

Capítulo 2. Generación, Representación y Principios de Procesamiento Digital de Imágenes

2.1 Cámara pinhole y formación de una imagen

Una cámara puede ser explicada de acuerdo a varios modelos, así como la formación de una imagen a partir de la cámara. El más utilizado es el modelo *pinhole*.

Una cámara *pinhole* es una cámara muy simple que no tiene lente y que tiene una sola y muy pequeña apertura, de ahí su nombre, *pinhole*, que es la traducción al inglés de la palabra “hoyo de aguja”. Esta cámara puede ser vista de otra manera como una caja a prueba de luz con un pequeño orificio en uno de sus lados, siendo ese orificio la única entrada de luz que permite la caja. Cuando la luz de una imagen pasa a través de este agujero entonces se forma una imagen invertida en el lado opuesto de la caja [8].

El ojo humano trabaja de esta misma forma y como estamos más familiarizados con la explicación del funcionamiento del ojo es posible generarse una mejor idea del funcionamiento de la cámara *pinhole* de esta manera.

En las cámaras *pinhole*, mientras más pequeño es el agujero por el que atraviesa la luz, más nítida y definida es la imagen que se forma en el interior de la misma. Sin embargo, la imagen se hace también más oscura conforme el orificio disminuye de tamaño. Puede concluirse que mientras más definida es la imagen, será más opaca.

Para el buen funcionamiento de este dispositivo se recomienda que el orificio sea del 1% o menos de la distancia entre el orificio y la pantalla (el fondo de la caja) [7,8].

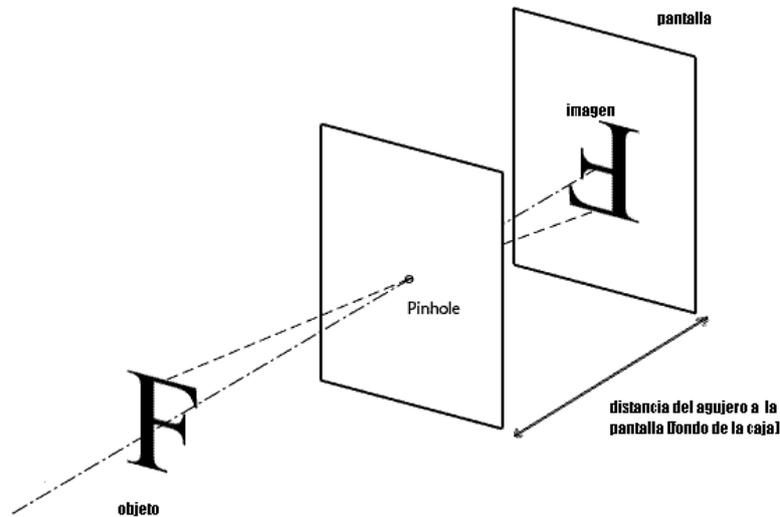


Figura 2.1 Cámara pinhole

Como puede observarse en la Figura 2.1, el comportamiento de esta cámara es similar al modelo de la formación de una imagen virtual visto en la sección 1.4.

2.2 Modelo de la cámara pinhole

El modelo de la cámara pinhole describe la relación matemática que existe entre las coordenadas de un punto en 3D y su proyección en el plano de imagen, donde la apertura de la cámara es descrita como un punto (infinitesimalmente pequeña) y no se utiliza ningún tipo de lente para enfocar la luz.

En otras palabras, este modelo de cámara puede solamente ser utilizado como una aproximación de primer orden en el trazo de un mapa de una escena 3D a una imagen 2D.

De esta manera, el modelo de la cámara pinhole es frecuentemente utilizado como una descripción razonable de cómo una cámara representa una escena en 3D, como en el caso de visión por computadora.

