



INSTITUTO SUPERIOR TÉCNICO  
Universidade Técnica de Lisboa

# RAY TRACING

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 Junho 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](mailto:brisson@ist.utl.pt)

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

# Ray Tracing

## 1 Introdução

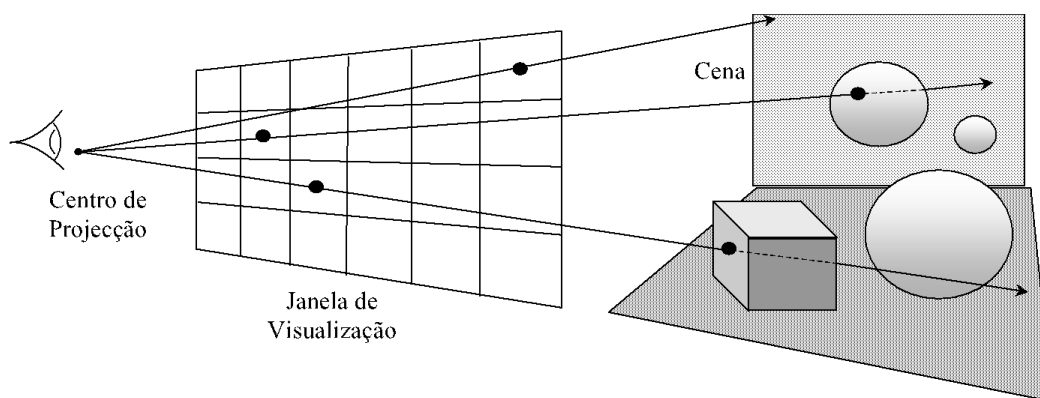
A designação de ray tracing aplica-se a um conjunto de algoritmos que foram desenvolvidos a partir do trabalho inicial de Appel publicado em 1968, e não apenas a um algoritmo particular. Estes algoritmos são dos mais empregues na geração de imagens sintéticas.

Começamos por apresentar o algoritmo fundamental de ray tracing, abordando de seguida o aspecto determinante da carga computacional que lhe está associada, o cálculo de intersecções entre os raios e os objectos a representar. A interacção entre a luz e as superfícies, que determina a cor local dos objectos, é analisada de seguida, sendo apresentada a extensão ao modelo de iluminação de Phong necessária para que este possa tratar também a transmissão de luz através de corpos transparentes e translúcidos. Da junção do algoritmo fundamental com o cálculo de intersecções e o modelo de iluminação estendido resulta o algoritmo completo cuja característica mais relevante é a recursividade. Apresentam-se então os fundamentos das estratégias de optimização destinadas a diminuir a carga computacional associada ao cálculo de intersecções. O final é dedicado à apresentação de técnicas para resolver problemas de perda de detalhe de objectos de pequena dimensão (*antialiasing*).

## 2 O Algoritmo Fundamental de Ray Tracing

Na base de todos os algoritmos de ray tracing está o trabalho de Appel desenvolvido em 1968 com o intuito de determinar sombras, nomeadamente se um ponto visível de uma cena estaria ou não no interior de uma zona de sombra. Entre 1968 e 1971, Goldstein e Nagel desenvolveram algoritmos para o cálculo de trajectórias balísticas e de partículas nucleares que, só depois, conjuntamente com os algoritmos desenvolvidos por Appel, foram aplicados à geração de imagens. Em 1979, Kay introduziu as modificações necessárias para o cálculo mais correcto da iluminação especular e da refacção da luz, tendo este último aspecto sido completado por Whitted em 1980.

Na sua origem, os algoritmos de ray tracing são algoritmos para determinar a visibilidade dos objectos de uma cena, operando ao nível de precisão da imagem. O algoritmo fundamental de ray tracing considera raios com origem no centro de projecção em que cada um dos raios passa pelo ponto correspondente a um pixel da imagem sobre o plano de projecção da cena a representar, como a figura 2.1 apresenta. Se um raio não intersectar nenhum objecto da cena, é atribuída ao pixel da imagem por onde o raio passa a cor de fundo da cena. Caso contrário, determinam-se as intersecções do raio com os objectos da cena e ao pixel correspondente ao raio é atribuída a cor do ponto de intersecção mais próximo da origem do raio.



**Figura 2.1 – Algoritmo fundamental de ray tracing. Os raios com origem no centro de projecção passam pelos píxeis respectivos na janela de visualização. A cor de cada pixel é a cor do objecto intersectado pelo seu raio no ponto de intersecção (ou a cor de fundo, se não existir qualquer intersecção).**

Este algoritmo realiza implicitamente a remoção de superfícies ocultas, pois só o ponto de intersecção mais próximo da origem do raio é visível a partir do centro de projecção<sup>1</sup>. A determinação da cor de cada pixel da imagem depende do cálculo da iluminação no ponto da cena que o raio atinge. Esta depende da iluminação directa deste ponto pelas fontes de luz da cena, da iluminação global (ambiente) da cena e da iluminação devida à luz reflectida pelos outros objectos da cena que atinge esse ponto. Isto permite o cálculo implícito de sombras pois estas são zonas de iluminação fraca. O emprego de modelos de iluminação mais completos do que o modelo de Phong permite ainda ter em conta a iluminação dos objectos devida à refacção da luz, tal como mais adiante se verá<sup>2</sup>. Esta contribuição adicional para a iluminação tem em conta a luz transmitida através do próprio objecto (quando este é transparente ou translúcido), ou a luz transmitida através de outros objectos, e que vai iluminar o ponto de intersecção do objecto com o raio. O resultado é a obtenção de imagens com maior realismo.

<sup>1</sup> Neste aspecto particular, o algoritmo fundamental de ray tracing encontra-se mais perto do algoritmo da linha de varrimento para remoção de superfícies ocultas do que do algoritmo de Z buffer.

<sup>2</sup> Poderemos ainda aplicar modelos de iluminação mais complexos para ter em conta outros efeitos ópticos.

Na sua forma mais simples, o algoritmo fundamental de ray tracing pode ser descrito da seguinte forma:

```

Seleccionar o centro de projecção e a janela de recorte sobre o plano de projecção
Para cada linha horizontal (de varrimento) da imagem
{
  Para cada pixel da linha de varrimento
  {
    Determinar o raio que une o centro de projecção com o pixel
    Para cada objecto da cena
      {
        Se o raio intersecta o objecto e o ponto de intersecção encontra-se
          mais próximo do centro de projecção do que o ponto de
          intersecção até agora encontrado
          Registrar o ponto de intersecção e o objecto intersectado
      }
    Atribuir ao pixel a cor do objecto intersectado no ponto de intersecção
    registado
  }
}

```

Note-se que todos estes cálculos são realizados no espaço dos objectos e não no espaço da imagem.

Uma vez determinada a intersecção de um raio com o objecto mais próximo, têm então lugar os cálculos de iluminação no ponto de intersecção para atribuir ao pixel da imagem a cor desse ponto. Estes cálculos têm por objectivo determinar a intensidade luminosa proveniente do ponto de intersecção que é propagada na direcção do centro de projecção, isto é, na direcção contrária à da direcção do raio. Esta luz é o resultado das contribuições de várias origens que atingem o ponto de intersecção. Uma das componentes é devida à reflexão da luz proveniente das fontes de luz existentes na cena e que iluminam directamente o objecto no ponto de intersecção. Neste caso, é necessário determinar se cada uma das fontes de luz da cena ilumina realmente o ponto de intersecção ou se existe um outro objecto que se interpõe entre os dois e faz com que o ponto se encontre na sombra em relação a essa fonte. O ponto de intersecção pode também ser iluminado pela luz reflectida por outros objectos. Finalmente, se o objecto intersectado é transparente ou translúcido, existe uma outra componente para a sua iluminação devida à transmissão (com refacção na superfície do objecto) da luz através do objecto. Assim, e em resumo, existem três componentes que contribuem para a iluminação:

- luz proveniente de outros objectos, onde foi reflectida na direcção do objecto cuja iluminação se pretende determinar;
- luz refractada transmitida através do objecto e que ilumina o ponto;
- luz proveniente directamente das fontes de luz e que é reflectida pelo objecto.

A determinação de cada uma destas componentes requer cálculos próprios, sendo necessário inspeccionar as origens prováveis de cada uma delas. Para isto empregam-se os chamados raio secundários, traçados em sentido inverso ao da propagação da luz correspondente a cada uma das componentes. Uma vez que cada um destes raios

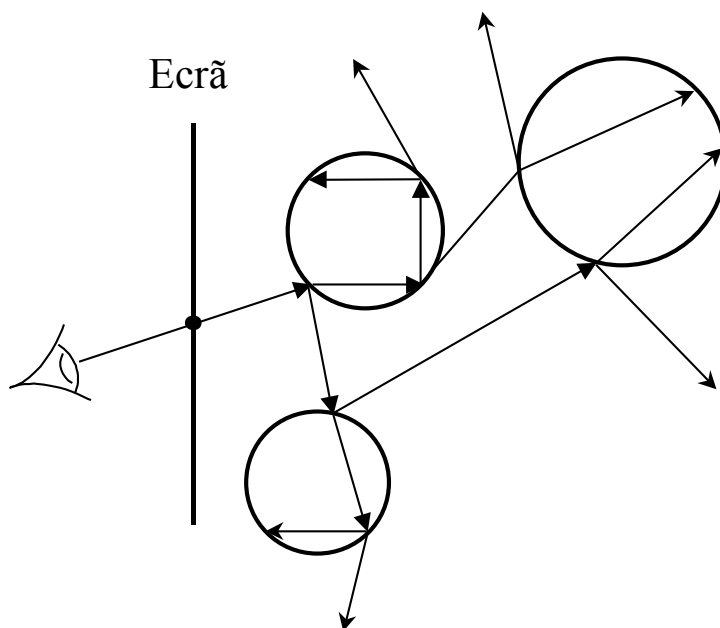
secundários tem objectivos diferentes e o seu cálculo é também diferente, convém então separá-los em três tipos:

- raios reflectidos;
- raios refractados;
- raios de iluminação directa ou de sombra (“shadow feelers”).

A determinação das intersecções destes raios secundários com os objectos da cena permite então calcular as características da luz proveniente desses objectos e a sua contribuição para a iluminação do ponto de intersecção do raio primário com o objecto original. Mas a iluminação dos pontos de intersecção dos raios secundários com os objectos da cena obriga a que se considerem novos raios secundários com origem nesses pontos de intersecção. Isto implica um processo de inspecção recursiva da cena através da geração de novos raios secundários tal como a figura 2.2 apresenta.

Forma-se assim uma árvore de ramos múltiplos, um por cada raio secundário gerado em cada ramificação. A questão que imediatamente se põe é saber quando é que este processo recursivo deve ser terminado, ou seja, até que nível de ramificação (profundidade) deve ser levado este processo.

O caso mais trivial de terminação de um ramo acontece quando um raio secundário sai para fora da cena sem intersectar qualquer objecto da mesma. Neste caso o raio secundário contribui para a iluminação com a cor de fundo da cena.



**Figura 2.2 – Cálculo recursivo da iluminação num ponto de intersecção de um raio primário com um objecto numa cena contendo objectos transparentes. Os raios secundários permitem calcular a iluminação no ponto de intersecção do raio primário devida à luz reflectida e à luz transmitida que atingem esse ponto.**

Uma outra situação, que permite terminar a inspecção de uma ramificação, verifica-se quando a respectiva contribuição para a iluminação do ponto de intersecção de partida, a raiz da árvore, se torna inferior a um valor mínimo previamente arbitrado. Nesta situação, pode-se terminar o processo de inspecção recursiva pois quaisquer ramificações que possamos considerar não irão contribuir de forma visível para a iluminação.

Os dois critérios anteriores podem, mesmo assim, não ser suficientes para terminar este processo recursivo. Para ter em conta esta eventualidade é então necessário arbitrar um critério de paragem que seja função de um nível (profundidade) máximo previamente arbitrado.

### 3 Cálculo de Intersecções

A maior parte da carga computacional dos algoritmos de ray tracing corresponde ao cálculo da intersecção de raios com os objectos da cena a representar. Whitted estimou que esta carga representa cerca de 90 a 95 % do tempo total de processamento, valores estes confirmados por medidas que posteriormente realizou. Esta estimativa mostra a necessidade do emprego de algoritmos eficientes no cálculo de intersecções.

Em primeiro lugar, os algoritmos para o cálculo de intersecções não devem realizar quaisquer cálculos significativos antes de verificar se existe ou não intersecção dos raios com os objectos. Na maior parte dos casos não existe intersecção e um raio só muito raramente intersectará todos os objectos de uma cena. Em segundo lugar, uma vez seleccionada a intersecção mais próxima da origem do raio, é necessário calcular a normal à superfície do objecto no ponto intersectado para efectuar os cálculos de iluminação e sombreamento nesse ponto. O cálculo desta normal deve ser o mais simples possível.

É pelas razões apontadas que, em ray tracing, se tenta circunscrever o cálculo das intersecções de um raio com os objectos da cena à intersecção de linhas rectas com formas geométricas simples, tais como esferas, polígonos e paralelepípedos. Esta secção apresenta de seguida o cálculo destas intersecções. Mas, antes disso, interessa realçar que a maioria dos objectos existentes numa cena não corresponde normalmente a essas formas geométricas simples. No entanto, é possível envolver cada um dos objectos ou grupos de objectos por uma daquelas formas geométricas simples e testar a intersecção dos raios com essas formas (volumes) envolventes. Se o volume envolvente não for intersectado, também o objecto ou grupo de objectos que contém não será intersectado e, conseqüentemente, poupar-se-á a carga computacional respectiva, que é normalmente pesada<sup>3</sup>. A esfera é o volume envolvente mais empregue devido às suas características isotrópicas e à simplicidade do cálculo da sua intersecção com uma linha recta. No entanto, uma esfera poderá não ser o volume envolvente mais adequado à forma do objecto ou grupo de objectos a envolver. Nesse caso, o emprego de volumes paralelepípedos ou contidos entre pares de planos paralelos constitui normalmente uma boa alternativa.

