

## 2. Fundamentos de la imagen digital

El campo del procesamiento digital de imágenes está construido sobre bases matemáticas y probabilísticas, pero la *intuición y análisis* humanos juegan un importante papel al momento de escoger una técnica u otra. Esta elección se basa usualmente en *juicios visuales subjetivos*.

Desarrollar un entendimiento básico de la percepción humana es entonces pertinente.

Nos dedicaremos a entender la mecánica y los parámetros de la formación de imágenes en el ojo humano, así como las limitaciones físicas de la vista humana en los mismos términos que usaremos para nuestras imágenes digitales. Se definirán algunos conceptos para entendernos en el mismo lenguaje. Nos plantearemos las siguientes preguntas: ¿Qué diferencias de intensidad podemos distinguir? ¿Cuál es la resolución espacial del ojo humano? ¿Qué tan precisamente podemos estimar y comparar distancias? ¿Cómo percibimos los colores? ¿Qué características nos ayudan a detectar y distinguir objetos?

### 2.1 Elementos de la percepción visual

#### 2.1.1 Estructura del ojo humano

La figura 2.1 muestra una imagen simplificada del ojo humano. El ojo humano es casi una esfera, de aproximadamente 20 mm de diámetro. Tres membranas lo cubren: Córnea, Esclerótica y Retina.

##### a. Córnea

Junto con la esclerótica forma la membrana exterior del ojo. Es un tejido transparente que cubre la superficie frontal del ojo.

##### b. Esclerótica

Es una membrana opaca que cubre el resto del globo ocular.

##### c. Coroides

Contiene una red de vasos sanguíneos que son la principal fuente de nutrición del ojo. La coroides está fuertemente pigmentada para reducir la cantidad de luz exterior que entra al ojo y la reflexión interna dentro de él. En su parte frontal, se divide en el cuerpo ciliar y el iris.

##### d. Cuerpo ciliar

A él están unidos los músculos ciliares, que controlan el cristalino.

**e. Diafragma Iris**, que se expande o contrae para controlar la cantidad de luz que entra en el ojo. La apertura central del iris, llamada *pupila*, varía su diámetro de 2 a 8mm. El frente del iris contiene el pigmento visible del ojo, y la parte trasera contiene un pigmento negro.

##### d. Cristalino

Está suspendido por fibras que lo atan al cuerpo ciliar. Tiene una pigmentación amarillosa que puede aumentar con la edad. El cristalino absorbe aproximadamente 8% del espectro visible, especialmente en longitudes de onda cortas. La luz ultravioleta e

infrarroja son absorbidas apreciablemente. El cristalino es “la lente” del ojo y sirve para enfocar, ayudado por los músculos ciliares.

### e. Retina

Es la membrana interior posterior del ojo. Cuando el ojo está enfocado, la imagen de un objeto exterior es proyectada en la retina. Sobre su superficie se encuentran los receptores que permiten la percepción visual: los conos y los bastones.

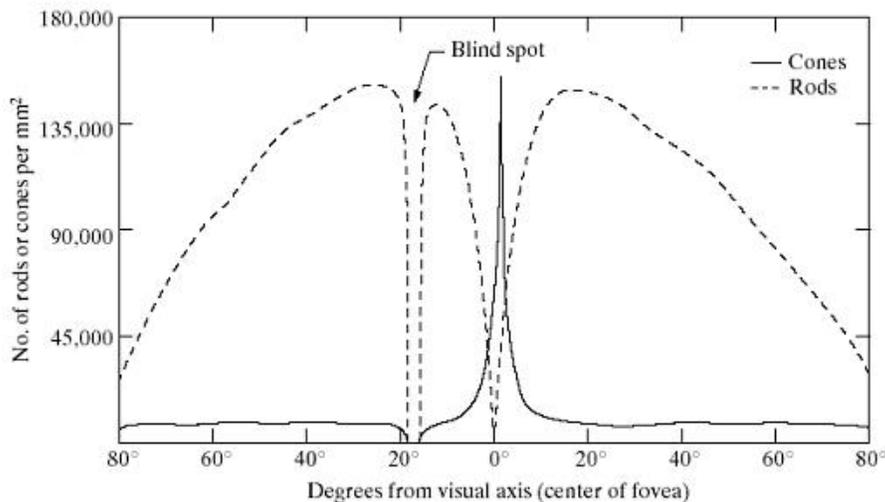
Los *conos* son entre 6 y 7 millones, se localizan principalmente en la parte central de la retina, llamada *fóvea*, y son muy sensibles al color. La *fóvea* es una región circular de la retina de aproximadamente 1.5 mm de diámetro.

Los músculos mueven al ojo hasta que la región de interés se localiza en la fóvea. Los humanos somos capaces de percibir detalles muy finos gracias a que cada uno de los conos está conectado a una terminal nerviosa.

La visión que utiliza los conos es llamada *fotópica*, o de luz brillante.

Los *bastones* son muchos más, entre 75 y 150 millones están distribuidos sobre la superficie retinal. Esta área de distribución tan grande y el hecho de que varios están conectados a una misma terminal nerviosa reduce su cantidad de detalle discernible. Sirven para dar una imagen general del campo de visión. No participan en la visión del color y son sensitivos a bajos niveles de iluminación (o visión *escotópica*, de luz tenue).

Existe una región del ojo en la que no hay receptores, ya que es la región donde surge el *nervio óptico*. Esta región se denomina *punto ciego*. Fuera de este punto, la distribución de receptores es radial con respecto a la fóvea, como se representa en la figura 2.2



**FIGURE 2.2**  
Distribution of rods and cones in the retina.

Con ciertas libertades de interpretación, podemos pretender que la fóvea es un arreglo de sensores de 1.5 mm de diámetro (un área de 1.767 mm<sup>2</sup>). La densidad de conos ahí es de aproximadamente 150,000 elementos por mm<sup>2</sup>. Basado en esas aproximaciones, el número de conos en la región de más agudeza en el ojo es de cerca de 265,000 elementos.

Comparando *solamente* el poder de resolución, un circuito de imágenes de dispositivo CCD (charge-coupled device) de resolución media puede tener este número de elementos en un arreglo receptor no mayor de  $5 \times 5 \text{ mm}^{-1}$ .

Recuérdese que los humanos integramos la inteligencia y la experiencia con la vista, por lo que esta comparación sólo es válida para mostrar el hecho de que la habilidad básica de resolución de detalle del ojo está ya desde hace algún tiempo casi en el mismo rango de los sensores electrónicos de imágenes.

### 2.1.2 Formación de imágenes en el ojo

La principal diferencia entre una lente óptica ordinaria y el cristalino es que este último es flexible, y su forma es controlada por la tensión de las fibras del cuerpo ciliar. Para enfocar objetos lejanos, se aplanan, para enfocar objetos cercanos, se ensanchan. La distancia entre el centro del cristalino y la retina (que llamaremos *distancia focal*), varía de aproximadamente 17mm a 14mm.

Con esta información podemos calcular el tamaño del objeto reflejado en la retina. Consideremos la figura 2.3.

El observador mira una palmera de 15m de altura desde una distancia de 100m (al punto focal C). Si  $h$  es la altura en mm del objeto en la imagen retinal, por geometría obtenemos que:

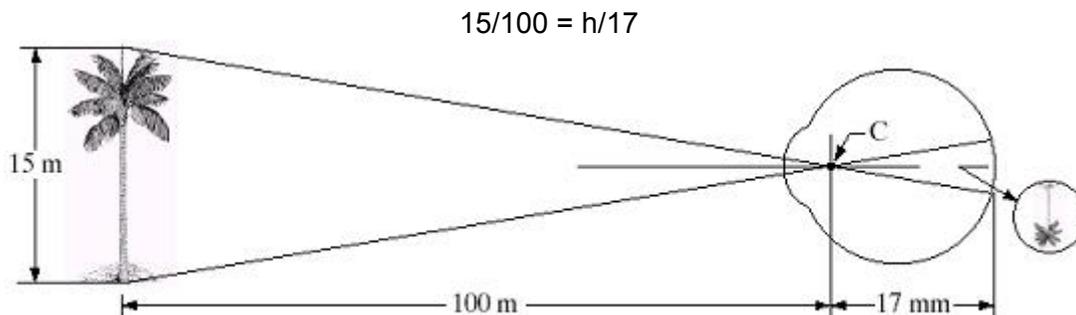


fig. 2.3 Obtención del tamaño de la imagen en la retina

Por lo tanto  $h=2.55 \text{ mm}$ , el tamaño de la palmera reflejada en la retina.

