



INSTITUTO SUPERIOR TÉCNICO
Universidade Técnica de Lisboa

RADIOSIDADE

por

João Manuel Brisson Lopes

Departamento de Engenharia Informática

texto elaborado para a disciplina de

Computação Gráfica

Licenciatura em Engenharia Informática e de Computadores

publicado em Julho de 2000

reeditado em Janeiro de 2009, Abril 2013

Este texto, elaborado no contexto da disciplina de Computação Gráfica da Licenciatura em Engenharia Informática e de Computadores do Instituto Superior Técnico, foi originalmente concebido para fazer parte de um conjunto de textos sobre Computação Gráfica, apresentando-se agora como um texto independente.

Contacto do autor: brisson@ist.utl.pt

© 2000, 2003, 2009, 2013 J. M. Brisson Lopes & IST

Radiosidade

1 Introdução

A síntese de imagens por ray tracing permite obter imagens de muito boa qualidade mas, simultaneamente, encontra-se sujeita a algumas limitações que impedem um maior grau de realismo dessas imagens. Os métodos de ray tracing, tal como o seu nome indica, baseiam-se no seguimento de raios¹. Este procedimento está mais de acordo com a reflexão e a transmissão especulares da luz do que com a reflexão e transmissão difusas². Desta forma, os métodos de ray tracing permitem obter muito bons resultados em cenas bem iluminadas contendo superfícies reflectoras apresentando um comportamento próximo do comportamento de superfícies reflectoras especulares.

Em muitas cenas, principalmente em cenas de interiores, existem zonas que não são directamente iluminadas pelas fontes de luz. A iluminação de tais zonas é produto da luz reflectida, uma ou mais vezes, por superfícies reflectoras não especulares. Para tratar estes casos, os métodos de ray tracing empregam um termo de iluminação ambiente constante cujo cálculo nem sempre é suficientemente preciso. O emprego desta técnica faz com que as superfícies indirectamente iluminadas aparentem uma iluminação uniforme em vez de uma variação de sombreamento gradual e suave, o que conduz a diferenças abruptas de iluminação entre zonas directamente iluminadas e zonas que lhes são contíguas e não são directamente iluminadas³. Daqui resulta que a iluminação de superfícies indirectamente iluminadas que, em geral, são superfícies reflectoras difusas, é incorrectamente calculada pelos métodos de ray tracing.

O método da radiosidade tem por objectivo o cálculo da iluminação e do sombreamento em cenas em que predominam superfícies reflectoras difusas, e deriva do cálculo das trocas de radiação térmica entre superfícies empregue em Transmissão de Calor adaptado à Computação Gráfica. Esta adaptação foi realizada por Goral, Torrance, Greenberg e Bataille em 1984, com a contribuição de Nishita e Nakamae em 1985.

¹ Na direcção inversa à direcção de propagação destes.

² Em ray tracing, os fenómenos de reflexão e transmissão difusas são tratados por meio de aproximações como a aproximação de Blinn.

³ Estes fenómenos são, poderemos dizer, a assinatura dos métodos de ray tracing nas imagens por eles geradas.

Convém aqui notar que o método da radiosidade, ao incidir unicamente sobre a radiação luminosa difusa, não permite o cálculo de quaisquer componentes especulares e, portanto, as imagens que permite gerar não apresentam quaisquer zonas de brilho⁴.

Na sua origem, o cálculo da radiosidade empregue em Transmissão de Calor não é mais do que a aplicação da lei da conservação da energia (radiante neste caso) a cada uma das superfícies de um recinto ou cena, e pressupõe a existência de equilíbrio térmico. Nestas circunstâncias, a energia radiante que cada superfície recebe terá que ser igual à energia radiante que abandona a mesma superfície. Ao atingir uma superfície (veja-se a figura 1.1), a energia incidente (a irradiação da superfície) pode ser reflectida, absorvida ou transmitida⁵. A energia radiante que abandona a superfície é constituída pela energia que a superfície reflecte e pela energia que a superfície emite.

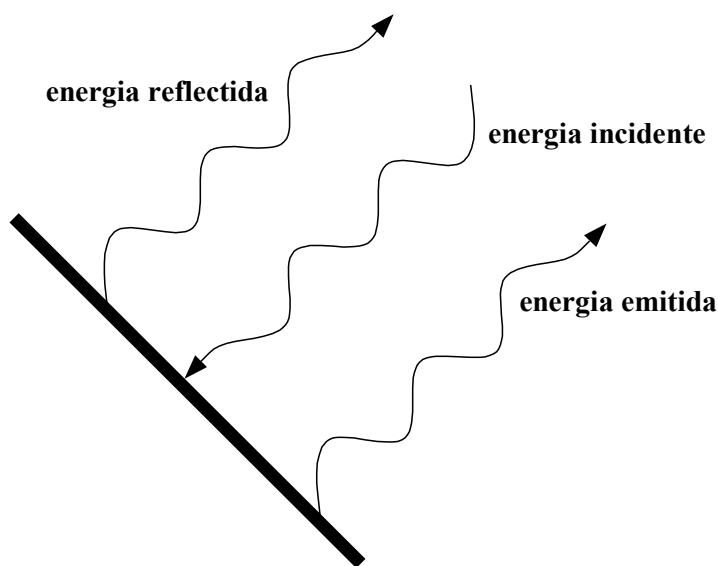


Figura 1.1 - Balanço de energia radiante numa superfície. A soma da energia reflectida com a energia emitida pela superfície é igual à energia incidente, a irradiação da superfície.

A primeira diferença entre o método da radiosidade em Transmissão de Calor e em Computação Gráfica está em que, na primeira, a radiação a ter em conta se encontra na zona do espectro electromagnético cuja radiação produz efeitos térmicos ao atingir os corpos, a zona dos infravermelhos, enquanto na segunda a radiação a tratar corresponde à radiação na zona visível do espectro, a luz. A segunda diferença consiste em que, em Computação Gráfica, o objectivo do método é determinar a luz proveniente de uma superfície, a radiosidade da superfície, enquanto em Transmissão de Calor o objectivo é determinar a temperatura das superfícies que trocam radiação térmica ou o saldo da radiação trocada entre superfícies.

Para calcular as radiosidades de uma cena, dividem-se as superfícies desta em áreas em que a radiosidade possa ser considerada uniforme e, seguidamente, estabelecem-se as equações de balanço para cada uma das áreas. A precisão dos resultados depende assim

⁴ A geração de imagens apresentando zonas de brilho pelo método da radiosidade necessita do recurso ao método de ray tracing para calcular a iluminação de tais zonas.

⁵ A radiosidade não trata, em geral, da transmissão de energia radiante.

do número de áreas consideradas, do tamanho destas áreas e da sua adequação em reproduzir os gradientes de sombreamento existentes na realidade⁶. Isto implica que o método da radiosidade deve reduzir o tamanho das áreas sempre que existam gradientes significativos e obriga a que os limites das áreas e, logicamente, a sua maior concentração, coincidam com os locais onde existem descontinuidades da iluminação local.

A radiosidade de cada uma das áreas em que as superfícies de uma cena são discretizadas depende da energia radiante que a atinge⁷ e que provém das outras áreas que se encontram dentro do seu campo de visão. Uma área pode igualmente ser emissora de radiação se for uma fonte de luz, caso em que, naturalmente, apresenta uma emissão uniforme.

Em princípio, uma área pode trocar radiação com todas as áreas da cena o que significa que a carga computacional do cálculo das radiosidades de uma cena é função do quadrado do número de áreas em que a cena se encontra dividida.

A troca de radiação entre superfícies não pressupõe necessariamente que a radiação se propague através de meios totalmente transparentes à radiação, podendo ser considerados meios que absorvem a radiação em profundidade. Daqui resultam efeitos visuais como o efeito de profundidade com o esbatimento e indefinição das superfícies dos objectos distantes. Este cálculo é bem conhecido em Transmissão de Calor pelo que a sua transposição para Computação Gráfica é relativamente simples embora o cálculo o não seja.

