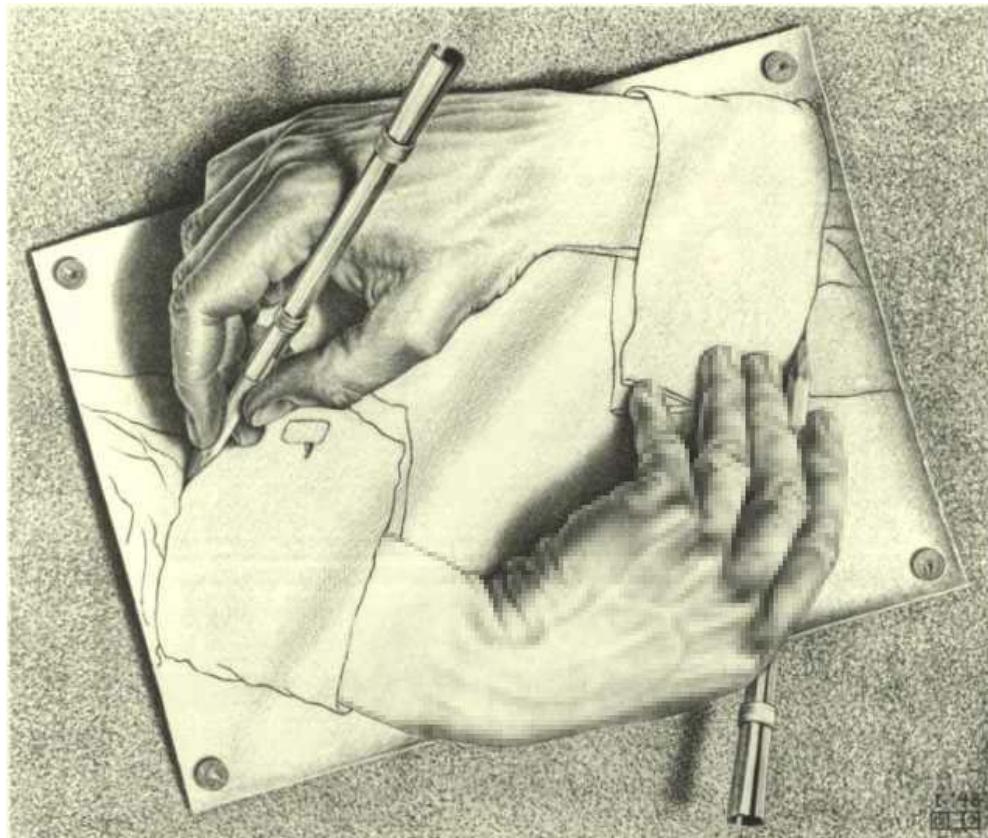


Apuntes de

Procesamiento Digital de Imágenes

(Primer borrador)

José Ramón Mejía Vilet
Área de Computación e Informática



Facultad de Ingeniería
UASLP

Apuntes de
Procesamiento Digital de Imágenes

Estos apuntes fueron elaborados durante el curso de agosto a diciembre de 2004 de la materia de Procesamiento Digital de Imágenes, en el Área de Computación e Informática de la Facultad de Ingeniería, Universidad Autónoma de San Luis Potosí. Los apuntes no están completos, ni he tenido tiempo de hacer una revisión exhaustiva, y están aún en versión “borrador”, por lo que agradeceré cualquier precisión, corrección o comentario enviada a la dirección jrmv@uaslp.mx.

San Luis Potosí, enero de 2005

* La portada es una versión editada de la obra de M. Escher “Manos dibujando”.

Índice

1. Introducción

- 1.1 Definiciones preliminares
 - ¿Qué entendemos por imagen?*
 - ¿Qué entendemos por imagen digital?*
 - ¿Qué es el procesamiento digital de imágenes?*
 - Procesos de Bajo Nivel.*
 - Procesos de Nivel Medio.*
 - Procesos de Alto Nivel.*
- 1.2 Antecedentes.
 - ¿Cuáles son los orígenes del procesamiento digital de imágenes?*
- 1.3 Ejemplos de campos de aplicación del PDI
 - 1.3.1 Rayos gamma
 - 1.3.2 Rayos X
 - 1.3.3 Ultravioleta
 - 1.3.4 Visible e infrarroja
 - 1.3.5 Microondas
 - 1.3.6 Ondas de radio
 - 1.3.7 Aplicaciones de otras fuentes de energía
- 1.4 Pasos fundamentales del procesamiento digital de imágenes
- 1.5 Componentes de un sistema de procesamiento de imágenes de propósito general

2. Fundamentos de la imagen digital

- 2.1 Elementos de la percepción visual
 - 2.1.1 Estructura del ojo humano
 - 2.1.2 Formación de imágenes en el ojo
 - 2.1.3 Adaptación a la iluminación y discriminación
 - Adaptación a la iluminación*
 - Adaptación a los cambios y cociente de Weber*
- 2.2 La luz y el espectro electromagnético
 - Ondas electromagnéticas y fotones.
 - Luz visible*
 - Luz monocromática*
 - Luz cromática*
 - 1. Radiancia
 - 2. Luminancia
 - 3. Brillo
- 2.3 Adquisición de imágenes
 - 2.3.1 Adquisición con un solo sensor
 - 2.3.2 Adquisición con bandas o líneas de sensores
 - 2.3.3 Adquisición por arreglos de sensores
 - 2.3.4 Un modelo simple de formación de imágenes
 - Nivel de gris y escala de grises*
- 2.4 Muestreo y cuantificación
 - 2.4.1 Conceptos básicos de muestreo y cuantificación
 - 2.4.2 Representación de imágenes digitales
 - 2.4.3 Resolución espacial y resolución en niveles de gris
 - 2.4.4 Aliasing y patrones de Moiré
 - 2.4.5 Cambiando el tamaño de la imagen
- 2.5 Algunas relaciones básicas entre píxeles
 - 2.5.1 Vecinos de un pixel
 - 2.5.2 Conectividad, regiones y fronteras
 - Conectividad*
 - Caminos, componentes conexas, conjuntos conectados, regiones*
 - Medidas de distancia*
 - 2.5.4 Operaciones entre píxeles
- 2.6 Operaciones lineales y no lineales

3 Mejora de la imagen

- 3.1 Definiciones previas
 - Vecindad*
 - Procesamiento puntual y Aumento del contraste*
 - Máscaras*
- 3.2 Algunas transformaciones del nivel de gris básicas
 - 3.2.1 Negativos de la imagen
 - 3.2.2 Transformaciones logarítmicas
 - 3.2.3 Transformaciones de función de potencia
 - 3.2.4 Funciones de transformación lineal por partes
 - 3.2.4.1 Aumento del contraste
 - 3.2.4.2 Fraccionamiento del nivel de gris
 - 3.2.4.3 Fraccionamiento de planos de bits
- 3.3 Procesamiento del histograma
 - 3.3.1 Ecualización del histograma

4 Mejora de la imagen en el dominio de la frecuencia

- 4.1 Antecedentes
 - Serie de Fourier*
 - Transformada de Fourier*
- 4.2 Introducción a la transformada de Fourier y el dominio de la frecuencia
 - 4.2.1 La transformada de Fourier en una dimensión y su inversa
 - El dominio de la frecuencia*
 - 4.2.2 La DFT bidimensional y su inversa
 - 4.2.3 Filtrado en el dominio de la frecuencia
 - Propiedades básicas del dominio de la frecuencia*
 - Ideas básicas del filtrado en el dominio de la frecuencia*
 - Algunos filtros básicos y sus propiedades*
 - 4.2.4 Correspondencia entre el filtrado en los dominios espacial y de frecuencia

5 Procesamiento del color

- 5.1 Fundamentos del color
 - Los conos, los colores primarios y los colores secundarios*
 - Tono, brillo y Saturación*
 - Valores de tristímulo*
 - El diagrama cromático de CIE*
- 5.2 Modelos de color
 - 5.2.1 El modelo RGB
 - 5.2.2 Los modelos CMY y CMYK
 - 5.2.3 El modelo HSI
 - Conversión de RGB a HSI*
 - Conversión de HSI a RGB*
 - Manipulación de los componentes HSI*
- 5.3 Procesamiento del falso color
 - 5.3.1 División del color
 - 5.3.2 Transformaciones de nivel de gris a color
- 5.4 Introducción al procesamiento de imágenes de color
- 5.5 Transformaciones de color
 - 5.5.2 Complementos de color
 - 5.5.3 Fraccionamiento del color
 - 5.5.4 Correcciones de color y tono
 - 5.5.5 Procesamiento del histograma
- 5.6 Suavizado y realce
 - 5.6.1 Suavizado de imágenes a color
 - 5.6.2 Realzado de imágenes a color
- 5.7 Segmentación de color
 - 5.7.1 Segmentación en HSI
 - 5.7.2 Segmentación en el espacio vectorial RGB

6 Compresión de imágenes

6.1 Fundamentos

Compresión y redundancia de datos

6.1.1 Redundancia de código

6.1.2 Redundancia entre pixeles

6.1.3 Redundancia psicovisual

6.1.4 Criterios de fidelidad

6.2 Modelos de compresión de imágenes

6.2.1 El codificador y el decodificador de fuente

6.2.2 El codificador y decodificador de canal

6.3 Elementos de la teoría de la información*

6.3.1 Midiendo la información

6.3.2 El canal de información

6.4 Compresión libre de errores

6.4.1 Codificación por longitud variable

El código de Huffman

Otros códigos de longitud variable casi óptimos

Huffman truncado

Código B

Shift codes

Codificación aritmética

Codificación LZW

* Este tema se verá sólo si queda tiempo.

Apuntes de

Procesamiento Digital de Imágenes

1. Introducción

1.1 Definiciones preeliminares

¿Qué entendemos por imagen?

Una imagen puede ser definida matemáticamente como una función bidimensional,

$$f(x,y),$$

donde x y y son coordenadas *espaciales* (en un plano), y f en cualquier par de coordenadas es la *intensidad* o *nivel de gris* de la imagen en esa coordenada.

¿Qué entendemos por imagen digital?

Cuando x, y , y los valores de f son todas cantidades finitas, discretas, decimos que la imagen es una *imagen digital*.

Una imagen digital se compone de un número finito de elementos, cada uno con un lugar y valor específicos. Estos elementos son llamados *pels*, o *pixels*.

¿Qué es el procesamiento digital de imágenes?

En el resto de la sección se intentará contestar esta pregunta basándose en el dominio que presentan González y Woods en el libro de texto de la materia¹.

La vista es nuestro sentido más avanzado, y no es sorprendente que las imágenes jueguen el papel más importante en la percepción humana.

Aunque los seres humanos estemos limitados a la banda visible del espectro electromagnético (EM), las máquinas pueden percibir casi el espectro completo, desde los rayos gamma, a las ondas de radio.

Las máquinas también pueden procesar imágenes generadas por fuentes que los humanos no asociamos con imágenes; como es el caso del ultrasonido, la microscopía de electrones, etc.

Los autores no se ponen de acuerdo para decir dónde termina el campo del Procesamiento Digital de Imágenes y dónde empiezan otros campos como el Análisis de Imágenes y la Visión por Computadora. La tarea de esta última disciplina es utilizar computadoras para emular la visión humana, incluyendo el aprendizaje, hacer inferencias y actuar basándose en entradas visuales. La Visión Computacional es, claramente, un área de la Inteligencia Artificial.

Para propósitos de nuestro curso, consideraremos tres tipos de procesos que comienzan en el PDI y terminan en la VC:

¹ González y Woods, Digital Image Processing, 2nd Edition .

Procesos de Bajo Nivel:

Utilizan operaciones como el *preprocesamiento de imagen* para reducir el ruido, *mejora del contraste*, y *filtros de enfoque*.

Se caracterizan por que sus entradas son imágenes y sus salidas también.

Procesos de Nivel Medio:

Operaciones como *segmentación* y *clasificación* de objetos individuales.

Se caracterizan por que sus entradas son generalmente imágenes, pero sus salidas son atributos extraídos de esas imágenes (*contornos*, *bordes*, *identidad de objetos individuales*).

Procesos de Alto Nivel:

Implica el obtener algún significado de un *conjunto de objetos reconocidos* – análisis de imágenes – y, finalmente, realizar las funciones *cognitivas* asociadas con la vista.
(ejemplo de símbolos de tráfico)

Como hemos visto en nuestra clasificación de procesos, el translape del PDI y el Análisis de Imágenes se da en el área de reconocimiento de objetos y regiones individuales en una imagen.

En este curso llamaremos *Procesamiento Digital de Imágenes* a los procesos cuyas entradas y salidas son imágenes (procesos de bajo nivel) y, además, a aquellos procesos que extraen atributos de imágenes, incluyendo el reconocimiento de objetos individuales (procesos de nivel medio). Todos los procesos se llevarán a cabo con la ayuda de una computadora digital.

Ejemplo: Análisis de texto

Para ilustrar lo anterior consideremos el análisis de texto. Los procesos de *adquisición* de la imagen del área que contiene el texto, el *preprocesamiento* de la imagen, la *extracción (segmentación)* de caracteres individuales, la *descripción* de los caracteres de una forma aceptable para el procesamiento computacional, y el *reconocimiento* de cada carácter individual entra en el campo de lo que llamaremos PDI (la *interpretación* de lo que dice el texto, o aún el *reconocimiento* de palabras, quedan fuera de los propósitos de este curso, y ya en el campo del Análisis de Imágenes y la Visión Computacional).

1.2 Antecedentes.

¿Cuáles son los orígenes del procesamiento digital de imágenes?

En nuestra definición de PDI aclaramos que éste conlleva el uso de una computadora digital, por lo tanto los orígenes de uno y la otra están íntimamente relacionados. Sin embargo hay antecedentes que vale la pena comentar.

Una de las primeras aplicaciones de las imágenes digitales se dio en la industria del periódico cuando se enviaron fotografías a través de cable submarino entre Londres y Nueva York, en la primera parte de la década de los veintes. El sistema Bartlane tomaba cerca de tres horas y constaba de equipo de impresión especializado que codificaba las fotografías para su envío por cable y las reconstruía del otro lado.



Fig. 1.1 Fotografía digital de 1921 obtenida con una impresora telegráfica con tipos especiales.

Esta técnica se abandonó rápidamente, favoreciendo otro tipo de reproducción fotográfica basada en cintas que se perforaban en la terminal telegráfica receptora (fig 1.2)

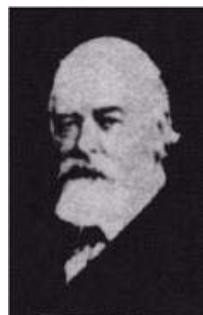


Fig 1.2 Fotografía digital de 1922 obtenida a partir de una cinta perforada después de que las señales cruzaron el atlántico dos veces. Se pueden apreciar errores.

Como ya mencionamos, aunque estas imágenes entran en nuestra definición de imagen digital, no podemos decir que son resultados del Procesamiento Digital de Imágenes pues no se usaron computadoras para crearlas. De hecho, las imágenes digitales necesitan tanto espacio de almacenamiento y capacidad de procesamiento que el avance del PDI ha dependido del desarrollo de las Computadoras digitales y tecnologías de apoyo que incluyen almacenamiento de datos, despliegue y transmisión.

Las primeras computadoras suficientemente potentes para desarrollar tareas de PDI significativas aparecieron en los comienzos de los sesentas, junto con el programa espacial estadounidense. El laboratorio Jet Propulsion de Pasadena California inició los trabajos en 1964 cuando un equipo procesó varias fotos de la luna transmitidas por el Ranger 7, para corregir varios tipos de distorsión que producía la cámara de a bordo (fig 1.3)

Estas investigaciones servirían de base para desarrollos posteriores en subsecuentes misiones como Surveyor, Mariner y Apollo.

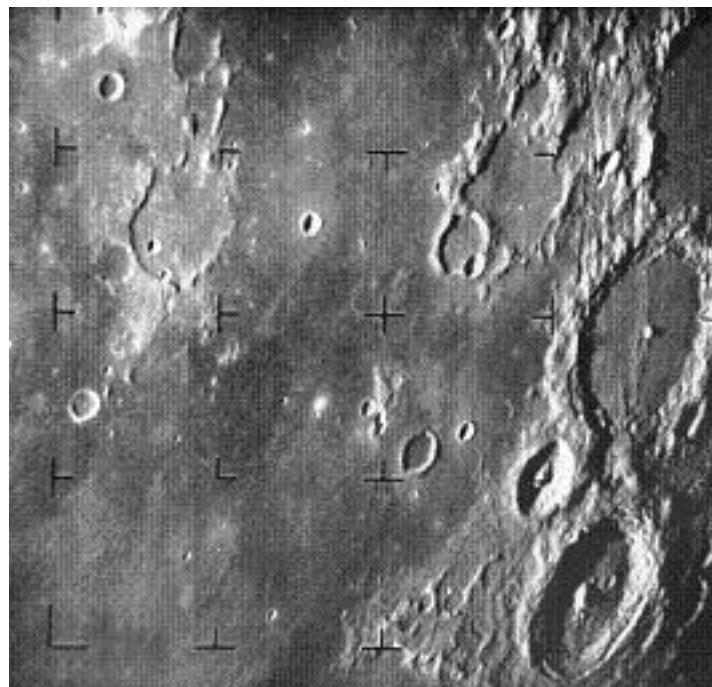


Fig.1.3 La primera fotografía de la luna de una nave espacial estadounidense (31 de julio de 1964 a las 9:09 am). Las marcas se usan para corregir la curvatura.

Paralelamente, en los años 60 y 70, se desarrollaron técnicas para imágenes médicas, observaciones remotas de la tierra y astronomía. La invención en los años 70s de la tomografía axial computarizada (CAT por sus siglas en inglés), fue uno de los más importantes eventos en la aplicación de PDI a diagnósticos médicos.

La CAT consta de un anillo que rodea al paciente mientras una fuente de rayos X concéntrica rota a su alrededor. Los rayos X pasan a través del paciente y se recolectan en el lado opuesto del anillo. El paciente va avanzando y se obtienen “rebanadas” a partir de las cuales se puede formar una imagen 3D del interior del paciente.

A partir de los años sesenta y con el desarrollo de las computadoras, el PDI ha avanzado a pasos impresionantes y sus algoritmos se empezaron a utilizar en un amplio rango de aplicaciones (Geólogos que estudian la contaminación con imágenes satelitales o aéreas, arqueólogos que restauran viejas fotografías de antiguos artefactos destruidos con el tiempo, aplicaciones en astronomía, biología, defensa, la industria, reconocimiento de caracteres, control de calidad, etc). El abaratamiento de equipo de alta tecnología y el advenimiento de la red Internet han creado oportunidades jamás vistas para el crecimiento de la PDI.

