassage

C.1.1 Description

L'aliassage est un problème inhérent à toutes les discrétisations de signaux continus, lorsque le taux d'échantillonnage est insuffisant pour reproduire correctement les hautes fréquences. Il est causé par un repliement du spectre fréquentiel du signal. Le théorème de Shannon [Sha49], bien connu en traitement du signal, stipule que la fréquence d'échantillonnage doit être au moins le double de la fréquence maximale que l'on souhaite reproduire.

Une image, qu'elle soit de synthèse ou non, est un signal continu à deux dimensions. L'affichage sur un écran est un échantillonnage, dont la finesse est liée à la résolution de l'écran et à la taille des pixels. Il introduit donc de l'aliassage (figure C.1), qui peut perturber fortement l'aspect de l'image.



(a) 128 x *128 pixels.* *(b) 256* x *256 pixels.* *(c) 512* x *512 pixels.*

FIG. C.1 – *Images d'un damier calculées à différents taux d'échantillonnage*

En synthèse d'images, l'aliassage se manifeste essentiellement par un aspect crénelé des lignes droites obliques et par un effet de moiré lors de la reproduction de motifs réguliers. De plus, les détails dont une des dimensions est comparable à la taille d'un pixel peuvent disparaître complètement parfois, ce qui est particulièrement génant lors de la réalisation d'animations (figure C.2). Ce phénomène est beaucoup plus sensible en synthèse d'images qu'en photographie, car la taille des pixels de l'écran est incomparablement plus grande que celle des grains d'une émulsion photographique.

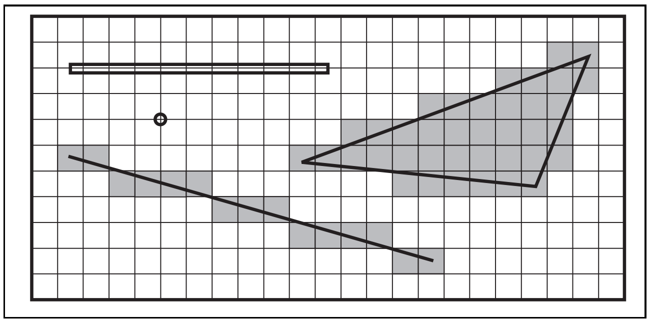


FIG. C.2 – *Exemples de disparition d'objets et de déformation des contours*

C.1.2 Principales méthodes d'antialiassage

C.1.2.1 Sur-échantillonnage

Une solution pour limiter les effets de l'aliassage, inspirée par le théorème de Shannon, est d'augmenter artificiellement le taux d'échantillonnage. L'image est calculée à une résolution plus grande que celle de l'écran, puis un lissage et une remise à l'échelle sont effectués avant l'affichage. Il se produit alors un très léger effet de flou, beaucoup moins gênant pour un observateur humain que l'aliassage.

En lancer de rayons, le sur-échantillonnage consiste à lancer plusieurs rayons à l'intérieur de chaque pixel, répartis de façon uniforme ou aléatoire. La couleur définitive affectée au pixel est alors la moyenne des couleurs obtenues pour chacun des rayons. On peut noter que cette notion de couleur moyenne n'a aucun sens physique dans un espace colorimétrique tel que le classique *rvb*. Les résultats sont cependant très bons, d'un point de vue perceptuel, car les rayons ne renvoient des valeurs très différentes que dans les zones frontières, soit en général pour un petit nombre de pixels.

C'est une méthode coûteuse, car elle augmente de façon très importante le nombre de rayons. Elle peut être fortement optimisée par un sur-échantillonnage adaptatif, c'est à dire se limitant aux endroits où les effets de l'aliassage risquent d'être visibles. Ce procédé sera détaillé plus loin.

C.1.2.2 Introduction de bruit

Il est aussi possible de perturber volontairement l'image en introduisant un léger bruit aléatoire. En lancer de rayons, c'est la méthode de *jittering* [CPC84]. Elle consiste à perturber aléatoirement la direction des rayons générés. L'aliassage est toujours présent, mais il est alors masqué par du bruit, auquel le système visuel humain est un peu moins sensible. La mise au point est assez délicate, car le niveau de bruit donnant les meilleurs résultats, tout en ne devenant pas lui-même dissuasif, dépend de l'image produite. Cette méthode est rapide, car aucun rayon supplémentaire n'est généré, mais la qualité des images produites n'est pas très satisfaisante. La présence de seulement quelques pixels erronés est suffisante pour perturber l'observateur, notamment à proximité des frontières entre les objets de couleurs contrastées.

C.1.2.3 Filtre passe-bas

Une autre possibilité est encore de lisser l'image complète, en la convoluant avec un filtre passe-bas, sans sur-échantillonnage. La valeur de chaque pixel est pondérée par une fraction de la valeur des pixels voisins. Il en résulte la disparition des transitions brusques, notamment sur les contours des objets où l'aliassage est visible, mais aussi lorsque ces transitions sont des informations utiles de l'image. Cette méthode simple et relativement rapide introduit de plus un effet de flou fortement perceptible sur toute l'image. Il semble malheureusement qu'il n'y ait pas de solution qui soit à la fois satisfaisante et peu coûteuse.

C.2 Antialiassage en post-traitement

C.2.1 Principe

Lors de la réalisation d'images fixes, l'aliassage n'est réellement gênant que dans les zones contrastées, d'un point de vue visuel. En effet, si l'on néglige arbitrairement le phénomène de disparition des petits objets, c'est l'apparition de “marches d'escaliers” qui perturbe le plus l'observateur, principalement sur les frontières entre objets de couleurs très différentes, mais aussi sur les surfaces fortement texturées. Il nous faut donc détecter les zones perturbées de l'image, pour appliquer la coûteuse procédure d'antialiassage par sur-échantillonnage uniquement lorsque l'aliassage risque d'être visible. Un calcul de gradient de luminosité permet de créer une carte des pixels à traiter, après qu'une première image ait été calculée classiquement, avec un seul rayon primaire par pixel.

Le principal avantage de cette méthode en deux passes est de fournir relativement rapidement un image provisoire aliassée. La procédure d'antialiassage est ensuite lancée, dont les paramètres peuvent être éventuellement modifiés sans remettre en cause la première phase.

