

Tema III

Síntesis de Imágenes: ray casting básico

Ricardo Ramos

Colaboradores: Enrique Sierra Blanco, Eloy Álvarez Fernández, Reyes Berguillos Moretón, M^a del Carmen Suárez Torrente, M^a Sandra García Peláez, Francisco J. Barroso Alfambra y M^a Félix de la Corte Martínez

Comienza el estudio la Síntesis de Imágenes viendo el algoritmo de ray casting (síntesis directa y un modelo local de iluminación) que representa la versión más sencilla de los procesos que buscan la información gráfica directamente en el modelo vectorial 3D.

1.1 Introducción

Según vimos en el primer tema, el algoritmo de **ray casting** se basa en un método directo de búsqueda de la información gráfica, y en un modelo de iluminación local. Puesto que por el momento sólo veremos una versión sencilla de este algoritmo, el modelo de iluminación que analizaremos será un *modelo local simplificado*. Los modelos simplificados no sólo son útiles el los renderizados simples, sino que además sirven como base para el estudio de modelos más sofisticados, circunstancia que aprovecharemos aquí.

1.2 Modelo de cámara oscura

El modelo de **cámara oscura**, en el cual se fundamentan las cámaras fotográficas, *sirve también como base para los métodos de búsqueda directa* de la información gráfica utilizados en la Síntesis de Imágenes.

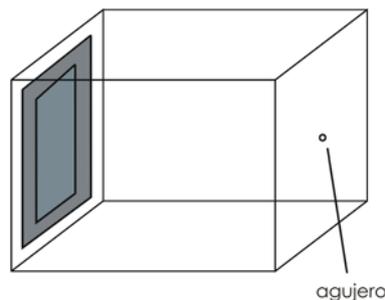


figura 1: modelo de cámara oscura

Una cámara oscura consiste en una caja completamente cerrada, a excepción de un pequeño orificio en el centro de una de sus caras (figura 1).

Dichas cámaras ya eran conocidas y utilizadas antes de la aparición de la fotografía. La luz entra por el agujero y se proyecta en la cara opuesta, formando así una imagen invertida (figura 2).

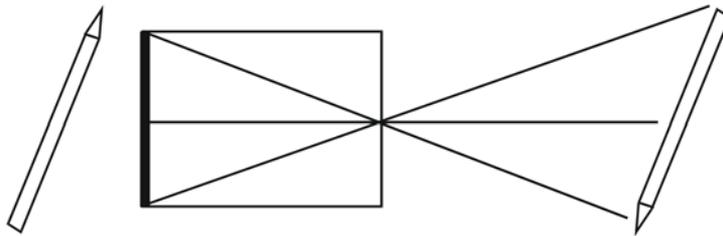


figura 2: la cámara oscura invierte la imagen proyectada

Este método de obtención de imágenes puede ser utilizado en la síntesis directa, aunque es conveniente aplicar la transformación que veremos a continuación para facilitar la implantación del modelo.

1.3 Modelo de Ventana

En el **modelo de ventana** (o **de marco**), lo que antes era el orificio de la cámara pasa a ser el ojo del observador, el cual se sitúa detrás del plano visual o pantalla. De esta forma se evita que la imagen que se forma en la pantalla quede invertida.

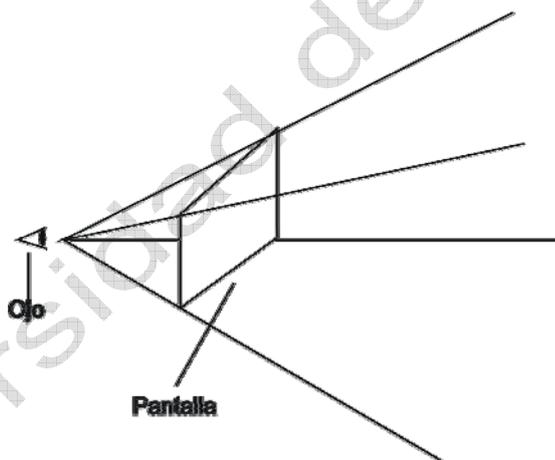


figura 3: modelo de ventana

Ver que los componentes de este nuevo modelo son los mismos que los que intervienen en la cámara oscura, sólo que posicionados en un orden diferente. El requerimiento anterior de que todos los rayos pasaran por el orificio, se traduce ahora en que *todos los rayos de luz que atraviesen la pantalla han de converger en el punto de visión (ojo)*.

Si se traza un rayo desde el punto de visión a cada píxel de la pantalla, y se prolongan dichos rayos hasta que intersequen con los objetos que se encuentren en su camino, entonces es posible averiguar con qué objetos intersecan y el color de las superficies donde intersecan. Si se pinta los píxeles correspondiente con el color de los puntos de intersección, se obtiene la imagen del objeto que se trata de visualizar (figura 4). Los objetos visualizados son aquellos que se encuentran dentro del volumen de la pirámide, cuyas aristas pasan por los vértices de la ventana, y convergen en el punto de vi-

sión, tal como muestra la figura 3.

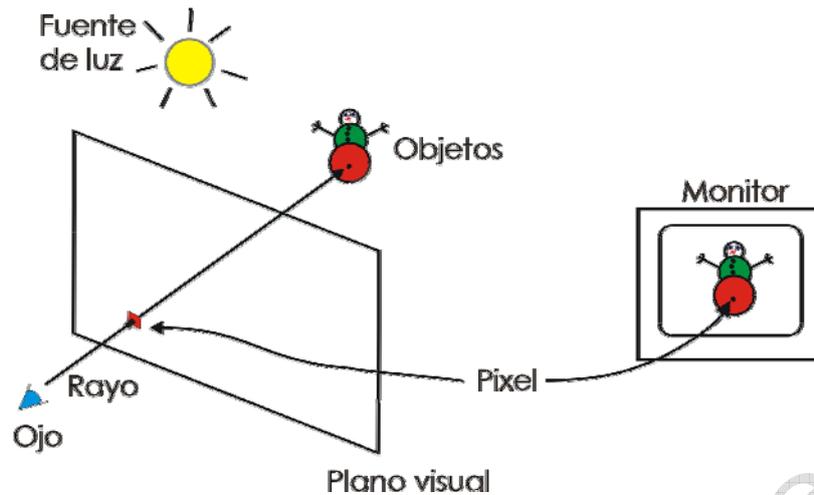


figura 4: trazado de los píxeles

En definitiva, el principio en el que se basa el algoritmo de ray casting es el de trazar una línea entre el punto de visión y el espacio de objetos (modelos), a través de cada píxel de la pantalla.

1.4 Trazado hacia adelante y hacia atrás

Según indicamos en el tema anterior, los sistemas gráficos basados en la síntesis directa emulan el comportamiento de los fotones de luz. Como sabemos, éstos salen del emisor (lámpara) y viajan por el espacio hasta que colisionan con la materia (objetos). Cuando esto ocurre, una gran parte de los fotones son filtrados por ésta, impidiendo el paso de unos (absorción) y dejando a otros pasar (transmisión) o escapar (reflexión), modificando su trayectoria en la mayoría de los casos. Las transformaciones que sufre la luz al incidir sobre materiales diferentes, se traduce en información visual que somos capaces de procesar cuando llega a nosotros.

Entonces, cuando el trazado de los rayos se implanta de forma que emule el viaje natural de los fotones, se trata del **trazado hacia adelante**. El trazado de rayos hacia adelante es teóricamente un método válido. Ocurre, sin embargo, que muchos fotones que salen de la lámpara nunca llegan al observador, es decir, que no juegan papel alguno en la construcción de la imagen. Dado que computacionalmente es muy caro trazar rayos que no sirven de nada, es necesario encontrar una solución a este problema. *La clave para evitar el coste que representa el cálculo inútil está en invertir el recorrido de los fotones.*

En el modelo de ventana, si un fotón llega al ojo del observador, necesariamente ha tenido que cruzar la pantalla siguiendo la dirección del rayo. Entonces, si se trazan exclusivamente los rayos de luz que parten del punto de visión, a través de los píxeles de la pantalla hacia el espacio vectorial (conocidos como **rayos primarios**), nos aseguramos de que el sistema gráfico trabaje únicamente con los rayos útiles. Cuando el trazado de rayos se realiza de esta forma, se trata del **trazado hacia atrás**.

Dado el ahorro de cálculos que implica el invertir el recorrido de los fotones, *el trazado de rayos hacia atrás es el procedimiento utilizado*. Por tanto, a partir de ahora, siempre que hablemos de *ray casting* o *ray tracing* supondremos que ambos algoritmos efectúan el trazado hacia atrás.

1.5 Fases en el algoritmo de ray casting

Teniendo en cuenta que vamos a seguir la pista de los fotones partiendo del punto de visión hacia los objetos y fuentes de luz, los pasos que ha de dar el algoritmo de ray casting son los siguientes:

- primera fase:** Suponiendo que llegasen fotones al punto de visión siguiendo la trayectoria de un rayo dado, es preciso saber de dónde viene dicha luz, es decir, *cuál es su punto de partida o procedencia*. Los puntos de procedencia pueden encontrarse en las superficies de los objetos, en las fuentes de luz o ser luz de fondo, es decir, carente de un punto concreto de procedencia.
- segunda fase:** En la fase anterior hemos supuesto que llega luz al punto de visión, pero no sabemos si realmente llega. Entonces, para una trayectoria dada, una vez que se conoce el punto de procedencia, se ha de averiguar si existen fotones que recorran el camino del rayo. Si al punto de procedencia llega luz desde de cualquier lugar, cabe la posibilidad de que parte de esa luz alcance también el punto de visión, siguiendo la trayectoria del rayo. Por tanto, *es necesario averiguar si llega luz al punto de procedencia y, en caso afirmativo, cuánta llega y de qué color*.
- tercera fase:** Una vez que se conoce la intensidad y color de la luz que llega al punto de procedencia, la última cuestión que ha de resolver el algoritmo es *averiguar qué cantidad de esta luz llega al observador, y cuál es su color*.

El trazado de un rayo cualquiera finaliza pintando el píxel cruzado por el rayo con el color e intensidad de la luz que llega al observador. Si por cada píxel de la pantalla se efectúan los pasos anteriores, al finalizar tendremos en la pantalla la imagen que ve el observador.

En el estudio del algoritmo de ray casting estudiaremos las técnicas empleadas para resolver cada una de las fases anteriores, en el orden indicado. Por tanto, comenzaremos por ver los métodos utilizados para encontrar los puntos de procedencia de los fotones, o lo que es igual, cómo se realiza el trazado de los rayos primarios.

1.6 Primera fase: ¿cuál es el punto de procedencia? Trazado de los rayos y cálculo de las intersecciones

A la hora de sintetizar la imagen de un objeto es obvio que los rayos de luz más importantes son los que proceden de las superficies de éstos. *Encontrar los puntos de procedencia de los fotones en las superficies de los objetos, no es otra cosa que averiguar los puntos de intersección entre los rayos de luz (rec-*

tas) y los objetos.

Los métodos y algoritmos de cálculo de los puntos de intersección rayo-objeto dependen principalmente de las características geométricas de los objetos. Debido a esto, las figuras geométricas más importantes tienen su propio algoritmo de intersección. Por ejemplo, las esferas, cilindros, conos, toros, etc., disponen de un algoritmo de intersección apropiado, basado en la ecuación o ecuaciones que definen dichos objetos. Por otro lado, los trozos de superficies (parches) también poseen su algoritmo de intersección, dependiendo de cómo hayan sido matemáticamente definidos. Así, el algoritmo de intersección de un polígono difiere del utilizado con las superficies curvas, y dentro de éstas, los algoritmos de intersección varían según sea la naturaleza de dichas superficies (ver Glas89, pg. 79).

Por lo que acabamos de decir podría suponerse que cada objeto que se diseñe necesita su propio algoritmo de intersección. Si esto fuese cierto, los sistemas gráficos generales serían inviables dada la infinidad de objetos que es posible diseñar. Afortunadamente, la utilización de esquemas de modelado como el CSG o el B-rep reducen el problema del cálculo de las intersecciones, al simple desarrollo e implantación de un *conjunto limitado de algoritmos de intersección, tantos como primitivas diferentes (polígonos, parches curvos, etc.) se utilicen en el modelado de los objetos.*

Por otra parte, *los algoritmos basados en la síntesis directa (ray casting y ray tracing) utilizan un gran porcentaje de su tiempo de cálculo (entorno al 95%) en la búsqueda de los puntos de intersección rayo-objeto.* Por tal motivo, *es muy importante que el grado de eficiencia de los algoritmos de intersección sea lo mayor posible.* A tal efecto, en el desarrollo de dichos algoritmos se buscan técnicas que minimicen el número de operaciones aritméticas, sobre todo las más costosas, como el cálculo de raíces, divisiones o multiplicaciones, reduciendo éstas, si es posible, a sumas, restas y comparaciones¹.

No vamos a realizar un estudio exhaustivo de los diferentes algoritmos de intersección, sino que simplemente nos limitaremos a ver los casos más representativos, como es el algoritmo de intersección recta-esfera y, más adelante, el de intersección recta-polígono.

1.6.1 Intersección rayo-esfera

La búsqueda de métodos eficientes para el cálculo de las intersecciones entre los rayos y las esferas es importante, no sólo porque la esfera sea un objeto ampliamente utilizado en el diseño de los modelos, sino porque también éstas sirven para acelerar el algoritmo de visualización, como veremos más adelante.

¹ Aunque aquí se mantienen los criterios expuestos en los libros de texto, sin embargo en los procesadores de última generación las operaciones en coma flotante llegan a superar en rapidez al cálculo con enteros y a los procesos de toma de decisión, por lo que en dichos procesadores queda en tela de juicio la eficacia de los algoritmos que evitan los cálculos aritméticos aplicando test basados en comparaciones.

Comenzaremos el estudio del algoritmo de intersección rayo-esfera, definiendo los elementos que intervienen en dicho algoritmo, es decir, la recta y la esfera.

A) Definición de la recta: ecuación explícita o paramétrica

Los parámetros que definen el rayo (recta) son los siguientes:

$R_0 = [X_0, Y_0, Z_0]$: punto de origen del rayo

$R_d = [X_d, Y_d, Z_d]$: vector de dirección del rayo

Partiendo de las expresiones anteriores, se pueden expresar las coordenadas de cualquier punto $[x, y, z]$ de la recta R (rayo), en función de “ t ”, como sigue:

$$X = X_0 + X_d * t$$

$$Y = Y_0 + Y_d * t$$

$$Z = Z_0 + Z_d * t$$

[1]

Dando distintos valores a “ t ”, se obtienen todos los puntos de la recta.

