

UNIVERSIDADE FEDERAL DO RIO DE JANEIRO
INSTITUTO DE MATEMÁTICA
CURSO DE BACHARELADO EM CIÊNCIA DA COMPUTAÇÃO

VICTOR LAMOSA BAUER

IMPLEMENTANDO UM PATH TRACER
INTERATIVO

RIO DE JANEIRO

2022

VICTOR LAMOSA BAUER

IMPLEMENTANDO UM PATH TRACER
INTERATIVO

Trabalho de conclusão de curso de graduação apresentado ao Departamento de Ciência da Computação da Universidade Federal do Rio de Janeiro como parte dos requisitos para obtenção do grau de Bacharel em Ciência da Computação.

Orientador: Prof. Paulo Roma Cavalcanti, D.Sc.

RIO DE JANEIRO

2022

CIP - Catalogação na Publicação

B344i Bauer, Victor Lamosa
Implementando um path tracer interativo / Victor
Lamosa Bauer. -- Rio de Janeiro, 2022.
75 f.

Orientador: Paulo Roma Cavalcanti.
Trabalho de conclusão de curso (graduação) -
Universidade Federal do Rio de Janeiro, Instituto
de Computação, Bacharel em Ciência da Computação,
2022.

1. Path tracing. 2. GPU. 3. Interativo. 4.
Physically based rendering. 5. Ray tracing. I.
Cavalcanti, Paulo Roma, orient. II. Título.

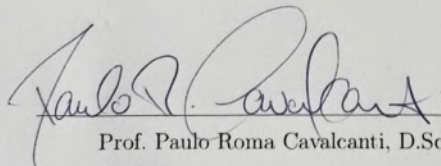
VICTOR LAMOSA BAUER

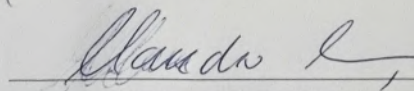
IMPLEMENTANDO UM PATH TRACER
INTERATIVO

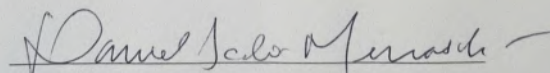
Trabalho de conclusão de curso de graduação apresentado ao Departamento de Ciência da Computação da Universidade Federal do Rio de Janeiro como parte dos requisitos para obtenção do grau de Bacharel em Ciência da Computação.

Aprovado em 14 de dezembro de 2022.

BANCA EXAMINADORA:


Prof. Paulo Roma Cavalcanti, D.Sc.


Prof. Claudio Esperança, D.Sc.


Prof. Daniel Sadoc Menasché, D.Sc.

AGRADECIMENTOS

Primeiramente, gostaria de agradecer aos meus pais; pela paciência e apoio diante todas as decisões que tomei até aqui.

Também gostaria de agradecer ao corpo docente e funcionários da Universidade Federal do Rio de Janeiro, por seus serviços prestados a todos que por aqui passaram.

Por último, gostaria de agradecer a todos que dedicaram tempo e esforço para nos fornecer excelentes materiais didáticos.

RESUMO

Por muito tempo, *path tracing* foi considerado caro demais para ser utilizado em renderização em tempo real, devido a sua complexidade computacional. Com o avanço da tecnologia de hardware e pesquisas inovadoras, parece que finalmente chegamos a um ponto onde é possível. Neste trabalho, irei estabelecer os fundamentos e técnicas utilizadas para implementar um *path tracer* interativo.

Palavras-chave: Path Tracer; Tempo Real; Renderização Baseada em Física.

ABSTRACT

For the longest time, path tracing has been deemed too expensive to be useable in real-time rendering, due to its computational complexity. With the advancement of hardware technology and innovative research, it seems we've finally reached a point where it's possible to do so with moderate success. In this work, I will lay down the foundations and techniques used in order to implement an interactive path tracer.

Keywords: Path Tracer; Real Time; Physically Based Rendering.

LISTA DE ABREVIATURAS E SIGLAS

PBR	Physically Based Rendering
ELF	Extreme Low Frequency
BRDF	Bidirectional Reflectance Distribution Function
MMC	Método de Monte Carlo
FDP	Função de Densidade de Probabilidade
EQM	Erro Quadrático Médio
IS	Importance Sampling
MIS	Multiple Importance Sampling
BVH	Bounding Volume Hierarchy
SAH	Surface Area Heuristic

SUMÁRIO

1	INTRODUÇÃO	9
2	PHYSICALLY BASED RENDERING	10
2.1	FOTORREALISMO	11
2.1.1	Modelagem da Luz	11
2.2	COLORIMETRIA	13
2.3	RADIOMETRIA	14
2.3.1	Energia Radiante	14
2.3.2	Fluxo Radiante	15
2.3.3	Irradiância	15
2.3.4	Intensidade Radiante	16
2.3.5	Radiância	17
2.3.6	Radiância Incidente e Sainte	18
3	TRANSPORTE DA LUZ	21
3.1	RAY TRACING	21
3.1.1	Câmera	22
3.2	SUPERFÍCIES	24
3.2.1	Modelo de Micro-superfícies	25
3.2.2	Dispersão Sob a Superfície	26
3.3	EQUAÇÃO DE RENDERIZAÇÃO	28
3.3.1	BRDF	30
3.3.2	Função de Distribuição Normal	32
3.3.3	Sombreamento da Micro-geometria	35
3.3.4	Fresnel	37
4	AMOSTRAGEM	40
4.1	MÉTODO DE MONTE CARLO	40
4.1.1	Estimadores	40
4.2	INTEGRAÇÃO DE MONTE CARLO	42
4.3	PATH TRACING	44

4.4	IMPORTANCE SAMPLING	47
4.4.1	Gerando Amostras	49
4.4.2	Iluminação Direta	49
4.4.3	Iluminação Indireta	52
4.5	MULTIPLE IMPORTANCE SAMPLING	59
4.5.1	Heurística Balanceada	61
4.5.2	Heurística da Potência	61
5	ESTRUTURAS DE ACELERAÇÃO	63
5.1	PARTICIONAMENTO DE OBJETOS	63
5.1.1	Bounding Box	63
5.1.2	Hierarquia de Volumes	64
5.1.3	Heurística da Área de Superfície	66
6	RESULTADOS E CONCLUSÕES	70
	REFERÊNCIAS	74

1 INTRODUÇÃO

Renderização pode ser descrita como um processo em que o objetivo é produzir uma imagem a partir de uma descrição tridimensional de uma cena [9]. A criação de imagens realistas é uma grande força motriz no ramo da computação gráfica desde sua concepção, no início da década de 1960. A busca por um maior realismo foi um importante fator para o ramo, inspirando importantes inovações [5].

No início da década de 70, os problemas mais importantes a serem resolvidos envolviam questões fundamentais como determinar a visibilidade de superfícies, e a representação de primitivas geométricas [9]. A quantidade de memória disponível era limitada e cara, tornando qualquer operação de renderização mais complexa, inviável. Conforme os computadores foram evoluindo, ficando mais potentes e acessíveis, tornou-se possível considerar abordagens computacionalmente mais exigentes para renderização.

Essa busca por um maior realismo também impulsionou a indústria de *hardware*, que buscava atender uma demanda crescente por aplicações interativas que exigiam cada vez mais recursos gráficos. Além dos componentes físicos, pesquisas inovadoras foram elaboradas para a área, permitindo uma melhor compreensão de como abordar o problema de renderizar a realidade.

Este trabalho pretende implementar um *path tracer* interativo. Para tal, iremos elaborar as bases teóricas do *Physically Based Rendering* e como acelerar o processo de renderização. Também veremos como resolver o problema da equação do transporte da luz pelo cenário, e técnicas para diminuir a quantidade de ruído nas imagens geradas.

2 PHYSICALLY BASED RENDERING

Em 1980, pesquisadores buscavam criar uma base teórica sólida para a renderização; uma que pudesse modelar de forma precisa as interações físicas entre a luz e a matéria. *Physically Based Rendering* (PBR)¹ é um conjunto de técnicas de renderização que visa simular o comportamento real da luz. Como o PBR visa imitar o comportamento da luz de uma maneira fisicamente plausível, ele parece mais realista comparado a modelos de sombreado tradicionais, como Phong e Blinn-Phong [4].

Outra grande vantagem é que materiais criados para um pipeline PBR são baseados em parâmetros físicos reais, fazendo com que a iluminação pareça correta sem precisar recorrer a truques e manipulações da cena (Figura 1). Importante ressaltar que as técnicas são meramente *baseadas* na Física. Portanto, os resultados são uma aproximação da realidade.

Figura 1: Materiais criados para um pipeline PBR parecem corretos em diferentes cenas, independente das condições de iluminação presentes.



Fonte: geekfaner.com/unity/images/506.PNG

¹Em tradução livre: Renderização Baseada na Física

2.1 FOTORREALISMO

Em seu artigo *Three Varieties of Realism in Computer Graphics* [5], James Ferwerda descreve três categorias quanto ao grau de realismo a qual podemos classificar uma imagem e quais critérios devem ser satisfeitos para tal. Dentre estes, podemos listar:

- Realismo físico
- Fotorrealismo
- Realismo funcional

Quando se trata de fotorrealismo, que será o foco deste trabalho, isso quer dizer que estamos interessados em criar uma imagem que seja indistinguível de uma fotografia. Apesar de não existir um consenso no meio acadêmico do porquê uma imagem pode ser considerada “real”, podemos ao menos partir de uma definição mais concreta ao dizer que uma imagem é fotorrealista se a mesma possa ser medida por fotometria [5].

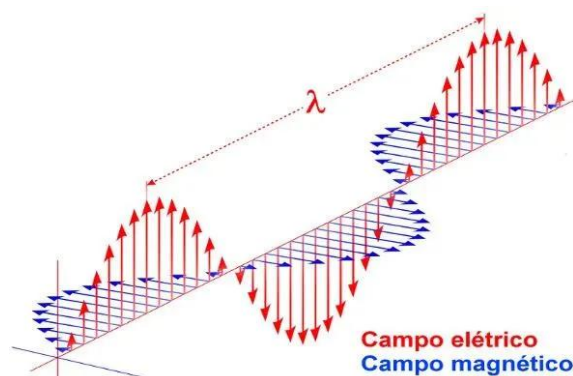
Fotometria é o ramo da óptica que procura medir a luz em relação a como seu brilho é percebido pelo olho humano. Isso quer dizer que uma imagem fotorrealista deve prover o mesmo estímulo visual que o cenário representado, mesmo que a energia provinda da imagem não seja fidedigna. Ao adotar este critério, podemos nos aproveitar da limitação visual do observador e simplificar o processo de renderização. A natureza tricromática da visão humana nos permite descrever a representação espectral da luz visível a partir de uma representação digital, geralmente descrita por um sistema de cores RGB ou CYMK.

2.1.1 Modelagem da Luz

A interação entre a luz e a matéria é a fundação do PBR. Para gerar imagens fotorrealistas, é necessário simular as propriedades e comportamentos luz. Na física óptica, a luz é modelada como uma onda eletromagnética transversal que oscila

campos elétricos e magnéticos, perpendiculares, na direção de propagação. A configuração mais simples que uma onda pode formar é a de um único comprimento de onda uniforme, denotado pela letra λ (Figura 2). Por essa razão, luzes compostas por um único comprimento de onda são ditas monocromáticas, resultando em uma única cor percebida, enquanto luzes compostas por múltiplos comprimentos de onda são ditas policromáticas (e.g. 'luz branca').

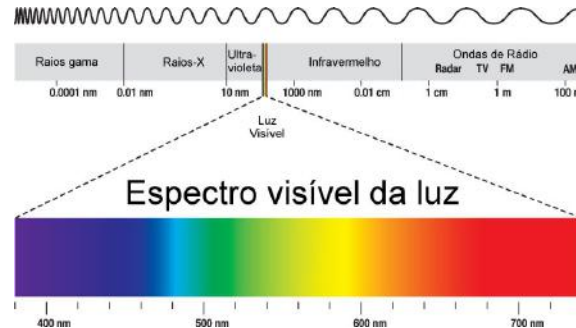
Figura 2: A luz, modelada como uma onda eletromagnética transversal. Os campos elétrico e eletromagnético oscilam perpendicularmente na direção de propagação. A onda retratada é a forma mais simples que a luz pode tomar; ela é tanto monocromática quanto linearmente polarizada.



Fonte: mundoeducacao.uol.com.br

Na natureza existem ondas eletromagnéticas em uma enorme faixa de frequência e comprimento de onda. Desde raios gama com um comprimento de onda menor que um centésimo de um nanômetro, até ondas de rádio de frequência extremamente baixa (ELF), com comprimentos maiores que milhares de quilômetros [9]. A luz visível compõe apenas uma pequena parcela desse espectro. Como pode ser observado na Figura 3, a cor percebida esta fortemente relacionada ao comprimento de onda da luz.

Figura 3: A faixa de frequência e comprimento de onda para a luz visível em relação a todo espectro eletromagnético.



Fonte: infoescola.com

Para efeitos práticos, será utilizado o modelo da óptica geométrica para descrever o comportamento da luz. Dessa maneira a luz é tratada como uma partícula, ignorando certas propriedades inerentes ao modelo da óptica física, como dispersão e interferência. A luz é propagada através do meio como um feixe de luz e interações macroscópicas entre a luz e a matéria, maiores que o comprimento de onda da luz, são suficientes para satisfazer os requisitos necessários para simular o transporte de energia [9].

2.2 COLORIMETRIA

Colorimetria é a ciência e o conjunto de técnicas envolvidas tanto na quantificação, quanto na investigação física do fenômeno da percepção de cores pelos seres humanos. Como mencionado na Seção 2.1, para simplificar o processo de renderização, iremos codificar a representação espectral da luz visível da cena como uma tripla em um sistema de cores RGB. Isso significa que podemos designar a cor de objetos, luzes, e outras entidades, por suas respectivas triplas. Efetivamente, isso implica que a representação espectral da luz não é inteiramente correta, devido à perda de informação. Geralmente essa simplificação funciona bem para a maioria dos casos [1].

2.3 RADIOMETRIA

No ramo da óptica, radiometria é o conjunto de técnicas para medição de radiação eletromagnética, incluindo a luz visível. A radiometria é baseada na abstração do modelo da luz pela óptica geométrica, tratando-a como uma partícula. Mesmo não sendo estritamente baseada no comportamento real da luz, existe uma conexão entre a radiometria e as equações de Maxwell, dando-a um forte embasamento na Física [9].

Ao adotar o modelo geométrico para descrever o comportamento da luz, algumas suposições podem ser feitas:

- Linearidade: a combinação entre dois efeitos distintos é feita de forma linear.
- Conservação de energia: quando a luz é refletida ou refratada por uma superfície, ou meio participativo, esse fenômeno não irá produzir mais energia do que quando começou.
- Polarização: não é adequada para o modelo proposto. Portanto, será ignorada.
- Fluorescência: assim como a polarização, tais efeitos trariam pouco valor para o sistema.
- Estado de equilíbrio: assume-se que as fontes de luz presentes no ambiente tenham atingido um estado de equilíbrio. Dessa forma, a distribuição de energia radiante não muda com a passagem do tempo.

2.3.1 Energia Radiante

Fontes de luz emitem fótons que carregam uma quantidade fixa de energia, medida em joules (J). Essa quantidade é derivada da frequência do espectro luminoso em que o fóton foi emitido. A energia carregada por um fóton é dada por

$$Q = hf = \frac{hc}{\lambda}, \quad (2.1)$$

em que c é a velocidade da luz, e h a constante de Planck. A quantidade mínima de energia que uma onda eletromagnética deve possuir corresponde a esse produto, e qualquer valor de energia para uma radiação eletromagnética deve ser um múltiplo inteiro desse produto.

2.3.2 Fluxo Radiante

Fluxo radiante (Φ), ou potência, é a quantidade total de energia que passa por uma superfície ou região do espaço, por unidade de tempo. Considerando um ambiente em estado de equilíbrio (Seção 2.2), estamos interessados em medir a quantidade de fluxo radiante em um instante de tempo qualquer:

$$\Phi = \lim_{\Delta t \rightarrow 0} \frac{\Delta Q}{\Delta t} = \frac{dQ}{dt}. \quad (2.2)$$

Sua medida é dada em joules/segundo (J/s) ou watts (W). Por exemplo, imagine uma fonte de luz que emite uma quantidade $Q = 200000\text{J}$ em um período de uma hora. Se a quantidade de energia emitida por unidade de tempo for constante, temos que o fluxo radiante dessa luz é calculado como

$$\Phi = \frac{200000\text{J}}{3600\text{s}} \approx 55.6\text{W}.$$

2.3.3 Irradiância

A irradiância (E) representa a quantidade de potência radiante que incide sobre uma superfície, real ou imaginária, por unidade de área (medida em W/m^2):

$$E = \frac{\Phi}{A} \quad (2.3)$$

Em geral, estamos interessados em calcular a irradiância para um ponto p em uma superfície de área dA , recebendo um fluxo radiante $d\Phi = E dA$. Tem-se então:

$$E(p) = \lim_{\Delta A \rightarrow 0} \frac{\Delta \Phi(p)}{\Delta A} = \frac{d\Phi(p)}{dA}. \quad (2.4)$$

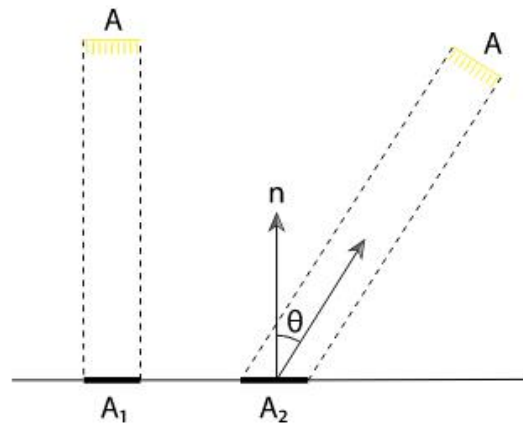
Considere um feixe de luz colimado incidindo sobre uma superfície (Figura 4). Seja Φ a potência radiante do feixe e A a área da seção reta do feixe. Se o feixe incide perpendicularmente sobre a superfície, então a área A_1 iluminada por esse feixe é igual a A . A irradiância medida em qualquer ponto em A_1 é igual a

$$E_1 = \frac{\Phi}{A}. \quad (2.5)$$

Contudo, se o feixe de luz incide com um ângulo θ , a área iluminada na superfície é maior. Se A é pequena, então a área A_2 receptora do fluxo é aproximadamente $A/\cos(\theta)$. Para qualquer ponto em A_2 , a irradiância é então:

$$E_2 = \frac{\Phi \cos(\theta)}{A}. \quad (2.6)$$

Figura 4: A irradiância em uma superfície varia conforme o ângulo de incidência. A irradiância é maximizada quando a superfície está orientada perpendicularmente ao feixe incidente ($\theta = 0$).