C.2.2 Analyse de l'image

La détection des zones contrastée nécessite tout d'abord la création d'une image de luminosité, où la valeur (R, V, B) de chaque pixel est remplacée par la luminosité Lp, correspondant à sa valeur sur une image en niveaux de gris. Il s'agit approximativement de la composante Y de l'espace de couleur XYZ :

$$Lp = 0.299\,R + 0.587\,V + 0.114\,B$$

Ensuite, un gradient $G_{i,j}$ est calculé pour chaque pixel (i, j) de l'image. Il indique le niveau de contraste relatif de chaque pixel par rapport à ses voisins, dans un voisinage de

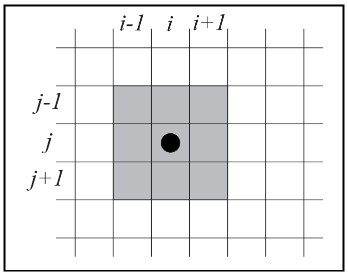


FIG. C.3 – *Voisinage V8 d'un pixel*

type $V8$ (figure C.3) :

$$\begin{cases} G_{i,j} = \dfrac{\max\limits_{x\in\{i-1,i,i+1\},y\in\{j-1,j,j+1\}} |Lp_{i,j} - Lp_{x,y}|}{\sum\limits_{x=i-1}^{i+1}\sum\limits_{y=j-1}^{j+1} Lp_{x,y}/9} & \text{si} \quad \sum\limits_{x=i-1}^{i+1}\sum\limits_{y=j-1}^{j+1} Lp_{x,y} \neq 0 \\ G_{i,j} = 0 \quad \text{sinon} & \end{cases}$$

Lorsque le gradient $G_{i,j}$ est supérieur à un seuil prédéterminé, le pixel est marqué comme devant être traité par la procédure d'anti-aliassage. De plus, afin de diminuer le risque de ne pas détecter quelques déformations des contours, la zone à traiter est étendue au voisinage $V8$ dans son entier. Ce n'est pas trop pénalisant en général, car la plupart des pixels du voisinage sont bien souvent déjà marqués, la zone courante étant perturbée.

Le choix du seuil, arbitraire, influe sur la qualité finale de l'image, ainsi que sur le nombre de pixels à subdiviser, et donc sur le nombre total de rayons à générer. En pratique, les valeurs comprises entre 0.4 et 1.2 donnent de bons résultats.

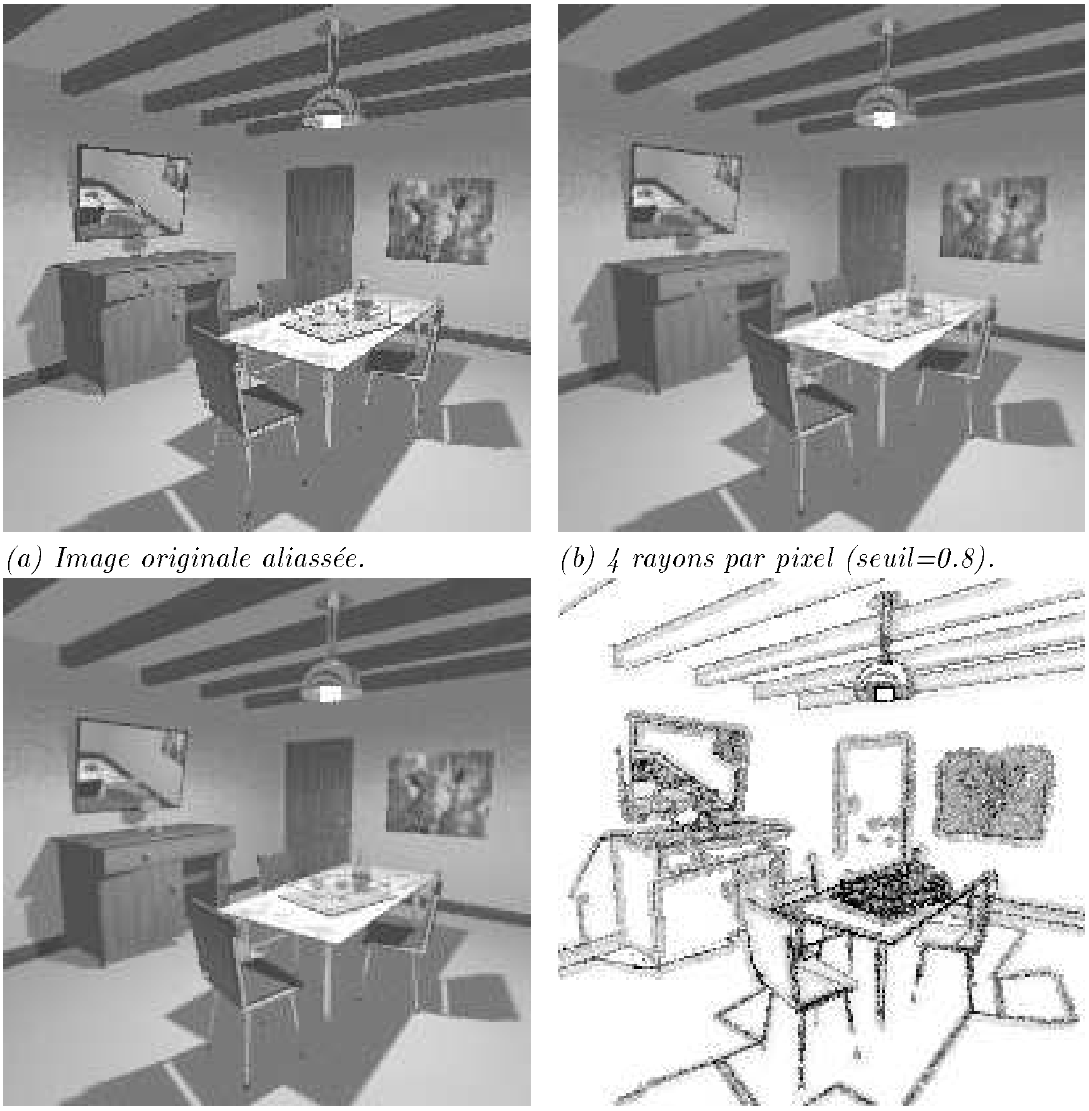
C.2.3 Sur-échantillonnage puis lissage

Les pixels marqués à l'étape précédente sont subdivisés en N sous-pixels, puis un rayon est généré en position aléatoire à l'intérieur de chacun d'eux, selon un processus d'échantillonnage stratifié. La couleur définitive affectée au pixel est obtenue en calculant la moyenne entre ces N couleurs, avec les restrictions déjà vues en ce qui concerne cette notion de couleur moyenne. Il s'agit donc d'un lissage, rudimentaire mais suffisant. La couleur initiale est ignorée, car sa prise en compte privilégierait sans justification la zone centrale du pixel.

Par commodité N est une puissance de 2, afin que la subdivision soit géométriquement très simple. En pratique, $N = 4$ donne des résultats globalement corrects, alors que $N = 9$ est préférable lorsque les zones aliassées sont fortement contrastées. Il semble inutile d'augmenter au delà cette valeur, suite aux diverses expériences que nous avons effectuées sur des simulations d'intérieurs. En effet, les différences ne sont alors pratiquement plus perceptibles visuellement.

C.3 Résultats

La figure C.4 présente un exemple d'antialiassage sur une image, avec un seuil de 0.8 et une subdivision en 4 (image (b)) et en 9 (image (c)) des pixels traités. Elles sont calculées à une



(a) Image originale aliassée.