O método da radiosidade é um método que opera no espaço dos objectos e não depende da vista da cena que se pretenda gerar. Uma vez calculadas as radiosidades das superfícies de uma cena basta então introduzir os resultados obtidos na sequência normal do pipeline de visualização gráfica, imediatamente antes do andar que realiza a transformação de visão, para que se obtenham as imagens pretendidas. Este facto constitui uma vantagem importante do método da radiosidade sobre os métodos de ray tracing. Com efeito, uma vez calculadas as radiosidades de uma cena é imediatamente possível obter qualquer vista da cena sem que seja necessário voltar a calcular as radiosidades, desde que não existam objectos que se desloquem na cena, o que alteraria a sua geometria.

2 Equação da Radiosidade

Na base do método da radiosidade encontra-se o balanço de energia entre a radiação que atinge uma superfície e a radiação que deixa a mesma superfície. Estas duas quantidades de energia devem ser iguais numa situação de equilíbrio. De seguida apresentaremos o resultado deste balanço sob a forma da equação da radiosidade aplicada à radiação luminosa. Antes, porém, é necessário apresentar algumas definições e considerações sobre os fenómenos radiactivos que ocorrem junto das superfícies.

⁶ Este facto está na base do refinamento por subdivisão de áreas que será tratado mais à frente.

⁷ Uma área pode receber radiação proveniente de si própria se for côncava.

2.1 Irradiação Total

A energia radiante total que atinge uma dada superfície, a irradiação total da superfície, pode ser por esta absorvida, reflectida ou transmitida em proporções que dependem das propriedades físicas da superfície e do respectivo corpo. A quantificação destas proporções está ligada a três propriedades da superfície que são

- **Absorsividade** (α) – fracção da irradiação total incidente que é absorvida;
- **Reflectividade** (ρ) – fracção da irradiação total incidente que é reflectida;
- **Transmissividade** (τ) – fracção da irradiação total incidente que é transmitida através do corpo da superfície.

cuja soma é, naturalmente, um.

$$\alpha + \rho + \tau = 1 \quad (2.1)$$

Estes parâmetros são parâmetros totais, isto é, são o resultado da aplicação da sua definição à irradiação total em todos os comprimentos de onda do espectro electromagnético e, conseqüentemente, correspondem a valores médios para todo o espectro. Na realidade, as superfícies apresentam variações significativas da absorsividade, reflectividade e transmissividade com o comprimento de onda. O exemplo mais claro é o das superfícies reflectoras selectivas, que são a maior parte das superfícies reais, e em que a reflectividade é maior para determinados comprimentos de onda e menor para outros⁸.

Os corpos opacos não transmitem radiação e assim teremos $\tau=0$. Por outro lado, podemos considerar que, desde que as distâncias sejam relativamente curtas, a atmosfera⁹ não participa na troca de radiação entre superfícies e, neste caso, teremos $\tau=1$ e, conseqüentemente, $\alpha=\rho=0$, ou seja, a atmosfera comporta-se como sendo o vácuo.

2.2 Radiosidade Total e Radiosidade

A energia radiante total proveniente de uma superfície é designada como a radiosidade total dessa superfície. Quando esta é opaca, a radiosidade total resulta apenas da adição de duas contribuições, a da energia radiante total emitida pela superfície, designada por poder emissivo total da superfície, e a da energia radiante total incidente que a superfície reflecte.

Deste modo podemos então definir os seguintes termos:

- **Irradiação** (I) – energia incidente numa superfície por unidade de área da superfície;
- **Poder Emissivo** (E) – energia emitida por uma superfície por unidade de área da superfície;
- **Radiosidade** (B) – energia radiante que abandona uma superfície por unidade de área da superfície.

⁸ Uma superfície de cor verde apresenta uma grande reflectividade na gama do espectro visível correspondente aos verdes e uma reflectividade muito menor nas restantes gamas.

⁹ Ou seja, os seus gases constituintes.

Deve notar-se que as definições anteriores correspondem a fluxos (intensidades).

Face a estas definições, a radiosidade total de uma superfície com a área A_s é

$$B A_s = E A_s + \rho I A_s \quad (2.2)$$

ou, por unidade de área,

$$B = E + \rho I \quad (2.3)$$

Devemos aqui ter em atenção que, para os fins que pretendemos, estas igualdades se restringem aos balanços realizados na gama do espectro electromagnético correspondente à luz visível.

O poder emissivo, E , é nulo para as superfícies que não emitem luz e diferente de zero para as que são fontes de luz¹⁰.

As expressões (2.2) e (2.3) referem-se a toda a gama de comprimentos de onda visíveis, não sendo tido em linha de conta que as superfícies são, na sua generalidade, reflectoras selectivas. Uma superfície verde, por exemplo, reflecte apenas a radiação visível na gama de comprimentos de onda correspondentes aos verdes apresentando uma reflectividade não nula nessa gama de comprimentos de onda, e absorve todos os outros comprimentos de onda o que corresponde a apresentar uma reflectividade nula ou quase nula para estes últimos comprimentos de onda. Para ter em consideração este facto há então que expandir (2.3) nas gamas de comprimento de onda que seja necessário considerar. O processo mais comum de realizar esta expansão consiste na decomposição segundo as componentes RGB e, assim, teremos,

$$\begin{aligned} B_R &= E_R + \rho_R I_R \\ B_G &= E_G + \rho_G I_G \\ B_B &= E_B + \rho_B I_B \end{aligned} \quad (2.4)$$

2.3 Factores de Forma

Resta-nos agora examinar em mais detalhe o termo correspondente à irradiação presente em (2.2) e (2.3). Consideremos então que a superfície para a qual obtivemos aquelas expressões é irradiada por um conjunto de outras superfícies, isto é, recebe radiação proveniente dessas superfícies. Cada uma destas superfícies apresenta uma radiosidade total de que uma fracção atinge a superfície que estamos a considerar (a que atribuiremos o índice i). Representando essa fracção por F_{j-i} , a irradiação total da superfície será

$$A_i I_i = \sum_j F_{j-i} A_j B_j \quad (2.5)$$

¹⁰ Neste caso, o valor da reflectividade pode ser considerado nulo.

pelo que (2.2) passará a ser

$$A_i B_i = A_i E_i + \rho_i \sum_j F_{j-i} A_j B_j \quad (2.6)$$

A fracção F_{j-i} , denominada factor de forma, representa a fracção da radiosidade total que abandona a superfície j e atinge a superfície i . Como estamos a tratar de superfícies com reflexão e emissão difusas, a radiosidade proveniente de uma tal superfície é uniforme e, como tal, é independente da direcção de propagação da luz. Deste modo, os factores de forma assumem um significado geométrico preciso correspondente à forma como cada superfície “vê” outra superfície. Isto corresponde ainda ao ângulo sólido do campo de visão da superfície que “vê” ocupado pela superfície que é “vista”.

Os factores de forma dependem unicamente da geometria das superfícies e da sua disposição espacial relativa. A sua expressão é (veja-se a figura 2.1 para compreensão do significado das varáveis presentes na expressão)

$$F_{i-j} = \frac{1}{A_i} \iint_{A_i A_j} \frac{\cos\theta_i \cos\theta_j}{\pi r^2} dA_j dA_i \quad (2.7)$$

Deixaremos para mais tarde a análise da determinação dos factores de forma, sublinhando por agora que todo o problema da sua determinação se encontra no cálculo do integral duplo de (2.7).

2.4 Propriedades dos Factores de Forma

Os factores de forma gozam de algumas propriedades importantes, entre as quais se encontram a reciprocidade, a aditividade e o fecho.

Reciprocidade

A expressão (2.7) para o factor de forma F_{i-j} pode ser aplicada para determinar o factor de forma inverso, F_{j-i} , e obteremos

$$F_{j-i} = \frac{1}{A_j} \iint_{A_j A_i} \frac{\cos\theta_j \cos\theta_i}{\pi r^2} dA_i dA_j \quad (2.8)$$

Os integrais duplos de superfície de (2.7) e (2.8) não dependem da ordem de integração e são consequentemente idênticos. Logo podemos concluir que

$$A_i F_{i-j} = A_j F_{j-i} \quad (2.9)$$

que exprime a propriedade da reciprocidade dos factores de forma.

