

## Capítulo 1

# Imágenes digitales y sus propiedades

---

### 1. Introducción

El sistema de visión humano nos permite percibir y comprender el mundo que nos rodea y nos suministra más de un 70% de la información total (imágenes, sonidos, sensaciones, etc.) que recibimos. Las escenas que percibimos suelen ser tridimensionales (3D) y cuando las capturamos mediante dispositivos (cámaras fotográficas o de vídeo, pantallas de rayos X, etc.) obtenemos imágenes bidimensionales (2D). Esta proyección a una dimensión inferior supone una gran pérdida de información. Por ello, dotar a los ordenadores de la capacidad de ver no es una tarea fácil. Las escenas dinámicas donde aparecen objetos en movimiento complican más todavía la visión por ordenador. Además, el análisis y la interpretación de muchas imágenes suele requerir un conocimiento amplio de la materia de la que trata la imagen. Por ejemplo, es difícil analizar e interpretar una imagen que muestre la capa de ozono sobre la Antártida enviada por un satélite. La conclusión cualitativa de que hay una tendencia a la disminución de la capa de ozono deducida a partir de los cambios de color en imágenes sucesivas no es tan simple y mucho menos la conclusión cuantitativa sobre cuánto esta disminuyendo.

Hay diferentes tipos de imágenes según el dispositivo utilizado para la formación de la imagen:

- Imágenes formadas mediante la *luz visible*, es decir, la luz cuya longitud de onda oscila entre 350 y 700 nanómetros, que corresponden a las imágenes capturadas por una cámara fotográfica, de vídeo o de TV.
- Imágenes formadas por *rayos infrarrojos*, que permiten la visión nocturna.
- Imágenes formadas por *rayos ultravioleta*, como **espectrogramas** (fotografía de un espectro luminoso obtenida mediante un espectrógrafo, sobre placa de cristal o película sensible).
- Imágenes formadas a partir de *campos magnéticos*, como la **resonancia magnética** utiliza un potente campo magnético para obtener imágenes detalladas del corazón o del tórax. Se coloca a la persona dentro de un gran electroimán que causa una vibración de los núcleos de los átomos del organismo, produciendo unas señales características que son convertidas en imágenes.
- Imágenes formadas a partir de *ultrasonidos*, como la **ecografía**, que forma una imagen por la reflexión de las ondas sonoras con determinadas partes del cuerpo. El **ecocardiograma**,

que utiliza ondas ultrasonoras de alta frecuencia que chocan contra las estructuras del corazón y de los vasos sanguíneos y, al rebotar, producen una imagen móvil que aparece en una pantalla de vídeo.

- Imágenes formadas a partir de una radiación de *rayos X*, como las **radiografías** de tórax o la **radioscopia** (fluoroscopia) que es una exploración continua con rayos X que muestra en una pantalla el movimiento del corazón y los pulmones. La tomografía computarizada que crea imágenes transversales de todo el tórax utilizando los rayos X y muestra la ubicación exacta de las anomalías (permite obtener una imagen tridimensional móvil del corazón). La **angiografía** muestra con detalle el aporte de sangre, por ejemplo, a los pulmones (se inyecta un líquido radiopaco dentro de un vaso sanguíneo que puede verse en las radiografías) o al cerebro. Detecta anomalías vasculares como la obstrucción de un vaso sanguíneo (ictus), aneurismas (bolsas en una arteria) o arteritis (inflamaciones).
- Imágenes formadas a partir de *impulsos eléctricos*, como el **electrocardiograma**, que amplifica los impulsos eléctricos del corazón y se registran en un papel en movimiento. El **electroencefalograma** es una imagen formada a partir de los impulsos eléctricos de 20 alambres (electrodos) colocados sobre el cuero cabelludo con el objeto de establecer el trazado y registro eléctrico de la actividad cerebral.
- Imágenes formadas a partir de *isótopos radiactivos* (indicadores o trazadores). Los indicadores se reparten por todo el cuerpo y se detectan con una gammacámara (detección de coágulos de sangre en los pulmones). En la técnica de **tomografía computarizada** por emisión de fotones simples, distintos tipos de cámaras de registro de radiaciones pueden grabar una imagen simple o producir una serie de imágenes de secciones transversales amplificadas por el ordenador. En las imágenes formadas por *emisión de positrones*, como la **tomografía** por *emisión de positrones*, se inyecta una sustancia en la sangre que se desplaza hasta las estructuras cerebrales, permitiendo medir la actividad que desarrolla el cerebro. También permite analizar, por ejemplo, el funcionamiento del corazón.