#### 3.1 Equação Paramétrica de um Raio

Um raio é uma semi-recta com origem num ponto  $\vec{R}_0$  e uma dada direcção  $\vec{R}_d$ . Sendo  $t$  o parâmetro da descrição paramétrica de uma recta, a representação vectorial paramétrica da linha sobre a qual assenta a semi-recta correspondente ao raio assume a expressão

$$\vec{R}(t) = \vec{R}_0 + \vec{R}_d t \quad (3.1)$$

<sup>3</sup> Para, por exemplo, um objecto descrito por uma malha poligonal fechada com 500 polígonos, haveria que determinar a intersecção do raio com cada um dos polígonos.

Nesta expressão, interessam-nos apenas para os valores positivos de  $t$ , dado que se trata de uma semi-recta. A expressão paramétrica de (3.1) para cada uma das coordenadas de um qualquer ponto sobre a recta será então

$$\begin{aligned}x &= x_0 + x_d t \\y &= y_0 + y_d t \\z &= z_0 + z_d t\end{aligned}\quad (3.2)$$

pois  $\vec{R}_0$  tem como componentes  $(x_0, y_0, z_0)$ , enquanto o vector  $\vec{R}_d$  tem como componentes  $(x_d, y_d, z_d)$ . Considerando o vector  $\vec{R}_1$ , que é  $\vec{R}_1 = \vec{R}_0 + \vec{R}_d$ , e cujas componentes são  $(x_1, y_1, z_1)$ , poderemos ainda obter a expressão paramétrica (3.2) em termos das componentes de  $\vec{R}_0$  e  $\vec{R}_1$

$$\begin{aligned}x &= x_0 + (x_1 - x_0)t \\y &= y_0 + (y_1 - y_0)t \\z &= z_0 + (z_1 - z_0)t\end{aligned}\quad (3.3)$$

Em ray tracing, a origem dos raios primários encontra-se no centro de projecção, enquanto  $\vec{R}_1$  corresponde ao centro do pixel cujas propriedades (cor) pretendemos determinar. O valor de  $t$  corresponderá então às seguintes situações:

$t < 0$	Pontos para trás do centro de projecção e que não pertencem ao raio
$0 \leq t \leq 1$	Pontos localizados entre o centro de projecção e o plano de projecção
$t > 1$	Pontos situados para lá do plano de projecção

### 3.2 Intersecção de um Raio com uma Esfera

A esfera é um dos objectos geométricos simples cuja intersecção com um raio é mais fácil determinar. Uma superfície esférica é o lugar geométrico dos pontos do espaço que se encontram à mesma distância de um dado ponto, o centro da esfera. Esta distância é o raio da esfera.

Seja então  $S_c$  o centro da esfera de coordenadas  $(x_c, y_c, z_c)$  e  $S_s$  um ponto qualquer da superfície esférica cujas coordenadas são  $(x, y, z)$ . A superfície da esfera é o lugar geométrico dos pontos tais que

$$|S_s - S_c| - r = 0 \quad (3.4)$$

ou

$$(x - x_c)^2 + (y - y_c)^2 + (z - z_c)^2 - r^2 = 0 \quad (3.5)$$

em que  $r$  é o raio da esfera. Desenvolvendo então (3.5), obteremos

$$x^2 - 2x_c x + x_c^2 + y^2 - 2y_c y + y_c^2 + z^2 - 2z_c z + z_c^2 - r^2 = 0 \quad (3.6)$$



e, introduzindo (3.2),

$$\begin{aligned} & x_0^2 + 2x_0 x_d t + x_d^2 t^2 - 2x_c(x_0 + x_d t) + x_c^2 + \\ & y_0^2 + 2y_0 y_d t + y_d^2 t^2 - 2y_c(y_0 + y_d t) + y_c^2 + \\ & z_0^2 + 2z_0 z_d t + z_d^2 t^2 - 2z_c(z_0 + z_d t) + z_c^2 - r^2 = 0 \end{aligned} \quad (3.7)$$

e, reagrupando,

$$\begin{aligned} & t^2 (x_d^2 + y_d^2 + z_d^2) + \\ & 2t [x_d(x_0 - x_c) + y_d(y_0 - y_c) + z_d(z_0 - z_c)] + \\ & (x_0 - x_c)^2 + (y_0 - y_c)^2 + (z_0 - z_c)^2 - r^2 = 0 \end{aligned} \quad (3.8)$$

Obtemos assim uma equação do segundo grau em t

$$A t^2 - 2 B t + C = 0 \quad (3.9)$$

em que os coeficientes A, B e C são

$$\begin{aligned} A &= x_d^2 + y_d^2 + z_d^2 \\ B &= x_d(x_c - x_0) + y_d(y_c - y_0) + z_d(z_c - z_0) \\ C &= (x_c - x_0)^2 + (y_c - y_0)^2 + (z_c - z_0)^2 - r^2 \end{aligned} \quad (3.10)$$

Se o vector  $\vec{R}_d$  for um vector unitário ( $|\vec{R}_d| = 1$ ), então  $A = 1$  e poderemos resolver (3.9) aplicando a fórmula resolvente simplificada para equações do segundo grau

$$t = B \pm \sqrt{B^2 - C} \quad (3.11)$$

Se o radicando  $B^2 - C$  for negativo, as duas raízes da equação são imaginárias conjugadas e, conseqüentemente, o raio não intersectará a esfera. Se o mesmo radicando for nulo, existe uma raiz dupla e o raio será tangente à esfera, caso este que também não nos interessa pois é equivalente a o raio não intersectar a esfera.

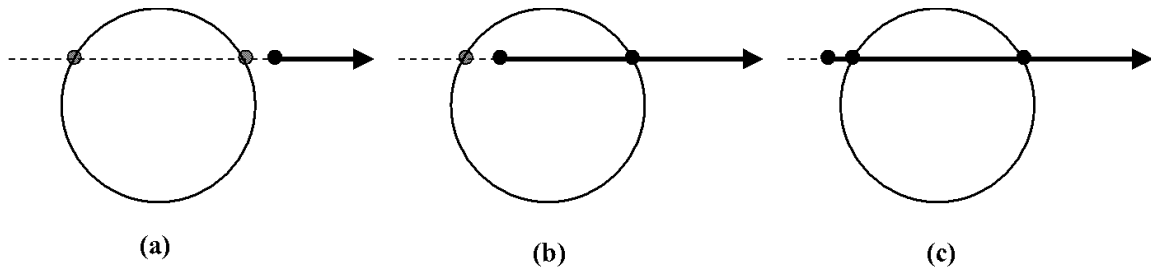
Quando (3.11) for positivo, a linha recta sobre a qual o raio existe intersectará a esfera em dois pontos correspondentes a duas raízes reais. Interessa-nos determinar o ponto de intersecção mais próximo da origem do raio que corresponderá, em princípio, à menor raiz, ou seja,  $t = B - \sqrt{B^2 - C}$ . No entanto, este valor poderá ser negativo, o que significa que a linha recta intersecta a esfera na semi-recta oposta ao raio. Há então que

determinar a outra raiz da equação para a qual  $t = B + \sqrt{B^2 - C}$ . Se esta raiz for positiva, o raio intersectará a esfera. Caso contrário o raio não intersectará a esfera, tal como a figura 3.1 apresenta.

No caso em que a origem do raio se encontra dentro da esfera (veja-se a figura 3.1(b), a raiz a calcular é  $t = B + \sqrt{B^2 - C}$ . Podemos evitar calcular primeiro a outra raiz se tivermos em conta que a distância da origem do raio ao centro da esfera é inferior ao raio da esfera. Esta condição é fácil de detectar pois corresponde a um valor negativo do coeficiente C de (3.10).

Existindo um ponto de intersecção em  $t = t_i$ , este ponto será

$$\vec{R}_i = \vec{R}_0 + \vec{R}_d t_i \quad (3.12)$$



**Figura 3.1 – Intersecção de um raio com uma esfera. A linha recta onde o raio existe intersecta a esfera, mas em (a) a esfera não é intersectada pelo raio, em (b) o raio intersecta a esfera no ponto correspondente à maior raiz e (c) o ponto de intersecção corresponde à menor raiz.**

cujas componentes são  $(x_0 + x_d t_i, y_0 + y_d t_i, z_0 + z_d t_i)$ . O vector normal exterior à esfera terá como componentes

$$\vec{N}_i = \left[ \frac{x_0 + x_d t_i - x_c}{r}, \frac{y_0 + y_d t_i - y_c}{r}, \frac{z_0 + z_d t_i - z_c}{r} \right] \quad (3.13)$$

Se o ponto de origem do raio se encontrar no interior da esfera, há ainda que calcular a normal contrária, isto é, o vector simétrico, para que a nova normal aponte para o interior da esfera em lugar de apontar para o seu exterior.

A dedução analítica da intersecção de um raio com uma esfera a partir da equação paramétrica da recta em que o raio se localiza apresenta uma desvantagem computacional importante. Esta consiste na detecção tardia (depois de vários cálculos) de que o raio não intersecta a esfera embora a linha recta a intersecte, como a figura 3.1(a) ilustra, pois o raio tem origem fora da esfera e aponta no sentido contrário à esfera. Esta situação só é detectada analiticamente depois de calculada a raiz quadrada de  $B^2 - C$  e de se verificar que ambas as raízes correspondem a valores negativos do

parâmetro  $t$ . O peso computacional do cálculo de uma raiz quadrada é elevado e, por isso, deverá ser possível detectar esta situação antes do cálculo desta raiz.

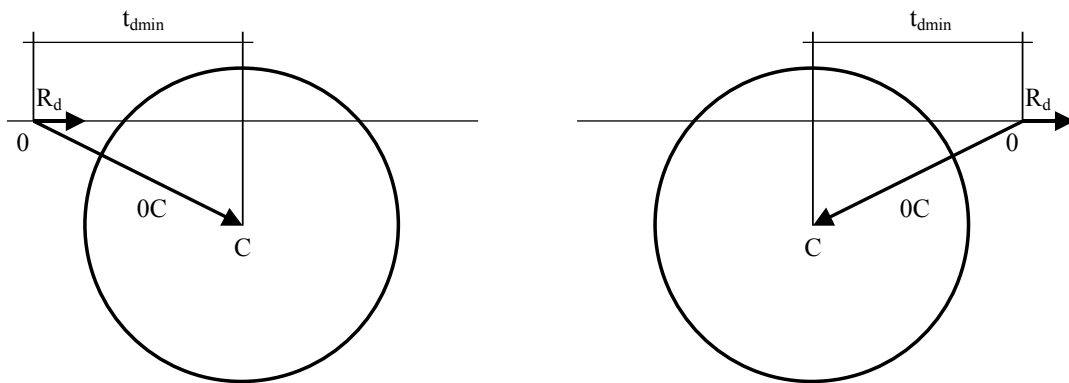
Para tal consideremos o vector que une a origem do raio ao centro da esfera (veja-se a figura 3.2) que é

$$\overrightarrow{OC} = \overrightarrow{S_c} - \overrightarrow{R_0} = [x_c - x_0, y_c - y_0, z_c - z_0] \quad (3.14)$$

O quadrado da distância da origem do raio ao centro da esfera é, por sua vez,

$$d_{OC}^2 = \overrightarrow{OC} \cdot \overrightarrow{OC} \quad (3.15)$$

onde  $\cdot$  é o operador produto interno (ou escalar) de dois vectores.



**Figura 3.2 – Intersecção de um raio com uma esfera quando a origem do raio é exterior à esfera. O valor de  $t_{dmin}$  ( $\overrightarrow{R_d} \cdot \overrightarrow{OC}$ ) é positivo quando a esfera se encontra na direcção do raio e negativo no caso contrário.**

Se  $d_{OC}^2 < r^2$ , a origem do raio encontra-se dentro da esfera e existe necessariamente intersecção do raio com a esfera. Se  $d_{OC}^2 > r^2$ , a origem do raio encontrar-se-á fora da esfera e o raio poderá ou não intersectá-la. Geometricamente, podemos concluir que o raio nunca poderá intersectar a esfera se a projecção do vector  $\overrightarrow{OC}$  sobre a linha recta que suporta o raio apresentar um sentido contrário ao sentido do raio. Mas, se os dois vectores tiverem o mesmo sentido, existe então a possibilidade de o raio intersectar a esfera. Esta verificação equivale a calcular o valor do parâmetro  $t$  correspondente ao ponto da linha recta em que o raio se apoia situado à menor distância do centro da esfera, conforme a figura 3.2 apresenta.

Para tal basta calcular o produto interno de  $\overrightarrow{OC}$  por  $\overrightarrow{R_d}$

$$t_{dmin} = \overrightarrow{OC} \cdot \overrightarrow{R_d} = x_d(x_c - x_0) + y_d(y_c - y_0) + z_d(z_c - z_0) \quad (3.16)$$

e, se  $t_{dmin}$  for negativo, concluir que não existe intersecção. A expressão (3.16) para  $t_{dmin}$  é a mesma que a do coeficiente B de (3.10).

Finalmente, se  $d_{OC}^2 = r^2$ , o raio tem origem sobre a superfície da esfera e não a intersectará a não ser que esteja orientado para o interior da esfera, o que corresponde ao caso de um raio secundário de refração que atravessa a esfera. Nestas circunstâncias, o valor do coeficiente B deverá ser positivo, tal como no caso em que a origem do raio se encontra no exterior da esfera.

Deste modo, podemos agora apresentar o algoritmo para o cálculo da intersecção de um raio com uma esfera, otimizado para evitar o cálculo de raízes quadradas quando não existe intersecção.