Escribiendo en formato vectorial las expresiones anteriores, obtenemos la **ecuación explícita de la recta**, es decir:

$$\mathbf{R}(t) = \mathbf{R}_0 + \mathbf{R}_d * t$$

[2]

siendo $t > 0$ para el rayo.

Es conveniente que el vector de dirección del rayo esté **normalizado**², es decir, que su módulo sea unitario ($x_d^2 + y_d^2 + z_d^2 = 1$). De esta forma “ t ” representa la distancia desde el origen del rayo, referida al sistema de coordenadas.

Ver que en ray casting es de suma importancia conocer las distancias entre el origen del rayo (punto de visión) y los objetos con los que interseca, ya que el punto de intersección que se encuentre más cerca del origen del rayo, normalmente será el único visible por el observador.

B) Definición de la esfera: ecuación implícita

Los parámetros que definen la esfera son:

²Recordemos que la normalización un vector $r_d = [x, y, z]$, se efectúa como sigue:

a) se calcula el módulo del vector: $|r_d| = \sqrt{x^2 + y^2 + z^2}$, y

b) se normaliza el vector haciendo: $R_d = \left[\frac{x}{|r_d|}, \frac{y}{|r_d|}, \frac{z}{|r_d|} \right]$

$S_c = [X_c, Y_c, Z_c]$: centro de la esfera

S_r : radio de la esfera

Todo punto $[X_s, Y_s, Z_s]$ que se encuentre en la superficie de la esfera, ha de cumplir la siguiente ecuación:

[3]

$$(X_s - X_c)^2 + (Y_s - Y_c)^2 + (Z_s - Z_c)^2 = S_r^2$$

La expresión anterior se conoce como **ecuación implícita de la esfera**.

Como vemos, a partir de [3] no se pueden obtener directamente los puntos pertenecientes a la superficie de la esfera (de la forma que hemos visto en la ecuación explícita del rayo, dando valores a “t”), pero *sirve para comprobar si un determinado punto pertenece a ella o no*. Por tanto, *los puntos de la recta que satisfagan la ecuación [3], serán los puntos de intersección rayo-esfera que busquemos*.

C) Cálculo de los puntos de intersección

Antes de estudiar el algoritmo que permite encontrar los puntos de intersección rayo-esfera veamos algunas consideraciones previas sobre las posibles situaciones de intersección, en función de la ubicación del origen del rayo.

A la hora de calcular las intersecciones rayo-esfera existen tres posibilidades: que el origen del rayo (R_0) se encuentre fuera de la esfera, que esté en la superficie, o bien, que se encuentre en el interior de ella.

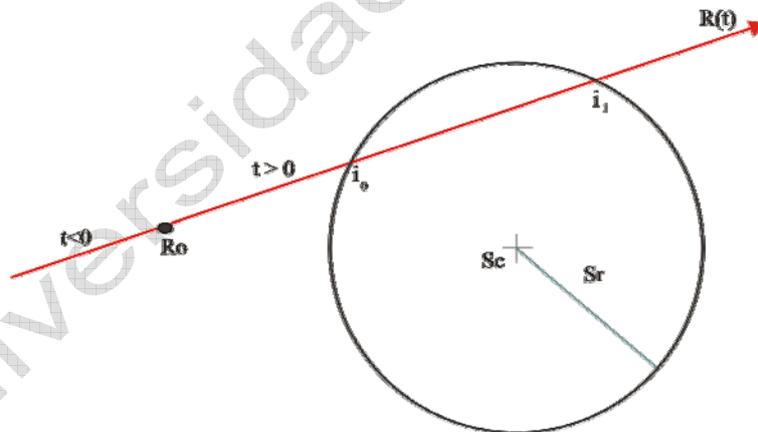


figura 5: origen del rayo fuera de la esfera

En la figura 5 se representa el caso de un rayo que interseca con la esfera, cuyo origen se encuentra fuera de ella. Como vemos, necesariamente ha de intersectarla en dos puntos i_0 e i_1 , siendo en ambos el valor de “t” positivo. En este caso interesa que el algoritmo devuelva las coordenadas del punto i_0 , ya que es el más cercano al origen del rayo y, según comentamos arriba, será el punto visible por el observador³. Además, también es necesari-

³En ocasiones también interesa conocer las coordenadas del punto i_1 , por ejemplo, cuando los modelos han sido definidos utilizando el esquema de modelado CSG.

rio conocer la normal a la superficie de la esfera en este punto i_0 , dado que posteriormente se ha de utilizar para calcular la intensidad de luz que se refleja en dicho punto, la cual depende de la orientación de la superficie.

La siguiente posibilidad es que $R_0 = i_0$, es decir, que el origen del rayo se encuentre en la superficie de la esfera. Si este es el caso, de la ecuación [2] se deduce que $t = 0$. Sin embargo, al definir dicha ecuación pusimos por condición que $t > 0$, por lo que queda descartada la posibilidad de que $t = 0$, sin antes tomar ciertas precauciones. *Esto es debido a los problemas de precisión que se originan al efectuar los cálculos en coma flotante.*

Por ejemplo, para realizar la segunda fase del algoritmo de ray casting, con frecuencia se han de trazar rayos hacia las fuentes de luz, con el origen en la superficie de los objetos, para averiguar si llega luz directa de las fuentes a los puntos de intersección visibles. A este tipo de rayos se les conoce como **rayos de sombra**.

Teóricamente se supone que el origen de los *rayos de sombra* se encuentra en las superficies de los objetos ($t = 0$), ya que sus coordenadas coinciden con las de los puntos de intersección de los rayos primarios. Sin embargo, en la práctica, debido a las imprecisiones en el cálculo de los puntos de intersección, es muy posible que “ t ” sea *casi* cero, pero no cero, o lo que es igual, que el origen del rayo de sombra se encuentre muy cercano a la superficie, pero no en la superficie. Esto puede dar lugar a que los rayos trazados con origen en una superficie intersequen con ella misma (como se muestra en la figura 6), lo que produciría manchas no deseadas en la superficie.

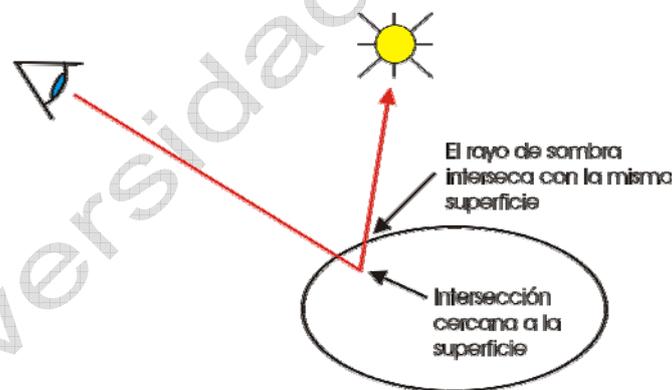


figura 6: problemas con la precisión de los cálculos cuando el origen está en la superficie

Entre otras soluciones para este problema, se puede utilizar una variable booleana de control que indique si el origen del rayo es un punto de la esfera, o bien dando unos límites de tolerancia a la igualdad con cero, en función del tamaño de la esfera (ver [Glas89, pg. 46]).

Por último, consideramos el caso en el que el origen del rayo se encuentra dentro de la esfera (figura 7). Como se puede apreciar, el valor de “ t ” para el punto i_0 será negativo, es decir, pertenece a la trayectoria del rayo, *pero no al rayo*, ya que se encuentra detrás de su origen (R_0). Por tanto, el algoritmo tendrá que devolvernos las coordenadas del punto i_1 , además de la normal en dicho punto; en este caso tendrá dirección hacia el centro de la

esfera.

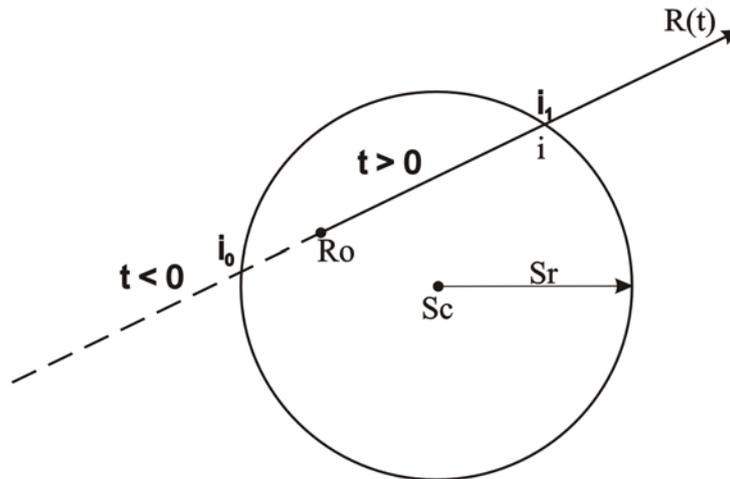


figura 7: origen del rayo dentro de la esfera

i.- Búsqueda algebraica de los puntos de intersección

La solución más simple para saber si el rayo interseca con la esfera consiste en sustituir la ecuación del rayo [1], en la ecuación implícita de la esfera [3], y calcular los valores de t (puntos del rayo) que verifican la ecuación de la esfera, es decir:

$$((X_0 + X_d * t) - X_c)^2 + ((Y_0 + Y_d * t) - Y_c)^2 + ((Z_0 + Z_d * t) - Z_c)^2 = S_r^2 \quad [4]$$

Operando y sacando factor común, ésta ecuación se convierte en una ecuación de segundo grado en t , es decir:

$$A * t^2 + B * t + C = 0 \quad [5]$$

donde:

$$\begin{aligned} A &= X_d^2 + Y_d^2 + Z_d^2 = 1 \\ B &= 2 * [X_d * (X_0 - X_c) + Y_d * (Y_0 - Y_c) + Z_d * (Z_0 - Z_c)] \\ C &= (X_0 - X_c)^2 + (Y_0 - Y_c)^2 + (Z_0 - Z_c)^2 - S_r^2 \end{aligned} \quad [6]$$

Teniendo en cuenta que $A = 1$, las soluciones a la ecuación [5] vienen dadas por:

$$\begin{aligned} t_0 &= \frac{-B - \sqrt{B^2 - 4C}}{2} \\ t_1 &= \frac{-B + \sqrt{B^2 - 4C}}{2} \end{aligned} \quad [7]$$

Dependiendo del valor que tome el discriminante “ $B^2 - 4C$ ” (positivo, cero o negativo) nos encontraremos ante los siguientes casos:

- $B^2 - 4C < 0$

Si el discriminante es negativo, las soluciones a la ecuación de segundo grado *son ambas complejas*. Dado que según la ecuación [2] la intersección rayo-esfera sólo es posible para valores reales y positivos de “ t ”, se deduce que *el rayo no interseca con la esfera*.

- $B^2 - 4C = 0$

Cuando el discriminante sea nulo, sustituyendo en [7] se obtiene que $t_0 = t_1$, es decir, los puntos de entrada y salida del rayo en la esfera coinciden. Esto significa que *el rayo es tangente a la esfera*.

- $B^2 - 4C > 0$

Por último, si el discriminante es positivo, las raíces t_0 y t_1 de la ecuación [5] serán reales. En este caso puede ocurrir que:

1) **Ambas sean positivas ($t_0 > 0$, $t_1 > 0$):** la solución que buscamos será $t_i = t_0$, ya que de las expresiones [7] se deduce fácilmente que $t_0 < t_1$. En consecuencia, el punto más cercano al observador (visible) será t_0 , y t_1 el más alejado. Como vimos, en este caso el origen del rayo se encuentra fuera de la esfera figura 5.

2) **Ambas negativas ($t_0 < 0$, $t_1 < 0$):** por definición de la ecuación paramétrica del rayo ([2]), *éste no interseca con la esfera*, aunque sí su trayectoria.

3) **Una positiva y una negativa ($t_0 < 0$, $t_1 > 0$):** de igual modo, por la definición de la ecuación [2] la solución que buscamos será $t_i = t_1$, ya que el origen del rayo se encuentra dentro de la esfera (figura 7).

* Si de los cálculos anteriores se obtiene un valor real y positivo para t_i , esto querrá decir que el rayo interseca la esfera, con lo cual pasaremos a calcular las coordenadas del *punto de intersección más cercano al origen del rayo*. Para ello, se sustituye el valor de t_i en la ecuación paramétrica del rayo ([2]):

[8]

$$R(t_i) = [x_i, y_i, z_i] = [X_0 + X_d * t_i, Y_0 + Y_d * t_i, Z_0 + Z_d * t_i]$$

ii.- Cálculo de la normal a la esfera en el punto de intersección

Según mencionamos con anterioridad, es interesante que el algoritmo de intersección encuentre también la normal a la superficie de la esfera en el punto de intersección, pues más tarde será necesaria en el cálculo de la intensidad de la luz.

Dado que la normal (r_n) en cualquier punto de la superficie de la esfera

tiene la misma dirección que el radio que pasa por dicho punto, para calcular la normal en el punto de intersección, sólo hace falta normalizar el vector del radio que pasa por éste. Por la expresión [3] vemos que S_r es el módulo del radio. Por tanto:

[9]

$$r_n = \left[\frac{(x_i - X_c)}{S_r}, \frac{(y_i - Y_c)}{S_r}, \frac{(z_i - Z_c)}{S_r} \right]$$

En el caso de que el rayo tenga su origen dentro de la esfera, la normal que se obtendrá estará dirigida hacia el centro de la esfera.

iii.- Algoritmo de intersección: solución algebraica

Como ya hemos comentado, evitar cálculos innecesarios en la búsqueda de los puntos de intersección es muy importante. En el algoritmo para la búsqueda algebraica de los puntos de intersección se gana velocidad si, por ejemplo, antes de calcular t_1 se encuentra t_0 y se prueba si es positivo. En caso afirmativo ya tenemos la solución, por lo que sólo será preciso calcular t_1 , si t_0 resulta negativo.

A continuación se establecen los distintos pasos que se han de dar para realizar el cálculo algebraico de las coordenadas de los puntos de intersección rayo-esfera. En cada paso se establece la ecuación que se utiliza para los cálculos.