## 2. Conceptos básicos

Cualquiera de estas imágenes se puede modelar por una función continua de dos o tres variables. En el caso de imágenes estáticas los dos argumentos de la función son las coordenadas (x,y) del plano cartesiano, mientras que si las imágenes van cambiando en el tiempo (dinámicas) necesitamos una tercera coordenada t que nos especifica el tiempo. Los valores de la función corresponden a la luminosidad, brillo o intensidad de la radiación de los puntos de la imagen. También pueden representar temperatura, presión, distancia al observador, etc.

$$\text{punto de la imagen} \equiv (x, y, t) \xrightarrow{f} f(x, y, t) \equiv \text{intensidad luminosa}$$

Una función continua tiene el dominio y el rango continuo; una función discreta tiene el dominio discreto y la **función digital** tiene tanto el dominio como el rango discretos.

Una imagen 2D es el resultado de una **proyección** de una escena 3D. Un punto de la escena 3D viene representado por su proyección en el plano imagen. Así, el punto  $P=(x,y,z)$  se proyecta en el punto  $P'=(x',y',f)$  cuyas coordenadas vienen dadas por las expresiones:

$$x' = \frac{xf}{z}, \quad y' = \frac{yf}{z}$$

que se obtienen directamente del teorema de Thales (ver la figura 1).

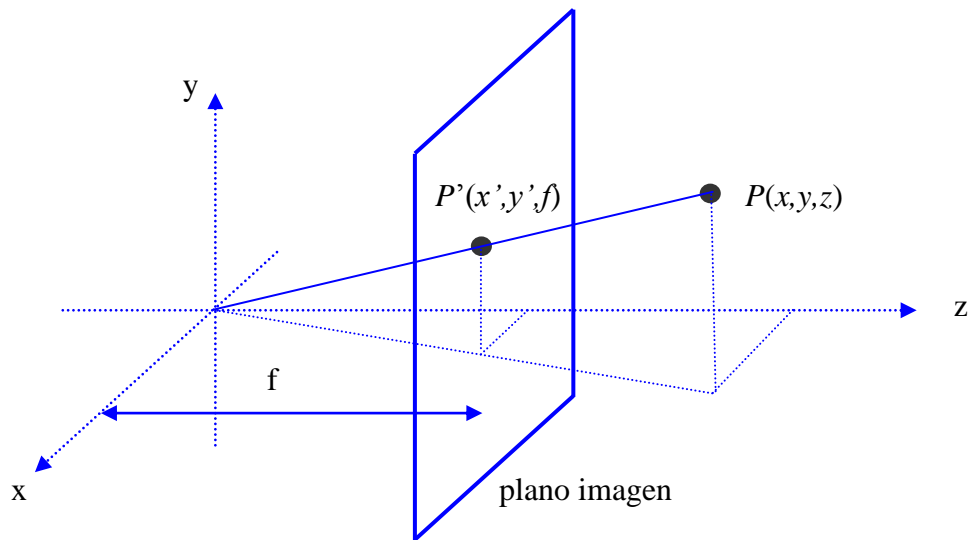


Figura 1. Proyección sobre el plano imagen.

Un **píxel** de una imagen digitalizada es un punto del dominio de la función digital correspondiente que tiene asociado el valor de dicha función y cuya posición viene determinada por sus coordenadas  $(x, y)$ . En el análisis de imágenes juegan un papel importante los píxeles de un entorno (píxel próximos) de cada píxel. Por ello, es necesario definir una función distancia entre píxeles.