En la siguiente sección se ilustrarán algunas de las áreas de aplicación del PDI.

1.3 Ejemplos de campos de aplicación del PDI

Hoy en día casi no hay áreas de enfoque técnico que no hayan sido impactadas de alguna manera por el procesamiento digital de imágenes.

Las áreas son tan abundantes y variadas que nos forzan a utilizar un criterio de clasificación para dar algunos ejemplos.

En este caso nuestro criterio de clasificación será por su fuente. La principal fuente de energía de las imágenes es el espectro electromagnético (otras fuentes son la acústica, la ultrasónica, y la electrónica). Las imágenes basadas en la radiación del espectro EM son las más familiares, especialmente las imágenes de rayos X y las bandas visuales del espectro.

Las ondas electromagnéticas pueden ser conceptualizadas como ondas sinusoidales de longitudes de onda variadas, o como una corriente de partículas sin masas viajando en un patrón de ondas y moviéndose a la velocidad de la luz. Su unidad es el fotón.

Agrupando las bandas espectrales de acuerdo a su energía por fotón, obtenemos el espectro de la figura, desde los rayos gamma (con mayor energía) hasta las ondas de radio.

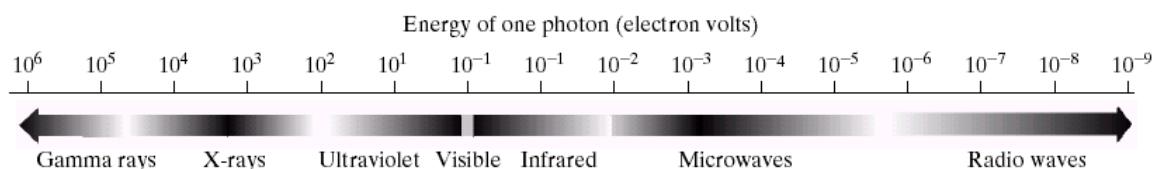


FIGURE 1.5 The electromagnetic spectrum arranged according to energy per photon.

Fig 1.4 Espectro EM en energía por fotón.

1.3.1 Rayos gamma

Sus aplicaciones más importantes: la medicina nuclear y las observaciones astronómicas.

Medicina nuclear

Se inyecta al paciente un isótopo radioactivo que emite rayos gamma cada vez más débiles, estos se detectan y se forma la imagen.

Otro caso es la Tomografía de Emisión de Positrones PET, similar a la tomografía de rayos X, pero en esta se inyecta al paciente un isótopo radioactivo que emite positrones. Los positrones chocan con los electrones del paciente, se destruyen ambos, y se producen 2 rayos gamma que se detectan para formar la imagen.

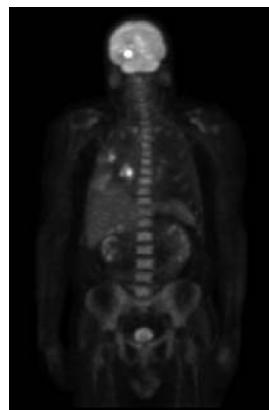


Fig 1.5 Tomografía por emisión de positrones (PET)

Astronomía

Se detecta la radiación natural de rayos gamma de los astros.

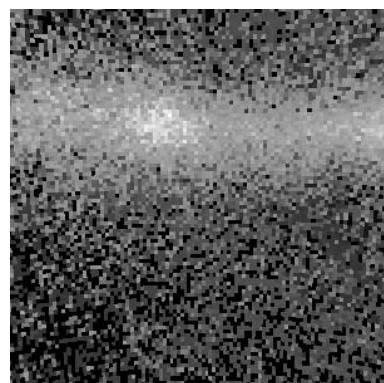


Fig 1.6 Imagen a partir de rayos gamma de Cygnus Loop

1.3.2 Rayos X

Los rayos X no sólo se usan en medicina, sino también en gran medida en la industria y otras áreas, como la astronomía.

(explicación de cómo funciona, ánodo y cátodo)

Angiografía

Es una aplicación de la radiografía de mejora de contraste. Se utiliza para obtener angiogramas, imágenes de venas y arterias. Un catéter se inserta en una de estas y se guía hasta el área a estudiar. Una vez ahí, se inyecta por medio de éste un medio que mejora el contraste y permite al radiólogo ver irregularidades o bloqueos.



fig 1.7 Angiograma de la aorta

CAT, Tomografía axial computarizada



fig 1.8 Tomografía de la cabeza

Aplicaciones en la industria

Generalmente en control de calidad, se utilizan rayos X de alta energía.

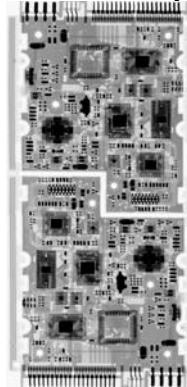


fig 1.9 Circuito integrado (rayos X)

1.3.3 Ultravioleta

Las aplicaciones incluyen litografía, inspección industrial, microscopía, lasers, imágenes biológicas y observaciones astronómicas.

Microscopía de fluorescencia

La fluorescencia se observó en el siglo diecinueve, al dirigir luz ultravioleta a un mineral con fluor se observó que este brillaba. La luz ultravioleta no es visible, pero cuando un fotón ultravioleta choca con un electrón de un átomo de material fluorescente, este electrón se eleva a un nivel más alto de energía. Después éste regresa a un nivel bajo y emite luz en forma de un fotón de baja energía en la región de luz visible (rojo).

La microscopía de fluorescencia utiliza este principio, dirigiendo una luz de excitación a un objeto preparado y separando la radiación ultravioleta producida (débil) de la producida por la luz de excitación (brillante).

Fig 1.8 a b
Ejemplos de imágenes ultravioleta.
a) Maíz sano.
b) Maíz enfermo (smut)

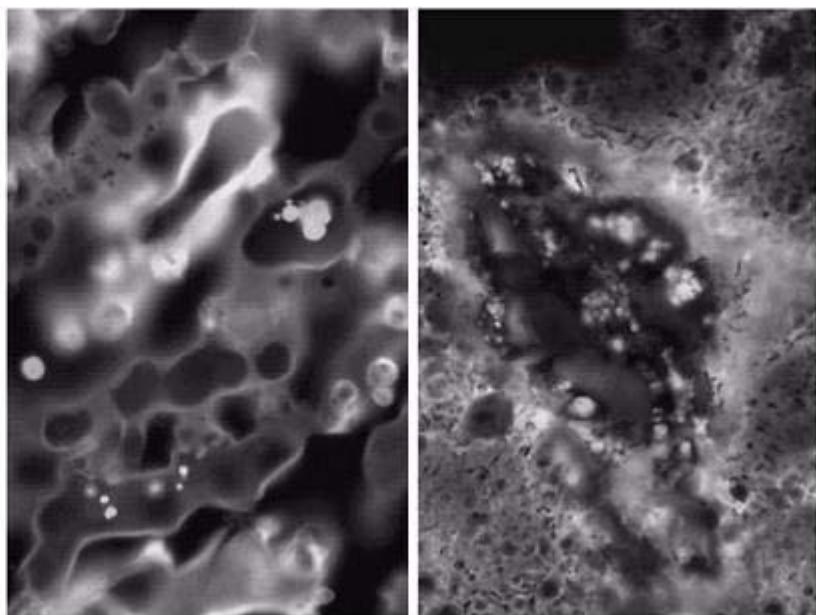


fig 1.10

1.3.4 Visible e infrarroja

Estas son, por mucho, las aplicaciones más numerosas. La banda infrarroja se utiliza usualmente en conjunto con la imagen visual.

Mencionamos algunos ejemplos de microscopía, astronomía, detección remota, industria, y policía.

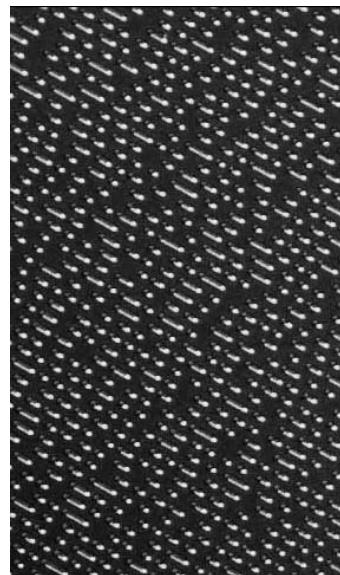


Fig 1.11 Superficie de un CD de audio al microscopio

Detección remota

Usualmente incluye varias bandas que detectan diferentes detalles.

Banda	Nombre	Longitud de onda	Usos y características
1	Azul visible	0.45 – 0.52	Penetración máxima del agua
2	Verde visible	0.52 – 0.60	Bueno para medir la salud de las plantas
3	Rojo visible	0.63 – 0.69	Discriminación de vegetación
4	Infrarrojo bajo	0.76 – 0.90	Mapeo de costas y biomasa
5	Infrarrojo medio	1.55 – 1.75	Contenido de humedad en el suelo y vegetación
6	Infrarrojo termal	10.4 – 12.5	Humedad del suelo y mapeo termal
7	Infrarrojo alto	2.08 – 2.35	Mapeo mineral

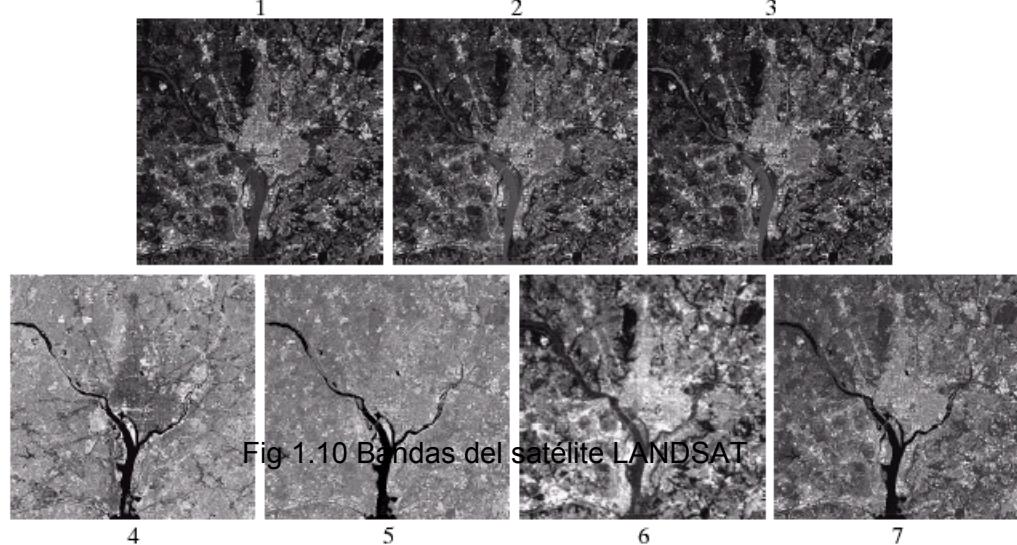


Fig 1.10 Bandas del satélite LANDSAT

FIGURE 1.10 LANDSAT satellite images of the Washington, D.C. area. The numbers refer to the thematic bands in Table 1.1. (Images courtesy of NASA.)

Fig 1.12

Estas imágenes pueden ser usadas para ver, en ciudades, movimientos o crecimiento de población, contaminación, u otros factores que dañan al ambiente. Otras aplicaciones incluyen observación y predicción del clima, o, como en el caso de la base de datos de fotografías Nighttime lights of the world, un inventario de asentamientos humanos mundiales (infrarrojo).

Control de calidad

Una aplicación muy utilizada de la luz visible e infrarroja es en el control de calidad, donde se utiliza para revisar productos de orden variado (niveles correctos en botellas, circuitos integrados con todos los componentes, errores de fabricación o material defectuoso).

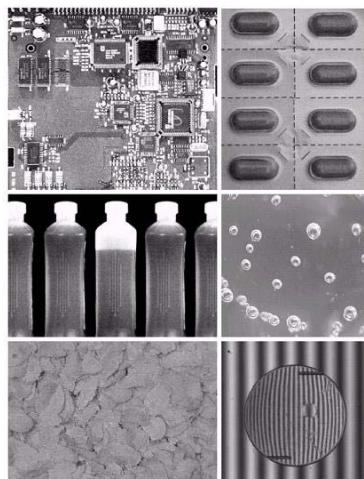


fig 1.13 Control de calidad en circuito integrado, empaque de píldoras, nivel de líquidos, burbujas en plástico transparente, cereal, y fallas en una lente intraocular.

1.3.5 Microondas

Su aplicación principal es el radar. El radar puede obtener datos de virtualmente cualquier región a cualquier hora, sin importar condiciones de clima o de luz ambiental (puede penetrar nubes, vegetación, hielo, arena).

Un radar utiliza su propia “iluminación” (pulsos de microondas) para obtener una imagen obtenida a partir de la energía reflejada.

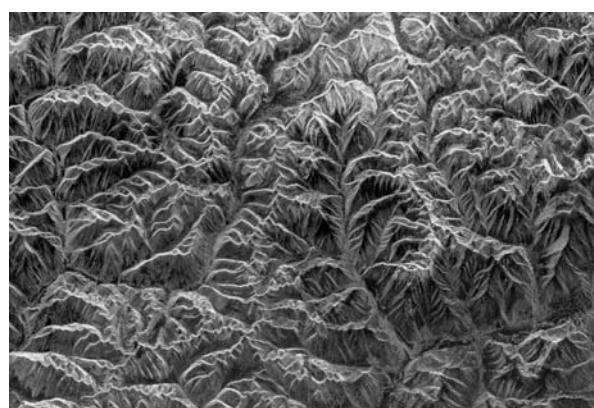


Fig 1.14 Imagen de radar del tibet.

1.3.6 Ondas de radio

Mayoritariamente se usa en medicina y astronomía.

Medicina

Resonancia magnética (MRI). Se introduce al paciente en un imán de gran poder y se pasan ondas de radio por su cuerpo en pulsos cortos. Cada pulso causa un pulso resultante de ondas de radio desde los tejidos del paciente. La computadora determina el lugar y fuerza de estos pulsos, y produce una imagen bidimensional de una sección del paciente.

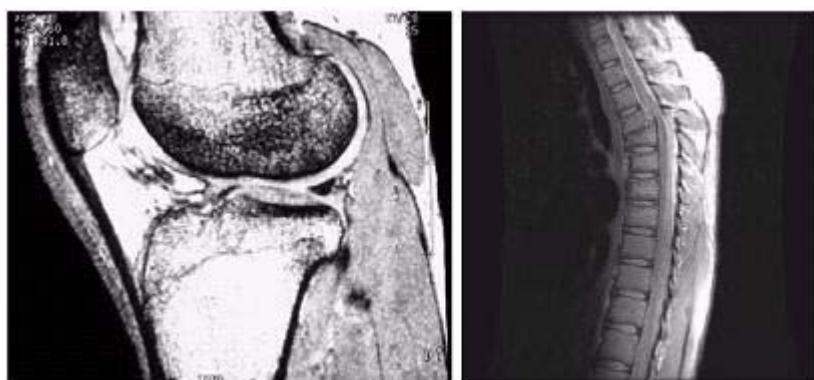


fig 1.15 Imágenes MRI de una rodilla y una espina.

1.3.7 Aplicaciones de otras fuentes de energía

Sonido

Se aplica en exploración geológica (petróleo y minerales), la industria, y medicina. En la primera se utiliza sonido en la parte baja del espectro (cientos de Hertz), y las otras áreas utilizan ultrasonido.

Por ejemplo, en geología se utiliza un gran camión para presionar un aplaca en el suelo, y se hace vibrar el camión a una frecuencia de cerca de 100 Hz. Las ondas que regresan se analizan.

Un método similar se utiliza para adquisición submarina, al disparar 2 pistolas de aire y recuperar las ondas de regreso para producir un mapa 3D del fondo del océano.

Microscopía de electrones

Un microscopio de transmisión de electrones (TEM) trabaja como un proyector de acetatos. El proyector transmite un rayo de electrones a través del espécimen y se proyecta en una pantalla de fósforo.

Un microscopio de escaneo de electrones escanea el rayo de electrones y registra la interacción del rayo y la muestra en cada lugar.

1.4 Pasos fundamentales del procesamiento digital de imágenes

Dividiremos el material de los siguientes capítulos en métodos en que la entrada y salida es una imagen, y métodos en los que la entrada es una imagen y la salida son datos. Las partes fundamentales del procesamiento digital de imágenes son las siguientes. Como lo indica la figura, después de adquirir la imagen, en un procesamiento no siempre se siguen todos los pasos.

1. Adquisición de imágenes

Generalmente incluye preprocessamiento (p. Ej. Escalar la imagen).

2. Mejora de la imagen

La idea de este paso es obtener detalles que no se veían, o simplemente subrayar ciertas características de interés. Que se “vea mejor”.

3. Restauración de la imagen.

También mejora la apariencia de la imagen, a diferencia de la mejora de la imagen, subjetiva, la restauración es objetiva, en el sentido en que las técnicas de restauración tienden a ser modelos probabilísticos o matemáticos de degradación de la imagen (¿Cómo era esta imagen antes de ser dañada?)

4. Procesamiento del color.

Procesamientos especiales para el color.

5. Ondeletas

Fundamentalmente utilizadas para representar imágenes en varios grados de resolución. Se utiliza principalmente en compresión.

6. Compresión

Reduce el almacenamiento requerido para guardar una imagen, o el ancho de banda para transmitirla.

7. Procesamiento morfológico

Herramientas para extraer componentes de la imagen útiles para la representación y descripción de formas.

8. Segmentación

Divide una imagen en sus partes constituyentes.

9. Representación y descripción.

Se toman decisiones tales como si la forma obtenida debe ser tratada como un frontera o una región, y extrae atributos que resultan en información cuantitativa de interés.

* Base de conocimiento. Nos sirve de almacenamiento de información: Ya sean los datos de las imágenes en cada paso, o heurísticas que nos ayudarán a obtener mejores imágenes en subsecuentes procesamientos.

