

Master: Ciencia de Datos e Ingeniería de Computadores

Curso: Visión por Computador

Rosa M^a Rodríguez Sánchez

Dpto. Ciencias de la Computación e I. A.

E.T.S. de Ingenierías Informática y de Telecomunicaciones

Universidad de Granada



Introducción a los sistemas de visión por Computador

Índice de contenido

1. Introducción.....	2
2. Elementos de la Percepción Visual.....	2
2.1. El ojo.....	3
3. Luz, Color y el espectro Electromagnético.....	4
3.1. La luz y el espectro electromagnético.....	4
3.2. Tipos de Imágenes.....	5
3.3. La luz y la percepción del color.....	6
3.4. Codificación del color y representación.....	8
3.4.1. Matiz.....	8
3.4.2. Saturación.....	9
4. El proceso de formación de imágenes.....	9
4.1. Sensores de Imágenes.....	9
4.2. Óptica de la Cámara.....	10
4.3. Digitalización de la imagen.....	11
4.3.1. Muestreo.....	12
4.3.2. Cuantificación.....	14
5. Representación de Imágenes Digitales.....	14
5.1. Imágenes Binarias.....	14
5.2. Imágenes de Color.....	15
5.2.1. Representación RGB.....	15
5.2.2. Representación Indexada.....	16
6. Formatos de Imágenes.....	16
7. Componentes de un sistema de procesamiento de imágenes.....	17
8. Bibliografía.....	18

1. Introducción

En este tema, vamos a describir una imagen como una representación de una escena del mundo 3D mediante una matriz 2D. Para que dicha imagen ocurra es necesario una fuente de luz que ilumine la escena. Desde este punto de vista una *imagen digital* es una representación de una imagen 2D usando un número finito de píxeles.

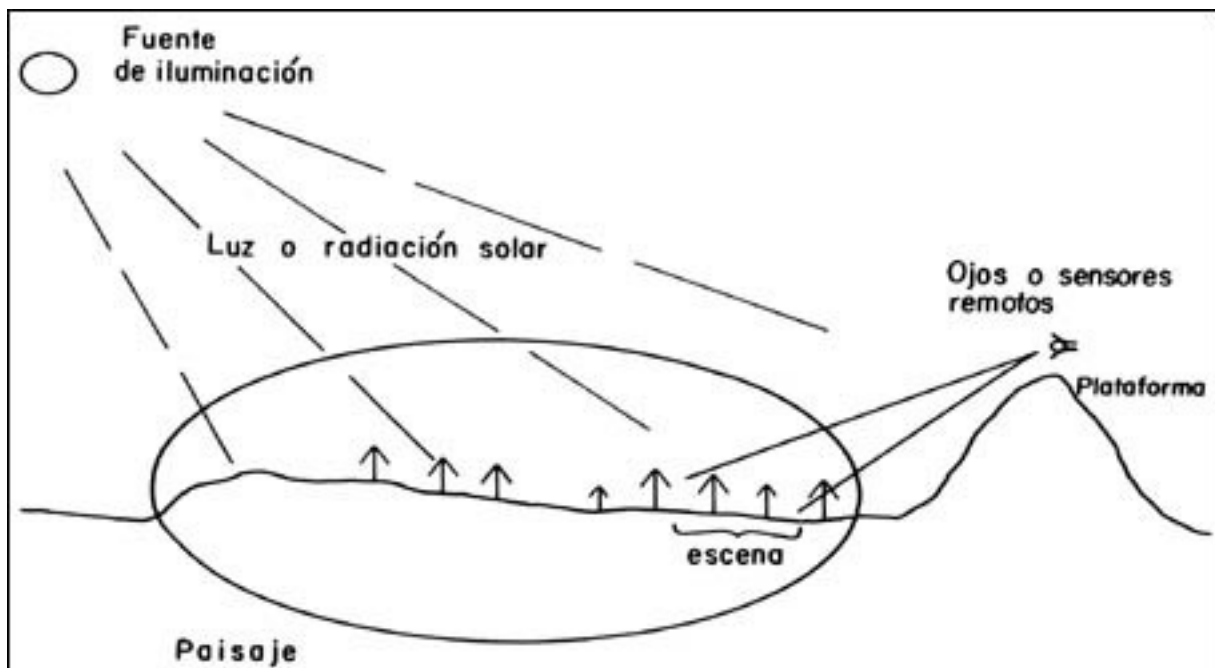
En primer lugar analizaremos los aspectos más relevantes involucrados en la adquisición y digitalización de una imagen:

- Los principios para formar una imagen, como efecto colateral de la reflexión de la luz sobre un objeto o escena.
- Sensores usados para capturar la energía reflejada
- Aspectos relacionados en el número óptimo de muestras y niveles de cuantificación

2. Elementos de la Percepción Visual

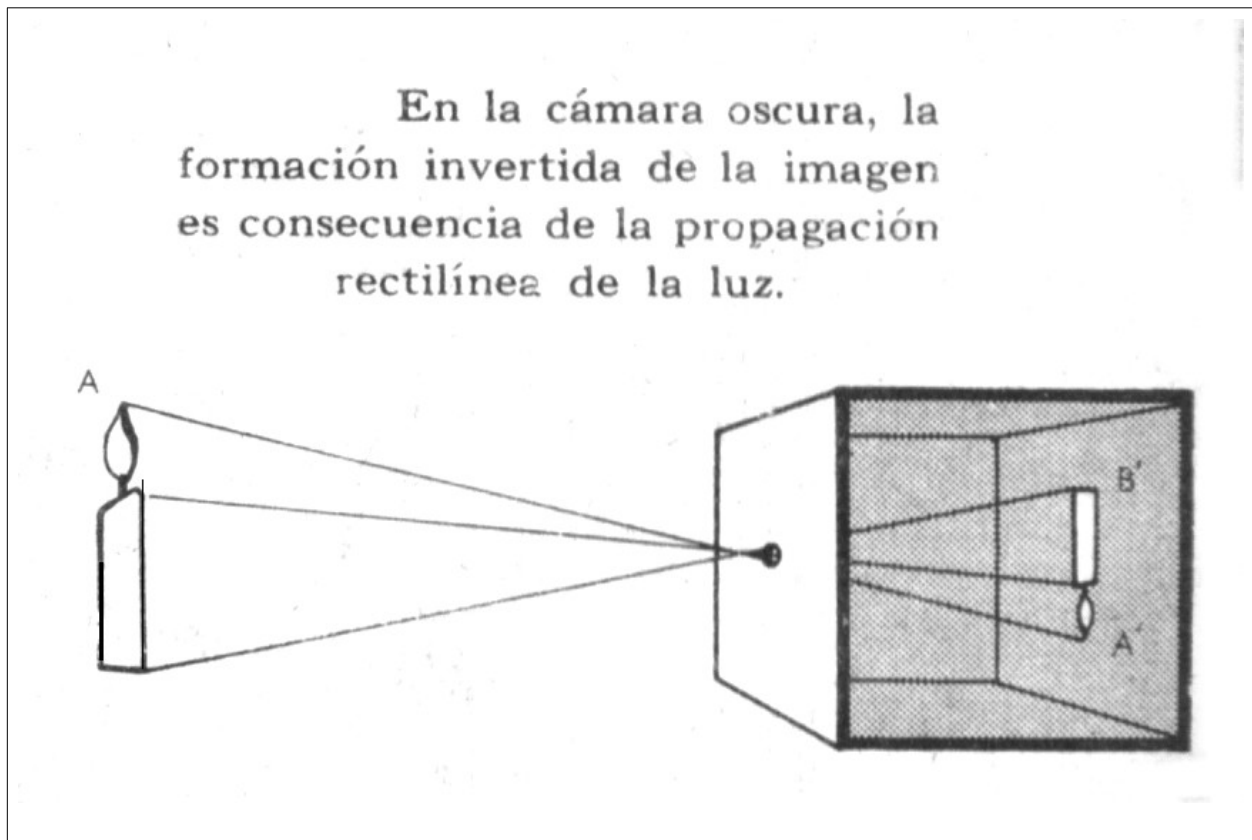
La percepción es el proceso que nos permite, a través de los sentidos, recibir, interpretar y elaborar información que proviene de nuestro entorno. Por lo tanto la *Percepción Visual* es el proceso mediante el cual captamos e interpretamos las imágenes. Para ello tenemos que tener en cuenta los siguientes aspectos:

- Para poder registrar una imagen deben darse condiciones de tipo físico (luz, tipo de iluminación, atmósfera, distancia, movimiento, etc) y de la existencia de un órgano o dispositivo sensible a la luz (el ojo, o cámara). Si estamos a oscuras o tenemos los ojos cerrados no vemos, no es posible la percepción visual.
- La interpretación de los datos visuales captados por el ojo la realiza el cerebro o computador.