1. Normalizar  $\vec{R}_d$  para que  $|\vec{R}_d| = 1$ .
2. Calcular o quadrado da distância da origem do raio ao centro da esfera
 
$$d_{OC}^2 = (x_c - x_0)^2 + (y_c - y_0)^2 + (z_c - z_0)^2$$
3. Calcular
 
$$B = x_d(x_c - x_0) + y_d(y_c - y_0) + z_d(z_c - z_0)$$
4. Se  $d_{OC}^2 \geq r^2$ , testar se B é negativo e, em caso afirmativo, abandonar os cálculos de intersecção (o raio aponta no sentido contrário ao da localização da esfera relativamente à origem do raio).
5. Calcular
 
$$R = B^2 - C = B^2 - d_{OC}^2 + r^2$$
 e abandonar os cálculos se R for negativo.
6. Determinar  $t_i$ 

$$t_i = \begin{cases} B - \sqrt{R} & \text{se } d_{OC}^2 > r^2 \\ B + \sqrt{R} & \text{se } d_{OC}^2 \leq r^2 \end{cases}$$
7. Calcular o ponto de intersecção
 
$$\vec{R}_i = [x_0 + x_d t_i, \quad y_0 + y_d t_i, \quad z_0 + z_d t_i]$$
8. Calcular a normal à superfície da esfera no ponto de intersecção
 
$$\vec{N}_i = \left[ \frac{x_i - x_c}{r}, \quad \frac{y_i - y_c}{r}, \quad \frac{z_i - z_c}{r} \right]$$
 trocando os sinais das componentes quando  $d_{OC}^2 \leq r^2$  (a origem do raio encontra-se no interior da esfera ou sobre a sua superfície).

### 3.3 Intersecção de um Raio com um Plano

A equação implícita de um plano é

$$A x + B y + C z + D = 0 \quad (3.17)$$

em que  $A^2 + B^2 + C^2 = 1$ , e cuja normal é um vector unitário tal que

$$\vec{N} = [A, \quad B, \quad C] \quad (3.18)$$

Substituindo em (3.17)  $x$ ,  $y$  e  $z$  pelas expressões da equação paramétrica de uma linha recta de (3.2)

$$A(x_0 + x_d t) + B(y_0 + y_d t) + C(z_0 + z_d t) + D = 0 \quad (3.19)$$

e, explicitando  $t$ , obtemos a expressão para o valor do parâmetro  $t_i$  no ponto de intersecção de uma recta com um plano

$$t_i = \frac{-(Ax_0 + By_0 + Cz_0 + D)}{Ax_d + By_d + Cz_d} \quad (3.20)$$

ou, em notação vectorial,

$$t_i = -\frac{\vec{N} \cdot \vec{R}_0 + D}{\vec{N} \cdot \vec{R}_d} \quad (3.21)$$

Por questões de eficiência, convém primeiro calcular o produto interno  $\vec{N} \cdot \vec{R}_d$  pois, se este for nulo, a recta será paralela ao plano e não existirá intersecção<sup>4</sup>. Caso contrário, existe intersecção entre a linha recta que suporta o raio e o plano, e poderemos calcular o valor de  $t_i$ . Se este valor for positivo, o raio intersectará o plano no ponto  $P_i$  de coordenadas<sup>5</sup>

$$\vec{R}_i = [x_i, y_i, z_i] = [x_0 + x_d t_i, y_0 + y_d t_i, z_0 + z_d t_i] \quad (3.22)$$

A normal ao plano no ponto de intersecção determinado será  $\vec{N}$ , se  $\vec{N} \cdot \vec{R}_d$  for negativo, e  $-\vec{N}$ , se for positivo.

Devemos considerar ainda o caso em que o plano é um plano em que existe um objecto planar de uma única face, pois, neste caso, o algoritmo apresentado pode ser simplificado dado que só nos interessará considerar raios com origem no semi-espaco positivo em relação ao plano. Nestas circunstâncias, poderemos eliminar liminarmente os raios para os quais  $\vec{N} \cdot \vec{R}_0 + D$  é negativo ou nulo (a origem do raio está por detrás do plano ou sobre ele) e, dos raios restantes, aqueles para os quais seja  $\vec{N} \cdot \vec{R}_d \geq 0$  (o sentido do raio é idêntico ao do vector da normal ao plano).

### 3.4 Intersecção de um Raio com um Polígono

Uma vez determinado o ponto de intersecção de um raio com o plano onde existe um polígono, esse ponto será o ponto de intersecção do raio com o polígono, se esta existir.

<sup>4</sup> O caso limite em que a recta existe no plano também não tem qualquer interesse.

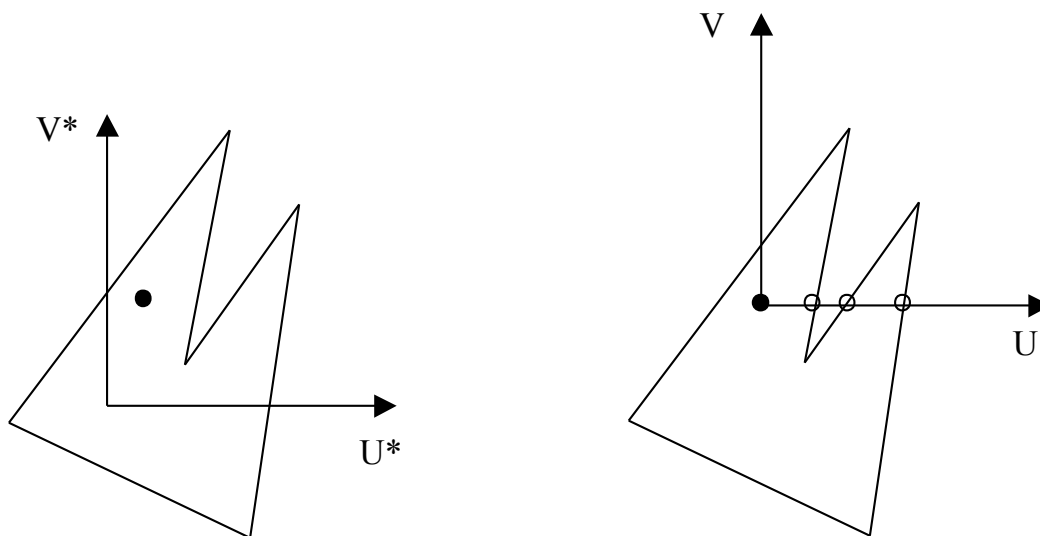
<sup>5</sup> Um valor de  $t_i$  negativo significa que o plano se encontra no sentido contrário ao do sentido do raio, enquanto que um valor nulo de  $t_i$  corresponde à origem do raio se encontrar sobre o plano. Ambos os casos não nos interessam.

Esta verificação consiste em determinar se o ponto pertence ao interior do polígono. Uma das formas mais eficientes de o fazer consiste no algoritmo de Sedgewick modificado por Haines e que se baseia no teorema da curva de Jordan. Segundo este teorema, um ponto pertencerá ao interior de um polígono se uma qualquer semi-recta com origem no ponto intersectar os lados do polígono um número ímpar de vezes. Este teorema permite tratar polígonos côncavos e autointersectantes.

O primeiro passo do algoritmo consiste em projectar o plano do polígono num dos planos principais do espaço XYZ. Este plano de projecção poderá ser qualquer dos planos XY, XZ ou YZ e a projecção será obtida pela simples eliminação de uma das coordenadas Z, Y ou X, respectivamente. A escolha da coordenada a eliminar deve ser feita examinando as componentes da normal ao plano e eliminando a direcção do espaço correspondente à maior componente em valor absoluto do vector normal ao plano.

O espaço projectado,  $U^*V^*$ , é um espaço bidimensional topologicamente semelhante ao espaço original, tal como a figura 3.3 apresenta. O passo seguinte consiste em deslocar a origem do espaço  $U^*V^*$  para o ponto correspondente à intersecção do raio com o plano, obtendo-se assim um novo sistema de coordenadas UV com origem na projecção do ponto de intersecção.

Considerando agora o semi-eixo U positivo<sup>6</sup>, resta determinar o número de intersecções deste semi-eixo com os lados do polígono. Existirá intersecção entre o semi-eixo e um dos lados do polígono sempre que os dois vértices do lado do polígono apresentem valores de U positivos e os valores de V sejam de sinal contrário ou quando, simultaneamente, os valores das coordenadas U dos vértices sejam de sinal contrário e os de V também, e a intersecção do lado do polígono com o eixo dos UU apresente um valor de U positivo. Para efeitos desta determinação considera-se que um valor nulo é positivo.



**Figura 3.3 – Algoritmo para verificar se um ponto pertence ao interior de um polígono através da determinação do número de intersecções com os lados do polígono de uma semi-recta com origem no ponto.**

<sup>6</sup> Poderíamos também considerar o semi-eixo U negativo ou qualquer dos outros dois semi-eixos de V.

Tomemos então um polígono com  $N$  vértices de coordenadas  $U_n, V_n$  em que  $n=0, \dots, N-1$ . O algoritmo para determinar o número de intersecções,  $NI$ , apresenta então os seguintes passos:

1. Inicializar  $NI$  a zero.
2. Testar o valor de  $V_0$  e inicializar o valor para teste de sinal,  $VS$ , de acordo com<sup>7</sup>

$$VS = \begin{cases} -1 & \text{se } V_0 < 0 \\ 1 & \text{se } V_0 \geq 0 \end{cases}$$

3. Repetir os passos 4 a 9 para  $n=0, \dots, N-1$
4. Calcular  $k = \text{resto da divisão de } n+1 \text{ por } N$
5. Testar o valor de  $V_k$  e atribuir o valor para teste local do sinal,  $VSL$ , de acordo com

$$VSL = \begin{cases} -1 & \text{se } V_k < 0 \\ 1 & \text{se } V_k \geq 0 \end{cases}$$

6. Se  $VSL$  for igual a  $VS$ , prosseguir para o passo 9.
7. Se tanto  $U_n$  como  $U_k$  forem positivos, incrementar o número de intersecções,  $NI$ , e prosseguir para o passo 9.
8. Se  $U_n$  e  $U_k$  forem de sinais contrários, determinar se o lado do polígono intersecta o semieixo positivo, isto é, se

$$U_n - V_n \frac{U_k - U_n}{V_k - V_n} > 0$$

e incrementar  $NI$  se esta condição se verificar.

9. Atribuir a  $VS$  o valor corrente de  $VSL$ .
10. Testar se o número de intersecções determinado,  $NI$ , é par ou ímpar. Se o valor for ímpar, o ponto pertence ao interior do polígono. Caso contrário o ponto não pertence ao interior do polígono.

### 3.5 Intersecção de um Raio com um Paralelepípedo Orientado Segundo os Eixos

Um dos processos para aumentar a eficiência dos cálculos de intersecção em ray tracing consiste em determinar se existe ou não intersecção entre o raio e um volume envolvente de um objecto ou conjunto de objectos, antes de determinar as intersecções desse raio com cada um dos objectos. Se não houver intersecção do raio com o volume envolvente, não haverá intersecção do raio com qualquer dos objectos contidos no volume e o cálculo será mais rápido, como mais adiante veremos.

Um dos volumes envolventes mais empregues é o correspondente ao de um paralelepípedo orientado segundo os eixos da cena e que delimita a região do espaço

<sup>7</sup> A possibilidade de um vértice poder ser contado duas vezes por se encontrar sobre o eixo dos  $UU$  é evitada ao se considerar que  $V$  é positivo sempre que  $V$  seja nulo.

compreendida entre  $x_{\min}$  e  $x_{\max}$ ,  $y_{\min}$  e  $y_{\max}$ , e  $z_{\min}$  e  $z_{\max}$ . O algoritmo de Kay e Kajiya permite determinar a intersecção de um raio com um paralelepípedo assim definido, calculando sucessivamente os dois pontos de intersecção do raio com o par de planos correspondente ao mínimo e ao máximo das coordenadas das faces do paralelepípedo segundo cada uma das direcções do espaço. Como os planos estão alinhados com as direcções do espaço XYZ, o cálculo da intersecção do raio com os planos segundo (3.21) fica consideravelmente simplificado. Para, por exemplo, os planos para os quais  $x = C^{te}$ , temos que o correspondente vector normal será

$$\vec{N} = [1, 0, 0] \quad (3.23)$$

pelo que (3.21) passa a ser

$$t_i = -\frac{x_0 - C^{te}}{x_d} = \frac{C^{te} - x_0}{x_d} \quad (3.24)$$

Se  $x_d$  for nulo, o raio é paralelo aos dois planos e só poderá existir intersecção do raio com o volume envolvente se a origem do raio se encontrar dentro do paralelepípedo, ou seja, o valor de  $x$ , neste caso, estiver compreendido entre  $x_{\min}$  e  $x_{\max}$ <sup>8</sup>.

O algoritmo de Kay e Kajiya devolve o valor lógico verdadeiro se houver intersecção e os valores de correspondentes aos pontos de intersecção mais próximo ( $t_{\text{prox}}$ ) e mais afastado ( $t_{\text{dist}}$ ) da origem do raio ou o valor lógico de falso quando não existir intersecção.

Os passos do algoritmo de Kay e Kajiya são os seguintes:

1. Inicializar  $t_{\text{prox}} = -\infty$  e  $t_{\text{dist}} = +\infty$ <sup>9</sup>.
2. Repetir os passos 3 a 7 para, sucessivamente, as direcções X, Y e Z, inicializando os valores de  $V_0$ ,  $V_d$ ,  $V_{\min}$  e  $V_{\max}$  para cada uma daquelas direcções de acordo com

	X	Y	Z
$V_0$	$x_0$	$y_0$	$z_0$
$V_d$	$x_d$	$y_d$	$z_d$
$V_{\min}$	$x_{\min}$	$y_{\min}$	$z_{\min}$
$V_{\max}$	$x_{\max}$	$y_{\max}$	$z_{\max}$

3. Se  $V_d$  for nulo, o raio é paralelo aos planos e, se  $V_0 < V_{\min}$  ou  $V_0 > V_{\max}$ , devolver falso e terminar. Se não, voltar ao passo 2 e prosseguir com a direcção do espaço seguinte.

<sup>8</sup> Se o raio for paralelo aos planos e  $x_0 < x_{\min}$  ou  $x_0 > x_{\max}$ , a origem do raio encontra-se fora do volume do paralelepípedo e o raio nunca o poderá intersectar.

<sup>9</sup> Para este efeito sugere-se o emprego das constantes pré-definidas FLT\_MAX ou DBL\_MAX da linguagem C.