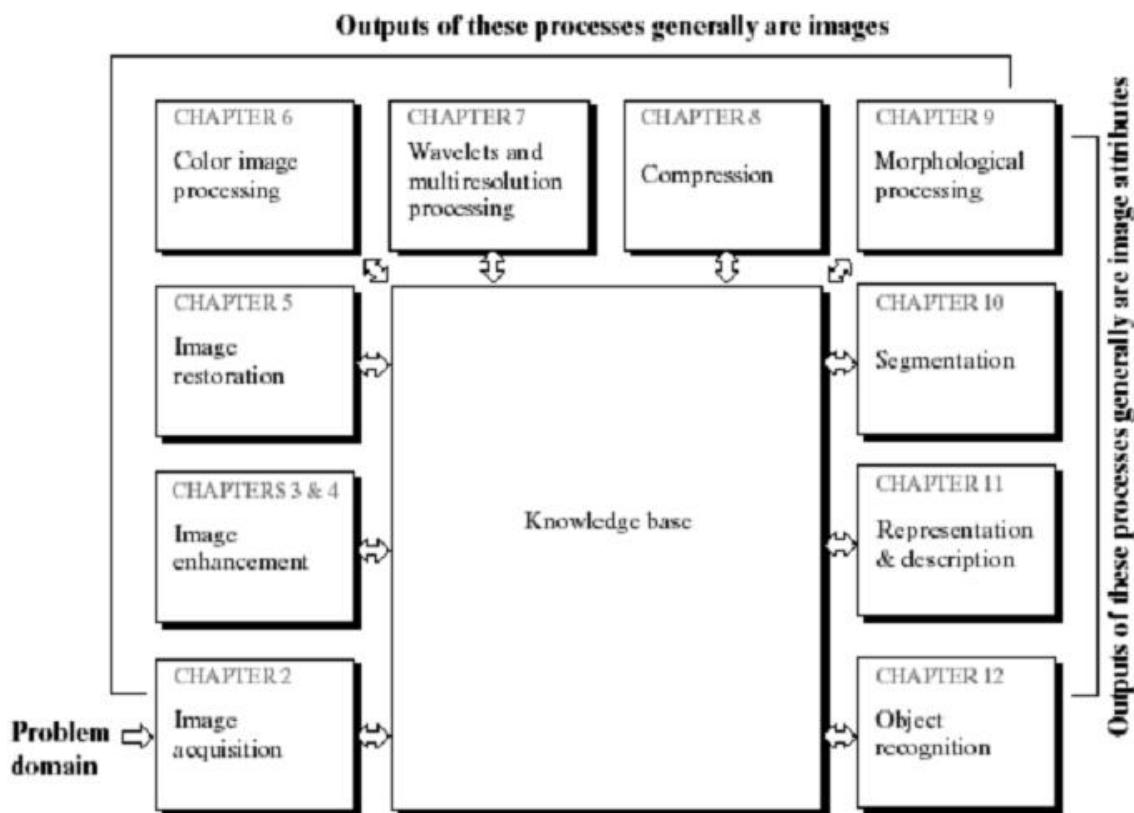


Fig 1.16 Pasos del procesamiento digital de imágenes

1.5 Componentes de un sistema de procesamiento de imágenes de propósito general

1. Sensor

Es un aparato físico sensible a la energía radiada por el objeto.

2. Hardware de procesamiento de imágenes especializado

Usualmente consta de un digitalizador que convierte la salida del sensor a formato digital, y de hardware que realiza otras operaciones, como una ALU, que procesa operaciones en paralelo de imágenes completas. Su característica más distintiva es la velocidad (realiza operaciones que requieren salidas de datos muy rápidas).

3. Computadora

En un sistema de propósito general puede ser desde una PC hasta una supercomputadora.

4. Software

Módulos especializados que llevan a cabo tareas específicas.

5. Almacenamiento en masa

Obligatorio en aplicaciones de procesamiento de imágenes. El almacenamiento se divide en tres categorías:

- a corto plazo, utilizado durante el procesamiento (ej. Memoria de la computadora, o memorias especiales “frame buffers”, éstas últimas se utilizan en zooms digitales)

- en línea, para carga relativamente rápida (discos magnéticos u ópticos)
- de archivo, de acceso poco frecuente (cintas magnéticas y discos ópticos).

6. Dispositivos de despliegue de imágenes

Por ejemplo, monitores a color (de preferencia planos, para evitar distorsiones), o dispositivos estéreo tales como lentes para estereo visión.

7. Dispositivos de copia dura (hardcopy)

Impresoras láser o tinta, cámaras de película, dispositivos sensibles al calor, unidades digitales como cd-rom.

8. Red

Casi una función por defecto en cualquier sistema computacional moderno, lo más importante es el ancho de banda.

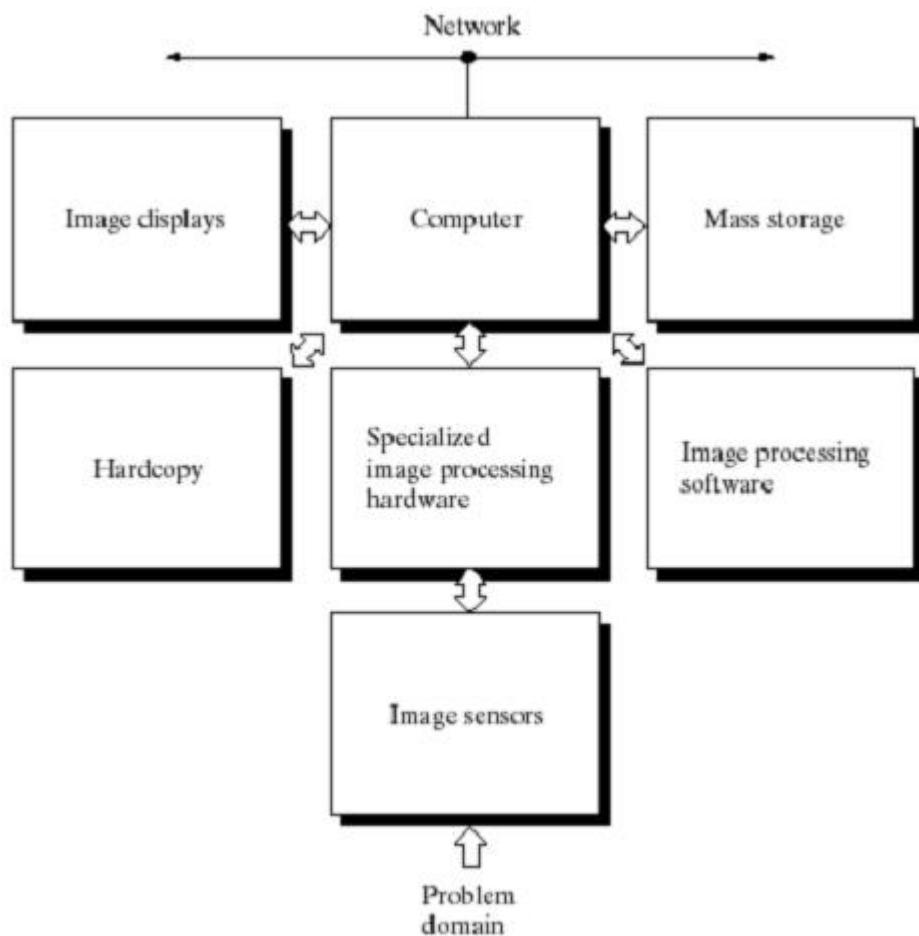


Fig 1.17 Componentes de un sistema de PDI de propósito general

Apuntes de

Procesamiento Digital de Imágenes

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.

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.

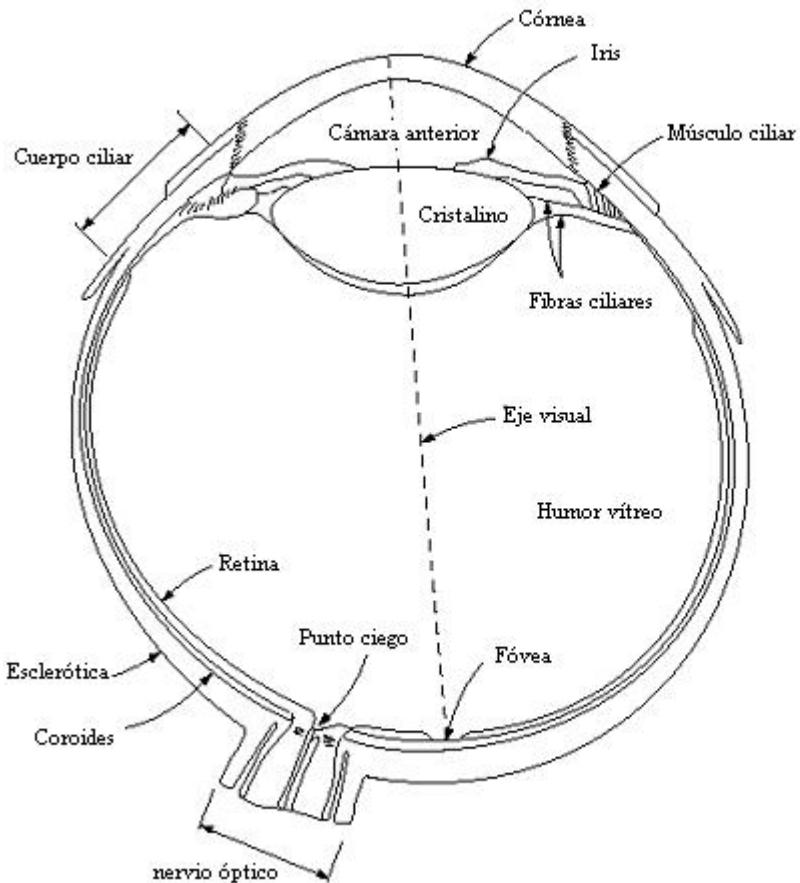


Figura 2.1
Diagrama
simplificado de un
corte del ojo humano

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. 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

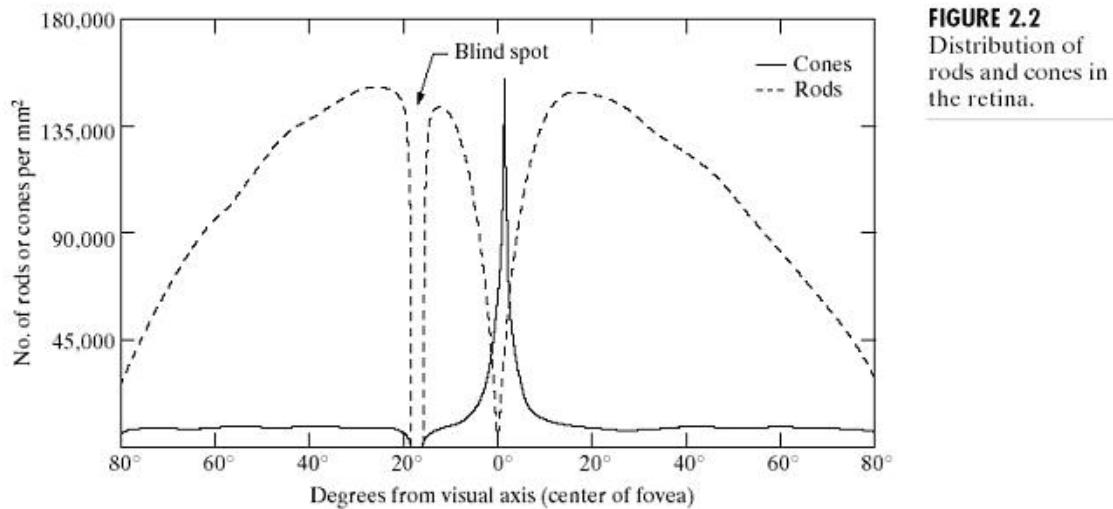


FIGURE 2.2
Distribution of rods and cones in the retina.

La fóvea es una región circular de la retina de aproximadamente 1.5 mm de diámetro.

Con ciertas libertades de interpretación, podemos pretender que la fóvea es un arreglo cuadrado de sensores de 1.5×1.5 mm. La densidad de conos ahí es de aproximadamente 150,000 elementos por mm^2 . Basado en esas aproximaciones, el número de conos en la región de más agudeza en el ojo es de cerca de 337,000 elementos.

Comparando *solamente* el poder de resolución, un chip de imágenes de dispositivo CCD (charge-coupled device) de resolución media puede tener este número de elementos en un arreglo receptor no mayor de 5×5 mm.

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 en el mismo rango de los sensores de imágenes electrónicos.

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 aplana, para enfocar objetos cercanos, se ensancha. 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. Consideraremos 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:

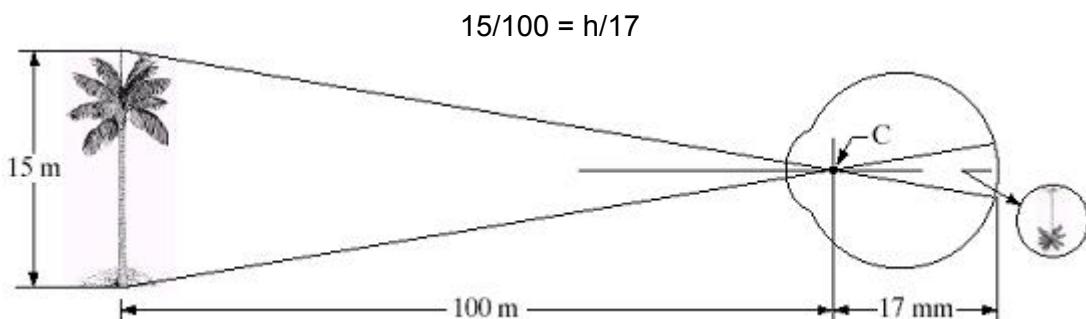


fig. 2.3 Obtención del tamaño de la imagen en la retina

Por lo tanto $h=2.55$ 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 decodificación².

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* (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).

² Para profundizar en este tema se puede consultar el libro de Dayan y Abbott, “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.

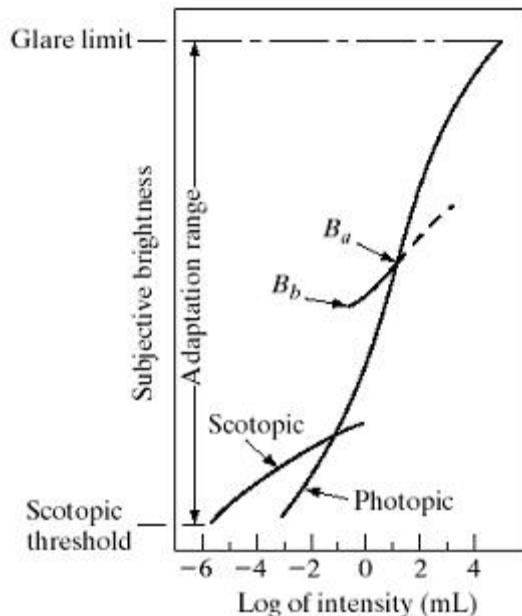


fig 2.4 Rango de sensaciones de iluminación subjetiva respecto a intensidad

Esto se ve 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 típicamente iluminada por la parte de atrás por una fuente cuya *intensidad I*, puede variar.

Se incrementa la iluminación, ΔI , con destellos de corta duración con forma de círculo en el centro del campo iluminado.

Si ΔI no es suficientemente brillante, el observador dirá "no". Mientras ΔI crece, si el observador lo percibe dirá "sí", cuando lo perciba. Finalmente, cuando ΔI sea lo suficientemente fuerte, el observador dirá "sí" todo el tiempo.

La cantidad $\Delta I_c / I$, donde ΔI_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.

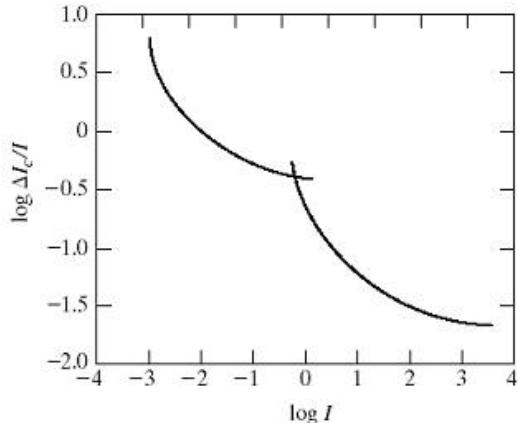


fig 2.4 Cociente de Weber como función de la intensidad

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 a flashazos, 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).

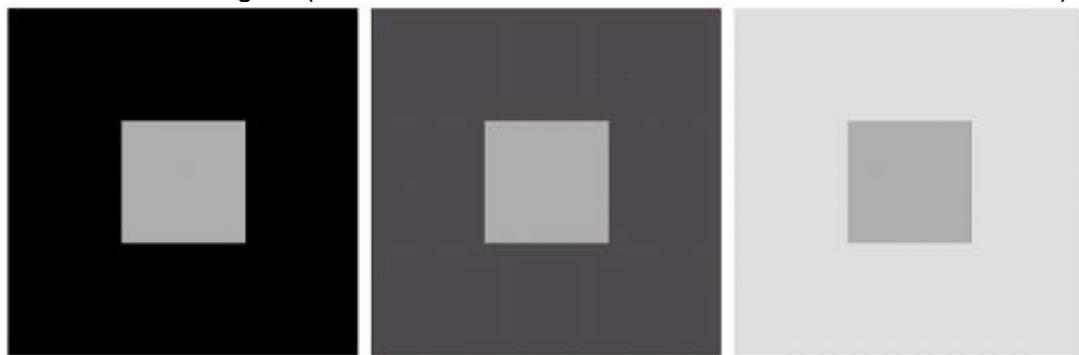
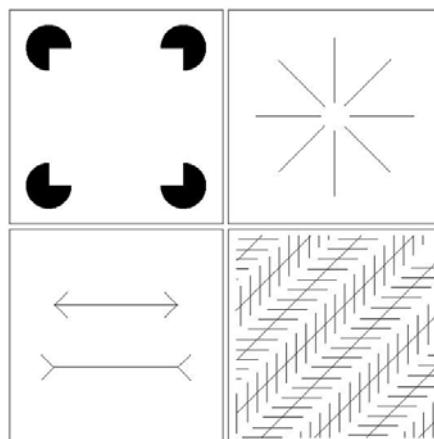


fig 2.7 Los cuadrados centrales parecen de distinto color y son exactamente el mismo

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.