2.1. El ojo

El ojo es el órgano de la visión. Funciona de un modo similar a una cámara fotográfica. Ambos se basan en el principio de la cámara oscura.



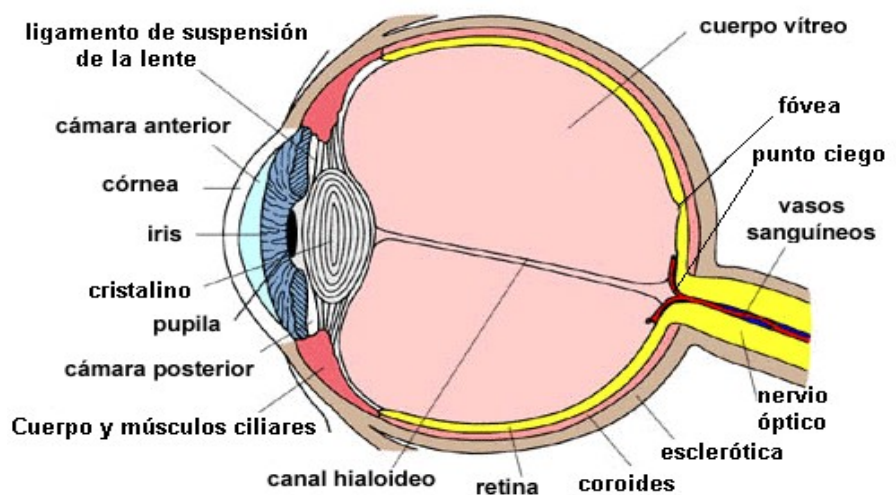
Ambos sistemas tienen mecanismos para regular la entrada de la luz, enfocar la imagen y una superficie recubierta de sustancias sensibles a la luz.

Nosotros tenemos dos ojos, nuestra visión es binocular así que, además de enfocar, tenemos que ajustar la dirección en la que mira cada ojo para que se dirijan al mismo punto y nuestro cerebro debe combinar las imágenes que envía cada ojo, que no son idénticas, y elaborar sólo una que es tridimensional y que nos permite reconocer el volumen de los objetos y la distancia a la que están situados.

El ojo tiene forma aproximadamente esférica y está rodeado por una membrana llamada **esclerótica** que por la parte anterior se hace transparente para formar la **córnea**.

Tras la córnea hay un diafragma, el **iris**, que posee una abertura, la **pupila**, por la que pasa la luz hacia el interior del ojo. El iris es el que define el color de nuestros ojos y el que controla automáticamente el diámetro de la pupila para regular la intensidad luminosa que recibe el ojo.

El **cristalino** está



unido por ligamentos al **músculo ciliar**. De esta manera el ojo queda dividido en dos partes: la posterior que contiene **humor vítreo** y la anterior que contiene **humor acuoso**. El índice de refracción del cristalino es 1,437 y los del humor acuoso y humor vítreo son similares al del agua.

El cristalino enfoca las imágenes sobre la envoltura interna del ojo, la **retina**. Esta envoltura contiene fibras nerviosas (prolongaciones del nervio óptico) que terminan en unas pequeñas estructuras denominadas conos y bastones muy sensibles a la luz. Existe un punto en la retina, llamado **fóvea**, alrededor del cual hay una zona que sólo tiene conos (para ver el color). La fóvea es la parte más sensible de la retina y sobre ella se forma la imagen del objeto que miramos.

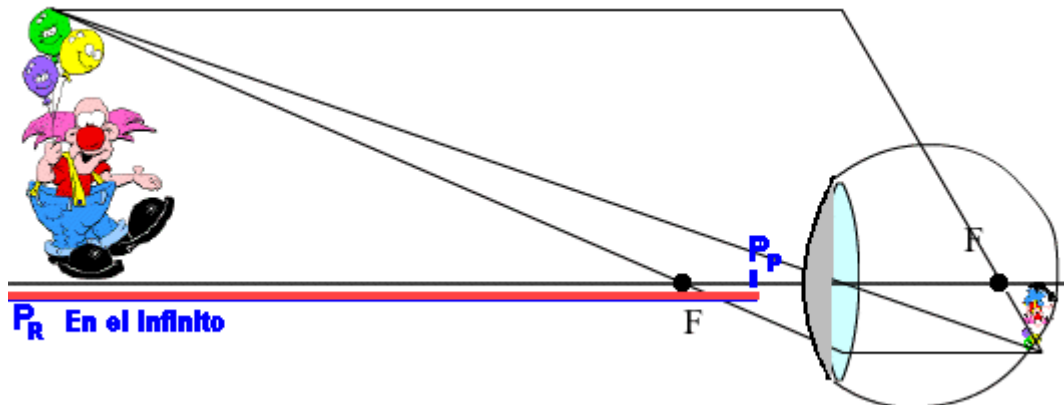
Los millones de nervios que van al cerebro se combinan para formar un **nervio óptico** que sale de la retina por un punto que no contiene células receptores. Es el llamado **punto ciego**.

La córnea refracta los rayos luminosos y el cristalino actúa como ajuste para enfocar objetos situados a diferentes distancias. De esto se encargan los músculos ciliares que modifican la curvatura de la lente y cambian su potencia. Para enfocar un objeto que está próximo, es decir, para que la imagen se forme en la retina, los músculos ciliares se contraen, y el grosor del cristalino aumenta, acortando la distancia focal imagen. Por el contrario si el objeto está distante los músculos ciliares se relajan y la lente adelgaza. Este ajuste se denomina **acomodación o adaptación**.

El ojo sano y normal ve los objetos situados en el infinito sin acomodación enfocados en la retina. Esto quiere decir que el foco está en la retina y el llamado punto remoto (P_r) está en el infinito.

Se llama **punto remoto** la distancia máxima a la que puede estar situado un objeto para que una persona lo distinga claramente y **punto próximo** a la distancia mínima.

Un ojo normal será el que tiene un punto próximo a una distancia " d " de 25 cm, (para un niño puede ser de 10 cm) y un punto remoto situado en el infinito. Si no cumple estos requisitos el ojo tiene algún defecto.



3. Luz, Color y el espectro Electromagnético

La existencia de luz o de otras formas de radiación electromagnética- es un requerimiento esencial para que una imagen pueda ser creada, captura y percibida.

3.1. La luz y el espectro electromagnético

La luz puede ser descrita en términos de ondas electromagnéticas o partículas, llamadas

fotones. Un fotón es un pequeño paquete de energía electromagnética que puede ser caracterizado por su longitud de onda o frecuencia. Las longitudes de onda se miden normalmente en metros. La frecuencia se mide en hercios (Hz). La longitud de onda (λ) y la frecuencia (f) están relacionados mediante la siguiente expresión:

$$\lambda = \frac{v}{f}$$

siendo v la velocidad con la que la onda viaja (normalmente se aproxima a la velocidad de la luz $c=2.998 \times 10^8 \text{ m/s}$).