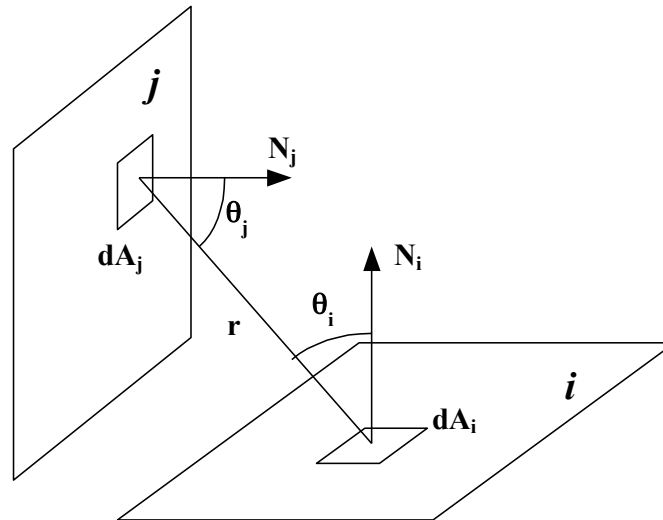


Figura 2.1 – Determinação do factor de forma entre a superfície i e a superfície j, apresentando as grandezas intervenientes.

Aditividade

Supondo que a área A_i é decomponível numa soma de n subáreas tais que

$$A_i = \sum_{k=1}^{k=n} A_{ik} \tag{2.10}$$

e, tendo em conta (2.7), teremos que para cada uma das subáreas o respectivo factor de forma é

$$A_{ik} F_{ik-j} = \iint_{A_{ik} A_j} \frac{\cos \theta_{ik} \cos \theta_j}{\pi r^2} dA_j dA_{ik} \tag{2.11}$$

Os integrais duplos são aditivos pelo que

$$\sum_{k=1}^{k=n} \iint_{A_{ik} A_j} \frac{\cos \theta_{ik} \cos \theta_j}{\pi r^2} dA_j dA_{ik} = \iint_{A_k A_j} \frac{\cos \theta_i \cos \theta_j}{\pi r^2} dA_j dA_k \tag{2.12}$$

o que significa que

$$\sum_{k=1}^{k=n} A_{ik} F_{ik-j} = A_i F_{i-j} \tag{2.13}$$

Aplicando agora a propriedade da reciprocidade dos factores de forma (2.9) aos dois membros da expressão anterior

$$\sum_{k=1}^{k=n} A_j F_{j-ik} = A_j F_{j-i} \quad (2.14)$$

e, eliminando A_j , teremos finalmente

$$F_{j-i} = \sum_{k=1}^{k=n} F_{j-ik} \quad (2.15)$$

que é a expressão da propriedade da aditividade dos factores de forma.

Fecho

A propriedade de fecho dos factores de forma é uma propriedade simples decorrente da aditividade. Se considerarmos um recinto totalmente fechado, o horizonte visual de qualquer das superfícies do recinto encontra-se totalmente preenchido pelas outras superfícies (incluindo a superfície considerada, se for côncava), pelo que

$$\sum_{k=1}^{k=n} F_{i-k} = 1 \quad (2.16)$$

2.5 Equação da Radiosidade

Estamos agora em condições de obter a forma final da equação da radiosidade. Retomando a expressão da radiosidade total de uma superfície (2.6) e introduzindo nela a expressão da propriedade da reciprocidade dos factores de forma (2.9), obteremos

$$A_i B_i = A_i E_i + \rho_i \sum_j F_{i-j} A_j B_j \quad (2.17)$$

que, eliminando A_i , se transforma em

$$B_i = E_i + \rho_i \sum_j F_{i-j} B_j \quad (2.18)$$

Esta equação tem que ser resolvida para cada uma das áreas em que a cena a calcular se encontra decomposta. Uma vez determinados os factores de forma, problema que abordaremos mais adiante, a consideração de todas as equações de balanço com a forma de (2.18) resulta num sistema de equações lineares com a seguinte forma

$$\begin{bmatrix} 1 - \rho_1 F_{1-1} & -\rho_1 F_{1-2} & \cdots & -\rho_1 F_{1-n} \\ -\rho_2 F_{2-1} & 1 - \rho_2 F_{2-2} & \cdots & -\rho_2 F_{2-n} \\ \cdots & \cdots & \cdots & \cdots \\ -\rho_n F_{n-1} & -\rho_n F_{n-2} & \cdots & 1 - \rho_n F_{n-n} \end{bmatrix} \begin{bmatrix} B_1 \\ B_2 \\ \cdots \\ B_n \end{bmatrix} = \begin{bmatrix} E_1 \\ E_2 \\ \cdots \\ E_n \end{bmatrix} \quad (2.19)$$

O sistema de equações lineares anterior deve ser decomposto ainda segundo gamas de comprimento de onda para ter em conta a natureza das superfícies presentes na cena que serão, na sua maioria, superfícies reflectoras selectivas. Tal como atrás foi afirmado, esta decomposição é realizada em três gamas, as gamas de vermelho, verde e azul (RGB). Cada uma das equações de balanço da radiosidade será então desdobrada em três equações tais que

$$\begin{aligned} B_{iR} &= E_{iR} + \rho_{iR} \sum_j F_{i-j} B_{jR} \\ B_{iG} &= E_{iG} + \rho_{iG} \sum_j F_{i-j} B_{jG} \\ B_{iB} &= E_{iB} + \rho_{iB} \sum_j F_{i-j} B_{jB} \end{aligned} \quad (2.20)$$

2.6 Solução Numérica do Sistema de Equações

O método de resolução mais apropriado para determinar a solução do sistema de equações lineares resultante da aplicação da equação de balanço da radiosidade é o método numérico de Gauss-Seidel. Este método é um método iterativo cuja convergência para a solução está assegurada desde que a matriz dos coeficientes seja uma matriz diagonalmente dominante, o que é o caso.

Se M for a matriz de coeficientes m_{ij} da expressão (2.19) tal que

$$M \times B = E \quad (2.21)$$

o algoritmo numérico de Gauss-Seidel consiste em realizar iterações sucessivas até que o critério de convergência estipulado seja satisfeito. Cada iteração calcula novos valores para todas as componentes do vector solução B , aplicando sucessivamente, para $i=1,2,\dots,n$, a expressão explícita para B_i da respectiva linha da matriz

$$B_i = E_i - \frac{1}{m_{ii}} \left[\sum_{j=1}^{j=i-1} m_{ij} B_j + \sum_{j=i+1}^{j=n} m_{ij} B_j \right] \quad (2.22)$$

Numa iteração realizada pelo algoritmo de Gauss-Seidel, o valor de B_i calculado para uma das linhas substitui o seu valor anterior logo que é calculado e é imediatamente empregue no cálculo da linha seguinte, isto é, não se espera pelo fim da iteração para substituir os antigos valores pelos novos valores.

3 Método do Hemicubo

O cálculo das trocas de radiosidade entre as superfícies de uma cena depende do cálculo dos factores de forma em jogo, tal como vimos anteriormente. De seguida abordaremos este tema, mostrando que este cálculo é bastante complexo. Daqui resulta a necessidade de encontrar processos de cálculo para os factores de forma que sejam menos complexos e permitam reduzir a carga computacional. Estes processos baseiam-se na analogia de Nusselt e têm a sua expressão em Computação Gráfica no método do hemicubo.

3.1 Cálculo de Factores de Forma

O factor de forma de uma superfície em relação a outra expressa a ocupação do horizonte visual dessa superfície por aquela outra. O seu cálculo implica o cálculo dos integrais duplos de superfície que figuram na sua expressão (2.7).

Existem expressões analíticas exactas para a determinação dos factores de forma de algumas geometrias. Estas expressões podem ser mais ou menos complexas, dependendo tal complexidade da geometria das superfícies e da sua posição relativa no espaço. Algumas destas expressões são relativamente simples, como é o caso da geometria constituída por dois círculos paralelos cujos centros se localizam sobre uma recta que é perpendicular aos círculos, tal como a figura 3.1 apresenta.