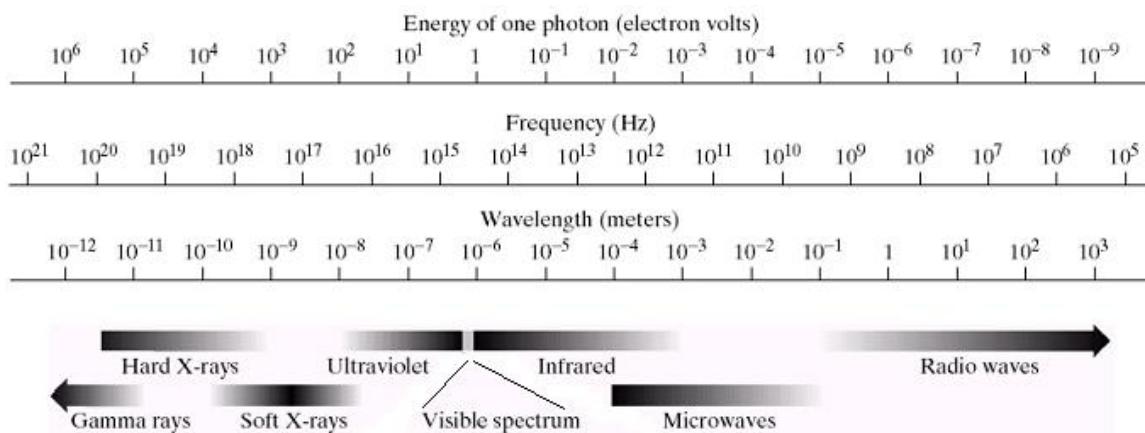


2.8 Ilusiones ópticas (un cuadrado que no existe, un círculo que no está delineado, dos líneas que aparentan ser de diferentes tamaños y líneas que siendo paralelas no lo aparentan)

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 (λ) en metros, frecuencia (v) en hertz, o energía (E) en electron volts. La longitud de onda y la frecuencia se relacionan por:

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

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

Y la energía es dada por

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

Donde h es la constante de Planck.

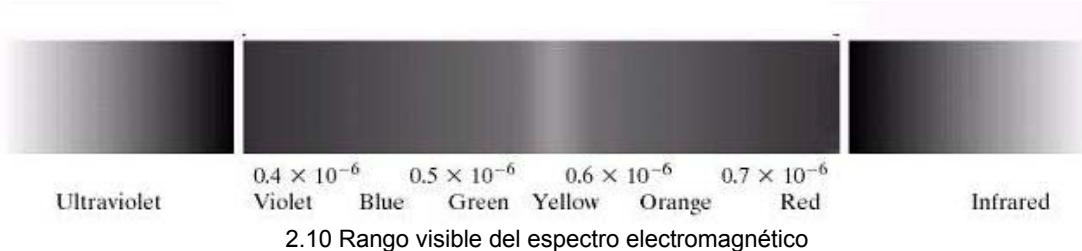
Ondas electromagnéticas y fotones.

Las ondas electromagnéticas se pueden visualizar como ondas sinusoidales propagadas de longitud de onda λ , 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 MICRÓm (violeta), a 0.79 MICRÓm (rojo).



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

Un cuerpo refeljante 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 (iluminación)

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

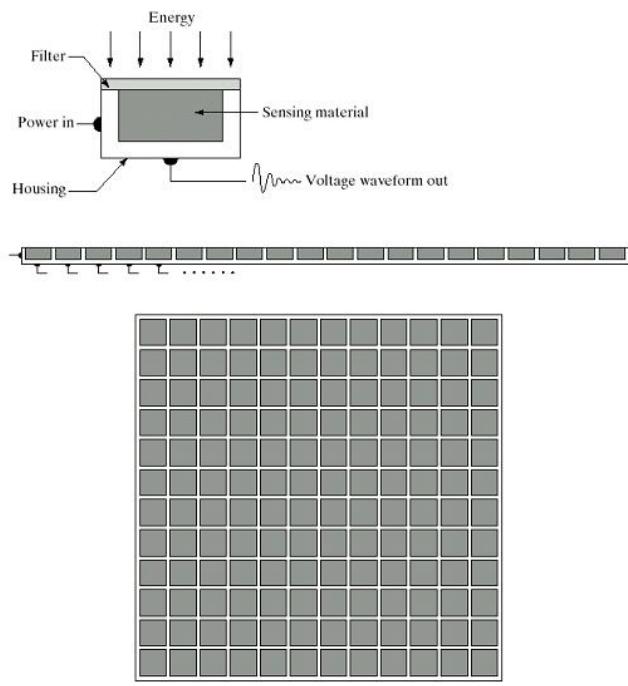
Los tipos de imágenes que nos interesan se generan por una combinación de una fuente de “iluminación” y la reflexión o absorción de energía de esta fuente por parte de los elementos de la escena.

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.

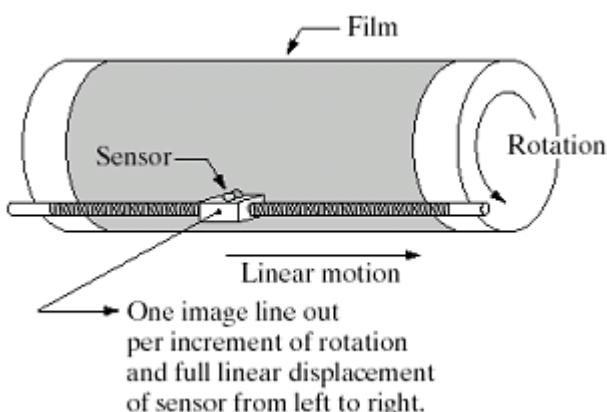
2.3.1 Adquisición con un solo sensor

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*.



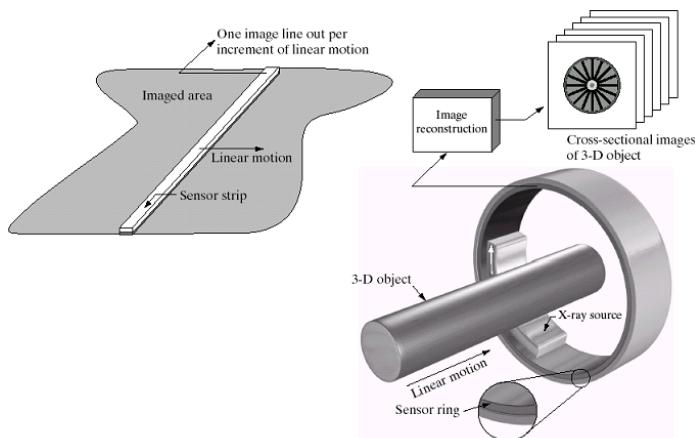
2.11 Se combina un sensor único con movimiento para generar una imagen 2D

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 (MRI).

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.

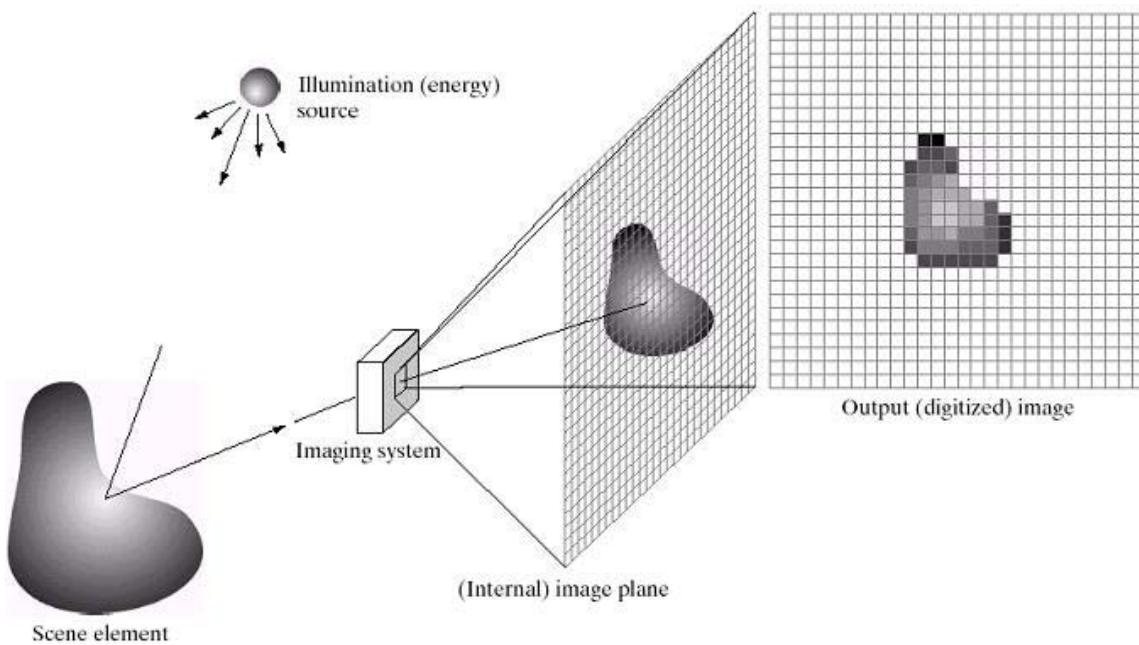
2.3.3 Adquisición por arreglos de sensores

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

Este es también el tipo de ordenamiento de sensores que encontramos en las cámaras digitales. Un sensor típico de estas cámaras es el arreglo CCD.

La respuesta de cada sensor es proporcional a la integral de la energía luminosa proyectada en la superficie del sensor. Esta propiedad se utiliza en aplicaciones astronómicas que requieren imágenes con bajo nivel de ruido. La reducción del ruido se obtiene dejando al sensor "integrar" señales luminosas por minutos y hasta horas.

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 descretas. 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.

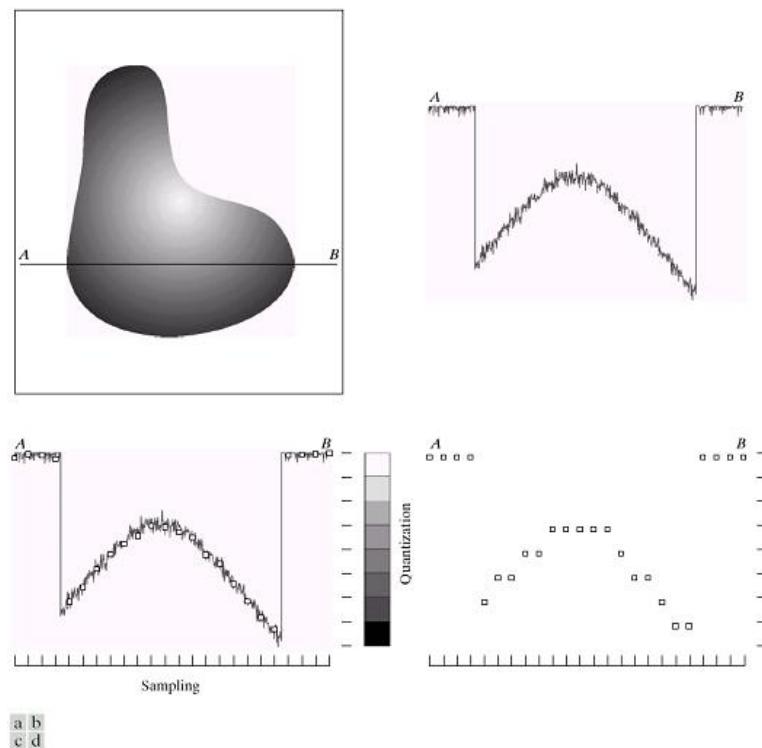


Fig 2.14 Muestreo y cuantificación: a) Imagen original b) Amplitud a lo largo de AB c) Muestreo d) Cuantificación

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 medida 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:

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

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):

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

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 $M = N$:

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

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 1024×1024 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, 512×512 .

Nótese que a partir de la tercera imagen, de 256×256 , 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 32×32 .

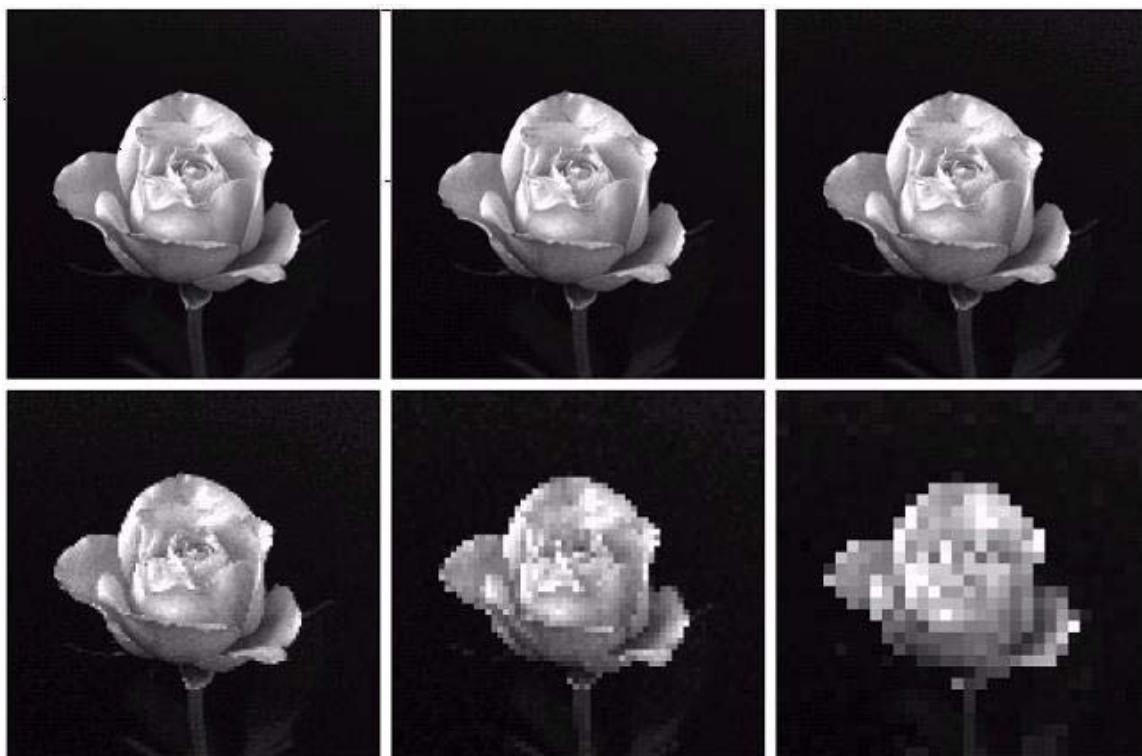


Fig 2.15 Imagen de 1024×1024 original y sus submuestreos (ampliados al tamaño de la primera) de 512×512 , 256×256 , 128×128 , 64×64 y 32×32

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*.

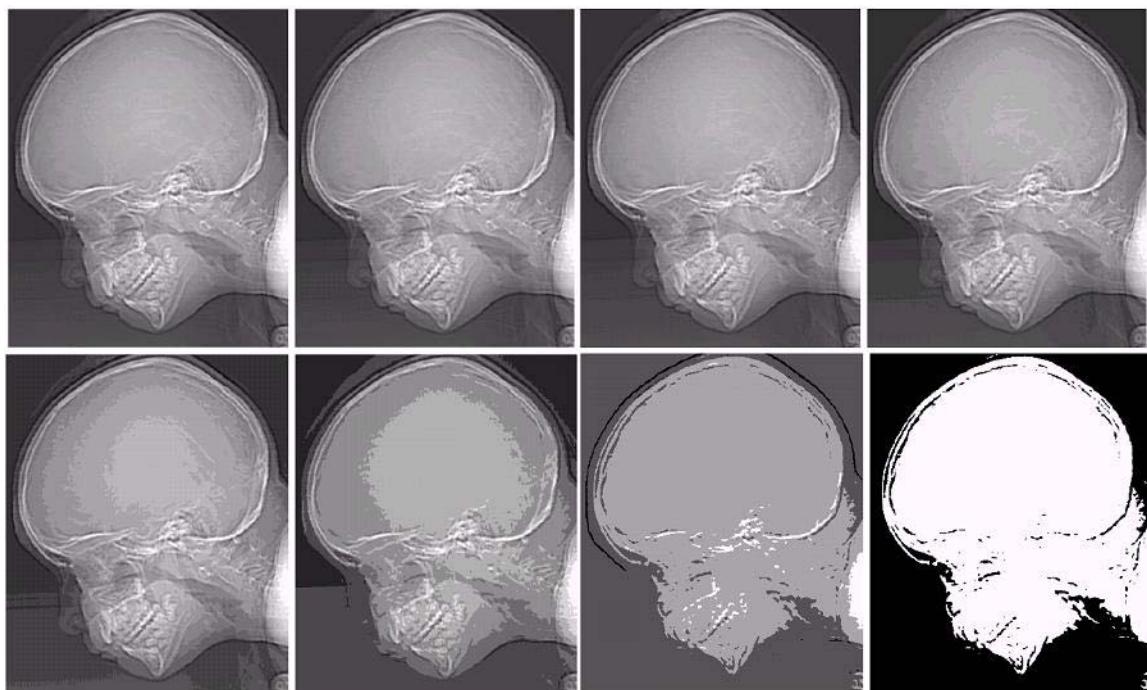


Fig 2.16 Imagen 472 x 374 en 256, 128, 64, 32, 16, 8, 4 y 2 niveles de gris

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 muestras³.

Si la función se submuestra, 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*).

³ 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 períodos de la función.

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*.

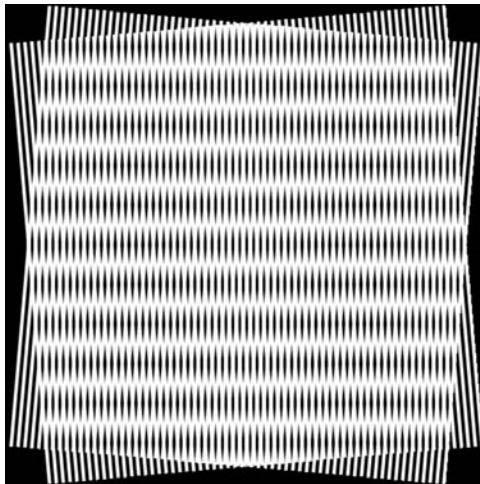


Fig 2.17 Ilustración del efecto de patrones de Moiré

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 sobreponida, buscamos el pixel más cercano en la imagen original y le asignamos su nivel de gris al pixel 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 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 pixeles

2.5.1 Vecinos de un pixel

Un pixel 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 pixeles, llamados los *4-vecinos* de p , se denota como $N_4(p)$. Cada pixel 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 pixeles es un concepto fundamental que simplifica la definición de numerosos conceptos de la imagen digital, como regiones y fronteras.

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

Digamos que en una imagen binaria (con valores 0 y 1) 2 pixeles 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:

- a) 4-conectividad. 2 pixeles p y q con valores en V son 4-adyacentes si q está en el conjunto $N_4(p)$.
- b) 8-conectividad. 2 pixeles p y q con valores en V son 8-adyacentes si q está en el conjunto $N_8(p)$.
- c) m-conectividad (conectividad mixta). 2 pixeles 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 $N_4(p) \cap N_4(q)$ no tiene pixeles 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.