4. Calcular as intersecções

$$t_{\min} = \frac{V_{\min} - V_0}{V_d}$$

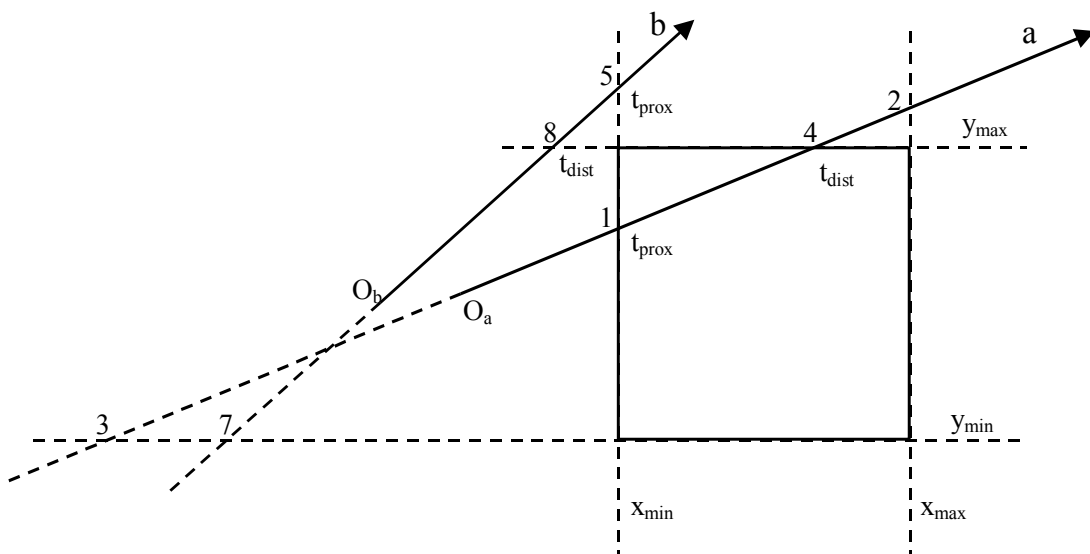
$$t_{\max} = \frac{V_{\max} - V_0}{V_d}$$

e trocar  $t_{\min}$  com  $t_{\max}$  se  $t_{\min} > t_{\max}$ . Actualizar então os valores de  $t_{\text{prox}}$  e  $t_{\text{dist}}$  de acordo com

$$\begin{array}{ll} \text{se} & t_{\min} > t_{\text{prox}} & - & t_{\text{prox}} = t_{\min} \\ \text{se} & t_{\max} < t_{\text{dist}} & - & t_{\text{dist}} = t_{\max} \end{array}$$

5. Se  $t_{\text{prox}} > t_{\text{dist}}$  o raio não intersecta os dois planos testados e o algoritmo deve parar, devolvendo falso.
6. Se  $t_{\text{dist}} < 0$ , o raio aponta na direcção contrária à da localização do paralelepípedo e, portanto, não o intersecta. O algoritmo deve parar e devolver falso.
7. Voltar ao passo 2 e prosseguir com a direcção seguinte do espaço.
8. O algoritmo encontrou as duas intersecções do raio com o paralelepípedo em que  $t_{\text{prox}}$  corresponde ao ponto de entrada do raio no paralelepípedo e  $t_{\text{dist}}$  ao ponto de saída.

A figura 3.4 ilustra os vários passos deste algoritmo, considerando apenas as direcções X e Y por razões de simplicidade. O raio a intersecta o volume, mas o raio b não. O raio a intersecta o plano para o  $x = x_{\min}$  no ponto 1 e o plano  $x = x_{\max}$  no ponto 2. Como  $t_1 < t_2$  não há lugar a troca e teremos  $t_{\text{prox}} = t_1$  e  $t_{\text{dist}} = t_2$ . Como as condições dos passos 5 e 6 se não verificam, podemos prosseguir determinando as intersecções do raio a com os planos para os quais  $y = y_{\min}$  e  $y = y_{\max}$ . Estas intersecções correspondem aos pontos 3 e 4 da figura, respectivamente. Teremos então  $t_{\min} = t_3$  e  $t_{\max} = t_4$ . Como  $t_{\min}$  é menor que  $t_{\text{prox}}$  este fica inalterado. Mas  $t_{\max}$  é menor que  $t_{\text{dist}}$  pelo que  $t_{\text{dist}} = t_4$ . O raio A intersecta então a região do espaço examinada, entrando nela em  $t_{\text{prox}} = t_1$  e saindo em  $t_{\text{dist}} = t_4$ .



**Figura 3.4 – Algoritmo de Kay e Kajiya num caso bidimensional, mostrando dois raios. O raio a intersecta o volume, o raio b não.**

O raio  $b$  intersecta o plano para o qual  $x = x_{\min}$  no ponto 5 e o plano de  $x = x_{\max}$  num ponto não representado na figura para o qual  $t_6 > t_5$ . Logo teremos  $t_{\min} = t_5$  e  $t_{\max} = t_6$ , sendo estes valores os valores a atribuir a  $t_{\text{prox}}$  e  $t_{\text{dist}}$ , respectivamente. As condições dos passos 5 e 6 não se verificam e, conseqüentemente, passa-se à determinação das intersecções do raio  $b$  com os planos correspondentes a  $y_{\min}$  e  $y_{\max}$ . Estas intersecções verificam-se nos pontos 7 e 8 e teremos  $t_{\min} = t_7$  e  $t_{\max} = t_8$ . O valor de  $t_{\text{prox}}$  não vai ser alterado porque  $t_{\min} < t_{\text{prox}}$ , mas  $t_{\text{dist}}$  é alterado para o valor de  $t_{\min}$  uma vez que este tem um valor menor. Obtemos assim  $t_{\text{prox}} = t_5$  e  $t_{\text{dist}} = t_8$ . O teste do passo 5 falha pois  $t_8 < t_5$  e, portanto, o raio não intersecta o volume.

## 4 Interação da Luz com as Superfícies

Como vimos anteriormente, a cor de um ponto de uma superfície depende de três contribuições:

- cor da luz reflectida pela superfície;
- cor da luz refractada;
- cor local devida à iluminação da superfície pela luz recebida directamente das fontes de luz da cena e pela luz ambiente.

A figura 4.1 representa estas três componentes. De seguida examinaremos cada uma delas.

### 4.1 Reflexão Especular Perfeita

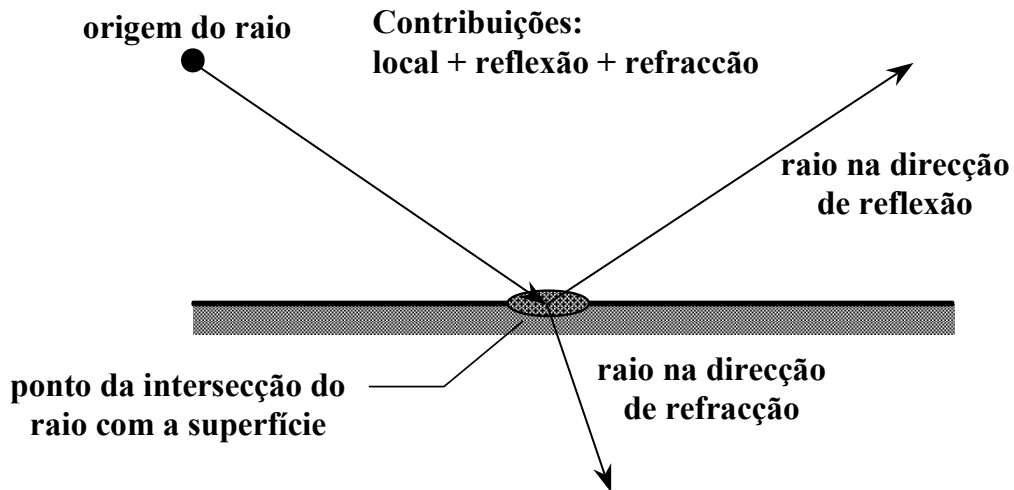
A reflexão especular perfeita ocorre em superfícies lisas e brilhantes, normalmente espelhadas, metálicas ou de vidro. Estas superfícies apresentam a característica comum de serem polidas. A reflexão será tanto mais perfeita quanto mais perfeito for o seu polimento. Isto significa que na realidade não existem superfícies reflectoras perfeitas. O modelo de reflexão especular perfeita corresponde a uma situação limite, ainda que, na maior parte das situações, providencie uma aproximação suficientemente exacta do que na realidade acontece.

A figura 4.2 apresenta a interpretação geométrica da reflexão especular perfeita de um raio de luz que incide sobre uma superfície reflectora. O raio de luz incidente  $I$  faz com a normal à superfície,  $N$ , um ângulo  $\theta_i$  denominado ângulo de incidência. Por seu lado, o raio reflectido  $R$  faz um ângulo  $\theta_r$ , o ângulo de reflexão, com a normal à superfície. O raio reflectido existe no plano definido pelo raio incidente e pela normal à superfície e o ângulo de reflexão é igual ao ângulo de incidência.

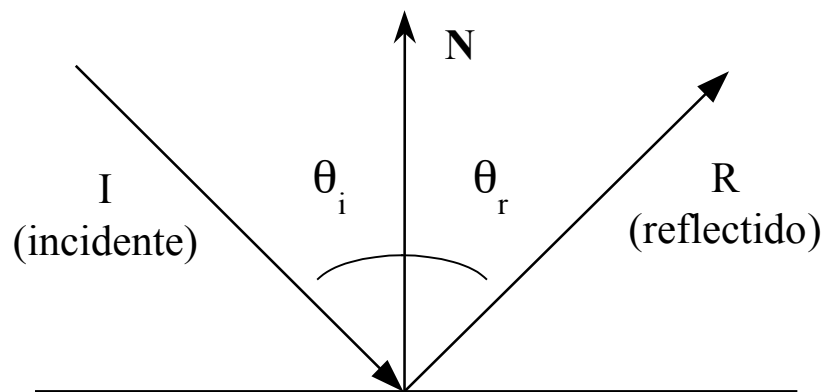
Se o raio reflectido existe no plano definido pelo raio incidente e pela normal (ambos normalizados), poderemos expressá-lo em função destes dois vectores e teremos

$$\vec{R} = \alpha \vec{I} + \beta \vec{N} \quad (4.1)$$

em que  $\vec{N}$  representa o vector unitário normal à superfície.



**Figura 4.1 – Iluminação de um ponto. A luz proveniente do ponto e detectada pelo observador deve-se à soma das contribuições da iluminação local (directa e ambiente), da luz reflectida e da luz transmitida.**



**Figura 4.2 – Reflexão especular perfeita. O raio incidente, o raio reflectido e o vector normal à superfície pertencem ao mesmo plano. Os ângulos de incidência e de reflexão são iguais.**

Os co-senos dos dois ângulos são iguais dado que os dois ângulos são iguais e como

$$\begin{aligned} \cos\theta_i &= -\vec{I} \cdot \vec{N} \\ \cos\theta_r &= \vec{R} \cdot \vec{N} \end{aligned} \tag{4.2}$$

logo

$$\begin{aligned} -\vec{I} \cdot \vec{N} &= \vec{R} \cdot \vec{N} \\ &= (\alpha \vec{I} + \beta \vec{N}) \cdot \vec{N} \\ &= \alpha (\vec{I} \cdot \vec{N}) + \beta (\vec{N} \cdot \vec{N}) \\ &= \alpha (\vec{I} \cdot \vec{N}) + \beta \end{aligned} \tag{4.3}$$

Note-se que, porque  $|\vec{N}| = 1$ , temos  $\vec{N} \cdot \vec{N} = 1$ . Arbitrando agora  $\alpha = 1$ , virá

$$\beta = -2(\vec{I} \cdot \vec{N}) \quad (4.4)$$

que, substituindo em (4.1), conduz a

$$\vec{R} = \vec{I} - 2(\vec{I} \cdot \vec{N})\vec{N} \quad (4.5)$$

ou ainda, tendo em conta (4.2),

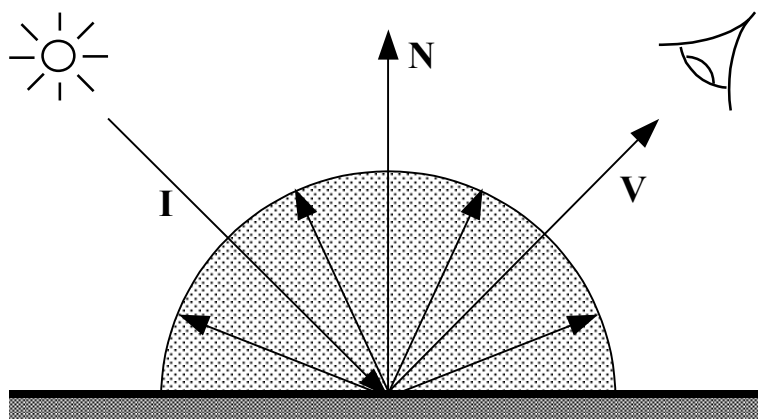
$$\vec{R} = \vec{I} + 2 \cos \theta \vec{N} \quad (4.6)$$

## 4.2 Reflexão Difusa Perfeita

Nas superfícies rugosas não existe uma direcção pré-determinada segundo a qual a luz incidente é reflectida. Pelo contrário, todas as direcções são igualmente prováveis e, segundo o modelo de reflexão difusa, a intensidade da luz reflectida é igual em todas as direcções.

A lei de Lambert para a reflexão difusa permite calcular os efeitos produzidos por feixes de luz incidente em função do seu ângulo de incidência. O fluxo de um feixe de luz incidente (veja-se a figura 4.3) é determinado tendo em conta a secção unitária perpendicular à direcção de propagação do feixe. O fluxo correspondente referido à superfície sobre a qual o feixe incide é então determinado projectando esta secção unitária sobre a superfície, ou seja, multiplicando a secção perpendicular ao feixe pelo co-seno do ângulo entre a secção e a superfície. Este ângulo é igual ao ângulo entre o feixe incidente e a normal à superfície e o respectivo co-seno não é mais do que o produto interno da direcção do feixe pela normal à superfície. Assim a intensidade luminosa que atinge a superfície será

$$-(\vec{I} \cdot \vec{N})\vec{N} \quad (4.7)$$



**Figura 4.3 – Reflexão difusa. A intensidade reflectida é igual em todas as direcções.**

### 4.3 Transmissão Especular Perfeita e Refracção

A luz propaga-se através dos objectos transparentes ou translúcidos<sup>10</sup>. É por esta razão que podemos observar objectos imersos em água quer directamente, quer através de paredes de vidro como em copos, oceanários, etc. Os objectos aparentam estar mais próximos do observador do que realmente se encontram. Por outro lado, se observarmos objectos parcialmente imersos, estes aparentam estar quebrados na superfície de separação entre o ar e a água. Estas observações são causadas pela ocorrência do fenómeno de refração na interface entre a água e o ar, fenómeno este que é devido às diferentes velocidades de propagação da luz nestes dois meios.