La percepción tiene lugar cuando los diferentes receptores son excitados, estos transforman la energía radiante a impulsos eléctricos que son enviados al cerebro para su decodificación<sup>2</sup>.

<sup>1</sup> Ver en 2.3.3 los ejemplos de sensores CCD y CMOS de kodak y sony.

<sup>2</sup> Para profundizar en el tema de cómo se transmite la información de los ojos al cerebro se puede consultar el libro de Dayan y Abbott [6], "Theoretical neuroscience", disponible en la red en su versión borrador. Aquí también se trata a fondo los modelos matemáticos más utilizados para modelar las redes neuronales y las señales neurológicas.

### 2.1.3 Adaptación a la iluminación y discriminación

Recordemos que las imágenes digitales son un conjunto discreto de intensidades. Por lo tanto tiene sentido hablar de la *capacidad de discriminación* del ojo entre diferentes niveles de intensidad.

El rango de intensidades de luz al que el ojo humano puede adaptarse es enorme (del rango de  $10^{10}$ ) desde el umbral escotópico hasta el límite de deslumbramiento.

Evidencia experimental ha demostrado que la *iluminación subjetiva o brillo* (la intensidad luminosa como la percibe el sistema visual humano) es una función logarítmica de la intensidad de luz que incide en el ojo.

En la figura 2.4 se ilustra esta característica. La línea larga sólida representa el rango de intensidades al que se puede adaptar el sistema visual. El rango en la visión fotópica es de cerca de  $10^6$ . La transición entre visión fotópica y escotópica es gradual entre los 0.001 y 0.1 millilambert (-3 a -1 mL en la escala logarítmica).

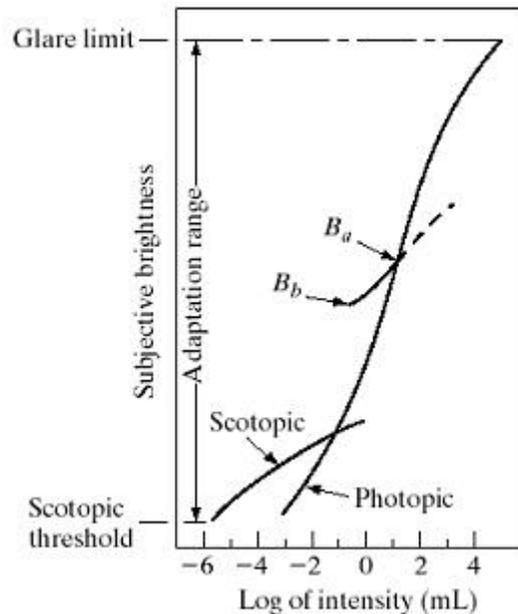


fig 2.4 Rango de sensaciones de iluminación subjetiva respecto a intensidad

Esto puede parecer impresionante, sin embargo la verdad es que el sistema visual no puede operar en todo este rango *simultáneamente*. En realidad logra esta gran variación con pequeños cambios de sensibilidad. A esto se le llama *Adaptación a la iluminación*.

#### *Adaptación a la iluminación*

El rango total de niveles de intensidad distintos que puede discriminar simultáneamente es relativamente pequeño cuando se le compara con el rango total de adaptación que hemos visto.

El nivel de adaptación para cierto conjunto de condiciones es llamado *Nivel de adaptación a la iluminación* (por ejemplo,  $B_a$  en la fig. 2.4). Este nivel está bastante

restringido, pues debajo del nivel  $B_b$  todo estímulo es percibido como negros indistinguibles unos de otros. El nivel superior, mostrado con guiones, no está restringido pero, si se extiende mucho pierde el significado, ya que intensidades mayores simplemente elevarían el nivel de adaptación  $B_a$ .

#### Adaptación a los cambios y cociente de Weber

La habilidad del ojo de discriminar entre cambios de intensidad de luz en cualquier nivel específico de adaptación también ha sido estudiada.

Un experimento clásico consiste en poner a un observador a mirar una gran área iluminada uniformemente que ocupa su campo visual entero. Esta área es iluminada por la parte de atrás por una fuente cuya *intensidad*  $I$ , puede variar.

Se incrementa la iluminación,  $\Delta I$ , con destellos de corta duración con forma de círculo en el centro del campo iluminado.

Si  $\Delta I$  no es suficientemente brillante, el observador dirá “no”. Mientras  $\Delta I$  crece, si el observador lo percibe dirá “sí”, cuando lo perciba. Finalmente, cuando  $\Delta I$  sea lo suficientemente fuerte, el observador dirá “sí” todo el tiempo.

La cantidad  $\Delta I_c / I$ , donde  $\Delta I_c$  es el incremento de iluminación que puede discernirse el 50% de las veces con iluminación  $I$ , se llama *cociente de Weber*.

Un valor pequeño de este cociente indica que un cambio de porcentaje pequeño en intensidad puede distinguirse, es decir, buena discriminación de iluminación. Un valor grande quiere decir que se necesita un gran cambio en intensidad para poder distinguirlo. Es decir, mala discriminación.

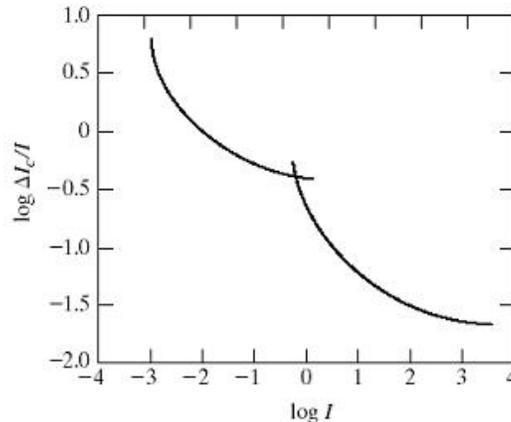


fig 2.4 Cociente de Weber como función de la intensidad

La figura muestra una gráfica de  $\log \Delta I_c / I$  como función de  $\log I$ . La curva muestra que la discriminación de iluminación es mala (el cociente de Weber es grande) con bajos niveles de iluminación, pero crece significativamente cuando se aumenta la iluminación de fondo.

Las dos ramas de la curva reflejan el hecho de que los niveles bajos de iluminación son manejados por bastones, y los altos – con mejor discriminación – por los conos.

Si la iluminación de fondo permanece constante y la intensidad de la otra fuente de luz varía incrementalmente en lugar de por destellos, el observador típico podrá discernir un total de entre una y dos docenas de cambios de intensidad.

Esto *no significa* que una imagen pueda ser representada por un número tan pequeño de valores de intensidad porque, mientras el ojo recorre la imagen, el fondo promedio cambia, permitiendo la detección de un conjunto diferente de cambios incrementales con cada nivel de adaptación.

La consecuencia es que el ojo es capaz de un rango bastante más amplio de discriminación de intensidad en total, como veremos más adelante.

Dos fenómenos pueden demostrar que la iluminación percibida no es simplemente una función de la intensidad.

El primero se basa en el hecho de que el sistema visual humano tiende a subestimar o sobreestimar cerca del límite de regiones de diferente intensidad.

En la figura se ven las *Bandas de Mach* (Ernst Mach, 1865), que demuestran este fenómeno.



fig 2.6 Cada banda de Mach tiene la misma intensidad, sin embargo se percibe un cambio en las fronteras

El segundo es llamado *contraste simultáneo*, y se relaciona con el hecho de que la iluminación percibida de una región no depende simplemente de su intensidad, como podemos ver en la figura (todos los cuadrados centrales tienen la misma intensidad).