Fonte: Physically Based Rendering: From Theory to Implementation

2.3.4 Intensidade Radiante

A intensidade radiante (I) descreve o fluxo radiante provindo de uma fonte, em uma dada direção, por unidade de ângulo sólido. Considere um ponto luminoso infinitesimal que esteja emitindo fótons. Se centralizarmos esse ponto em uma esfera unitária imaginária, a intensidade radiante dessa fonte é dada por

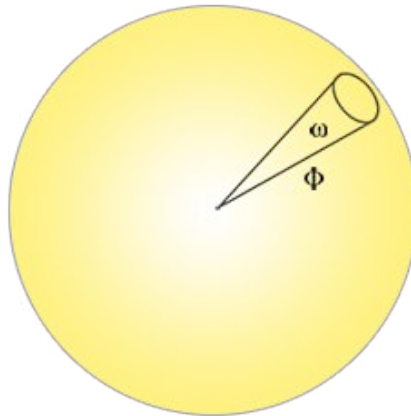
$$I = \frac{\Phi}{4\pi}, \quad (2.7)$$

medida em W/sr (esterradianos). Queremos determinar a intensidade para um cone de direções, dado pelo ângulo sólido $d\omega$:

$$I = \lim_{\Delta\omega \rightarrow 0} \frac{\Delta\Phi}{\Delta\omega} = \frac{d\Phi}{d\omega}. \quad (2.8)$$

Note que, conforme o ângulo sólido tende a zero, menor o cone de direções se torna. No limite, podemos descrever a intensidade como a quantidade de fluxo em uma única direção. Em outras palavras, a quantidade de fluxo radiante medida em uma direção vetorial.

Figura 5: Uma luz omnidirecional apresenta a mesma intensidade em todas as direções. Para uma esfera unitária, a intensidade pode ser medida por $I = \Phi/4\pi$. Para qualquer direção, a intensidade é medida em relação a um cone de direções, representado pelo ângulo sólido ω .



Fonte: learnopengl.com/PBR/Theory

2.3.5 Radiância

A última, e mais importante medida radiométrica, é a radiância (L). Enquanto irradiância mede a quantidade de fluxo por unidade de área de uma superfície, ela não faz nenhuma distinção quanto a direção do fluxo. Intensidade, mede a quantidade de fluxo em um cone de direções, mas não diz nada a respeito da área de incidência medida. Radiância dá esse último passo, e mede irradiância em relação a um ângulo sólido de direção. Radiância é definida como:

$$L(p, \omega) = \lim_{\Delta\omega \rightarrow 0} \frac{\Delta E_\omega(p)}{\Delta\omega} = \frac{dE_\omega(p)}{d\omega} \quad (2.9)$$

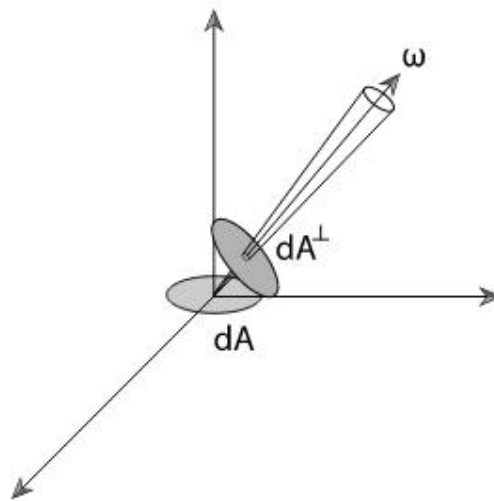
E_ω denota a irradiância na superfície perpendicular a direção ω . Em outras palavras, radiância não é medida em relação a irradiância presente na superfície onde um ponto p se encontra, e sim em uma área projetada perpendicularmente à direção de ω . Essa mudança serve para eliminar o fator de atenuação, descrito pela lei do cosseno de Lambert, da equação.

Radiância é a densidade de fluxo por unidade de área, por ângulo sólido. Em relação ao fluxo, ela é definida por

$$L = \frac{d\Phi}{d\omega dA^\perp}. \quad (2.10)$$

Sendo dA^\perp a área projetada de dA , em direção à ω (Figura 6).

Figura 6: Radiância L é definida como fluxo por unidade de ângulo sólido $d\omega$ e por unidade de área projetada dA^\perp .



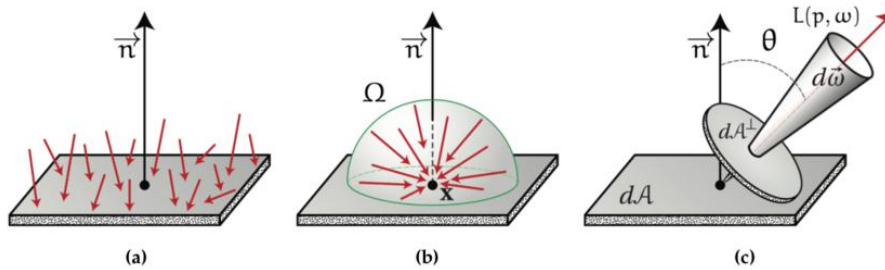
Fonte: Physically Based Rendering: From Theory to Implementation

2.3.6 Radiância Incidente e Sainte

Radiância é o que sensores medem, sendo de suma importância para o PBR. Se a radiância é dada, todas as outras medidas podem ser calculadas. Considere uma superfície de área A . Para calcular a irradiância E_i em um ponto p contido nessa área, podemos integrar a radiância incidente L_i provinda de todas as direções possíveis sobre o domínio de integração (o hemisfério que engloba o ponto, no caso):

$$E_i(p) = \int_{2\pi} L_i(p, \omega) \cos(\theta_i) d\omega$$

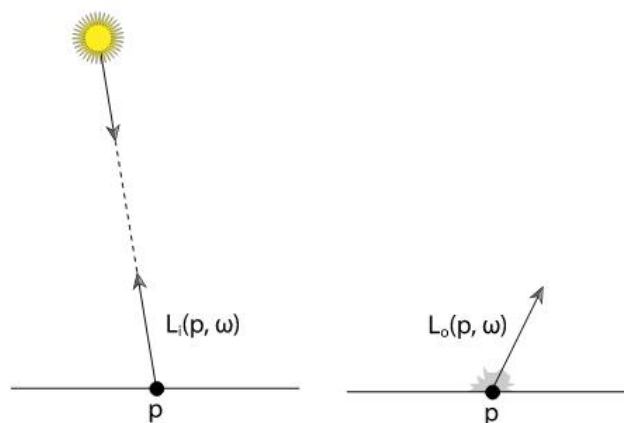
Figura 7: Ilustração gráfica do fluxo radiante (a) sobre a área, irradiância (b) e radiância (c), previamente descritos.



Fonte: researchgate.net

Todavia, é conveniente criar uma distinção entre a radiância medida que chega em um ponto (e.g. devido à iluminação por uma fonte de luz) e a que sai (e.g. devido à reflexão pela superfície). Desse modo, denotamos $L_i(p, \omega)$ como a radiância incidente, e $L_o(p, \omega)$ como a radiância saínte.

Figura 8: À direita, a radiância incidente $L_i(p, \omega)$ descreve a distribuição de radiância que chega em um ponto, em função de posição e direção. À esquerda, a radiância saínte $L_o(p, \omega)$ descreve a distribuição de radiância que sai desse ponto. Note que, para ambos os casos, a direção ω é orientada a apontar para longe da superfície. Nesse caso, $L_i(p, -\omega)$ descreveria a radiância incidente provinda de uma direção abaixo da superfície.



Fonte: Physically Based Rendering: From Theory to Implementation

Uma propriedade interessante é de que, assumindo não haver nenhuma interferência pelo meio participativo (i.e. o raio de luz viaja pelo vácuo), para um ponto p isolado no espaço a radiância L é contínua. Isto é:

$$L_o(p, \omega) = L_i(p, -\omega) = L(p, \omega)$$

3 TRANSPORTE DA LUZ

No capítulo anterior, desenvolvemos as medidas radiométricas necessárias para quantificar a luz. Contudo, ainda é necessário formular como o transporte da luz se dará pelo meio participativo¹, e como isso se traduzirá para a composição da imagem renderizada.

3.1 RAY TRACING

Ray tracing é uma técnica utilizada para computar a visibilidade entre dois pontos através da projeção de 'raios' pelo espaço. Apesar de a ideia por trás do *ray tracing* não ser nova (ela pode ser datada do século XVI, quando foi inicialmente descrita por Albrecht Dürer [7]), o termo como é popularmente conhecido foi imortalizado por Turner Whitted em 1980 [14].⁰ Turner foi o primeiro a demonstrar o uso de um *ray tracer* recursivo para gerar efeitos realistas que não eram possíveis com técnicas tradicionais, como a rasterização (e.g. reflexão, refração, inter-reflexão, etc.). O algoritmo de *ray tracing* recursivo de Turner reformulou o conceito de renderização; de uma questão de determinar a visibilidade entre superfícies, para uma questão de resolver o transporte da luz pela cena. Seu artigo inspirou uma série de trabalhos subsequentes, incluindo a formulação do *path tracing* e a solução para o problema da iluminação global, permitindo a geração de imagens fotorrealistas por um computador.

Como mencionado na Seção 2.1.1, iremos modelar a luz pela sua representação geométrica, e a distribuição espectral da cena como uma tripla RGB. Nesse caso, o uso do *ray tracing* é conveniente para simular o transporte da luz. Um feixe de luz, ou raio, pode ser descrito como uma entidade que possui uma origem espacial e direção de propagação. Podemos achar um ponto P de interseção entre um raio e uma primitiva de forma que:

$$P_{xyz} = O_{xyz} + t * Dir_{xyz}. \quad (3.1)$$

¹Meio participativo é qualquer meio homogêneo por onde a luz é propagada (e.g. água, vidro, vácuo, etc.).

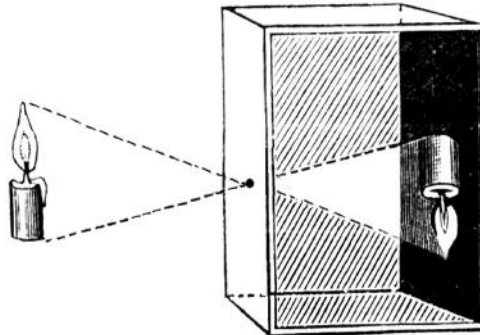
Tal qual P_{xyz} é o ponto de interseção, O_{xyz} o ponto de origem do raio e Dir_{xyz} sua direção, todas em coordenadas cartesianas globais. t é a distância entre a origem do raio e o ponto de interseção.

3.1.1 Câmera

Para gerar uma imagem fotorrealista, é necessário medir a radiância incidente em algum ponto de um medidor radiométrico. No mundo físico real, a luz viaja pelo espaço, desde sua emissão por uma fonte luminosa, até chegar a um observador. No caso do *ray tracing*, o processo é invertido; a luz é propagada a partir da origem do medidor (também denominado como câmera ou observador), até a fonte de luz.

Geralmente, o modelo mais simples para descrever a câmera é o de uma câmara estenopeica, conhecido como modelo de câmera pinhole (Figura 9). Fotografias produzidas por uma câmera pinhole real são geradas invertidas verticalmente. Isso é devido ao fato do filme onde a imagem é gerada estar localizado atrás do centro de projeção da câmera. Para uma modelagem virtual da câmera, esse fenômeno pode ser substituído por uma abstração mais simples. Ao invés de gerar a imagem no filme localizado na câmara estenopeica, podemos projetar um novo plano em frente à câmera, e gerar a imagem partindo desse novo plano. Dessa maneira, o buraco de abertura da câmera, que também é o centro de projeção, é considerado como a posição do observador (ou origem da câmera), e o plano de projeção é a imagem observada.

Figura 9: Ilustração do funcionamento de uma câmera pinhole.



Fonte: markellisimagery.com/pinhole-camera

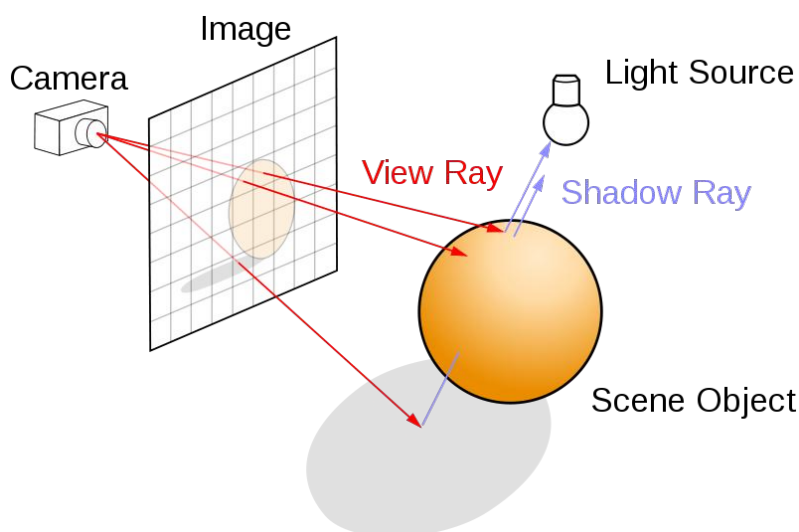
O plano de projeção é particionado em pequenos segmentos chamados píxeis, com tamanho dependente da resolução da imagem. Para cada píxel p_i da imagem, iremos traçar um raio r_i , denominado raio primário, partindo da origem da câmera O_c em direção ao píxel. Podemos achar pontos de interseções entre esses raios primários e os objetos da cena pela Equação 3.1. Se algum ponto de interseção P existir, a radiância sainte L_o nesse ponto deve ser calculada. Para raios primários, o direcionamento da radiância sainte é do ponto de interseção para a origem da câmera. Lembre-se de que, assumindo que o meio participativo não interfere no transporte da energia, a radiância sainte medida em P é a mesma que a radiância incidente sobre o píxel p_i , por onde o raio r_i passa. Sob a perspectiva do observador, essa radiância incidente sobre a câmera é interpretada através da fotometria e colorimetria como a cor do píxel.

O processo é então repetido recursivamente, gerando um novo raio com origem em P , e com direção dependente das propriedades da superfície atingida. A radiância incidente nesse ponto, por sua vez, vai depender da intensidade luminosa provinda na direção do novo raio. No caso de Turner, podemos gerar também outro raio a partir de P , chamado de 'raio de sombra'. Esse raio é direcionado a uma fonte de luz da cena, e é útil para determinar o fator de visibilidade para o ponto de interseção, além de auxiliar na quantificação da radiância sobre o ponto. Se existir qualquer interseção entre P e a fonte de luz, a superfície encontra-se sombreada. Do contrário,

ela está iluminada e a radiância provinda dessa direção deve ser calculada.

Todos os raios gerados continuam a propagar pela cena até serem finalizados (por diversos fatores), ou por não encontrarem mais interseções e atingirem o "fundo da cena".

Figura 10: Ilustração do transporte de luz utilizando *ray tracing*. Um raio primário é gerado para cada píxel da imagem, a partir de uma origem comum. Os raios são propagados recursivamente pela cena até serem finalizados. A radiância incidente em um píxel é acumulada e propagada pelo raio, realizando o caminho inverso da luz: da origem da câmera, até a fonte luminosa.



Fonte: en.wikipedia.org/wiki/Raytracing

3.2 SUPERFÍCIES

Sob uma perspectiva óptica, a superfície de um objeto é uma interface bidimensional que separa dois volumes com diferentes índices de refração. Em uma configuração típica para renderização, o volume externo é normalmente assumido como ar, cujo índice de refração é aproximadamente 1. Já o índice de refração do volume interno depende da substância da qual o objeto é composto [1].

Quando a luz incide sobre uma superfície, uma porção é refratada, enquanto a outra é refletida de volta para o meio. Dois aspectos do material que compõe a superfície tem efeitos importantes para o resultado: as substâncias em ambos os

meios, e a configuração geométrica da superfície.

No caso da refração, a luz sofre uma mudança na velocidade de propagação ao passar de um meio para o outro. Durante a refração, o comprimento de onda da luz muda, enquanto a sua frequência permanece constante. A refração pode ou não ser acompanhada de uma mudança na direção da propagação da luz. Denotamos o índice de refração para o volume externo como η_1 . Para o volume interno, por onde a luz será transmitida ao passar pela superfície, como η_2 . A refração sempre ocorre quando o índice de refração relativo entre dois meios difere de 1.