La **distancia Euclídea** entre los píxeles  $(i, j)$  y  $(h, k)$  viene dada por la expresión:

$$D_E[(i, j), (h, k)] = \sqrt{(h-i)^2 + (k-j)^2}$$

La **distancia rectangular** entre los píxeles  $(i, j)$  y  $(h, k)$  viene dada por la expresión:

$$D_4[(i,j), (h,k)] = |h-i| + |k-j|$$

Si se permiten movimientos en diagonal, como en el tablero de ajedrez, la distancia entre los píxeles  $(i,j)$  y  $(h,k)$  viene dada por la expresión:

$$D_8[(i,j), (h,k)] = \max\{|h-i|, |k-j|\},$$

y se llama distancia de Tchebychev o del tablero de ajedrez.

En una **imagen monocromática** (en blanco y negro o en tonos de gris) el rango de  $f$  tiene un valor mínimo (negro) y un valor máximo (blanco). Si se trata de una imagen digital el rango viene dado por un conjunto finito de valores, como puede ser el conjunto  $\{0, 1, \dots, L-1, L\}$ , donde el 0 corresponde al negro y el valor  $L$  a blanco. En este caso diremos que la imagen tiene  $L+1$  niveles o **tonos de gris**. El dominio de la función digital es también una región acotada del plano,  $D$ , de la forma:

$$D = \{ (x, y): x=1,2,\dots,M, y=1,2,\dots,N \}.$$

En la figura 2 se muestra una imagen monocromática que corresponde a un vaso capilar de una mamografía (región de interés).

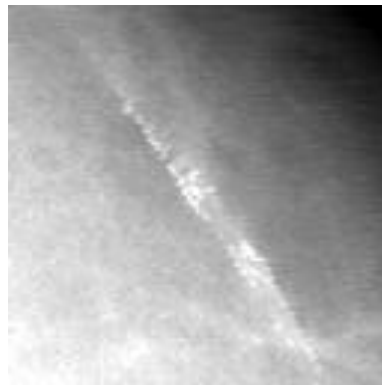


Figura 2. Región de interés en una mamografía.

Por lo tanto, una imagen digital monocromática viene dada por una matriz  $f$  de tamaño  $M \times N$ :

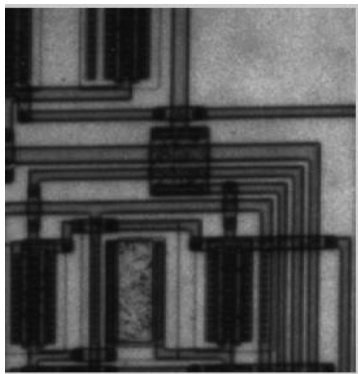
$$f = \begin{bmatrix} f(1,1) & f(1,2) & \dots & f(1,N) \\ f(2,1) & f(2,2) & \dots & f(2,N) \\ \dots & \dots & \dots & \dots \\ f(M,1) & f(M,2) & \dots & f(M,N) \end{bmatrix}$$

Cada uno de sus elementos se llama *píxel* (picture element) o *pel* y  $f(m, n)$  nos da el tono de gris del píxel que ocupa la posición  $(m, n)$ .

La **resolución espacial** viene dada por la proximidad de las muestras de la imagen en el plano imagen. En el caso de imágenes digitalizadas es el número de puntos del dominio de la imagen, es decir,  $M \times N$ . La **resolución espectral** viene dada por el ancho de banda de las frecuencias de la radiación luminosa capturadas por el sensor. La **resolución radiométrica** (o

profundidad) viene dada por el número de niveles o tonos de gris distinguibles y la **resolución temporal** viene dada por el intervalo de tiempo entre muestras consecutivas en las que la imagen se captura.