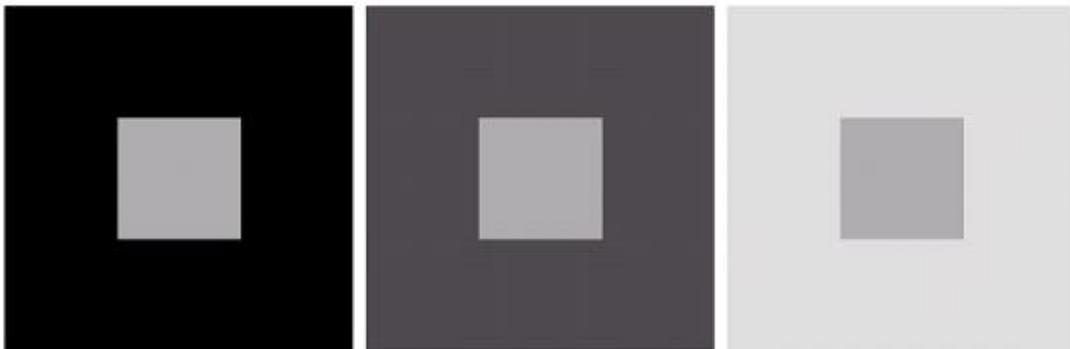
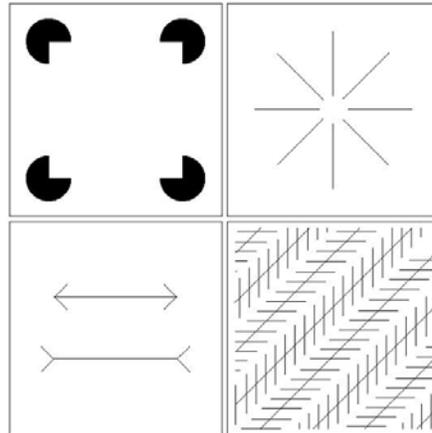


fig 2.7 Los cuadrados centrales parecen de distinto color y son exactamente el mismo

Otros ejemplos de fenómenos de percepción humana son las ilusiones ópticas, en las que el ojo da información que no existe o percibe erróneamente propiedades geométricas. En la figura vemos un cuadrado delineado que no está definido por línea alguna, un círculo con el mismo efecto, dos líneas paralelas que son exactamente del mismo tamaño y las líneas a 45° son perfectamente paralelas.

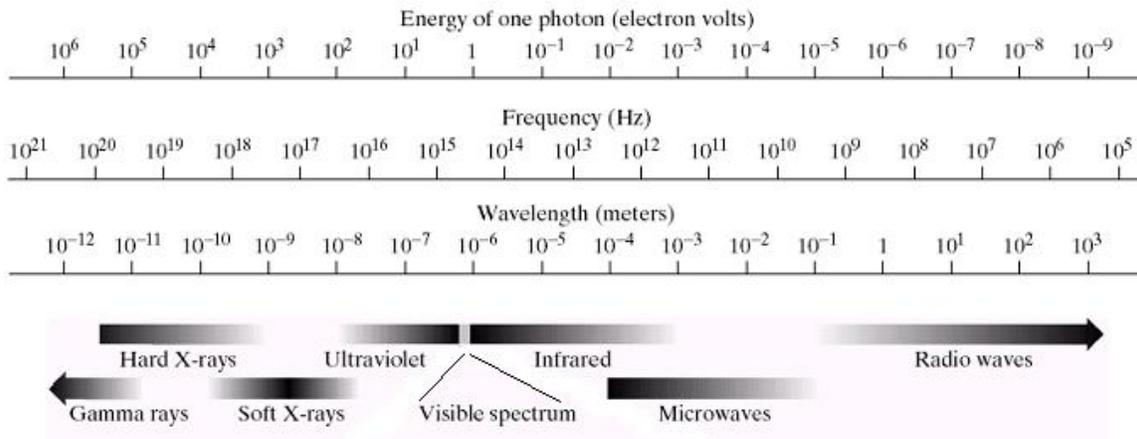


2.8 Ilusiones ópticas (un cuadrado que no existe, un círculo que no está delineado, dos líneas que aparentan ser de diferentes tamaños y líneas que siendo paralelas no lo aparentan)

## 2.2 La luz y el espectro electromagnético

En 1666, Isaac Newton descubrió que al pasar un rayo de luz por un prisma de vidrio, el rayo de salida no es blanco sino que está compuesto de un espectro de colores continuo, que va desde el violeta hasta el rojo.

El rango de colores que percibimos no es sino una muy pequeña parte del espectro electromagnético (EEM). En un extremo del espectro se encuentran las ondas de radio con longitudes de onda mil millones de veces más largas que los de la luz visible. En el otro están los rayos gama con longitudes de onda millones de veces más pequeños que los de la luz visible.



2.9 Espectro electromagnético en unidades de energía, frecuencia y longitud de onda

Como vemos en la figura, el espectro puede definirse en términos de longitud de onda ( $\lambda$ ) en metros, frecuencia ( $\nu$ ) en hertz, o energía ( $E$ ) en electron volts. La longitud de onda y la frecuencia se relacionan por:

$$\lambda = c / \nu \quad (\text{ec. 2.2-1})$$

Donde  $c$  es la velocidad de la luz ( $2.998 \times 10^8$  m/s).

Y la energía está dada por

$$E = h\nu \quad (\text{ec. 2.2-2})$$

Donde  $h$  es la constante de Planck ( $6.626 \times 10^{-34}$  J.s).

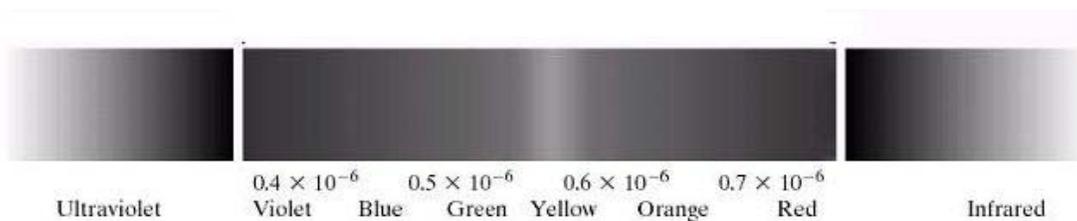
### Ondas electromagnéticas y fotones.

Las ondas electromagnéticas se pueden visualizar como ondas sinusoidales propagadas de longitud de onda  $\lambda$ , o como una corriente de partículas sin masa viajando en un patrón con forma de onda y moviéndose a la velocidad de la luz. Cada partícula sin masa contiene cierta cantidad (o paquete) de energía. Los paquetes de energía se llaman fotones.

La energía es proporcional a la frecuencia (ecuación 2.2-2): a más frecuencia, más energía por fotón.

### Luz visible

La luz es un tipo particular de radiación electromagnética que puede ser detectada por el ojo humano. La banda visible del EEM se extiende en un rango de aproximadamente  $0.43 \mu\text{m}$  (violeta), a  $0.79 \mu\text{m}$  (rojo).



2.10 Rango visible del espectro electromagnético

Los colores que percibimos en un objeto son determinados por la naturaleza de la luz *reflejada* por dicho objeto.

*Un cuerpo reflejante que está balanceado en todas las longitudes de onda visibles aparece blanco para el observador. Un cuerpo que refleja un rango en particular del espectro visible se ve de cierto color (p. Ej. Los objetos que se ven de color verde absorben la energía en todas las longitudes de onda excepto en la de los tonos de verde, de 500 a 570 nm, esta la reflejan).*

### Luz monocromática

La luz que no tiene color se llama acromática o monocromática. Su único atributo es su intensidad o cantidad. En general usamos el término *nivel de gris* para definir la intensidad monocromática porque ésta va desde el negro hasta el blanco pasando por una gama de grises.

## Luz cromática

Es la luz de color, ya descrita. Se usan 3 cantidades para describir la calidad de una fuente cromática:

### 1. Radiancia

Cantidad total de energía que fluye de una fuente de luz. Se mide en watts.

### 2. Luminancia

Cantidad de energía que un observador percibe de una fuente de luz. Se mide en lúmenes.

### 3. Brillo

Es la iluminación subjetiva, ya vista anteriormente. Es casi imposible de medir y engloba la noción de intensidad.

## Limitación fundamental de la capacidad de los sensores

En principio, si se puede desarrollar un sensor capaz de detectar energía radiada por una banda del EEM, se pueden obtener imágenes de esa banda.

Sin embargo, se requiere que la longitud de onda de una onda EM sea del mismo tamaño o más pequeña que el objeto a detectar (*por ejemplo, una molécula de agua tiene un diámetro del orden de  $10^{-10}$  m. Entonces se requiere una fuente capaz de emitir luz en el extremo lejano del ultravioleta o los rayos X suaves*).