Los argumentos de entrada serán los parámetros que se han establecido para definir el rayo y la esfera, es decir:

$R_0 = [X_0, Y_0, Z_0]$: punto de origen del rayo
 $R_d = [X_d, Y_d, Z_d]$: vector de dirección del rayo normalizado
 $S_c = [X_c, Y_c, Z_c]$: centro de la esfera
 S_r : radio de la esfera

```
interseccion_algebraica(R_0, R_d, S_c, S_r)
{
/* paso 1: Calcular los coeficientes A, B y C de la ecuación de
segundo grado */
```

$$A = X_d^2 + Y_d^2 + Z_d^2 = 1$$

$$B = 2[X_d(X_0 - X_c) + Y_d(Y_0 - Y_c) + Z_d(Z_0 - Z_c)]$$

$$C = (X_0 - X_c)^2 + (Y_0 - Y_c)^2 + (Z_0 - Z_c)^2 - S_r^2$$

$N = 1$ //Sentido de la normal por defecto

```
/* paso 2: Calcular el discriminante y comprobar si es > 0 */
```

$$D = (B^2 - 4C)$$

Si $D \leq 0$ entonces SALIDA: RAYO NO INTERSECA LA ESFERA

/ paso 3: Calcular t_0 y comprobar si es positiva */*

$$t_0 = \frac{-B - \sqrt{B^2 - 4C}}{2}$$

*Si $t_0 > 0$ entonces $t_i = t_0$ // origen del rayo fuera de la esfera
sino
{ /* paso 4: Calcular t_1 y comprobar si es positiva */*

$$t_1 = \frac{-B + \sqrt{B^2 - 4C}}{2}$$

*Si $t_1 < 0$ entonces SALIDA: RAYO NO INTERSECA LA ESFERA
sino*

{

$$t_i = t_1$$

N = -1 // Se cambia el sentido de la normal

}

/ paso 5: Coordenadas del punto de intersección */*

$$x_i = X_0 + X_d \cdot t_i$$

$$y_i = Y_0 + Y_d \cdot t_i$$

$$z_i = Z_0 + Z_d \cdot t_i$$

$$\text{SALIDA: } r_i = [x_i, y_i, z_i]$$

/ paso 6: Calcular la normal en el punto de intersección */*

$$r_n = \left[\frac{(x_i - X_c)}{S_r}, \frac{(y_i - Y_c)}{S_r}, \frac{(z_i - Z_c)}{S_r} \right]$$

$$\text{SALIDA: } r_n = [x_n, y_n, z_n] \cdot N$$

} / Fin del algoritmo interseccion_algebraica() */*

Ejemplos numéricos sobre el algoritmo de intersección algebraico

Veamos a continuación dos ejemplos numéricos. En el primero se muestra el caso en el que el origen del rayo está **fuera** de la esfera (figura 5) y el segundo, cuando se encuentra **dentro** (figura 7).

Ejemplo 1: El origen del rayo está **fuera** de la esfera:

Sea el rayo: $\begin{cases} R_0 = [1, -2, -1] \\ R_d = [1, 2, 4] \end{cases}$ El vector de dirección **no** está normalizado.

Lo normalizamos, quedando:

$$|r_d| = \sqrt{1^2 + 2^2 + 4^2} = \sqrt{21}$$

$$R_d = \left[\frac{1}{\sqrt{21}}, \frac{2}{\sqrt{21}}, \frac{4}{\sqrt{21}} \right] = [0.218, 0.436, 0.873] \text{ y por tanto}$$

el rayo será: $\begin{cases} R_0 = [1, -2, -1] \\ R_d = [0.218, 0.436, 0.873] \end{cases}$

Por otro lado, sea la esfera:
$$\begin{cases} S_c = [3, 0, 5] \\ S_r = 3 \end{cases}$$

Aplicación del algoritmo algebraico

/ paso 1: Calcular los coeficientes A, B y C */*

$$A = X_d^2 + Y_d^2 + Z_d^2$$

A=1 //por estar el vector de dirección del rayo normalizado

$$B = 2[X_d(X_0 - X_c) + Y_d(Y_0 - Y_c) + Z_d(Z_0 - Z_c)]$$

$$B = 2 \cdot [0.218 \cdot (1-3) + 0.436 \cdot (-2-0) + 0.873 \cdot (-1-5)]$$

$$B = -13.092$$

$$C = (X_0 - X_c)^2 + (Y_0 - Y_c)^2 + (Z_0 - Z_c)^2 - S_r^2$$

$$C = (1-3)^2 + (-2-0)^2 + (-1-5)^2 - 3^2$$

$$C = 35$$

$$N = 1$$

/ paso 2: Calcular el discriminante y ver si es positivo */*

$$D = (B^2 - 4C) = ((-13.092)^2 - 4 \cdot 35)$$

$$D = 31.4$$

Como $D > 0$ el rayo **puede** intersectar la esfera

/ paso 3: Calcular t_0 y comprobar si es positivo */*

$$t_0 = \frac{-B - \sqrt{B^2 - 4C}}{2} = \frac{13.092 - \sqrt{31.4}}{2}$$

$$t_0 = 3.744$$

Como $t_0 > 0$ entonces $t_i = t_0$ */* no tendremos que calcular t_1 */*

/ paso 5: Calcular las coordenadas del punto de intersección*/*

$$x_i = X_0 + X_d \cdot t_i = 1 + 0.218 \cdot 3.744$$

$$y_i = Y_0 + Y_d \cdot t_i = -2 + 0.436 \cdot 3.744$$

$$z_i = Z_0 + Z_d \cdot t_i = -1 + 0.873 \cdot 3.744$$

SALIDA: $r_i = [1.816, -0.368, 2.269]$

/ paso 6: Calcular la normal en el punto de intersección */*

$$r_n = \left[\frac{(x_i - X_c)}{S_r}, \frac{(y_i - Y_c)}{S_r}, \frac{(z_i - Z_c)}{S_r} \right]$$

$$r_n = \left[\frac{(1.816 - 3)}{3}, \frac{(-0.368 - 0)}{3}, \frac{(2.269 - 5)}{3} \right]$$

SALIDA: $r_n = [-0.395, -0.123, -0.910] \cdot 1$

Ejemplo 2: El origen del rayo se encuentra **dentro** de la esfera

Sea el rayo: $\begin{cases} R_0 = [3, -2, 2] \\ R_d = [1, 2, 3] \end{cases}$ El vector de dirección no está normalizado

$$|R_d| = \sqrt{1^2 + 2^2 + 3^2} = \sqrt{14}$$

Lo normalizamos:
$$R_d = \left[\frac{1}{\sqrt{14}}, \frac{2}{\sqrt{14}}, \frac{3}{\sqrt{14}} \right] = [0.267, 0.534, 0.802]$$

El rayo normalizado es:
$$\begin{cases} R_0 = [3, -2, 2] \\ R_d = [0.267, 0.534, 0.802] \end{cases}$$

Sea la esfera:
$$\begin{cases} S_c = [4, 1, 3] \\ S_r = 4 \end{cases}$$

Aplicación del algoritmo algebraico

/ paso 1: Calcular los coeficientes A, B y C */*

$$A = X_d^2 + Y_d^2 + Z_d^2$$

$A = 1$ */* por estar el vector de dirección normalizado */*

$$B = 2[X_d(X_0 - X_c) + Y_d(Y_0 - Y_c) + Z_d(Z_0 - Z_c)]$$

$$B = 2 \cdot [0.267 \cdot (3 - 4) + 0.534 \cdot (-2 - 1) + 0.802 \cdot (2 - 3)]$$

$$B = -5.344$$

$$C = (X_0 - X_c)^2 + (Y_0 - Y_c)^2 + (Z_0 - Z_c)^2 - S_r^2$$

$$C = (3 - 4)^2 + (-2 - 1)^2 + (2 - 3)^2 - 4^2$$

$$C = -5$$

$$N = 1$$

/ paso 2: Calcular el discriminante y ver si es positivo */*

$$D = (B^2 - 4C) = ((-5.344)^2 - 4 \cdot (-5))$$

$$D = 48.55$$

Como $D > 0$ el rayo **puede** intersectar la esfera

/ paso 3: Calcular t_0 y comprobar si es positivo */*

$$t_0 = \frac{-B - \sqrt{B^2 - 4C}}{2} = \frac{5.344 - \sqrt{48.55}}{2}$$

$$t_0 = -0.812$$

Como $t_0 < 0$ entonces tenemos que calcular t_1

/ paso 4: Calcular t_1 y comprobar si es positivo */*

$$t_1 = \frac{-B + \sqrt{B^2 - 4C}}{2} = \frac{5.344 + \sqrt{48.55}}{2}$$

$$t_1 = 6.156$$

Como $t_1 > 0$ entonces $t_i = t_1$, $N = -1$

/ paso 5: Calcular las coordenadas del punto de intersección */*

$$\begin{aligned}x_i &= X_0 + X_d \cdot t_i = 3 + 0.267 \cdot 6.156 \\y_i &= Y_0 + Y_d \cdot t_i = -2 + 0.534 \cdot 6.156 \\z_i &= Z_0 + Z_d \cdot t_i = 2 + 0.802 \cdot 6.156\end{aligned}$$

SALIDA: $r_i = [4.645, 1.290, 6.935]$

/ paso 6: Calcular la normal en el punto de intersección */*

$$r_n = \left[\frac{(x_i - X_c)}{S_r}, \frac{(y_i - Y_c)}{S_r}, \frac{(z_i - Z_c)}{S_r} \right]$$

$$r_n = \left[\frac{(4.645 - 4)}{4}, \frac{(1.290 - 1)}{4}, \frac{(6.935 - 3)}{4} \right]$$

SALIDA: $r_n = [0.161, 0.073, 0.983] \cdot (-1)$

1.7 Segunda fase: ¿Llega luz? ¿Cuánta, y de qué color?

Para averiguar si llega luz o no a un punto de intersección (punto de procedencia), se han de consultar los elementos del escenario que podrían aportar luz a dicho punto. Éstos son conocidos como **fuentes de luz** o de **iluminación**.

1.7.1 Fuentes de luz

En los modelos de iluminación se consideran dos fuentes básicas de luz:

- **Fuentes de luz emitida**, como por ejemplo una lámpara o el sol.
- **Fuentes de luz reflejada**, constituidas por las superficies de los objetos circundantes, tales como las paredes, espejos, etc., que reflejan la luz que reciben, y por tanto se convierten en reemisores.

Las fuentes de luz emitida se clasifican en dos grandes grupos, dependiendo de su geometría. Cuando el tamaño de la fuente emisora es pequeño comparado con el de los objetos, se considera una **fente puntual**, como por ejemplo una bombilla o el sol, ya que éste está lo suficientemente lejos como para considerarlo como un punto. En otros casos se trata de **fuentes de luz distribuida**, como ocurre con las barras de neón, que comparativamente no pueden ser consideradas como un punto cuando iluminan a los objetos cercanos (figura 8).

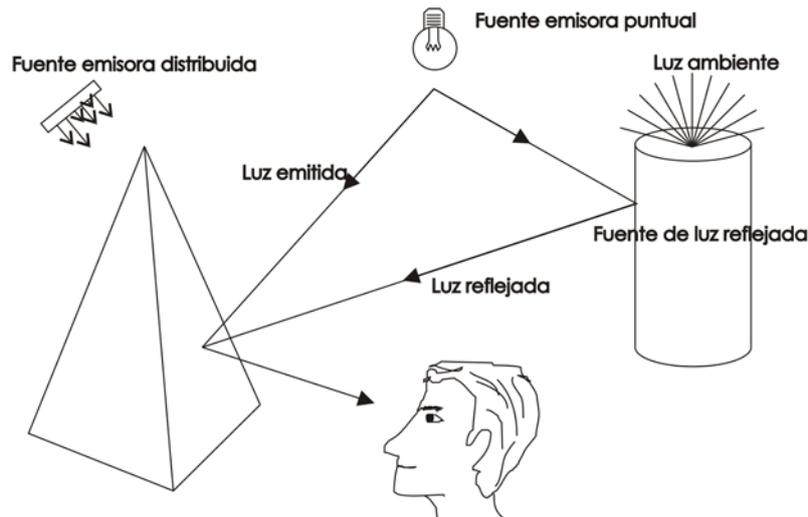


figura 8: fuentes emisoras (puntual y distribuida) y reflejada

1.7.2 Las fuentes de luz en los modelos de iluminación locales

Según vimos en el tema anterior, los *modelos locales de iluminación* no tienen en cuenta la luz reflejada en los objetos circundantes. Por tanto, *las únicas fuentes de luz que se tienen en consideración son las fuentes emisoras*.

Sin embargo, al evaluar la luz que llega a un punto se tiene presente una forma de luz reflejada, llamada **luz ambiente**, que se supone *uniforme* en todo el espacio que envuelve a los objetos. Esta uniformidad hace que *la aportación lumínica debida a la luz ambiente sea igual en todos los puntos de las superficies*, independientemente de su orientación.

Aunque la luz ambiente sea reflejada (pues se debe a la combinación de las múltiples reflexiones de la luz en los objetos circundantes), en su modelización *está totalmente desvinculada de las superficies donde se ha reflejado*. Por tanto, su inclusión en los modelos de iluminación no implica que dichos modelos tengan en cuenta la aportación directa de las fuentes de luz reflejada. En definitiva, los modelos locales que consideran la luz ambiente, no por eso dejan de serlo.

Resumiendo, *en un modelo de iluminación local, la luz que llega a un punto dado (luz incidente) viene directamente de las fuentes emisoras, o bien se trata de luz ambiente*. Ambas opciones no son excluyentes, o sea, tanto la luz ambiental como la procedente de las fuentes emisoras pueden estar presentes en la iluminación de las superficies, si bien la presencia de esta última no está garantizada. Así, *si entre un punto en la superficie de un objeto y la fuente emisora se encuentra un objeto opaco, entonces éste produce sombra*, o lo que es igual, la fuente de luz no contribuye a la iluminación del punto. En la versión simplificada del algoritmo de ray casting no suele incluirse el cálculo de las sombras, por lo que dejaremos el estudio del *sombreado* para el próximo tema, y posteriores.

1.7.3 Modelos de iluminación locales simplificados

Existen ocasiones, como por ejemplo en las primeras fases de diseño, en las que se requiere máxima velocidad en la síntesis, lo que normalmente se traduce en una calidad inferior de las imágenes obtenidas.

En estas situaciones, en las que la velocidad prima sobre la calidad, conviene aplicar modelos de iluminación simplificados que faciliten alcanzar la velocidad deseada, a la vez que proporcionen imágenes aceptables. Para ello, se adopta un modelo de iluminación local (en vez de uno semiglobal o global), simplificándolo, más o menos, en función de la calidad de las imágenes que se pretenda obtener. En lo que resta de tema iremos indicando las simplificaciones que introducimos en el modelo local que se describe.

En los modelos locales simplificados, *se excluye normalmente la posibilidad de la existencia de sombras (simplificación 4)*. Hay una razón para ello. Dado que se pretende simplificar el modelo, *se suele utilizar una única fuente emisora de luz (simpl. 1), de tipo puntual (simpl. 2), posicionada normalmente en el punto de visión o cerca de él (simpl. 3)*, como por ejemplo, en el origen de coordenadas del marco del visor (figura 9). De esta forma, *si un punto es visible, también ha de estar iluminado por la fuente de luz, por lo que no puede encontrarse en sombra*.