(b) 4 rayons par pixel (seuil=0.8).

(c) 9 rayons par pixel (seuil=0.8).

(d) Image de différence entre (b) et (c).

FIG. C.4 – *Exemple d'antialiassage d'une image*

faible résolution (200x200), afin d'accentuer les effets de l'aliassage. L'image (*d*) est l'image de différence entre les images (*b*) et (*c*), la couleur blanche indiquant les pixels identiques. La figure C.5 indique pour la même image les zones contrastées détectées, avec différentes valeurs du seuil.

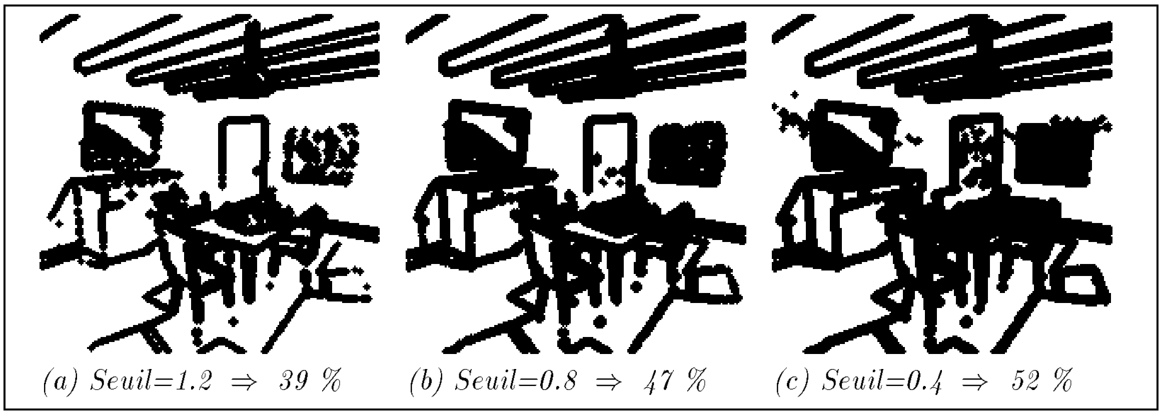


FIG. C.5 – *Detection des zones contrastées (Seuil et taux de pixels détectés)*

C.4 Commentaires

C.4.1 Disparition des petits objets

Cette méthode ne résoud pas le problème de la disparition des petits objets, car rien ne peut garantir qu'un rayon infiniment petit rencontre un objet, si celui-ci ne recouvre pas entièrement le pixel. Il faudrait avoir une connaissance explicite de leur position dans la scène pour déterminer des régions à sur-échantillonner, et de leur taille relativement à celle des pixels pour estimer le niveau de subdivision optimal. La représentation des scènes par arbre *CSG* que nous utilisons ne nous permet pas d'accéder naturellement à ces informations.

Seul le *lancer de pyramides*, dans ses versions adaptatives [Gha92], offre une solution fiable à ce problème. Chaque rayon est remplacé par une pyramide ayant l'œil pour sommet et englobant initialement le pixel. Si son intersection avec la scène est complexe, c'est à dire en présence de plusieurs objets visibles ou d'une zone texturée, elle est subdivisée récursivement. La complexité des algorithmes d'intersection pénalise sensiblement les temps de calcul. En effet, il faut détecter l'ensemble des objets partiellement visibles parmi tous ceux que contient la pyramide, et non pas le seul premier objet vu par un rayon infiniment petit. De plus, la réflexion des pyramides sur les surfaces non planes, ainsi que la réfraction dans le cas général, ne sont pas des opérations linéaires. L'image géométrique par réflexion d'une pyramide n'est pas toujours une pyramide, dans le cas où les bordures n'ont pas de sommet commun. Il faut donc prévoir des critères de subdivision supplémentaires, pour que les volumes réfléchis puissent être assimilés à des pyramides simples.

C.4.2 Développements possibles

Il est dommage de ne pas réutiliser la valeur du pixel initialement calculée. Un découpage judicieux des sous-pixels et/ou des coefficients de pondération adaptés devraient permettre de l'inclure dans la phase de lissage.

Une amélioration simple serait d'adapter automatiquement le taux de sur-échantillonnage au niveau de contraste détecté, en modifiant le nombre N de subdivisions. Le seul paramètre à founir par l'utilisateur serait alors la valeur du seuil. Par exemple :

$$\begin{cases} 0 \leq G_{i,j} < Seuil & \Rightarrow & N = 0 \\ Seuil \leq G_{i,j} < 2 * Seuil & \Rightarrow & N = 4 \\ G_{i,j} \geq 2 * Seuil & \Rightarrow & N = 9 \end{cases}$$

Le choix du seuil est peut-être lui aussi automatisable, en se plaçant dans un espace de couleur perceptuel, où la notion de contraste serait bien définie. Nous n'avons pas poussé plus avant cette étude, car il faudrait introduire une approche psycho-visuelle complexe [Rou98]. C'est toujours actuellement un domaine de recherche très actif, où de nombreux paramètres sont encore inconnus ou incompris.

Annexe D

Calcul des angles solides

Les angles solides interviennent dans la plupart des équations décrivant la propagation de la lumière. Il est parfois nécessaire d'évaluer explicitement leur valeur, ce qui n'est pas toujours aisé. Cette section propose une méthode générique, qui ne résoud hélas pas tous les cas, et une expression analytique exacte pour quelques configurations particulières. Les démonstrations s'appuient sur les livres de Berger [Ber79] et Pusterle [Pus91].

D.1 Définitions et notations

Un angle solide est l'équivalent volumique d'un angle plan. Pour un point O donné, il est une mesure de l'ouverture du cône englobant exactement une surface S (figure D.1). L'unité de mesure est le *stéradian* (sr). Une sphère vue depuis son centre sous-tend un angle solide de 4π sr.

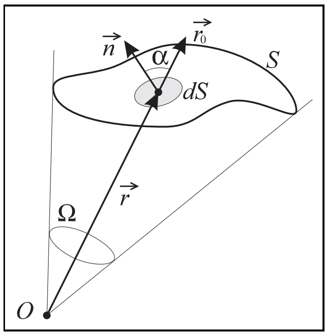


FIG. D.1 – *Géométrie d'un angle solide*

L'angle solide Ω se calcule en intégrant sur la surface S la mesure d'un angle solide élémentaire $d\omega = \frac{\cos\alpha}{r^2} dS$. Sa valeur est aussi l'aire de la surface S projetée sur une sphère unitaire centrée en O:

$$\Omega = \iint_S \frac{\cos\alpha}{r^2}\, dS \qquad \text{avec} \quad r = \|\vec{r}\| \tag{D.1}$$

Dans le cas d'une intégration sur un hémisphère échantillonné (figure D.2), que nous rencontrons souvent en synthèse d'images, l'angle solide élémentaire $d\omega$ sous-tendu par une cellule s'exprime en fonction de l'angle zénithal θ et de l'angle azimutal ϕ : $d\omega = \sin\theta\, d\theta\, d\phi$.