## 2.3 Adquisición de imágenes

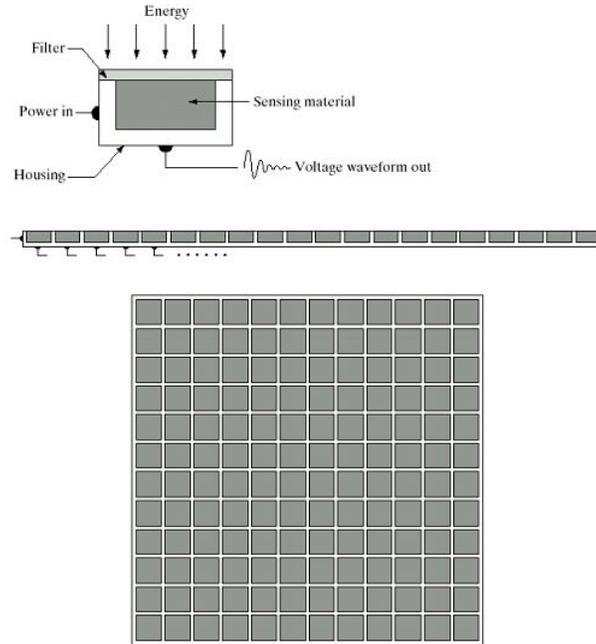
Los tipos de imágenes que nos interesan se generan por una combinación de una fuente de “iluminación” y la reflexión o absorción de energía de esta fuente por parte de los elementos de la escena.

Recordemos que esta “iluminación” puede ser también radar, infrarrojo, o alguna fuente que no es parte del EEM., como ultrasonido.

Dependiendo de la naturaleza de la fuente, la energía “luminosa” puede ser *reflejada o transmitida* por los objetos.

En algunas aplicaciones, la energía transmitida o reflejada se enfoca a un fotoconvertidor (p. Ej. Una pantalla de fósforo) que convierte la energía a luz visible.

En la figura se muestran los 3 principales arreglos de sensores utilizados para transformar energía luminosa en imágenes digitales.



2.11 Tipos de sensores: sencillo, en línea y en arreglo

El proceso en todos ellos es simple: La energía entrante se transforma a un voltaje por la combinación de electricidad de entrada y el material del sensor, sensible al tipo de energía que se quiere detectar. La onda de voltaje de salida es la respuesta del sensor, y una cantidad digital se obtiene de cada sensor digitalizando su respuesta.

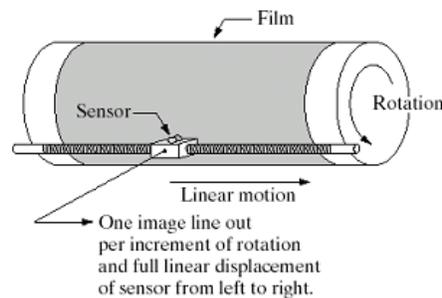
### 2.3.1 Adquisición con un solo sensor

El sensor más familiar de este tipo es el fotodiodo, de silicón, cuya voltaje de salida es proporcional al de la luz. A veces se utiliza un filtro frente al sensor para mejorar la selección de ciertas longitudes de onda (p. Ej. Un filtro verde favorece la banda verde del espectro).

Para generar una imagen 2D, se requieren desplazamientos relativos en las direcciones x y y del área a capturar. En la figura vemos un sensor montado en un tornillo que le da movimiento en dirección perpendicular. Este método es una manera barata pero lenta de obtener imágenes de alta resolución.

Arreglos similares usan una “cama” plana (flat bed), con el sensor moviéndose en 2 direcciones lineales.

A este tipo de sensores se les llama *microdensitómetros*.



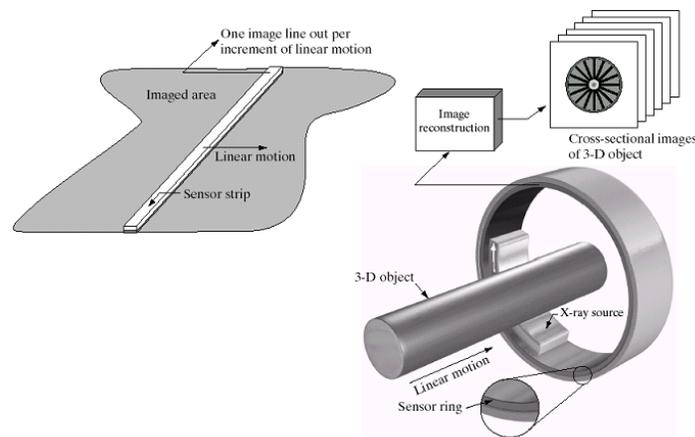
2.11 Se combina un sensor único con movimiento para generar una imagen 2D

Otro tipo de sensor único utiliza una fuente de láser coincidente con el sensor, con espejos utilizados para controlar el láser de salida en un patrón de escaneo, y para dirigir la señal reflejada al sensor.

### 2.3.2 Adquisición con bandas o líneas de sensores

Es un arreglo de sensores en línea, formando una banda que provee elementos de la imagen en una dirección. La dirección perpendicular se obtiene por el movimiento de la banda. Este es el tipo de arreglo utilizado en la mayoría de los escáneres de “cama”.

Se utilizan rutinariamente en aplicaciones de imágenes aéreas, en las que el movimiento perpendicular es dado por el movimiento del avión. Se utilizan lentes u otro tipo de esquemas de enfoque para proyectar el área a escanear en los sensores.



2.12 Sensores en línea y en línea circular.

También existe la configuración de anillo, en la que la fuente rota y los sensores están en el lado opuesto, colectando la energía que pasa a través del objeto. Esta es la base de la tomografía Axial Computarizada (CAT), la resonancia magnética (MRI), y la tomografía por emisión de positrones (PET).

Nótese que la salida de los sensores debe ser procesada por algoritmos de reconstrucción cuyo objetivo es transformar los datos en imágenes de cortes que tengan sentido.

### 2.3.3 Adquisición por arreglos de sensores

Numerosos aparatos de sentido electromagnéticos y algunos ultrasónicos frecuentemente se ordenan en forma de arreglos.

Un punto histórico en la tecnología de sensores de imágenes fue la invención de arreglos de semiconductores fotodetectores. El más común es el CCD (Charge-coupled Device), que consiste en un gran número de elementos fotosensibles ordenados en forma de arreglo. Estos funcionan en 2 fases.

En la fase de acumulación, cada elemento recoge cargas eléctricas generadas por fotones absorbidos (la carga acumulada es proporcional a la iluminación). En la fase de

lectura, estas cargas se transportan secuencialmente en todo el circuito de sensor a sensor y son finalmente convertidas a un voltaje eléctrico.

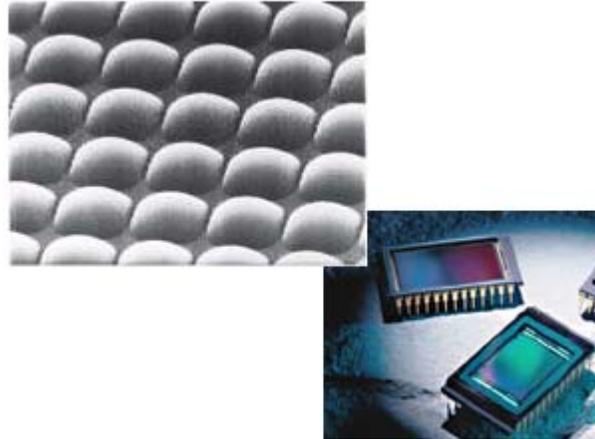


fig 2.13 Vista al microscopio de un CCD de Sony y 2 presentaciones del CCD de Kodak

Dispositivo	Pixeles (HxV)	Tamaño de pixel (H x V $\mu$ m)
KAF-0261E	512 x 512	20.0 x 20.0
KAF-0402E(/ME)	768 x 512	9.0 x 9.0
KAF-1001E	1024 x 1024	24.0 x 24.0
KAF-1301E(/LE)	1280 x 1024	16.0 x 16.0
KAF-1402E(ME)	1320 x 1037	6.8 x 6.8
KAF-1602E(/LE)	1536 x 1024	9.0 x 9.0
KAF-1603ME	1536 x 1024	9.0 x 9.0
KAF-3200E(ME)	2184 x 1472	6.8 x 6.8
KAF-4301E	2084 x 2084	24.0 x 24.0
KAF-4320E	2092 x 2092	24.0 x 24.0
KAF-6303E(/02LE)	3088 x 2056	9.0 x 9.0
KAF-16801E(/LE)	4096 x 4096	9.0 x 9.0

tabla 2.1 Algunos sensores CCD comerciales de Kodak (“full frame”, 2004)

Recientemente los sensores CMOS han atraído la atención por la calidad de su imagen y la uniformidad de la sensibilidad de sus elementos, que se acercan a la calidad del CCD. Sus ventajas sobre el CCD son que consumen menos energía, sus subáreas se accesan rápidamente, y se pueden añadir a circuitos (cámaras de un solo chip). También son más baratos.