Quando um raio luminoso atravessa a interface (a superfície separadora) entre dois meios, a direcção de propagação do raio altera-se. O ângulo de incidência  $\theta_i$  do raio incidente  $\vec{I}$  em relação à normal no ponto de incidência é diferente do ângulo  $\theta_t$  do raio emergente (refractado)  $\vec{T}$  no outro meio, conforme a figura 4.4 apresenta. A lei de Snell estabelece a relação entre os dois ângulos e os índices de refração ( $\eta$ ) dos dois meios que é

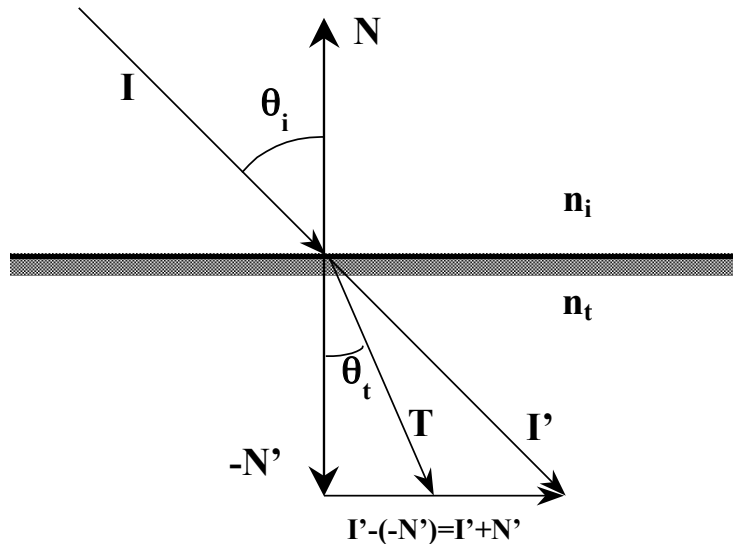
$$\frac{\sin \theta_i}{\sin \theta_t} = \frac{\eta_t}{\eta_i} = \eta_{ii} \quad (4.8)$$

em que  $\eta_{ii}$  é o índice de refração do meio onde se propaga o raio refractado em relação ao meio onde se propaga o raio incidente. O índice de refração de um meio depende das propriedades ópticas desse meio e varia de meio para meio. A tabela 4.1 apresenta os valores do índice de refração para alguns meios mais comuns.

Meio	Índice de refração referido ao vácuo
Ar (293,15 K, 101,325 kPa)	1,00029
Água	1,33
Vidro crown	1,52
Vidro denso	1,66
Vidro cristal	2
Quartzo	1,46
Safira	1,77
Diamante	2,417
Cristais de iodo	3,34
Poliestireno	1,55
Álcool etílico	1,36

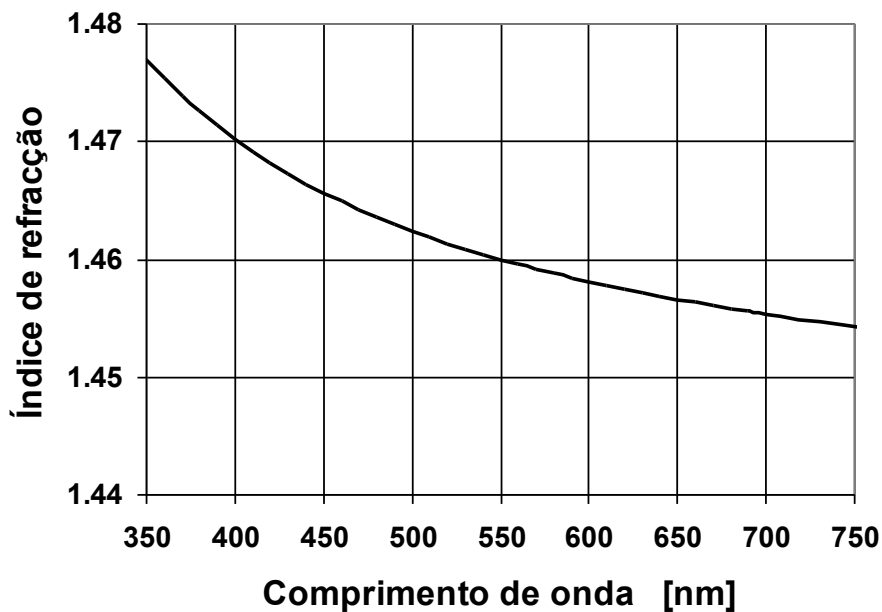
**Tabela 4.1 – Valores do índice de refração de alguns meios ópticos.**

<sup>10</sup> Em geral, um objecto só é transparente para determinadas gamas de comprimento de onda da radiação electromagnética. O vidro é transparente à luz, mas não permite a transmissão da radiação infravermelha.



**Figura 4.4 – Refracção de um raio luminoso na superfície separadora de dois meios de densidades ópticas diferentes (o vector  $-\vec{N}'$  é a projecção do vector do raio refractado,  $\vec{T}$ , sobre a direcção normal à superfície de separação dos dois meios).**

O valor do índice de refração de um meio depende também da frequência (ou do comprimento de onda) da luz. Este facto está na origem do conhecido fenómeno da separação da luz branca nas suas várias componentes cromáticas que se observa quando a luz branca atravessa um prisma. A figura 4.5 apresenta a variação do índice de refração dos cristais de quartzo em função do comprimento de onda. Embora esta variação seja pequena, ela é todavia suficiente para que o fenómeno seja observado.



**Figura 4.5 – Variação do índice de refração dos cristais de quartzo em função do comprimento de onda da luz.**

Consideremos então a figura 4.4. O vector  $-\vec{N}'$  é a projecção do vector representando o raio refractado,  $\vec{T}$ , sobre a direcção da normal à superfície de separação dos dois meios. Por sua vez, o vector  $\vec{I}'$  é um vector com origem no ponto de incidência do vector incidente,  $\vec{I}$ , na mesma superfície, com a mesma direcção deste vector e tal que a sua projecção sobre a direcção da normal à superfície é também igual a  $-\vec{N}'$ . A diferença entre  $\vec{I}'$  e  $-\vec{N}'$  é o vector

$$\vec{I}' - (-\vec{N}') = \vec{I}' + \vec{N}' \quad (4.9)$$

Escalando de um valor  $k$  conveniente ( $0 < k < 1$ ) este vector diferença e somando o resultado ao vector  $-\vec{N}'$ , obteremos o vector correspondente ao raio reflectido que será

$$\begin{aligned} \vec{T} &= -\vec{N}' + k(\vec{I}' + \vec{N}') \\ &= k\vec{I}' + (k-1)\vec{N}' \end{aligned} \quad (4.10)$$

Como a projecção de  $\vec{I}'$  sobre a direcção da normal à superfície é igual à projecção de  $\vec{T}$  sobre a mesma direcção e sendo os módulos das suas projecções iguais ao módulo do vector  $\vec{N}'$ , teremos

$$|T|\cos\theta_t = |I'|\cos\theta_i = |N'| \quad (4.11)$$

ou, tendo em conta que  $|T|=1$ , porque se trata de um versor (vector unitário),

$$|I'| = \frac{\cos\theta_t}{\cos\theta_i} \quad |N'| = \cos\theta_t \quad (4.12)$$

pelo que

$$\vec{I}' = |I'| \vec{I} = \frac{\cos\theta_t}{\cos\theta_i} \vec{I} \quad \vec{N}' = |N'| \vec{N} = \cos\theta_t \vec{N} \quad (4.13)$$

O módulo do vector  $\vec{I}' + \vec{N}'$  é

$$|\vec{I}' + \vec{N}'| = |I'| \operatorname{sen}\theta_i \quad (4.14)$$

e o módulo do vector  $k(\vec{I}' + \vec{N}')$  é, por sua vez,

$$k |\vec{I}' + \vec{N}'| = |T| \operatorname{sen}\theta_t = \operatorname{sen}\theta_i \quad (4.15)$$

porque  $|T| = 1$ . Logo teremos  $k |I'| \text{sen}\theta_i = \text{sen}\theta_t$  e, portanto,

$$k = \frac{1}{|I'|} \frac{\text{sen}\theta_t}{\text{sen}\theta_i} \quad (4.16)$$

que, tendo em conta (4.12), é também

$$k = \frac{\text{sen}\theta_t \cos\theta_i}{\text{sen}\theta_i \cos\theta_t} \quad (4.17)$$

Introduzindo agora as expressões (4.16) e (4.17) em (4.10), teremos

$$\begin{aligned} \vec{T} &= \frac{1}{|I'|} \frac{\text{sen}\theta_t}{\text{sen}\theta_i} \vec{I}' + \left( \frac{\text{sen}\theta_t \cos\theta_i}{\text{sen}\theta_i \cos\theta_t} - 1 \right) \vec{N}' \\ &= \frac{\text{sen}\theta_t}{\text{sen}\theta_i} \vec{I} + \left( \frac{\text{sen}\theta_t \cos\theta_i}{\text{sen}\theta_i \cos\theta_t} - 1 \right) \cos\theta_t \vec{N} \\ &= \frac{\text{sen}\theta_t}{\text{sen}\theta_i} \vec{I} + \left( \frac{\text{sen}\theta_t}{\text{sen}\theta_i} \cos\theta_i - \cos\theta_t \right) \vec{N} \end{aligned} \quad (4.18)$$

e, introduzindo a lei de Snell, teremos finalmente

$$\vec{T} = \eta_{it} \vec{I} + (\eta_{it} \cos\theta_i - \cos\theta_t) \vec{N} \quad (4.19)$$

#### 4.4 Reflexão Total na Interface entre Dois Meios

Em geral, um raio luminoso sofre uma refração ao atravessar a interface entre dois meios de densidade óptica diferente. Mas, quando a passagem se dá de um meio mais denso para outro menos denso, o raio pode ser reflectido na interface em vez de ser refractado, tal como a figura 4.6 apresenta. Este fenómeno é denominado de reflexão total e a sua ocorrência depende do ângulo de incidência do raio luminoso e dos índices de refração dos dois meios.

Explicitando na lei de Snell o ângulo de emergência do raio (o ângulo do raio refractado), obtemos

$$\text{sen}\theta_t = \frac{\eta_i}{\eta_t} \text{sen}\theta_i \quad (4.20)$$

Se o meio do raio incidente é opticamente mais denso do que o meio do raio refractado, teremos  $\eta_i > \eta_t$  e, conseqüentemente, o seno do ângulo de refração será maior que o seno do ângulo de incidência, o que significa que

$$\theta_t > \theta_i \quad (4.21)$$



O valor máximo que o ângulo  $\theta_i$  pode assumir é  $90^\circ$ . Existirá então um valor limite do ângulo  $\theta_i$  a partir do qual deixa de se dar a refração e passa a verificar-se a reflexão dos raios incidentes na interface entre os dois meios. Este ângulo limite é chamado ângulo crítico. O seu valor é apenas função da relação entre os índices de refração absolutos (em relação ao vácuo) dos dois meios e é

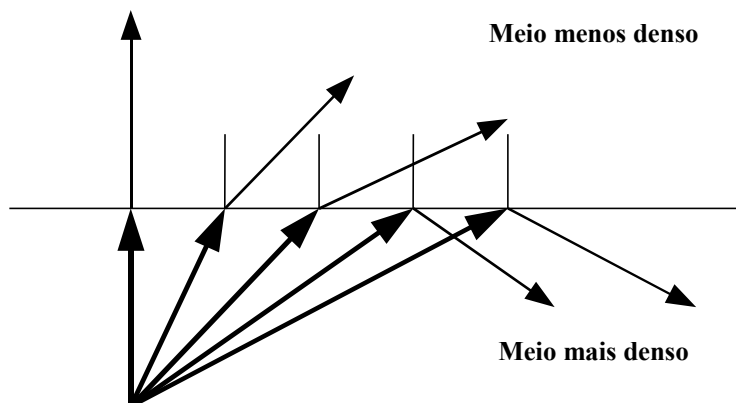
$$\theta_i = \arcsen\left(\frac{\eta_t}{\eta_i} \operatorname{sen}90^\circ\right) = \arcsen\left(\frac{\eta_t}{\eta_i}\right) \quad (4.22)$$

A tabela 4.5 apresenta os valores do ângulo crítico para algumas combinações de dois meios.

O fenómeno de reflexão total não pode ocorrer na passagem de um meio opticamente menos denso para um meio mais denso, porque, tendo em atenção (4.20) e que neste caso é  $\eta_i < \eta_t$ , o ângulo de refração será sempre inferior ao ângulo de incidência.