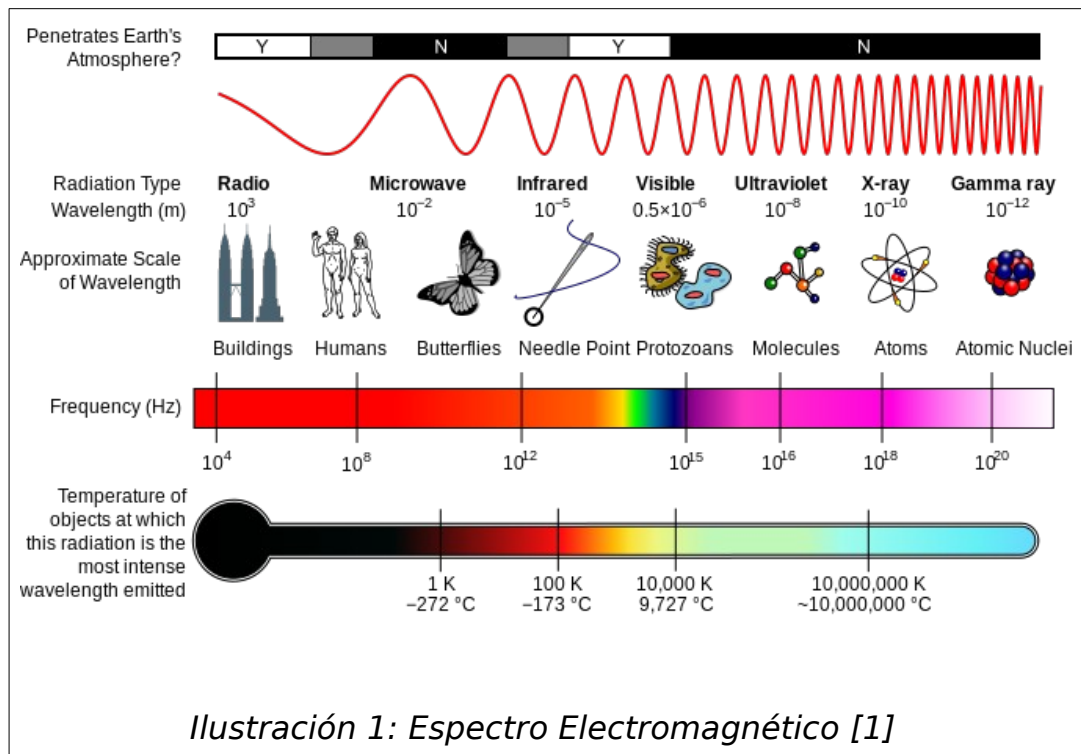
El sistema visual humano (HVS) es sensible a fotones de longitudes de onda entre 400 a 700nm ($1\text{nm}=10^{-9}\text{m}$). Como se puede observar en la Figura 1 es una porción del espectro muy pequeño. El rango del espectro abarca desde las ondas radio (con un longitud de onda de 1m o más) hasta los rayos gamma (con longitudes de onda de 0.01nm o menos).

Muchas aplicaciones (p.e médicas, militares) trabajan con imágenes que han sido capturadas con sensores especializados para obtener datos del espectro no visible.

3.2. Tipos de Imágenes

Las imágenes pueden clasificarse en tres categorías de acuerdo al el tipo de interacción entre la fuente de la radiación y la posición relativa del sensor de imagen:

- Imágenes por reflexión. Estas imágenes es el resultado de la radiación que ha sido reflejada desde la superficie de los objetos. La radiación puede ser ambiental o artificial. La mayoría de las imágenes que percibimos son imágenes reflejadas. Las características que se pueden extraer de este tipo de imágenes son por ejemplo: forma, color y textura.
- Imágenes por emisión. Estas son el resultado de objetos que emiten luminosidad (son fuentes de luz, p.e las estrellas) en el rango del espectro visible e incluso más allá del rango visible como son las imágenes termale o infrarrojas.
- Imágenes por absorción. Son el resultado de la radiación que pasa a través de un objeto y este proceso da como resultado una imagen que produce información acerca de la estructura interna del objeto. Las imágenes más comunes de este tipo son las imágenes de rayos-X.



3.3. La luz y la percepción del color

Los colores percibidos por los humanos se determinan por la naturaleza de la luz reflejada por el objeto. Esta luz reflejada está en función de las propiedades espectrales de la fuente de luz así como de la luz absorbida y reflejada por el objeto.

En el año 1666, Sir Isaac Newton descubrió que un rayo de luz pasando a través de un prisma se descompone en un espectro continuo de componentes. Cada una de esas componentes produce una percepción de color diferente, que está entre el rojo hasta el violeta. Así la luz del sol se compone de diferentes "colores" de luz.

La radiación de un fuente de luz se expresa en términos de su distribución de potencia espectral (SPD).

La percepción del color empieza con un fuente de luz cromática, capaz de emitir una radiación electromagnética con longitudes de onda entre 400 a 700nm. Parte de esa radiación se refleja sobre la superficie de los objetos en una escena y la luz reflejada alcanza el ojo humano, dando lugar a la sensación o percepción del color. Un objeto que refleja luz de manera equitativa en todas las longitudes de ondas en el espectro visible se percibe como blanco, mientras que un objeto que absorbe la mayoría de luz que le llega, se percibe como negro. La percepción de los diferentes niveles de gris, entre blanco y negro, se refiere como una percepción acromática o falta de color. El rango del espectro que un objeto refleja determina el nombre del color que etiqueta al objeto. Por ejemplo, un objeto que refleja la mayor parte de energía entre longitudes de onda comprendidas entre 565-590 nm se considera amarillo.

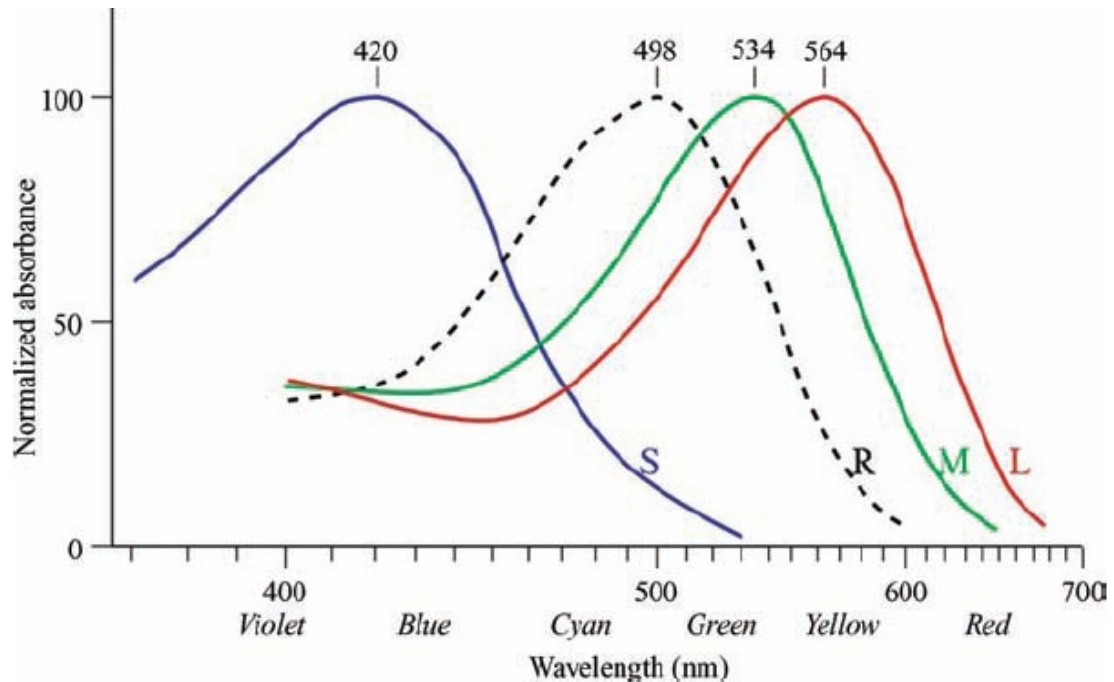
Una luz cromática puede ser descrita por tres cantidades:

- Intensidad: la cantidad total de energía que fluye desde la fuente de luz (se mide en vatios (W)).
- Luminancia: la cantidad de información que un observador percibe de la fuente de luz, se miden en lumen (lm).
- Brillo. La percepción subjetiva de la intensidad luminosa (acromática).

La retina humana esta formada por receptores fotosensitivos de dos tipos diferentes: conos y bastones (rods). Los bastones no responden al color pero si responden a niveles de luminancia baja y se activan incluso en condiciones de oscuridad. Los conos son los responsables de que veamos color y operan solamente bajo condiciones de brillo (en palabras cotidianas diríamos que tiene que haber claridad para percibir color).