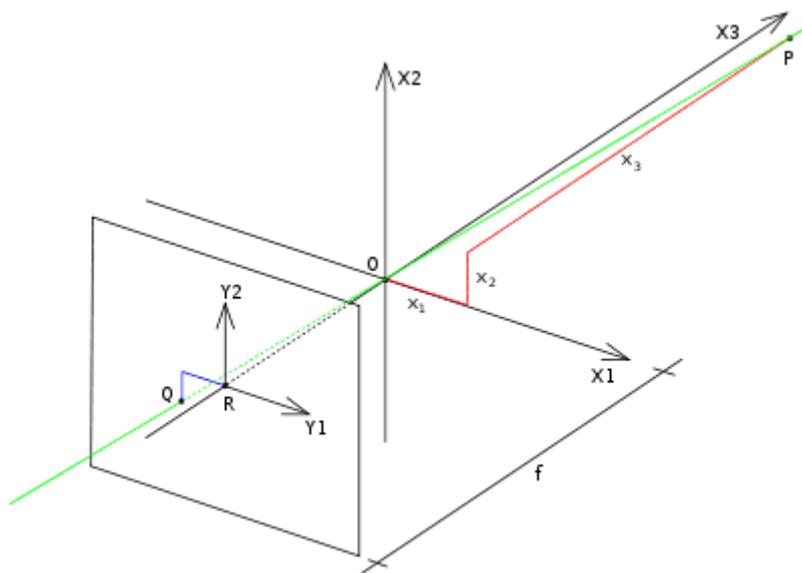


Figura 2.2 Modelo de la cámara pinhole

La Figura 2.2 muestra el modelo de la cámara pinhole, conteniendo lo siguiente:

- Un sistema coordenado en 3D con origen en O . Este es el punto en donde se encuentra localizada la apertura de la cámara, es decir, el orificio. Los ejes de este sistema son X_1 , X_2 y X_3 , apuntando este último en la dirección en la que ve la cámara y es llamado el eje óptico o eje principal.

-Un plano de imagen donde el mundo en 3D es proyectado a través de la apertura en la cámara. Este plano es paralelo a los ejes X_1 y X_2 , y se encuentra a una distancia f desde el origen O en dirección negativa del eje óptico. f es conocida como la distancia focal de la cámara pinhole.

-Un punto R en la intersección del eje óptico y el plano de imagen, llamado punto principal o centro de imagen.

-Un punto P en algún lugar en el mundo exterior con coordenadas (x_1, x_2, x_3) relativas a los ejes X_1, X_2 y X_3 .

-Una línea de proyección del punto P en la cámara, que pasa a través de los puntos P y O .

-Un punto Q que es la proyección del punto P en el plano de imagen, el cual está dado por la intersección de la línea de proyección y el plano de imagen.

-Un sistema coordenado en 2D en el plano de imagen con origen en R y con ejes Y_1 y Y_2 , los cuales son paralelos a X_1 y X_2 respectivamente. Las coordenadas del punto Q relativas a este sistema coordenado son (y_1, y_2) .

Cabe mencionar que el “hoyo de aguja” del que se habla para este modelo se considera infinitesimalmente pequeño, o sea, un punto. Éste punto en 3D, es referido en la literatura como centro óptico (de la cámara o de la lente) [9].

Ahora bien, para entender la dependencia de las coordenadas (y_1, y_2) del punto Q con las coordenadas (x_1, x_2, x_3) del punto P , véase la Figura 2.3:

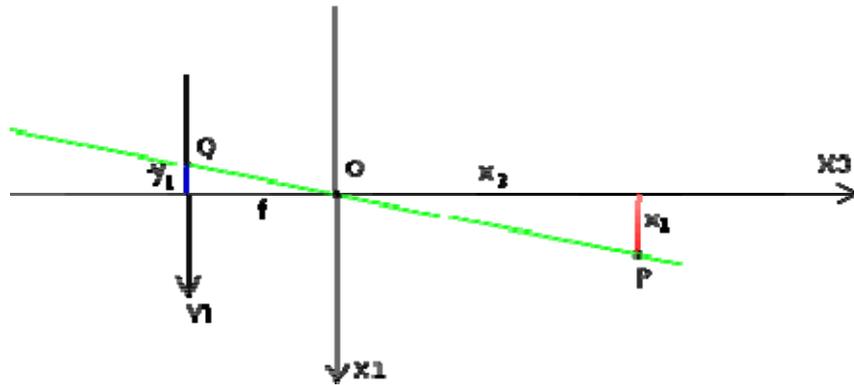


Figura 2.3 Modelo de la cámara pinhole en perspectiva del eje X2

La Figura 2.3 es la misma que la Figura 2.2 pero vista desde arriba ó en dirección negativa del eje X2.