Se r_1 e r_2 forem os raios dos dois círculos e L a distância entre os seus centros, o factor de forma desta geometria apresenta a seguinte expressão analítica

$$F_{1-2} = \frac{1}{2} \left[X - \sqrt{X^2 - 4 \left(\frac{R_2}{R_1} \right)^2} \right] \quad (3.1)$$

em que

$$X = 1 + \frac{1 + R_2^2}{R_1^2} \quad R_1 = \frac{r_1}{L} \quad R_2 = \frac{r_2}{L}$$

A figura 3.2 apresenta graficamente esta geometria.

A expressão anterior é relativamente simples. No entanto, para outras geometrias simples, como são os casos de dois rectângulos paralelos de dimensões iguais e de dois rectângulos perpendiculares que partilham um lado comum (veja-se a figura 3.3), as respectivas expressões analíticas tornam-se mais complexas.

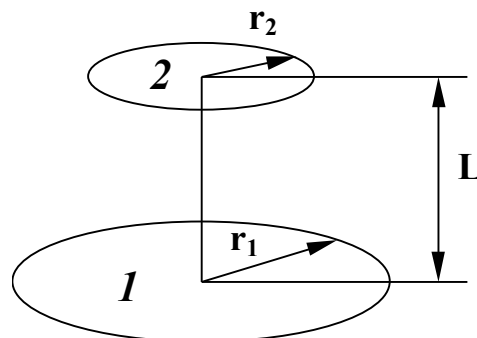


Figura 3.1 – Geometria com dois círculos paralelos cujos centros se encontram sobre a mesma linha recta.

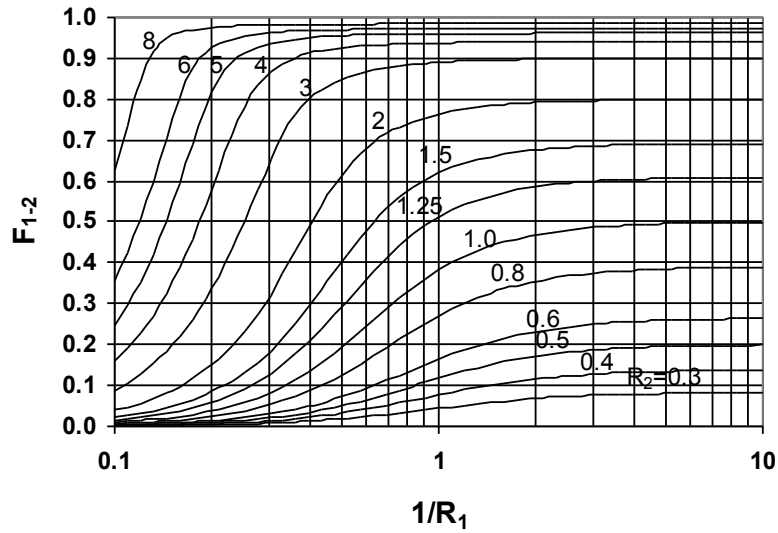


Figura 3.2 – Representação gráfica da função analítica do factor de forma entre dois círculos.

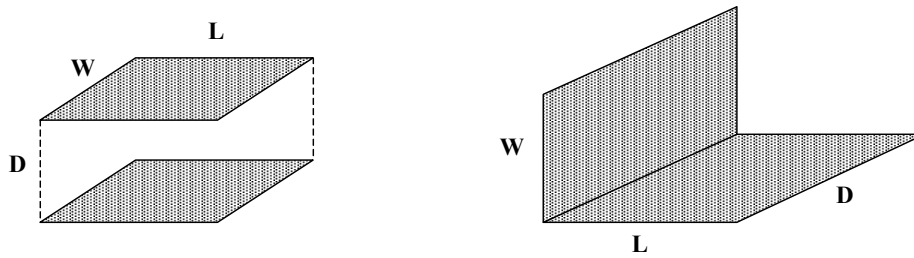


Figura 3.3 – Geometrias contendo dois rectângulos: rectângulos iguais e paralelos (à esquerda) e rectângulos perpendiculares partilhando um dos lados (à direita).

No caso dos dois rectângulos iguais e paralelos (veja-se a figura 3.3, à esquerda) a expressão analítica do factor de forma é

$$\begin{aligned}
 F_{1-2} = \frac{1}{\pi} & \left[\frac{1}{R_1 R_2} \ln \frac{(1+R_1^2)(1+R_2^2)}{1+R_1^2+R_2^2} - \frac{2}{R_1} \operatorname{arctg} R_2 - \frac{2}{R_2} \operatorname{arctg} R_1 \right. \\
 & \left. + 2 \sqrt{1+\frac{1}{R_1^2}} \operatorname{arctg} \frac{R_2}{\sqrt{1+R_1^2}} + 2 \sqrt{1+\frac{1}{R_2^2}} \operatorname{arctg} \frac{R_1}{\sqrt{1+R_2^2}} \right] \quad (3.2)
 \end{aligned}$$

em que

$$R_1 = \frac{L}{D} \quad R_2 = \frac{W}{D}$$

Para a geometria constituída por dois rectângulos perpendiculares com um lado comum que a figura 3.3 (à direita) apresenta, a expressão analítica do respectivo factor de forma é

$$F_{1-2} = \frac{1}{\pi R_1} \left\{ R_1 \operatorname{arctg} \frac{1}{R_1} + R_2 \operatorname{arctg} \frac{1}{R_2} - \sqrt{R_1^2 + R_2^2} \operatorname{arctg} \frac{1}{\sqrt{R_1^2 + R_2^2}} \right. \\ \left. + \frac{1}{4} \ln \left[\frac{(1 + R_1^2)(1 + R_2^2)}{(1 + R_1^2 + R_2^2)} \times \left(\frac{R_2^2 (1 + R_1^2 + R_2^2)}{(1 + R_2^2)(R_1^2 + R_2^2)} \right)^{R_2^2} \right. \right. \\ \left. \left. \times \left(\frac{R_1^2 (1 + R_1^2 + R_2^2)}{(1 + R_1^2)(R_1^2 + R_2^2)} \right)^{R_1^2} \right] \right\} \quad (3.3)$$

em que

$$R_1 = \frac{L}{D} \quad R_2 = \frac{W}{D}$$

Como se vê, as expressões dos factores de forma (3.2) e (3.3) são já complexas e computacionalmente algo pesadas¹¹ para geometrias ainda relativamente simples e reflectem um dos maiores problemas que afectam o método da radiosidade: o cálculo do factor de forma entre duas superfícies arbitrárias. Com efeito, mesmo para geometrias simples, pode acontecer que não sejam conhecidas ou não existam mesmo expressões analíticas para o cálculo dos respectivos factores de forma, sendo necessário recorrer a métodos de integração numérica para a sua determinação¹².

Como as geometrias que se encontram nas cenas não se limitam a casos simples, há então que recorrer a outras formas de cálculo dos factores de forma. O método mais empregue, o método do hemicubo, tem por base a analogia de Nusselt que a seguir abordamos.

3.2 Analogia de Nusselt

A analogia de Nusselt pode ser enunciada da seguinte forma:

Se duas superfícies diferentes ocupam a mesma parte do horizonte visual de uma terceira superfície, os factores de forma desta última para cada uma daquelas são iguais.

A figura 3.4 apresenta graficamente o significado desta analogia. Quando se pretende determinar o factor de forma entre a superfície elementar A e a superfície B da figura, basta ter em conta que o arco de visão da superfície A ocupado pela superfície B corresponde simplesmente ao arco subtendido pelas linhas que unem os extremos de B com a superfície elementar A. Para determinar o factor de forma bastará então calcular a fracção desse arco em relação à semicircunferência.

¹¹ Atente-se no cálculo da função trigonométrica arctg.

¹² Para alguns casos existem expressões que permitem o cálculo numérico dos valores dos factores de forma por meio de técnicas de integração numérica como a quadratura de Gauss.