Existen tres tipos de conos:

- Conos L: se activan a longitudes de onda de alrededor de 610nm.
- Conos M: se activan a longitudes de onda entorno a 560nm
- Conos S: se activan con longitudes de onda alrededor de 430nm



La existencia de tres tipos especializados de conos en el ojo humano fue confirmado experimentalmente por Thomas Young con su teoría tri-cromática de la visión en 1802. La teoría de Young explica solamente parte del proceso de visión del color. Por ejemplo no explica, por qué es posible calificar a un objeto de color verde azulado pero no amarillo azulado. Esto fue explicado por Edward Hering en 1872, con su teoría de la oponencia del color. Los colores a los que los conos responden mas fuertemente son los colores primarios de luz y han sido estandarizados por la CIE (Commission Internationale de L'Éclairage). Así el rojo 700nm, verde 546.1nm y azul 435.8nm.

Los colores secundarios de luz, se obtienen como mezclas aditivas de los primarios, dos colores en cada momento, estos son: el magenta o púrpura=red + azul, cyan (o turquesa)=azul + verde y el amarillo=verde+rojo (ver Figura 2(a)).

Cuando usamos lapices, o acuarelas, los colores

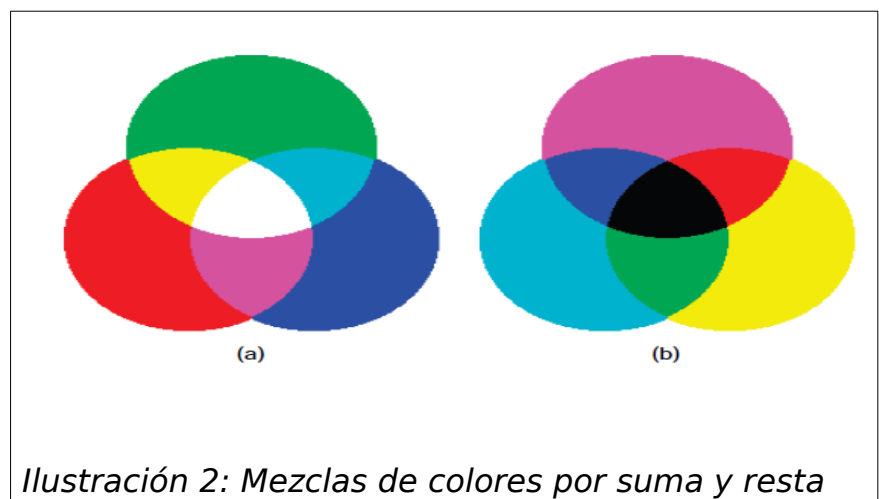


Ilustración 2: Mezclas de colores por suma y resta

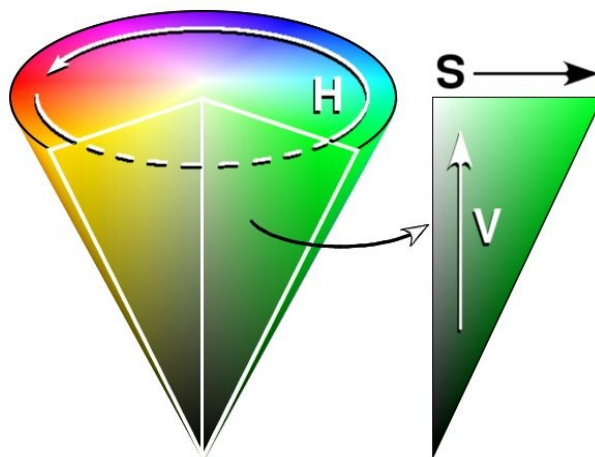
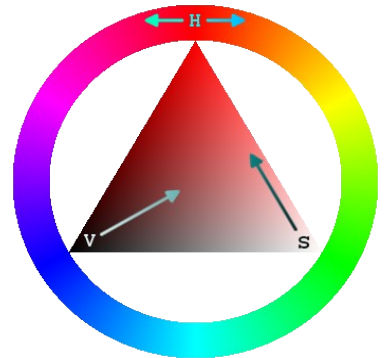
primarios son el magenta, cyan y amarillo y los secundarios son el rojo, verde y azul (ver Figura 2(b)). Es importante destacar que cuando usamos acuarelas un color se nombra después de la porción del espectro que absorbe, mientras que por la luz un color se define basado sobre la porción del espectro que emite. Consecuentemente, mezclar los tres colores primarios de luz resulta el blanco, mientras que mezclar los tres color primarios según las acuarelas, obtenemos negro (todos los colores han sido absorbidos, y no se refleja ninguna luz).

3.4. Codificación del color y representación

El color puede ser codificado usando tres componentes. La colorimetría es la ciencia que trata con el estudio cuantitativo de la percepción del color. Este estudio se centra en la representación de los valores de tres estímulos, desde los que se puede derivar la percepción del color. La forma más simple de codificar el color en cámaras y pantallas es usando los valores del nivel de rojo, verde y azul de cada pixel (sistema RGB).

Los humanos perciben la luz, y consecuentemente el color, describiendo tres parámetros (HSV):

- Brillo (V): La percepción subjetiva de la intensidad de luminosidad (un objeto aparece mas o menos luminoso).
- Matiz (Hue H). A que color es más similar: rojo, verde o azul o una combinación de estos.
- Saturación (S). Si es un color más brillante o menos. P.e rojo claro o rojo oscuro.



3.4.1. Matiz o Hue (color)

Se representa como un grado de ángulo cuyos valores posibles van de 0 a 360° (aunque para algunas aplicaciones se normalizan del 0 al 100%). Cada valor corresponde a un color. Ejemplos: 0 es rojo, 60 es amarillo y 120 es verde.

De forma intuitiva se puede realizar la siguiente transformación para conocer los valores básicos RGB:

Disponemos de 360 grados dónde se dividen los 3 colores RGB, eso da un total de 120° por color, sabiendo esto podemos recordar que el 0 es rojo RGB(1, 0, 0), 120 es verde RGB(0, 1, 0) y 240 es azul RGB(0, 0, 1). Para colores mixtos se utilizan los grados intermedios, el amarillo, RGB(1, 1, 0) está entre rojo y verde, por lo tanto 60°. Se puede observar como se sigue la secuencia de sumar 60 grados y añadir un 1 o quitar el anterior:

- $0^\circ = \text{RGB}(1, 0, 0)$
- $60^\circ = \text{RGB}(1, 1, 0)$
- $120^\circ = \text{RGB}(0, 1, 0)$
- $180^\circ = \text{RGB}(0, 1, 1)$
- $240^\circ = \text{RGB}(0, 0, 1)$
- $300^\circ = \text{RGB}(1, 0, 1)$
- $360^\circ = 0^\circ$

Esta transformación permite saber los tonos de matices de colores puros que contienen alguna cantidad (o ninguna) de los colores R, G y B. Para el color blanco se puede poner cualquier color y saturación, siempre que se establezca el valor de brillo (o luminosidad) máximo. Asimismo, para el color negro se puede poner cualquier color y saturación, siempre que se ponga un valor de 0 en el brillo.

3.4.2. Saturación

Se representa como la distancia al eje de brillo negro-blanco. Los valores posibles van del 0 al 100%. A este parámetro también se le suele llamar "pureza" ([pureza colorimétrica](#) de la colorimetría). Cuanto menor sea la saturación de un color, mayor tonalidad grisácea habrá y más decolorado estará.

4. El proceso de formación de imágenes

Las imágenes digitales se forman mediante la captación de cierta energía por un dispositivo. Esas medidas se toman en varios puntos a lo largo de un retículo (matriz) 2D para formar la imagen. Los dispositivos se denominan sensores. Los sensores pueden responder (o activarse) a diferentes rangos de luz del espectro electromagnético, energía acústica, rayos de electrones, láser, o cualquier otra señal que pueda ser medida.