A mudança de velocidade da luz ao atravessar do meio externo para o meio interno pode ocasionar um deslocamento lateral da luz. Essa mudança acontece se o raio de luz em questão incidir perpendicularmente à superfície, na mesma direção da reta normal ao plano. A reta normal é usada como referência para as medidas dos ângulos de incidência e refração. O desvio angular sofrido pelo raio de luz refratado é descrito pela lei de Snell-Descartes:

$$\eta_1 \sin(\theta_i) = \eta_2 \sin(\theta_r) \quad (3.2)$$

Sendo θ_i o ângulo de incidência, e θ_r o ângulo de refração após atravessar o meio.

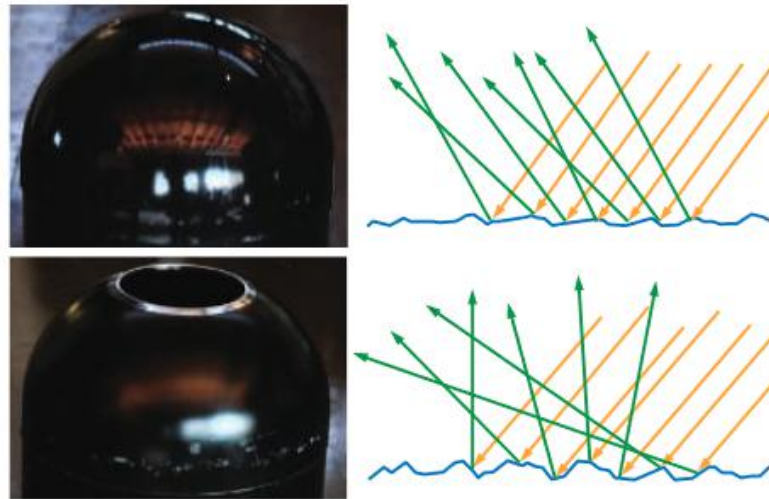
3.2.1 Modelo de Micro-superfícies

Toda superfície possui algum tipo de irregularidade. Contudo, irregularidades menores que o comprimento da onda não afetam a luz. Muitas abordagens para uma modelagem geométrica são baseados na ideia de que uma superfície áspera é composta por um conjunto de 'micro-superfícies' [1]. Costumeiramente, as direções de reflexão e refração dependem da reta normal à macro-superfície. Uma modelagem de micro-superfícies, por sua vez, tem como efeito a reflexão da luz considerando a reta normal a uma micro-superfície qualquer.

Superfícies compostas por micro-superfícies são geralmente modeladas como campos de altura, em que a distribuição da orientação da micro-superfície é descrita estatisticamente. Mesmo que um ponto em específico reflita a luz em uma única direção, cada píxel cobre uma porção de micro-superfícies que refletem a luz

em várias direções. Como resultado, a propagação da luz é modelada como um cone de direções. A largura desse cone depende da variância estatística da distribuição do direcionamento das retas normais das micro-superfícies. Em outras palavras, depende da aspereza do material que compõe a superfície.

Figura 11: À esquerda, podemos ver duas fotografias de dois objetos. À direita, os diagramas das composições de suas respectivas micro-geometrias. Materiais mais lisos refletem a luz de maneira coerente, enquanto materiais mais ásperos refletem a luz de maneira divergente. A nitidez da imagem refletida depende de quão áspera é a superfície.



Fonte: Real-Time Rendering

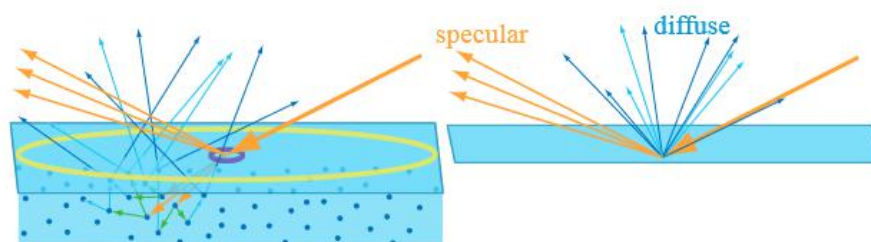
3.2.2 Dispersão Sob a Superfície

Apesar de o fenômeno da refração estar associado às materiais translúcidos, ele também ocorre em superfícies opacas. Quando a refração acontece em objetos opacos, a luz é dispersa pelo interior do objeto e sofre absorção, causando perda de energia. No caso de metais, o interior do objeto pode conter elétrons livres, causando uma grande absorção da luz. É por essa razão que metais tem alta refletividade. Em contrapartida, materiais dielétricos podem apresentar uma variedade de comportamentos quanto a absorção e dispersão da luz pela substância do objeto. Além de absorção, a luz também pode atravessar o objeto intacta, saindo por outro ponto e retornando ao meio exterior.

A distância entre o ponto de entrada e saída depende da densidade e propriedades da substância do objeto. Materiais com pouca dispersão e absorção são transparentes, transmitindo qualquer luz refratada por todo o objeto. A relação entre essa distância e a escala da superfície analisada para a renderização (o tamanho do particionamento do píxel, por exemplo) é importante. Se a distância entre a entrada e a saída for pequena comparada ao tamanho do píxel, os efeitos de absorção e dispersão podem ser assumidos como efetivamente nulos [1]. Isso permite com que a dispersão sob a superfície possa ser combinada com a reflexão da luz sobre a superfície, em um único modelo de sombreamento. A direção de propagação da luz em um ponto na superfície vai depender apenas da direção da luz incidente, no mesmo ponto. Como a luz dispersa sob a superfície tem uma aparência diferente da luz refletida sobre a superfície, é conveniente separá-las em termos distintos. Da-se o nome de *especular* para descrever a luz refletida sobre a superfície, e *difusa* para a luz dispersa sob a superfície.

Contudo, se a distância entre o ponto de entrada e saída for maior que um píxel, é necessário recorrer às técnicas especializadas que capturem os efeitos visuais da transmissão da luz pelo meio, saindo por outro ponto do objeto. A diferença entre os modelos de dispersão sob a superfície local e global podem ser observados na Figura 11.

Figura 12: À esquerda, uma representação da dispersão sob a superfície e reflexão sobre a superfície. Duas áreas de análise são mostradas (a área que um píxel abrange), em amarelo e roxo. Para o círculo amarelo, a distância entre os pontos de entrada e saída da luz estão contidos em uma área menor que a do píxel analisado. Dessa maneira, efeitos como absorção e dispersão podem ser ignorados, e o modelo de reflexão da luz pode ser simplificado, como visto na figura à direita. Para o círculo roxo, a distância entre os pontos é maior que área do píxel. Técnicas de sombreamento especializadas são necessárias para reproduzir os efeitos visuais descritos por esse fenômeno.



Fonte: Real-Time Rendering

3.3 EQUAÇÃO DE RENDERIZAÇÃO

Podemos quantificar a iluminação através da equação que descreve o total de irradiância presente em cada ponto da cena. Essa equação é denominada equação de renderização, ou também equação do transporte da luz.

O princípio fundamental da equação do transporte da luz é o balanço de energia. Qualquer mudança de energia deve ser carregada por um processo e retransmitida pela cena. Como assumimos que a iluminação é um processo linear (Seção 2.3), a diferença entre a quantidade de energia entrando e saindo de um sistema deve ser igual à diferença entre a quantidade de energia emitida e absorvida. Em uma escala macroscópica, temos a conservação de potência dada por:

$$\Phi_o - \Phi_i = \Phi_e - \Phi_a. \quad (3.3)$$

Assim, para que o equilíbrio de energia permaneça estável, a radiância sainte de uma superfície deve ser igual à radiância emitida mais a porção de radiância dispersa pela superfície. A equação de renderização é formulada de forma que:

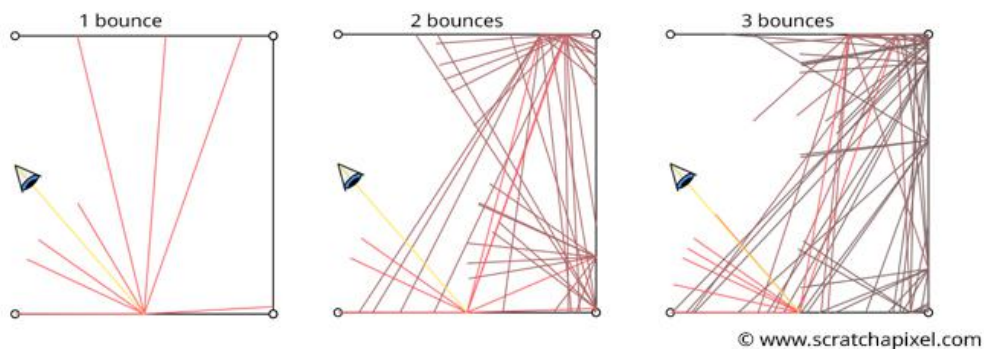
$$L_o(p, \omega_o) = L_e(p, \omega_o) + \int_{S^2} f(p, \omega_o, \omega_i) L_i(p, \omega_i) |\cos(\theta_i)| d\omega_i. \quad (3.4)$$

Nas quais:

- $L_o(p, \omega_o)$ é radiância sainte de um ponto p na direção ω_o .
- $L_e(p, \omega_o)$ é a radiância emitida pela superfície na direção ω_o .
- $f(p, \omega_o, \omega_i)$ é a função que descreve o modelo de reflexão da luz.
- $L_i(p, \omega_i)$ é a radiância incidente vinda de uma direção ω_i .
- $\cos(\theta_i)$ é o fator de atenuação da radiância incidente, dado pela lei de Lambert.

Repare que, a quantidade de radiância sainte de p depende da quantidade total de radiância incidente no ponto, sobre todo o domínio (uma esfera unitária que engloba o ponto, para a forma completa da equação). Podemos dizer que a radiância sainte depende da irradiância medida no ponto. Essa, por sua vez, depende da intensidade luminosa medida em todas as direções possíveis sobre todo o domínio. Faz-se observável a natureza recursiva da equação, uma vez que para calcular a irradiância em p , é necessário recorrer novamente à equação para calcular a radiância sainte de algum outro ponto, que conseqüentemente ilumina p .

Figura 13: A quantidade de radiância que chega a um observador depende da quantidade da irradiância mais a quantidade de radiância emitida por um ponto p . A irradiância medida em p depende da intensidade luminosa medida em todas as direções, quantificados pelas radiâncias saintes de outros pontos da cena.

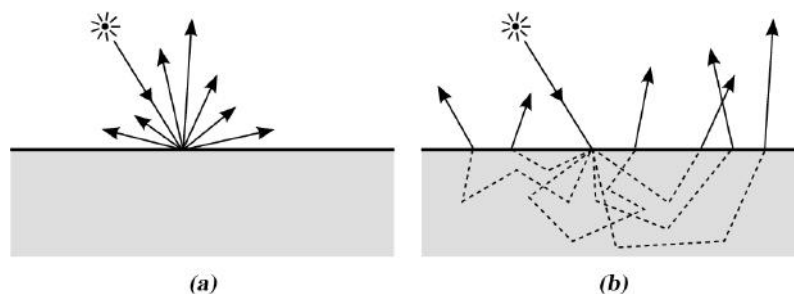


Fonte: scratchapixel.com

A função de reflexão é responsável por avaliar a influência que a superfície tem sobre a luz, dependendo das propriedades do material envolvido, configuração da geometria local, e direcionamento da luz. Existem diversos modelos, cada qual dependendo do tipo de substância que compõe o objeto.

Modelos como a *bidirectional scattering-surface distribution function* (BSSDF) visam avaliar o comportamento da luz quando esta é propagada no interior do objeto, saindo em algum outro ponto. Ela é usada para modelar substâncias como a cera, por exemplo. Para materiais como vidro, água, etc., temos a *bidirectional transmittance distribution function* (BTDF), que avalia a transmissão da luz diretamente pelo objeto, sem a interferência do meio no processo. Para materiais opacos, podemos utilizar a *bidirectional reflectance distribution function* (BRDF), que unifica os efeitos de reflexão e refração.

Figura 14: Dois modelos de reflexão: (a) BRDF e (b) BSSDF.



Fonte: Physically Based Rendering: From Theory to Implementation

3.3.1 BRDF

A BRDF aproxima as propriedades reflexivas do material baseadas no modelo de micro-superfícies previamente descrito. Uma BRDF construída para um pipeline PBR deve satisfazer três requisitos importantes:

- Positividade: $f_r(p, \omega_o, \omega_i) \geq 0$.
- Reciprocidade de Helmholtz: para todos os pares de direções ω_o e ω_i , a avaliação da BRDF deve ser a mesma, independentemente da ordem das direções consideradas. Isso é, $f_r(p, \omega_o, \omega_i) = f_r(p, \omega_i, \omega_o)$.

- Conservação da energia: o total de energia sante deve ser menor ou igual à quantidade de energia incidente.

Existem diversos modelos de reflexão. Um dos mais populares é o modelo de Cook-Torrance [3]. A BRDF de Cook-Torrance é baseada na unificação dos efeitos de reflexão e refração, vistos na Seção 3.2.2:

$$f_r = k_d f_{diffuse} + k_s f_{specular}$$

$f_{diffuse}$ é a função que avalia as propriedades do material para a porção difusa da luz. Similarmente, $f_{specular}$ é a função para a porção especular, e k_d e k_s são as proporções da luz refratada e refletida, respectivamente ($k_d + k_s \leq 1$). O modelo de Cook-Torrance assume que cada micro-superfície se comporte como um espelho ideal, refletindo a luz perfeitamente pelo ângulo de reflexão.

A porção difusa é modelada como uma superfície difusa ideal, também chamada de Lambertiana. Isto é, ela é modelada de tal forma que o raio incidente é refletido em múltiplas direções. Uma superfície Lambertiana apresentará uma luminosidade uniforme, percebida da mesma forma de todas as direções em torno da normal da macro-superfície.

$$f_{diffuse} = \frac{c}{\pi}$$

Em que c é o *albedo*, ou cor base do material que compõe a superfície. A divisão por π serve como um fator de normalização, assegurando a conservação de energia. Já a porção especular é mais complexa, uma vez que ela depende da aspereza do material, e do ângulo entre o observador e o ponto de análise. Ela é descrita como:

$$f_{specular} = \frac{DFG}{4(\omega_o \cdot n)(\omega_i \cdot n)}$$

A porção especular de Cook-Torrance é composta por três funções, em que cada uma aproxima uma propriedade reflexiva da superfície. Essas são: a função de distribuição normal (D), a equação de Fresnel (F) e a função para o sombreamento da micro-geometria (G).

3.3.2 Função de Distribuição Normal

A função de distribuição normal (NDF²) descreve estatisticamente a distribuição das retas normais das micro-superfícies, sobre uma área coberta pela macro-superfície. Em termos simples, a NDF retorna a densidade de uma direção em particular estar alinhada com os direcionamentos das demais retas normais das micro-superfícies que compõe essa distribuição, descritas por um cone de direções.

Formalmente, dado um cone de direções $d\omega_h$ centrado em torno da reta normal n da macro-superfície, e uma área infinitesimal A que abrange uma parcela da macro-superfície, então $dA_h = D(h)d\omega_h A$ é a área total da porção da micro-superfície cuja reta normal está alinhada com a distribuição das retas normais da micro-geometria contida nesta área, e $D(h)$ é a NDF.

Como a NDF descreve uma função de distribuição, temos que nos assegurar que seja normalizada. Pode-se pensar que, considerando que dA_h corresponde a uma pequena parcela da área A , temos a normalização da NDF dada por

$$\frac{1}{A} \int dA_h = \int_{\Omega} D(h)d\omega_h = 1.$$

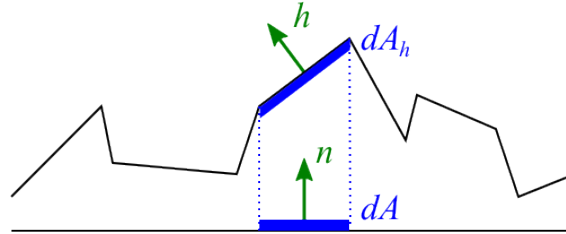
Porém essa presunção é incorreta, pois ignora a distorção espacial da micro-superfície cuja área é dA_h . O que queremos é a área correspondente a projeção dessa micro-superfície sobre a área A , ou seja, $dA = (n \cdot h)dA_h$ (Figura 15). A função normalizada correta é então

$$\frac{1}{A} \int dA = \int_{\Omega} D(h)(n \cdot h)d\omega_h = 1. \quad (3.5)$$

Temos que a NDF é, portanto, uma função de densidade da área da micro-superfície em relação ao ângulo sólido e área da macro-superfície.

²Normal Distribution Function

Figura 15: dA corresponde a área projetada da micro-superfície com área dA_h sobre a macro-superfície.



Fonte: reedbeta.com/blog/how-the-ndf-really-defined

Existem diversos modelos de NDF que aproximam estatisticamente o alinhamento geral das micro-superfícies por algum coeficiente α (0-100%), que descreve a aspereza do material. Nesse trabalho, usaremos o modelo conhecido como Trowbridge-Reitz, ou GGX³.

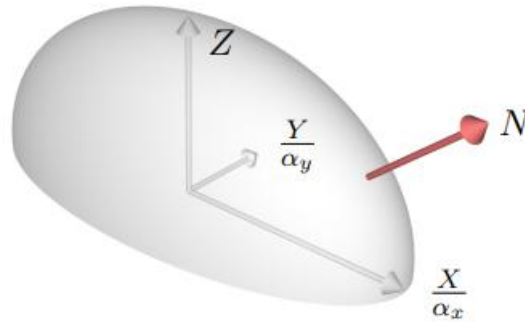
3.3.2.1 GGX

A GGX é a distribuição das normais de um elipsoide. Isto é, ela mede a densidade da orientação de uma reta normal à superfície do elipsoide. GGX é medida apenas no hemisfério superior do elipsoide, como vista na Figura 16. A aspereza do material é descrita por dois parâmetros α_x e α_y , representando um fator de deformação do elipsoide nos eixos X e Y, respectivamente.