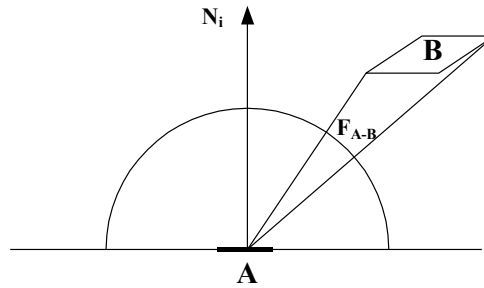


Figura 3.4 – Analogia de Nusselt. O factor de forma F_{A-B} pode ser calculado determinando a fracção da semicircunferência correspondente ao arco subtense.

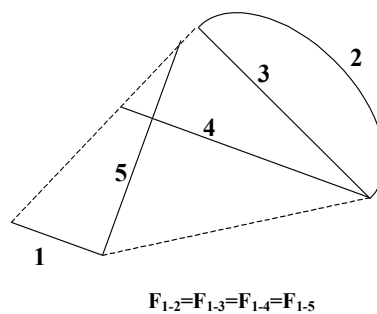


Figura 3.5 – Aplicação da analogia de Nusselt. Os factores de forma da superfície 1 para qualquer das outras superfícies são iguais, o que permite determinar o factor de forma de 1 para 2 através do factor de forma de 1 para 4 ou 1 para 5 cujo cálculo é mais simples (as superfícies são representadas de perfil na figura).

A analogia de Nusselt permite desta forma substituir um cálculo complexo de um factor de forma por outro mais simples, preferencialmente por um que possua expressão analítica.

Tomemos agora como exemplo o cálculo do factor de forma F_{1-2} para a geometria representada na figura 3.5. Em lugar de tentarmos calcular directamente o seu valor, poderemos considerar a superfície 3 que ocupa a mesma fracção do horizonte visual da superfície 1 que a superfície 2 ocupa. Se os planos das superfícies 1 e 3 fossem paralelos, o que não é o caso, poderíamos calcular o factor de forma F_{1-3} através do emprego da expressão (3.2) associada a dois rectângulos de iguais dimensões subdivididos de tal forma que as superfícies 1 e 3 correspondessem a duas das subdivisões e assim permitissem a aplicação da propriedade da aditividade dos factores de forma (2.15). No entanto, como os planos que contêm as superfícies 1 e 3 não são paralelos, não o podemos fazer. Porém, se considerarmos agora a superfície 4 que, esta sim, é paralela à superfície 1, já poderemos aplicar esta estratégia no cálculo do factor de forma F_{1-2} que é igual ao factor de forma F_{1-4} . Uma alternativa consistiria em considerar ainda a superfície 5 perpendicular à superfície 1, que é uma geometria para a qual existe uma expressão analítica (3.3).

3.3 O Método do Hemicubo

A analogia de Nusselt permite substituir o cálculo de factores de forma de geometrias complexas pelo cálculo de factores de forma de geometrias mais simples, recorrendo a geometrias para as quais existam expressões analíticas e cuja carga computacional não seja demasiado elevada. A primeira hipótese a considerar seria a de envolver as superfícies por semiesferas. No entanto, o cálculo resultante desta estratégia não é tão simples como à partida aparenta e há então que considerar outras alternativas. Destas a mais empregue e mais popular em radiosidade é a do método do hemicubo.

Consideremos então duas superfícies para as quais se pretende determinar os factores de forma, tal como a figura 3.6 apresenta, e defina-se um cubo centrado numa dessas superfícies, envolvendo-a completamente, e em que um dos eixos coincide com a direcção da normal à superfície. Para efeitos da determinação dos factores de forma, consideraremos apenas a metade superior do cubo e descartaremos a outra metade¹³.

A face superior do hemicubo é paralela à superfície e as outras faces são-lhe perpendiculares. O cálculo do factor de forma da superfície envolvida para uma parte das faces do hemicubo reduzir-se-á então à aplicação de expressões já conhecidas, empregando exhaustivamente a propriedade da aditividade. Para calcular o factor de forma da superfície envolvida para uma outra superfície basta então determinar que áreas das faces do hemicubo se interpõem entre as duas superfícies e, novamente empregando a propriedade da aditividade dos factores de forma, adicionar os factores de forma correspondentes a essas áreas em que as faces do hemicubo se encontram subdivididas, conforme a figura 3.7 apresenta.

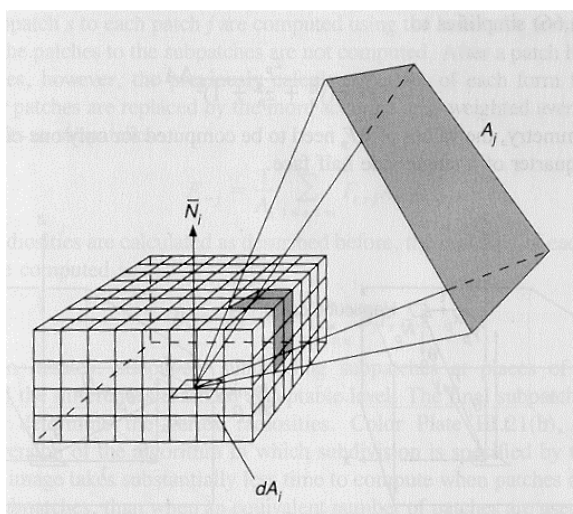


Figura 3.6 – Hemicubo apresentando o seu posicionamento face à superfície para a qual se pretendem determinar os factores de forma e a discretização das suas faces. O hemicubo inferior foi removido.

¹³ A consideração da metade superior do cubo está na origem da designação de hemicubo para este método.

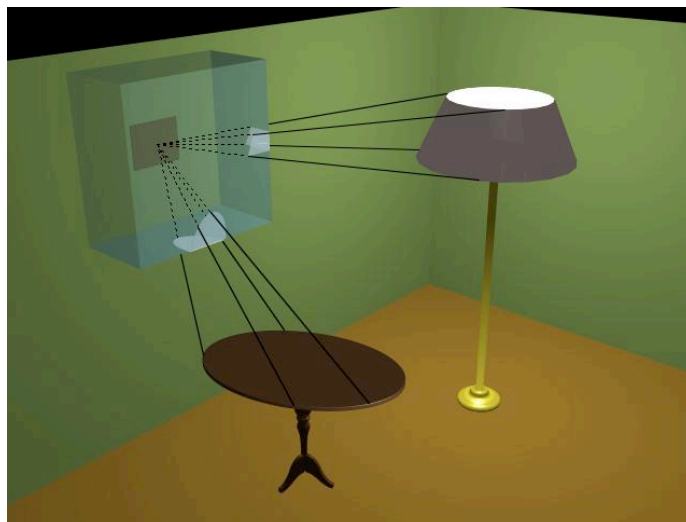


Figura 3.7 – Aplicação do método do hemicubo. Os factores de forma da superfície localizada na parede, para a mesa e para o candeeiro, são calculados determinando que subáreas do hemicubo se interpõem no campo de visão e somando os respectivos factores de forma.

4 Optimização

Os algoritmos anteriormente descritos para o cálculo das radiosidades de uma cena resolvem alguns dos principais problemas encontrados, mas deixam ainda por resolver dois tipos de problemas:

- Grau de discretização da cena;
- Velocidade de convergência do método de resolução do sistema de equações lineares a resolver.

O grau de discretização da cena influencia directamente a ordem do sistema de equações lineares a resolver para determinar todas as radiosidades da cena. Uma cena discretizada por uma malha grossa conduz a um sistema de equações lineares de ordem menor e, conseqüentemente, menos pesado em termos de carga computacional do que o sistema de equações lineares resultante do emprego de uma malha mais fina. Porém, os resultados serão muito menos precisos, principalmente nas zonas da cena onde existam grandes gradientes de radiosidade.

Por outro lado, a velocidade de convergência do algoritmo numérico de Gauss-Seidel empregue na resolução do sistema de equações lineares (2.19) é tanto menor quanto maior for a ordem do sistema de equações a resolver e, portanto, é necessário encontrar modificações ao algoritmo que permitam apressar a convergência.