En esta figura es posible distinguir dos triángulos similares¹ ambos compartiendo la línea de proyección como hipotenusa. Los catetos del triángulo del lado izquierdo son $-y_1$ y f y los catetos del triángulo del lado derecho son x_1 y x_3 . Como estos triángulos son similares entonces se puede decir que:

$$\frac{-y_1}{f} = \frac{x_1}{x_3} \quad \text{Ecuación 2.1}$$

Y de manera similar, viendo en la dirección negativa del eje X1, se obtiene:

$$\frac{-y_2}{f} = \frac{x_2}{x_3} \quad \text{Ecuación 2.2}$$

Y ambos resultados se pueden resumir en:

¹ Triángulos similares: Se dice que dos triángulos son similares si sus ángulos son iguales dos a dos.

$$\begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = -\frac{f}{x_3} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \quad \text{Ecuación 2.3}$$

Ésta es ya una expresión que describe la relación entre las coordenadas del punto **P** y las coordenadas del punto **Q**. Como se puede ver, esta expresión es una proyección en perspectiva pero además la sigue una rotación de 180° en el plano de imagen. Esto es normal como se nota en la sección 2.1, donde se ve la formación de la imagen en una cámara *pinhole* real. Se nota de igual forma que el tamaño de los objetos proyectados depende de la distancia al punto focal y el tamaño de la imagen depende en la distancia f entre el plano de imagen y el punto focal, como se había mencionado en la sección 1.4.

Por último, para que la imagen con la que se trabaja esté derecha, entonces se sigue algún método para voltearla. Existen varias formas de lograrlo. Si hablamos de una cámara digital, por ejemplo, sólo se necesitan leer los píxeles en determinado orden para tener la imagen en su posición original. Por ende, la ecuación que se obtiene es la misma que la ecuación 2.3, pero sin el signo negativo:

$$\begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = \frac{f}{x_3} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \quad \text{Ecuación 2.4}$$

2.2.1 Modelo de la cámara como matriz

Otra manera de ver el modelo de la cámara mencionado en la sección anterior es por medio de una matriz.

La matriz de una cámara (también conocida como matriz de proyección de una cámara) es una matriz 3×4 que describe el trazo de un mapa de los puntos en 3D en el espacio (mundo) a los puntos 2D en el plano (imagen) que hace la cámara *pinhole*.

Sea \mathbf{x} una representación de un punto 3D en coordenadas homogéneas² y sea \mathbf{y} una representación de la imagen de este punto en la cámara pinhole (un punto en 2D), entonces se puede decir que:

$$\mathbf{y} \sim \mathbf{C}\mathbf{x} \quad \text{Ecuación 2.5}$$

donde \mathbf{C} es la matriz de la cámara y \sim significa que el lado derecho e izquierdo son iguales dependiendo de una multiplicación por un escalar diferente de cero.

Como se vio en la sección anterior la relación que existe entre los puntos en el espacio y los puntos en el plano de imagen dada por el modelo de la cámara pinhole es:

$$\begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = \frac{f}{x_3} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \quad \text{Ecuación 2.4}$$

Para poder pasar del modelado anterior a una matriz, es necesario hacer uso de coordenadas homogéneas. El vector en 2D ahora se convertirá en uno 3D al agregar una coordenada extra de valor unitario. Además la igualdad se cambia

² Coordenadas Homogéneas: Aquellas en las que todo punto bidimensional, está definido en tres coordenadas. De tal forma un punto en x, y equivale a una representación en tripleta de la forma: xw, yw, w y las coordenadas x y y se hallan dividiendo entre w . Normalmente $w=1$.

por una equivalencia a multiplicación por un escalar diferente de cero como se había dicho. De esta manera se llega a la siguiente expresión:

$$\begin{pmatrix} y_1 \\ y_2 \\ 1 \end{pmatrix} \sim -\frac{f}{x_3} \begin{pmatrix} x_1 \\ x_2 \\ -\frac{x_3}{f} \end{pmatrix} \sim \begin{pmatrix} x_1 \\ x_2 \\ -\frac{x_3}{f} \end{pmatrix} \quad \text{Ecuación 2.6}$$