Para uma superfície isotrópica, temos que a superfície é deformada igualmente em ambos os eixos, tal qual $\alpha_x = \alpha_y = \alpha$. Como o elipsoide é truncado para o hemisfério superior em direção ao eixo $Z = (0, 0, 1)$, para $\alpha = 1$ temos o elipsoide com o formato de uma esfera. Em termos práticos, isso quer dizer que densidade para qualquer direção no hemisfério dessa distribuição é a mesma quando o coeficiente de aspereza é máximo.

³Ambas as nomenclaturas se referem ao mesmo modelo, mas em seu artigo *Microfacet models for refraction through rough surfaces*[13], Walter et al. cunhou o termo GGX, como é mais popularmente conhecido.

Figura 16: GGX é a distribuição das normais de um elipsoide. O elipsoide é truncado para o hemisfério superior, em direção a $Z = (0, 0, 1)$. α_x e α_y representam a deformação do elipsoide nos eixos X e Y, respectivamente.



Fonte: Eric Heitz - Sampling the GGX Distribution of Visible Normals

A distribuição GGX é dada por:

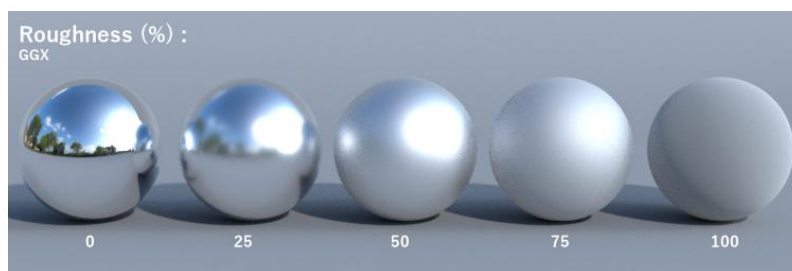
$$D_{GGX}(n, h, \alpha) = \frac{\alpha^2}{\pi((n \cdot h)^2(\alpha^2 - 1) + 1)^2} \quad (3.6)$$

Uma maneira de imaginar como a distribuição afeta visualmente a superfície de um objeto é em termos do coeficiente de aspereza, α . Uma superfície polida, como porcelana, vidro, etc., apresentará uma distribuição de micro-superfícies mais alinhada com a reta normal da macro-superfície, parecendo refletir a luz perfeitamente pelo ângulo de reflexão.

Um valor baixo para α indica que a aspereza do material é "baixa", indicando uma superfície lisa. Pela construção do elipsoide da distribuição GGX, isso faz sentido, uma vez que um valor baixo para α acabaria por comprimir o elipsoide, transformando a distribuição das normais mais alinhada com o eixo Z. Qualquer direcionamento de h que não esteja alinhado nesse sentido acabaria resultando em uma baixa densidade. Para um material perfeitamente liso, como um espelho ideal, isso significa que qualquer direção que não esteja perfeitamente alinhada com a normal n da macro-superfície, tem probabilidade nula de ocorrer. Se fossemos utilizar a GGX para determinar o ângulo de reflexão de um raio incidente, então o direcionamento do raio refletido seria governado pela aspereza do material.

Similarmente, uma superfície áspera, como tijolo, giz, etc., apresentaria um modelo de reflexão mais parecido com uma superfície Lambertiana, uma vez que um valor alto de α acabaria por refletir a luz em múltiplas direções.

Figura 17: A aspereza do material determina como a luz é refletida pela superfície. Materiais mais lisos apresentam uma distribuição do direcionamento das micro-superfícies mais alinhadas com a normal da superfície, enquanto materiais mais ásperos apresentam uma distribuição mais divergente.

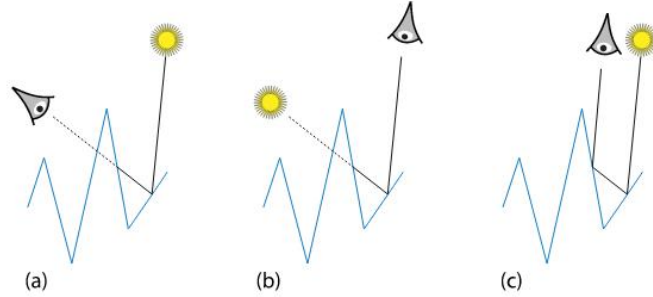


Fonte: clarissewiki.com/3.6/reflection2.html

3.3.3 Sombreamento da Micro-geometria

Assim como na macro-geometria, efeitos de reflexão entre a luz e a micro-geometria devem ser contabilizados (Figura 18). A função de sombreamento da micro-geometria descreve a atenuação da luz devido à obstrução da luz pelas micro-superfícies. Novamente, como no caso da NDF, a função é uma aproximação estatística. A função mede a probabilidade de que, em um ponto, a visibilidade das micro-superfícies seja obstruída uma pelas outras, ou que a luz seja refletida múltiplas vezes entre as micro-superfícies, perdendo energia no processo antes de chegar ao observador.

Figura 18: Três efeitos importantes que devem ser considerados para a BRDF. (a) *Máscara*: a micro-superfície considerada não está visível para o observador devido à oclusão por outra micro-superfície. (b) *Sombreamento*: analogamente, a luz não alcança a micro-superfície. (c) *Interreflexão*: a luz é refletida entre as micro-superfícies antes de chegar ao observador.



Fonte: Physically Based Rendering: From Theory to Implementation

A função utilizada é uma combinação entre a GGX e a aproximação de Schlick-Beckmann, conhecida como Schlick-GGX [10]:

$$G_{Schlick-GGX}(n, v, k) = \frac{n \cdot v}{n \cdot v(1 - k) + k}. \quad (3.7)$$

Sendo v o vetor que representa a direção oposta ao raio incidente e $k = \alpha \sqrt{\frac{2}{\pi}}$. Para efetivamente aproximar o sombreamento da micro-geometria, é necessário considerar tanto a direção do raio incidente v (máscara), como a direção do raio refletido l (sombreamento). Esses efeitos são contabilizados pela função de máscara-sombreamento de Smith, que retorna a fração de micro-superfícies que estão visíveis pelo ponto de vista do observador:

$$G(n, v, l, k) = G_S(n, v, k)G_S(n, l, k). \quad (3.8)$$

Usando a função de Smith com G_S substituída pela $G_{Schlick-GGX}$, temos os efeitos ilustrados pela Figura 19.

Figura 19: Da esquerda para a direita, o coeficiente de aspereza α variando de 0% a 100%. A função de Smith age como um multiplicador, retornando um valor entre [0.0, 1.0]. Para 1.0, todas as micro-superfícies estão visíveis, não apresentando sombreamento. Para 0.0, o sombreamento é máximo.



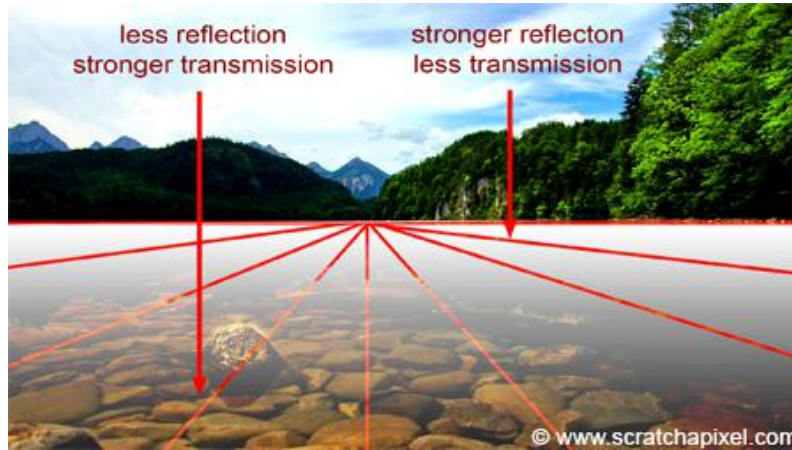
Fonte: learnopengl.com/PBR/Theory

3.3.4 Fresnel

Quando a luz atinge uma interface entre um meio com índice de refração η_1 , e um segundo meio com índice de refração η_2 , a luz pode sofrer tanto reflexão como refração. A equação de Fresnel (ou coeficiente de Fresnel), descreve a porção da luz que é refletida.

A equação de Fresnel assume que a interface entre os meios é um plano liso, e que os meios participativos envolvidos são homogêneos e isotrópicos. Todo material possui algum nível base de refletividade quando visto perpendicularmente. Contudo, conforme o ângulo entre o observador e o plano aumenta, a porção refletida da luz se torna mais aparente. Esse fenômeno é conhecido como *Fresnel*, descrito pela equação de Fresnel.

Figura 20: A porção da luz refletida aumenta conforme o ângulo entre observador e o plano aumenta. Esse fenômeno é conhecido como *Fresnel*.



Fonte: scratchapixel.com

A equação de Fresnel é complexa e difere dependendo do tipo de material envolvido. Existem duas classes principais de materiais para a renderização PBR: dielétricos e condutores. Materiais dielétricos não conduzem eletricidade e possuem índice de refração na faixa de 1 a 3. Condutores, como materiais metálicos, absorvem uma grande parcela da luz refratada. Geralmente, a absorção total da luz ocorre bem próxima da superfície do objeto [9]. Efetivamente, isso significa que metais não possuem a componente difusa da reflexão da luz, apenas a especular.

Existe uma aproximação para a equação de Fresnel, formulada por Schlick [10], que pode ser avaliada rapidamente e utilizada tanto para materiais dielétricos, quanto para metálicos:

$$F_{Schlick}(h, v, F_0) = F_0 + (1 - F_0)(1 - (h \cdot v))^5 \quad (3.9)$$

$$F_0 = \left(\frac{\eta_1 - \eta_2}{\eta_1 + \eta_2} \right)^2 \quad (3.10)$$

Tal que F_0 é o nível base de refletividade quando o material é visto perpendicularmente. Para materiais dielétricos a função é adequada, mas quando se trata de materiais metálicos, é necessário um certo cuidado.

Como metais absorvem a luz completamente, pelas propriedades eletromagnéticas da luz, isso se traduz de forma que a porção refletida é composta exclusivamente pelo espectro eletromagnético do comprimento de onda que não foi absorvido pelo metal. É por esse motivo que a refletividade de metais possui uma certa coloração, relativa a essa porção da luz não absorvida. O modelo geométrico proposto não é capaz de simular tal efeito, mas podemos contornar essa restrição se definirmos F_0 como uma tripla, em que cada elemento representa um coeficiente de reflexão para cada espectro de cor no sistema RGB.

Com todos os componentes do modelo de Cook-Torrance formulados, podemos substituir o modelo de reflexão na equação de renderização:

$$L_o(p, \omega_o) = L_e(p, \omega_o) + \int_{\Omega} f_r(p, \omega_o, \omega_i) L_i(p, \omega_i) \cos(\theta_i) d\omega_i$$

$$L_o(p, \omega_o) = L_e(p, \omega_o) + \int_{\Omega} \left(k_d \frac{c}{\pi} + k_s \frac{DFG}{4(\omega_o \cdot n)(\omega_i \cdot n)} \right) L_i(p, \omega_i) \cos(\theta_i) d\omega_i$$

$$L_o(p, \omega_o) = L_e(p, \omega_o) + \int_{\Omega} \left(k_d \frac{c}{\pi} + \frac{DFG}{4(\omega_o \cdot n)(\omega_i \cdot n)} \right) L_i(p, \omega_i) \cos(\theta_i) d\omega_i.$$

Repare que k_s é excluído na última parte da equação. Isso é porque o *Fresnel* já contabiliza a porção da luz refletida. Podemos achar a porção refratada por $k_d = 1 - F$.

4 AMOSTRAGEM

4.1 MÉTODO DE MONTE CARLO

Idealmente, gostaríamos de resolver a equação de renderização analiticamente. Contudo, para além de casos mais simples, ela é intratável devido à sua natureza recursiva. Para resolver esse problema, é necessário recorrer a uma aproximação numérica. Desenvolvido para a simulação de experimentos físicos durante o *projeto Manhattan*, o Método de Monte Carlo (MMC) é ideal para lidar com essa classe de problemas.

4.1.1 Estimadores

Uma variável aleatória contínua X é uma variável que pode assumir qualquer valor numérico x em um determinado intervalo, ou série de intervalos. Isto é, uma variável aleatória contínua é uma variável para a qual o conjunto amostral Ω é infinito e não enumerável. Os valores de x podem ser quantitativamente descritos pela função densidade de probabilidade (FDP) p . A probabilidade de x assumir um valor em um intervalo a e b é dada por

$$Pr\{a \leq X \leq b\} = \int_a^b p(x)dx. \quad (4.1)$$

A função densidade de probabilidade $p(x)$ deve satisfazer dois critérios:

- É sempre positiva: $p(x) \geq 0$
- É normalizada: $\int_{-\infty}^{\infty} p(x)dx = 1$

Dessa forma, podemos definir a média, ou valor esperado, de uma distribuição amostral de uma variável aleatória contínua X como:

$$E[X] = \int_{\Omega} xp(x)dx. \quad (4.2)$$

E sua variância:

$$V[X] = E[(X - E[X])^2]. \quad (4.3)$$

Para computar um valor aproximado para uma quantidade desconhecida Q (i.e. uma estimativa do valor de Q), uma função F_N de variáveis aleatórias X_1, \dots, X_N é dita um estimador para Q se o valor esperado $E[F_N]$ é uma aproximação adequada de Q :

$$F_N = F_N(X_1, \dots, X_N). \quad (4.4)$$

Em geral, estamos interessados em um bom estimador; um que calcule uma boa estimativa o mais rápido possível. Um estimador robusto deve atender certos requisitos:

- Erro

Para uma amostra de valor x , o erro para um estimador F_N é definido como:

$$e(x) = F_N(X = x) - Q. \quad (4.5)$$

- Erro Quadrático Médio

O erro quadrático médio (EQM) de F_N é o valor esperado dos erros ao quadrado, isto é:

$$EQM(F_N) = E[(F_N(X) - Q)^2]. \quad (4.6)$$

O EQM é útil para indicar o quão distante, em média, o conjunto de estimativas está do parâmetro a ser estimado.

- Viés

Um viés β é o valor esperado do erro:

$$\beta[F_N] = E[F_N - Q]. \quad (4.7)$$

Um estimador é dito não enviesado se $\beta[F_N] = 0$ para uma amostragem de tamanho N :

$$E[F_N] = Q, \text{ para todo } N \geq 1. \quad (4.8)$$

Uma vantagem de um estimador não enviesado é a garantia de que ele irá convergir para o valor correto de Q , dada uma quantidade suficiente de amostras [2]. Isso significa que podemos escrever o EQM do estimador como:

$$EQM[F_N] = V[F_N] + \beta[F_N]^2. \quad (4.9)$$

Para estimadores não enviesados, o EQM é igual à variância. Para estimadores enviesados, o erro é mais difícil de se calcular. Mesmo com uma quantidade infinita de amostras examinadas, não há garantia que um estimador enviesado convergirá para o resultado correto. Na prática, estimadores enviesados podem apresentar certos atributos desejáveis, como a geração de uma menor variância amostral, o que os tornam atraentes para a síntese de imagens. A exemplo de renderização, ruído é a manifestação da variância das amostras geradas. Enquanto mais amostras diminuem o ruído presente na imagem, um estimador enviesado pode produzir menos ruído utilizando o mesmo número de amostras, mesmo que o resultado não seja inteiramente correto.

- Consistência

Um estimador F_N é dito consistente se, conforme o número de amostras coletadas aumenta, o valor esperado do estimador converge para o resultado correto. Matematicamente, para um erro $\epsilon > 0$, não importando quão pequeno seja, pela Lei Fraca dos Grandes Números, temos:

$$\lim_{N \rightarrow \infty} Pr\{|F_N - Q| < \epsilon\} = 1. \quad (4.10)$$

Isso quer dizer que, se usarmos um estimador consistente, temos a garantia que o resultado convergirá para o valor estimado, conforme o número de amostras aumenta. Consistência é uma condição mais forte do que exigir que o estimador seja não enviesado. Apesar de que, ainda é possível que um estimador não enviesado seja inconsistente, o que geraria uma variância infinita. Um estimador enviesado é consistente se seu viés β tende a 0 conforme o número de amostras N aumenta.

4.2 INTEGRAÇÃO DE MONTE CARLO

Quando se trata de avaliar uma integral I , a ideia por trás do Método de Monte Carlo é achar uma aproximação para essa integral a partir da geração de amostras aleatórias e independentes uma das outras, sobre um espaço amostral Ω . Retornando

ao exemplo anterior, temos a integral para o valor esperado de uma variável aleatória como:

$$I = \int_{\Omega} f(x)p(x)dx \quad (4.11)$$

Dadas N amostras independentes X_1, X_2, \dots, X_N , geradas a partir de uma função densidade de probabilidade p , uma aproximação para I é dada por um estimador \hat{I}_N :

$$\hat{I}_N = \frac{1}{N} \sum_{i=1}^N f(X_i), \quad X_i \sim p \quad (4.12)$$

Conforme o número N de amostras aumenta, o erro estimado diminui. Queremos escolher N de forma que tenhamos confiança de que o estimador não enviesado \hat{I}_N é uma boa aproximação para a I .

Note que não foram feitas quaisquer presunções sobre como $f(x)$ é formulada. Até agora, assumimos que a chance de escolher um ponto qualquer para a variável aleatória X é equiprovável sobre todo o espaço amostral Ω . Em outras palavras, assumimos que qualquer amostra tenha a mesma chance de ser escolhida dentre as demais. Infelizmente, problemas reais raramente são tão simples assim [2]. Por exemplo, a função $f(x)$ pode ter valores muito baixos em algumas regiões, e valores muito altos em outras. Se fossemos coletar amostras uniformemente sobre esse espaço amostral, poderíamos acabar com uma variância enorme. Para resolver esse problema podemos gerar amostras a partir de outra função densidade, q . Re-escrevendo a equação Equação 4.11, temos então:

$$I = \int_{\Omega} f(x)p(x)dx = \int_{\Omega} \frac{f(x)p(x)}{q(x)}q(x)dx. \quad (4.13)$$

Dessa maneira, a contribuição de uma amostra é ajustada pelo 'peso' dessa nova função densidade.