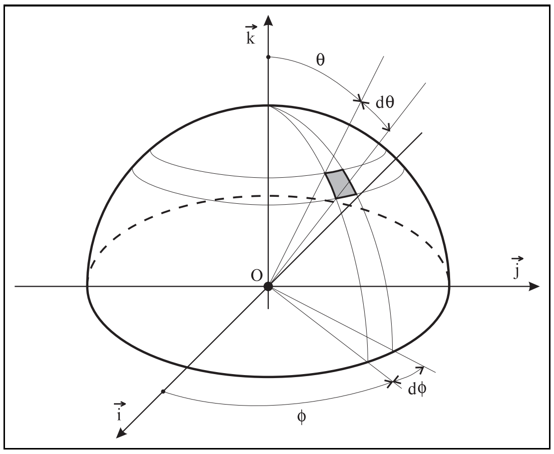


FIG. D.2 – *Angle solide élémentaire sur un hémisphère échantillonné*

D.2 Expression analytique

D.2.1 Transformation en intégrale de contour

Théorème de Stokes *Le flux du rotationnel d'un vecteur à travers une surface ouverte est égal à la circulation de ce vecteur le long du contour qui délimite cette surface.*

$$\iint_S rot(\vec{P}) \cdot \vec{n}\, dS \;=\; \oint_C \vec{P} \cdot \vec{da}$$

Ce théorème va nous permettre de transformer l'intégrale de surface D.1 en une intégrale de contour de degré inférieur, après en avoir modifié l'écriture :

$$\Omega \;=\; \iint_S \frac{\vec{r_0} \cdot \vec{n}}{r^2}\, dS \qquad \text{avec} \quad \vec{r_0} = \frac{\vec{r}}{\|\vec{r}\|} \tag{D.2}$$

L'expression D.2 indique que l'angle solide Ω est aussi le flux du champ vectoriel de valeur $\frac{\vec{r_0}}{r^2}$ à travers la surface S. Il existe alors un champ vectoriel $\vec{P}$ tel que la valeur de l'angle solide soit égale à la valeur de la circulation de $\vec{P}$ autour du contour C de la surface S. L'intégrale de surface peut donc être transformée en une intégrale de contour :

$$\Omega \;=\; \oint_C \vec{P} \cdot \vec{da}$$

Dans ce cas, le rotationnel de $\vec{P}$ doit être égal au vecteur $\frac{\vec{r_0}}{r^2}$:

$$rot(\vec{P}) \;=\; \frac{\vec{r_0}}{r^2} \tag{D.3}$$

D.2.2 Solution particulière

Une solution particulière pour $\vec{P}$ de l'équation différentielle D.3 peut être déterminée dans le cas où la surface est une calotte sphérique de rayon r. Le vecteur $\frac{\vec{r_0}}{r^2}$ est normal à

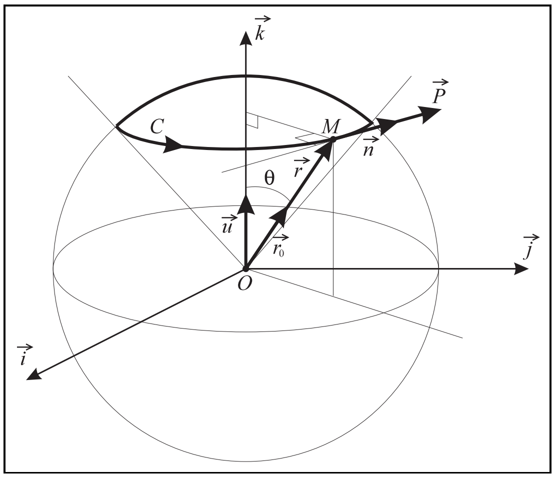


FIG. D.3 – *Cas d'une calotte sphérique*

la surface S en tout point. L'aire de cette surface est $A = 2\pi r^2(1-\cos\theta)$. Le flux de $\dfrac{\vec{r_0}}{r^2}$ à travers S est donc :

$$\Phi \;=\; 2\pi r^2(1-\cos\theta)\;\frac{1}{r^2} \;=\; 2\pi(1-\cos\theta)$$

$\vec{n}$ est le vecteur normé tel que $\vec{u}\wedge\vec{r_0} = \sin\theta\;\vec{n}$. En choisissant $\vec{P}$ tangent au cercle C suivant $\vec{n}$ et de norme constante P, la circulation est :

$$\Gamma \;=\; 2\pi r\sin\theta\, P$$

Le théorème de Stokes implique $\Phi = \Gamma$, on en déduit donc :

$$\vec{P} \;=\; \frac{1-\cos\theta}{r\sin\theta}\,\vec{n} \;=\; \frac{\sin\theta}{r(1+\cos\theta)}\,\vec{n} \;=\; \frac{1}{r}\tan\frac{\theta}{2}\,\vec{n}$$

Soit encore, sous forme vectorielle :

$$\vec{P} \;=\; \frac{\vec{u}\wedge\vec{r_0}}{r+\vec{u}\cdot\vec{r}} \tag{D.4}$$

où $\vec{u}$ est un vecteur unitaire quelconque servant uniquement de référenciel pour la mesure des angles.

D.2.3 Cas général

En utilisant le résultat D.4, le calcul d'un angle solide devient le calcul de l'intégrale de contour :

$$\Omega \;=\; \oint_C \frac{\vec{u}\wedge\vec{r_0}}{r+\vec{u}\cdot\vec{r}}\cdot\vec{da} \tag{D.5}$$

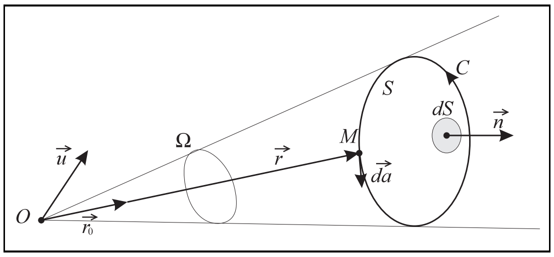


FIG. D.4 – *Transformation en intégrale de contour*

D.2.4 Expression cartésienne

Le vecteur $\vec{r}$ décrit la courbe C, et le vecteur $\vec{u}$ est un vecteur unitaire quelconque.