4.1. Sensores de Imágenes

El principal objetivo de un sensor de imagen es convertir la energía electromagnética en señales eléctricas que pueden ser procesadas, visualizadas, e interpretadas como imágenes. La forma en la que este proceso se realiza es variable en función de la tecnología usada.

Dos de los dispositivos más usados para la captación de imágenes son las cámaras digitales y los escáneres. Las cámaras normalmente usan sensores 2D CCD (en inglés **Charge-Coupled Devices**), mientras que los escáner 1D CCDs que se mueven a lo largo de la imagen fila a fila.

Un sensor CCD está formado por una matriz de células sensibles a la luz llamadas "*photosites*". Cada célula produce un voltaje proporcional a la intensidad de la luz que caen en el área de la célula. Un "*photosite*" tiene una capacidad finita de la cantidad de luz a la que puede responder (alrededor 10^6), lo que impone un límite superior del brillo de los objetos a ser visualizados. Un "*photosite*" saturado puede desbordarse, corrompiendo su entorno y causando un defecto llamado fluorescencia ("*blooming*").

La **resolución nominal** de un sensor CCD es el número de elementos de la escena que se asocian con un único pixel de la imagen. Por ejemplo, si un cuadrado de 20x20 cm forma la escena y queremos visualizarla en una imagen digital de 500x500 entonces la resolución nominal es de 0.04cm.

El **campo de vista (FOV)** de un sensor de imagen se mide como el área de la escena que puede ver, por ejemplo, 10cmx10cm. Ya que esto puede variar debido a la profundidad, normalmente es más informativo expresarlo con el ángulo de visión por ejemplo, $55^\circ \times 40^\circ$. El ángulo de visión de una cámara es una función del tamaño de la superficie que registra la imagen y de la longitud focal y la distorsión de la lente.

Normalmente las cámaras CCD incorporan una tarjeta, memoria, denominada frame buffer,

siendo esta una memoria muy rápida y almacena imágenes capturadas por la cámara. Después de ser capturadas y temporalmente almacenadas en la memoria, las imágenes pueden ser procesadas o copiadas a un dispositivo de memoria no volátil.

4.2. Óptica de la Cámara

Una cámara usa una lente para enfocar parte de la escena sobre el sensor de imagen. Los parámetros más importantes de una lente es la cantidad de aumento y su capacidad de capturar luz. La cantidad de aumento se especifica por el factor de aumento (m), que se define como la razón entre el tamaño de imagen y el tamaño del objeto (ver Figura 3):

$$m = \frac{v}{s}$$

siendo s la distancia del objeto a la lente y v la distancia de la lente al plano de imagen.

El poder de aumento de una lente se expresa en términos de su distancia focal, f (en mm), que se define como la distancia desde la lente a el punto al cual el paralelo incidente converge (se cruza con el eje óptico) dando lugar a la siguiente igualdad:

$$\frac{1}{f} = \frac{1}{s} + \frac{1}{v}$$

Combinando las dos ecuaciones resulta que f se puede expresar como

$$f = \frac{sm}{m+1}$$

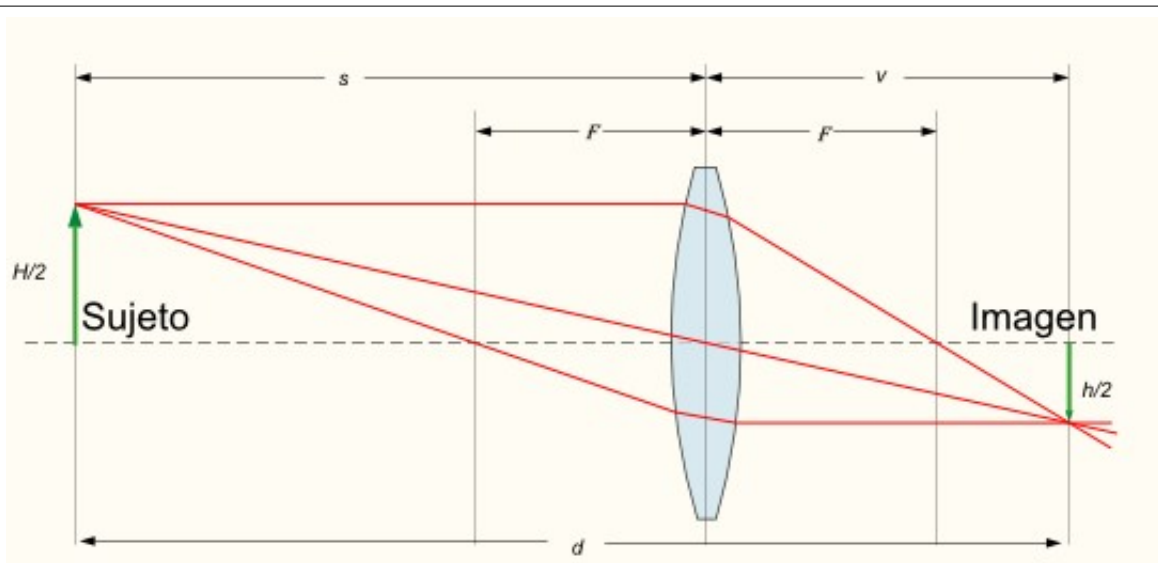


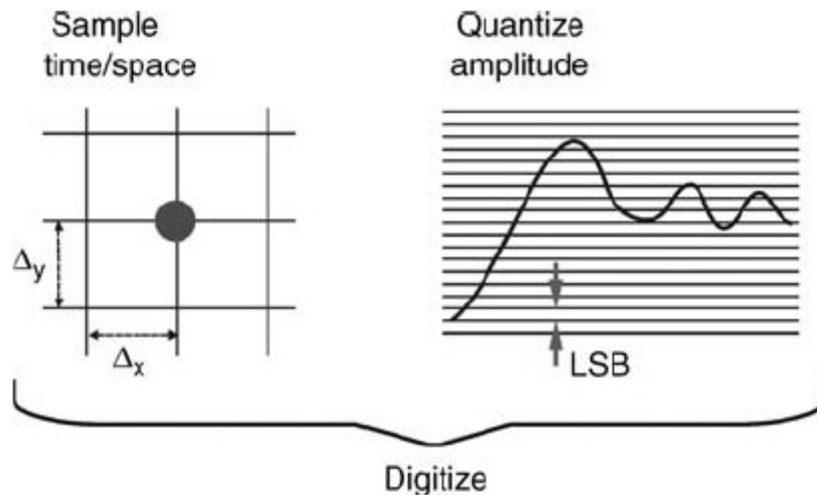
Ilustración 3: Formación de imágenes usando una lente

La capacidad de capturar luz una lente de cámara se determina por su apertura que se expresa normalmente por la razón entre la longitud focal y el diámetro de apertura.

Las lentes pueden sufrir de aberraciones, las cuales afectan a la calidad de la imagen y generan distorsiones no deseables en la imágenes resultantes.