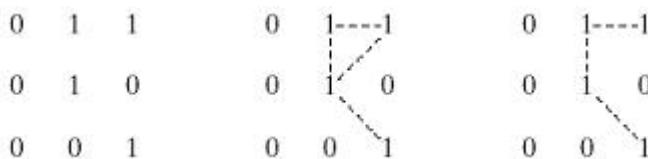


fig 2.19 La ambigüedad en 8-conectividad desaparece al utilizar 4-conectividad

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 conexa* de S. Si sólo tiene un componente conexo, entonces el conjunto S se denomina *conjunto conectado*.

Sea R un subconjunto de pixeles en una imagen. Llamamos R una *región 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) $D(p, q) \geq 0$ ($D(p, q) = 0$ si y sólo si $p = q$)
- (b) $D(p, q) = D(q, p)$, y
- (c) $D(p, z) \leq D(p, q) + D(q, z)$

La *distancia euclídea* 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 pixeles 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 pixeles 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 pixeles 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 pixeles con distancia D_8 menor o igual al valor r forman un cuadrado centrado en (x, y) . Por ejemplo, los pixeles 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 pixeles 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 pixeles no sólo depende de los valores de los pixeles en el camino, sino también de los valores de sus vecinos.

Por ejemplo, considérese el siguiente arreglo de pixeles 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

p ₃	p ₄
p ₁	p ₂
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 $p \ p_1 \ p_2 \ p_3 \ p_4$.

2.5.4 Operaciones entre píxeles

En las operaciones aritmético – lógicas, por ejemplo, al sumar 2 imágenes, lo que haremos será aplicar la operación, en este caso la suma, del pixel p en la imagen 1 con el pixel q en la misma posición en la imagen 2 y obtener una nueva imagen.

Las operaciones aritméticas serán adición, sustracción, multiplicación y división.

Las operaciones lógicas siguen un planteamiento idéntico, pero se aplicarán en su mayoría en imágenes binarias (AND, OR, NOT, XOR, etc).

Como veremos en el capítulo siguiente, las operaciones aritmético-lógicas también se podrán aplicar como *máscaras* entre imágenes.

2.6 Operaciones lineales y no lineales

Sea H un operador cuyas entradas y salidas son imágenes. Se dice que H es un *operador lineal* si, para cualesquiera 2 imágenes f y g y escalares a y b ,

$$H(af + bg) = aH(f) + bH(g) \quad (2.6 - 1)$$

En otras palabras, el resultado de aplicar un operador lineal a la suma de dos imágenes es el mismo obtenido al aplicarlo a cada una de las imágenes y luego sumar los resultados.

Los operadores que no se comportan de esta manera son llamados *no lineales*, y no son siempre predecibles.

3 Mejora de la imagen

El principal objetivo de la mejora es procesar una imagen para que el resultado sea más conveniente que la imagen original para una aplicación **específica**.

Un método conveniente para mejorar radiografías no necesariamente será el mejor para mejorar fotografías de marte transmitidas desde el espacio.

La mejora de la imagen se divide en 2 categorías: métodos del dominio espacial y métodos del dominio de la frecuencia. Los métodos del dominio espacial trabajan sobre el plano de la imagen, y en éste se manipulan directamente los píxeles de una imagen. En los métodos del dominio de la frecuencia se modifica la transformada de Fourier de una imagen. Existen técnicas que se basan en combinaciones de métodos de ambas categorías.

No hay una teoría general de mejora de la imagen. Cuando la imagen se procesa para interpretación visual, el observador es el que juzga qué tan bueno es un método: la evaluación visual de una imagen es un proceso *altamente subjetivo*. Cuando la imagen se procesa para ser percibida por una máquina, la evaluación es más fácil: el mejor procesamiento de la imagen es aquél que provoca un mejor reconocimiento por parte de la máquina.

3.1 Definiciones previas

El término *dominio espacial* se refiere al conjunto de píxeles que compone una imagen. Los procesos en el dominio espacial se denotarán por la expresión

$$g(x, y) = T[f(x, y)] \quad (3.1)$$

donde $f(x, y)$ es la imagen de entrada, $g(x, y)$ es la imagen procesada, y T es un operador de f , definido en alguna vecindad de (x, y) . T también puede operar en un conjunto de imágenes de entrada (p. Ej. La suma pixel por pixel de k imágenes para reducir el ruido).

Vecindad

Generalmente se define la vecindad respecto a un punto (x, y) utilizando un cuadrado o una subimagen o área rectangular⁴ centrada en (x, y) , como se muestra en la figura.

El centro de la subimagen se mueve de pixel a pixel comenzando, por ejemplo, en la esquina superior izquierda. El operador T se aplica a cada lugar (x, y) para producir g en ese lugar. El proceso utiliza solamente los píxeles en el área de la imagen que ocupa la vecindad.

⁴ Se pueden utilizar otras formas, como un círculo. Los cuadrados y rectángulos son las formas más usuales.

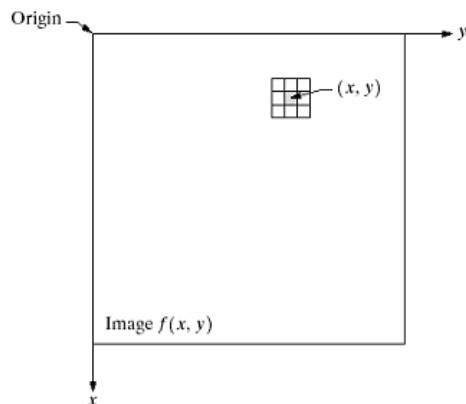


Fig 3.1 Una vecindad de 3×3 respecto a un punto (x, y) de la imagen.

Procesamiento puntual y Aumento del contraste

La forma más simple de T es cuando la vecindad es de tamaño 1×1 (un sólo pixel). En ese caso, g depende sólo del valor de f en (x, y) y T se convierte en una *función de transformación del nivel de gris* de la forma

$$s = T(r) \quad (3.2)$$

donde r y s denotan respectivamente el nivel de gris de $f(x, y)$ y $g(x, y)$ en cualquier punto (x, y) .

Por ejemplo, si $T(r)$ tiene la forma mostrada en la figura 3.2 (a), el efecto de la transformación sería producir una imagen de más alto contraste que la original al oscurecer los niveles bajo m y aclarar los niveles sobre m de la imagen original. Esta técnica se denomina *aumento del contraste*. En el caso límite (fig 3.2 (b)) $T(r)$ produce una imagen en 2 niveles de gris (binaria). Un mapeo de este tipo se denomina función *de umbral*.

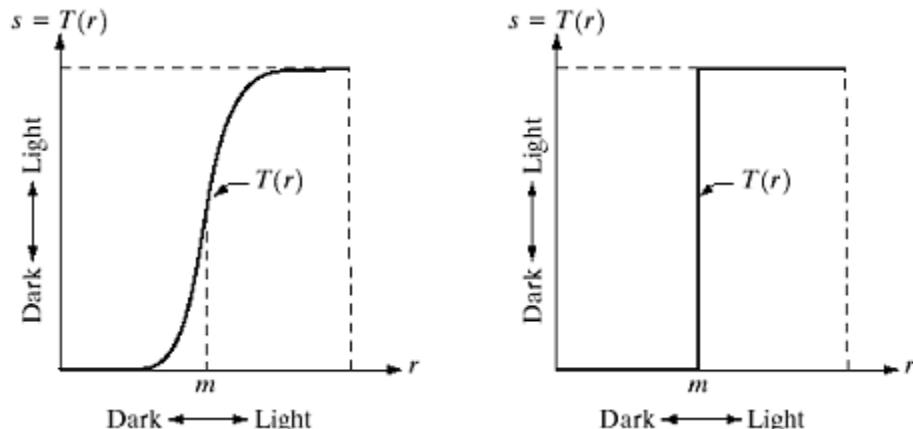


Fig 3.2 Funciones de transformación del nivel de gris para mejora del contraste

Muchas técnicas simples pero poderosas pueden formularse con transformaciones de nivel de gris. A estas técnicas se les llama de *procesamiento puntual*.

Máscaras

Las vecindades más grandes se prestan a mayor flexibilidad. Uno de los principales enfoques se basa en el uso de *máscaras* (también llamadas filtros, kernels o ventanas). Básicamente, una máscara es un arreglo pequeño (digamos, de 3 x 3) de dos dimensiones, como el de la figura 3.1, en el que los valores de los coeficientes de la máscara determinan la naturaleza del proceso, como el realzado de la imagen. Este enfoque se denomina *procesamiento por máscara*, o simplemente *filtrado*.

3.2 Algunas transformaciones del nivel de gris básicas

En este caso, los valores de los píxeles antes y después de procesar, se denotarán por r y s , respectivamente. Como ya se dijo, estos valores se relacionan por la expresión $s = T(r)$, donde T es una transformación que mapea un pixel de valor r a un pixel de valor s .

Considérese la figura 3.3, que muestra 3 tipos básicos de funciones utilizadas frecuentemente para la mejora de la imagen: lineal (negativo y transformación identidad), logarítmica (log y log inverso), y de potencia (n potencia y n raíz).

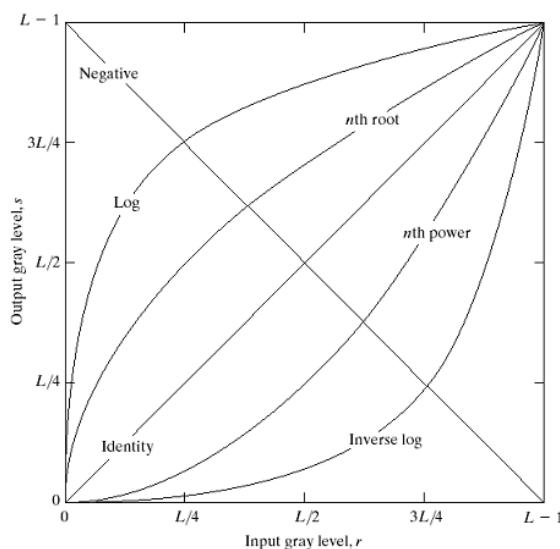


Fig 3.3 Algunas funciones de transformación del nivel de gris básicas utilizadas para mejora de la imagen

3.2.1 Negativos de la imagen

El negativo de una imagen con niveles de gris en el rango $[0, L - 1]$ se obtiene al utilizar la transformación de negativo mostrada en la figura 3.3, dada por la expresión

$$s = L - 1 - r \quad (3.3)$$

Invertir los niveles de intensidad de esta manera produce el equivalente de un negativo fotográfico. Este tipo de procesamiento es útil para mejorar niveles de blanco o gris en regiones oscuras de la imagen, especialmente cuando las áreas negras dominan en tamaño. Se muestra como ejemplo (fig 3.4) la imagen de un mamograma con una

pequeña lesión. Aunque el contenido visual es el mismo, nótese que es mucho más fácil analizar el tejido en la imagen negativa, para este caso en particular.

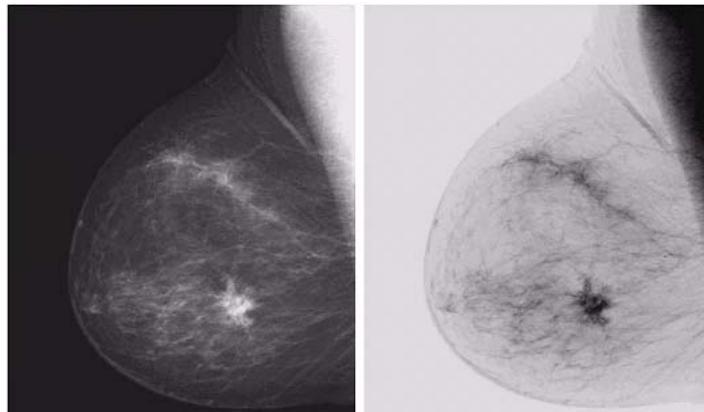


Fig 3.4 Un mamograma digital y su negativo obtenido por la transformación 3.3

3.2.2 Transformaciones logarítmicas

La forma general de la transformación logarítmica mostrada en la figura 3.3 es:

$$s = c \log (1 + r) \quad (3.4)$$

donde c es una constante, y se asume que $r \geq 0^5$. El efecto de la transformación es mapear un pequeño rango de valores bajos de nivel de gris a un rango más amplio de niveles de salida, al tiempo que lo contrario ocurre con los valores de entrada altos.

Se usa una transformación de este tipo para expandir los valores de píxeles oscuros de una imagen, mientras se comprime los valores de alto valor. La transformación logarítmica inversa lleva a cabo la transformación contraria.

La transformación logarítmica comprime el rango dinámico en imágenes con variaciones grandes en los valores de pixel. En la figura 3.5 se muestra un ejemplo donde la transformación logarítmica ayuda a visualizar niveles de gris antes no distinguibles.

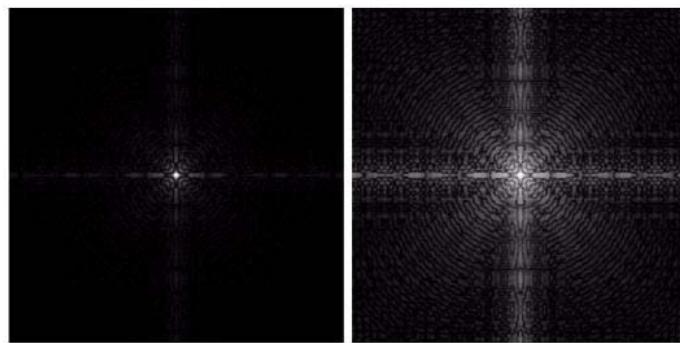


Fig 3.5 Espectro de Fourier y el resultado de una transformación logarítmica con $c = 1$

⁵ Para $c = 1$, el valor aproximado de s cuando $r = 255$ (el nivel de gris máximo en una escala de niveles de gris de 8 bits) es de 2.4. Si queremos ocupar el rango dinámico completo de 256 niveles de gris de salida, tenemos que convertir los valores de s de una escala de 0 a 2.4 a una escala de 0 a 255 (o utilizar $c = 106.25$).

3.2.3 Transformaciones de función de potencia

Las transformaciones de función de potencia tienen la forma básica

$$s = cr^\gamma \quad (3.5)$$

donde c y γ son constantes positivas. Algunas veces la ecuación se escribe como $s = c(r + \epsilon)^\gamma$ para tomar en cuenta un offset de calibración.

Al variar la γ obtenemos las diversas transformaciones que se grafican en la figura 3.6.

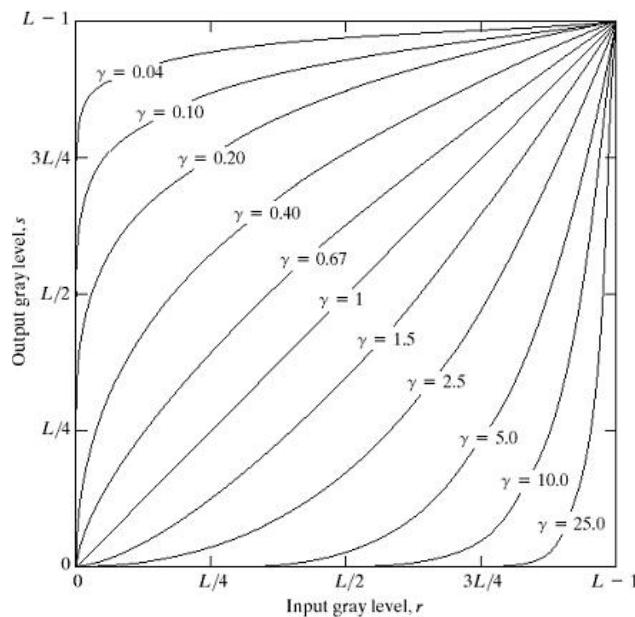


Fig 3.6 Gráficas de la ecuación $s = cr^\gamma$ para varios valores de γ ($c = 1$)

Como en el caso de la transformación logarítmica, las curvas de función de potencia con valores fraccionarios de γ mapean un rango pequeño de valores oscuros en un rango amplio de valores de salida y lo contrario con los valores claros de entrada. Sin embargo, en el caso de esta transformación, aquí obtenemos una familia de curvas de transformación simplemente variando γ ⁶.

Las curvas con valores de $\gamma > 1$ tienen un comportamiento contrario a las curvas con valores de $\gamma < 1$. Cuando $\gamma = 1$ la transformación es la transformación identidad (si $c = 1$).

Esta transformación también se utiliza para manipular el contraste de una imagen.

⁶ Muchos dispositivos utilizados para captura, despliegue e impresión de imágenes rigen su respuesta por esta ecuación. El proceso utilizado para corregir este fenómeno se conoce como corrección de gama. La corrección de gama es crítica en procesos que requieren un despliegue de imagen preciso.

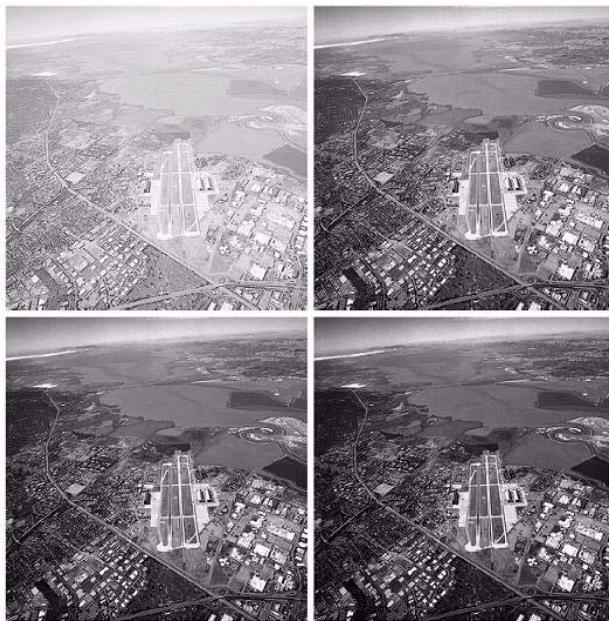


Fig 3.7 Imagen aérea y los resultados de aplicar la transformación de la ecuación 3.5 con $c = 1$ y $\gamma = 3.0$, 4.0 y 5.0 , respectivamente (NASA)

3.2.4 Funciones de transformación lineal por partes

La principal ventaja de las funciones de transformación lineal por partes es que pueden ser arbitrariamente complejas (pueden tener la forma que queramos). Su desventaja es que su especificación requiere más datos de entrada por parte del usuario.

3.2.4.1 Aumento del contraste

La idea del aumento del contraste es estirar el rango dinámico de los niveles de gris que se están procesando. En la figura 3.8 (a) vemos una transformación típica de aumento del contraste. Los puntos de control (r_1, s_1) y (r_2, s_2) determinan la forma de la función de transformación.