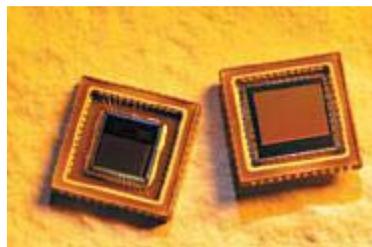


fig 2.14 Sensores CMOS

Dispositivo	Pixeless (HxV)	Tamaño de pixel (H x V $\mu$ m)
KAC-9618	648 x 488	7.5 x 7.5
KAC-9619	648 x 488	7.5 x 7.5
KAC-9628	648 x 488	7.5 x 7.5
KAC-9630	128 x 98	20 x 20
KAC-9637	648 x 488	6 x 6
KAC-9638	1,288 x 1,032	6 x 6
KAC-9647	648 x 488	6 x 6
KAC-9648	1,288 x 1,032	6 x 6

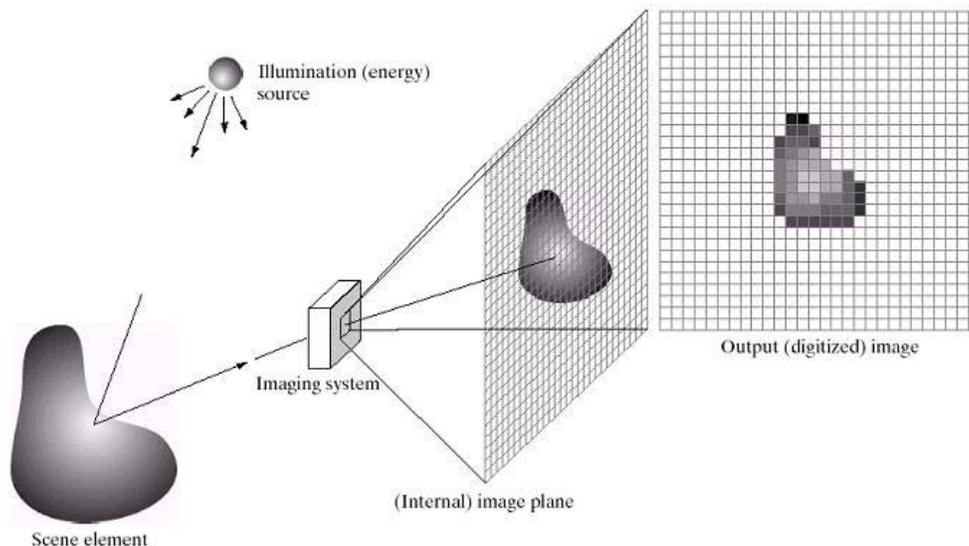
tabla 2.2 Algunos sensores CMOS comerciales de Kodak (2004)

### Características de los sensores semiconductores

- La distorsión geométrica es inexistente.
- Son pequeños y aguantadores. Soportan campos magnéticos, vibraciones, etc.
- Tienen una alta sensibilidad. Algunos CCD, si son enfriados de manera especial, pueden ser expuestos por horas y son cerca de 100 veces más sensibles que el material fotográfico.
- Hay una amplia variedad (ver tablas).
- No están limitados al rango visible (los sensores de silicón tienen un rango entre los 200 nm (ultravioleta) y los 1100 nm (Infrarrojo cercano). Sin embargo sensores de otros materiales pueden detectar incluso rayos X).

**Ejercicio 2.2** Dados los datos de la sección 2.1 y los de las tablas anteriores, compare la densidad de receptores (pixeles) por  $\text{mm}^2$  de un sensor KAF-1402E ME (Kodak 2004) con la densidad de receptores de color por  $\text{mm}^2$  de la fovea.

La ventaja de los arreglos es que una imagen completa puede obtenerse con sólo enfocar el patrón de energía en la superficie del arreglo (no requiere movimiento).



2.13 Ejemplo del proceso de adquisición de una imagen digital

El proceso se explica en la figura:

La energía de la fuente de iluminación es reflejada o transmitida por un elemento de la escena, la energía obtenida se colecta y se enfoca al plano focal(en el caso de luz, por medio de una lente).

El arreglo de sensores produce salidas proporcionales a la integral de la luz recibida en cada sensor.

Circuitería analógica y digital convierte estas salidas a señal de video, que se digitaliza para dar como resultado una imagen digital.

### 2.3.4 Un modelo simple de formación de imágenes

Recordemos que una imagen puede ser representada por una función bidimensional de la forma  $f(x, y)$ , en donde el valor o amplitud de  $f$  en las coordenadas espaciales  $(x,y)$  da la intensidad de la imagen en ese punto.

Cuando la imagen se genera a partir de un proceso físico, sus valores son proporcionales a la energía radiada por la fuente física (p. Ej. Ondas EM). Como consecuencia,  $f(x,y)$  debe ser diferente de cero y finita. Esto es:

$$0 < f(x,y) < \infty \quad (2.3.4 - 1)$$

La función  $f(x,y)$  se caracteriza por dos componentes:

- 1) *Iluminación*: la cantidad de luz incidente procedente de la fuente sobre la escena.
- 2) *Reflectancia*: la cantidad de luz reflejada por los objetos de la escena.

Las dos se combinan para formar el producto:

$$f(x,y) = i(x,y)r(x,y) \quad (2.3.4 - 2)$$

donde

$$0 < i(x,y) < \infty \quad (2.3.4 - 3)$$

y

$$0 < r(x,y) < 1 \quad (2.3.4 - 4)$$

La ecuación 2.3.4 – 4 indica que la reflectancia está acotada entre 0 (absorción total) y 1 (reflexión total).

La naturaleza de  $i(x,y)$  está determinada por la fuente de iluminación, y la de  $r(x,y)$ , por las características de los objetos.

Esto se aplica también al caso en que las imágenes se forman por la transmisión de la “iluminación” por el medio, como en los rayos X. En ese caso la segunda componente sería de *capacidad de transmisión* y no *reflectancia*.

### Nivel de gris y escala de grises

A la intensidad de una imagen monocromática  $f$  en las coordenadas  $(x, y)$  se le denomina *nivel de gris* ( $I$ ) de la imagen en ese punto.

$$I = f(x_0, y_0) \quad (2.3.4 - 5)$$

de las ecuaciones anteriores se deduce que  $I$  está en el rango

$$L_{\min} \leq I \leq L_{\max} \quad (2.3.4 - 6)$$

Al intervalo  $[L_{\min}, L_{\max}]$  se le denomina *escala de grises*. Una práctica habitual consiste en desplazar este intervalo hasta el intervalo  $[0, L]$ , donde  $I = 0$  se considera negro y  $I = L - 1$  se considera blanco (todos los valores intermedios son tonos de gris).

## 2.4 Muestreo y cuantificación

Ya hemos visto que hay muchas maneras de adquirir imágenes, pero nuestro objetivo siempre es el mismo: generar imágenes digitales desde datos detectados con sensores.

Para crear una imagen digital, necesitamos convertir los datos detectados a forma digital, esto involucra los procesos de *muestreo* y *cuantificación*.

### 2.4.1 Conceptos básicos de muestreo y cuantificación

La idea básica detrás de estos dos conceptos se ilustra en la figura.

Una imagen puede ser continua tanto respecto a sus coordenadas  $x$  y  $y$ , como a su amplitud. Para convertirla a forma digital, hay que digitalizarla en los dos aspectos (espacialmente y en amplitud).

La digitalización de las coordenadas espaciales ( $x$ ,  $y$ ) se denomina *muestreo de la imagen* y la digitalización de su amplitud se conoce como *cuantificación*.

La función bidimensional mostrada en la segunda figura es una gráfica de los valores de amplitud (el nivel de gris) de la imagen continua en la primera figura a lo largo del segmento de línea AB. Las variaciones aleatorias se deben a ruido de la imagen.

Para muestrear esta función, tomamos muestras a espacios iguales a lo largo de AB, indicadas por los cuadritos blancos. El conjunto de estos cuadritos nos da la función muestreada.

Sin embargo los valores de las muestras aún se encuentran en un rango continuo de valores de niveles de gris. Para obtener una función digital, debemos convertir (cuantificar) los valores de gris a cantidades discretas. Esto se hace simplemente asignando uno de los ocho valores de la figura a cada muestra.