Nesta secção apresentaremos algumas respostas para estes dois problemas. Em primeiro lugar analisaremos o grau de discretização da cena e introduziremos o refinamento adaptativo da respectiva malha e, seguidamente, abordaremos a questão de como apressar a convergência por meio das técnicas de radiosidade progressiva.

4.1 Subdivisão Adaptativa

A carga computacional do método da radiosidade é proporcional ao quadrado do número de áreas em que cada cena se encontra dividida, pelo que há que manter este número tão pequeno quanto possível. Por outro lado, porque a radiosidade em cada uma destas áreas é constante, a sua dimensão deverá ser tão pequena quanto o necessário para que se possa reproduzir adequadamente os sombreados naturais das cenas, o que acarreta que o número de áreas seja elevado. Estas condições contraditórias só são satisfeitas se nas cenas só existirem gradientes de radiosidade pequenos. Mas, em geral, existem gradientes de radiosidade elevados, principalmente nas transições de zonas de luz para zonas de sombra que implicam que as áreas nessas zonas sejam de menor dimensão do que em outros locais da cena. Uma discretização demasiado fina indiscriminada resolve este problema, mas implica uma carga computacional demasiado elevada. Este problema é tanto mais grave quanto, à partida, se desconhece em que zonas se irão localizar as áreas de maiores gradientes de radiosidade, o que não permite que, com absoluta certeza, se determinem em que zonas a discretização deverá ser mais fina e qual o grau de discretização que aí deverá ser empregue. Os dados para esta decisão só se tornam conhecidos à medida que o processo de solução iterativa progride e as radiosidades de cada área são conhecidas com precisão suficiente.

A resposta para este problema é a subdivisão (refinamento) adaptativa que consiste em, no decorrer da solução, identificar as áreas de maiores gradientes de radiosidade e subdividi-las quando se verifica que a divisão inicial foi demasiado grosseira face aos gradientes calculados.

A figura 4.1 ilustra este refinamento progressivo realizado em função dos resultados calculados. Numa primeira fase, detecta-se a existência de uma zona de separação entre uma zona de luz e uma zona de sombra que atravessa algumas das áreas, através da constatação da existência de fortes gradientes de radiosidade entre essas áreas e as áreas vizinhas. O processo de subdivisão adaptativa consiste então em subdividir as áreas em causa em áreas mais pequenas, recalcular as radiosidades para cada uma das novas áreas e em analisar novamente os gradientes de radiosidade. As subáreas para as quais ainda existam gradientes de radiosidade significativos poderão por sua vez ser novamente subdivididas num processo recorrente. É claro que este processo recorrente deve estar sujeito a limites para o grau de subdivisão para que a carga computacional total se mantenha dentro de limites razoáveis.

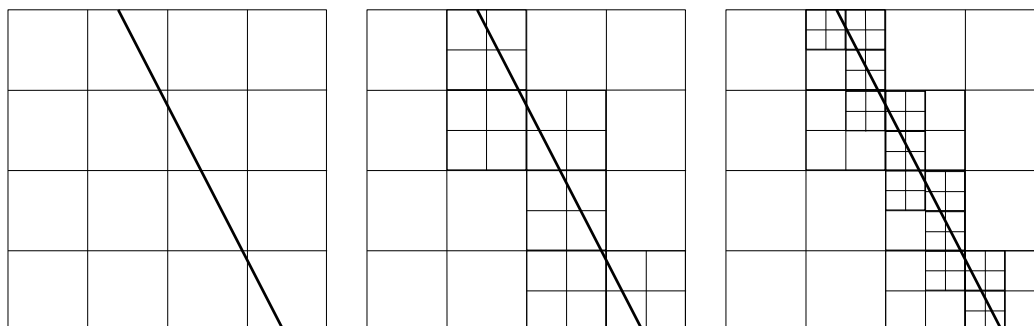


Figura 4.1 – Subdivisão adaptativa sucessiva nas áreas em que existem gradientes elevados de radiosidade (a linha inclinada assinala a fronteira entre uma zona iluminada e uma zona de sombra).

4.2 Radiosidade Progressiva

O algoritmo de Gauss-Seidel empregue na resolução numérica do sistema de equações lineares (2.19) é estável e convergente, uma vez que a matriz dos coeficientes deste sistema é diagonalmente dominante. Mas a sua velocidade de convergência diminui com o aumento da ordem dos sistemas a resolver, o que implica a realização de mais iterações sempre que esta ordem aumente. São portanto necessárias variantes algorítmicas que permitam aumentar a velocidade de convergência do algoritmo original. Começaremos por analisar o algoritmo de Gauss-Seidel, e o significado físico que assume, quando aplicado ao sistema de equações lineares constituído pelas equações de balanço da radiosidade.

Colecção (Gathering)

O algoritmo de Gauss-Seidel calcula e altera sucessivamente o valor da radiosidade, B_i , associada a cada área A_i . A ordem seguida é a ordem das equações no sistema, isto é, a ordem das linhas na matriz do sistema.

A expressão da radiosidade de uma área em função das radiosidades das outras áreas da cena é

$$B_i = E_i + \rho_i \sum_j F_{i-j} B_j \quad (4.1)$$

expressão esta que possui um significado preciso. O novo valor da radiosidade para a área A_i é função das radiosidades das outras áreas da cena, ou seja, a área i colecta as radiosidades das outras áreas emitidas na sua direcção. A contribuição da área j para a radiosidade da área i é

$$B_i(B_j) = \rho_i F_{i-j} B_j \quad (4.2)$$

Deste processo de colecção (gathering) das radiosidades resulta a alteração sucessiva das áreas na imagem a ser composta, exactamente pela ordem com que as respectivas equações de balanço figuram no sistema de equações lineares.

Dispersão (Shooting)

O processo de colecção altera a radiosidade de cada área, uma área de cada vez, e só ao fim de uma iteração completa é que a imagem reflecte coerentemente todas as alterações. Isto corresponde a resolver o sistema de equações lineares proposto por linhas.

A alternativa consiste em adoptar um processo de resolução por colunas em vez de linhas. Com efeito, se tomarmos a área A_i , a sua radiosidade contribui para as radiosidades das outras áreas com o termo

$$B_j(B_i) = \rho_j F_{j-i} B_i \quad (4.3)$$

Fisicamente, isto corresponde a dispersar a radiosidade da área A_i pelas outras áreas, determinando cada uma das contribuições da radiosidade da área A_i para a radiosidade das outras áreas. Esta estratégia é então uma estratégia de dispersão (shooting).

No cálculo de cada uma das contribuições da área i para as outras áreas figura o factor de forma F_{j-i} em lugar do factor de forma F_{i-j} , calculado através do hemicubo local da área A_i . É necessário então transformar (4.3) tendo em conta a propriedade da reciprocidade dos factores de forma (2.9), pelo que teremos

$$B_j(B_i) = \rho_j F_{i-j} \frac{A_i}{A_j} B_i \quad (4.4)$$

Esta modificação do algoritmo permite observar a alteração progressiva e simultânea da radiosidade de todas as áreas da imagem que esteja a ser gerada, em lugar de uma área de cada vez. Além disso, testes realizados permitiram determinar que esta estratégia apressa a convergência da solução do sistema de equações lineares, embora o faça de uma forma que ainda não é totalmente satisfatória.

Solução Ordenada

As estratégias de colecção (gathering) e de dispersão (shooting) seguem a ordem pela qual as linhas e colunas, respectivamente, se encontram na matriz do sistema de equações lineares a resolver, sem atender a que existem áreas cuja radiosidade contribui mais para a iluminação da cena do que as outras. No caso da estratégia de dispersão, os valores calculados para as radiosidades das áreas recipientes encontrar-se-ão mais próximos dos respectivos valores finais se primeiro forem dispersas as radiosidades das áreas apresentando maior radiosidade total, as áreas para as quais o produto $B_i A_i$ é maior. Como se trata de um processo iterativo, isto equivale a dispersar primeiro as radiosidades das áreas cuja variação de radiosidade na iteração anterior foi maior. A convergência do processo iterativo é assim apressada, pois as primeiras iterações vão produzir aproximações mais próximas da solução final.