Meios	Ângulo crítico
Água - Ar	48° 46'
Vidro (crown) - Ar	41° 9'
Vidro (crown) - Água	61° 3'
Diamante - Ar	24° 27'
Safira - Ar	34° 25'

**Tabela 4.2 – Valores do ângulo crítico na passagem de um meio opticamente mais denso para um meio opticamente menos denso.**



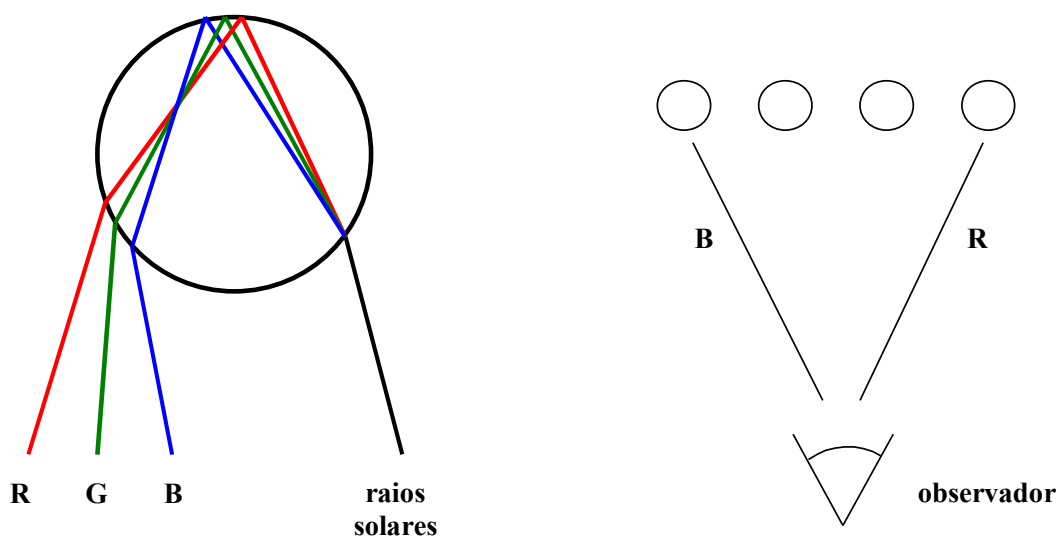
**Figura 4.6 – Reflexão total. A luz é reflectida quando, ao tentar passar de um meio opticamente mais denso para outro menos denso, o seu ângulo de incidência é superior ao ângulo crítico.**

## 4.5 Refracção e Reflexão Total Combinadas

Os fenómenos de refracção e reflexão total podem ocorrer simultaneamente, dando origem a fenómenos naturais com interesse em Computação Gráfica. O mais interessante de entre estes fenómenos é o arco-íris, cuja explicação se deve a Descartes.

O arco-íris é o resultado da refracção da luz solar ao entrar e sair de gotas de água em queda (chuva) ou suspensas na atmosfera, com pelos menos uma ou mais reflexões totais no interior das gotas, conforme a figura 4.7 apresenta. A refracção da luz desvia os vários comprimentos de onda da luz solar de uma quantidade angular diferente, pois o índice de refracção ar-água varia com o comprimento de onda da luz, tal como sucede num prisma. O desvio angular é menor para os comprimentos de onda maiores (zona do vermelho) e maior para os pequenos comprimentos de onda (zona do azul). À saída das gotas de água, os raios de cada cor (e portanto já separados) emergem das gotas em direcções diferentes. Um observador, que deverá estar de costas para o Sol, com este à sua esquerda ou à sua direita, verá apenas os raios que saíam das gotas na sua direcção.

Deste modo (veja-se a figura 4.7), quando a luz solar vinda da direita e por detrás incide nas gotas de água, o observador verá a gama de vermelhos do arco-íris do lado direito do arco e a gama dos azuis do lado esquerdo do mesmo arco, tal como a figura apresenta. As cores do arco-íris permutam as posições anteriores se a luz solar incidir nas gotas vinda da esquerda do observador.



**Figura 4.7 – Refracção e reflexão da luz solar em gotas de água suspensas na atmosfera dando origem ao fenómeno do arco-íris. O observador vê raios de cores diferentes provenientes de locais diferentes do espaço.**

Considerando agora planos oblíquos de visão do observador, haverá lugar à alteração do ângulo de visão segundo o qual o observador vê as gotas que lhe parecem azuis e as que lhe parecem vermelhas. O resultado é o aparecimento de um arco em que os tons azuis se encontram colocados no interior.

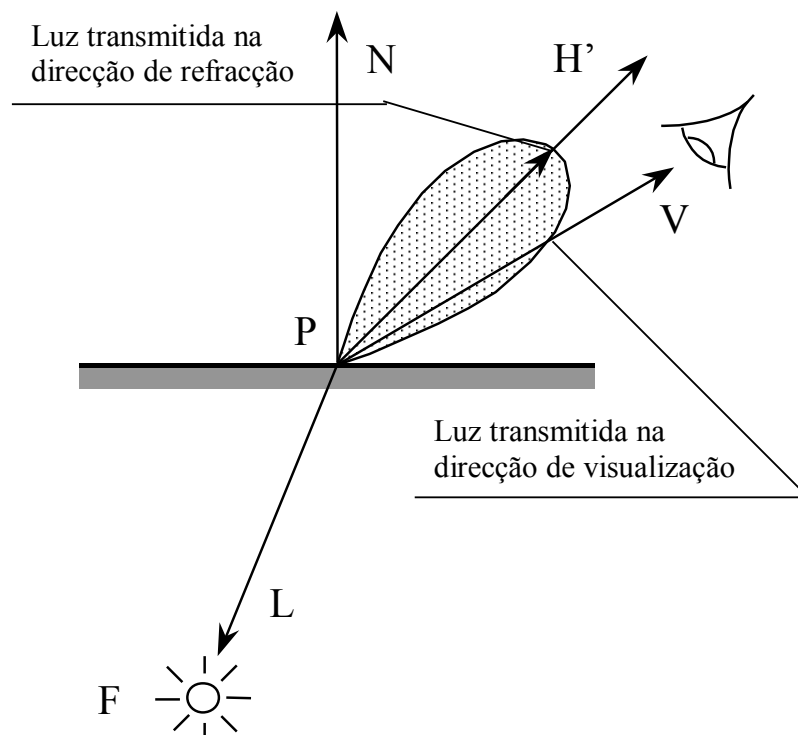
## 4.6 Extensão do Modelo de Iluminação de Phong

O modelo de iluminação de Phong não inclui a contribuição da luz transmitida através de objectos transparentes ou translúcidos para a iluminação. É necessário estender este modelo para que possa ser empregue pelos algoritmos de ray tracing, adicionando um termo correspondente à luz assim transmitida.

A dedução deste termo adicional faz apelo à noção de micro facetas virtuais do modelo físico de iluminação de Torrance-Sparrow. Considere-se o ponto P da figura 4.8 situado sobre a superfície de separação entre dois meios de densidades ópticas diferentes. A normal a esta superfície no ponto P é  $\vec{N}$ . O ponto recebe luz proveniente da fonte de luz F que, ao atingir a superfície é refractada segundo  $\vec{H}'$ . Como a direcção de visualização,  $\vec{V}$ , não coincide com  $\vec{H}'$ , há então que calcular a iluminação de P percebida segundo a direcção de visualização empregando um processo semelhante à aproximação de Blinn para a reflexão especular. A contribuição da luz refractada para a iluminação de P será então

$$I_{transmitida} K_t (\vec{N} \cdot \vec{H}')^n \quad (4.23)$$

em que  $K_t$  é o coeficiente de transmissividade.



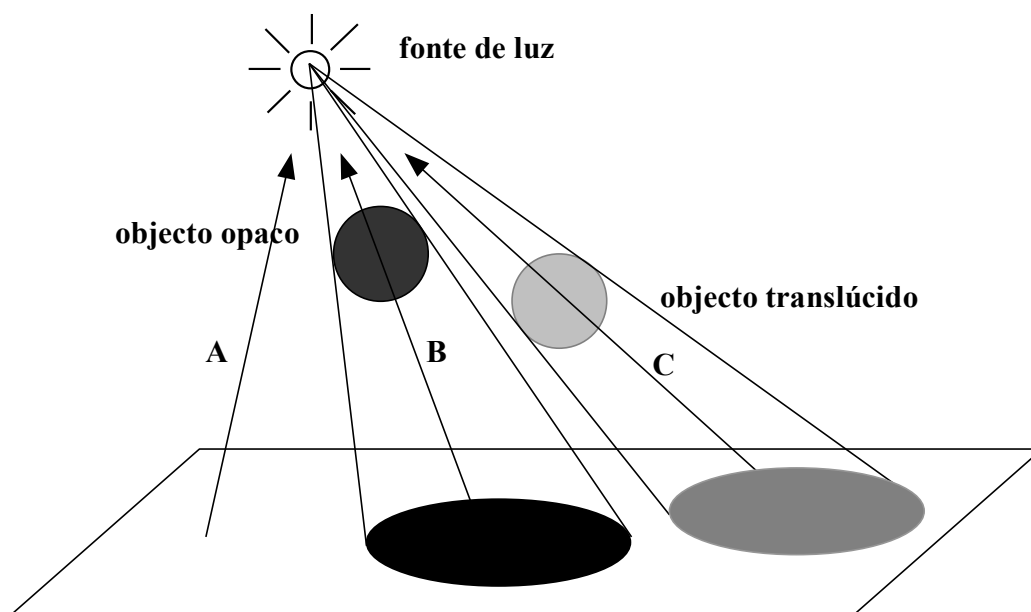
**Figura 4.8 – Refracção difusa.**

## 4.7 Iluminação Directa e Determinação de Sombras

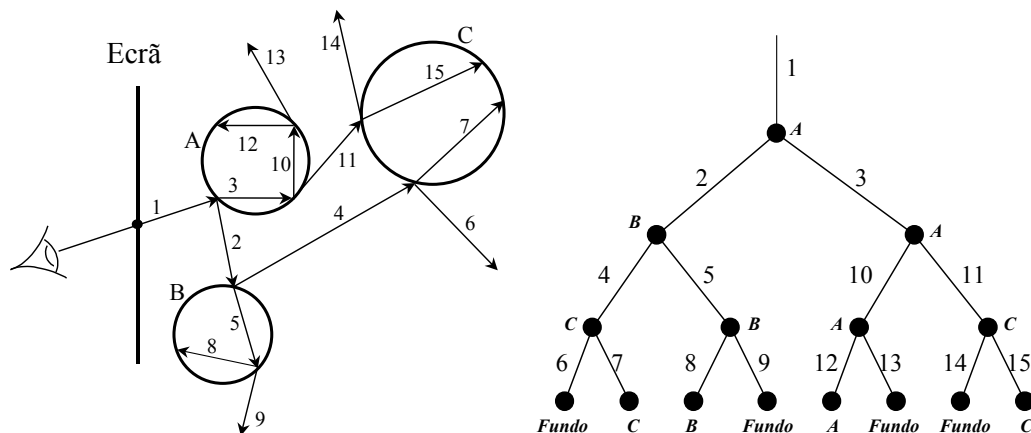
A iluminação do ponto de intersecção de um raio primário com um objecto da cena a representar pode ser directa se o objecto for directamente iluminado por uma ou mais das fontes de luz existentes na cena. Para verificar se esta condição ocorre é necessário determinar se o ponto de intersecção se encontra em linha directa de visão com cada uma das fontes de luz. Se tal acontecer, o ponto será directamente iluminado pela fonte de luz e a respectiva contribuição para a iluminação do ponto deve então ser calculada. Caso contrário, o objecto encontrar-se-á na sombra em relação a essa fonte de luz.

Para determinar se um ponto de um objecto se encontra ou não em linha directa de visão com uma fonte de luz, empregam-se raios secundários do tipo raios de iluminação (ou de sombra). Estes raios são também correntemente denominados por “shadow feelers” em língua inglesa, dado que permitem verificar a ocorrência de sombras.

Os raios de iluminação (ou de sombra) unem o ponto de intersecção de um raio com um objecto com cada uma das fontes de luz existentes na cena, tal como a figura 4.9 apresenta. Se não existir nenhum objecto no caminho do raio antes de este atingir a fonte de luz, como é o caso do raio A da figura, a fonte de luz ilumina directamente o ponto de origem do raio secundário. Nos casos dos raios B e C da figura, esses raios intersectam objectos localizados entre as suas origens e a fonte de luz com resultados diferentes. No caso do raio B o objecto intersectado é opaco e, conseqüentemente, a fonte de luz não contribui para a iluminação do ponto de origem do raio B, que estará à sombra da fonte de luz. No caso do objecto C estamos em presença de um objecto transparente ou translúcido. Isto significa que a fonte de luz ilumina directamente o ponto de origem do raio, mas que a intensidade de iluminação é atenuada. Daqui resulta uma penumbra e não uma sombra no local de origem do raio C.



**Figura 4.9 – Raios secundários de iluminação (shadow feelers). O raio A ilumina o seu ponto de origem (não existe qualquer objecto no caminho do raio), a origem do raio B está na sombra da fonte de luz (o objecto intersectado é opaco) e a origem do raio C está numa penumbra (objecto translúcido intersectado).**



**Figura 4.10 – Recursividade em ray tracing. Os raios considerados (à esquerda) encontram-se organizados numa estrutura de árvore (à direita).**

## 5 O Algoritmo Completo

O algoritmo básico de ray tracing apresentado no início encontra-se incompleto porque se limita a atribuir a cada pixel da imagem a cor do ponto de intersecção do raio primário correspondente ao pixel com o objecto mais próximo da origem do raio, sem atender a outras contribuições para a iluminação desse ponto de intersecção que não seja a sua iluminação directa. Para ter em conta todas as outras contribuições para a iluminação é necessário inspecionar recursivamente a cena através da geração de raios secundários.

A figura 4.10 apresenta uma cena onde existem três esferas semi-transparentes que reflectem e refractam a luz. O raio primário (1) intersecta a esfera A num ponto em que existe uma componente de iluminação devida à refacção da luz proveniente da própria esfera e outra componente devida à luz reflectida pela esfera B na direcção da esfera A. Geram-se assim os raios secundários 3 e 2, respectivamente. No ponto em que o raio secundário 2 atinge a esfera B há que ter em conta novamente duas componentes para a iluminação: luz reflectida por B e que provém da esfera C (raio 4) e luz refractada pela própria esfera B (raio 5). Este último raio intersecta a esfera B num ponto em que o raio refractado inverso (raio 9) sai da cena e, portanto, contribui para a iluminação com a cor de fundo. O raio 8 (reflexão no interior da esfera) corresponde por sua vez a um quarto nível de recursividade e, como este foi o limite imposto à partida, podemos concluir que este raio não contribui para a iluminação do ponto de intersecção original.

Esta pesquisa recursiva constrói a estrutura de árvore dos raios secundários que a figura 4.10 apresenta. Esta estrutura deve ser explorada recursivamente até ao quarto nível de recursividade, o nível máximo arbitrado. Estamos agora na situação em que podemos completar o algoritmo traçando raios secundários sucessivamente até que estes saiam da cena ou se atinja o nível máximo permitido.