Si $r_1 = s_1$ y $r_2 = s_2$ la transformación es una función lineal que no produce cambios en el nivel de gris.

Si $r_1 = r_2$, $s_1 = 0$ y $s_2 = L - 1$, la transformación se vuelve una *función de umbral* cuyo resultado es una imagen binaria (3.8 (d)).

Valores intermedios producen distintos grados de distribución del nivel de gris, afectando el contraste.

La figura 3.8 (c) muestra el resultado de aumentar el contraste dados $(r_1, s_1) = (r_{\min}, 0)$ y $(r_2, s_2) = (r_{\max}, L - 1)$ donde r_{\min} y r_{\max} denotan los valores mínimo y máximo de niveles de gris en la imagen original. La función de transformación estira el rango original de niveles de gris al rango máximo $[0, L-1]$.

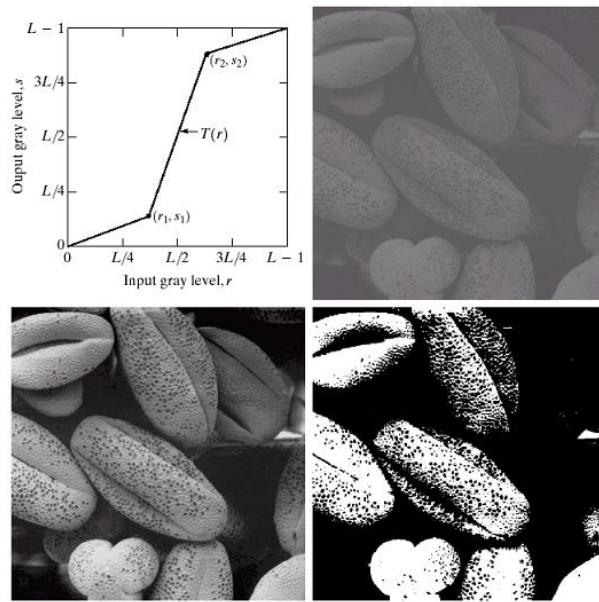


Fig 3.8 Aumento del contraste (a) función de transformación (b) imagen con bajo contraste (c) resultado de aumento del contraste (d) función de umbral (para umbral = media)

3.2.4.2 Fraccionamiento del nivel de gris

A veces se desea mostrar un rango específico de niveles de gris (p. Ej. Masas de agua en fotos satelitales). Una manera de lograr esto es con la transformación lineal por partes llamada fraccionamiento del nivel de gris. Se muestra un valor alto para todos los valores dentro del rango de interés y un valor bajo para todos los demás. Ejemplos de este tipo de transformación se muestran en la figura 3.9

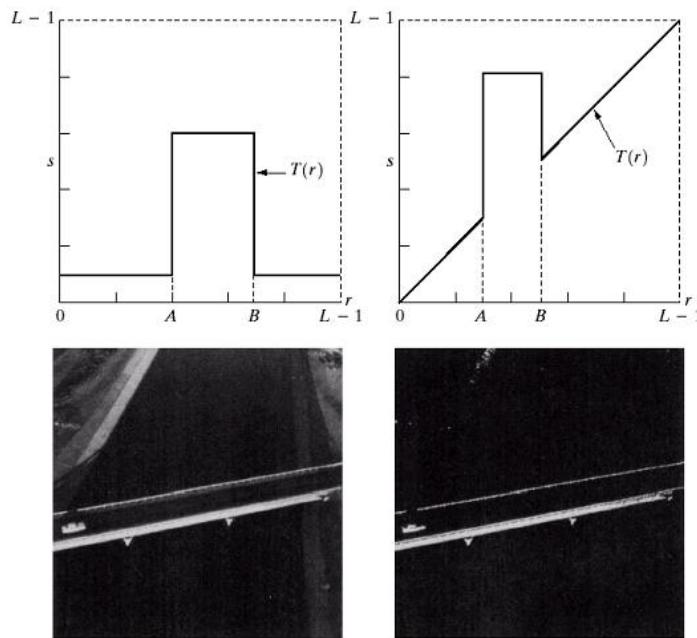


Fig 3.9 Fraccionamientos del nivel de gris para el rango de A a B, imagen original y el efecto de aplicar la segunda transformación en la imagen

3.2.4.3 Fraccionamiento de planos de bits

Para algunas aplicaciones, es deseable conocer la contribución de cada bit específico en la apariencia total de la imagen.

Supóngase que cada pixel de una imagen es representado por 8 bits. Imagínese que la imagen se compone de 8 planos de 1 bit, desde el plano de bits 0 (el bit menos significativo, LSB), hasta el plano de bits 7 (más significativo MSB). El plano de bits cero contiene todos los LSB de cada byte de cada pixel en la imagen y el plano 7 contiene los MSB de cada byte de cada pixel de la imagen.

Con esto obtenemos 7 imágenes binarias (una por cada plano). La figura 3.10 muestra una imagen y sus planos de bits.

Se puede observar que los 4 planos de bits superiores son los que contienen los datos más significativos visualmente. Separar una imagen en planos de bits es útil para analizar la importancia relativa de cada bit en una imagen. Este proceso ayuda a determinar la cantidad de bits necesaria para *cuantificar* una imagen. También es útil para la compresión de imágenes, como se verá después.

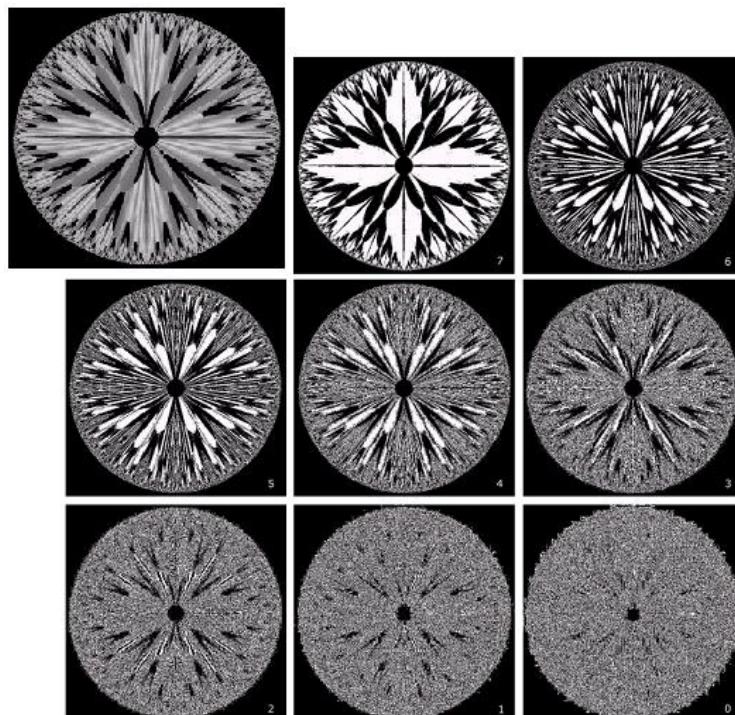


Fig 3.10 Una imagen de 256 niveles de gris y sus ocho planos binarios.

3.3 Procesamiento del histograma

El histograma de una imagen digital con niveles de gris en el rango $[0, L - 1]$ es una función discreta $h(r_k) = n_k$ donde r_k es el nivel de gris k-ésimo y n_k es el número de píxeles en la imagen con nivel de gris r_k .

Para *normalizar* el histograma se divide cada uno de sus valores entre el número total de píxeles en la imagen, n . Así, el histograma normalizado se obtiene de $p(r_k) = n_k / n$, para $k = 0, 1, \dots, L-1$. En breve, $p(r_k)$ nos da una estimación de la probabilidad de que

aparezca el nivel de gris r_k (la suma de los componentes de un histograma normalizado es igual a 1).

Los histogramas son la base de muchas técnicas de procesamiento de la imagen en el dominio espacial. En la figura 3.11 vemos 4 ejemplos de histogramas para 4 imágenes: oscura, clara, con bajo contraste y alto contraste. El eje horizontal de los histogramas es el valor de los niveles de gris, r_k y el eje vertical corresponde a los valores de $h(r_k) = n_k$ o $p(r_k) = n_k / n$ si los valores están normalizados.

La figura nos muestra la distribución del histograma de acuerdo a las características de las imágenes.

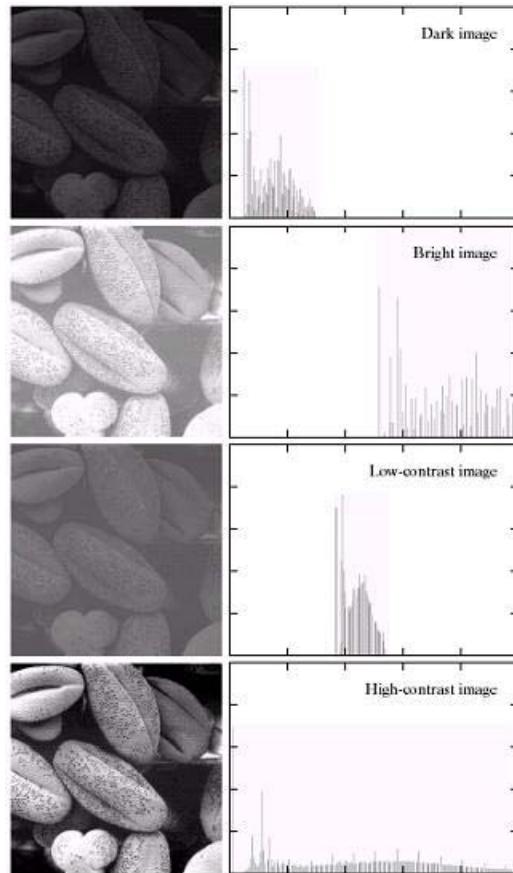


Fig 3.11 4 ejemplos de histograma para 4 tipos de imágenes.

3.3.1 Ecualización del histograma

Intuitivamente, es razonable concluir que una imagen cuyos pixeles tienden a ocupar el rango entero de posibles valores de gris y, además, tiende a estar uniformemente distribuido, tendrá una apariencia de alto contraste y exhibirá una gran variedad de tonos de gris.

Es posible obtener este histograma y al proceso se le llama *ecualización del histograma*. La función de transformación está dada por

$$s_k = T(r_k) = \sum_{j=0}^k p_r(r_j) = \sum_{j=0}^k n_j / n \quad (3.6)$$

Donde n es el número total de pixeles en la imagen, n_k es el número de pixeles que tienen nivel de gris r_k , y L es el número total de posibles niveles de gris en la imagen.

De esta transformación se obtiene una imagen procesada que mapea un pixel de nivel r_k de la imagen de entrada a un pixel de nivel s_k de la imagen de salida.

La transformación inversa se denota por

$$r_k = T^1(s_k) \quad k = 0, 1, 2, \dots, L-1 \quad (3.7)$$

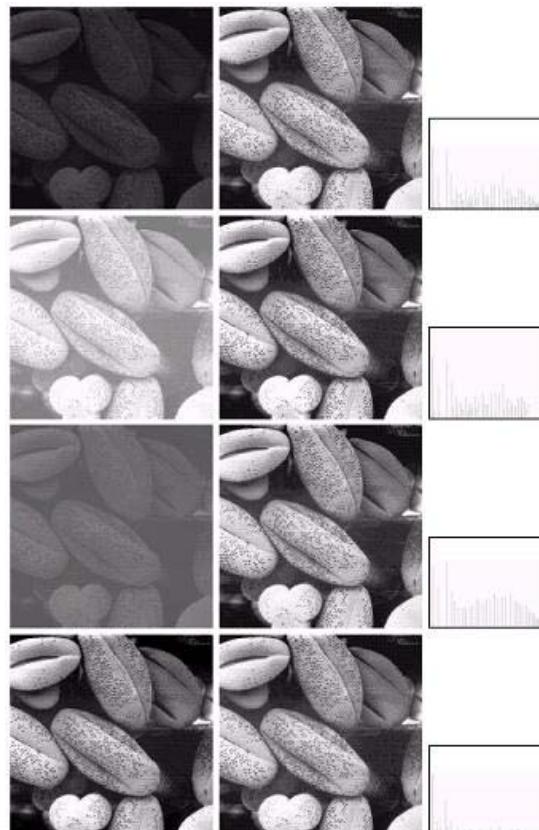


Fig 3.12 Distintas imágenes y el resultado de la transformación de ecualización de histograma

(El resto del capítulo fue impartido por los alumnos y no se elaboraron apuntes)

4 Mejora de la imagen en el dominio de la frecuencia

Comprender bien el área de la mejora de la imagen es imposible sin conocimiento del funcionamiento de la transformada de Fourier y los usos del dominio de la frecuencia en el procesamiento de imágenes.

Se intentará enfocarse en los fundamentos y su relevancia, enfatizando en la conexión entre las características de la imagen y las herramientas matemáticas utilizadas para representarlas.

Se verán brevemente los orígenes de la transformada de Fourier y su impacto, después una introducción a la transformada de Fourier y al dominio de la frecuencia, seguido de secciones que tratarán las mismas técnicas del capítulo anterior, pero en el dominio de la frecuencia.

4.1 Antecedentes

En el año de 1768 nació el matemático Jean Baptiste Joseph Fourier. Su contribución más recordada fue publicada en 1822 bajo el título *La théorie Analitique de la Chaleur* (teoría analítica del calor).

Series de Fourier

La idea más importante de este trabajo es que *toda función que se repite periódicamente puede ser expresada como la suma de senos y/o cosenos de diferentes frecuencias, cada uno multiplicado por un coeficiente diferente*. Ahora llamamos a esta suma *Serie de Fourier*.

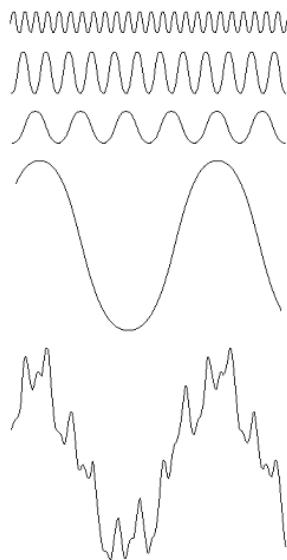


Fig 4.1 Descomposición de una onda en una suma de funciones senoidales y cosenoidales

Transformada de Fourier

Aún funciones que no son periódicas (pero con un área finita bajo la curva) pueden ser expresadas como la integral de senos y/o cosenos multiplicada por una función de

ponderación. Esta es la *transformada de Fourier*, y su utilidad es aún más grande que la de las series de Fourier en muchos problemas prácticos.

Las 2 representaciones comparten la importante característica de que una función, expresada en series de Fourier o la transformada, pueden ser reconstruidas (recobradas) completamente por un proceso inverso sin perder información.

La utilidad de las series y transformada de Fourier para resolver problemas prácticos las ha hecho ampliamente utilizadas y estudiadas como herramientas fundamentales.

El advenimiento de la computación digital y el “descubrimiento” del algoritmo la transformada rápida de Fourier (FFT) a finales de los cincuentas revolucionó el campo del procesamiento de señales.

Se utilizarán solamente funciones de duración finita (imágenes), así que nos concentraremos en la transformada de Fourier.

4.2 Introducción a la transformada de Fourier y el dominio de la frecuencia

En esta sección se introduce la transformada de Fourier en una y dos dimensiones. Nos enfocaremos en especial en la formulación discreta de la transformada continua y algunas de sus propiedades.

4.2.1 La transformada de Fourier en una dimensión y su inversa

La transformada de Fourier $F(u)$ de una función continua de una sola variable, $f(x)$, se define con la ecuación

$$F(u) = \int_{-\infty}^{\infty} f(x) e^{-j2\pi ux} dx \quad (4.2-1)$$

donde $j = \text{SQRT}(-1)$. De manera correspondiente, dada $F(u)$, podemos obtener $f(x)$ por medio de la transformada de Fourier inversa

$$f(x) = \int_{-\infty}^{\infty} F(u) e^{j2\pi ux} du \quad (4.2-1)$$

Estas 2 ecuaciones comprenden el *par de transformadas de Fourier*. Es decir que una función puede ser recuperada a partir de su transformada.

Como veremos más tarde, estas ecuaciones pueden fácilmente extenderse a 2 variables, u y v :

$$F(u, v) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y) e^{-j2\pi(ux + vy)} dx dy \quad (4.2-3)$$

y, similarmente para la transformada inversa

$$f(x, y) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} F(u, v) e^{j2\pi(ux + vy)} du dv \quad (4.2-4)$$

Nosotros estamos interesados en funciones discretas. La transformada de Fourier de una función discreta de una variable, $f(x)$, cuando $x = 0, 1, 2, \dots, M-1$, está dada por la ecuación

$$F(u) = \frac{1}{M} \sum_{x=0}^{M-1} f(x) e^{-j2\pi ux/M} \quad (u = 0, 1, \dots, M-1) \quad (4.2-5)$$

Esta *transformada de Fourier discreta* (o DFT, por sus siglas en inglés), es la base del resto de este capítulo. Similarmente, dada F(u), podemos obtener la función original usando la DFT inversa:

$$f(x) = \sum_{u=0}^{M-1} F(u) e^{j2\pi ux/M} \quad (x = 0, 1, \dots, M-1) \quad (4.2-6)$$

Para obtener F(u) en (4.2-5) comenzamos por sustituír u = 0 en el término exponencial y después sumamos para TODOS los valores de x. Después sustituímos u = 1 en el exponencial y repetimos la suma para todos los valores de x. Se repite este proceso para los M valores de u y de esta manera se obtiene la DFT.

Como f(x), la transformada es una cantidad discreta, y tiene el mismo número de componentes que f(x).

El mismo proceso aplica para calcular la DFT inversa.

Ejercicio 4.1

*Calcular la transformada discreta de Fourier de
f(x) = x para x = 0, 1
y la DFT inversa del resultado F(u)*

Una propiedad importante del par de transformadas discretas es que *la transformada discreta de Fourier y su inversa siempre existen*.

El dominio de la frecuencia

El concepto de dominio de la frecuencia, se puede derivar fácilmente de la fórmula de Euler:

$$e^{j\theta} = \cos \theta + j \sin \theta \quad (4.2 - 7)$$

Sustituyendo en (4.2-5) y recordando que $\cos -\theta = \cos \theta$ y $\sin -\theta = -\sin \theta$, obtenemos

$$F(u) = \frac{1}{M} \sum_{x=0}^{M-1} f(x) [\cos 2\pi ux / M - j \sin 2\pi ux / M] \quad (4.2-8)$$

para $u = 0, 1, 2, 3, \dots, M - 1$. Así, observamos que cada término de la transformada de Fourier (es decir, el valor de F(u) para cada valor de u) se compone de la suma de todos los valores de la función f(x).