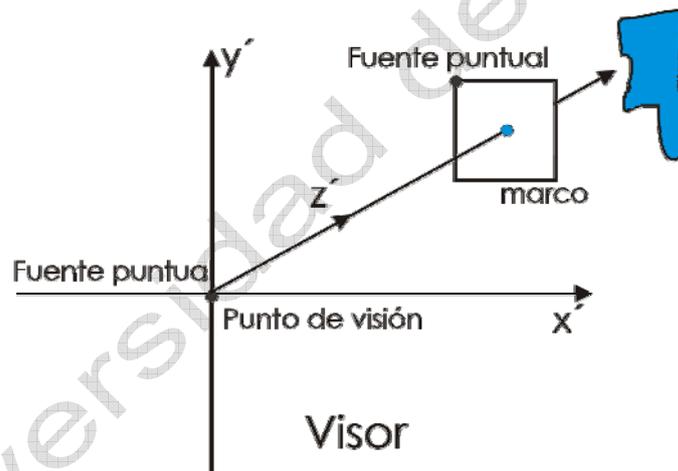


figura 9: posición de las fuentes de luz en los modelos sencillos

1.7.4 Intensidad de la luz incidente en los modelos simplificados

En los modelos de iluminación simplificados, la intensidad de la luz que llega a las superficies procedente de las fuentes puntuales se considera de igual intensidad que la emitida por dichas fuentes, es decir, *no se tiene en cuenta la atenuación debida a la distancia que recorre (simpl. 5)*.

Por tanto, si a la intensidad de la luz ambiente la llamamos I_a , y la que sale de la fuente puntual emisora I_i , la intensidad total de luz incidente en cada punto viene dada en los modelos simplificados por:

$$I_a + I_i$$

⊙ Ya conocemos la procedencia e intensidad de la luz incidente en un punto, pero nada sabemos de su color.

Aunque todo el mundo tiene una idea más o menos intuitiva sobre la forma en que los televisores y monitores obtienen su gama de colores, vamos a estudiar los **modelos de color** para entender mejor cómo se establece el color de la luz incidente, y cómo se calcula el color de la luz reflejada en un punto, una vez que ha interactuado con la superficie.

1.7.5 Modelos de color

La finalidad de los modelos de color es averiguar el color de la luz reflejada en un punto, a partir de las propiedades ópticas de la materia (punto) y las características de la luz incidente.

Son varios los modelos de color, aunque se pueden clasificar en dos grupos principales:

a) *modelos basados en la combinación de tres colores*, a partir de los cuales se obtienen los demás. En este grupo encontramos, entre otros, los modelos RGB y CMY.

Los modelos de este grupo pueden ser **aditivos** o **substractivos**. Cuando la gama de colores que abarca el modelo se obtiene *agregando colores* (como el RGB), los modelos son aditivos. Si la gama de colores se consigue *eliminando colores*, entonces los modelos son substractivos, como en el caso del modelo CMY.

b) *grupo de modelos de color basados en la “paleta de los pintores”*, que obtienen las combinaciones de color partiendo de un color puro, añadiendo blanco o negro, para obtener las diversas tonalidades. Este tipo de modelo es de uso más intuitivo. Dentro de este grupo están los modelos HSV y HLS.

El siguiente diagrama muestra una clasificación más completa de los modelos de color:

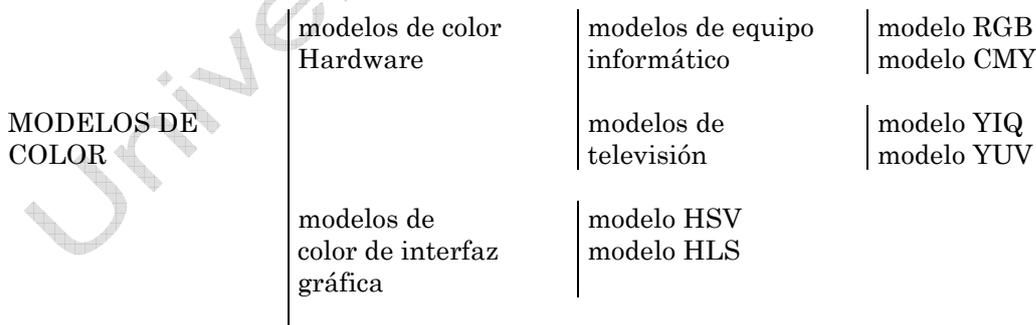


diagrama 1

Debido a nuestra política simplificadora, de momento nos limitaremos a ver los modelos RGB y CMY, dejando los restantes modelos para el siguiente tema. La elección de estos modelos, clasificados a veces como **modelos hardware**, se debe a que son los utilizados por los periféricos gráficos (monitores e impresoras), y por tanto los que más urge conocer. Los modelos HSV y HLS han sido pensados para facilitar al usuario la manipulación de

los colores, de ahí que se les conozca como **modelos de interfaz gráfica**.

Antes de describir los modelos RGB y CMY conviene conocer las bases teóricas en que se fundamentan, por lo que no queda más remedio que comenzar desde el principio viendo qué entendemos por luz y color.

A) Luz y color

El **color** es un concepto frecuentemente utilizado para describir el aspecto de un determinado objeto. Aunque en la mayoría de las ocasiones el describir el color de un objeto es una tarea sencilla, el concepto de color es bastante más complicado ya que no sólo depende de la naturaleza del objeto, sino que también depende de las características de la luz que lo ilumina, y de la fisiología del observador.

La **luz** está formada por infinidad de partículas cuánticas, denominadas fotones, los cuales son emitidos principalmente por los electrones de los átomos al pasar de un estado de energía a otro inferior. Dado que las partículas cuánticas son al mismo tiempo ondas electromagnéticas, normalmente se considera y estudia la luz como radiación de este tipo. En el espectro electromagnético, la luz visible se corresponde con una banda de frecuencias que va desde los $4,3 \times 10^{14}$ hasta los $7,5 \times 10^{14}$ hertz. Dentro de esta banda, cada frecuencia es detectada e interpretada por nuestro sistema visual como un color diferente, siendo la frecuencia más baja la correspondiente al rojo y la más alta al violeta. Entre estas dos frecuencias extremas del espectro de la luz, el ojo humano puede distinguir cerca de 400.000 colores diferentes.

Según demostró Newton mediante su famoso experimento del prisma, la luz del Sol (que consideramos “luz blanca”) no es otra cosa que la combinación en nuestros ojos de las diferentes frecuencias del espectro visible que, una vez separadas (clasificadas) por el prisma, se nos manifiestan de forma más o menos individual agrupadas entorno a 7 frecuencias fundamentales que denominamos violeta, añil, azul, verde, amarillo, naranja y rojo.

Cuando la luz de una fuente (el Sol, por ejemplo) incide sobre un objeto, algunas frecuencias son reflejadas y otras absorbidas por éste. *La combinación de frecuencias presentes en la luz reflejada determina lo que nosotros percibimos como color del objeto.*

B) Características de la luz

Acabamos de describir la luz como un conjunto de “ondas electromagnéticas dentro de una banda de frecuencias determinada”. Sin embargo, nos interesa también hablar de la luz en términos de “longitudes de onda”, lo cual no representa problema alguno, pues en las ondas electromagnéticas ambas magnitudes están relacionadas por la ecuación

[10]

$$c = \lambda f$$

siendo c la velocidad de la luz, λ la longitud de onda y f la frecuencia.

Podemos decir entonces que la luz visible está formada por ondas electromagnéticas de diferentes longitudes de onda, en el espectro comprendido entre los 400 (violeta) y los 700 nanómetros (roja).

Si decimos que un objeto es de color rojo, la luz que estamos percibiendo tiene una **frecuencia dominante** o **longitud de onda dominante** que es la que se corresponde con el color rojo. La frecuencia dominante es, por lo tanto, lo que se denomina **color** o **matiz** de la luz.

Las características de la luz se pueden medir o evaluar desde la perspectiva subjetiva del ojo humano, o bien desde un punto de vista más objetivo, como el de la **colorimetría**. Este campo de la física considera asociados a la luz los parámetros de **pureza** o **saturación** y **luminancia**. El primero se refiere al número de frecuencias diferentes que forman la luz. *Cuantas menos frecuencias diferentes posea, más pura es la luz*; por tanto, la luz blanca es la menos pura de todas. Por otra parte, la **luminancia** está relacionada con la medida de la intensidad de la luz que se percibe. Por último, la palabra **cromaticidad** o **cromatismo** se utiliza para indicar las características de la luz mediante los dos parámetros que la definen: *longitud de onda dominante y pureza*. La terminología anterior sirve tanto para las fuentes de luz emitida, como para las de luz reflejada.

Las características de la luz que percibimos de dos o más fuentes quedan determinadas por éstas. Dos fuentes de luz emitiendo en diferentes colores pueden ser utilizadas para generar una gran diversidad de colores mediante la variación de las intensidades de ambas fuentes. Si como resultado de la combinación resulta el color blanco, se dice que los colores de ambas fuentes son **complementarios**. Algunas de estas combinaciones son el rojo con el cian, el azul con el amarillo y el verde con el magenta.

Con una selección apropiada de dos o más colores de partida se puede obtener otros muchos colores. Normalmente se utilizan tres colores para este fin, y el conjunto de colores obtenido mediante su combinación se conoce como **gama de colores** del modelo. Los dos o tres colores mezclados para producir otros se conocen como **colores primarios** del modelo.

C) Colores primarios estándar

Existen muchos grupos de colores que al combinarlos pueden dar lugar a una amplia gama de colores, lo que significa que hay muchos colores que podrían considerarse **primarios**.

Como ocurre siempre que hay muchas posibilidades donde elegir, pronto se planteó la cuestión sobre *cuál era el número mínimo de colores necesarios para generar todos los demás, y cuáles eran los más eficientes*. La respuesta fue que con **tres** colores era suficiente para conseguir una amplia gama de colores; sin embargo, en el espectro visible *no hay ningún grupo de tres colores capaces de generar todos los colores del espectro*, es decir, entre los conjuntos de *colores primarios* no hay ninguno universal.

El problema de la estandarización de los colores primarios se resolvió airosamente eligiendo *tres colores primarios ficticios* (inexistentes), los cua-

les generan una gama de colores superior a la generada por cualquier trio de colores reales.

En 1931, la Comisión Internacional de Iluminación (*Commission Internationale de l'Eclairage* (**CIE**)) estableció un estándar internacional para los tres colores primarios, los cuales quedaron determinados mediante funciones matemáticas; así, cualquier color puede ser definido como una combinación de los colores primarios estándar, si se conoce la cantidad (intensidad) que se precisa de cada uno.

El conjunto de colores primarios CIE normalmente se representa mediante **X**, **Y**, **Z** representando al rojo, verde y azul. **X** se considera como un rojo con una saturación (pureza) mayor que cualquier rojo real. Algo parecido ocurre con los otros dos colores, siendo **Z** e **Y** un azul y verde supersaturados, respectivamente.

Si (**X**, **Y**, **Z**) se representan como *vectores unitarios en los ejes de un espacio de coordenadas tridimensional*, entonces cualquier color C_λ puede expresarse como:

$$C_\lambda = \mathbf{X}X + \mathbf{Y}Y + \mathbf{Z}Z \quad [11]$$

donde X, Y, y Z son las cantidades necesarias de colores primarios para obtener el color C_λ .

Para analizar las propiedades del color es conveniente normalizar las cantidades X, Y, Z con respecto a la luminancia, es decir, la cantidad total de luz (color) que se utiliza ($X + Y + Z$). Queda entonces que:

$$x = \frac{X}{X+Y+Z} \quad y = \frac{Y}{X+Y+Z} \quad z = \frac{Z}{X+Y+Z} \quad [12]$$

siendo $x + y + z = 1$.

D) El diagrama de cromatismo CIE.

Utilizando los valores normalizados anteriores x , y , z , es posible definir los colores con sólo dos cantidades, por ejemplo x e y , ya que $z = 1 - (x + y)$. Estos valores (x , y), suelen conocerse como **valores** o **parámetros de cromatismo**, ya que al haber sido normalizados dividiendo por la luminancia, sólo dependen del tono de la luz (frecuencia dominante) y de su pureza, siendo independientes de la cantidad de energía luminosa (luminancia). Si se representan en el plano los valores (x , y) para todos los colores visibles del espectro, se obtiene una curva conocida como **diagrama de cromatismo CIE**, como muestra la figura 10.

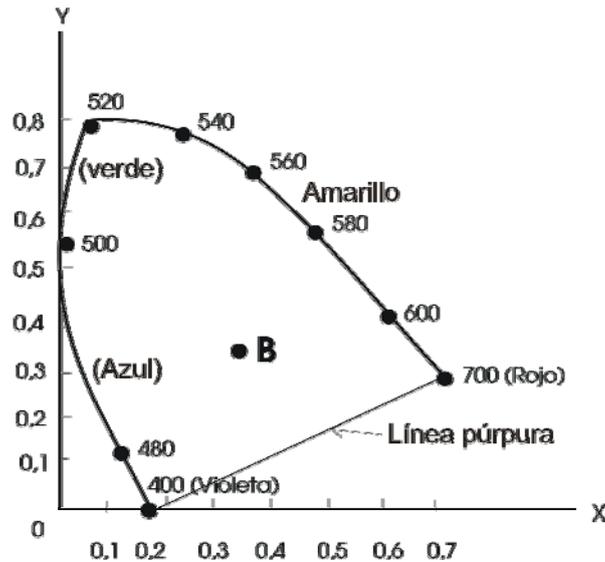


figura 10: diagrama de cromatismo CIE

Los puntos de la curva se corresponden con los colores puros del espectro, etiquetados a lo largo de ésta con la longitud de onda en nanómetros desde el rojo hasta el violeta. La línea que une el rojo con el violeta, conocida como *línea púrpura*, no forma parte del espectro. Por último, los puntos interiores representan todas las posibles combinaciones de colores del espectro. El punto **B** representa a la luz blanca del Sol.

El diagrama de cromatismo se utiliza para:

- Determinar la longitud de onda dominante y la pureza de un color dado.
- Comparar gamas de colores generadas por ternas diferentes de colores primarios.
- Identificar colores complementarios.

© Determinación de la pureza y longitud de onda dominante

Si la curva se denominada “diagrama de cromatismo” es porque ha sido construida a partir de los valores x e y (valores de cromatismo), que como vimos dependen únicamente de la longitud de onda dominante y de la pureza del color. Por tanto, el diagrama proporciona un método eficaz para definir cuantitativamente la pureza y la longitud de onda dominante de los colores.