4.3. Digitalización de la imagen

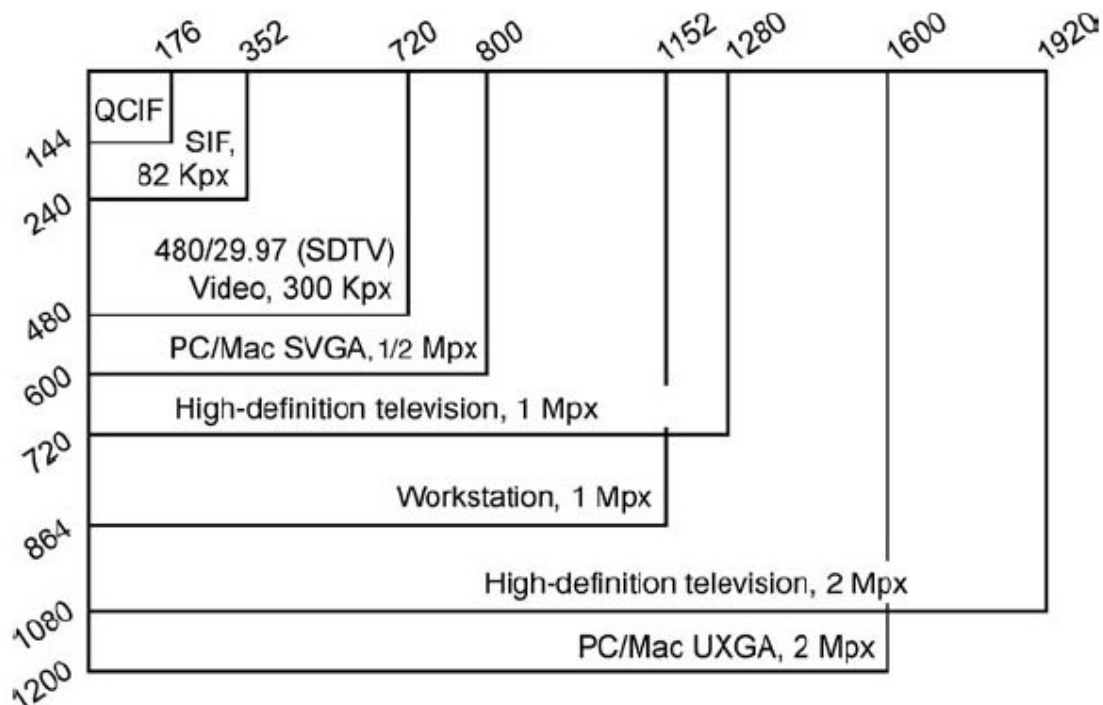
La etapa de la digitalización establece el puente entre el mundo real analógico, desde el cual las escenas son adquiridas, y el formato digital que son manipulables por los algoritmos en los computadores.



El proceso de digitalización involucra dos procesos: Muestreo y Cuantificación.

El muestreo implica seleccionar un número finito de puntos en un intervalo, mientras que la cuantificación implica asignar valores de amplitud (en un rango de posibles valores) a cada punto. El resultado de la digitalización es una matriz de píxeles, cuyos valores corresponden a sus intensidades o a sus componentes de color (imágenes de color).

Es común representar el tamaño de la matriz como el producto del numero de píxeles en cada dimensión. En la siguiente figura se muestra los tamaños para diferentes formatos:



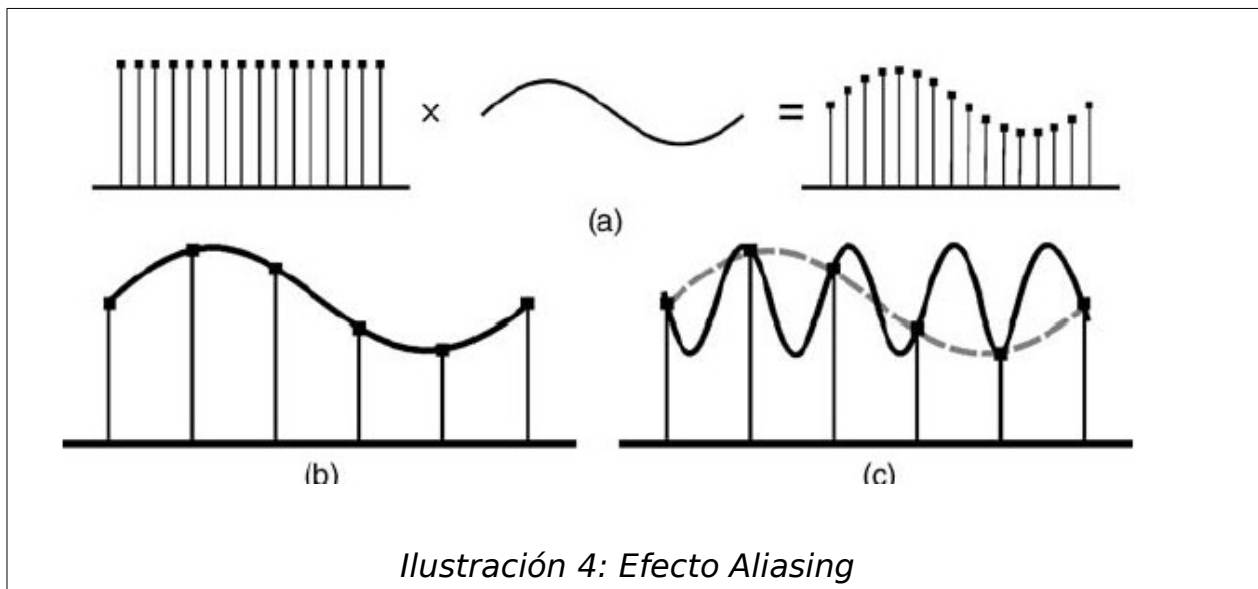
4.3.1. Muestreo

Los parámetros que se deben tener en cuenta cuando se muestrea una imagen son:

1. La razón de muestreo es el numero de muestras tomadas a lo largo del alto y ancho de la imagen. Si la razón es muy alta la calidad de la imagen será mayor. Si la razón no es suficiente puede dar lugar al conocido problema de aliasing (pérdida de información).
2. El patrón de muestreo es la forma física de las muestras. El patrón rectangular es el más común, en el cual los píxeles son alineados horizontal y verticalmente. Otros tipos de formas son la hexagonal y la polar.

Si el muestreo se realiza a una razón más baja que 2 veces la frecuencia más alta de la señal (criterio de Nyquist), no habrá suficientes muestras (puntos) para asegurar una reconstrucción sin pérdida de la señal.

En la Figura 4 se muestra el efecto que produce el aliasing sobre una señal 1d: en (a) se muestra como el producto del tren de impulsos de muestras y la señal analógica que es muestreada; en (b) se puede ver el resultado de reconstruir la señal con un numero apropiado de muestras; y en (c) se observa que si usamos el mismo numero de muestras cuando la señal tiene un frecuencia más alta es insuficiente para hacer una buena reconstrucción de la señal.



El efecto de aliasing sobre imágenes se percibe en la forma del patrón de Moiré. En la Figura 5 se ve este patrón sobre todo en el mantel (d).