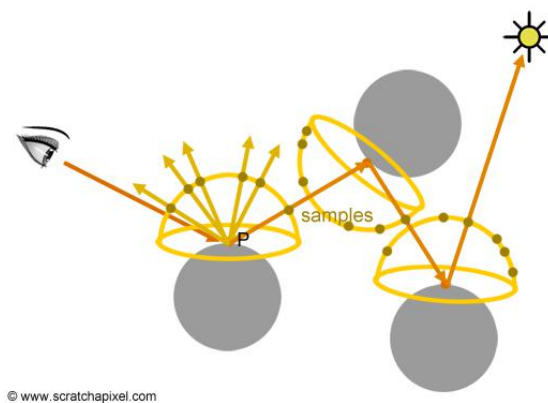
$$\hat{I}_p = \frac{1}{N} \sum_{i=1}^N f(X_i) \frac{p(X_i)}{q(X_i)} \quad X_i \sim q \quad (4.14)$$

4.3 PATH TRACING

A equação de renderização foi desenvolvida e apresentada pela primeira vez por James Kajiya, em 1986 [8]. A equação de renderização é concisa, mas não é fácil de resolver. Cenas de computação gráfica variam desde configurações bem simples, até grandes construções envolvendo bilhões de triângulos e diversos materiais. Não há como resolver a equação de renderização diretamente.

Kajiya mostrou que técnicas estatísticas poderiam ser usadas para resolver o problema; mesmo que não seja resolvida diretamente, é possível resolvê-la ao longo dos caminhos de raios individuais. Se ela for resolvida até certo ponto que consiga aproximar a iluminação na cena com precisão, imagens fotorrealistas podem ser geradas. A essa técnica ele deu o nome de *path tracing*.

Figura 21: Podemos aproximar a iluminação da cena utilizando o *path tracing*. A ideia por trás da técnica é seguir o caminho de raios individuais até sua finalização. A iluminação da cena é aproximada pela média da iluminação quantificada por esses raios.

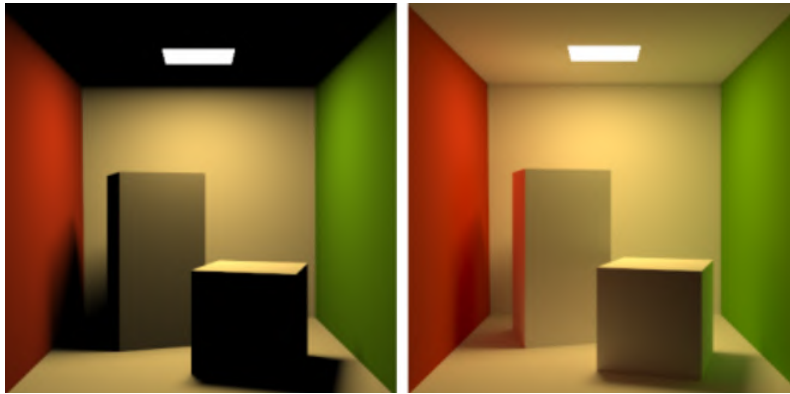


Fonte: scratchapixel.com

Algoritmos que visam gerar imagens fotorrealistas estão interessados em resolver o problema da iluminação global. Tais algoritmos consideram não apenas a iluminação provinda de uma fonte de luz que ilumina uma superfície diretamente (iluminação direta), mas também os casos subsequentes em que os raios de luz da mesma fonte são refletidos por outras superfícies da cena, sejam refletivas ou não (iluminação indireta).

Teoricamente, reflexões, refrações e sombreamento são exemplos de iluminação global, uma vez que quando simulados, um objeto deve afetar a renderização dos demais. Na prática, apenas a simulação de inter-reflexões difusas e cásticas são classificadas como iluminação global [15].

Figura 22: Cornell Box - A direita, a cena renderizada considerando apenas os efeitos da iluminação direta. Partes da cena que não são iluminadas diretamente pela luz, como o teto, não são bem definidas. A esquerda, a mesma cena renderizada considerando a iluminação global. Repare na fusão de cores entre as paredes e as caixas (esse fenômeno é conhecido como *color bleeding*).



Fonte: victorphellipe.wordpress.com/2010/11/13/iluminacao-global-2

Com o método de Kajiya, conseguimos alcançar a iluminação global se resolvermos a equação de renderização para cada ponto da cena. O algoritmo consiste em seguir o caminho de um raio, e a cada novo ponto de interseção, avaliar a equação de renderização nesse ponto. Pela natureza recursiva da equação, não conseguimos resolvê-la analiticamente, mas através do MMC podemos achar uma aproximação.

A fim de exemplificação, considere a equação de renderização com o modelo de reflexão de Cook-Torrance (Seção 3.3), visto no capítulo 3:

$$L_o(p, \omega_o) = L_e(p, \omega_o) + \int_{2\pi} f_r(p, \omega_o, \omega_i) L_i(p, \omega_i) \cos(\theta_i) d\omega_i$$

Para resolver a equação com o MMC, deve ser possível gerar amostras para todas as direções possíveis sobre o espaço amostral (o hemisfério que engloba o ponto p de análise, nesse caso). Dessa maneira, podemos gerar amostras a partir de uma FDP

$p(x)$ que respeite esta condição. Se consideramos que a probabilidade de gerar uma amostra em uma determinada direção é equiprovável sobre todo o espaço amostral, temos que $p(x)$ é igual a uma constante C para qualquer amostra gerada.

$$\int_{\omega=0}^{2\pi} p(\omega) d\omega = 1.$$

Como ela é constante sobre todo o espaço amostral, podemos rescrever a equação de forma que

$$C \int_{\omega=0}^{2\pi} d\omega = 1.$$

Portanto,

$$C \int_{\omega=0}^{2\pi} d\omega = C(2\pi - 0) = 1.$$

Temos então que

$$C = p(\omega) = \frac{1}{2\pi}.$$

Uma aproximação para a integral é então dada por:

$$L_o(p, \omega_o) \approx L_e(p, \omega_o) + \frac{1}{N} \sum_{i=1}^N \frac{f_r(p, \omega_o, \omega_i) L_i(p, \omega_i) \cos(\theta_i) d\omega_i}{p(\omega_i)} =$$

$$L_o(p, \omega_o) \approx L_e(p, \omega_o) + \frac{2\pi}{N} \sum_{i=1}^N f_r(p, \omega_o, \omega_i) L_i(p, \omega_i) \cos(\theta_i) d\omega_i$$

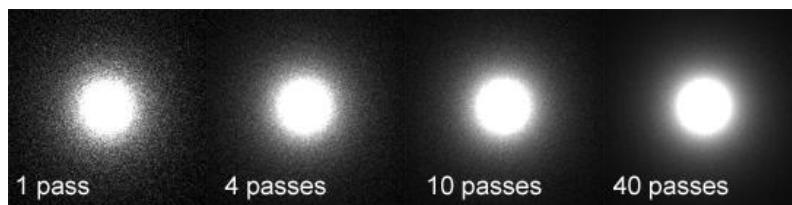
Estamos interessados em computar a radiância incidente que chega ao observador. Para raios primários, conseguimos calcular a irradiância presente no primeiro ponto de interseção pela aproximação; uma nova direção de propagação para o raio é gerada, escolhendo uma direção aleatória sobre o hemisfério superior que engloba a reta normal a superfície, e o trajeto do raio é continuado. A irradiância é encontrada pela média da soma das radiâncias incidentes que chegam a esse ponto.

No entanto, é necessário estabelecer um critério de parada, visto que não há como prever até quando um raio de luz será propagado. Geralmente, encerramos a recursão quando o raio de luz atinge uma fonte luminosa ($L_e > 0$), ou sai da cena, ou ultrapassa um limite pre-estabelecido de quantas vezes ele pode ser refletido, etc.

4.4 IMPORTANCE SAMPLING

Uma grande desvantagem do MMC é sua relativa demora de convergência. O erro quadrático médio converge a uma taxa $O(1/\sqrt{N})$, necessitando do quádruplo de N amostras para reduzir o erro pela metade (Figura 23) [2]. Idealmente, gostaríamos de um estimador que produza pouca variância, e seja computacionalmente eficiente.

Figura 23: O estimador converge ao resultado correto conforme o número de amostras geradas aumenta. No caso da renderização, o erro estimado é interpretado como ruído. É necessário o quádruplo de N amostras para diminuir o erro pela metade.

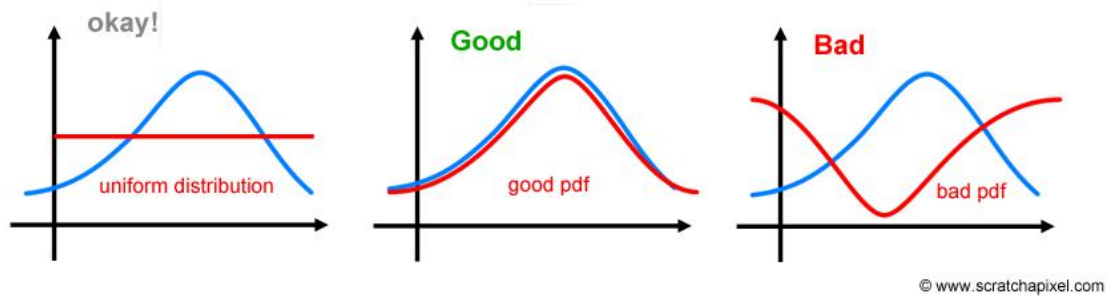


Fonte: scratchapixel.com

Como amostras são geradas a partir da FDP $p(x)$, não há nenhuma garantia quanto à qualidade dessas amostras, apenas que elas serão ajustadas conforme sua densidade. Não seria então mais eficiente escolher uma FDP que gere boas amostras, ou pelo menos próximo disso? Essa técnica é chamada *importance sampling*¹, e faz parte de uma família de técnicas de redução de variância. Se escolhermos $p(x)$ de forma inteligente, a variância do estimador será reduzida. Mas deve-se ter cuidado; uma FDP ruim pode aumentar a variância.

¹Em tradução livre: amostragem por importância

Figura 24: Para reduzir a variância, deve-se escolher uma FDP (em vermelho) que mais se aproxime da função $f(x)$ (em azul). Uma boa escolha para $p(x)$ diminui a variância, enquanto uma má escolha, aumenta.



Fonte: scratchapixel.com

A melhor escolha possível para a FDP é $p^*(x) = c(x)$, de forma que c é proporcionalmente constante

$$c = \frac{1}{\int_{\Omega} f(x) dx}. \quad (4.15)$$

A constante c garante que $p(x)$ é normalizada. A FDP $p^*(x)$ fornece um estimador com variância nula, uma vez que o valor estimado seria igualmente constante. Na prática, essa densidade não pode ser usada, pois é necessário saber de antemão o valor da integral avaliada para-se calcular c (e não faria sentido usar o MMC, nesse caso). Contudo, se escolhermos uma $p(x)$ com uma forma similar à $f(x)$, a variância será reduzida (Figura 24). Podemos escolher alguma das alternativas para contornar esse problema:

1. Aproximar alguma porção de $f(x)$ tal qual a função $g(x) = f(x)p(x)$ possa ser integrada analiticamente.
2. Construir uma aproximação discreta para $f(x)$.
3. Aproximar $f(x)$ por uma série de Taylor.

4.4.1 Gerando Amostras

Até o momento, assumimos que a geração de amostras em direções aleatórias sobre o hemisfério é suficiente para aproximar a equação de renderização. Conceitualmente, esta abordagem funciona como esperado, na prática, ela é ineficiente.

Dependendo das configurações da cena, muitos caminhos podem ser gerados sem sequer atingir uma fonte de luz até sua finalização, aumentando a variância e o tempo de convergência. Pior, se levarmos em consideração as qualidades do material, e dependendo de como a micro-geometria está orientada, pode levar um bom tempo até que um caminho válido seja gerado. Além disso, mesmo que um raio seja direcionado para uma região de grande radiância, sua contribuição pode ser ínfima, dependendo da angulação do raio em relação à superfície, e as propriedades do material. Igualmente, direcionar o raio levando em conta a configuração da micro-geometria não garante que uma fonte de luz será atingida. Diversas técnicas podem ser utilizadas para uma convergência mais rápida do resultado, variando conforme a configuração da cena e dos objetos.

4.4.2 Iluminação Direta

Fontes de luz e objetos emissivos dominam a irradiância presente na cena. Podemos aumentar a eficiência do estimador enviando raios diretamente em direção as luzes. Considerando caminhos individuais dos raios, a cada interseção encontrada projetamos um novo raio de sombra em direção a uma fonte de luz. Como a recursão é imediatamente finalizada quando uma luz é encontrada, podemos quantificar a radiância nessa direção analiticamente contanto que a luz escolhida não se encontre obstruída por outros objetos da cena.

Para cada nova interseção de um caminho de um raio, podemos dividir a radiância calculada em duas categorias: a radiância emitida pela superfície do objeto, encontrada pela análise do raio de sombra, e a radiância incidente, encontrada pela análise da continuação do caminho do raio refletido.

$$L_o(p, \omega_o) = \underbrace{L_e(p, \omega_o)}_{\text{radiância emitida}} + \underbrace{L_r(p, \omega_o)}_{\text{radiância incidente}}$$

Partimos da premissa de que apenas luzes contribuem para a quantificação da irradiância, sendo o restante dos objetos da cena, não emissivos.

- Objetos não emissivos: $L_e(p, \omega_o) = 0, L_r(p, \omega_o) > 0$
- Fontes de luz: $L_e(p, \omega_o) > 0, L_r(p, \omega_o) = 0$

Por conseguinte, podemos calcular a irradiância para objetos não emissivos como:

$$L_o(p, \omega_o) = \int_{2\pi} f_r(p, \omega_o, \omega_i) (L_e(p, \omega_i) + L_i(p, \omega_i)) \cos(\theta_i) d\omega_i$$

Uma vez que os resultados da iluminação direta e indireta são independentes uns dos outros, podemos separar ambos para uma maior eficiência computacional:

$$L_o(p, \omega_o) = \int_{2\pi} f_r(p, \omega_o, \omega_i) L_i(p, \omega_i) \cos(\theta_i) d\omega_i + \int_{2\pi} f_r(p, \omega_o, \omega_i) L_e(p, \omega_i) \cos(\theta_i) d\omega_i$$

Note que existe a possibilidade do raio refletido ser direcionado na mesma direção da luz escolhida para o raio de sombra da iluminação direta. Deve-se ter cuidado, uma vez que estaríamos adicionando a contribuição da iluminação da mesma luz, duas vezes (veremos como resolver esse problema mais a frente).

Contudo, nem todas as luzes são tratadas igualmente. Luzes como *point lights*, *directional lights*, etc., não descrevem entidades reais, mas sim abstrações de efeitos luminosos para uma simplificação do processo de sombreamento. Desconsiderando esse fator, ainda podemos incorporá-las na equação da iluminação direta, bastando a avaliação de uma única amostra para resolver o problema.

O modelo mais comumente utilizado em uma renderização PBR é o de *area lights*², que como o nome sugere, é descrita pela área de sua superfície. Para esse caso, é conveniente computar a iluminação direta pela forma da área da integral (este fato advém da própria definição do ângulo sólido):

$$L_o(p, \omega_o) = \int_{2\pi} f_r(p, \omega_o, \omega_i) L_i(p, \omega_i) \cos(\theta_i) d\omega_i$$

²Em tradução livre: luz de área

$$+ \int_A f_r(p, \omega_o, \omega_i) L_e(p', -\omega_i) \frac{\cos(\theta_i) \cos(\theta')}{\|p' - p\|^2} V(p, p') dA$$

Onde p' é o ponto de interseção sobre a superfície da luz escolhida, θ' o ângulo entre a direção do raio de sombra e a reta normal à superfície da luz, e $V(p, p')$ o fator de visibilidade:

$$V(p, p') = \begin{cases} 0, & \text{se a luz não estiver visível a partir de } p. \\ 1, & \text{do contrário.} \end{cases}$$

Dada uma FDP q , que descreve a densidade de probabilidade de escolhermos um ponto sobre a superfície da luz, podemos estimar a contribuição da iluminação direta por:

$$L_{direct}(p, \omega_i) \approx \frac{1}{N} \sum_{i=1}^N \frac{1}{q(p')} f_r(p, \omega_o, \omega_i) L_e(p', -\omega_i) \frac{\cos(\theta_i) \cos(\theta')}{\|p' - p\|^2} V(p, p')$$

Como estamos escolhendo um ponto qualquer sobre a superfície da luz de área, temos que a densidade é uniforme sobre todo o espaço amostral. Dessa maneira, podemos rescrever a FDP dessa distribuição tal qual $q(p') = 1/A$:

$$L_{direct}(p, \omega_i) \approx \frac{A}{N} \sum_{i=1}^N f_r(p, \omega_o, \omega_i) L_e(p', -\omega_i) \frac{\cos(\theta_i) \cos(\theta')}{\|p' - p\|^2} V(p, p')$$

Um detalhe importante a ser considerado é que todas as luzes possuem um peso diferente de contribuição, dependendo das configurações da cena. Algumas são mais brilhantes, outras possuem uma área de superfície maior ou estão localizadas mais distante, etc. Novamente, podemos amostrar todas as luzes uniformemente, mas devemos considerar que cada uma pode ter uma distribuição diferente de como os pontos são espalhados sobre a superfície. Sendo $q_L(k)$ a função massa de probabilidade (FMP) usada para selecionar uma fonte de luz k dentre N_L luzes, temos que a FDP de um ponto p' ser selecionado a partir dessa luz é dependente da escolha de k . Isto é, $q(y|k)$:

$$L_{direct}(p, \omega_i) \approx \frac{1}{N} \sum_{i=1}^N \frac{1}{q_L(k)q(y|k)} f_r(p, \omega_o, \omega_i) L_e(p', -\omega_i) \frac{\cos(\theta_i) \cos(\theta')}{\|p' - p\|^2} V(p, p')$$

Como estamos amostrando as luzes uniformemente, temos que $q_L(k) = \frac{1}{N_L}$ e $q(y|k) = \frac{1}{A_{L_k}}$:

$$L_{direct}(p, \omega_i) \approx \frac{N_L}{N} \sum_{i=1}^N A_{L_k} f_r(p, \omega_o, \omega_i) L_e(p', -\omega_i) \frac{\cos(\theta_i) \cos(\theta')}{\|p' - p\|^2} V(p, p')$$

Outra maneira de abordar o problema é escolher a luz através de uma heurística que melhor se adapte para cada caso. Mas para cenas mais simples, escolher a luz de maneira uniforme funciona bem.