$$\vec{r} = \begin{pmatrix} x \\ y \\ z \end{pmatrix} \qquad \vec{u} = \begin{pmatrix} u_1 \\ u_2 \\ u_3 \end{pmatrix} \qquad r = \sqrt{x^2 + y^2 + z^2}$$

$$\vec{r_0} = \frac{\vec{r}}{r} \qquad \vec{u} \wedge \vec{r} = \begin{pmatrix} u_2 z - u_3 y \\ u_3 x - u_1 z \\ u_1 y - u_2 x \end{pmatrix} \qquad \vec{da} = \begin{pmatrix} dx \\ dy \\ dz \end{pmatrix}$$

$$\Omega = \oint_C \frac{\vec{u} \wedge \vec{r_0}}{r + \vec{u} \cdot \vec{r}} \cdot \vec{da} = \oint_C \frac{\vec{u} \wedge \vec{r}}{r^2 + r\, \vec{u} \cdot \vec{r}} \cdot \vec{da}$$

Une expression générique de l'angle solide est donc :

$$\Omega = \oint_C \frac{(u_2 z - u_3 y)dx + (u_3 x - u_1 z)dy + (u_1 y - u_2 x)dz}{x^2 + y^2 + z^2 + (u_1 x + u_2 y + u_3 z)\sqrt{x^2 + y^2 + z^2}} \tag{D.6}$$

Le choix du vecteur unitaire $\vec{u}$ peut dans certains cas conduire à des simplifications de cette expression. Il faut choisir la formulation la mieux adaptée à l'équation du contour de la surface, afin de réduire au mieux le dénominateur. Exemples :

$$\vec{u} = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} \quad \Rightarrow \quad \Omega = \oint_C \frac{y dz - z dy}{x^2 + y^2 + z^2 + x\sqrt{x^2 + y^2 + z^2}}$$

$$\vec{u} = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} \quad \Rightarrow \quad \Omega = \oint_C \frac{z dx - x dz}{x^2 + y^2 + z^2 + y\sqrt{x^2 + y^2 + z^2}}$$

$$\vec{u} = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} \quad \Rightarrow \quad \Omega = \oint_C \frac{x dy - y dx}{x^2 + y^2 + z^2 + z\sqrt{x^2 + y^2 + z^2}}$$

Le calcul analytique de cette intégrale n'est cependant pas possible dans le cas général. Il est alors nécessaire d'utiliser une méthode de calcul numérique. On peut noter que les méthodes de Monte Carlo sont bien adaptées à la résolution de ce problème.

D.3 Formulaire pour quelques cas particuliers

D.3.1 Sphère

L'angle solide sous-tendu par une sphère de rayon R, dont le centre est situé à une distance $d \geq R$ du point O est :

$$\Omega = 2\,\pi \left[1 - \sqrt{1 - \left(\frac{R}{d}\right)^2} \right] \tag{D.7}$$

Cette formule est obtenue en calculant l'aire de la calotte sphérique délimitée par la projection de la sphère sur la sphère unitaire centrée en O.

D.3.2 Polygone

D.3.2.1 Trièdre

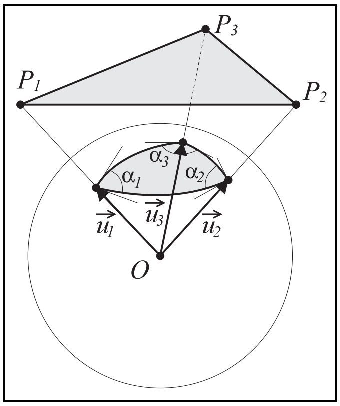


FIG. D.5 – *Angle solide sous-tendu par un trièdre*

On considère l'aire du triangle sphérique obtenu en projetant le triangle (P_1, P_2, P_3) sur la sphère unitaire centrée en O :

$$\Omega = \alpha_1 + \alpha_2 + \alpha_3 - \pi \tag{D.8}$$

avec :

$$\vec{u}_1 = \frac{\overrightarrow{OP_1}}{\|\overrightarrow{OP_1}\|} \qquad \vec{u}_2 = \frac{\overrightarrow{OP_2}}{\|\overrightarrow{OP_2}\|} \qquad \vec{u}_3 = \frac{\overrightarrow{OP_3}}{\|\overrightarrow{OP_3}\|}$$

$$\alpha_1 = \cos^{-1}\left(\frac{(\vec{u}_1 \wedge \vec{u}_2) \cdot (\vec{u}_1 \wedge \vec{u}_3)}{\| \vec{u}_1 \wedge \vec{u}_2 \| \; \| \vec{u}_1 \wedge \vec{u}_3 \|} \right)$$

$$\alpha_2 = \cos^{-1}\left(\frac{(\vec{u}_2 \wedge \vec{u}_3) \cdot (\vec{u}_2 \wedge \vec{u}_1)}{\| \vec{u}_2 \wedge \vec{u}_3 \| \; \| \vec{u}_2 \wedge \vec{u}_1 \|} \right)$$

$$\alpha_3 = \cos^{-1}\left(\frac{(\vec{u}_3 \wedge \vec{u}_1) \cdot (\vec{u}_3 \wedge \vec{u}_2)}{\| \vec{u}_3 \wedge \vec{u}_1 \| \; \| \vec{u}_3 \wedge \vec{u}_2 \|} \right)$$

D.3.2.2 Polygone quelconque

La formule (D.8) se généralise facilement pour un polygone à N côtés, même non plan :

$$\Omega \;=\; \left(\sum_{i=1}^{N} \alpha_i\right) \;-\; (N-2)\,\pi \tag{D.9}$$

avec :

$$\alpha_i = \cos^{-1}\left(\frac{(\vec{u_i} \wedge \overrightarrow{u_{i+1}}) \;\cdot\; (\vec{u_i} \wedge \overrightarrow{u_{i-1}})}{\|\,\vec{u_i} \wedge \overrightarrow{u_{i+1}}\|\;\|\,\vec{u_i} \wedge \overrightarrow{u_{i-1}}\|}\right)$$

Cette formule permet en pratique de résoudre la plupart des cas, éventuellement après une discrétisation du contour de la surface.

D.3.3 Disque centré sur un axe

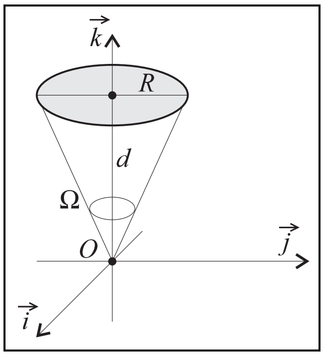


FIG. D.6 – *Angle solide sous-tendu par un disque centré sur un axe*

L'angle solide sous-tendu par un disque de rayon R, situé à une distance d du point O, et centré sur un axe passant par O est :

$$\Omega \;=\; 2\,\pi\,\left(1 \;-\; \frac{d}{\sqrt{R^2+d^2}}\right) \tag{D.10}$$

Cette formule est obtenue en considérant l'aire de la calotte sphérique délimitée par la projection du disque sur la sphère unitaire centrée en O.

Remarque Il n'existe pas de solution analytique exploitable lorsque le disque est vu depuis un point quelconque situé en dehors de l'axe central, dès lors que sa projection devient une ellipse.