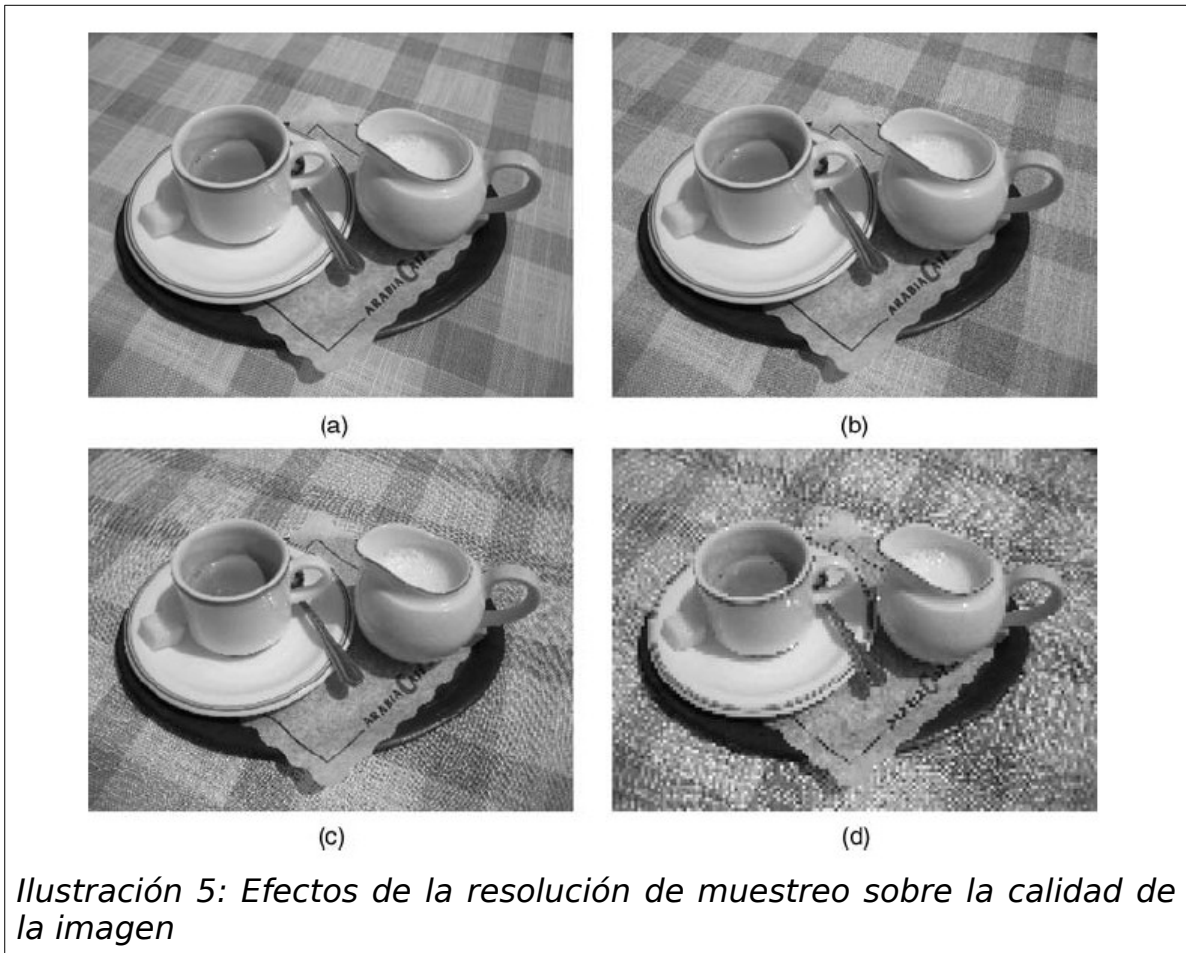


Ilustración 5: Efectos de la resolución de muestreo sobre la calidad de la imagen

4.3.2. Cuantificación

La cuantificación es el proceso de sustituir una función continua con un conjunto discreto de niveles (niveles de cuantificación). Para imágenes de niveles de gris normalmente los niveles de cuantificación van de [0-255]. Siendo 0 el negro y 255 el valor blanco.

5. Representación de Imágenes Digitales

Un imagen digital - que fue obtenida mediante un proceso de muestreo y cuantificación de una señal analógica- puede ser representada como una matriz 2D de números reales. Así $f(x,y)$ se refiere a una imagen, vamos a suponer que es monocroma (p.e una imagen de niveles de gris), de tamaño $M \times N$, donde x denota las filas e y las columna. Así

$$f(x,y) = \begin{pmatrix} f(1,1) & f(1,2) & \dots & f(1,N) \\ f(2,1) & f(2,2) & \dots & f(2,N) \\ \vdots & \vdots & \ddots & \vdots \\ f(M,1) & f(M,2) & \dots & f(M,N) \end{pmatrix}$$

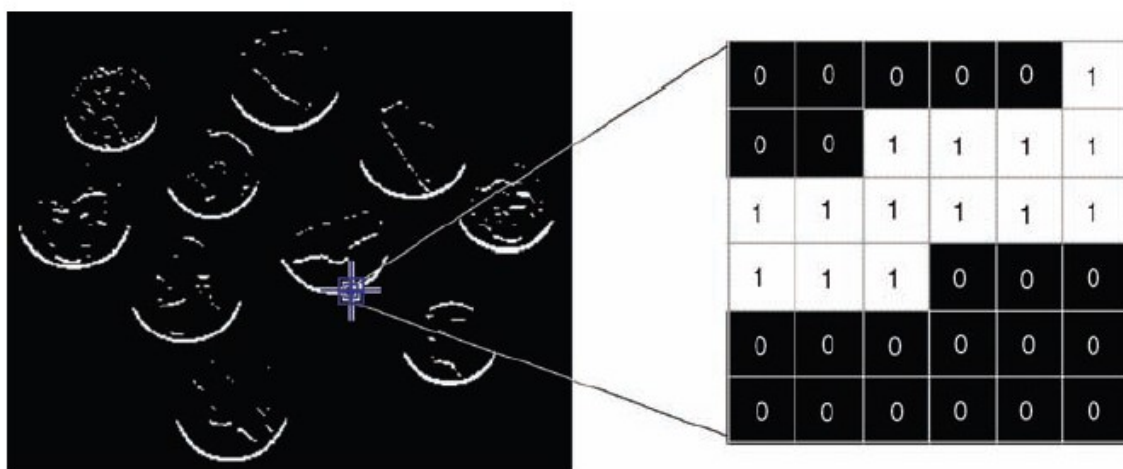
El valor de la función $f(x,y)$ en un pixel (x_0, y_0) se define como el nivel de gris o intensidad de la imagen en ese pixel. El valor máximo y mínimo que pueda adoptar un pixel depende del tipo de dato que se haya usado. Los rangos más comunes son: 0.0 (negro) a (1.0) blanco para imágenes de tipo doble o 0 (negro) y 255 (blanco) para imágenes de tipo unsigned char.

Las imágenes se representan de forma digital de diferentes formas. A un nivel muy básico,

existen dos formas de codificar las imágenes digitales 2D: raster or bitmap y vectorial. La representación bitmap usa uno o más de dos matrices de píxeles. Mientras que la representación vectorial usa una serie de comandos de dibujo para representar la imagen. Cada método tiene sus ventajas y desventajas. La principal ventaja del formato bitmap es la calidad y rapidez, su desventaja más importante es la gran necesidad de memoria que necesita. La representación vectorial requiere menos memoria y permite reescalar y manipular geométricamente la imagen sin introducir artefactos.

5.1. Imágenes Binarias

Las imágenes binarias se codifican como una matriz 2D, usando 1 bit por pixel. Un 0 codifica el valor negro y un 1 codifica el valor blanco. La principal ventaja de esta representación es que ocupa muy poca memoria. En la imagen siguiente se muestra una imagen binaria y a la derecha una porción de la matriz de una área 6x6.



5.2. Imágenes de Color

Normalmente las imágenes de color se representan mediante la representación RGB en la que cada pixel se codifica con 24 bit (8 bit por color); y la otra forma es mediante la representación indexada, en la que la matriz 2D contiene los índices a una paleta de colores.

5.2.1. Representación RGB

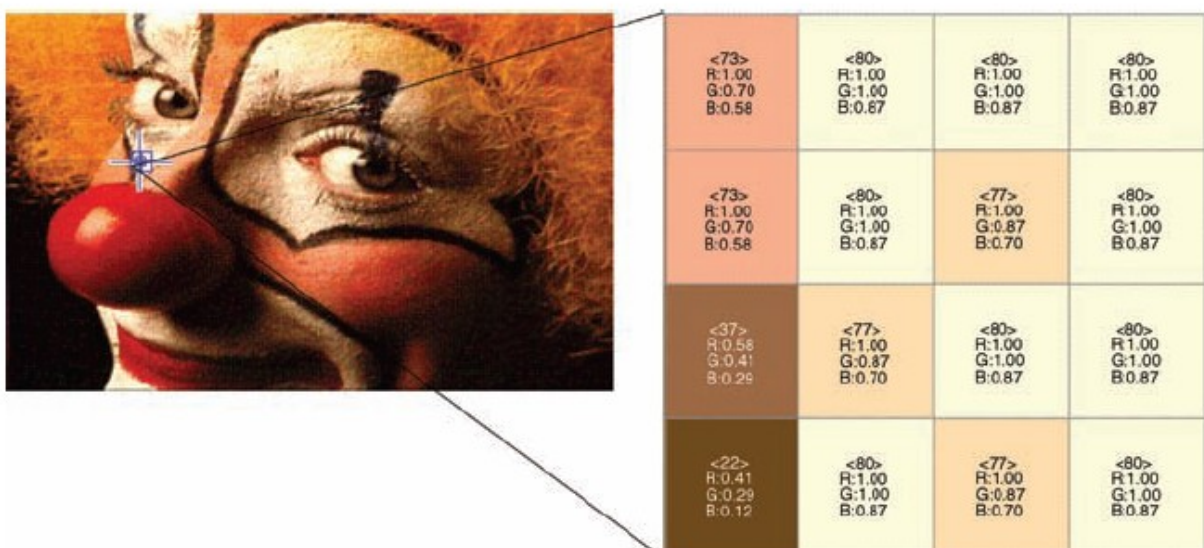
Las imágenes de color con esta representación usa normalmente 3 matrices: una para la banda roja (R) , otra para la verde (G) y otra para la azul (B).