4.4.3 Iluminação Indireta

Analisando a equação de renderização para a porção da iluminação indireta

$$L_{indirect}(p, \omega_i) = \int_{2\pi} f_r(p, \omega_o, \omega_i) L_i(p, \omega_i) \cos(\theta_i) d\omega_i, \quad (4.16)$$

notamos que ela é governada por três fatores: a BRDF, a radiância incidente e o fator de atenuação de Lambert. Tal como no caso da iluminação direta, gostaríamos de gerar amostras direcionadas para regiões com um maior potencial de contribuição para a iluminação. Mas diferente da iluminação direta, tais regiões são desconhecidas para a iluminação indireta, a princípio.

Ademais, como a análise de cada um dos três fatores depende do direcionamento do raio gerado, não há garantia de que o resultado retornado seja o melhor possível, mesmo que a amostra vise maximizar um dos fatores. A qualidade da amostra não dependerá apenas de um componente da equação, como no caso da iluminação direta. Dependendo das propriedades do material de que o objeto é composto, diferentes estratégias podem ser adotadas para a geração da amostra:

- Amostragem uniforme: ineficiente para superfícies polidas.
- Pelo fator do cosseno: ideal para superfícies difusas.
- Pela BRDF: ajustável à micro-geometria da superfície. Útil para a porção especular da iluminação.

- Proporcional a radiância incidente: desconhecida a princípio, mas pode ser pré-computada ou encontrada utilizando certas técnicas (e.g. mapeamento de fótons).
- Uma combinação das demais: prática, pois a cena pode conter uma variedade de materiais e objetos.

Para os propósitos deste trabalho, iremos gerar amostras a partir de uma combinação entre o fator do cosseno e a BRDF. Relembrando o modelo de Cook-Torrance, sabemos que a reflexão da luz é descrita tanto pela porção especular, quanto pela porção difusa:

$$f_r(p, \omega_i, \omega_o) = k_d \frac{c}{\pi} + \frac{DFG}{4(\omega_o \cdot n)(\omega_i \cdot n)}$$

Contudo, aqui há um dilema: devido à complexidade recursiva do algoritmo, só podemos nos dar o luxo de gerar uma amostra por iteração! Em outras palavras, devemos escolher entre gerar uma nova amostra para a porção difusa, ou uma nova amostra para a porção especular.

Muitos renderizadores, como os usados em motores gráficos de *video games*, contornam este problema realizando um mapeamento prévio da irradiância presente na cena. Esta técnica amortiza o custo da avaliação da porção difusa, uma vez que esta independe do ângulo entre o observador e a superfície. Para achar a radiância em uma dada direção, basta uma checagem do mapa de irradiância. Dessa maneira, sobra apenas a porção especular para ser calculada em tempo de execução.

Neste trabalho, optamos por avaliar ambas em tempo de execução. Apesar dessa escolha ser mais custosa que a opção anterior, há a vantagem de que qualquer mudança na iluminação da cena pode ser recalculada dinamicamente. Todavia, ainda não nos livramos da restrição de que só conseguimos gerar uma amostra por iteração, necessitando da avaliação de no mínimo dois 'caminhos' para os raios primários.

4.4.3.1 Método da Transformada Inversa

Podemos gerar amostras para uma variável aleatória X através do Método da Transformada Inversa.

Dada uma variável aleatória X com uma função de distribuição acumulada $F_X(X)$, então uma variável aleatória $\xi = F_X(X)$ tem uma distribuição uniforme em um intervalo de $[0, 1]$. A ideia por trás do método é inverter a distribuição. Primeiro geramos ξ através de uma distribuição uniforme $U(0, 1)$, e então calculamos $F_X^{-1}(X)$, a transformada inversa de $F_X(X)$. Consequentemente, $F_X^{-1}(\xi) = X$ apresenta a mesma distribuição que da variável aleatória X .

Mais precisamente, podemos gerar amostras para uma variável aleatória contínua X com FDP $p(x)$ seguindo os passos:

1. Computar a FDA $F_X(x) = \int_0^x p(x')dx'$.
2. Computar a inversa $F_X^{-1}(x)$.
3. Gerar ξ por uma distribuição uniforme $U(0, 1)$.
4. Computar $X = F_X^{-1}(\xi)$.

4.4.3.2 Transformada Multidimensional

Suponha que queiramos gerar amostras para uma função conjunta de variáveis aleatórias contínuas X e Y , com FDP $p(x, y)$. Algumas vezes, densidades multidimensionais são separáveis e podem se expressas como um produto entre duas funções de densidades independentes $p(x, y) = p_x(x)p_y(y)$. Nesse caso, podemos achar valores para X e Y gerando amostras por p_x e p_y , respectivamente. Muitas densidades não são separáveis, no entanto.

Dada uma função densidade multidimensional $p(x, y)$, a função densidade marginal $p(x)$ pode ser obtida fixando uma das dimensões:

$$p(x) = \int p(x, y)dy. \quad (4.17)$$

Outra maneira de pensar, é que $p(x)$ é a FDP para X apenas. Ou, em outras palavras, é a densidade média para um valor particular x sobre todos os valores possíveis para y .

A função densidade condicional $p(y|x)$ é a função densidade para y , dado que um valor particular para x foi escolhido:

$$p(y|x) = \frac{p(x, y)}{p(x)}. \quad (4.18)$$

Para gerarmos amostras para uma função conjunta de variáveis aleatórias, devemos primeiro computar a função densidade marginal para alguma das variáveis aleatórias. Com essa função densidade, podemos então gerar amostras para a variável aleatória, isolando-a em um valor em particular. Com esse valor fixado, podemos então computar a função densidade condicional para as demais variáveis aleatórias que não foram escolhidas, e então gerar amostras para essas variáveis.

4.4.3.3 Amostragem da Difusa

Pode-se pensar que, a princípio, uma amostragem uniforme sobre o horizonte que engloba a reta normal à superfície é a melhor escolha para a porção difusa. Entretanto, toda radiância incidente é suavizada pelo fator do cosseno (Equação 4.16).

Uma fonte de luz de grande intensidade luminosa pode acabar apresentando uma radiância menor do que uma fonte de luz mais fraca, se essa estiver angulada quase que paralelamente à superfície. Uma ideia promissora é gerar amostras levando-se em conta onde o fator do cosseno é maior. Isto é, gerar amostras que estejam mais alinhadas com a reta normal ao plano.

Para gerar uma amostra alinhada com a distribuição do cosseno, é preciso uma FDP que seja proporcional à função (Figura 24):

$$p(\omega) \propto \cos(\theta)$$

No entanto, descrever o direcionamento de um raio pelo ângulo sólido não é prático, mesmo que a área projetada sobre a superfície da esfera unitária seja infinitesimalmente pequena. É conveniente descrever a direção do raio por sua representação em

coordenadas cartesianas. Para tal, podemos nos valer da noção de que um ponto $P(r, \theta, \phi)$, descrito em coordenadas esféricas, possa ser convertido em sua respectiva representação cartesiana, tal qual:

$$x = r \sin(\theta) \cos(\phi)$$

$$y = r \sin(\theta) \sin(\phi)$$

$$z = r \cos(\theta)$$

Sendo θ o ângulo da colatitude em relação à base, e ϕ representa o ângulo da longitude. Com as variáveis definidas, conseguimos gerar uma direção válida sobre o hemisfério, e o direcionamento do raio pode então ser descrito por um vetor normalizado $v = (x, y, z)^T$, em coordenadas cartesianas.

Uma derivação importante é que a diferencial do ângulo sólido em relação a sua representação em coordenadas esféricas é dada por

$$d\omega = \sin \theta d\theta d\phi^3,$$

e a função densidade $p(\omega)$ passa a ser descrita em termos de θ e ϕ , podendo ser derivada:

$$p(\omega)d\omega = p(\theta, \phi)d\theta d\phi$$

$$p(\omega) \sin \theta = p(\theta, \phi).$$

Com a função densidade $p(\theta, \phi)$, estamos prontos para gerar amostras através do Método da Transformada Inversa, mas antes é necessário garantir que a função esteja em sua forma normalizada (Equação 4.15).

$$c = \frac{1}{\int_{\Omega} f(x)dx}, f(x) = p(\omega) \propto \cos(\theta)$$

Assim,

$$\int_0^{2\pi} \int_0^{\frac{\pi}{2}} c \cos \theta \sin \theta d\theta d\phi = 1$$

$$2\pi c \int_0^{\frac{\pi}{2}} \cos \theta \sin \theta d\theta = 1$$

$$c = \frac{1}{\pi}$$

³Formulação aqui omitida, pois vai além do escopo dessa seção, mas pode ser averiguada em [11].

Portanto,

$$p(\theta, \phi) = \frac{1}{\pi} \sin \theta \cos \theta.$$

O próximo passo é isolar as variáveis para podermos encontrar suas respectivas funções de densidade:

$$p(\theta) = \int_0^{2\pi} p(\theta, \phi) d\phi = 2 \sin \theta \cos \theta$$

$$p(\phi|\theta) = \frac{p(\theta, \phi)}{p(\theta)} = \frac{\frac{\sin \theta \cos \theta}{\pi}}{2 \sin \theta \cos \theta} = \frac{1}{2\pi}$$

Agora podemos calcular as funções de distribuição acumulada para θ e ϕ , integrando suas funções de densidade sobre os respectivos domínios:

$$F_\theta(\theta) = \int_0^\theta p(\theta) d\theta = \int_0^\theta 2 \sin \theta \cos \theta d\theta = \sin^2 \theta$$

$$F_\phi(\phi|\theta) = \int_0^\phi p(\phi|\theta) d\phi = \int_0^\phi \frac{1}{2\pi} d\phi = \frac{\phi}{2\pi}$$

Finalmente, geramos duas variáveis aleatórias uniformes ξ_0 e ξ_1 e invertemos as funções.

$$F_\theta(\theta) = \sin^2 \theta = \xi_0 \rightarrow \theta = \sin^{-1}(\sqrt{\xi_0}) = \cos^{-1}(\sqrt{1 - \xi_0})$$

$$F_\phi(\phi|\theta) = \frac{\phi}{2\pi} = \xi_1 \rightarrow \phi = 2\pi\xi_1$$

4.4.3.4 Amostragem da Especular

Diferente da porção difusa, o ângulo entre o observador e o ponto de análise tem grande influência no resultado da especular. Idealmente, gostaríamos de uma função densidade que consiga gerar amostras que estejam alinhadas com esse ângulo.

Como visto no capítulo 3, sabemos que a luz é refletida pela reta normal a uma micro-superfície. A avaliação da BRDF, e conseqüentemente o resultado da iluminação especular, vão depender inteiramente do quão alinhada a reta normal da micro-superfície está com a distribuição das retas normais da área da macro-superfície. A solução é construir uma função densidade a partir da própria BRDF. Mais precisamente, a partir da função da distribuição normal, D .

Intuitivamente, há um mérito nessa ideia. Lembre-se de que a NDF é a função densidade da área da micro-superfície em relação ao ângulo sólido e área da macro-superfície. Estamos interessados nas regiões onde a densidade é maior. Ou, em outras palavras, nas regiões onde a reta normal da micro-superfície está mais alinhada com a distribuição. A ideia, portanto, é inverter a lógica da avaliação da NDF. Ao invés de avaliarmos uma direção h em relação à distribuição das retas normais das micro-superfícies, iremos gerar essa nova reta normal para a micro-superfície, e refletiremos a luz a partir dela.

Figura 25: Diferentes estratégias para a renderização de um mesmo objeto metálico e polido. O modelo da direita foi renderizado levando-se em conta a BRDF para gerar amostras para a porção especular, enquanto o da esquerda utilizou-se da abordagem padrão de gerar amostras aleatórias sobre o hemisfério.



Fonte: agraphicsguynotes.com/posts/sample_microfacet_brdf

Lembre-se que a distribuição GGX é dada por:

$$D_{GGX}(n, h, \alpha) = \frac{\alpha^2}{\pi((n \cdot h)^2(\alpha^2 - 1) + 1)^2}.$$

Como mencionado, já derivamos a função densidade normalizada da função da distribuição normal, no capítulo 3:

$$\int_{\Omega} D(h)(n \cdot h) d\omega = 1.$$

Por conseguinte, temos que

$$p(\omega) = \frac{\alpha^2 \cos \theta}{\pi((\alpha^2 - 1) \cos^2 \theta + 1)^2}.$$

Novamente, estamos interessados em gerar amostras com respeito às coordenadas esféricas dadas pelas variáveis θ e ϕ , e não pelo ângulo sólido:

$$p(\theta, \phi) = \frac{\alpha^2 \cos \theta \sin \theta}{\pi((\alpha^2 - 1) \cos^2 \theta + 1)^2}$$

E mais uma vez, podemos usar o Método da Transformada Inversa para achar as variáveis ξ_0 e ξ_1 , e gerar amostras. Primeiro, devemos isolar uma das variáveis e achar sua função densidade:

$$p(\theta) = \int_0^{2\pi} p(\theta, \phi) d\phi = \int_0^{2\pi} \frac{\alpha^2 \cos \theta \sin \theta}{\pi((\alpha^2 - 1) \cos^2 \theta + 1)^2} d\phi = \frac{2\alpha^2 \cos \theta \sin \theta}{((\alpha^2 - 1) \cos^2 \theta + 1)^2}$$

Para a variável ϕ , temos que

$$p(\phi|\theta) = \frac{p(\theta, \phi)}{p(\theta)} = \frac{1}{2\pi}$$

Repare que a função densidade para ϕ é a mesma que no caso da porção difusa! Com as funções de densidade, podemos calcular as funções de distribuição acumulada para as variáveis:

$$\begin{aligned} F_\theta(\theta) &= \int_0^\theta p(\theta) d\theta \\ &= \int_0^\theta \frac{2\alpha^2 \cos t \sin t}{(\cos^2 t(\alpha^2 - 1) + 1)^2} dt \\ &= \int_\theta^0 \frac{\alpha^2}{(\cos^2 t(\alpha^2 - 1) + 1)^2} d(\cos^2 t) \\ &= \frac{\alpha^2}{\alpha^2 - 1} \int_0^\theta d \frac{1}{\cos^2 t(\alpha^2 - 1) + 1} \\ &= \frac{\alpha^2}{\alpha^2 - 1} \left(\frac{1}{\cos^2 \theta(\alpha^2 - 1) + 1} - \frac{1}{\alpha^2} \right) \\ &= \frac{\alpha^2}{\cos^2 \theta(\alpha^2 - 1)^2 + (\alpha^2 - 1)} - \frac{1}{\alpha^2 - 1} \end{aligned}$$

Invertendo as funções:

$$F_\theta(\theta) = \xi_0 \rightarrow \theta = \arccos \left(\sqrt{\frac{1 - \xi_0}{\xi_0(\alpha^2 - 1) + 1}} \right) = \arctan \left(\alpha \sqrt{\frac{\xi_0}{1 - \xi_0}} \right)$$

$$F_\phi(\phi|\theta) = \xi_1 \rightarrow \phi = 2\pi\xi_1$$

4.5 MULTIPLE IMPORTANCE SAMPLING

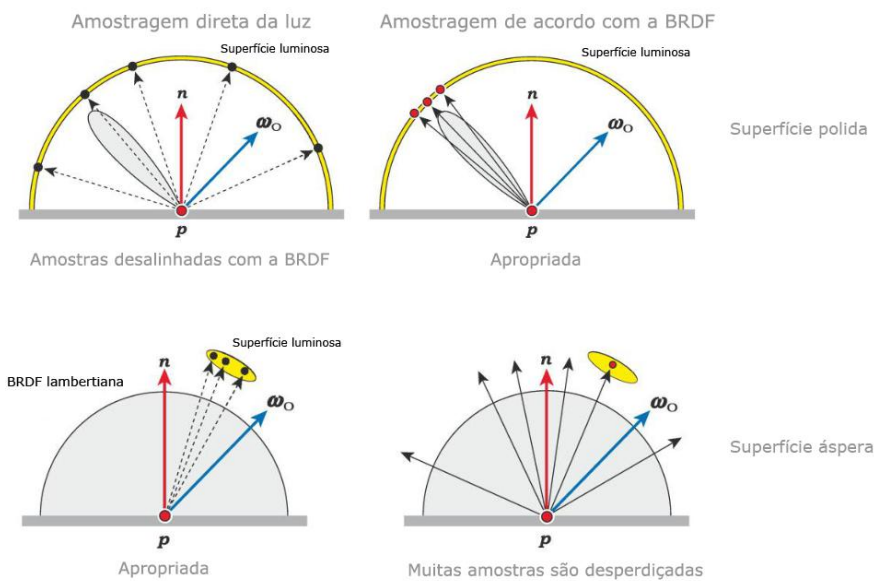
Agora que conseguimos gerar amostras para as iluminações direta e indireta, uma questão pertinente é como balanceá-las. Se fossemos adicionar a contribuição de cada

uma de forma independente, poderíamos acabar contabilizando uma quantidade excessiva de radiância. Uma primeira abordagem a ser considerada é atribuir o mesmo peso para ambas as partes. Dessa maneira, temos que a iluminação total contabilizada por uma amostra é dada pela média da iluminação direta e indireta

$$L_o(p, \omega_o) \approx \frac{L_{indirect} + L_{direct}}{2}.$$

Apesar de ser uma escolha mais sensata do que simplesmente contabilizar cada porção separadamente, ainda há o problema de que a configuração da cena tem grande influência no produto final da iluminação (Figura 26).