Por ejemplo, sea **A** el color representado en el gráfico

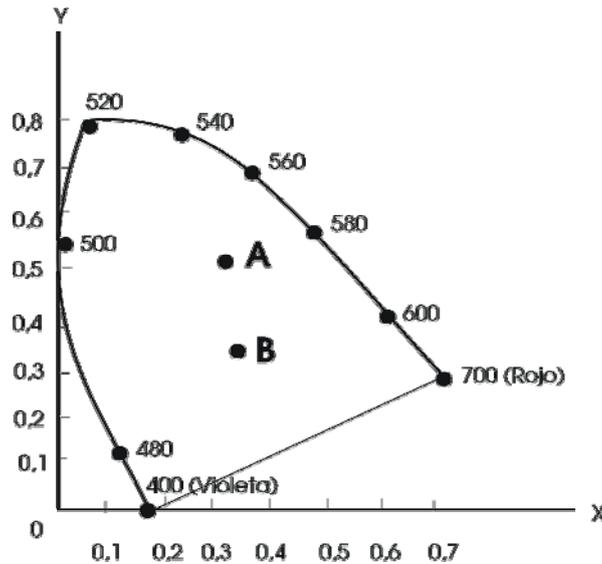


figura 11: representación del color A en el diagrama cromático

Para este punto de color definiremos:

a) pureza del color:

Viene determinada por la distancia relativa del punto de color (A en este caso) al punto de luz blanca (B), tomando como referencia la recta que une ambos puntos con la curva.

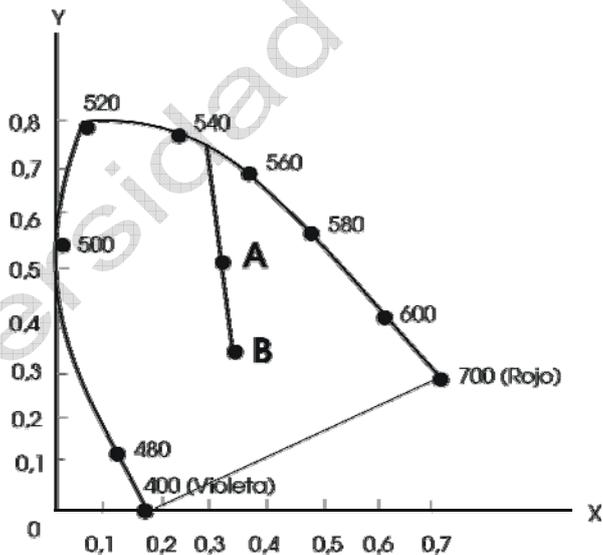


figura 12: búsqueda de la pureza de un color

En este caso, el color A es alrededor de un 40% puro ya que no está situado en la mitad de la recta (en cuyo caso sería puro al 50 %) sino algo más cerca del punto de color blanco (B). Cuanto más cerca está un color del blanco (punto B), más frecuencias diferentes forman dicho color, y por lo tanto menos puro es.

b) longitud de onda dominante:

La longitud de onda dominante de un color A, queda determinada por

el punto de la curva donde interseca la línea que une B con A. Para el punto A, la longitud de onda viene determinada por la longitud de onda del punto C, ya que es el punto en el cual se produce la intersección con la recta que pasa por A y B.

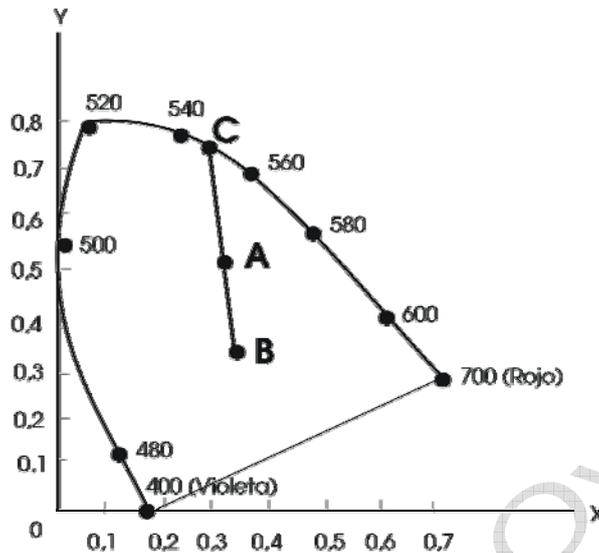


figura 13: búsqueda de la longitud de onda dominante en un color

© Determinación de la gama de colores

Sean D y E los colores primarios que van a ser mezclados en cantidades variables para proporcionar otros colores. La gama de colores resultante de la mezcla de los dos anteriores viene representada en el diagrama de cromatismo por la línea que une los colores D y E.

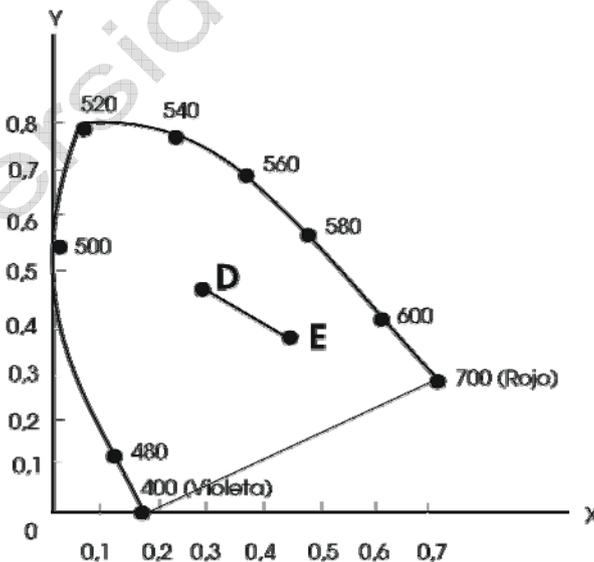


figura 14: gama de colores combinando dos primarios

Si se incluye un tercer color F para mezclar con los anteriores, entonces los colores de la gama son los comprendidos en el triángulo formado por la unión de los tres colores.

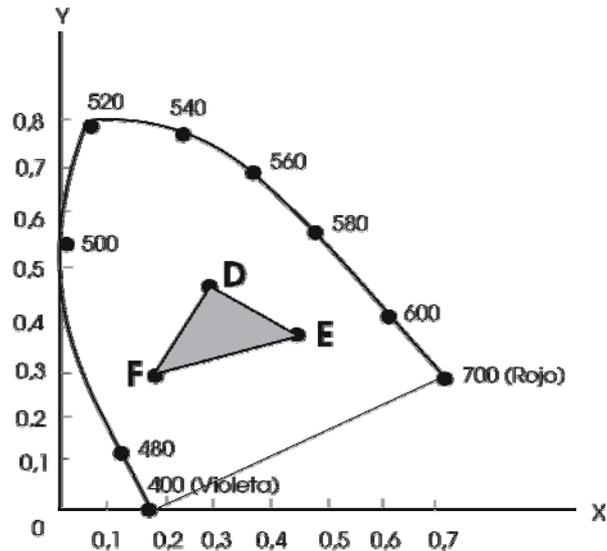


figura 15: gama de colores combinando tres primarios

El diagrama anterior permite *explicar por qué ningún conjunto de tres colores puede dar lugar a todos los colores del espectro*. La razón se debe a que *no es posible formar ningún triángulo que abarque todos los colores del diagrama de cromatismo*. Por ejemplo, si formamos uno de los triángulos más amplios que se puede conseguir en el diagrama de cromatismo (figura 16), la gama de los colores es la comprendida en el área oscura, pero quedan unos cuantos colores (como el G) que no pueden ser generados a partir de los colores primarios seleccionados.

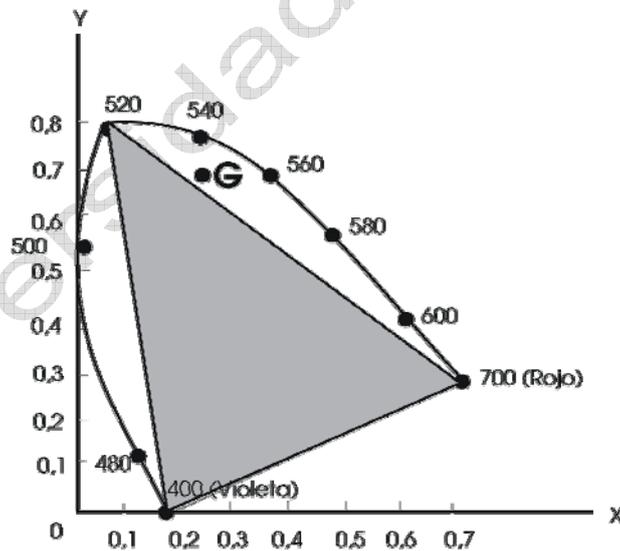


figura 16: no puede existir un triángulo que abarque todos los colores

© Determinación de los colores complementarios

Como la gama de colores debida a dos colores primarios está formada por todos los colores de la línea que los une, dos colores primarios complementarios (mezclados dan el blanco) han de estar posicionados en los extremos de una línea que pase por el punto B, como muestra el ejemplo de la figura 17. Mezclando las cantidades apropiadas de los colores C_1 y C_2 , se ha

de obtener el color blanco.

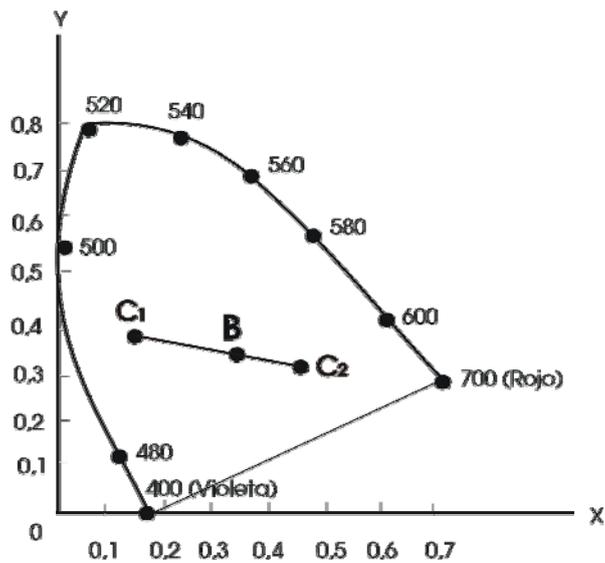


figura 17: colores primarios complementarios

Por último, para finalizar la lista de utilidades del modelo CIE, al ser un modelo teórico que abarca todos los colores del espectro visible, *se puede utilizar como base de comparación (medida) de otros modelos de color*. Mediante transformaciones lineales, los colores primarios de cualquier modelo pueden transformarse en colores X, Y, Z del modelo CIE, y viceversa. Por tanto, este modelo proporciona un enlace entre todos los posibles modelos. Estudiemos ahora los modelos RGB y CMY cuya filosofía es muy similar a la del estándar CIE, aunque más realista.

E) Modelo de color RGB

Este modelo se basa en la *teoría del triestímulo* sobre la visión, que dice que percibimos los colores debido a tres pigmentos situados en unas células de la retina denominadas conos, que son sensibles al rojo (630 nm), verde (530 nm) y azul (450 nm). Haciendo caso a esta teoría, el modelo RGB (**R**ed, **G**reen, **B**lue) adopta el rojo el verde y el azul como colores primarios. Este modelo es la base para la emisión de color en una pantalla de rayos catódicos, utilizando estos tres colores primarios.

Teniendo en cuenta las longitudes de onda anteriores, la gama de colores del modelo RGB queda representada en el diagrama cromático como se ve en la figura 18.

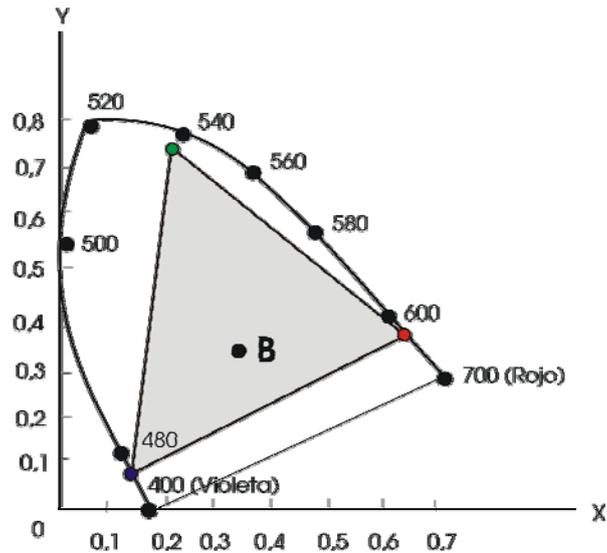


figura 18: gama de colores del modelo RGB

La gama de colores del modelo RGB, que según muestra la figura anterior se trata de un subconjunto de los colores que puede percibir el ojo humano, puede quedar matemáticamente representada mediante un cubo unitario (normalizado). Los vértices en los ejes ortogonales representan los colores primarios puros, y los vértices opuestos están asociados a los colores complementarios de cada uno de los primarios.

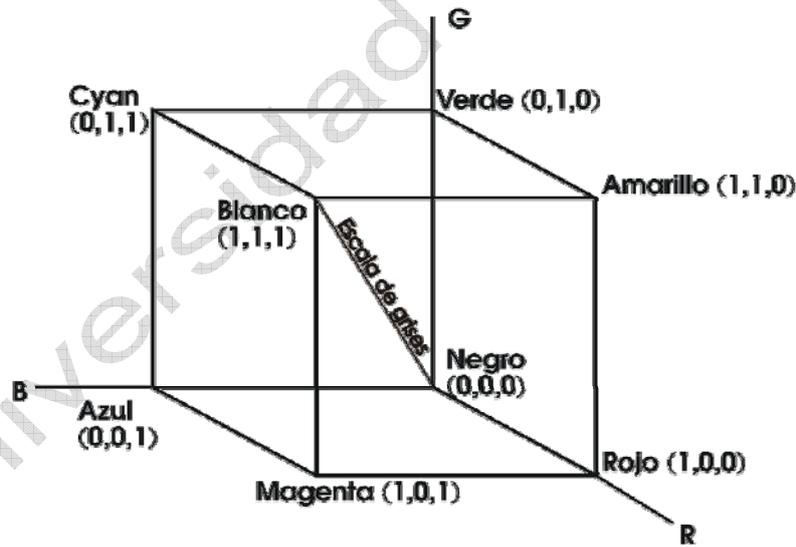


figura 19: representación del espacio de color del modelo RGB

Como indicamos al comienzo, el RGB se trata de un *modelo aditivo*, pues las intensidades de los colores primarios se suman para dar lugar a otros colores. Los colores se consiguen combinando distintos valores de los componentes rojo, verde y azul. El magenta, por ejemplo, viene representado por (1,0,1) y se obtiene sumando el azul (0,0,1) y el rojo (1,0,0).