Una representación alternativa usa una cuarta matriz para la transparencia o canal alfa.

5.2.2. Representación Indexada

En este caso tenemos una matriz 2D del mismo tamaño que la imagen. Los elementos de esta matriz son valores que se corresponden con las posiciones de una tabla de valores (la paleta). En esta tabla se codifican los colores.



6. Formatos de Imágenes

La mayoría de los formatos de archivo de imágenes representan las imágenes usando una especificación bitmap que consiste de una cabecera seguida de los datos de los píxeles. La cabecera almacena información acerca de la imagen, tales como dimensiones, número de bandas, número de bits por píxel, y alguna información indicando el tipo de fichero. En formatos más complejos la cabecera también puede contener información sobre el tipo de compresión usada y otros parámetros que son necesarios para visualizarla.

Los formatos de archivo más simples son el formato BIN y el PPM. El formato BIN (o RAW) consiste de una fila de datos para los píxeles, sin cabecera. Consecuentemente, el usuario de un archivo BIN debe conocer los parámetros más relevantes de la imagen para poder usarla.

El formato PPM y sus variantes (PBM para imágenes binarias, PGM para imágenes de niveles de gris, y PNM para cualquiera de ellas) son las más usadas. La cabecera de este tipo de imagen contiene una cadena mágica de 2-byte, las dimensiones, el número de bandas, y el máximo valor de intensidad que puede almacenar.

El formato "Microsoft Windows bitmap BMP" es también un formato muy usado por su simplicidad, y consiste también de una cabecera y los datos en formato raw.

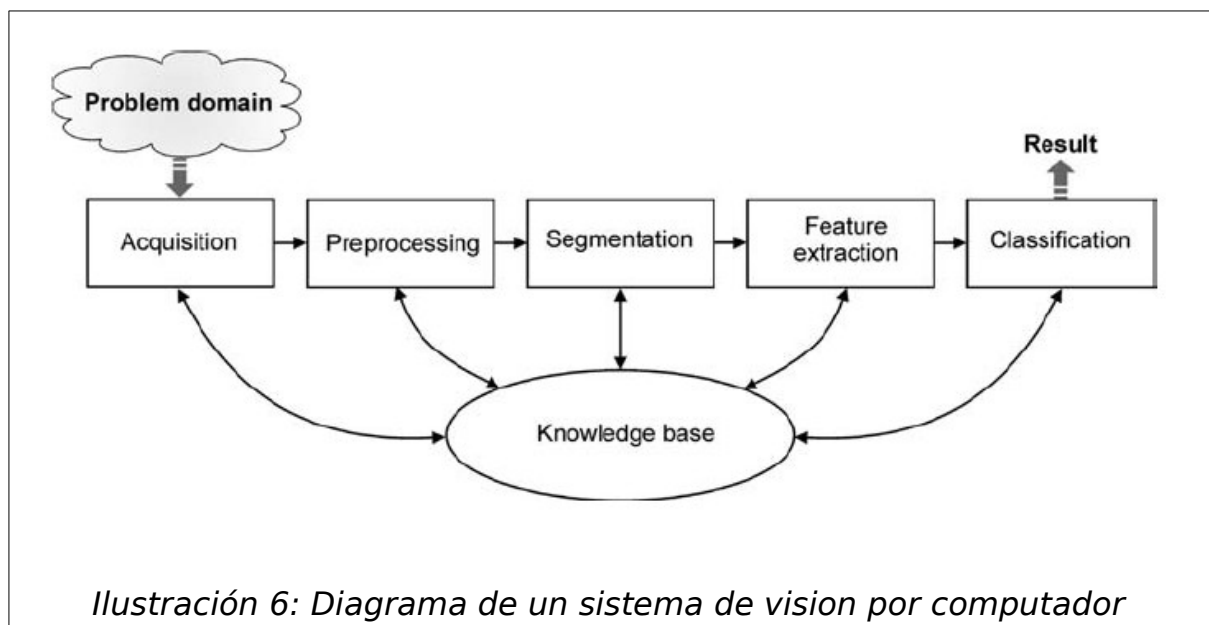
El formato JPEG es el formato más conocido. Se caracteriza por almacenar las imágenes con una alta razón de compresión y con una baja pérdida de calidad.

Otros formatos de interés son: GIF (Graphics Interchange Format) y TIFF (Tagged Image File Format). GIF usa una representación indexada para las imágenes de color (con una paleta de 256 colores). Además usa el algoritmo de compresión LZW (Lempel-Ziv-Welch) junto con una cabecera de 13 bytes. TIFF es el formato más sofisticado con muchas opciones y capacidades, incluyendo la capacidad de representar el color verdadero con 24 bpp y soporta cinco esquemas diferentes de compresión.

PNG (Portable Network Graphics) soporta tanto la representación indexada como la representación de verdadero color.

7. Componentes de un sistema de procesamiento de imágenes

En la Figura 16, se puede ver las etapas fundamentales de un sistema de procesamiento de imágenes. Para describir cada etapa supongamos que queremos aplicarlo al reconocimiento de matrículas.



En el primer paso nos encontramos con el dominio del problema (contexto). En nuestro ejemplo es ser capaz de extraer el contenido alfanumérico de la matrícula del vehículo, sin necesidad de la intervención humana.

En el bloque de adquisición tiene como objetivo obtener una o más imágenes del vehículo incluyendo la matrícula. Esta fase puede ser implementada usando una cámara CCD y controlando las condiciones de luz para asegurarnos que la imagen obtenida nos sirve para las siguientes fases de procesamiento. La salida de este bloque es una imagen digital que contiene una vista del vehículo.

El objetivo del preprocesamiento es mejorar la calidad de la imagen adquirida. Posibles algoritmos que se pueden aplicar en esta fase son: mejora del contraste, corrección del brillo y eliminación del ruido.

El bloque de segmentación es responsable para dividir la imagen en sus principales componentes: objetos en primer plano, objetos en el fondo. La salida en esta fase es un número de regiones etiquetadas o subimágenes. Es posible que en este caso particular la segmentación se realice a dos niveles: (1) extracción de la matrícula del resto de la imagen; y (2) segmentación de caracteres en el área de la matrícula.

La segmentación automática de una imagen es uno de las tareas mas difíciles en los sistemas de visión por computador.

El bloque de extracción de rasgos (también conocido como representación y descripción) consiste de algoritmos responsables para codificar los contenidos de las imágenes en una forma concisa y descriptiva. Rasgos típicos son: color, intensidad, textura, forma de los objetos más relevantes, etc. Esos rasgos se agrupan en un vector de rasgos que pueden ser usados como un indicador numérico de la imagen, usados para las siguientes fases, por ejemplo para clasificación.

Una vez que los rasgos más relevantes se han extraído y han sido codificados en un vector, el siguiente paso es usar este k-dimensional vector como entrada a un clasificador de patrones (también conocido como reconocimiento e interpretación). En este punto el procesamiento de imágenes usa algoritmos clásicos de reconocimiento de patrones como: clasificadores de distancia mínima, probabilísticos, redes neuronales, y muchos más. El último objetivo de este bloque es clasificar cada carácter individual, produciendo una cadena como salida, conteniendo la matrícula de nuestro vehículo.

En la Figura 6 todos los módulos están conectados a un bloque denominado base de conocimiento. Estas conexiones significan que para llegar a la solución óptima al problema de reconocimiento de matrículas, dependerá de cuanto conocimiento acerca del dominio del problema ha sido codificado y almacenado en el sistema de visión por computador.

8. Bibliografía

- [1] Wikipedia (<http://wikipedia.org>)
- [2] <http://www.marcadeagua.net/ilusiones/Ilusiones.pdf>
- [3] <http://teleformacion.edu.aytolacoruna.es/FISICA/document/fisicaInteractiva/OptGeometrica/Instrumentos/ollo/ollo.htm>
- [4] Practical Image and Video Processing Using Matlab. Oge Marques. ED. Wiley.
- [5] Computer Imaging. Digital Image Analysis and Processing. Scott E Umbaugh. A CRC Press Book