La última figura muestra las muestras digitales que resultan del muestreo y cuantificación.

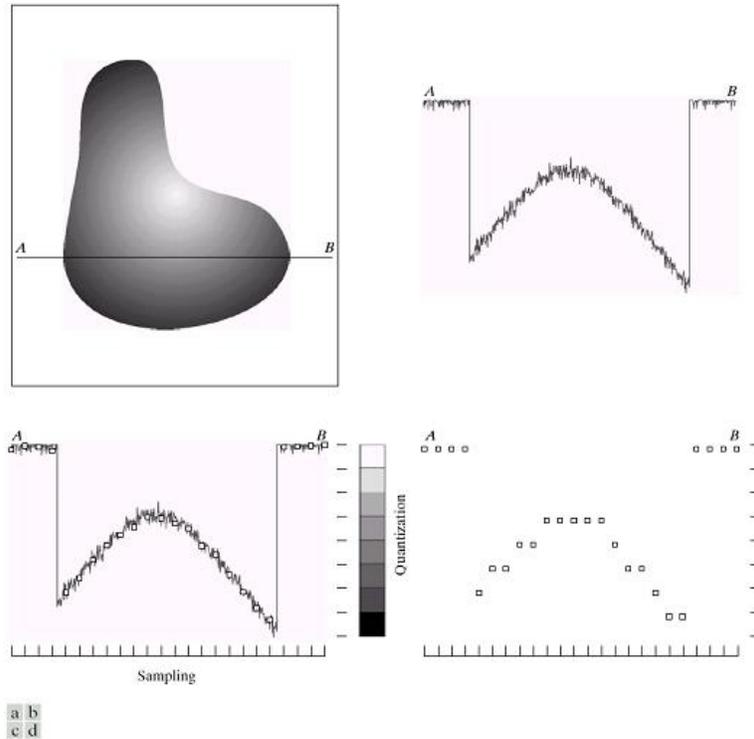


Fig 2.14 Muestreo y cuantificación: a) Imagen original b) Amplitud a lo largo de AB c) Muestreo d) Cuantificación

En la práctica, el método de muestreo está determinado por el orden de los sensores utilizados para generar la imagen. Por ejemplo, si se utiliza una banda de sensores, el número de sensores en la banda establece las limitaciones de muestreo en una de las direcciones. Si se usa un arreglo de sensores, el número de sensores del arreglo establece las limitaciones de muestreo en ambas direcciones.

Claramente, la calidad de una imagen digital se determina en gran manera por el número de muestras y niveles de gris utilizados en el muestreo y cuantificación.

## 2.4.2 Representación de imágenes digitales

La salida del muestreo y cuantificación es una matriz de números reales. Los representaremos de 2 maneras:

Una imagen continua  $f(x,y)$  se describe de forma aproximada por una serie de muestras igualmente espaciadas organizadas en forma de una matriz  $N \times M$  como se indica en la ecuación (2.4.2 – 1), donde cada elemento de la matriz es una cantidad discreta:

$$f(x, y) = \begin{bmatrix} f(0,0) & f(0,1) & \dots & f(0, N-1) \\ f(1,0) & f(1,1) & \dots & f(1, N-1) \\ \vdots & \vdots & \dots & \vdots \\ f(M-1, 0) & f(M-1,1) & \dots & f(M-1, N-1) \end{bmatrix}$$

Donde el término de la derecha representa lo que comúnmente se denomina una *imagen digital*. Cada uno de sus elementos es un *elemento de la imagen*, o *pixel*.

Algunas veces lo representaremos con una notación de matrices más tradicional (ec 2.4.2 -2):

$$A = \begin{bmatrix} a_{0,0} & a_{0,1} & \dots & a_{0, N-1} \\ a_{1,0} & a_{1,1} & \dots & a_{1, N-1} \\ \vdots & \vdots & \dots & \vdots \\ a_{M-1,0} & a_{M-1,1} & \dots & a_{M-1, N-1} \end{bmatrix}$$

No se requiere un valor especial de M y N, salvo que sean enteros positivos. En el caso del número de niveles de gris, éste es usualmente una potencia entera de 2:

$$L = 2^k \quad (2.4.2 - 3)$$

El número b de bits necesarios para almacenar una imagen digitalizada es :

$$b = N \times M \times k \quad (2.4.2 - 4)$$

cuando M = N:

$$b = N^2 k \quad (2.4.2 - 5)$$

### 2.4.3 Resolución espacial y resolución en niveles de gris

El muestreo es el factor principal para determinar la *resolución espacial* de una imagen. Básicamente, la *resolución espacial* es el grado de detalle discernible en una imagen.

La *resolución de nivel de gris* se refiere al más pequeño cambio discernible en nivel de gris aunque, como ya lo habíamos señalado, medir los cambios discernibles en niveles de intensidad es un proceso altamente subjetivo.

La potencia de 2 que determina el número de niveles de gris es usualmente 8 bits, es decir, 256 diferentes niveles de gris. Algunas aplicaciones especializadas utilizan 16 bits.

Usualmente decimos que una imagen digital de tamaño  $M \times N$  con  $L$  niveles de gris tiene una *resolución espacial* de  $M \times N$  píxeles y una *resolución de nivel de gris* de  $L$  niveles.

En el primer ejemplo vemos una imagen con *resolución espacial* de  $1024 \times 1024$  y 8 bits para representar los niveles de gris. Las imágenes siguientes han sido submuestreadas a partir de esta primera imagen, y se han agrandado para comparar los resultados.

El submuestreo consistió en eliminar un número apropiado de columnas y renglones de la imagen original. Por ejemplo, en la segunda imagen se borraron una columna y un renglón sí y uno no, obteniéndose una imagen de la mitad del tamaño,  $512 \times 512$ .

Nótese que a partir de la tercera imagen, de  $256 \times 256$ , aparece un fino patrón de “tablero de ajedrez”, también llamado “pixelado” en los bordes de la flor. Este patrón se va haciendo más notorio hasta llegar a la última imagen, de  $32 \times 32$ .

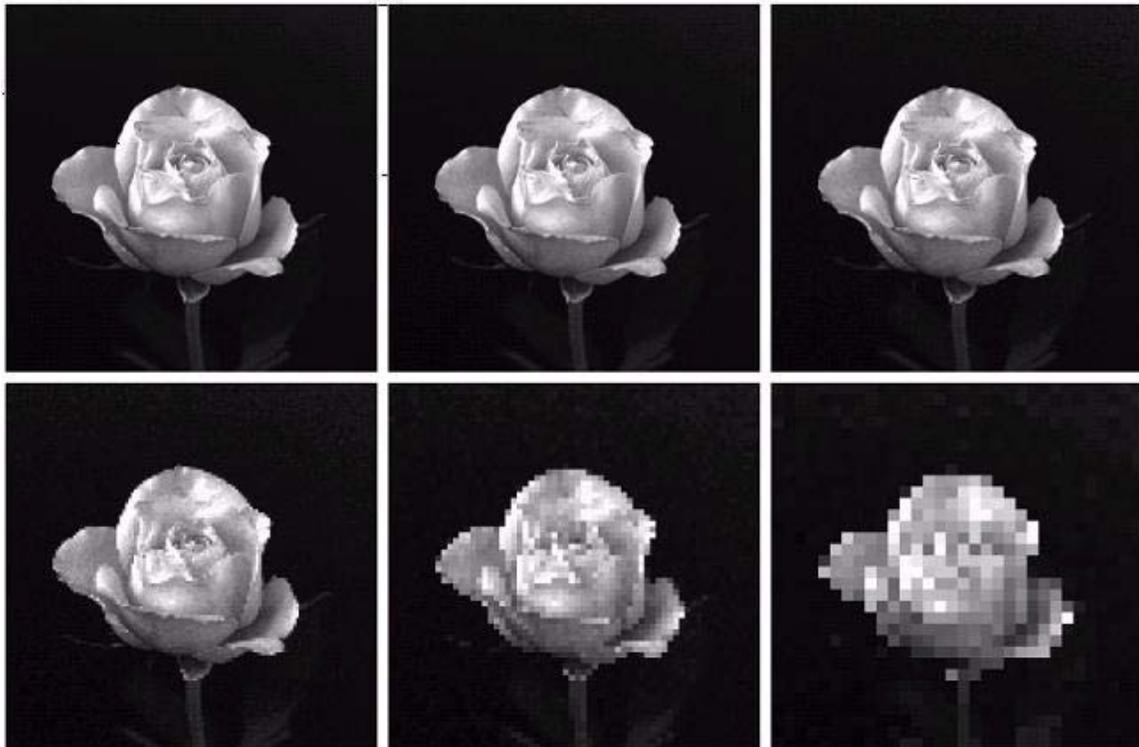


Fig 2.15 Imagen de  $1024 \times 1024$  original y sus submuestreos (ampliados al tamaño de la primera) de  $512 \times 512$ ,  $256 \times 256$ ,  $128 \times 128$ ,  $64 \times 64$  y  $32 \times 32$

En el segundo ejemplo tenemos una imagen CAT con el mismo número de muestras (*resolución espacial* constante), pero reducimos el número de niveles de gris desde 256 a 2, en potencias enteras de 2. Nótese a partir de la cuarta imagen, con 32 niveles, un casi imperceptible conjunto de “manchas” delimitadas en áreas suaves de niveles de gris. Este efecto, ya bastante notable en la figura con 8 niveles, es llamado *falso contorno*.