A su vez, los valores de f(x) son multiplicados por senos y cosenos en varias frecuencias.

El dominio (valores de u) para el que el rango son los valores de F(u) es llamado, apropiadamente, *dominio de la frecuencia*, porque u determina la frecuencia de los componentes de la transformada.

Cada uno de los M términos de F(u) se llama *componente de frecuencia* de la transformada.

El uso de los términos dominio de la frecuencia y componentes de frecuencia es equivalente a dominio del tiempo y componentes del tiempo que usaríamos para expresar el dominio y valores de $f(x)$ si x fuera una variable de tiempo⁷.

En general, vemos en las ecuaciones (4.2-5) y (4.2-8) que los componentes de la transformada de Fourier son cantidades complejas. A veces será conveniente manejar $F(u)$ en coordenadas polares:

$$F(u) = |F(u)|e^{j\Phi(u)} \quad (4.2 - 9)$$

donde

$$|F(u)| = [R^2(u) + I^2(u)]^{1/2} \quad (4.2 - 10)$$

es llamado *magnitud o espectro* de la transformada de Fourier, y

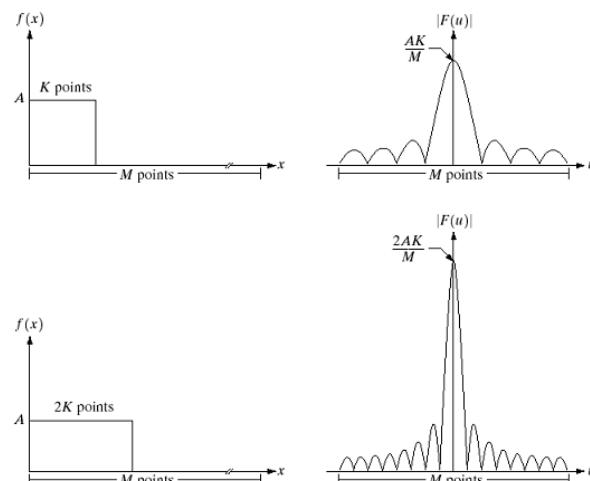
$$\Phi(u) = \tan^{-1} [I(u) / R(u)] \quad (4.2 - 11)$$

es llamado *ángulo de fase o espectro de fase* de la transformada. En las ecuaciones (4.2 – 10 y 4.2 – 11) $R(u)$ e $I(u)$ son las partes real e imaginaria de $F(u)$, respectivamente.

En términos de mejora de la imagen nos conciernen primariamente las propiedades del espectro.

La *densidad espectral* es el cuadrado del espectro de Fourier:

$$P(u) = |F(u)|^2 = R^2(u) + I^2(u) \quad (4.2-12)$$



4.2 Una función discreta de M puntos, con k puntos con valor A , y su transformada y una función con el doble de puntos con valor A y su transformada

⁷ Una analogía muy útil es comparar la transformada de Fourier a un prisma de vidrio. El prisma es el dispositivo físico que separa la luz en sus componentes de color, cada uno dependiendo de su contenido en longitud de onda (o frecuencia). La transformada de Fourier puede ser vista como un “prisma matemático” que separa una función en sus componentes, también basada en sus frecuencias.

Similarmente, la transformada nos permite caracterizar una función por su contenido en frecuencia.

Dada la relación inversa entre una función y su transformada ilustrada en la figura 4.2, no es sorprendente que Δx y Δu estén inversamente relacionadas por la expresión

$$\Delta u = 1 / M \Delta x$$

Esta medida es útil cuando las medidas son importantes en las imágenes procesadas. Por ejemplo, en una aplicación de microscopía electrónica las muestras de la imagen podrían estar espaciadas una micra entre ellas, y ciertas características del dominio de la frecuencia (como los componentes periódicos) pueden tener implicaciones con respecto a la estructura de la muestra física.

4.2.2 La DFT bidimensional y su inversa

La transformada discreta de Fourier se extiende fácilmente a 2 dimensiones:

$$F(u, v) = \frac{1}{MN} \sum_{x=0}^{M-1} \sum_{y=0}^{N-1} f(x, y) e^{-j2\pi(ux/M+vy/N)} \quad (u=0, 1, \dots, M-1; v=0, 1, \dots, N-1) \quad (4.2-16)$$

su inversa:

$$f(x, y) = \sum_{x=0}^{M-1} \sum_{y=0}^{N-1} F(u, v) e^{j2\pi(ux/M+vy/N)} \quad (x=0, 1, \dots, M-1; y=0, 1, \dots, N-1) \quad (4.2-17)$$

Donde u y v son las variables de transferencia o frecuencia y x e y son las variables espaciales o de imagen.

El espectro de Fourier, ángulo de fase y espectro de frecuencia se deducen también con facilidad:

$$|F(u, v)| = [R^2(u, v) + I^2(u, v)]^{1/2} \quad (4.2 - 18)$$

$$\Phi(u, v) = \tan^{-1} [I(u, v) / R(u, v)] \quad (4.2 - 19)$$

$$P(u) = |F(u)|^2 = R^2(u) + I^2(u) \quad (4.2- 20)$$

Usualmente se multiplica la imagen por $(-1)^{x+y}$. Esto resulta en:

$$\mathcal{F}[f(x, y) (-1)^{x+y}] = F(u - M/2, v - N/2) \quad (4.2 - 21)$$

Esta ecuación nos dice que el origen de la transformada de Fourier de $f(x, y) (-1)^{x+y}$ se localiza en $u = M/2$, $v = N/2$, lo que pone el origen al centro del área $M \times N$ que ocupa la DFT bidimensional. Esta área es llamada *rectángulo de frecuencia*.

Se requiere que M y N sean impares para que las coordenadas del centro sean enteros.

El valor de la transformada en $(u, v) = (0, 0)$ es:

$$F(0, 0) = \frac{1}{MN} \sum_{x=0}^{M-1} \sum_{y=0}^{N-1} f(x, y) \quad (4.2-22)$$

el cuál es claramente el promedio de $f(x, y)$, lo que quiere decir que el valor de la DFT en el origen es igual al nivel de gris promedio de la imagen.

Como las frecuencias son cero en el origen, $F(0, 0)$ a veces es llamado el componente dc (de corriente directa, cero frecuencia).

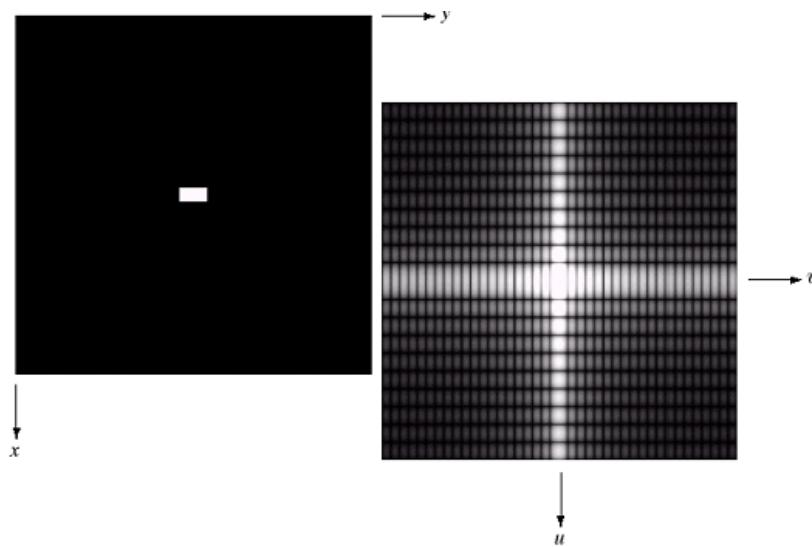
Además, el espectro de la transformada de Fourier es simétrico

$$|F(u, v)| = |F(-u, -v)| \quad (4.2 - 23)$$

y también se cumplen las siguientes relaciones:

$$\Delta u = 1 / M\Delta x$$

$$\Delta v = 1 / N\Delta y$$



4.3 Un rectángulo blanco sobre fondo negro y el espectro de Fourier de la imagen premultiplicada por $(-1)^{x+y}$ y posprocesada con el logaritmo para ver detalles de los niveles oscuros de gris.

4.2.3 Filtrado en el dominio de la frecuencia

Como ya se ha dicho, el dominio de la frecuencia no es más que el espacio definido por los valores de la transformada de Fourier y sus variables de frecuencia (u, v).

Propiedades básicas del dominio de la frecuencia

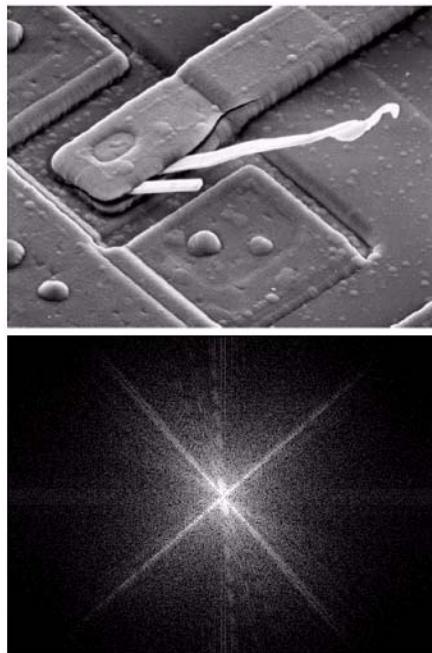
Recuérdese que cada uno de los término de $F(u, v)$ contiene todo los valores de $f(x, y)$, modificados por los valores de los términos exponentiales.

Debido a esto, es usualmente imposible hacer asociaciones directas entre componentes específicas de una imagen y su transformada.

Sin embargo se pueden hacer algunas observaciones generales acerca de la relación entre los componentes de frecuencia de la FT y características especiales de la imagen:

Ya que la frecuencia se relaciona directamente con la velocidad de cambio, no es difícil asociar intuitivamente frecuencias de la transformada de Fourier con patrones de variación de intensidad de una imagen.

- El componente de frecuencia que varía más lentamente ($u = v = 0$) corresponde al nivel de gris promedio.
- Al alejarnos del origen, las frecuencias bajas corresponden a componentes que varían lentamente.
- Lejos del origen, las frecuencias altas corresponden a cambios cada vez más rápidos en el nivel de gris e la imagen (p. Ej. Bordes, o ruido).



4.4 Imagen de un circuito integrado con daño termal inducido y su espectro de Fourier

En la imagen podemos ver un circuito integrado dañado (la parte blanca es óxido resultado del daño termal inducido). Varios detalles de la imagen pueden relacionarse con lo que vemos en su espectro de Fourier.

Los bordes prominentes corriendo a 45° y las 2 prominencias blancas de óxido corresponden a los componentes del espectro con direcciones de 45° y, a la izquierda del eje vertical un componente que sale un poco a la izquierda.

Ideas básicas del filtrado en el dominio de la frecuencia

El filtrado en el dominio de la frecuencia consiste en los siguientes pasos:

1. Multiplicar la imagen de entrada por $(-1)^{x+y}$ para centrar la transformada.
2. Calcular $F(u, v)$, la DFT de la imagen en el paso 1.
3. Multiplicar $F(u, v)$ por una función de filtro $H(u, v)$
4. Calcular la transformada inversa del resultado del paso 3.
5. Obtener la parte real del resultado en 4.
6. Multiplicar el resultado en 5 por $(-1)^{x+y}$

$H(u, v)$ es llamado filtro porque suprime ciertas frecuencias de la transformada, pero deja otras sin cambio.

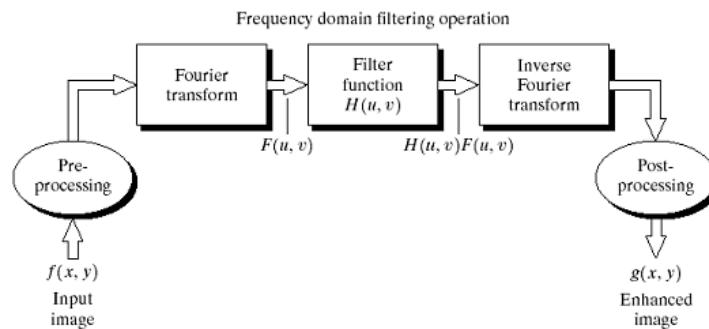
La transformada de Fourier de la imagen de salida es entonces dada por la siguiente ecuación:

$$G(u,v) = H(u,v)F(u,v)$$

La multiplicación de H y F se hace entre funciones bidimensionales y está definida elemento por elemento. Esto es, el primer elemento de H se multiplica por el primer elemento de F, y así.⁸

La imagen filtrada se obtiene tomando la parte real de este resultado y multiplicando por $(-1)^{x+y}$. La transformada inversa es en general, compleja, sin embargo, si la imagen de entrada y la función de filtro son reales, los componentes imaginarios de la transformada inversa deberían ser cero⁹.

Además de esta multiplicación se suele practicar algo más de procesamiento, como cortar la imagen para que tenga dimensiones pares, escalación de nivel de gris, conversión a punto flotante, etc.



4.5 Pasos básicos del filtrado

Algunos filtros básicos y sus propiedades

Veremos algunos filtros específicos y cómo éstos afectan a las imágenes.

Supóngase que se requiere forzar el promedio del valor de la imagen a cero. De acuerdo con lo visto, el valor promedio está dado por $F(0,0)$. Si ponemos este término en cero en el dominio de la frecuencia y tomamos la transformada inversa, el valor promedio de la imagen resultante será cero. Asumiendo que la transformada se ha centrado, podemos llevar a cabo esta operación multiplicando todos los valores de $F(u,v)$ por la función de filtro:

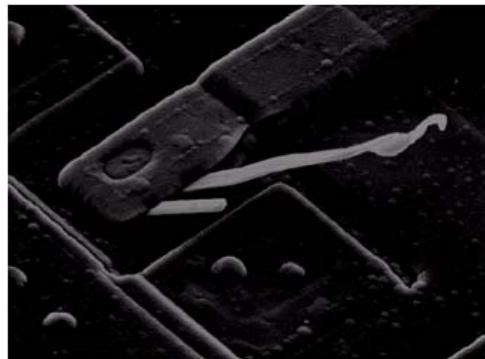
$$0 \quad \text{si } (u, v) = (M/2, N/2)$$

⁸ En general, los componentes de F son cantidades complejas pero los filtros con los que trataremos en el libro serán, típicamente, reales. En este caso, cada componente de H multiplica los valores real e imaginario de la componente correspondiente en F. Estos filtros son llamados *filtros de cambio de fase cero* porque no alteran la fase de la transformada.

⁹ En la práctica, la DFT inversa generalmente tiene componentes imaginarios parásitos por redondeo computacional. Ignoraremos estos componentes.

$$H(u, v) = \begin{cases} 1 & \text{en los demás casos.} \end{cases} \quad (4.2-29)$$

Lo que hará que $F(0,0)$ sea cero y los demás componentes queden igual. Este filtro es llamado filtro de muesca (*notch filter*) porque es una función constante con un agujero en el origen.



4.6 Imagen filtrada con el filtro de muesca de la imagen 4.4

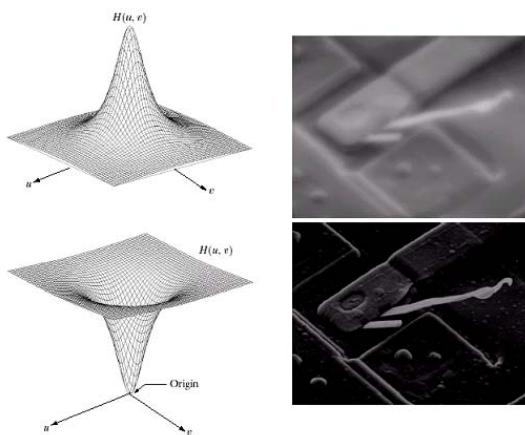
En el ejemplo podemos ver el efecto de este filtro: en general disminuye el nivel de gris y los bordes prominentes resaltan más.

Las bajas frecuencias en la transformada de Fourier son responsables de la apariencia general de nivel de gris de una imagen en las áreas suaves, mientras que las altas frecuencias son responsables de los detalles, como los bordes y el ruido.
Más adelante se verán estos ejemplos con más detalle.

Un filtro que atenúa las altas frecuencias mientras deja pasar las bajas se llama filtro pasabajas.

Un filtro con las características opuestas es llamado pasaaltas.

Un filtro pasabajas tendrá detalles menos finos que el original porque se han atenuado las altas frecuencias. Una imagen filtrada con pasabajas tendrá menos variaciones suaves de nivel de gris porque se ha enfatizado el detalle transicional de nivel de gris.



4.7 Filtros pasabajas y pasaaltas y las imágenes obtenidas con ellos a partir de 4.4

Los filtros mostrados $H(u,v)$ son simétricamente circulares.

En la primer imagen podemos observar su equivalencia con el filtro de suavizado anteriormente visto.

La segunda acentúa detalles finos, con $F(0,0)$ en cero y sus consecuencias. Usualmente se suma una constante para no eliminar $F(0,0)$ completamente.

4.2.4 Correspondencia entre el filtrado en los dominios espacial y de frecuencia

La relación fundamental entre el dominio espacial y el de frecuencia está establecido por un resultado llamado el *teorema de convolución*.

El proceso por el cuál movemos una máscara de pixel a pixel en una imagen y calculamos una cantidad predefinida para cada pixel, es el fundamento del proceso de convolución.

Formalmente la convolución discreta de 2 funciones $f(x, y)$ y $h(x,y)$ de tamaño $M \times N$ se denota por $f(x, y) * h(x, y)$ y se define por la expresión

$$f(x,y) * h(x,y) = \frac{1}{MN} \sum_{x=0}^{M-1} \sum_{y=0}^{N-1} f(x, y) h(x - m, y - n) \quad (4.2-16)$$

Los signos menos significan que la función h está reflejada con respecto al origen.

Esta ecuación es básicamente una implementación de

1 – reflejar una función respecto al origen

2 – mover la función respecto a la otra al cambiar los valores de (x, y)

3 – calcular la suma de los productos de todos los valores de m y n , para cada desplazamiento x, y (los desplazamientos x, y son incrementos enteros que se detienen cuando las funciones ya no están traslapadas).

(El resto del capítulo fue impartido por los alumnos y no se elaboraron apuntes)

5 Procesamiento del color

El uso del color en el procesamiento de imágenes está motivado por 2 factores:

El color es un poderoso descriptor que simplifica la identificación de objetos y su extracción de una escena. En segundo lugar, los seres humanos podemos discernir entre miles de tonalidades de color, comparado a cerca de 2 docenas de niveles de gris.