Bibliographie

[AH93] L. Aupperle and P. Hanrahan, *A hierarchical illumination algorithm for surfaces with glossy reflection*, Computer Graphics **27** (1993), 155–162.

[AMP98] D. Arques, S. Michelin, and B. Piranda, *Extending the zonal method to specular surfaces*, Proceedings of WSCG'98 (1998), 27–34.

[App68] Arthur Appel, *Some techniques for shading machine rendering*, AFIPS 1968 Spring Joint Computer Conference **32** (1968), 37–45.

[Arv86] James Arvo, *Backward ray tracing*, Developements in Ray Tracing. ACM SIGGRAPH '86 Course Notes **12** (1986), 259–263.

[Arv94] James Arvo, *The irradiance jacobien for partially occluded polyhedral sources*, Computer Graphics **28** (1994), 343–350.

[Arv95] James Arvo, *Applications of irradiance tensors to the simulation of non-lambertian phenomena*, Computer Graphics **29** (1995), 335–342.

[Ash93] I. Ashdown, *Near-field photometry: a new approach*, Journal of the Illuminating Engineering Society (1993), 163–180.

[Bei88] Michel Beigbeder, *Un développement pour la modélisation et la visualisation en synthèse d'images : Castor*, Ph.D. thesis, Ecole des Mines de Saint-Etienne, April 1988.

[Ben75] Jon Louis Bentley, *Multidimensional binary search trees used for associative searching*, Communications of the ACM **18** (1975), no. 9, 509–517.

[Ben90] J. L. Bentley, *K-d trees for semidynamic point sets*, Proceedings of the 6th Annual Symposium on Computational Geometry (SCG '90) (Berkeley, CA) (ACM-SIGACT ACM-SIGGRAPH, ed.), ACM Press, June 1990, pp. 187–197.

[Ber79] Marcel Berger, *Géometrie, la sphère pour elle-même*, vol. 5, Cedic-Nathan, Paris, France, 1979.

[BKP91] Jeffrey C. Beran-Koehn and Mark J. Pavicic, *A Cubic Tetrahedral Adaptation of the Hemi-Cube Algorithm*, Graphics Gems II (Boston, MA) (James Arvo, ed.), Academic Press Professional, Boston, MA, 1991, pp. 299–302.

[Bli77] James F. Blinn, *Models of light reflection for computer synthesized pictures*, Computer Graphics **11** (1977), no. 2, 192–198.

[Bli78] James F. Blinn, *Simulation of wrinkled surfaces*, Computer Graphics **12** (1978), no. 3, 286–292.

[BN76] J.F. Blinn and M.E. Newell, *Texture and reflection in computer generated images*, Communications of the ACM **19** (1976), no. 10, 542–547.

[BS63] P. Beckmann and A. Spizzichino, *The scattering of electromagnetic waves from rough surfaces*, Pergamon Press, 1963.

[Cal98] P. Callet, *Couleur-lumière, couleur-matière*, Diderot Editeur, France, 1998.

[Cat74] E. Catmull, *A subdivision algorithm for computer display curved surfaces*, Ph.D. thesis, University of Utah, December 1974.

[CB96] J-M. Coulomb and C. Brusque, *Propriétés de réflexion des surfaces, le goniore-flectomètre du lcpc, performances et perspectives d'études*, Bulletin des Laboratoires des Ponts et Chaussées **204** (1996), no. 4056, 25–35.

[CCWG88] M. Cohen, S. Chen, J. Wallace, and D. Greenberg, *A progressive refinement approach to fast radiosity image generation*, Computer Graphics **22** (1988), no. 3, 75–84.

[CG85] M. Cohen and D. Greenberg, *The hemi-cube: A radiosity solution of complex environments*, Computer Graphics **19** (1985), no. 3, 31–41.

[CMRT91] S. Chen, G. Miller, H. Rushmeier, and D. Turner, *A progressive multi-pass method for global illumination*, Computer Graphics **25** (1991), no. 4, 165–174.

[Col94] Steven Collins, *Adaptative splatting for specular to diffuse light transport*, Fifth Eurographics Workshop on Rendering (1994), 119–135.

[Col95] Steven Collins, *Reconstruction of illumination from area luminaires*, Rendering Techniques (Proceedings of the Sixth Eurographics Workshop on Rendering) (1995), 274–283.

[Col97] Steven Collins, *Wavefront tracking for global illumination solutions*, Ph.D. thesis, Trinity College Dublin, January 1997.

[CPC84] R. L. Cook, T. Porter, and L. Carpenter, *Distributed ray tracing*, Computer Graphics **18** (1984), no. 3, 137–144.

[CT81] R. L. Cook and K. E. Torrance, *A reflectance model for computer graphics*, Computer Graphics **15** (1981), no. 3, 307–316.

[CW93] M.F. Cohen and J.R. Wallace, *Radiosity and realistic image synthesis*, Academic Press, Cambridge, 1993.

[DG95] J-M. Dischler and D. Ghazanfarpour, *A geometrical based method for highly complex structured textures generation*, Computer Graphics Forum **14** (1995), no. 4, 203–215.

[Dut96] Philip Dutré, *Mathematical frameworks and monte carlo algorithms for global illumination in computer graphics*, Ph.D. thesis, Katholieke Universiteit Leuven, September 1996.

[DW94] Philip Dutre and Yves D. Willems, *Importance-Driven Monte Carlo Light Tracing*, Fifth Eurographics Workshop on Rendering (Darmstadt, Germany), June 1994, pp. 185–194.

[DW95] Philip Dutre and Yves D. Willems, *Potential-Driven Monte Carlo Particle Tracing for Diffuse Environments with Adaptive Probability Density Functions*, Rendering Techniques '95 (Proceedings of the Sixth Eurographics Workshop on Rendering) (New York, NY) (P. M. Hanrahan and W. Purgathofer, eds.), Springer-Verlag, 1995, pp. 306–315.

[FLM96] M. Fontoynont, P. Laforgue, and R. Mitanchey, *Precise daylighting calculations in architecture and city planning*, Proceedings of the International Daylighting Workshop: building with daylight, Energy Efficient Design, Perth, Australie (1996), 6–16.

[Gar84] G. Gardner, *Simulation of natural scenes using textured quadric surfaces*, Computer Graphics **18** (1984), no. 3, 11–20.

[GD95] D. Ghazanfarpour and J-M. Dischler, *Spectral analysis for automatic 3d texture generation*, Computer and Graphics **19** (1995), no. 3, 413–422.

[Gha92] Djamchid Ghazanfarpour, *Visualisation réaliste par lancer de pyramides et subdivision adaptative*, Actes de MICAD 92, Paris (1992).

[Gla95] Andrew S. Glassner, *Principles of digital image synthesis*, Morgan Kaufmann Publishers, San Francisco, California, 1995.