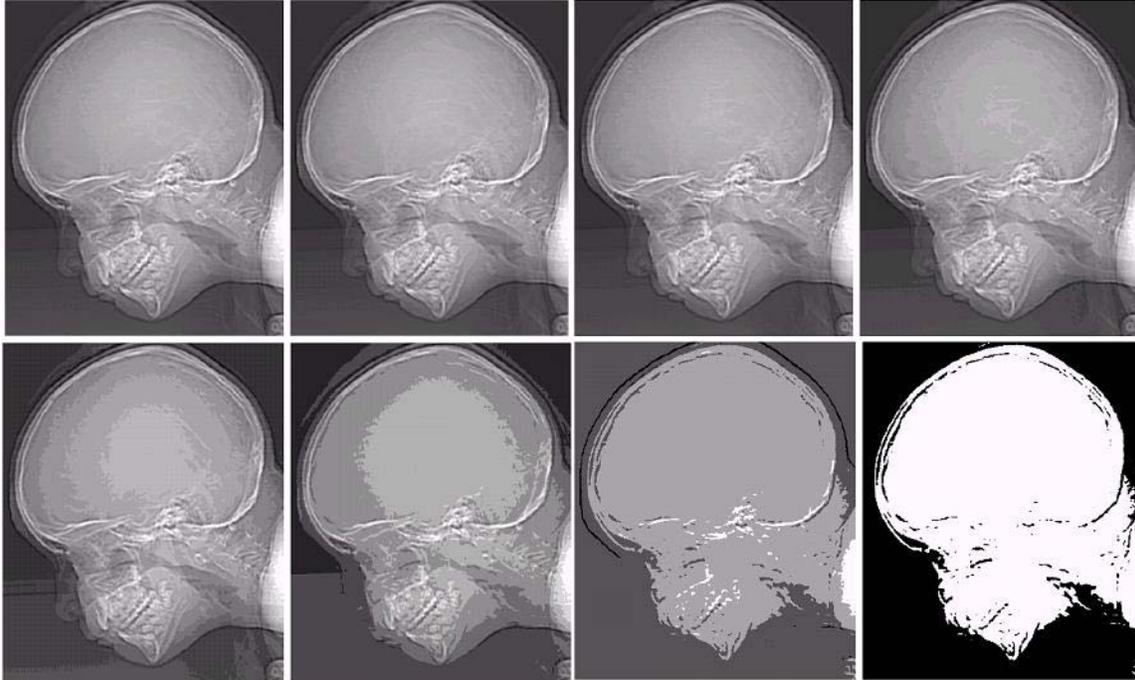


Fig 2.16 Imagen 472 x 374 en 256, 128, 64, 32, 16, 8, 4 y 2 niveles de gris

Una mínimo tamaño aproximado en el que una imagen estará libre de falsos contornos y “pixelados” será 256 x 256 con 64 niveles de gris.

#### 2.4.4 Aliasing y patrones de Moiré

Las funciones cuya área bajo la curva es finita, pueden ser representadas en términos de senos y cosenos de varias frecuencias (Fourier). El componente seno/coseno de más alta frecuencia determina el contenido en frecuencia más alto de la función.

*Suponga que la más alta frecuencia es finita y que la función no tiene una duración limitada (lo que se llama una función limitada en banda). El teorema de muestreo de Shannon nos dice que, si la función se muestrea a un ritmo igual o mayor que el doble de su más alta frecuencia, es posible recobrar completamente la función original a partir de sus muestras <sup>3</sup>.*

Si la función se submuestrea, el fenómeno conocido como *aliasing* corrompe la imagen muestreada (esta corrupción tiene la forma de componentes adicionales de frecuencia que se introducen en la función muestreada, llamadas *frecuencias de aliasing*).

El ritmo de muestreo en imágenes es el número de muestras tomadas (en las dos direcciones) por unidad de distancia.

El efecto de las frecuencias de aliasing puede ser apreciado bajo ciertas condiciones en la forma de los llamados *patrones de Moiré*.

<sup>3</sup> Es imposible satisfacer el teorema de muestreo en la práctica. Sólo podemos trabajar con datos muestreados de duración finita. Sin embargo hay un caso especial. Cuando la función es periódica, puede ser muestreada a un ritmo mayor o igual al doble de su frecuencia más rápida, y es posible recobrar la función a partir de sus muestras si el muestreo captura exactamente un número entero de periodos de la función.

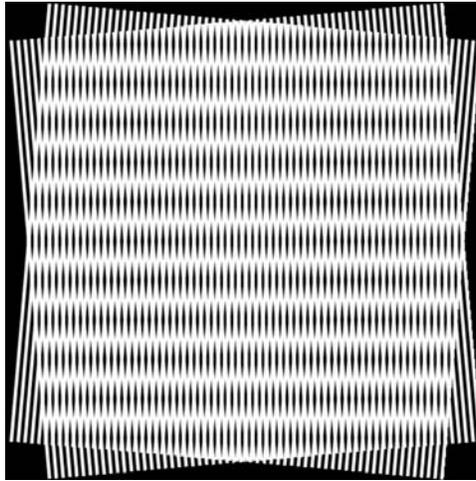


Fig 2.17 Ilustración del efecto de patrones de Moiré

### 2.4.5 Cambiando el tamaño de la imagen

Concluimos el tratamiento de muestreo y cuantificación con una breve descripción de cómo ampliar (zoom) y reducir una imagen digital.

Ampliar una imagen puede ser visto como sobremuestreo, y encogerla como submuestreo. Sin embargo, la diferencia entre estas dos operaciones y el muestreo y cuantificación de una imagen continua es que el cambio de tamaño se aplica a una imagen digital.

Ampliar requiere 2 pasos: la creación de nuevos lugares para píxeles, y la asignación de niveles de gris a esos nuevos píxeles.

Supongamos que tenemos una imagen de 500x500 que queremos agrandar 1.5 veces a 750x750 píxeles. Una de las maneras más fáciles de visualizar el proceso es sobreponer una malla imaginaria de 750x750 a la imagen original. Para asignar niveles de gris a cada punto en la malla sobrepuesta, buscamos el píxel más cercano en la imagen original y le asignamos su nivel de gris al píxel de la malla.

Este método de asignación de nivel de gris se llama *interpolación por vecino más cercano*.

Un caso especial de la interpolación del vecino más cercano es la *replicación de píxeles*. Se puede aplicar cuando queremos aumentar el tamaño de la imagen un número entero de veces.

Por ejemplo, si queremos doblar el tamaño de la imagen, duplicamos cada columna y a partir de la imagen resultante duplicamos cada renglón.

La reducción de la imagen se logra con un proceso parecido. Si queremos reducirla a la mitad, borramos una columna sí y una no y lo mismo con los renglones. Si queremos reducirla por un número no entero, vamos a expandir la malla de que se habló antes, de manera que sea del mismo tamaño que la imagen original, y aplicamos la interpolación por vecino más cercano.

Usualmente se desenfoca un poco la imagen antes de reducirla de tamaño, para reducir efectos de aliasing. Veremos cómo desenfocar la imagen en el siguiente capítulo.