De igual forma, el vector que anteriormente era 3D se expresa en coordenadas homogéneas y se obtiene la siguiente expresión:

$$\begin{pmatrix} y_1 \\ y_2 \\ 1 \end{pmatrix} \sim \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & \frac{-1}{f} & 0 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ 1 \end{pmatrix} \quad \text{Ecuación 2.7}$$

donde **C** es la matriz de la cámara y en este caso está dada por:

$$\mathbf{C} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & \frac{-1}{f} & 0 \end{pmatrix} \quad \text{Ecuación 2.8}$$

Y se puede expresar también de la siguiente manera:

$$\mathbf{C} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & \frac{-1}{f} & 0 \end{pmatrix} = \begin{pmatrix} f & 0 & 0 & 0 \\ 0 & f & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix} \quad \text{Ecuación 2.9}$$

Esta correspondencia existe para cumplir con lo establecido en la ecuación 2.6.

2.3 Representación digital de una imagen en el IPT

Una imagen se puede definir como una función bidimensional, $f(x, y)$, donde x y y son las *coordenadas espaciales* (en un plano) y la magnitud de f , en cualquier par de coordenadas (x, y) , es llamada *intensidad* de la imagen en ese punto. El término *nivel de gris* es usualmente utilizado para referirse a la intensidad en imágenes monocromáticas³. Las imágenes de color están formadas por una combinación de imágenes en 2D. Por ejemplo, en el sistema de color RGB una imagen de color consiste de tres imágenes: una roja, una verde y una azul. RGB son las iniciales en inglés de *red*, *green* y *blue*, respectivamente, que, traducidos al español, son los colores anteriormente mencionados. Por esta razón, muchas de las técnicas desarrolladas para imágenes monocromáticas pueden ser extendidas a imágenes de color al procesar las tres componentes que la constituyen individualmente.

Una imagen puede ser continua con respecto a las coordenadas x y y , así como en amplitud. Convertir esta imagen a forma digital requiere que las coordenadas, así como la amplitud, sean digitalizadas. Digitalizar los valores de las coordenadas se denomina “muestreo”, mientras que digitalizar los valores de la amplitud se denomina “cuantización”. Entonces, cuando x , y y las amplitudes de f son todas cantidades finitas y discretas, la imagen se llama *imagen digital*.

³ Imágenes Monocromáticas: Son aquellas imágenes en las cuales el rango de color consiste en matices de un sólo color o un tono. Las imágenes monocromáticas en colores neutrales son también conocidas como imágenes en escala de grises o en blanco y negro.

El resultado del proceso de digitalización es una matriz de números reales. Supóngase que después de digitalizar una imagen $f(x, y)$ la imagen resultante tiene M filas y N columnas. Entonces se dice que la imagen es de tamaño $M \times N$. Los valores de las coordinas (x, y) son cantidades discretas que por notación y conveniencia son valores enteros.

El IPT usa la notación (r, c) para indicar filas (*rows*) y columnas (*columns*). Para el IPT el origen del sistema coordenado en una imagen está en $(r, c) = (1, 1)$; y así r está en el rango desde 1 hasta M y c desde 1 hasta N , en incrementos enteros. La convención mencionada se puede ver en la Figura 2.4 [22].

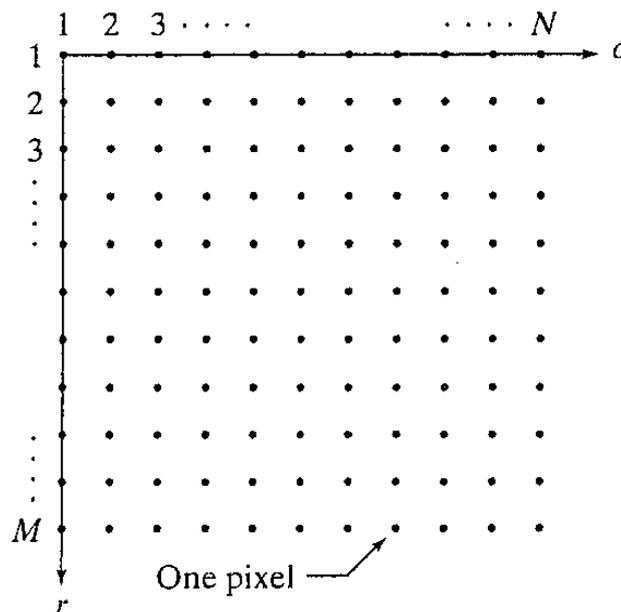


Figura 2.4 Convención para las coordenadas en el IPT

Las coordenadas de la Figura 2.4 son referidas en el IPT como “coordenadas píxel” ó “píxel”. Una imagen, entonces, puede ser naturalmente representada como una matriz en MATLAB:

$$f = \begin{bmatrix} f(1, 1) & f(1, 2) & \dots & f(1, N) \\ f(2, 1) & f(2, 2) & \dots & f(2, N) \\ \vdots & \vdots & & \vdots \\ f(M, 1) & f(M, 2) & \dots & f(M, N) \end{bmatrix}$$

Ecuación 2.6

Donde $f(1, 1)$ es el origen de la imagen y la notación $f(p, q)$ denota al elemento localizado en la fila p y en la columna q . Por ejemplo $f(8, 3)$ es el elemento en la octava fila y en la tercera columna de la matriz f [23].

2.4 Procesamiento digital de imágenes

El procesamiento digital de imágenes se refiere a procesar imágenes digitales por medio de una computadora digital. Para ser un poco más precisos, a veces se puede decir que el campo del procesamiento digital de imágenes se limita a ser una disciplina en las que tanto la entrada como la salida de un proceso son imágenes. Esta aclaración se hace debido a que las máquinas o artefactos que actualmente trabajan con imágenes son capaces de utilizar casi el completo espectro electromagnético, mientras que las imágenes visibles por el ojo humano ocupan una parte mucho menor de éste. Por lo tanto no existe un límite específico entre el procesamiento digital de imágenes y, por ejemplo, el análisis de imágenes (también llamado entendimiento o comprensión de imágenes) o la visión por computadora; sin embargo, un paradigma que puede utilizarse para unir estos tres conceptos es considerar tres niveles de procesos computarizados: bajo, medio y alto. En el nivel bajo se hacen sólo operaciones primitivas de procesamiento para reducir ruido o mejorar contraste y brillo. En este caso la entrada es una imagen y la salida sigue siendo una imagen, como es conocida generalmente por un humano. En el nivel medio, los procesos

incluyen tareas como la segmentación y la clasificación; por ejemplo reconocer objetos. En este segundo caso la entrada del proceso es una imagen pero la salida son generalmente algunos atributos extraídos de aquellas imágenes tales como límites, contornos o esquinas. Por último, el nivel alto se refiere a los procesos que incluyen “entender” los resultados obtenidos (interpretarlos) para luego aplicarles operaciones similares a las que realiza la visión humana [24]. En la presente tesis se hizo uso de un poco de los tres niveles.

2.4.1 Espacios de color RGB, L*a*b

Un espacio de color, también conocido como modelo de color, es un modelo matemático abstracto que describe cómo los colores pueden ser representados como números [32]. De esta manera es posible trabajar computacionalmente con los colores.

2.4.1.1 Modelo de color RGB

La descripción RGB (del inglés *red, green, blue*; "rojo, verde, azul") de un color, hace referencia a la composición del color en términos de la intensidad de los colores primarios con que se forma: el rojo, el verde y el azul. Es un modelo de color basado en la síntesis aditiva, con el que es posible representar un color mediante la mezcla por adición de los tres colores primarios. Para indicar con qué proporción mezclamos cada color, se asigna un valor a cada uno de los colores primarios, de manera que el valor 0 significa que un color no interviene en la mezcla y a medida que ese valor aumenta, se entiende que aporta más intensidad a la mezcla. Aunque el intervalo de valores podría ser cualquiera

(valores reales entre 0 y 1, valores enteros entre 0 y 37, etc.), es frecuente que cada color primario se codifique con un byte (8 bits). Así, de manera usual, la intensidad de cada una de las componentes se mide según una escala que va del 0 al 255.

Por lo tanto, el rojo se obtiene con (255, 0, 0), el verde con (0, 255, 0) y el azul con (0, 0, 255), obteniendo, en cada caso un color resultante monocromático. La ausencia de color (lo que se conoce como color negro) se obtiene cuando las tres componentes son 0, (0, 0, 0). La combinación de dos colores a nivel 255 con un tercero en nivel 0 da lugar a tres colores intermedios. De esta forma el amarillo es (255, 255, 0), el cyan (0, 255, 255) y el magenta (255, 0, 255). Obviamente, el color blanco se forma con los tres colores primarios a su máximo nivel (255, 255, 255). El conjunto de todos los colores se puede representar en forma de cubo. Cada color es un punto de la superficie o del interior de éste. La escala de grises estaría situada en la diagonal que une al color blanco con el negro de la Figura 2.5 [25].