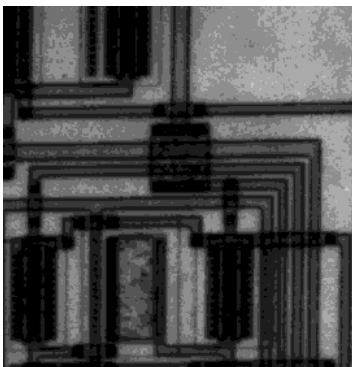
En la figura 3(a) se muestra una imagen con resolución espacial (tamaño)  $280 \times 272$  y con 256 tonos de gris (resolución radiométrica); en la figura 3(b) se muestra dicha imagen con una resolución espacial de  $28 \times 72$  píxeles y 256 tonos de gris; en la figura 3(d) se muestra una ampliación de la imagen 3(b) para apreciar la pérdida de calidad y en la figura 3(c) se muestra la imagen con una resolución espacial de  $280 \times 272$  píxeles y sólo 10 tonos de gris.



(a)



(b)



(c)



(d)

Figura 3. (a) Imagen de un circuito con  $280 \times 272$  píxeles y 256 tonos de gris. (b) Imagen de un circuito con una resolución espacial de sólo  $28 \times 27$  píxeles y 256 tonos de gris. (c) Imagen de un circuito con  $280 \times 272$  píxeles y sólo 10 tonos de gris. (d) Imagen (b) ampliada.

El **color** es una propiedad de gran importancia en la percepción visual humana. El color está asociado con la capacidad de los objetos de reflejar ondas electromagnéticas de diferente longitud de onda. El ser humano detecta los colores como combinaciones de tres colores primarios, el rojo, el verde y el azul. Las pantallas de nuestros ordenadores suelen utilizar el modelo **RGB** (red-green-blue), donde cada píxel tiene asociado un vector tridimensional  $(r, g, b)$  que nos conduce a un color determinado; el vector  $(0, 0, 0)$  es el negro,  $(L, L, L)$  es el blanco,  $(L, 0, 0)$  es el rojo “puro”, y así sucesivamente (ver la figura 3). Esto implica que tenemos  $L^3$

colores diferentes que no todos serán albergados por el dispositivo, por lo que es común especificar un subconjunto de estos, llamado **paleta** de colores.

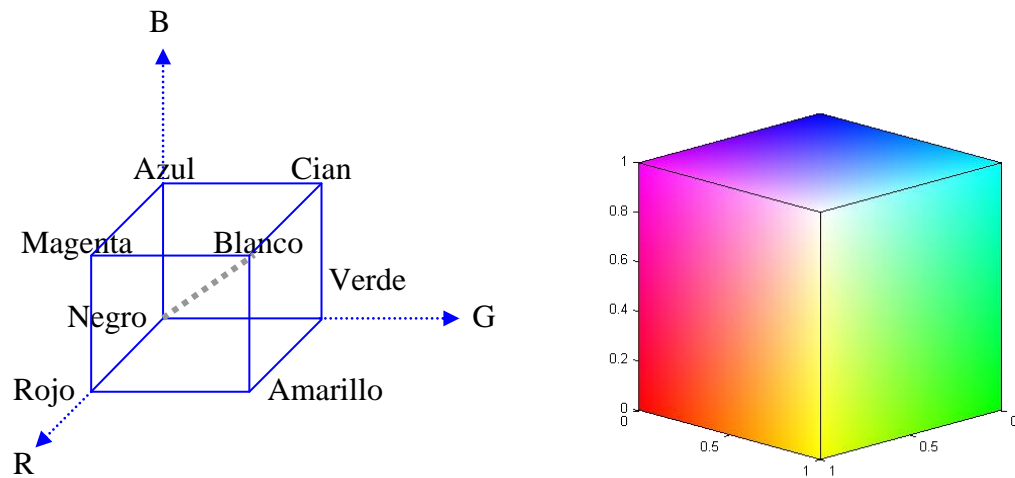


Figura 3. Espacio de colores RGB.