## 2.5 Algunas relaciones básicas entre píxeles

### 2.5.1 Vecinos de un píxel

Un píxel  $p$  en las coordenadas  $(x,y)$  tiene cuatro vecinos horizontales y verticales dados por

$$(x + 1, y), (x - 1, y), (x, y + 1), (x, y - 1)$$

Este conjunto de píxeles, llamados los *4-vecinos* de  $p$ , se denota como  $N_4(p)$ . Cada píxel es una unidad de distancia desde  $(x,y)$ , y algunos de los vecinos de  $p$  están fuera de la imagen digital si  $(x,y)$  está en el borde de la imagen.

Los cuatro vecinos diagonales de  $p$  tienen las coordenadas:

$$(x + 1, y + 1), (x + 1, y - 1), (x - 1, y + 1), (x - 1, y - 1)$$

y se denotan como  $N_D(p)$ . Estos puntos, junto con los 4 vecinos, son llamados los 8 vecinos de  $p$ , y se escriben  $N_8(p)$ . Algunos de ellos estarán fuera de la imagen si  $(x,y)$  está en el borde de la imagen.

### 2.5.2 Conectividad, regiones y fronteras

La conectividad entre píxeles es un concepto fundamental que simplifica la definición de numerosos conceptos de la imagen digital, como regiones y fronteras.

Para establecer si 2 píxeles están conectados, se debe determinar si son vecinos y si sus niveles de gris satisfacen un criterio predeterminado de similitud (p. Ej. Si sus niveles de gris son iguales).

Digamos que en una imagen binaria (con valores 0 y 1) 2 píxeles pueden ser 4-vecinos, pero sólo se dice que están conectados si tienen el mismo valor.

#### Conectividad

Sea  $V$  el conjunto de valores de niveles de gris que definen la conectividad. Consideremos 3 tipos de conectividad:

- 4-conectividad. 2 píxeles  $p$  y  $q$  con valores en  $V$  son 4-adyacentes si  $q$  está en el conjunto  $N_4(p)$ .
- 8-conectividad. 2 píxeles  $p$  y  $q$  con valores en  $V$  son 8-adyacentes si  $q$  está en el conjunto  $N_8(p)$ .
- $m$ -conectividad (conectividad mixta). 2 píxeles  $p$  y  $q$  con valores de  $V$  son  $m$ -adyacentes si
  - $1 - q$  está en  $N_4(p)$ , o
  - $2 - q$  está en  $N_D(p)$  y el conjunto  $N_4(p)$  INTERSECCIÓN  $N_4(q)$  no tiene píxeles con valores de  $V$

La conectividad mixta es una modificación de la 8-conectividad que se introduce para eliminar las ambigüedades que surgen al utilizar 8-conectividad. Por ejemplo,

considérese el arreglo de pixeles de la figura para  $V = \{1\}$ . Los 3 pixeles en la parte superior muestran múltiple 8-conectividad (ambigua), como se indica con las líneas punteadas. Esta ambigüedad desaparece al utilizar m-conectividad.

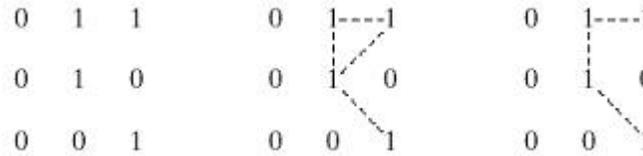


fig 2.19 La ambigüedad en 8-conectividad desaparece al utilizar 4-conectividad

### Caminos, componentes conexas, conjuntos conectados, regiones

Un *camino* de un pixel  $p$  con coordenadas  $(x, y)$  a un pixel  $q$  con coordenadas  $(s, t)$  es la secuencia de distintos pixeles con coordenadas

$$(x_0, y_0), (x_1, y_1), \dots, (x_n, y_n)$$

donde  $(x_0, y_0) = (x, y)$  y  $(x_n, y_n) = (s, t)$  y los pixeles  $(x_i, y_i)$  y  $(x_{i-1}, y_{i-1})$  son adyacentes para todo  $1 \leq i \leq n$ . En este caso,  $n$  es la longitud del camino. Si  $(x_0, y_0) = (x_n, y_n)$ , el camino es un *camino cerrado*.

Sea  $S$  un subconjunto de pixeles en una imagen. Se dice que 2 pixeles  $p$  y  $q$  están *conectados* en  $S$  si existe un camino entre ellos que consiste enteramente de pixeles en  $S$ . Para cada pixel  $p$  en  $S$ , el conjunto de pixeles que está conectado a él en  $S$  se denomina *componente conexas* de  $S$ . Si sólo tiene un componente conexas, entonces el conjunto  $S$  se denomina *conjunto conectado*.

Sea  $R$  un subconjunto de pixeles en una imagen. Llamamos  $R$  una *region de la imagen* si  $R$  es un conjunto conectado. La *frontera* (también llamada borde o contorno), de una región  $R$  es el conjunto de pixeles de la región que tienen uno o más vecinos que no están en  $R$ .

### Medidas de distancia

Para los pixeles  $p, q$  y  $z$  con coordenadas  $(x,y), (s,t)$  y  $(v, w)$ , respectivamente,  $D$  es una *función de distancia* o *métrica* si:

- (a)  $D(p, q) \geq 0$  ( $D(p,q) = 0$  sí y sólo sí  $p = q$ )
- (b)  $D(p, q) = D(q, p)$ , y
- (c)  $D(p, z) \leq D(p, q) + D(q, z)$

La *distancia euclidiana* entre  $p$  y  $q$  se define como

$$D_e(p, q) = [(x - s)^2 + (y - t)^2]^{1/2} \quad (2.5.2 -1)$$

Lo que significa que los pixels que tienen una distancia menor o igual al valor de  $r$  desde  $(x,y)$  son los que contiene un disco de radio  $r$  con centro en  $(x,y)$ .

La *distancia  $D_4$  (city-block)* entre  $p$  y  $q$  se define como

$$D_4(p,q) = |x - s| + |y - t| \quad (2.5.2 - 2)$$

En este caso, los píxeles que tienen una distancia  $D_4$  desde  $(x,y)$  de menor o igual, valor a  $r$  forman un diamante centrado en  $(x, y)$ . Por ejemplo, los píxeles con distancia  $D_4 \leq 2$  desde  $(x, y)$  (el punto central) forman los contornos con distancia constante siguientes:

```

2
2 1 2
2 1 0 1 2
2 1 2
2
```

Los píxeles con  $D_4 = 1$  son los 4-vecinos de  $(x,y)$ .

La *distancia*  $D_8$  (también llamada distancia de tablero de ajedrez) entre  $p$  y  $q$  se define como

$$D_8(p,q) = \max(|x - s|, |y - t|) \quad (2.5.2 - 3)$$

En este caso, los píxeles con distancia  $D_8$  menor o igual al valor  $r$  forman un cuadrado centrado en  $(x, y)$ . Por ejemplo, los píxeles con distancia  $D_8 \leq 2$  desde  $(x,y)$ , el punto central, forman los contornos con distancia constante siguiente:

```

2 2 2 2 2
2 1 1 1 2
2 1 0 1 2
2 1 1 1 2
2 2 2 2 2
```

Los píxeles con distancia  $D_8$  son los 8-vecinos de  $(x,y)$ .

Nótese que las distancias  $D_4$  y  $D_8$  entre  $p$  y  $q$  son independientes de cualquier camino que pueda existir entre los puntos porque estas distancias involucran solamente las coordenadas de los puntos. Si se considera  $m$ -conectividad, la distancia  $D_m$  entre 2 puntos se define como el  $m$ -camino más corto entre los puntos. En este caso la distancia entre 2 píxeles no sólo depende de los valores de los píxeles en el camino, sino también de los valores de sus vecinos.

Por ejemplo, considérese el siguiente arreglo de píxeles y asúmase que  $p, p_2$  y  $p_4$  tienen valor 1 y  $p_1$  y  $p_3$  pueden tener valor de 0 o 1

```

      p3 p4
    p1 p2
    p   .
```

Supongamos que consideramos la conectividad de píxeles que valen 1 ( $V = \{1\}$ ). Si  $p_1$  y  $p_3$  son 0, la longitud del  $m$ -camino más corto (la distancia  $D_m$ ) entre  $p$  y  $p_4$  es 2. Si  $p_1$  es 1, entonces  $p_2$  y  $p$  no serán  $m$ -adyacentes y la longitud del  $m$ -camino más corto es ahora 3. Finalmente, si  $p_1$  y  $p_3$  valen 1, la longitud del  $m$ -camino más corto entre  $p$  y  $p_4$  es 4. En este caso, el camino pasa por la secuencia de puntos  $p \ p_1 \ p_2 \ p_3 \ p_4$ .