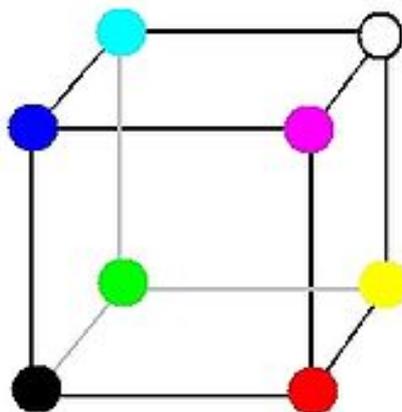


Figura 2.5 Cubo de color

Las imágenes de color que son cargadas en MATLAB generalmente se interpretan basándose en este modelo de color, por lo que al trabajar con una imagen es necesario trabajar con las tres componentes de ésta por separado.

2.4.1.2 Espacio de color L^*a^*b

A diferencia del espacio RGB, el espacio L^*a^*b intenta describir el color en términos de luminosidad, saturación y tono, que es una aproximación más cercana a cómo el ojo humano distingue el color [15]. Los tres parámetros del modelo representan específicamente la luminosidad del color L^* ($L^* = 0$ indica negro y $L^* = 100$ indica blanco), su tonalidad entre magenta y verde a^* (valores negativos indican verde mientras valores positivos indican magenta) y su tonalidad entre amarillo y azul b^* (valores negativos indican azul y valores positivos indican amarillo) como se ve en la Figura 2.6 [16].

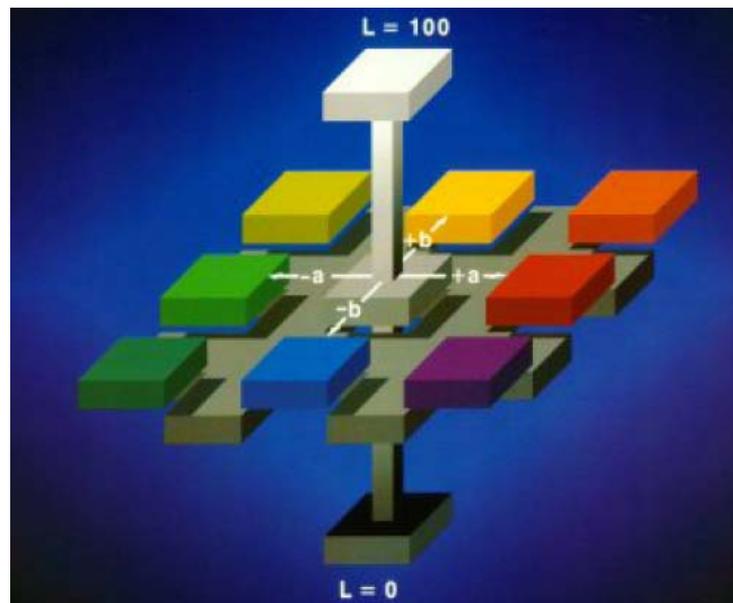


Figura 2.6 Representación del espacio L^*a^*b [17].

El espacio de color L^*a^*b nació con el propósito de producir un espacio de color que sea más “perceptivamente lineal” que otros espacios de color, es decir, que un cambio de la misma cantidad en un valor de color debe producir un cambio casi de la misma importancia visual. Contrario a RGB, los valores L^*a^*b no definen colores absolutos, sin embargo la cantidad de colores que puede representar es mayor que la que logra el espacio RGB, siendo esta una de sus mayores ventajas. En efecto, es capaz de describir prácticamente todos los colores que percibe el ojo humano [26]. Cabe mencionar que este espacio es tridimensional y sólo puede ser adecuadamente representado en un espacio tridimensional.

El modelo será de gran utilidad en esta tesis como se verá en el Capítulo 4 para detectar un objeto mediante clasificación de colores, usándolo en vez del modelo RGB que está predeterminado en MATLAB.