O algoritmo assim modificado é constituído pelos seguintes passos:

1. Inicializar $B_i = \Delta B_i = 0$ para todas as áreas que não sejam fontes de luz e $B_i = \Delta B_i = E_i$ para as áreas que são fontes de luz.
2. Repetir os passos 3 a 6 até atingir a convergência.
3. Seleccionar a área i com a maior variação de radiosidade total não distribuída, isto é, a área em que $\Delta B_i A_i$ é máximo.
4. Calcular os factores de forma F_{i-j} ¹⁴.
5. Para cada área A_j , calcular o incremento de radiosidade proveniente da área A_i que é recebida pela área A_j

$$\Delta rad = \rho_j F_{i-j} \frac{A_i}{A_j} \Delta B_i$$
 e adicioná-lo à radiosidade da área A_j que ainda não foi dispersa

$$\Delta B_j = \Delta B_j + \Delta rad$$
 e à sua radiosidade

$$B_j = B_j + \Delta rad$$
6. Assinalar que ΔB_i já foi dispersa, fazendo $\Delta B_i = 0$.

¹⁴ Se a malha não for alterada de iteração para iteração, estes factores podem ser mantidos sem necessidade de os recalcularem.

Termo Ambiente

Se, ao empregar o algoritmo de solução ordenada, as imagens correspondentes a cada passo forem sendo apresentadas num dispositivo de saída gráfica, verificaremos que estas serão inicialmente escuras e que, após cada passo, todas as áreas irão ficando sucessivamente mais claras até que a cena fique completamente iluminada. As maiores variações de luminosidade verificar-se-ão durante os primeiros passos em que as radiosidades provenientes das fontes de luz são dispersas pela cena. No decurso do processo, as áreas não expostas à iluminação directa das fontes de luz irão sucessivamente clareando, à medida que as áreas que reflectem luz de forma mais directa se tornam elas próprias mais iluminadas.

A estratégia do termo ambiente não apressa a convergência da solução ordenada, mas introduz o chamado termo ambiente que permite que as imagens das primeiras iterações aparentem encontrar-se muito próximas da solução final. Esta estratégia consiste em, no fim de cada passo da solução ordenada e em lugar de apresentar uma imagem correspondente à solução corrente, apresentar uma imagem em que à solução corrente é adicionado um termo correctivo derivado da estimativa da luz que ainda falta dispersar pela cena. A estratégia do termo ambiente actua unicamente ao nível da imagem a ser apresentada e não altera a solução obtida em cada passo e, portanto, não apressa a convergência do processo iterativo.

O termo ambiente é calculado em função das radiosidades ainda por dispersar em cada área da cena, redistribuindo-as uniformemente pela cena. Considerando globalmente a cena, podemos calcular uma reflectividade média para toda a sua superfície

$$\rho_{med} = \frac{\sum_i \rho_i A_i}{\sum_i A_i} \quad (4.5)$$

Calcula-se então o factor de reflexão total, R , a partir de (4.5) para ter em conta as reflexões sucessivas. A expressão do valor de R é

$$R = 1 + \rho_{med} + \rho_{med}^2 + \rho_{med}^3 + \dots = \frac{1}{1 - \rho_{med}} \quad (4.6)$$

Ponderando agora a radiosidade por dispersar de todas as áreas pelo quociente da área de cada uma delas pela área total¹⁵, obteremos a seguinte estimativa para o termo ambiente

$$B_{amb} = R \frac{\sum_i \Delta B_i A_i}{\sum_i A_i} \quad (4.7)$$

e, ao gerar a imagem, corrige-se a radiosidade de cada área adicionando-lhe o termo ambiente de modo a apresentar a radiosidade corrigida

$$B'_i = B_i + \rho_i B_{amb} \quad (4.8)$$

¹⁵ Este quociente corresponde a uma estimativa grosseira dos factores de forma que, no entanto, é suficiente para o fim em vista.

Encontramo-nos agora em condições para apresentar o algoritmo de solução ordenada com apresentação de imagens corrigidas por meio do termo ambiente. Os passos deste algoritmo são:

1. Inicializar $B_i = \Delta B_i = 0$ para todas as áreas que não sejam fontes de luz e $B_i = \Delta B_i = E_i$ para as áreas que são fontes de luz.

2. Calcular a área total da cena, o factor de reflexão total e a radiosidade ambiente por dispersar

$$A_{total} = \sum_i A_i$$

$$R = \frac{A_{total}}{A_{total} - \sum_i \rho_i A_i}$$

$$B_{amb} = R \frac{\sum_i \Delta B_i A_i}{A_{total}}$$

3. Repetir os passos 4 a 9 até atingir a convergência.
4. Seleccionar a área A_i , com a maior variação de radiosidade total não distribuída, isto é, a área em que $\Delta B_i A_i$ é máximo e fazer

$$\Delta E = \Delta B_i A_i$$

5. Calcular os factores de forma F_{i-j} .
6. Para cada área A_j , calcular o incremento de radiosidade proveniente da área A_i que é recebida pela área A_j

$$\Delta rad = \rho_j F_{i-j} \frac{A_i}{A_j} \Delta B_i$$

adicioná-lo à radiosidade da área A_j que ainda não foi dispersa

$$\Delta B_j = \Delta B_j + \Delta rad$$

e à sua radiosidade

$$B_j = B_j + \Delta rad$$

subtraindo a radiosidade total recebida pela área j da radiosidade total da área i ainda por dispersar

$$\Delta E = \Delta E - \Delta rad A_j$$

7. Assinalar que ΔB_i já foi dispersa, fazendo $\Delta B_i = 0$.
8. Corrigir o termo ambiente, retirando-lhe a radiosidade total distribuída pelo passo corrente

$$B_{amb} = B_{amb} - R \frac{\Delta E}{A_{total}}$$

9. Apresentar a imagem com as radiosidades corrigidas pela adição do termo ambiente

$$B'_i = B_i + \rho_i B_{amb}$$

Como se depreende do algoritmo, o termo ambiente irá diminuindo à medida que o algoritmo convergir para a solução. O valor do termo ambiente pode mesmo ser empregue como critério de paragem do processo iterativo, pois representa

indirectamente a radiosidade total ainda não dispersa pela cena. Um critério de paragem mais correcto, e que possui significado físico preciso, tem por base a radiosidade total não dispersa ao fim de cada passo que é

$$rad_{dist} = \sum_i \Delta B_i A_i \quad (4.9)$$

Como o valor de (4.9) é inicialmente igual à radiosidade total a distribuir, o seu valor ao fim de cada passo pode então ser comparado com esse valor inicial, e o processo iterativo pode ser dado como terminado quando o seu valor corrente for inferior a uma fracção pré-determinada do seu valor inicial.

Comparação entre as Várias Estratégias

Vimos já que a estratégia de dispersão (shooting) é mais eficiente que a estratégia de colecção (gathering). O emprego da solução ordenada aliada à estratégia de dispersão permite, por sua vez, o aumento da velocidade de convergência e a obtenção de imagens mais próximas da imagem final, com um menor número de passos efectuados. Por seu lado, a adição do termo ambiente, embora não acelere o processo de convergência, permite que, logo nos passos iniciais, as imagens aproximadas sejam ainda mais próximas dos resultados finais.

Estes resultados foram obtidos por um estudo com vários casos de teste realizado por Chen, Cohen, Wallace e Greenberg, que a figura 4.2 documenta, comparando imagens resultantes das várias estratégias ao fim do cálculo de 100 passos (áreas).