Los colores rojo, verde y azul, que corresponde a vértices de cubo son los *colores primarios* de la luz y los colores cian, magenta y amarillo de los otros tres vértices son los *colores secundarios*.

Una **imagen de color** RGB viene definida por tres matrices de tamaño  $M \times N$ , de manera que cada píxel corresponde a una posición específica en las mismas y tiene asociado una tripleta de valores,  $(r, g, b)$ , que indican la tonalidad corresponde a cada uno de los colores rojo, verde y azul. La primera matriz nos da la componente roja, la segunda la componente verde y la tercera la componente azul. En la figura 4 se muestran las tres componentes de una imagen en color.



Figura 4. (a) Imagen en color. (b) Componente roja. (c) Componente verde. (d) Componente azul.

Cuando se da como entrada a un monitor estas tres matrices (imágenes monocromáticas) se genera la imagen en color. Por lo tanto, una imagen en color viene dada por una función vectorial de la forma.

$$f(x,y)=(f_1(x,y),f_2(x,y),f_3(x,y))$$

donde cada componente corresponde a un color determinado.

También se puede representar una imagen mediante dos componentes: una matriz  $X$  de tamaño  $M \times N$  constituida por valores enteros del conjunto  $\{1,2,\dots,p\}$  y una matriz, llamada *mapa de colores*, de tamaño  $p \times 3$ , constituida por valores reales del intervalo  $[0,1]$ . El número  $p$  nos da el número de colores diferentes a utilizar. Cada fila de la matriz *mapa de colores* especifica la componente roja, verde y azul de un color específico. Cada elemento de la matriz  $X$  especifica una fila de la matriz del mapa de colores, es decir, un color específico. Por lo tanto, los elementos de la matriz  $X$  son apuntadores hacia el mapa de colores. Una imagen así representada se dice que es una **imagen indexada**.

### 3. Modelos o espacios de color

Un color se puede representar de dos maneras diferentes dependiendo del medio en que se reproduce:

a) Mediante un **sistema aditivo** que consiste en añadir colores al negro para crear nuevos colores. La presencia de todos los colores primarios es suficiente para crear un blanco puro, mientras que la ausencia de los colores primarios crea un negro puro. Los colores primarios de luz o colores primarios aditivos son el rojo, el verde y el azul. La suma de los tres colores primarios da lugar al blanco ( la luz blanca está compuesta por luz roja, verde y azul a partes iguales). Los colores secundarios de la luz son el magenta (rojo + azul), el cian (verde + azul) y el amarillo (rojo + verde).

b) Mediante un **sistema sustractivo** en el que los colores primarios son sustraídos del blanco para formar nuevos colores. Cuantos más colores se mezclen se obtendrá algo más parecido al negro. Teóricamente, la presencia de todos los colores primarios sustractivos nos da el negro y su ausencia total el blanco. Un color primario de pigmento (sustractivo) es aquel que absorbe un color primario de luz y refleja los otros dos, por tanto, son el cian, el magenta y el amarillo. Cuando la luz incide sobre un color primario de pigmento lo que vemos es la combinación de los otros dos colores primarios de luz. Así, además del modelo RGB podemos tener otros modelos de color.

El espacio de color **CMY** (Cian, Magenta y Amarillo) se basa en los colores secundarios de la luz, es decir, en los colores primarios de pigmentos. La luz blanca pura se obtiene como diferencia entre el amarillo y el azul. la mayoría de los dispositivos que depositan pigmentos coloreados sobre el papel, como las impresoras o las fotocopadoras, utilizan entrada de datos en la configuración CMY o realizan una conversión de RGB a CMY mediante la transformación simple:

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

donde se supone que todos los colores han sido normalizados en el rango [0,1].

Esta ecuación muestra que la luz reflejada en una superficie pintada de color cian puro no contiene rojo ( $C = 1 - R$ ); la reflejada en una superficie pintada de color magenta puro no contiene el verde ( $C = 1 - R$ ) y la reflejada una superficie pintada de color amarillo puro no contiene el azul ( $C = 1 - R$ ).