[Gou71] Henri Gouraud, *Continuous shading of curved surfaces*, IEEE Transactions on Computers **20** (1971), no. 6, 623–629.

[GP91] Djamchid Ghazanfarpour and Bernard Péroche, *A high quality filtering using forward texture mapping*, Computer and Graphics **15** (1991), no. 4, 569–577.

[GTGB84] C. M. Goral, K. E. Torrance, D. P. Greenberg, and B. Battaile, *Modeling the interaction of light between diffuse surfaces*, Computers Graphics **18** (1984), no. 3, 213–222.

[Hec90] Paul S. Heckbert, *Adaptative radiosity textures for bidirectional ray tracing*, Computer Graphics **24** (1990), no. 4, 145–154.

[HH64] J.M. Hammersley and D.C. Handscomb, *Monte carlo methods*, Chapman and Hall, London, Great Britain, 1964.

[HSA91] P. Hanrahan, D. Salzman, and L. Aupperle, *A rapid hierarchical radiosity algorithm*, Computers Graphics **25** (1991), no. 4, 197–206.

[HTSG91] X. D. He, K. E. Torrance, F. X. Sillion, and D. P. Greenberg, *A comprehensive physical model for light reflection*, Computer Graphics **25** (1991), no. 4, 175–186.

[JC98] Henrik Wann Jensen and Per H. Christensen, *Efficient simulation of light transport in scenes with participating media using photon maps*, Computer Graphics Proceedings, Annual Conference Series, 1998 (SIGGRAPH '98), July 1998, pp. 311–320.

[Jen95] Henrik Wann Jensen, *Importance Driven Path Tracing Using the Photon Map*, Rendering Techniques '95 (Proceedings of the Sixth Eurographics Workshop on Rendering) (New York, NY) (P. M. Hanrahan and W. Purgathofer, eds.), Springer-Verlag, 1995, pp. 326–335.

[Jen96] H. W. Jensen, *Global illumination using photon maps*, Seventh Eurographics Workshop on Rendering, Porto (1996), 22–31.

[Jen97] H. W. Jensen, *Rendering caustics on non-lambertian surfaces*, Computer Graphics Forum **16** (1997), no. 1, 57–64.

[JvW83] F. W. Jansen and J. J. van Wijk, *Fast previewing techniques in raster graphics*, Eurographics '83 (P. J. W. ten Hagen, ed.), North-Holland, 1983, pp. 195–202.

[Kaj86] J. T. Kajiya, *The rendering equation*, Computer Graphics **20** (1986), no. 4, 143–150.

[KJ91] A. Kok and F. W. Jansen, *Source selection for the direct lighting computation in global illumination*, Second Eurographics Workshop on Rendering, Barcelona, Spain (1991), 75–82.

[KJ92] A. Kok and F. W. Jansen, *Adaptive sampling of area light sources in ray tracing including diffuse interreflection*, Computer Graphics Forum (Eurographics'92) **11** (1992), no. 3, 289–298.

[Laf96] Eric Lafortune, *Mathematical models and monte carlo algorithms for physically based rendering*, Ph.D. thesis, Katholieke Universiteit Leuven, February 1996.

[Lew93] R.R. Lewis, *Making shaders more physically plausible*, Fourth Eurographics Workshop on Rendering (1993), 47–62.

[LFTG97] E. Lafortune, S. Foo, K. Torrance, and D. Grennberg, *Non-linear approximation of reflectance functions*, Computer Graphics **31** (1997), no. 3, 117–126.

[LM97] C. Lemaire and J.M. Moreau, *Amélioration de la complexité en moyenne de l'algorithme de lee et schachter par division et fusion bi-directionnelles*, Revue internationale de CFAO et d'informatique graphique **12** (1997), no. 4, 317–336.

[LTG92] D. Lischinski, F. Tampieri, and D. Greenberg, *Discontinuity meshing for accurate radiosity*, IEE Computer Graphics and Applications **12** (1992), no. 6, 25–39.

[LW94] E. P. Lafortune and Y. D. Willems, *A theorical framework for physically based rendering*, Computer Graphics Forum **13** (1994), no. 2, 97–107.

[LW95] E. P. Lafortune and Y. D. Willems, *A 5d tree to reduce the variance of monte carlo ray tracing*, Sixth Eurographics Workshop on Rendering, Dublin, Ireland (1995), 11–20.

[Mai96] J.-L. Maillot, *Pseudo-réalisme et progressivité pour le tracé de rayons*, Ph.D. thesis, Ecole des Mines de Saint-Etienne, September 1996.

[MCP92] J.-L. Maillot, L. Carraro, and B. Peroche, *Progressive ray tracing*, Third Eurographics Workshop on Rendering, Bristol, England (1992), 9–19.

[MH92] Don P. Mitchell and Pat Hanrahan, *Illumination from curved reflectors*, Computer Graphics SIGGRAPH'92 **26** (1992), 283–291.

[Mit96] Richard Mitanchey, *Synthèse d'images appliquée à l'éclairage intérieur des bâtiments*, Ph.D. thesis, Ecole des Mines de Saint-Etienne, January 1996.

[NN85] T. Nishita and E. Nakamae, *Continuous tone representation of three-dimensional objects taking account of shadows and interreflection*, Computer Graphics **19** (1985), no. 3, 23–30.

[NRH+77] F. Nicodemus, J. Richmond, J. Hsia, I. Ginsberg, and T. Limperis, *Geometric considerations and nomenclature for reflectance*, Monograph 160, National Bureau of Standards, 1977.

[PB95] S.N. Pattanaik and K. Bouatouch, *Equations adjointes et tracé de particules pour le calcul d'illumination globale*, Rapport de recherche INRIA, IRISA Rennes, France (1995), no. 2474, 1–25.

[Per85] Ken Perlin, *An image synthesizer*, Computer Graphics **19** (1985), no. 3, 287–296.

[PGMR98] B. Péroche, D. Ghazanfarpour, D. Michelucci, and M. Roelens, *Informatique graphique, méthodes et modèles*, Editions Hermes, Paris, France, 1998, 2ème edition.

[PH89] K. Perlin and E. Hoffert, *Hypertexture*, Computer Graphics **23** (1989), no. 3, 253–262.

[Pho75] B. T. Phong, *Illumination for computer generated picture*, Communications of the ACM **18** (1975), no. 6, 311–317.

[PM93] S.N. Pattanaik and S.P. Mudur, *Efficient potential equation solutions for global illumination computation*, Computer and Graphics **17** (1993), no. 4, 387–396.

[PM95] S.N. Pattanaik and S.P. Mudur, *Adjoint equations and random walks for illumination computation*, ACM Transactions on Graphics **14** (1995), 77–102.

[PPS97] F. Perez, X. Pueyo, and F.X. Sillion, *Global illumination techniques for the simulation of participating media*, 8th Eurographics Workshop on Rendering, St-Etienne, France (1997), 309–320.