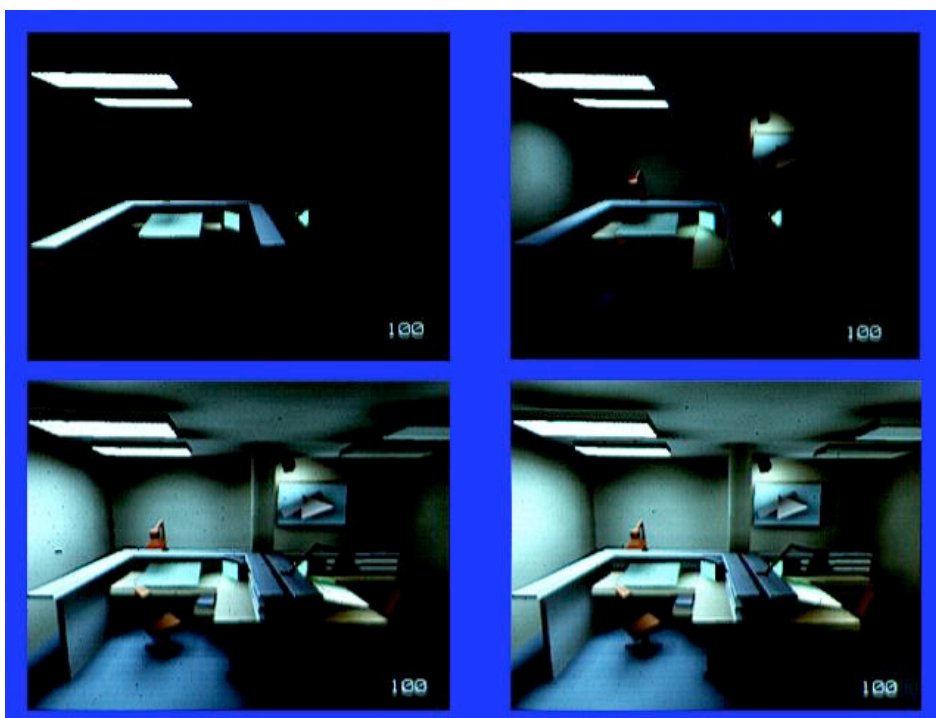


Figura 4.2 – Resultados produzidos pelas estratégias de colecção (em cima à esquerda), dispersão (em cima à direita), solução ordenada (em baixo à esquerda) e solução ordenada com termo ambiente (em baixo à direita) ao fim do cálculo de 100 áreas.

A figura mostra que tanto a estratégia de colecção como a estratégia de dispersão produziram imagens ainda um pouco escuras, mas que esta última se apresenta um pouco mais clara. Muito mais evidente é a diferença entre as imagens produzidas entre estas duas estratégias e a imagem correspondente à solução ordenada que se apresenta já como próxima da imagem final. Comparando esta imagem com a equivalente obtida pela adição do termo ambiente, verifica-se que existem apenas algumas diferenças pelo que se pode concluir que, embora as diferenças sejam ainda detectáveis, a importância do termo ambiente é já bastante reduzida.

A figura 4.3 apresenta uma comparação qualitativa da qualidade das imagens obtidas pelas várias estratégias ao fim de cada passo, comparando-as com a solução final. A estratégia de distribuição converge muito lentamente, enquanto a estratégia de dispersão, embora convirja de forma semelhante, apresenta uma convergência entrecortada por aumentos súbitos da qualidade da imagem devido ao processamento de áreas correspondentes às fontes de luz. Estes aumentos súbitos concentram-se nos passos iniciais da estratégia por solução ordenada e, conseqüentemente, esta estratégia produz resultados mais correctos com um menor número de passos efectuados. A adição do termo ambiente tem por seu lado um efeito de melhoria na qualidade logo a partir do primeiro passo e, naturalmente, acaba por se confundir com a solução ordenada sem termo ambiente logo que grande parte da radiosidade total proveniente das fontes de luz se encontra já razoavelmente distribuída.

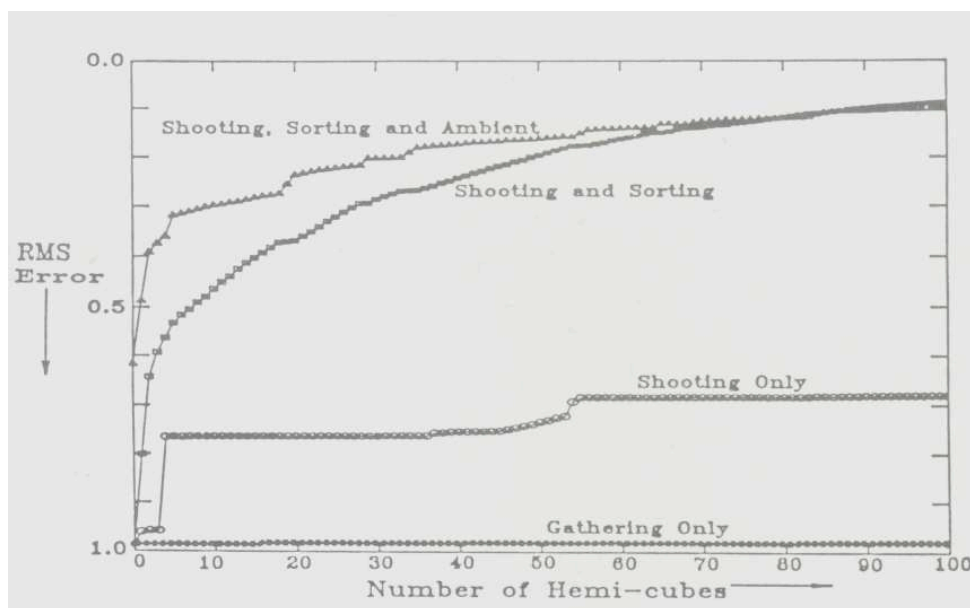
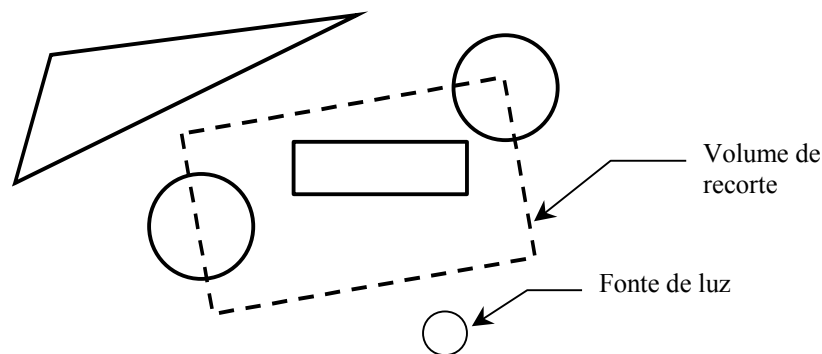


Figura 4.3 – Comparação do erro cometido pelas estratégias de colecção (gathering), dispersão (shooting), solução ordenada (shooting and sorting) e com emprego do termo ambiente (shooting, sorting and ambient).

Exercícios

- 1 Apresente o fundamento do método da radiosidade. Comente o que se verifica quando as superfícies são emissoras.
- 2 Explique a razão pela qual os algoritmos de radiosidade separam a luz nas suas componentes vermelha, verde e azul.
- 3 A emissão e a reflexão da luz podem ser difusas e especulares. No método da radiosidade é possível ter em conta estas duas componentes?
- 4 Apresente duas vantagens e duas desvantagens significativas da geração de imagens pelo método da radiosidade sobre o método de ray tracing.
- 5 Explique o que são os factores de forma e o papel que desempenham no método da radiosidade.
- 6 O que é a analogia de Nusselt? Como é que esta analogia é aplicada ao método do hemicubo nos algoritmos da radiosidade?
- 7 Os cálculos de uma cena por meio do método da radiosidade são ou não independentes da transformação de visualização? Explique porquê.
- 8 A figura seguinte apresenta, em planta, uma cena cuja imagem irá ser calculada pelo método da radiosidade. Na planta encontra-se representado a tracejado o volume de recorte a empregar. Será possível calcular as radiosidades recortando a cena antes de realizar os cálculos das radiosidades? Explique porquê.



- 9 Explique o objectivo do algoritmo da subdivisão adaptativa no contexto dos cálculos da radiosidade.
- 10 A radiosidade progressiva, intimamente ligada à solução do sistema de equações lineares representativo dos balanços de radiosidade numa dada cena, apresenta várias estratégias. Que estratégias são essas e que relação apresentam com a forma de resolver os sistemas de equações?
- 11 Compare o desempenho das várias estratégias denominadas globalmente por radiosidade progressiva. Explique a razão do emprego do chamado termo ambiente.