Cada punto de color en el cubo puede ser representado por una terna (R,G,B) donde R, G y B toman valores comprendidos entre 0 y 1. Así, un co-

lor cualquiera C_λ , viene dado por:

[13]

$$C_\lambda = \mathbf{R}\mathbf{R} + \mathbf{G}\mathbf{G} + \mathbf{B}\mathbf{B}$$

siendo $(\mathbf{R}, \mathbf{G}, \mathbf{B})$ los vectores unitarios de cada eje, y R, G, B valores escalares en el rango $[0, 1]$.

Si los tres componentes (rojo, verde y azul) de un color son iguales (y distintos de 0 y 1) el resultado es un color en la gama de los grises. Por lo tanto, la escala de grises se representa en este modelo mediante la diagonal que une los vértices de los colores blanco y negro. Un gris a mitad de camino entre el negro y el blanco quedaría representado por la terna:

$$(0,5, 0,5, 0,5)$$

A un monitor llegan diferentes señales de intensidad (R,G,B) donde (0,0,0) corresponde al color negro y (1,1,1) al blanco; en este último los colores rojo, verde y azul alcanzan su máxima intensidad (luminancia), que nosotros percibimos como máximo brillo. Entre estos dos extremos se obtiene un rango de intensidades que, en un monitor de 24 bits por píxel, es de 256 para cada color.

F) Modelo de color CMY

Este modelo se utiliza para describir los colores que se obtienen sobre papel blanco mezclando pigmentos de color.

Los colores primarios utilizados son el cian, el magenta y el amarillo, que son los colores complementarios del rojo, verde y azul.

Los tres colores primarios del CMY son llamados **subtractivos** ya que su efecto es sustraer color de la luz blanca. Por ejemplo, el cian procede de una mezcla de verde y azul. Cuando la luz blanca se refleja en algo coloreado de cian, la luz reflejada no tendrá componente rojo. Esto significa que el rojo es absorbido (sustraído) por la tinta color cian. Una combinación de cian y magenta produce azul porque los componentes rojo y verde de la luz incidente son absorbidos. Siguiendo el mismo proceso se obtienen otras combinaciones de color.

Dado que los colores son ondas electromagnéticas de frecuencias diferentes, el modelo CMY se entiende más fácilmente si en vez de hablar de “mezcla de colores”, se habla de “mezcla o combinación de filtros”. Así, C sería el filtro del rojo (R), M el filtro del verde (G) e Y el del azul (B).

En otras palabras, si de la luz blanca (RGB) se elimina la luz roja, quedarían la verde y la azul (GB), que al combinarse en nuestros ojos producirían el color cian, que es precisamente el color que atribuimos al filtro de la luz roja. El filtrado de la luz blanca quedaría esquemáticamente como sigue

$$\text{RGB (luz blanca)} - \text{R (luz roja)} \rightarrow \text{GB} \rightarrow \text{C (color cian)}$$

$$\text{RGB (luz blanca)} - \text{G (luz verde)} \rightarrow \text{RB} \rightarrow \text{M (color magenta)}$$

$$\text{RGB (luz blanca)} - \text{B (luz azul)} \rightarrow \text{RG} \rightarrow \text{Y (color amarillo)}$$

La representación matemática de este modelo consiste, al igual que en el modelo anterior, en un cubo cuyas coordenadas en cada eje son las complementarias de los colores del modelo RGB.

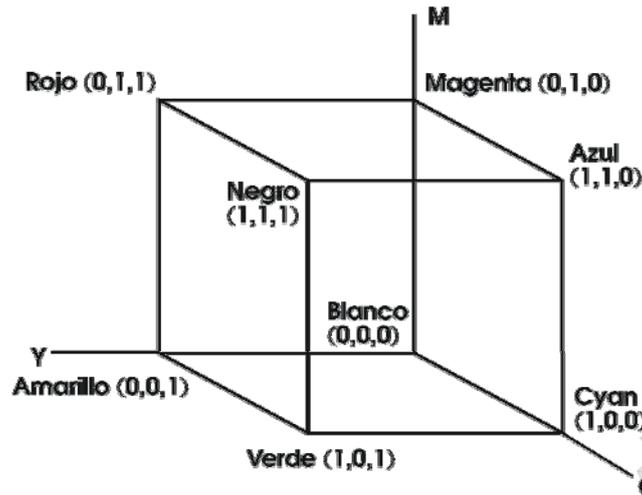


figura 20: representación del espacio de color del modelo CMY.

El origen del cubo de color (0,0,0) representa el blanco. Aumentando la cantidad de cada uno de los colores (filtros) primarios se tiende hacia el negro en la esquina opuesta del cubo y tiene por coordenadas (1,1,1); por lo tanto, la gama de los grises es representada mediante la diagonal que une ambos vértices.

Los dispositivos de impresión utilizan 3 fuentes (depósitos) de color, para los colores cian, magenta y amarillo. Además, utilizan una cuarta fuente para el color negro, ya que la unión de los tres colores primarios no proporciona negro, sino gris oscuro.

G) Conversión entre ambos modelos de color

Una imagen en color generada en un monitor se rige por el modelo RGB. Si se desea imprimir ésta imagen, los valores de cada color en el modelo RGB han de ser transformados en valores CMY ya que es el modelo utilizado en las impresoras de color.

Dado que los colores de ambos modelos son complementarios, si un color tiene componentes (r, g, b), en el modelo CMY tendrá por componentes (1-r, 1-g, 1-b). De la misma manera, un color (c, m, y) tendrá por componentes (1-c, 1-m, 1-y) en el modelo RGB. Esto significa que un valor (R,G,B) sustraído de la luz blanca proporcionará el valor (C,M,Y). De la misma manera, dado un color (C,M,Y), se obtendrá el correspondiente (R,G,B) restando (C,M,Y) del color negro. Las ecuaciones vectoriales de conversión quedan

[14]

$$\begin{bmatrix} C \\ M \\ Y \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} - \begin{bmatrix} R \\ G \\ B \end{bmatrix}$$

$$\begin{bmatrix} R \\ G \\ B \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} - \begin{bmatrix} C \\ M \\ Y \end{bmatrix}$$

1.8 Tercera fase: ¿De qué intensidad y color es la luz que llega al observador?

En la tercera y última fase del algoritmo de ray casting, se ha de averiguar cuál es la intensidad y el color de la luz que parte de los puntos de intersección (puntos de procedencia) hacia el observador. Para ello, se utilizan los llamados **modelos de intensidad** y los modelos de color.

1.8.1 Modelos de Intensidad

Algunos autores no hacen diferencia entre los modelos de iluminación y los de intensidad [Hear97], y otros, en cambio, prefieren considerar a estos últimos como una parte de los modelos de iluminación y viceversa. Sin embargo, ambos tienen sus funciones bien delimitadas. Así, los modelos de iluminación establecen (modelan) la procedencia, intensidad, color, etc. de la luz que llega a las superficies, mientras que los modelos de intensidad son los que deciden qué cantidad de esa luz llega al observador, así como su color. Desde luego, los resultados del modelo de intensidad dependen totalmente de modelo de iluminación utilizado, pero de cualquier modo sus cometidos son diferentes.

La intensidad y el color de la luz que percibe el observador depende en gran medida de las *propiedades ópticas* de los objetos, es decir, de si son opacos o translúcidos, brillantes o mates, rugosos o lisos, etc. Además, se han de tener en cuenta las posiciones relativas de los objetos y el observador.

Dado que, en principio, *el modelo de intensidad se ha de aplicar en cada punto de intersección de los rayos con las superficies*, es muy importante que el tiempo de cálculo del modelo sea pequeño. Esta imposición ha llevado a que, en mayor o menor grado, los modelos de intensidad adopten *planteamientos empíricos*, en vez de físicos, en la realización de los cálculos.

En el caso de los modelos de iluminación locales, según vimos en el modelo simplificado (pg. 17), el total de luz incidente viene dado por I_a e I_i , es decir, la luz ambiente y la que llega directamente de las fuentes emisoras. *Al modelo de intensidad le toca averiguar qué parte de cada clase de luz incidente llega al observador.*

Una fórmula típica en los modelos de intensidad empíricos para establecer lo que ocurre con la luz que incide en un objeto es la siguiente

$$\text{Luz incidente} = \text{Luz reflejada} + \text{Luz dispersada} + \text{Luz absorbida} + \text{Luz transmitida}$$

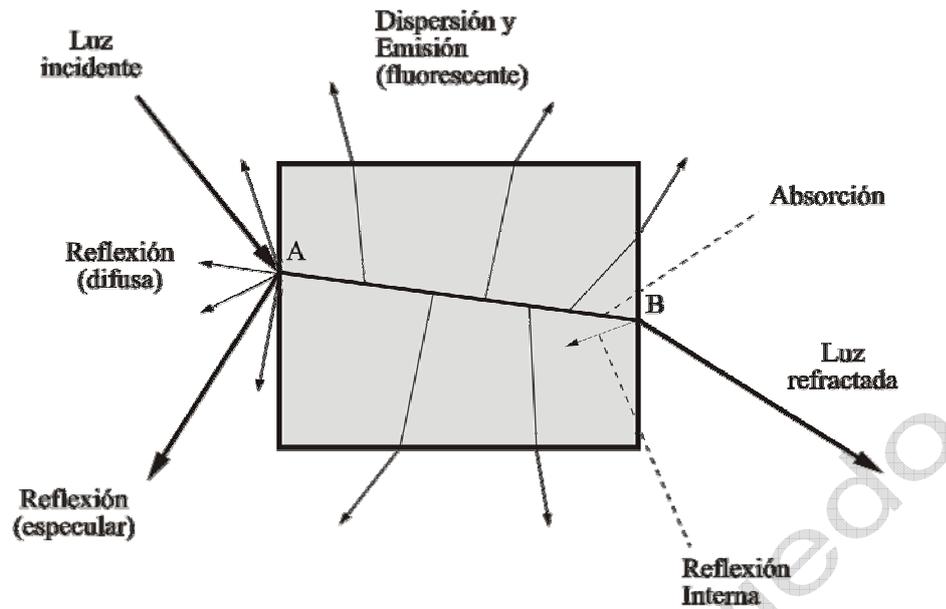


figura 21: distribución de la luz incidente

Dentro de los modelos de intensidad, a cada uno de los términos de la ecuación anterior se le suele dar el nombre de “*componente*”, encontrándonos así con la *componente ambiental*, la *componente difusa*, etc. *El cálculo de la intensidad aportada por cada una de estas componentes suele hacerse de modo independiente.*

1.8.2 Modelo de Intensidad Simplificado

Igual que se hizo en la fase anterior, incluiremos algunas simplificaciones en el modelo de intensidad para lograr el modelo de iluminación que pretendemos. En primer lugar, queda descartada la posibilidad de que existan en la escena objetos translúcidos, es decir, *el modelo no va a permitir la transmisión de la luz (simpl. 6)*. *Tampoco se tratará la aplicación de texturas o color falso (simpl. 7)*. Así pues, lo único que va a calcular el modelo que veremos será la intensidad de la luz reflejada que llega al observador.

Para comprender mejor la forma en que se modeliza (calcula) la cantidad de luz que se refleja hacia el observador, conviene analizar un poco más los factores que intervienen en la reflexión de la luz. Hemos comentado arriba que estos factores son dos: las *propiedades ópticas* de las superficies donde se refleja la luz y las *posiciones relativas* entre los rayos de luz incidente y las superficies.

A falta de nada mejor, las propiedades ópticas de las superficies que influyen sobre la reflexión quedan representadas en los modelos de intensidad por los **coeficientes de reflexión (K)**. Éstos toman normalmente un valor normalizado en el intervalo $[0, 1]$. De acuerdo con las propiedades lumínicas de las superficies, se les asocia un coeficiente de reflexión alto ($K \cong 1$) si son reflectoras, o uno bajo ($K \cong 0$), cuando absorben gran parte de luz incidente (poco reflectoras).

Por otro lado, las posiciones relativas entre los rayos de luz incidente y

las superficies también son importantes a la hora de calcular la intensidad de la luz reflejada. En la figura 22 claramente se puede ver que *las superficies perpendiculares a los rayos incidentes (a) están más iluminadas que aquellas que se encuentran en ángulo con respecto a la dirección de la luz (b)*.

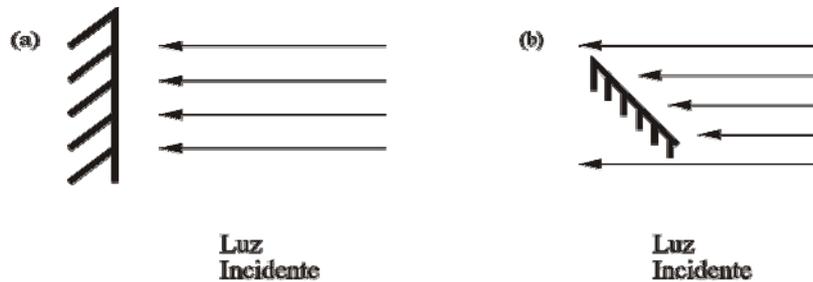


figura 22: variación de la iluminación según sea la orientación de las superficies

El ángulo que forman la normal a la superficie reflectora y la trayectoria de la luz incidente (figura 23) se denomina **ángulo de iluminación** o **ángulo de incidencia**.

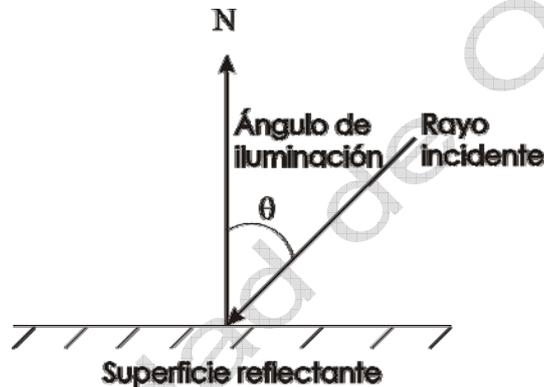


figura 23: ángulo de iluminación o incidencia

A) Modelización de la reflexión de la luz directa

Con la luz que llega de las fuentes emisoras (*luz directa*), normalmente los modelos de intensidad suelen modelar dos tipos diferentes de reflexión: la **reflexión difusa** y la **reflexión especular**. En el modelo de intensidad simplificado *sólo estudiaremos la modelización empírica de la reflexión difusa (simpl. 8)*.

i.- Reflexión difusa

El cálculo de la reflexión difusa en un punto dado, debida a la luz procedente de una fuente puntual, se basa en la **ley del coseno de Lambert**, la cual establece que *la intensidad de la luz difusa es proporcional al coeficiente de reflexión (K) y al coseno del ángulo de iluminación (θ)*.