[Pus91] Charles Pusterle, *Analyse vectorielle des champs*, Masson, Paris, France, 1991.

[Ris96] Philippe Ris, *Parallélisation du lancer de rayons par évaluation dynamique de la topologie de la scène*, Ph.D. thesis, Université de Franche-Comté, 1996.

[RKJ96] Erik Reinhard, Arjan J. F. Kok, and Frederik W. Jansen, *Cost prediction in ray tracing*, Eurographics Rendering Workshop 1996 (New York City, NY) (Xavier Pueyo and Peter Schröder, eds.), Eurographics, Springer Wein, June 1996, pp. 41–50.

[Roe93] Marc Roelens, *Un environnement pour le tracé de rayons utilisant une modélisation par arbre de construction*, Ph.D. thesis, Ecole des Mines de Saint-Etienne, February 1993.

[Rou98] Gilles Rougeron, *Problèmes liés à la couleur en synthèse d'images*, Ph.D. thesis, Ecole des Mines de Saint-Etienne, January 1998.

[RP97] G. Rougeron and B. Péroche, *An adaptive representation of spectral data for reflectance computations*, 8th Eurographics Workshop on Rendering, St-Etienne, France (1997), 127–138.

[Rub81] Reuven Y. Rubinstein, *Simulation and the monte carlo method*, John Wiley and Sons, New York, USA, 1981.

[Sam84] H. Samet, *The quadtree and related hierachical data structures*, Computing Surveys **16** (1984), no. 2, 187–260.

[SAWG91] F. X. Sillion, J. Arvo, S. H. Westin, and D. P. Greenberg, *A global illumination solution for general reflectance distribution*, Computer Graphics **25** (1991), no. 4, 187–196.

[Sch92] C. Schlick, *Divers éléments pour une synthèse d'images réaliste*, Ph.D. thesis, Université de Bordeaux I, November 1992.

[Sch94] C. Schlick, *An inexpensive brdf model for physically-based rendering*, Computer Graphics Forum **13** (1994), no. 3, 149–162.

[Sha49] C. Shannon, *Communication in the presence of noise*, Proceedings IRE **37** (1949), 10–21.

[Shi90a] P. Shirley, *Physically based lighting calculations for computer graphics*, Ph.D. thesis, Univ. of Illinois, November 1990.

[Shi90b] P. Shirley, *A ray tracing method for illumination calculation in diffuse-specular scenes*, Proceedings of Graphics Interface'90 (1990), 205–212.

[Sil89] François Sillion, *Simulation de l'éclairage pour la synthèse d'images : réalisme et interactivité*, Ph.D. thesis, Université de Paris-Sud, June 1989.

[Sil95] Francois X. Sillion, *A Unified Hierarchical Algorithm for Global Illumination with Scattering Volumes and Object Clusters*, IEEE Transactions on Visualization and Computer Graphics **1** (1995), no. 3, 240–254.

[Smi67] B.G. Smith, *Geometrical shadowing of a random rough surface*, IEEE Transactions of Antennas and Propagation **15** (1967), no. 5, 668–671.

[SP89] F. Sillion and C. Puech, *A general two-pass method for integrating specular and diffuse reflection*, Computer Graphics **23** (1989), no. 3, 335–344.

[SP94] F.X. Sillion and C. Puech, *Radiosity and global illumination*, Morgan Kaufmann Publishers, San Francisco, California, 1994.

[SWZ96] P. Shirley, C. Wang, and K. Zimmerman, *Monte carlo techniques for direct lighting calculations*, ACM Transactions on Graphics **15** (1996), no. 1, 1–36.

[TNK+93] Katsumi Tadamura, Eihachiro Nakamae, Kazufumi Kaneda, Masshi Baba, Hideo Yamashita, and Tomoyuki Nishita, *Modeling of skylight and rendering of outdoor scenes*, Eurographics '93 (Oxford, UK) (R. J. Hubbold and R. Juan, eds.), Eurographics, Blackwell Publishers, 1993, pp. 189–200.

[TS67] K. Torrance and E. Sparrow, *Theory for off-specular reflection from roughened surfaces*, Journal of the Optical Society of America **57** (1967), no. 9, 1105–1114.

[Vea97] Eric Veach, *Robust monte carlo methods for light transport simulation*, Ph.D. thesis, Stanford University, December 1997.

[Ved93] Christophe Vedel, *Calcul de l'illumination par des sources etendues : Espace objet et espace image*, Ph.D. thesis, Ecole Polytechnique, September 1993.

[VG95] E. Veach and L. J. Guibas, *Optimally combining sampling techniques for Monte Carlo rendering*, Computer Graphics **29** (1995), no. Annual Conference Series, 419–428.

[VG97] E. Veach and L. Guibas, *Metropolis light transport*, Computer Graphics **31** (1997), 65–76.

[War91] G. J. Ward, *Adaptive shadow testing for ray tracing*, Second Eurographics Workshop on Rendering, Barcelona, Spain (1991), 11–20.

[War92] G. J. Ward, *Measuring and modeling anisotropic reflection*, Computer Graphics **26** (1992), no. 2, 265–272.

[War94] G. J. Ward, *The radiance lighting simulation and rendering system*, Computer Graphics (1994), 459–472.

[WCG87] J. R. Wallace, M. F. Cohen, and D. P. Greenberg, *A two-pass solution to the rendering equation : a synthesis of ray-tracing and radiosity methods*, Computer Graphics **21** (1987), no. 4, 311–320.

[WH92] G. J. Ward and P. S. Heckbert, *Irradiance gradients*, Third Eurographics Workshop on Rendering, Bristol, England (1992), 85–98.

[Whi80] T. Whitted, *An improved illumination model for shaded display*, Communications of the ACM **23** (1980), 343–349.

[WRC88] G. J. Ward, F. M. Rubinstein, and R. D. Clear, *A ray tracing solution for diffuse interreflection*, Computer Graphics **22** (1988), no. 4, 85–92.

[ZP97] J. Zaninetti and B. Péroche, *Modélisation vectorielle des éclairements indirects en lancer de rayons*, Revue internationale de CFAO et d'informatique graphique **12** (1997), no. 4, 441–453.

[ZS95] Kurt Zimmerman and Peter Shirley, *A Two-Pass Realistic Image Synthesis Method for Complex Scenes*, Rendering Techniques '95 (Proceedings of the Sixth Eurographics Workshop on Rendering) (New York, NY) (P. M. Hanrahan and W. Purgathofer, eds.), Springer-Verlag, 1995, pp. 284–295.

[ZSP98] J. Zaninetti, X. Serpaggi, and B. Péroche, *A vector approach for global illumination in ray tracing*, Computer Graphics *Forum* (Proceedings of Eurographics'98) **17** (1998), no. 3, 149–158.

Table des figures

Table des matières