Este cálculo pode ser realizado de acordo com o procedimento recursivo que é seguidamente apresentado.

```

Procedure TraceRay( pontoOrigem, direccao, nivel, cor)
{ var pontoInterseccao, direccaoReflexao, direccaoTransmissao,
  corLocal, corReflexao, corTransmitida;
  if ( nivel > nivelMaximo )
    then cor = negro;
  else
    { pontoInterseccao = (intersecção mais próxima de pontoOrigem);
      if ( !intersecção )
        then cor = corFundo;
      else
        { corLocal=(contribuição da iluminação directa);
          (calcular direccaoReflexao);
          TraceRay( pontoInterseccao, direccaoReflexao, nivel+1,
                    corReflexao);
          (calcular direccaoTransmissao);
          TraceRay( pontoInterseccao, direccaoTransmissao, nivel+1,
                    corTransmitida);
          (combinar( cor, corLocal, corReflexao, pesoReflexao,
                    corTransmitida, pesoTransmissao));
        }
    }
}

```

O algoritmo recursivo apresentado parte do nível zero com o valor negro atribuído à cor e com a direcção e origem do raio primário correspondente ao pixel cuja cor se pretende determinar. Se o raio não intersectar qualquer objecto, então a cor do pixel será a cor de fundo da cena e o processo recursivo terminará. Caso contrário, determina-se a cor local devida à iluminação directa através de raios de iluminação (ou de sombra). Seguidamente determinam-se os raios secundários correspondentes às direcções cujos raios reflectido e transmitido coincidem com a direcção do raio corrente e calculam-se as respectivas contribuições para a cor do ponto corrente que, devidamente ponderadas, serão somadas com a cor local à cor anteriormente determinada para o raio que está a ser calculado.

Este procedimento apresenta no entanto um inconveniente que consiste em realizar os cálculos dos raios secundários de reflexão e transmissão, e nas respectivas chamadas recursivas ao procedimento, sem previamente testar se o nível de recursividade corrente atingiu o nível máximo arbitrado. Daqui resulta uma carga computacional adicional que pode ser evitada. Para tal basta agrupar num bloco a determinação destes raios e as chamadas recursivas ao procedimento. Este bloco só será executado se o nível de recursividade corrente for menor ou igual ao nível de recursividade máximo arbitrado. Caso contrário, atribuir-se-á a cor negra às contribuições correspondentes à reflexão e à transmissão. Esta modificação é implementada pelo procedimento seguinte

```

Procedure TraceRay( pontoOrigem, direccao, nivel, cor)
{
  var pontoInterseccao, direccaoReflexao, direccaoTransmissao,
    corLocal, corReflexao, corTransmitida;
  pontoInterseccao = (intersecção mais próxima de pontoOrigem);
  if ( !intersecção )
    then cor = corFundo;
  else
    { corLocal=(contribuição da iluminação directa);
      if ( nivel != nivelMaximo )
        { (calcular direccaoReflexao);
          TraceRay( pontoInterseccao, direccaoReflexao, nivel+1,
                    corReflexao);
          (calcular direccaoTransmissao);
          TraceRay( pontoInterseccao, direccaoTransmissao, nivel+1,
                    corTransmitida);
        }
      else
        { corReflexao = negro;
          corTransmitida = negro;
        }
      (combinar( cor, corLocal, corReflexao, pesoReflexao,
                corTransmitida, pesoTransmissao));
    }
}

```

Tal como o anterior, este procedimento não contempla a verificação da importância relativa da contribuição da cor local em relação à cor do ponto de intersecção do raio primário original. Se esta contribuição for inferior a um dado valor é perfeitamente possível terminar o processo recursivo, pois a cor local em pouco ou nada contribuirá para a cor do pixel que queremos determinar. A modificação para incluir este critério de paragem consiste em alterar o teste de paragem e na adição de um argumento ao procedimento. Este novo argumento contém o produto dos pesos sucessivos. Assim, o teste de paragem, além de comparar o nível de recursividade corrente com o nível de recursividade máximo arbitrado, verificará também se o peso a atribuir à cor do ponto corrente é inferior ao valor arbitrado.

## 6 Estratégias de Optimização

A carga computacional da geração de imagens pelo algoritmo de ray tracing é bastante pesada devido ao cálculo das intersecções dos raios com os objectos da cena. Uma imagem de 1024 x 1024 píxeis de uma cena que contém 100 objectos implica, no pior dos casos, a determinação de 104 857 600 intersecções. Este número é demasiado elevado e, portanto, deverão ser empregues estratégias que permitam reduzi-lo e, assim, reduzir a carga computacional. Para atingir este objectivo existem quatro tipos de solução:

São elas:

- Transformação da cena segundo o raio
- Emprego de volumes envolventes
- Hierarquização do espaço
- Partição do espaço

## 6.1 Transformação da Cena Segundo o Raio

A localização espacial das intersecções de um raio com os objectos de uma cena só é conhecida depois de essas intersecções terem sido calculadas e, portanto, nada se pode afirmar à partida sobre a sua posição. Porém, se o raio coincidir com o eixo dos  $ZZ$  da cena, sabemos que todas as intersecções corresponderão a pontos para os quais  $X = 0$  e  $Y = 0$ . Poderemos, então, fazer uso deste conhecimento, não procurando intersecções do raio com todos os objectos que não possuam pontos com valor nulo das coordenadas  $X$  e  $Y$ .

A optimização do cálculo de intersecções por transformação da cena emprega esta propriedade. A cena é transformada para cada raio por forma a que o eixo dos  $ZZ$  coincida com o raio<sup>11</sup> e determinam-se os pontos de intersecção. À intersecção de menor valor de  $Z$  aplica-se então a transformação inversa. Esta estratégia de optimização é facilitada se os objectos se encontrarem ordenados por profundidade crescente na cena, uma vez que interessa determinar a intersecção mais próxima da origem do raio.

Esta estratégia apresenta no entanto a desvantagem computacional devida à necessidade de aplicar uma transformação específica para cada raio, e de cada transformação ter que ser aplicada a todos os objectos da cena.

## 6.2 Intersecção com Volumes Envolventes

Embora haja a tendência para modelar os objectos de uma cena com base em formas geométricas simples (esferas, paralelepípedos, poliedros convexos, etc.), existem muitos objectos cuja forma não pode ser tão simplesmente descrita. Consequentemente, a determinação de intersecções torna-se complexa, tal como sucede com um objecto descrito através de uma malha poligonal. Neste caso é necessário determinar se o raio intersecta o objecto, testando a intersecção do raio com cada um dos polígonos da malha que o descreve. Na prática, a maioria dos testes de intersecção termina imediatamente porque não existe intersecção entre o raio e o objecto, a não ser que existam objectos de grandes dimensões ocupando quase totalmente a cena.

Uma estratégia de optimização deve poder determinar rapidamente se existe ou não possibilidade de intersecção entre um raio e um objecto. Em vez de realizar o teste com um objecto cuja descrição é complexa, podemos testar a existência de intersecção entre o raio e o volume envolvente do objecto cuja intersecção com o raio é mais fácil e rapidamente determinada. Se o resultado for afirmativo, existe grande probabilidade de existência de intersecção entre o raio e o objecto e, então, deverão ser realizados cálculos mais detalhados de intersecção agora com os elementos descritores do objecto.

---

<sup>11</sup> O cálculo da matriz desta transformação é idêntico ao cálculo da matriz de rotação da cena segundo as direcções dos eixos da câmara virtual.



Caso contrário, poderemos passar ao objecto seguinte da lista de objectos da cena e teremos poupado cálculos demorados que iriam contribuir para tornar o processamento da imagem muito mais demorado.

Para ilustrar a importância de que se reveste a estratégia de optimização por volumes envolventes, tomemos como exemplo uma cena onde existe um único objecto descrito através de uma malha poligonal contendo 46 polígonos. Se a dimensão da imagem a calcular for de 128 por 128 píxeis, ou seja a imagem é constituída por  $128 \times 128 = 16.384$  píxeis, será necessário determinar  $16.384 \times 46 = 753.664$  intersecções quando não se recorrer a volumes envolventes. Admitamos agora que apenas 1/8 dos 753.664 testes de intersecção realizados revelam a existência de intersecção entre os 16.384 raios primários e os 46 polígonos do objecto, e consideremos um volume envolvente que englobe todos estes 46 polígonos. O número de cálculos de intersecção dos raios primários com o volume envolvente passará a ser de apenas 16.384. Para 1/8 destes raios, que efectivamente intersectam o volume envolvente, há então que testar a sua intersecção com cada um dos 46 polígonos, ou seja, há que realizar mais  $16.384 \times 1/8 \times 46 = 94.208$  testes de intersecção. O emprego de volumes envolventes reduziu assim o volume dos cálculos de intersecção de 753.664 para 110.592 ( $16.384 + 94.208$ ) e permitiu que a imagem pudesse ser sintetizada cerca de 6,8 vezes mais depressa.

A escolha do volume envolvente é um factor importante para o sucesso desta estratégia de optimização. O volume envolvente deve ser tal que o objecto o preencha o mais completamente possível, para que não ocorram muitas situações em que, existindo intersecção entre o raio e o volume envolvente do objecto, não exista intersecção do raio com o objecto envolvido pelo volume. Assim, não deve ser escolhida uma esfera para volume envolvente de um objecto de forma achatada pois muitos dos pontos pertencentes à esfera não pertencerão ao objecto. Um volume envolvente paralelepípedo constituirá uma muito melhor escolha para este caso.

Uma vez determinada a existência de intersecção entre o raio e o volume envolvente de um objecto<sup>12</sup>, há então que determinar os pontos de intersecção entre o raio e o objecto. Esta determinação depende do tipo de representação geométrica que descreve o objecto. O caso mais comum é o de representação da fronteira através de uma malha poligonal. A determinação de intersecções passa por determinar a intersecção do raio com cada um dos polígonos da malha. Este exemplo demonstra de forma clara a vantagem do emprego de volumes envolventes. Como na maioria dos casos não há intersecção entre o raio e o objecto, esta estratégia permite poupar a determinação da intersecção do raio com cada um dos polígonos da malha representativa da fronteira do objecto.

### 6.3 Hierarquização do Espaço

A estratégia empregando volumes envolventes dos objectos permite diminuir a carga computacional de cálculo de intersecção entre um raio e os objectos, pois elimina imediatamente o objecto contido num volume envolvente se este volume não for intersectado. No entanto, é necessário determinar se existe ou não intersecção entre o raio e cada um dos volumes envolventes.

Os volumes envolventes podem ser organizados hierarquicamente, construindo uma árvore onde os volumes envolventes de cada objecto singular correspondem às folhas de

---

<sup>12</sup> Não interessa em que pontos, mas sim unicamente determinar se há ou não intersecção.

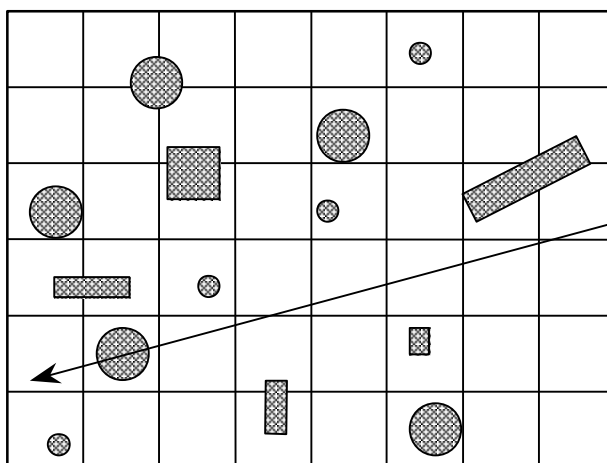
cada ramo e em que cada nó intermédio representa um volume que envolve todos os volumes situados nos ramos que dele dependem. Se num dado nó existir intersecção entre o respectivo volume envolvente e o raio, dever-se-á então inspeccionar todos os ramos com origem nesse nó. Mas, o que é o caso mais provável, se não existir intersecção com o volume envolvente a esse nível, então não existirão quaisquer intersecções do raio com os volumes correspondentes aos ramos com origem nesse nó e a determinação de intersecções pode ser imediatamente terminada para o ramo corrente.

Esta estratégia permite diminuir consideravelmente o número de volumes envolventes para os quais é realizada a determinação de intersecções. Persiste no entanto o problema da construção da hierarquia, embora se possam considerar hierarquias de volumes envolventes com base na distância crescente dos volumes à origem do raio. Esta condição é fácil de satisfazer para os raios primários, uma vez que todos os raios primários têm origem no mesmo ponto, o centro de projecção, mas deixa de o ser para os raios secundários que podem ter origem em um qualquer ponto da cena.

## 6.4 Partição do Espaço

Uma hierarquia de volumes envolventes como a anteriormente descrita organiza os volumes envolventes segundo uma lógica ascendente (de baixo para cima). A lógica descendente alternativa (de cima para baixo) consiste em dividir o espaço da cena em volumes envolventes paralelepípedicos dispostos segundo uma grelha tridimensional, tal como a figura 6.1 apresenta. Cada partição do espaço assim obtida poderá ou não conter objectos, não sendo necessário que cada objecto esteja totalmente contido numa única partição.

Dado um raio, esta estratégia consiste então em determinar a intersecção do raio com as partições seguindo uma ordem de teste por distância crescente das partições à origem do raio.