Expresada en modo matemático, la ley de Lambert queda como sigue:

[15]

$$I_d = I_i \cdot K \cdot \cos \theta$$

con $0 \leq \theta \leq \pi/2$, siendo I_i la intensidad de la luz incidente que llega de la fuente puntual.

Según la expresión anterior, la intensidad debida a la reflexión difusa (I_d), será máxima cuando $\cos \theta$ sea 1, es decir, cuando la luz sea perpendicular a la superficie ($\theta = 0^\circ$). A medida que el ángulo de iluminación aumenta, la luz incidente que llega a la superficie es menor. Vemos que [15] es una expresión simple y sencilla de calcular, por lo que ha sido adoptada por los modelos de intensidad empíricos.

Es lógico pensar que las propiedades ópticas de una superficie sean las mismas para toda la luz blanca que incida en ella, *independientemente de su procedencia*. Sin embargo, *en los modelos de intensidad se suelen diferenciar los coeficientes de reflexión, dependiendo de cuál sea la procedencia de la luz y el efecto lumínico que se esté modelando*. De este modo, se dispone de un mayor número de parámetros para controlar la calidad de las imágenes.

Por este motivo, para el cálculo la reflexión difusa se introduce el **coeficiente de reflexión difusa** (K_d), por lo que la expresión [15] queda como sigue:

[16]

$$I_d = I_i \cdot K_d \cdot \cos \theta$$

⊙ Los rayos incidentes de una fuente cercana a la superficie no son paralelos (figura 24-a). Por tanto, si la fuente puntual se encuentra cerca de los objetos, va a ser difícil modelar la reflexión difusa ya que, de unos rayos a otros, varía el ángulo de incidencia. Para evitar este problema, *se supone que las fuentes puntuales están lejos de la superficie, de forma que los rayos incidentes sean paralelos (simpl. 9)*⁴.

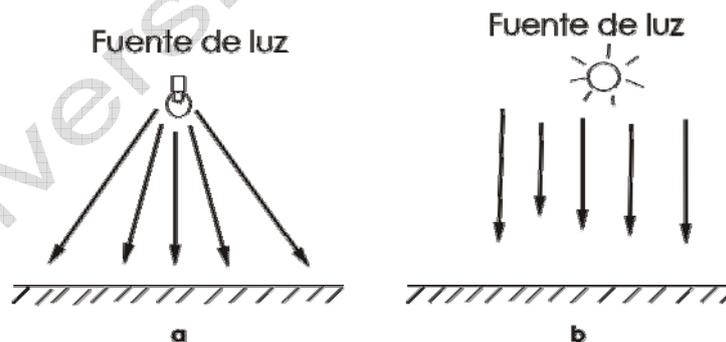


figura 24: distanciamiento de las fuentes

La simplificación anterior permite considerar que toda la luz que llega

⁴ Con la incorporación de esta nueva simplificación, ya no se puede considerar la fuente de luz puntual posicionada en el punto de visión o en el origen de coordenadas del marco [ver **Modelos de iluminación locales simplificados** pg. 17], dado que estos lugares se encuentran cerca de los objetos. *Ahora simplemente se ha de suponer que la luz que llega de la fuente viene de esa dirección.*

de una fuente puntual tiene la misma dirección. Por tanto, si \mathbf{L} es un vector unidad con dirección hacia la fuente de luz y \mathbf{N} es el vector unidad normal a una superficie (figura 25), según la definición de producto escalar ha de ocurrir que:

$$\mathbf{N} \cdot \mathbf{L} \cos \theta = \mathbf{N} \cdot \mathbf{L}$$

y por tanto, al ser los módulos igual a 1, queda:

[17]

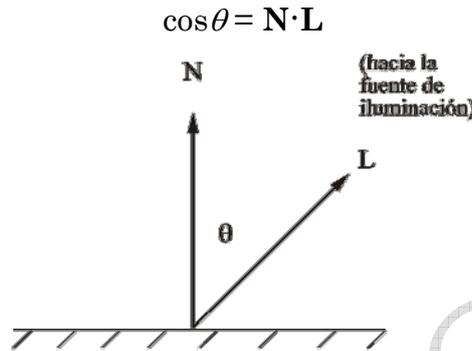


figura 25: vectores de referencia en la componente difusa

Sustituyendo en [16], queda:

[18]

$$I_d = I_i \cdot K_d \cdot \mathbf{N} \cdot \mathbf{L}$$

Si hubiese más de una fuente puntual de luz, entonces tendríamos:

[19]

$$I_d = K_d \sum_n I_{i,n} (\mathbf{N} \cdot \mathbf{L}_n)$$

siendo n el total de fuentes de luz.

Para finalizar, se ha de tener en cuenta que una superficie es iluminada por una fuente puntual sólo si el ángulo de incidencia entre ellas está entre 0 y 90 grados⁵. Cuando $\cos \theta$ es negativo, la fuente de luz está detrás de la superficie.

B) Modelización de la reflexión de la luz ambiente

Como dijimos en la pg. 17, la principal característica de la luz ambiente es que *se supone uniforme* a lo largo y ancho de toda la escena. Esta suposición (que es empírica y no del todo cierta) se debe a que, gracias a las múltiples reflexiones en las paredes y objetos circundantes somos capaces de ver éstos, aunque no reciban luz directa de las fuentes emisoras. Como calcular las reflexiones que producen la luz de fondo no está al alcance de los modelos

⁵Este hecho va a ser muy importante a la hora de aplicar criterios para acelerar el proceso de visualización, como veremos en los temas que siguen.

sencillos, se opta por la solución más viable, que es la de suponer que dicha luz es constante en todas las direcciones del espacio.

Esta uniformidad permite calcular empíricamente la luz ambiente que se refleja hacia el observador *sin tener en cuenta el ángulo de incidencia*, pues llega con igual intensidad de todas las direcciones. En definitiva, *quien determina la cantidad de luz de fondo reflejada son las propiedades ópticas de las superficies*.

La constante de reflexión para la luz ambiental (K_a) se denomina **coeficiente de reflexión ambiental**. Teniendo entonces presente la independencia sobre el ángulo de incidencia, la intensidad de luz ambiente reflejada viene dada por:

$$I_a = I_a K_a \quad [20]$$

C) Suma de las intensidades de luz reflejada

Acabamos de modelar por separado la intensidad de luz reflejada que llega al observador debido a la reflexión difusa de la luz directa, y a la reflexión de la luz ambiental. Estas dos intensidades se unen en los ojos del observador, por lo que la intensidad total que percibe será la suma de ambas. Queda por tanto:

$$I = I_a + I_d = I_a K_a + I_i K_d \text{ NL} \quad [21]$$

En los modelos de iluminación simplificados la intensidad de la luz que sale de las superficies se considera igual que la que llega al observador. Por tanto, *no se tiene en cuenta la atenuación de la luz debida a la distancia que recorre (simpl. 10)*.

D) Determinación del color de la luz reflejada

Hasta el momento hemos modelado el cálculo de las intensidades suponiendo que la luz incidente fuese toda de la misma longitud de onda. Sin embargo, vimos en [**Características de la luz**, pg. 19] que esto no es así, es decir, que a las superficies llega luz de frecuencias diferentes.

Dado que las superficies no se comportan de igual modo ante la luz con longitud de onda distinta, *es preciso modelar las propiedades ópticas de las superficies, en función de las longitudes de onda que llegan*. Afortunadamente, al utilizar modelos de color como el RGB, solamente se precisa tener en cuenta las propiedades ópticas de la materia con relación al rojo, el verde y el azul. Para ello, el coeficiente de reflexión (K) se descompone en una terna de valores escalares K_R , K_G , K_B (normalizados en el intervalo [0, 1]), cada uno de los cuales modela empíricamente el comportamiento de la materia en presencia de luz roja, verde y azul, respectivamente.

Teniendo presente el tipo de coeficiente de reflexión de que se trata y la

definición de la luz incidente, las ecuaciones [18] y [20] quedan como sigue:

$$I_{dr} = I_{iR} \cdot K_R \cdot \mathbf{N} \cdot \mathbf{L}$$

$$I_{dg} = I_{iG} \cdot K_G \cdot \mathbf{N} \cdot \mathbf{L}$$

$$I_{db} = I_{iB} \cdot K_B \cdot \mathbf{N} \cdot \mathbf{L}$$

y

$$I'_{ar} = I_{aR} \cdot K_{aR}$$

$$I'_{ag} = I_{aG} \cdot K_{aG}$$

$$I'_{ab} = I_{aB} \cdot K_{aB}$$

Según estas expresiones, la ecuación [21] puede quedar como:

$$I_{(r, g, b)} = I'_{a(R, G, B)} \cdot K_{a(R, G, B)} + I_{i(R, G, B)} \cdot K_{d(R, G, B)} \cdot \mathbf{N} \cdot \mathbf{L} \quad [22]$$

Dado que las ternas RGB que definen la luz incidente indican la intensidad de la luz de cada color primario, cada uno de los coeficientes $K_{(R, G, B)}$ puede interpretarse como *la proporción de luz que se refleja de cada uno de los colores primarios*, siendo por tanto $(1 - K_{(R, G, B)})$ la proporción de luz absorbida. Por tanto, la terna $K_{(R, G, B)}$ establece el color RGB de los objetos. De hecho, cuando se describen sus características, el color queda indicado mediante ternas de coeficientes de reflexión. Por ejemplo, si se quiere que la superficie de un objeto sea roja, podemos definir las ternas:

$$(K_R \ K_G \ K_B)_d = (1, 0, 0), \text{ y}$$

$$(K_R \ K_G \ K_B)_a = (1, 0, 0)$$

Si la luz incidente ha sido establecida como $I_a = (0,5, 0,5, 0,5)$ y $I_i = (1, 1, 1)$, y es perpendicular a la superficie ($\mathbf{N} \cdot \mathbf{L} = 1$), la luz reflejada será:

$$I_r = I_{aR} \cdot K_{aR} + I_{iR} \cdot K_{dR} = 0,5 \cdot 1 + 1 \cdot 1 = 1,5$$

$$I_g = I_{aG} \cdot K_{aG} + I_{iG} \cdot K_{dG} = 0,5 \cdot 0 + 1 \cdot 0 = 0$$

$$I_b = I_{aB} \cdot K_{aB} + I_{iB} \cdot K_{dB} = 0,5 \cdot 0 + 1 \cdot 0 = 0$$

En definitiva, una vez aplicada la ecuación [22], el color de la luz reflejada por un punto queda definido por la siguiente terna RGB:

$$(I_r, I_g, I_b)$$

En el ejemplo anterior vemos que $I_r > 1$. Como las ternas RGB han de estar normalizadas en el intervalo $[0, 1]$, cualquier valor que sobrepase los límites se le asigna el valor del límite. Otra solución mejor, aunque más costosa, consiste en calcular primero las ternas (I_r, I_g, I_b) de todos los píxeles, y después normalizarlas dividiendo por el $\text{máx}(I_r)$, $\text{máx}(I_g)$ y $\text{máx}(I_b)$. De esta manera las intensidades máximas quedan a 1, y las restantes proporcionales a ellas.

1.8.3 Modelo de Intensidad de Phong

Una versión ampliada del modelo de intensidad que acabamos de ver, que incorpora la intensidad aportada por la *componente especular*, y opcional-

mente la *atenuación* de la luz debida a la distancia, se conoce como el **Modelo de Intensidad de Phong**, el cual fue desarrollado por *Phong Bui Tuong*.

A) Reflexión Especular

Cuando la luz se refleja en las superficies sin ser absorbida y reemitida por éstas se produce la **reflexión especular**. Ésta suele percibirse como destellos de color similar al de la luz incidente. En la modelización empírica de la reflexión especular se tienen en cuenta tres factores: el *coeficiente de reflexión*, la *intensidad* y el *ángulo de incidencia*.

Para un reflector ideal (reflector perfecto), el ángulo de incidencia y el de la reflexión especular son iguales (figura 26). El vector unidad **R** representa la dirección para la reflexión especular; el vector unidad **V** apunta en la dirección del observador y el vector unidad **L** hacia la fuente de luz.

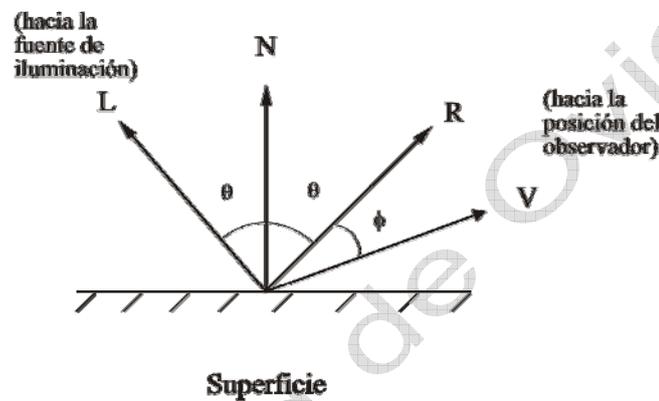


figura 26: vectores de referencia en el modelo de Phong

En un reflector perfecto, la intensidad debida a la reflexión especular llegará al observador solamente cuando **V** y **R** coincidan ($\phi = \theta$, es decir, si el ángulo de visión es igual al de reflexión).

*Los objetos reales, dado que no son reflectores perfectos, muestran la reflexión especular sobre un rango de direcciones del vector **R**.*

Así, las superficies brillantes, que se aproximan a los reflectores perfectos, tienen un campo de reflexión pequeño; en cambio, en las mates el campo es amplio (ver la figura 27)

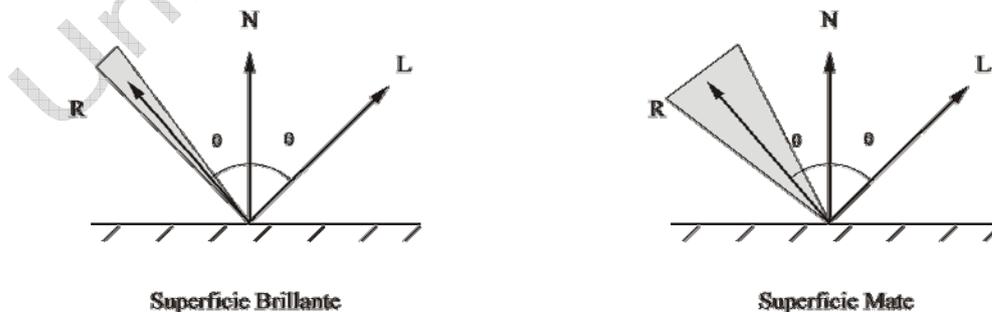


figura 27: rangos de dispersión

Dentro del espectro de trayectorias que pueden seguir los fotones después de colisionar con las superficies, las más probables son las que se encuentren más próximas a la trayectoria de reflexión teórica (**R**). Por lo tanto,

la intensidad irá disminuyendo gradualmente a medida que el ángulo de visión (ϕ) y el de reflexión (θ) difieran.