Figura 26: Dependendo da configuração da cena, cada porção da iluminação terá um peso maior de contribuição do que a outra.



Fonte: Matthias Teschner - Advanced Computer Graphics Path Tracing

Uma configuração ideal seria aonde cada parte da iluminação tenha sua contribuição ajustada pelo peso de importância, variando conforme o cenário. Uma solução proposta por Veach [12], é a técnica conhecida como *multiple importance sampling* (MIS).

MIS parte do princípio de que podemos combinar m diferentes estratégias de amostragem para produzir um único estimador não enviesado, combinando e ajustando cada estratégia por sua função densidade. Matematicamente, isso pode ser

expresso como:

$$F_{mis} = \sum_{i=1}^m \frac{1}{n_i} \sum_{j=1}^{n_i} w_i(X_{i,j}) \frac{f(X_{i,j})}{p_i(X_{i,j})}, \quad (4.19)$$

nas quais $X_{i,j}$ são variáveis aleatórias independentes escolhidas por uma função densidade p_i , e w_i é a heurística utilizada como peso de ajuste em relação à função densidade escolhida.

Para o estimador, as seguintes condições devem ser verdadeiras:

- $\sum_{i=1}^n w_i(x) = 1$, se $f(x) \neq 0$.
- $w_i(x) = 0$, se $p_i(x) = 0$.

4.5.1 Heurística Balanceada

Para o estimador, achar uma boa heurística é crucial para o balanceamento das amostras. Um modelo sugerido por Veach é a Heurística Balanceada, definida como

$$w_i(x) = \frac{n_i p_i(x)}{\sum_k m_k p_k(x)}. \quad (4.20)$$

Temos então a seguinte formulação para o estimador:

$$F_{mis} = \sum_{i=1}^m \sum_{j=1}^{n_i} \frac{f(X_{i,j})}{\sum_k m_k p_k(x)}$$

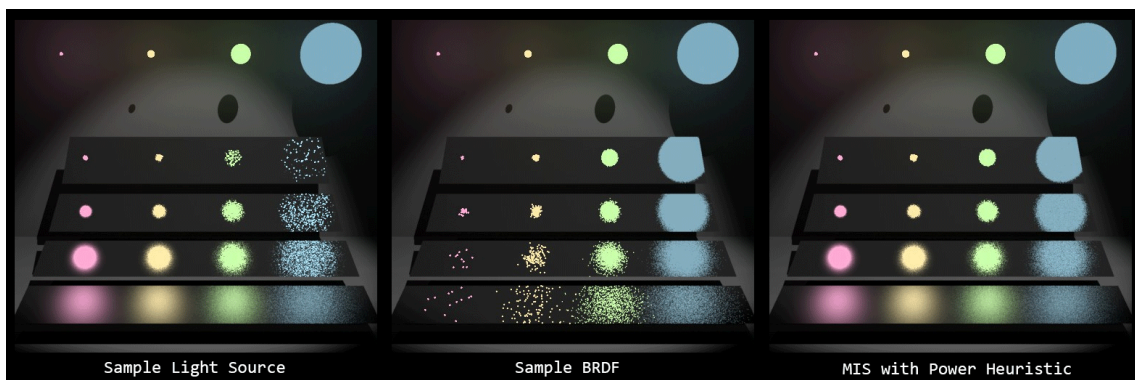
4.5.2 Heurística da Potência

Outra heurística sugerida por Veach, é a Heurística da Potência:

$$w_i(x) = \frac{[n_i p_i(x)]^\beta}{\sum_k [m_k p_k(x)]^\beta}. \quad (4.21)$$

Note que a Heurística Balanceada é um caso especial da Heurística da Potência, quando $\beta = 1$. De acordo com Veach, a Heurística da Potência provê uma taxa de convergência mais rápida do que a Heurística Balanceada. Ele também constatou que $\beta = 2$ é o suficiente para a maioria dos casos, em que valores maiores começam a apresentar retornos decrescentes.

Figura 27: Os efeitos do *multiple importance sampling*, visualizados. No exemplo das imagens, temos 4 fontes de luz variando em tamanho, e 4 placas metálicas variando conforme seus respectivos coeficientes de aspereza (α). Na imagem a esquerda, a amostragem da iluminação é feita considerando-se apenas a iluminação direta da luz. Na imagem ao centro, a amostragem da iluminação é feita considerando-se apenas a iluminação indireta, pela BRDF. Na imagem a direita, as amostras das iluminações são balanceadas e combinadas conforme a MIS, usando a Heurística da Potência.



Fonte: Veach e Guibas - Optimally Combining Sampling Techniques for Monte Carlo Rendering

5 ESTRUTURAS DE ACELERAÇÃO

Apesar dos avanços tecnológicos mais recentes no setor de hardware, *Path tracing* ainda é uma técnica computacionalmente cara. Mesmo com os conceitos e estratégias derivados até aqui, ainda teríamos problemas em renderizar cenas em taxas interativas. De acordo com Fujimoto [6], uma grande parte do tempo do algoritmo é gasta realizando-se o cálculo de interseções entre os raios e os objetos, chegando até 95% para cenas mais complexas. Uma prática padrão para resolver este problema é utilizar algum tipo de estrutura de aceleração, tornando o processo mais rápido.

5.1 PARTICIONAMENTO DE OBJETOS

Testes de interseção são realizados sobre as primitivas de que o objeto é composto. Um método simples, mas eficaz, é enclausurar as primitivas em grupos, geralmente envoltas por algum tipo de volume. Assim, passamos a testar primeiro a interseção entre o raio e o volume. Se o raio não interceptar o volume, temos a garantia de que também não intercepta nenhuma primitiva contida no volume, economizando tempo de computação em testes desnecessários.

5.1.1 Bounding Box

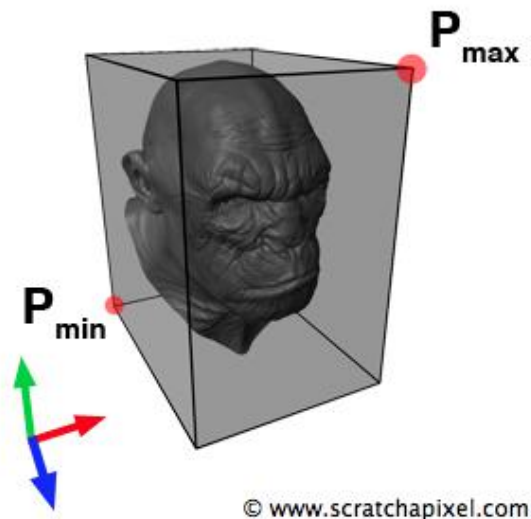
Um *bounding volume* é um volume fechado que encapsula completamente um objeto. Uma *bounding box*¹, como o nome sugere, é um *bounding volume* em formato de uma caixa.

Geralmente, estamos interessados em delimitar a caixa com o menor volume possível que encapsula o objeto. Em muitas aplicações, a *bounding box* é alinhada com os eixos das coordenadas cartesianas, conhecida como *axis-aligned bounding box* (AABB). AABBs são mais eficientes para realizar testes de interseção do que *bounding boxes* convencionais, mas com a desvantagem de que certas transformações espaciais aplicadas no objeto necessitem recalcular o dimensionamento da caixa.

¹Em tradução livre: caixa delimitadora

Achar a AABB é relativamente simples; podemos vasculhar todos os vértices do objeto, e armazenar as coordenadas dos vértices com os menores e maiores valores alinhados com os eixos. Ao final da busca, teremos dois pontos distintos P_{min} e P_{max} , representando os pontos com os menores e maiores valores, respectivamente. Podemos utilizar os pontos mínimo e máximo para delimitar o dimensionamento da caixa. Qualquer outro ponto no espaço que se encontre entre P_{min} e P_{max} , deve estar contido na caixa.

Figura 28: A AABB do modelo da cabeça de um gorila.



Fonte: scratchapixel.com

5.1.2 Hierarquia de Volumes

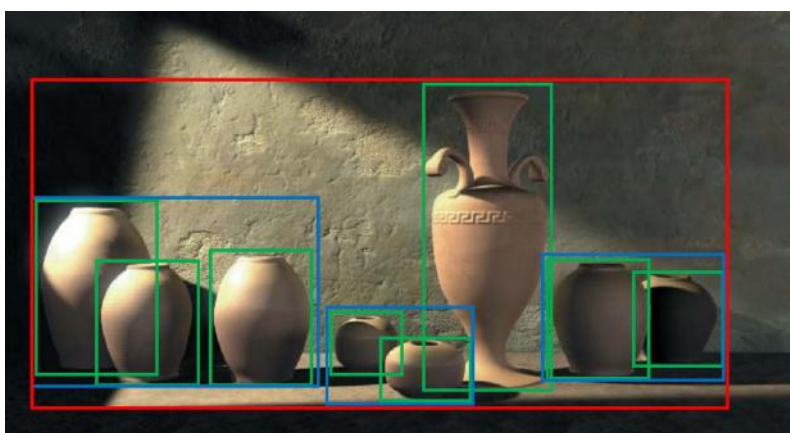
Uma cena é geralmente composta por múltiplos objetos, mas também há casos em que um único objeto é composto por múltiplos objetos menores (ou uma grande magnitude de primitivas). Estendendo a ideia anterior, podemos enclausurar esses objetos menores em suas próprias AABBs. Subindo na hierarquia dos objetos, passamos a enclausurar as AABBs dentro de outras AABBs, criando um conjunto de "caixas dentro de caixas". Essa hierarquia é conhecida como *Bounding Volume Hierarchy*² (BVH).

²Hierarquia de Volumes

A BVH é uma estrutura de dados do tipo árvore, onde cada folha contém uma caixa. Dentro destas caixas, temos um número arbitrário máximo de objetos ou primitivas que podem ser contidas em uma única caixa (veremos mais a frente como esse número influencia o desempenho do renderizador). No nó raiz da BVH, temos a caixa com o menor volume que enclausura completamente o objeto principal. Nos nós filhos, temos caixas com volumes menores do que o do pai, cada qual enclausurando grupos distintos de objetos menores.

Assim como uma busca padrão em uma árvore, podemos simplificar os testes de interseção realizando uma busca pela BVH. Se um raio não intercepta uma caixa de um certo nó, este raio também não intercepta as caixas contidas nos filhos desse nó. Ao final da busca, podemos descartar a verificação de todas as caixas cujos pais não são interceptados pelo raio. Ou, se ao final da busca averiguarmos que o raio intercepta um nó folha, passamos a testar as interseções entre o raio e os objetos/primitivas contidas na caixa desta folha.

Figura 29: Uma representação 2D de uma hierarquia de volumes. A caixa vermelha enclausura todos os objetos da cena. As caixas azuis enclausuram grupos distintos de objetos menores. As caixas verdes, contidas nas caixas azuis, enclausuram esses objetos individualmente.

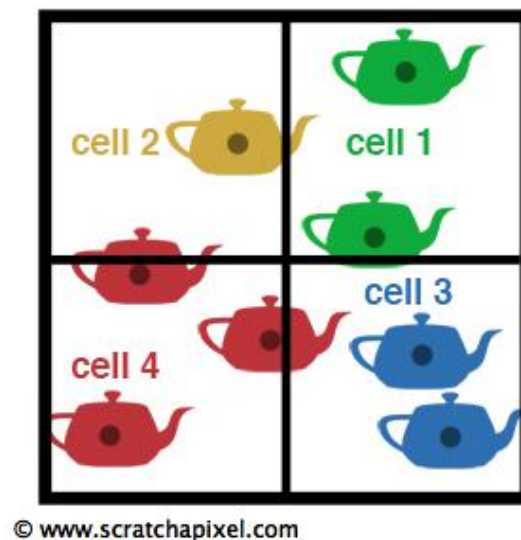


Fonte: jacco.ompf2.com/2022/04/13/how-to-build-a-bvh-part-1-basics/

5.1.3 Heurística da Área de Superfície

Um grande fator que influencia a busca pela BVH, e conseqüentemente o desempenho do renderizador, é como a hierarquia é organizada. Em uma primeira abordagem, poderíamos dividir os volumes pela mediana das dimensões espaciais das caixas. Se os objetos são igualmente espaçados pela cena, esta abordagem é aceitável (Figura 30). Contudo, podemos acabar encontrando uma situação de um *teapot in a stadium*³, onde um objeto muito denso encontra-se localizado em uma pequena parte do cenário. Nesse caso, os benefícios da estrutura de aceleração seriam praticamente irrelevantes.

Figura 30: Divisão da hierarquia de volumes pela mediana das dimensões espaciais das caixas.



Fonte: scratchapixel.com

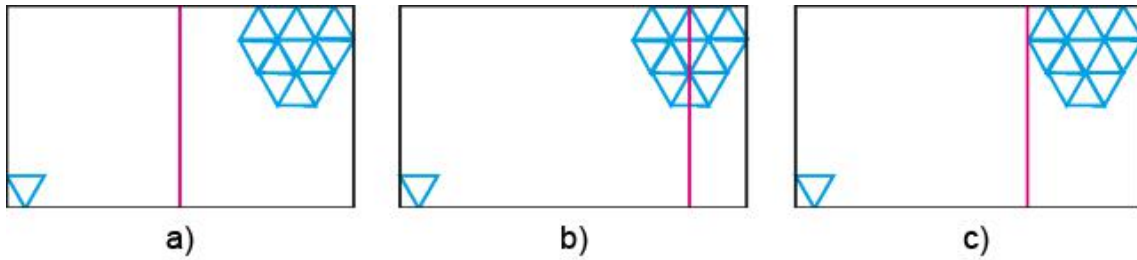
Podemos contornar esse problema particionando os objetos de uma forma mais inteligente. O modelo de heurística da área de superfície (SAH⁴) almeja estimar o custo computacional ao se dividir um volume a partir de um certo ponto, considerando o tempo que um raio levaria para atravessar um nodo, e o tempo para realizar os testes de interseção sobre as primitivas contidas no nodo. O objetivo é achar o

³Bule de chá em um estádio.

⁴Surface Area Heuristic.

melhor ponto de divisão que minimize o custo.

Figura 31: Exemplos hipotéticos para o particionamento do volume em três pontos diferentes: a) pela mediana da caixa. b) pela média das posições das primitivas. c) pela SAH, considerando o custo para o raio atravessar os novos volumes mais o custo de se avaliar as interseções sobre as primitivas em cada nodo.



Fonte: medium.com/@bromanz/how-to-create-awesome-accelerators-the-surface-area-heuristic-e14b5dec6160

A premissa por trás da SAH é que, a qualquer ponto que fossemos dividir um volume, poderíamos parar a divisão e criar uma folha na hierarquia considerando a região e geometria em questão. Nesse caso, qualquer raio que passe por essa região terá de ser testado para interseções para todas as primitivas locais, com um custo

$$c = \sum_{i=1}^N t_{\text{isect}}(i).$$

De forma que N é o número de primitivas contidas no volume, e $t_{\text{isect}}(i)$ é o tempo para se avaliar a interseção entre o raio e a i -ésima primitiva. A outra opção seria continuar a divisão, com custo

$$c(A, B) = t_{\text{trav}} + p_A \sum_{i=1}^{N_A} t_{\text{isect}}(a_i) + p_B \sum_{i=1}^{N_B} t_{\text{isect}}(b_i), \quad (5.1)$$

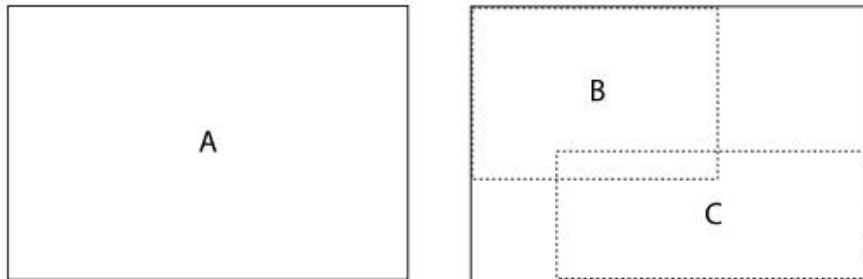
tal que t_{trav} é o tempo para atravessar a região e determinar por quais sub-regiões o raio passa, p_A e p_B são as probabilidades do raio passar por cada uma das sub-regiões, respectivamente, e N_A e N_B são os números de primitivas contidas nas sub-regiões A e B. A escolha do ponto para o particionamento afeta tanto as probabilidades do raio atravessar ambas as regiões, como o número de primitivas contidas em cada uma.