**Figura 6.1 – Partição do espaço da cena. O espaço é dividido em partições iguais que desempenham o papel de volumes envolventes. Um objecto pode localizar-se em mais do que uma partição.**

Uma partição intersectada pelo raio é liminarmente eliminada se não contiver quaisquer objectos. Caso contrário, calculam-se então as intersecções do raio com os objectos contidos na partição para, no final, reter a intersecção correspondente à menor distância

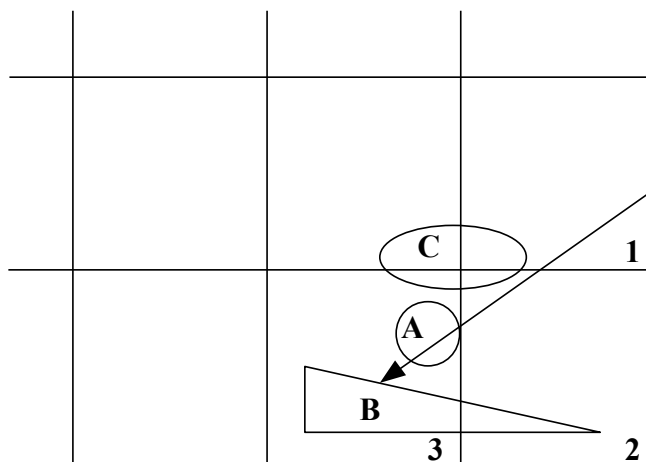
à origem do raio, aplicando a cada partição uma lógica em tudo semelhante à aplicada a toda a cena quando se não considera a sua partição.

No caso em que um objecto se encontra contido em mais do que uma partição do espaço é necessário verificar se a intersecção determinada ocorre na partição corrente e que deu origem ao desencadear dos cálculos de intersecção do raio com os objectos contidos na partição. A razão desta verificação adicional é que a intersecção com o objecto poderá ocorrer numa outra partição que não a partição para a qual se realizaram os cálculos de intersecção. Igualmente, poderá existir ainda uma terceira partição que apresente uma intersecção do raio com outro objecto que seja mais próxima da origem do raio. Neste caso é necessário continuar a examinar as partições até que se tenha a certeza de qual é a intersecção mais próxima da origem do raio. Esta situação encontra-se representada na figura 6.2 onde o algoritmo detecta uma intersecção do raio com o objecto B na partição 2. Este objecto estende-se pelas partições 2 e 3 pelo que há que continuar a testar novas partições. Ao ser testada a partição 3, verifica-se que o raio intersecta o objecto A nesta partição. O objecto A está totalmente contido na partição 3 e, portanto, a intersecção agora determinada é segura. Como esta intersecção corresponde a um ponto mais próximo da origem do raio do que a intersecção do raio com o objecto B anteriormente encontrada, a intersecção que procurávamos é a intersecção do raio com o objecto A e não a com o objecto B.

A situação anteriormente descrita presta-se ainda a uma observação importante relativa ao cálculo da intersecção do raio com o objecto B. Este cálculo foi realizado quando a partição 2 foi examinada e não faz sentido voltar a realizá-lo quando se examinar a partição 3. Para isto é necessário associar ao objecto B uma estrutura com a informação das intersecções do objecto com o raio já calculadas. Esta técnica, redescoberta inúmeras vezes, tem o nome de mailbox, ray caching, intersection caching, ray tagging ou hit memory, consoante os autores. A informação guardada deve indicar tanto a existência de intersecções como a sua não existência. Com efeito, não faz sentido voltar a verificar se existe intersecção entre o raio e o objecto C da figura nas partições 2 e 3 se na partição 1 foi já determinado que o raio não intersecta este objecto.

A estratégia de cálculo de intersecções por partição do espaço da cena pode também ser adaptativa, recorrendo à divisão de partições por meio de árvores de octantes. Este refinamento apresenta vantagem sempre que exista um número elevado de objectos numa dada partição. O seu emprego permite eliminar os octantes que não intersectem o raio, diminuindo assim a complexidade do problema. Os octantes não eliminados deverão então ser inspeccionados por ordem crescente da sua distância à origem do raio.

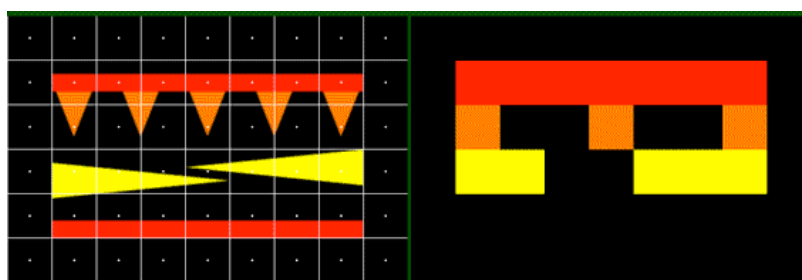
Uma alternativa à divisão em octantes de cada partição do espaço consiste em combinar a estratégia de partição do espaço com a estratégia de hierarquização de volumes envolventes, agora aplicada a cada partição. Com efeito, cada uma das partições poderá desempenhar o papel de volume envolvente pai de uma hierarquia de volumes envolventes dos objectos localizados na partição.



**Figura 6.2 – A intersecção de um raio com um objecto existente em mais do que uma partição pode não corresponder à intersecção mais próxima da origem do raio.**

## 7 Antialiasing em Ray Tracing

Como vimos anteriormente, a precisão dos algoritmos de ray tracing depende da resolução da quadrícula empregue. Esta, como se sabe, encontra-se no espaço de imagem. Por este motivo, os algoritmos de ray tracing estão sujeitos ao aparecimento de fenómenos em que detalhes de pequena dimensão podem não ser capturados na imagem ou surgir agrupados em estruturas que, na realidade, não existem. A figura 7.1 apresenta um exemplo de tais fenómenos, denominados de *aliasing*. A adopção de técnicas de *antialiasing* do tipo de pré-filtragem permite reduzir o impacto destes fenómenos dado que os algoritmos de ray tracing exploram o espaço dos objectos e, portanto, têm a possibilidade de obter mais informação sobre um objecto na área centrada no ponto de intersecção de um raio primário com o objecto.<sup>13</sup>



**Figura 7.1 – *Aliasing* devido à atribuição a cada pixel da cor do ponto de intersecção do raio primário com o objecto sem atender à área que o circunda.**

Em ray tracing, a cor de um pixel é determinada calculando a cor no ponto de intersecção do raio primário correspondente ao pixel com o objecto intersectado no

<sup>13</sup> Os fenómenos de *aliasing*, suas origens e as técnicas de *antialiasing* são abordados em profundidade no opúsculo de Rasterização.

espaço do objecto. Em lugar de se empregar apenas um raio primário por cada pixel, poder-se-ão empregar vários raios primários cobrindo uniformemente a área correspondente ao pixel. A cor do pixel será, então, a média das cores determinadas por cada um dos raios, o que equivale a realizar uma pré-filtragem por sobre amostragem.

Com mais do que um raio primário por cada pixel podemos obter uma ideia mais precisa do que acontece na área correspondente ao pixel. No entanto, a sobre amostragem assim realizada não faz desaparecer todos os fenómenos de aliasing que possam ocorrer, pois ao aumentar o limite imposto pelo teorema de Nyquist estaremos apenas a diminuir os seus efeitos. O emprego de raios primários uniformemente distribuídos, independentemente do que possa suceder nas áreas sobre amostradas, pode corresponder ao emprego de sobre amostragem em áreas de que dela não necessitam, realizando cálculos desnecessários e, o que é mais grave, não empregando uma amostragem adequada nas zonas em que tal seja necessário. Um factor a ter em conta é a carga computacional da sobre amostragem, uma vez que esta é directamente proporcional ao número de raios primários empregue.

## 7.1 Amostragem Adaptativa

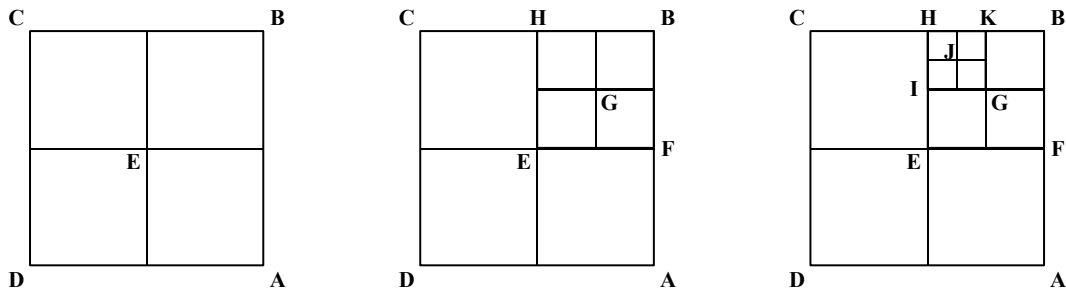
Uma forma mais racional de empregar a sobre amostragem é concentrar mais raios nas áreas em que estes são efectivamente necessários em vez de os distribuir uniforme e cegamente. A técnica de *amostragem adaptativa* emprega inicialmente menos raios primários por cada área, normalmente em número de cinco, em que um deles se localiza no centro da área e os outros quatro nos quatro cantos dessa área<sup>14</sup>. Se as cores determinadas por cada um dos cinco raios forem semelhantes, poderemos concluir que os raios intersectam o mesmo objecto e empregar a cor média para cor do pixel da imagem correspondente a essa área.

Se as cores determinadas diferirem mais do que um dado valor limite, proceder-se-á então à subdivisão da área em causa em quatro áreas mais pequenas e repetir-se-á a emissão de raios em cada uma das subáreas em que se verifiquem diferenças significativas de cor. Em cada uma delas (veja-se a figura 7.2) só é necessário calcular três novos raios porque as cores dos outros dois raios são já conhecidas. Esta subdivisão pode ser aplicada recursivamente até que as diferenças de cor deixem de ser significativas ou se atinja um nível máximo de subdivisões pré-determinado.

A determinação da cor de cada pixel é então feita através do cálculo da média dos valores das cores determinadas pelos cinco raios de cada área ou subárea. Se não houver divisão de uma área (veja-se a figura 7.2 à esquerda), a cor do pixel será

$$\frac{1}{4} \left( \frac{A+E}{2} + \frac{B+E}{2} + \frac{C+E}{2} + \frac{D+E}{2} \right) = \frac{1}{4} \left[ \frac{1}{2} (A+B+C+D) + 2E \right] \quad (7.1)$$

<sup>14</sup> Esta disposição permite a reutilização dos resultados dos cálculos destes quatro raios no cálculo das áreas adjacentes.



**Figura 7.2 – Sobre amostragem adaptativa em ray tracing. A cor do pixel é o resultado da média das cores dos vários raios calculados com aumento do número de raios empregue nas zonas onde existam maiores diferenças de cor entre raios.**

Para o caso representado ao centro da mesma figura, a cor da área passaria a ser

$$\frac{1}{4} \left[ \frac{A+E}{2} + \frac{C+E}{2} + \frac{D+E}{2} + \frac{1}{4} \left( \frac{E+G}{2} + \frac{F+G}{2} + \frac{B+G}{2} + \frac{H+G}{2} \right) \right] \quad (7.2)$$

onde existem apenas mais três pontos, os pontos F, G e H. Finalmente, para a situação apresentada à direita da figura, teríamos

$$\frac{1}{4} \left\{ \frac{A+E}{2} + \frac{C+E}{2} + \frac{D+E}{2} + \frac{1}{4} \left[ \frac{E+G}{2} + \frac{F+G}{2} + \frac{B+G}{2} + \frac{1}{4} \left( \frac{G+J}{2} + \frac{K+J}{2} + \frac{H+J}{2} + \frac{I+J}{2} \right) \right] \right\} \quad (7.3)$$

Um problema desta técnica reside em arbitrar quando é que existe ou não uma diferença significativa entre cores. Por outro lado, não é óbvio que a amostragem seja significativa apenas porque as cores dos cinco raios de uma área são idênticas ou semelhantes, pois pode acontecer que exista na área um objecto de pequena dimensão que não seja detectado por qualquer dos cinco raios. A causa deste problema está na natureza regular da malha definida pelos cinco raios.

## 7.2 Amostragem Estocástica

Uma forma de resolver o problema anterior consiste em distribuir aleatoriamente os raios pela área, mantendo o número de raios originalmente gerados por cada pixel, tendo o cuidado de distribuí-los de maneira uniforme. Desaparece assim a grelha regular e com ela desaparecem também os efeitos de aliasing devidos à amostragem regular. Esta técnica, denominada de “*ray tracing aleatório*”, “*ray tracing estocástico*” (*stochastic ray tracing*) ou “*ray tracing distribuído*” (*distributed ray tracing*), permite calcular cores médias mais correctas, pois os raios encontram-se melhor distribuídos dentro da área correspondente a cada pixel, mas introduz ruído devido à distribuição aleatória. No entanto, este ruído encontra-se mais ou menos uniformemente distribuído pela imagem e, como a visão humana é muito mais tolerante ao ruído aleatório do que a singularidades locais, esta técnica produz resultados de melhor qualidade visual, principalmente em zonas de sombra parcial (penumbra) onde não existem arestas vivas.

Persistem, no entanto, os problemas associados ao número de raios a calcular e à correspondente carga computacional. Com efeito, poderão existir píxeis em que a cor é uniforme e que dispensam o cálculo de um número de raios maior, pois dois ou três raios são suficientes. Por outro lado, podem existir píxeis para os quais nove raios primários são ainda insuficientes, como é, por exemplo, o caso de um pixel a que corresponda uma área de padrão axadrezado que contenha 16 cores diferentes. Os nove raios detectariam cores com diferenças estatisticamente significativas entre si<sup>15</sup>. A solução consiste no aumento sucessivo do número de raios primários até que, estatisticamente, se possa determinar que a cor calculada é a cor correcta com uma dada probabilidade de certeza.

---

<sup>15</sup> Tal como para a sobreamostragem adaptativa, existe o problema de arbitrar o que é uma diferença estatisticamente significativa.





existem na cena 5 corpos esféricos em que a luz pode ser reflectida e refractada. Considere um nível máximo de profundidade igual a 4.

- 11 Numa imagem de 128x128 píxeis, sabe-se que apenas 1/8 dos raios primários intersectam o único objecto existente na cena que tem a forma esférica e é constituído por 46 polígonos regulares. Calcule o número de intersecções que deverá calcular sem empregar um volume envolvente esférico e o número de intersecções que calcularia empregando um volume envolvente esférico, apresentando as razões para a diferença dos valores que encontrar.
- 12 A estratégia de optimização por partição do espaço em ray tracing permite que um objecto se localize em mais do que uma partição do espaço. Apresente uma modificação desta estratégia que faça com que o cálculo da intersecção de um raio com um destes objectos se realize apenas na primeira partição atravessada pelo raio em que se detecta que há que determinar a intersecção do raio com o objecto.