B) Modelado de la Reflexión Especular

Entre los posibles métodos para modelar el comportamiento de la reflexión especular, el presentado por Phong es uno de los más sencillos.

El **modelo de Phong** establece que la intensidad de la reflexión especular es proporcional al $\cos^n \phi$. El valor asignado a n determina el tipo de superficie que se está observando ($n \geq 1$). En las superficies brillantes n toma valores grandes (200 o mayor), mientras que en las mates los valores de n son pequeños. Para un reflector perfecto, n debería ser infinito.

La figura 28 muestra el comportamiento de $\cos^n \phi$, para distintos valores de n . Vemos que, a medida que aumenta n , disminuye el rango de valores de ϕ que permiten el paso de una intensidad de luz apreciable.

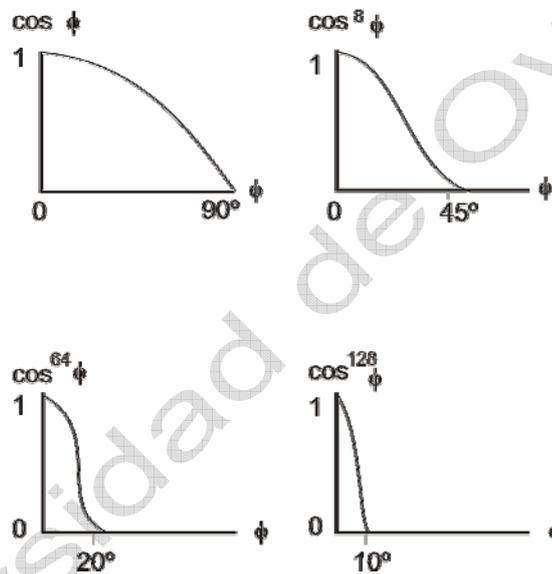


figura 28: comportamiento de $\cos^n \phi$

Como vimos, el *ángulo de incidencia* también influye en la reflexión especular. En general, la intensidad aumenta a medida que lo hace el ángulo de incidencia. Este efecto se incorpora al modelo de reflexión especular haciendo la intensidad proporcional a una función de reflexión $\mathbf{W}(\theta)$. De esta forma, el modelo completo de reflexión especular se escribe como:

[23]

$$I_e = I_i \cdot W(\theta) \cdot \cos^n \phi$$

La forma de la función $\mathbf{W}(\theta)$ depende del material de la superficie. Algunos materiales, como el cristal, muestran muy poca reflexión especular para ángulos de incidencia pequeños, aumentando a medida que θ se acerca a 90° . Para estos materiales, \mathbf{W} ha de variar desde un valor cercano a 0 hasta 1, a medida que el ángulo de incidencia varía de 0° a 90° . Por otro lado, hay materiales que tienen reflexiones especulares casi constantes para todos los ángulos de incidencia, por lo que \mathbf{W} podría tener asignado un valor cons-

tante en el intervalo [0,1].

Como \mathbf{V} y \mathbf{R} son vectores unidad en las direcciones de visión y reflexión, en la ecuación [23] ocurre que $\cos \phi = \mathbf{V} \cdot \mathbf{R}$. Además, en muchas ocasiones se puede simplificar los cálculos de intensidad sustituyendo $\mathbf{W}(\theta)$ por un valor constante K_e , llamado **Coefficiente de reflexión especular**.

Teniendo presentes las simplificaciones anteriores, si a la componente especular se le añade la componente difusa y la ambiente, puede escribirse el modelo de intensidad de Phong (para una sola fuente puntual) como:

[24]

$$I = K_a I_a + I_i \cdot [K_d (\mathbf{N} \cdot \mathbf{L}) + K_e (\mathbf{V} \cdot \mathbf{R})^n]$$

En este modelo, K_d y K_e son constantes para cada superficie iluminada.

Phong utilizó los vectores \mathbf{L} y \mathbf{N} para calcular el vector \mathbf{R} aplicando la siguiente fórmula:

[25]

$$\mathbf{R} = (2\mathbf{N} \cdot \mathbf{L})\mathbf{N} - \mathbf{L}$$

C) Reducción del tiempo de cálculo

El costo de la ecuación [24] puede reducirse considerablemente si se efectúan algunas suposiciones y aproximaciones geométricas.

En primer lugar, si la fuente de luz y el observador se consideran en el infinito, entonces \mathbf{L} y \mathbf{V} son constantes en toda la escena, lo que supone un considerable ahorro de tiempo de calculo, ya que \mathbf{L} y \mathbf{V} *han de calcularse una sola vez para todos los puntos de la escena*, mientras no varíe la posición de la fuente y/o del observador.

Otra mejora substancial en el tiempo de cálculo, consiste sustituir el vector \mathbf{R} por el **vector medio** entre el \mathbf{V} y el \mathbf{L} . A dicho vector, mostrado en la figura 29, se le denomina normalmente por la letra \mathbf{H} (*Halfway vector*).

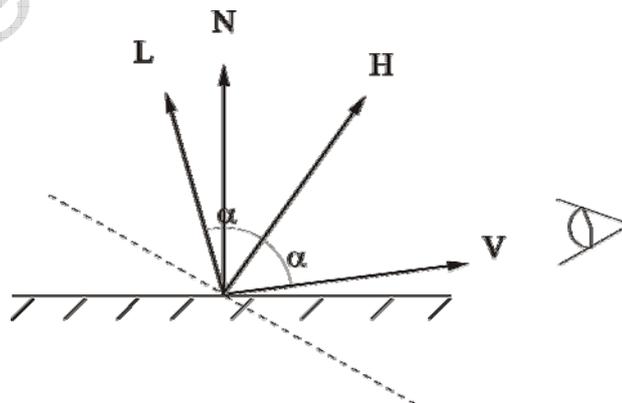


figura 29: vector \mathbf{H} , definido como el bisector del ángulo

\mathbf{H} es el vector unidad bisectriz del ángulo que forman los vectores \mathbf{V} y \mathbf{L} , y viene dado por

[26]

$$\mathbf{H} = \frac{\mathbf{L} + \mathbf{V}}{|\mathbf{L} + \mathbf{V}|}$$

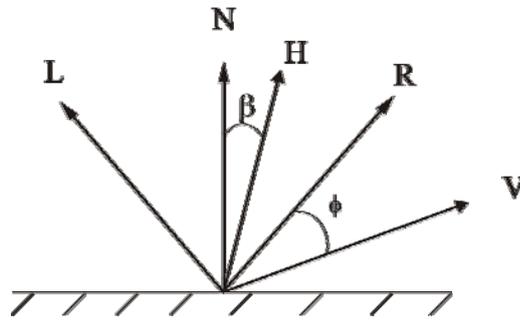


figura 30: relación entre β y ϕ

La importancia del vector \mathbf{H} radica en que, si \mathbf{L} , \mathbf{V} y \mathbf{R} son coplanares, entonces se cumple (ver figura 30) que

[27]

$$\phi = 2\beta$$

Gracias a esta relación entre los ángulos, se puede sustituir en la ecuación de Phong el producto escalar $\mathbf{V} \cdot \mathbf{R}$ por el de $\mathbf{N} \cdot \mathbf{H}$. Como el coste computacional de \mathbf{H} [26], es inferior al de \mathbf{R} [25], utilizando el vector medio se reduce el tiempo de cálculo en la ecuación de Phong. Además, y más importante, como \mathbf{H} se calcula en función de \mathbf{L} y \mathbf{V} , al ser éstos constantes, \mathbf{H} también es constante. Esto implica que, al igual que \mathbf{L} y \mathbf{V} , sólo ha de calcularse una vez por escena, mientras que el cálculo de \mathbf{R} depende de la normal a la superficie \mathbf{N} [25], por lo que ha de ser calculado para cada punto de las superficies en la escena.

Sustituyendo en la ecuación [24] $\mathbf{V} \cdot \mathbf{R} (\cos\phi)$ por $\mathbf{N} \cdot \mathbf{H} \left(\cos\frac{\phi}{2} \right)$ se obtiene

[28]

$$I = K_a I_a + I_i \cdot [K_d (\mathbf{N} \cdot \mathbf{L}) + K_e (\mathbf{N} \cdot \mathbf{H})^n]$$

Como el ángulo entre \mathbf{R} y \mathbf{V} es dos veces el ángulo entre \mathbf{N} y \mathbf{H} (figura 30), es obvio que el resultado de la ecuación de Phong utilizando $(\mathbf{V} \cdot \mathbf{R})$ no puede ser el mismo que si aplica $(\mathbf{N} \cdot \mathbf{H})$

Dado que $\cos\frac{\phi}{2} > \cos\phi$, al sustituir $(\mathbf{V} \cdot \mathbf{R})$ por $(\mathbf{N} \cdot \mathbf{H})$ la intensidad calculada aumenta. Si se quiere que el nuevo resultado se aproxime al dado por la ecuación original (la que utiliza $\mathbf{V} \cdot \mathbf{R}$), se puede compensar el incremento de intensidad aumentando el valor de n , ya que $\left(\cos\frac{\phi}{2} \right)^n < \left(\cos\phi \right)^n, n \geq 2$.

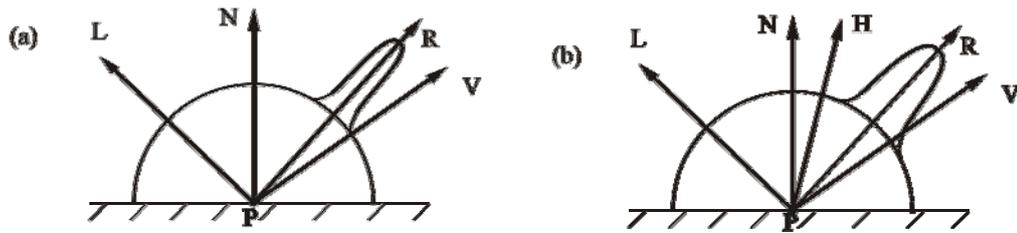


figura 31: compensación de la intensidad especular

Desde el punto de vista de la reflectividad de las superficies, la figura 31 muestra la diferencia en el reflejo para un único valor de \mathbf{n} , utilizando $\mathbf{V} \cdot \mathbf{R}$ (31-a), o $\mathbf{N} \cdot \mathbf{H}$ (31-b). Vemos que al aplicar $\mathbf{N} \cdot \mathbf{H}$ las superficies se vuelven más difusoras, por lo que para lograr el aspecto correcto (original) han de hacerse más reflectoras, lo que nos lleva a la conclusión de que se ha de aumentar n .

D) Atenuación

En los modelos de intensidad se puede añadir un parámetro que disminuya la intensidad de la luz, en función de la distancia entre la superficie y la fuente luminosa. De esta forma, superficies del mismo color se verán de distinta manera, dependiendo de la distancia a la que se encuentren de las fuentes de luz.

Si \mathbf{d} representa la distancia desde la fuente de luz a un punto de una superficie, el factor de atenuación que debería aplicarse sería $\frac{1}{d^2}$.

Sin embargo, la expresión anterior no siempre produce los efectos buscados, ya que *se producen grandes variaciones de intensidad cuando \mathbf{d} es pequeño, y poca variación cuando \mathbf{d} es grande*, debido principalmente a la simplificación de los modelos de iluminación y fuentes de luz.

Para lograr efectos más realistas, se suelen aplicar *funciones de atenuación* como:

$$f(\mathbf{d}) = \frac{1}{(\mathbf{d} + \mathbf{d}_0)}$$

donde \mathbf{d}_0 es una cte. que se incluye para calibrar la atenuación y prevenir la división por 0 cuando \mathbf{d} es muy pequeña.

Ver que para el cálculo de la reflexión especular conviene suponer que tanto *el observador como las fuentes de luz se encuentran en el infinito*. En cambio, *en el cálculo de la atenuación las fuentes se encuentran a una distancia \mathbf{d}* . Esto no supone incongruencia alguna, ya que 1) *los modelos de intensidad son empíricos*, y cualquier cosa que sirva para mejorar la calidad de las imágenes es válida, y 2) la aportación de intensidad debida a las diferentes componentes del modelo (ambiente, difusa, etc.) se calculan de modo individual, estableciendo las condiciones apropiadas independientemente en cada caso.

Otra función de atenuación, un poco más elaborada que la anterior, es la siguiente

$$f(d) = \frac{1}{a_0 + a_1 d + a_2 d^2}$$

En esta ocasión, jugando con los coeficientes a_0 , a_1 y a_2 se puede obtener una gran variedad de efectos en la escena [Hear97, pg. 506].

E) Ecuación de Phong para varias fuentes de luz

La ecuación [28] muestra el modelo de intensidad Phong (sin atenuación) para una sola fuente de luz. Veamos ahora como queda esta ecuación cuando en la escena hay dos o más fuentes, teniendo además en cuenta la función de atenuación

$$I = K_a I_a + \sum_{j=1}^m f(d_j) I_{ij} \cdot [K_d (\mathbf{N} \cdot \mathbf{L}_j) + K_e (\mathbf{N} \cdot \mathbf{H}_j)^{n_e}]$$

En la ecuación anterior, sólo los términos que tengan el subíndice j , dependen de la intensidad y/o posición de las fuentes de luz.

F) Determinación del color en el modelo de intensidad de Phong

Si a lo ya visto en *Determinación del color de la luz reflejada*, (pg. 35) añadimos las componentes introducidas en el modelo de Phong, las ecuaciones para determinar la intensidad de cada uno de los colores básicos (R, G, B) quedan como sigue

$$I_R = K_{aR} I_{aR} + \sum_{j=1}^m f(d_j) I_{iRj} \cdot [K_{dR} (\mathbf{N} \cdot \mathbf{L}_j) + K_{eR} (\mathbf{N} \cdot \mathbf{H}_j)^{n_e}]$$

$$I_G = K_{aG} I_{aG} + \sum_{j=1}^m f(d_j) I_{iGj} \cdot [K_{dG} (\mathbf{N} \cdot \mathbf{L}_j) + K_{eG} (\mathbf{N} \cdot \mathbf{H}_j)^{n_e}]$$

$$I_B = K_{aB} I_{aB} + \sum_{j=1}^m f(d_j) I_{iBj} \cdot [K_{dB} (\mathbf{N} \cdot \mathbf{L}_j) + K_{eB} (\mathbf{N} \cdot \mathbf{H}_j)^{n_e}]$$