Na prática, o tempo para testar a interseção $t_{sect}(i)$ não é linear para todas as primitivas. Mas levar esse fator em consideração traria pouco benefício para melhorar o desempenho do programa. Dessa maneira, assumimos que o tempo é o mesmo para todas as primitivas de uma mesma classe. As probabilidades p_A e p_B podem ser calculadas pelo conceito de probabilidades geométricas. Para um volume convexo A contido em outro volume convexo B, a probabilidade condicional de que um raio que atravesse B também atravesse A é dada pela proporção das áreas de suas superfícies, s_A e s_B :

$$p(A|B) = \frac{s_A}{s_B}.$$

Como estamos interessados em determinar o custo de um raio atravessar essas regiões, podemos usar esse resultado diretamente.

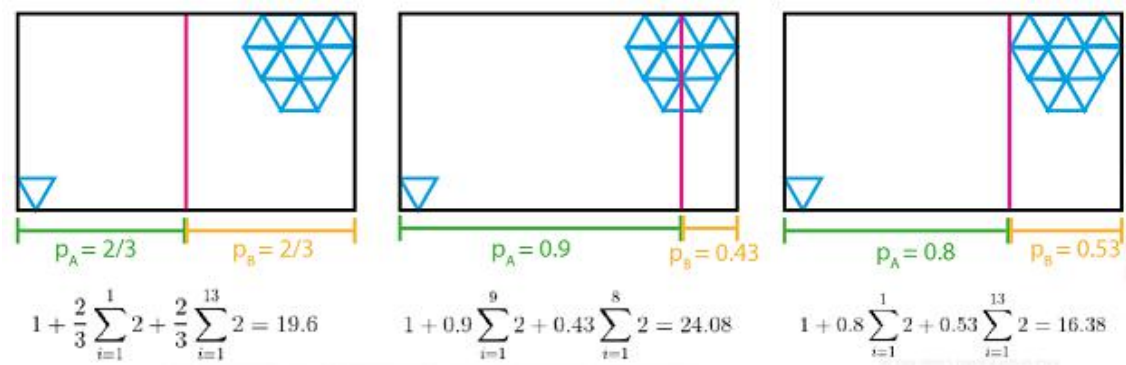
Figura 32: Se um volume com área de superfície s_A for particionado em duas sub-regiões com áreas s_B e s_C , então a probabilidade de um raio que atravesse A também atravesse B e C é dada por s_B/s_A e s_C/s_A , respectivamente.



Fonte: Physically Based Rendering: From Theory to Implementation

Voltando ao exemplo anterior (Figura 31), agora conseguimos calcular o custo de dividir a hierarquia para cada ponto. A fim de exemplificação, assumindo que $t_{trav} = 1$ e $t_{sect}(i)=2$, temos:

Figura 33: O ponto de divisão utilizando a SAH é a melhor escolha para minimizar o custo de particionamento.



Fonte: medium.com/@bromanz/how-to-create-awesome-accelerators-the-surface-area-heuristic-e14b5dec6160

6 RESULTADOS E CONCLUSÕES

Desenvolver um *path tracer* é um processo contínuo de aprimoramento. Há sempre um novo patamar a ser alcançado, ainda mais com os avanços tecnológicos e acadêmicos no ramo da computação gráfica nos últimos anos.

Quando iniciei este trabalho, só conseguia imaginar que todos os métodos e técnicas vistos até aqui seriam suficientes para alcançar o objetivo de criar um *path tracer* interativo. Apesar de não ter abordado alguns tópicos mais a fundo, especialmente quando se trata do estado da arte, devo dizer que fiquei satisfeito com os resultados.

Abaixo, podemos ver algumas cenas geradas pelo meu próprio renderizador. As especificações do hardware utilizado também estão disponíveis. O programa foi escrito em C++ e GLSL, linguagem de *shader* usada para rodar o código diretamente na placa de vídeo. A tradução do algoritmo do transporte da luz e simulação de Monte Carlo para GPU é ajustada para acomodar o código ao novo ambiente paralelo. Os detalhes da implementação podem ser conferidos no repositório: github.com/victorlbauer/GPU-Path-Tracer

Todas as cenas foram renderizadas em uma resolução de 1280 *pixels* de largura por 720 de altura. A profundidade máxima dos raios é 8 rebatidas e a cada iteração é acumulada 1 amostra por *frame*.

Especificações	
Processador	Intel(R) Core(TM) i5-6500 @ 3.20GHz (4 CPUs)
DRAM	8192GB
Placa Mãe	H110M-C/BR
Sistema Operacional	Windows 10 Pro 64-bit
Placa de Vídeo	Gigabyte Radeon R9 280x
VRAM	3GB
OpenGL	4.3

Figura 34: Um crânio metálico e polido (aprox. 10.000 triângulos), renderizada com uma taxa média de 40 quadros por segundo (FPS).

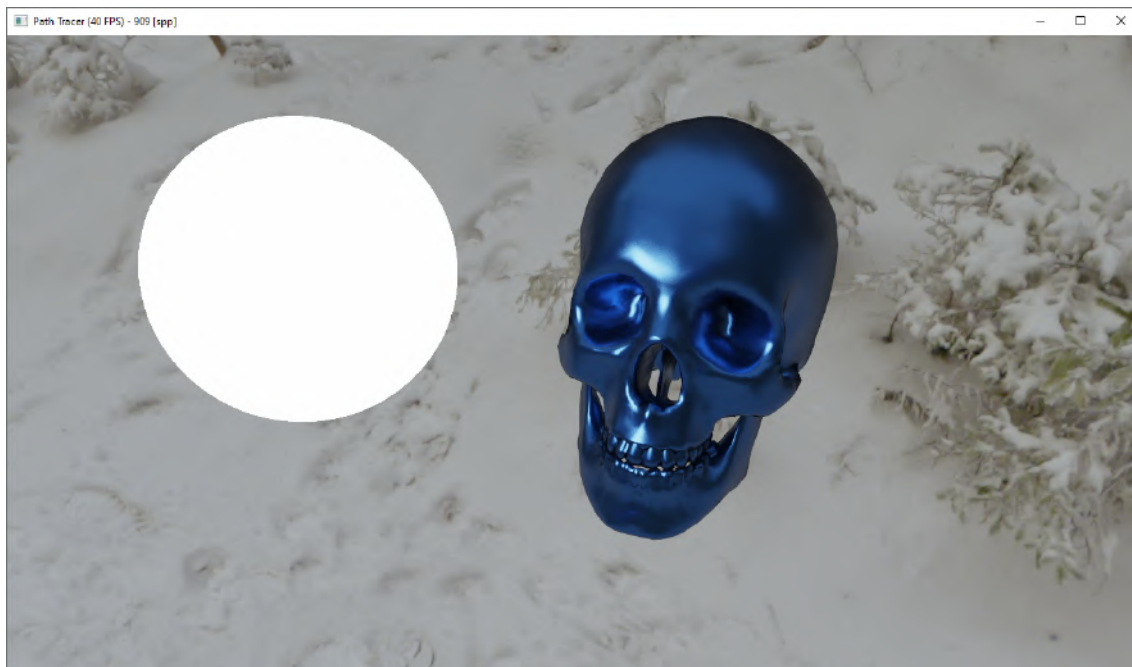


Figura 35: Um modelo de alta resolução do Dragão de Stanford (aprox. 900.000 triângulos), renderizado com uma taxa média de 18 FPS.



Figura 36: Um mesmo modelo do *Utah Teapot*, alinhados horizontalmente. Da esquerda para a direita, o coeficiente de aspereza do material é aumentado, fazendo com que a imagem refletida fique mais nítida.



Figura 37: O modelo *Suzanne*, renderizado sem uma imagem de fundo. A combinação da iluminação direta e indireta (MIS) torna a imagem nítida mais rapidamente.

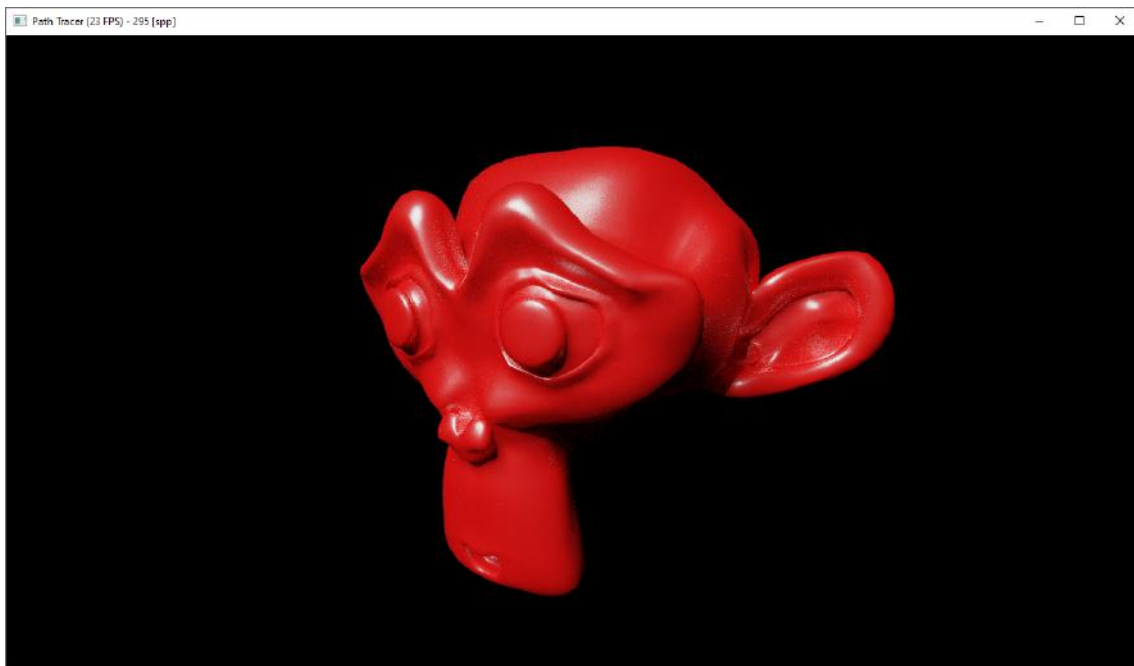


Figura 38: O clássico modelo de uma *Cornell Box*, demonstrando as capacidades do renderizador, mesmo sem uma boa estrutura para organizar objetos de uma mesma classe.

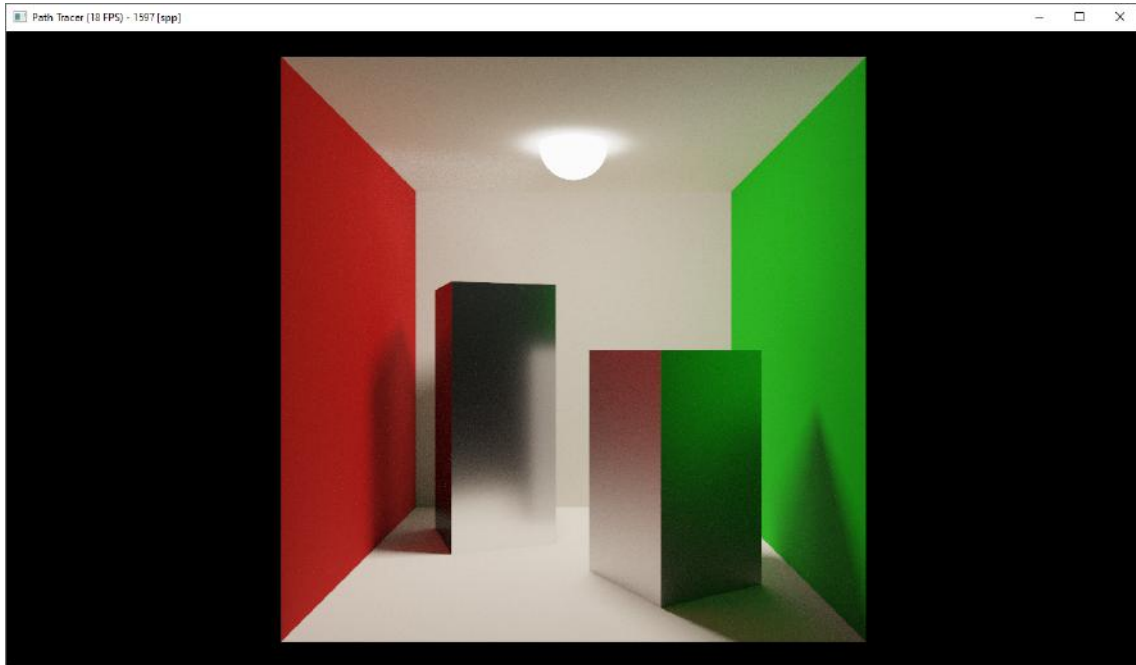
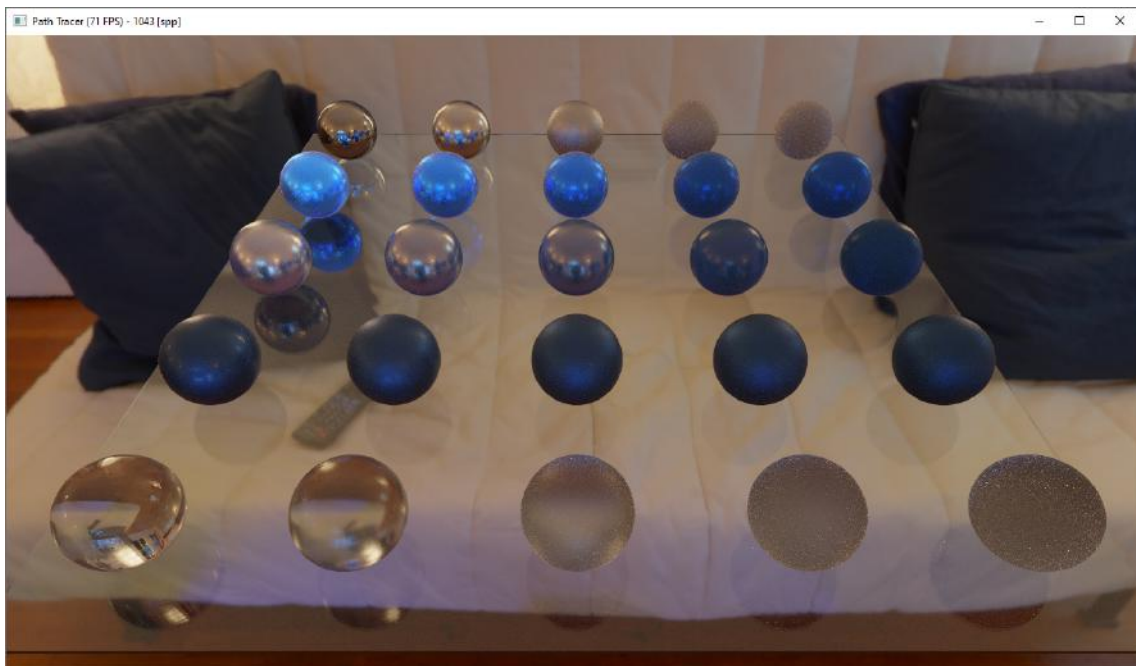


Figura 39: Demonstração de diferentes materiais suportados pelo renderizador. A modelagem de materiais como vidro (BTDF ao invés da BRDF) não foram abordados neste trabalho.



REFERÊNCIAS

- [1] AKENINE-MILLER, T., HAINES, E., E HOFFMAN, N. *Real-Time Rendering, Fourth Edition*, 4th ed. A. K. Peters, Ltd., USA, 2018.
- [2] COLBERT, M., E DIGITAL, W. Importance sampling for production rendering.
- [3] COOK, R. L., E TORRANCE, K. E. A reflectance model for computer graphics. *ACM Trans. Graph.* 1, 1 (jan 1982), 7–24.
- [4] DE VRIES, J. *Learn OpenGL - Graphics Programming*, 1 ed. 17 Junho 2020.
- [5] FERWERDA, J. A. Three varieties of realism in computer graphics. In *Human Vision and Electronic Imaging VIII (2003)*, B. E. Rogowitz e T. N. Pappas, Eds., vol. 5007, International Society for Optics and Photonics, SPIE, pp. 290 – 297.
- [6] FUJIMOTO, A., E IWATA, K. Accelerated ray tracing. In *Computer Graphics (Tokyo, 1985)*, T. L. Kunii, Ed., Springer Japan, pp. 41–65.
- [7] HOFMANN, G. R. Who invented ray tracing? vol. 6, pp. 120 – 124.
- [8] KAJIYA, J. T. The rendering equation. In *Computer Graphics (1986)*, pp. 143–150.
- [9] MATT PHARR, WENZEL JAKOB, G. H. *Physically Based Rendering: From Theory to Implementation*, 3 ed. 2018.
- [10] SCHLICK, C. An inexpensive brdf model for physically-based rendering. *Computer Graphics Forum* 13, 3 (1994), 233–246.
- [11] SOCIETY OF JAPAN, M., IYANAGA, S., KAWADA, Y., MAY, K., E MATHEMATICAL SOCIETY, A. *Encyclopedic Dictionary of Mathematics: Networks to Zeta functions*. No. v. 2. MIT Press, 1977.
- [12] VEACH, E., E GUIBAS, L. J. Optimally combining sampling techniques for monte carlo rendering. In *Proceedings of the 22nd Annual Conference on Com-*

puter Graphics and Interactive Techniques (New York, NY, USA, 1995), SIGGRAPH '95, Association for Computing Machinery, p. 419–428.

- [13] WALTER, B., MARSCHNER, S. R., LI, H., E TORRANCE, K. E. Microfacet models for refraction through rough surfaces. In *Proceedings of the 18th Eurographics Conference on Rendering Techniques* (Goslar, DEU, 2007), EGSR'07, Eurographics Association, p. 195–206.
- [14] WHITTED, T. An improved illumination model for shaded display. *Commun. ACM* 23, 6 (jun 1980), 343–349.
- [15] WIKIPEDIA CONTRIBUTORS. Global illumination — Wikipedia, the free encyclopedia, 2022. [Online; accessed 03-August-2022].