Teóricamente, con igual cantidad de pigmentos primarios, cian, magenta y amarillo, se producirá el color negro, pero en realidad, debido a las impurezas en las tintas, da lugar a un color café pardo; de ahí que para conseguir un color negro de calidad es conveniente añadir este color a las impresoras y fotocopadoras, pues es el que más utilizan, y se obtiene así el espacio de color CMYK.

El sistema de color **NTSC** se usa para la televisión en Estados Unidos. Una de las principales ventajas de este formato es que la información en tonos de gris se separa de los datos de color, y así la misma señal se puede utilizar para televisiones en blanco y negro y televisiones en color. En este sistema la imagen consta de tres componentes: La *luminancia* ( $Y$ ), el *matiz* o *tonalidad* ( $I$ ) y la *saturación* ( $Q$ ). La elección de las letras **YIQ** es convencional. La luminancia, luminosidad o brillo es el flujo luminoso o intensidad luminosa emitido por un objeto con luz propia (emitida y no reflejada), tal como una bombilla o el Sol. El *matiz* o *tonalidad* es un atributo asociado con la longitud de onda dominante en una mezcla de ondas de luz y representa el color percibido por el observador. Cuando llamamos a un objeto naranja, azul o marrón estamos especificando su tono. La *Saturación* se refiere a cómo de puro es el color, es decir, la cantidad de luz blanca que se mezcla con él. Se parte del color blanco hasta llegar al color totalmente saturado (ver figura 5).



Figura 5. Escala de saturación del color rojo.



La luminancia representa la información en tonos de gris y las otras dos componentes contienen la información de color en una señal de televisión. El modelo **YIQ** se obtiene a partir del modelo RGB mediante la siguiente transformación que es una transformación lineal:

$$\begin{pmatrix} Y \\ I \\ Q \end{pmatrix} = \begin{pmatrix} 0.299 & 0.587 & 0.114 \\ 0.596 & -0.275 & -0.321 \\ 0.212 & -0.523 & 0.311 \end{pmatrix} \begin{pmatrix} R \\ G \\ B \end{pmatrix}$$

El Sistema de color **PAL** (Phase Alternation Line), o modelo **YUV**, es el utilizado para la televisión en Europa. Fue desarrollado por los Laboratorios de Telefunken en Hannover (Alemania). En el año 1967, la República Federal Alemana y Reino Unido comenzaron a usarlo y posteriormente lo implantaron la mayoría de los países europeos que comenzaron a transmitir televisión en color, y algunos de Sudamérica. Las coordenadas que transmite son la *luminancia*  $Y$ , y dos coordenadas denominadas  $U$  y  $V$ , derivadas de las señales diferencia de color  $R-Y$  y  $B-Y$ . Se obtiene del modelo RGB mediante la transformación lineal:

$$\begin{aligned} Y &= 0.30 R + 0.59 G + 0.11 B \\ U &= 0.493 (B-Y) \\ V &= 0.877 (R-Y) \end{aligned}$$

El espacio de color alternativo de más relevancia para el procesamiento de imágenes es el modelo **HSI** (Hue, Saturation e Intensity). Los espacios de color hasta ahora estudiados no son demasiado adecuados para describir los colores en términos de su interpretación humana. Por ejemplo, nadie describe el color de su coche en términos de la composición de sus colores primarios. El ser humano suele describir un objeto de color por su brillo, su tonalidad y su saturación. Sin embargo, el brillo es una descripción subjetiva que es difícil de medir. Por ello, en este modelo se separa el brillo de la información sobre el color (el matiz y la saturación corresponden a la percepción humana). Como consecuencia, este modelo es una herramienta muy útil para el desarrollo de algoritmos en procesamiento de imágenes. Existe una transformación no lineal para pasar de RGB a HSI.

Las imágenes digitales se guardan como ficheros según un cierto formato, como puede ser el formato TIFF, JPEG, JPG, GIF, BMP, PNG, XWD, etc.