El procesamiento del color se divide en 2 áreas: *Procesamiento del color* y *procesamiento del seudocolor*. En el primer caso las imágenes se adquieren con un sensor de color como una cámara de televisión o un scanner. En el segundo caso se trata de asignar un color a una intensidad o un rango de intensidades monócromas.

Descubriremos que algunas de las técnicas de escala de gris vistas hasta ahora son directamente aplicables a imágenes de color.

5.1 Fundamentos del color

Básicamente, los colores que los seres humanos y otros animales perciben en un objeto están determinados por la naturaleza de la luz *reflejada* desde el objeto. La luz visible se compone de una delgada banda de frecuencias del espectro electromagnético. Un cuerpo que refleja luz balanceada en todas las longitudes de onda visibles se verá blanco. Sin embargo, un cuerpo que favorezca la reflectancia en un rango limitado del espectro visible mostrará algunos tonos de color.

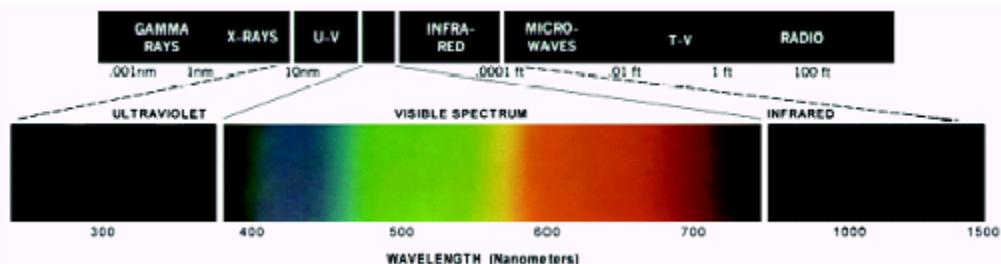


Fig 6.1 Longitudes de onda del espectro visible

La luz cromática ocupa el espectro electromagnético desde aproximadamente los 400 a los 700 nm.

Se utilizan 3 cantidades básicas para describir la calidad de una fuente de luz cromática:

- *Radiancia*. Es el total de energía proveniente de la fuente de luz, se mide en Watts.
- *Luminancia*. Es una medida de la cantidad de energía percibida por un observador. Se mide en lúmenes.
- *Brillo*. Engloba la noción acromática de intensidad. Es una medida altamente subjetiva.

Por ejemplo, luz emitida de una fuente que opera en el infrarrojo lejano tendría una radiancia bastante significativa (energía), pero un observador apenas la percibiría (su luminancia sería casi cero).

Los conos, los colores primarios y los colores secundarios

Como hemos visto, los conos son los sensores del ojo responsables de la visión a color. Experimentalmente se ha determinado que los 6 a 7 millones de conos del ojo humano pueden subdividirse en 3 categorías principales de sensibilidad, que corresponden vagamente al rojo, verde y azul. Aproximadamente 65% son sensibles a la luz roja, 33% a la verde y cerca de 2% a la azul (aunque los conos de luz azul son los más sensibles).

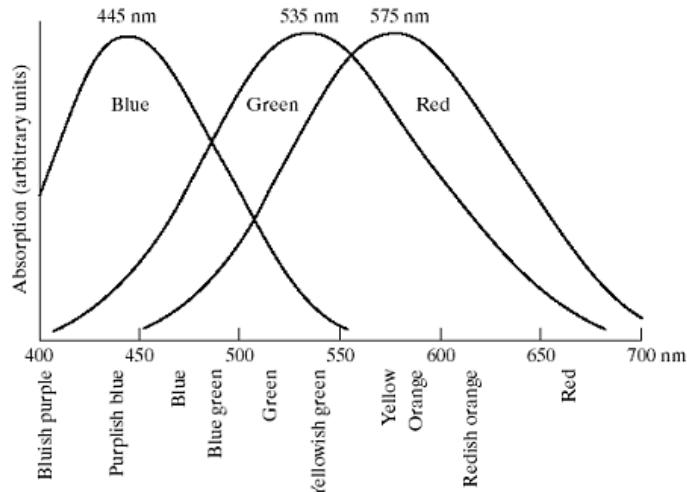


Fig 5.2 Absorción de luz de los conos rojo, verde y azul en función de la longitud de onda

Debido a estas características de absorción, los colores se ven como combinaciones variables de los llamados *colores primarios* rojo (R), verde (G), y azul (B). En los años 30 se definieron estándares de longitud de onda para los 3 colores.

Podemos notar en las figuras 5.1 y 5.2 que no podemos llamar rojo, verde o azul a un solo color. También debemos recordar que tener 3 longitudes de onda que designan a los colores primarios por propósitos de estandarización no significa que estos 3 colores fijos pueden generar todos los colores del espectro.

Los colores primario pueden ser sumados para producir los *colores secundarios* de la luz: magenta (rojo mas azul), cian (azul mas verde) y amarillo (rojo mas verde). Mezclando los 3 primarios, o un secundario con su primario opuesto en las intensidades correctas, podemos obtener luz blanca.

Es importante diferenciar entre los colores primarios de la luz y los de los pigmentos o colorantes. En estos últimos, un color primario se define como aquel que absorbe o sustrae un color primario de luz y refleja y transmite los otros dos. Por lo tanto los colores primarios de los pigmentos son magenta, cian y amarillo, y los secundarios rojo, verde y azul.

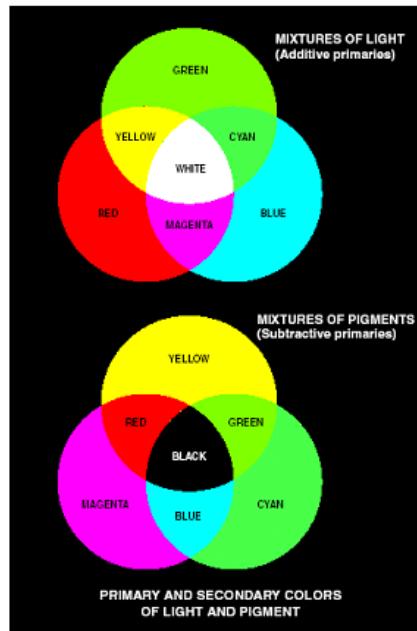


Fig 5.3 Los colores primarios de la luz y del pigmento

Tono, brillo y Saturación

Las características que generalmente utilizamos para distinguir un color de otro son el brillo, el tono y la saturación.

Brillo. Noción cromática de intensidad.

Tono. Es la longitud de onda dominante en una mezcla de ondas de luz. El color dominante que percibe un observador.

Saturación. La pureza relativa y la cantidad de luz blanca mezclada con el tono. Los colores puros del espectro están totalmente saturados.

Se llama *cromaticidad* al tono y saturación juntos. Entonces podemos caracterizar un color por su cromaticidad y su brillo.

Valores de tristímulo

Las cantidades de rojo, verde y azul necesarias para formar un color en particular se llaman valores de tristímulo y se denotan X, Y y Z, respectivamente. Se especifica entonces un color por sus *coeficientes tricromáticos*, definidos como:

$$x = X / (X + Y + Z) \quad (5.1)$$

$$y = Y / (X + Y + Z) \quad (5.2)$$

$$z = Z / (X + Y + Z) \quad (5.3)$$

Nótese que de estas ecuaciones:

$$x + y + z = 1 \quad (5.4)$$

El diagrama cromático de CIE

Otro enfoque para especificar colores es el uso del diagrama de cromaticidad del CIE, que muestra la composición de colores como una función de x (rojo) y y (verde). Para cada valor de x y y , el valor correspondiente de z (azul) se obtiene de la ec. 5.4 despejando

$$z = 1 - (x + y) \quad (5.5)$$

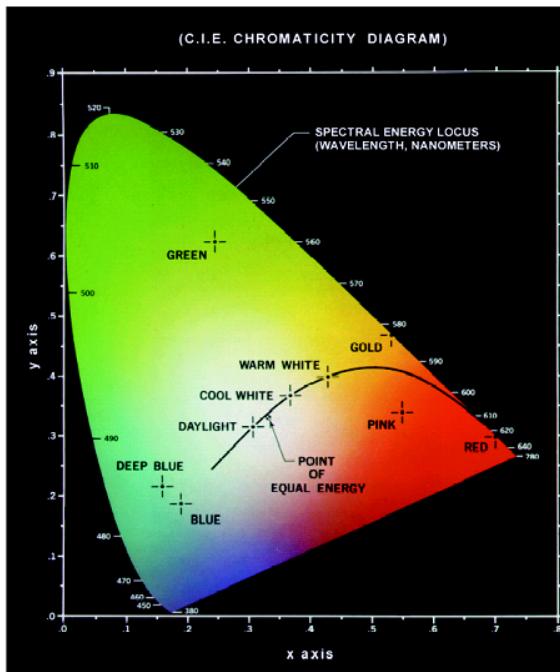


Fig 5.4 Diagrama de cromaticidad del CIE (Centro Internacional de Iluminación)

El punto de igual energía corresponde a fracciones iguales de los 3 colores primarios; representa el estándar de CIE para la luz blanca.

Cualquier punto en el borde de la gráfica cromática está saturado totalmente.

La saturación en el punto de igual energía es cero.

El diagrama de cromaticidad es útil para la mezcla de colores porque un segmento de línea recta utilizado para unir 2 puntos en el diagrama define todas las variaciones de color que pueden ser obtenidas combinando estos colores de manera aditiva.

Una línea desde el punto de igual energía a cualquier punto en el borde de la gráfica definirá todas las tonalidades de ese color del espectro en particular.

Para determinar el rango de colores que pueden ser obtenidos a partir de cualesquiera 3 colores dados del diagrama de cromaticidad, simplemente conectamos los 3 puntos. El triángulo resultante envuelve cualquier color que puede ser producido por combinaciones varias de los colores en los vértices.

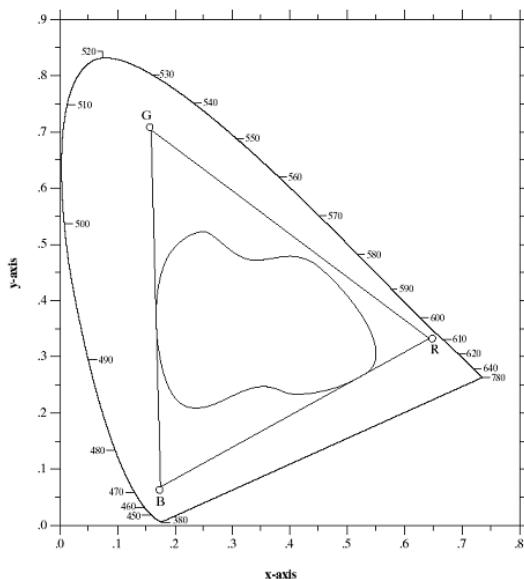


Fig 5.5 Gamut de monitores (triángulo) e impresoras (región irregular)

El triángulo de la figura muestra un rango típico de colores (llamado *gamut de color*) producido por monitores RGB. La región irregular dentro del triángulo representa el gamut de color de los dispositivos de impresión de alta calidad. Es irregular porque la impresión a color es una combinación de mezcla aditiva y sustractiva.

5.2 Modelos de color

El propósito de un modelo de color (también llamado espacio de color o sistema de color) es facilitar la especificación de colores de un modo estándar o aceptado en general. Un modelo de color es la especificación de un sistema de coordenadas y el subespacio dentro de él donde cada color puede ser representado por un solo punto.

La mayoría de los modelos de color en uso se orientan ya sea al hardware, o a aplicaciones donde la manipulación de color es el objetivo.

Los modelos orientados a hardware más comunes son el RGB (rojo, verde, azul), para monitores a color y cámaras de video; el CMY (cian, magenta, amarillo) y CMYK para impresión a color; y el HSI (tono, saturación, intensidad), que se acerca a la manera como el ojo humano describe e interpreta el color. El modelo HSI también tiene la ventaja de que separa la información del color y de la escala de gris de una imagen, haciéndolo conveniente para muchas técnicas de escala de gris desarrolladas en este curso.

5.2.1 El modelo RGB

En el modelo RGB, cada color aparece en sus componentes primarios del espectro de rojo, verde, y azul. Este modelo se basa en un sistema de coordenadas cartesianas que forma un cubo en el que los valores RGB están en 3 esquinas. Los valores de cian, magenta y amarillo están en las otras 3 esquinas, el negro está en el origen y el blanco está en el vértice más alejado del origen.

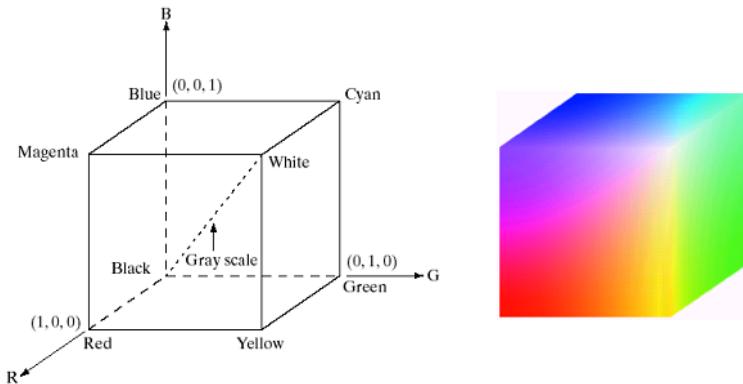


Fig 5.6 Diagrama esquemático del cubo RGB y el cubo a 24 bits

La escala de gris (puntos con valores de RGB iguales) se extiende del negro al blanco en la línea que une los puntos negro y blanco. Los diferentes colores del modelo son puntos sobre o dentro del cubo, y se definen por medio de vectores que se extienden desde el origen.

Las imágenes representadas en el modelo RGB consisten en 3 imágenes componentes, cada una de un color primario.

El número de bits utilizado para representar cada pixel en el espacio RGB se llama *profundidad de pixel*.

Considérese una imagen en la que cada una de las imágenes roja, azul y verde es una imagen de 8 bits. En estas condiciones se dice que cada pixel de color RGB (una tupla (R,G,B)) tiene una profundidad de 24 bits.

El número total de colores en una imagen RGB de 24 bits es de $(2^8)^3 = 16,777,216$.

Una imagen a color puede ser adquirida utilizando 3 filtros, sensitivos al rojo, verde y azul, respectivamente. Cuando observamos una escena de color con una cámara monócula equipada con uno de estos filtros, el resultado es una imagen monócula cuya intensidad es proporcional a la respuesta del filtro.

Si se repite el proceso con cada filtro se producen 3 imágenes monóculas que son los componentes RGB de la escena en color (en la práctica los sensores RGB integran el proceso en un solo dispositivo).

5.2.2 Los modelos CMY y CMYK

Como se indica en la sección 6.1, el cian, magenta y amarillo son los colores secundarios de la luz o, alternativamente, los colores primarios de los pigmentos. Por ejemplo, cuando una superficie cubierta por pigmento cian se ilumina con luz blanca, no se refleja luz roja de la superficie; es decir, el cian sustrae luz roja de la luz blanca reflejada.

La mayoría de los dispositivos que depositan pigmentos de color en el papel, como las impresoras y copiadoras, requieren entrada de datos en CMY o realizan una conversión interna de RGB a CMY. Esta conversión se lleva a cabo con la operación:

$$\begin{vmatrix} C \\ M \\ Y \end{vmatrix} = \begin{vmatrix} 1 \\ 1 \\ 1 \end{vmatrix} - \begin{vmatrix} R \\ G \\ B \end{vmatrix} \quad (5.6)$$

Esta ecuación muestra que la luz reflejada de una superficie cubierta de cian puro no contiene rojo ($C = 1 - R$ en la ecuación).

De igual manera, el magenta puro no refleja verde, el amarillo puro no refleja azul.

Los valores de RGB pueden ser obtenidos fácilmente del conjunto CMY sustrayendo los valores individuales CMY de ¹⁰

5.2.3 El modelo HSI

Como hemos visto, crear los colores en los modelos RGB y CMY y cambiar de un modelo a otro es fácil, además de que estos sistemas son muy convenientes para implementaciones de hardware.

El sistema RGB también es compatible con el hecho de que el ojo humano es altamente perceptivo a los primarios rojo verde y azul.

Desgraciadamente este tipo de modelos de color no son convenientes para describir colores en los términos prácticos para la interpretación humana. Cuando un ser humano ve un objeto de color, lo describe por su tono, saturación y brillo. El modelo que se presenta en esta sección, llamado *modelo de color HSI* (hue, saturation, intensity: tono, saturación e intensidad) separa el componente de intensidad de la información de color en una imagen de color. Como resultado, este modelo es una herramienta ideal para desarrollar algoritmos basados en descripciones de color naturales e intuitivas para los seres humanos. El modelo RGB, después de todo, es ideal para generación de imágenes a color, pero su uso para la descripción de color es limitado.

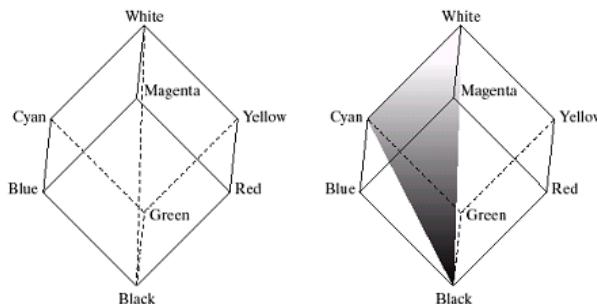


Fig 5.7 Deducción del modelo HSI a partir del cubo RGB

Como hemos visto, una imagen RGB puede ser conceptualizada como 3 imágenes de intensidad monocromas (representando el rojo, verde y azul). Así pues, se puede obtener la intensidad de una imagen RGB. Esto está claro si tomamos el cubo de color de la figura 5.6 y lo apoyamos sobre el vértice negro, con el vértice blanco directamente encima de él, como vemos en la figura.

La intensidad estará dada por la línea que une a estos 2 vértices. Así, si queremos determinar el componente de intensidad de cualquier punto de color en la figura,

¹⁰ De acuerdo a la figura 5.3, cantidades iguales de los primarios de pigmento (cian, magenta y amarillo) deberían producir negro. En la práctica, al combinar estos colores para imprimir se produce un negro de color lodoso. Para producir el verdadero negro (el color predominante en la impresión), se añade al modelo un cuarto color, *negro*, lo que nos da el modelo de color CMYK. Este modelo se llama impresión a 4 colores.

simplemente debemos pasar un plano perpendicular al eje de intensidad que contenga el punto de color. La intersección del plano con el eje nos dará un punto con el valor de intensidad en el rango [0, 1].

La saturación (pureza) de un color aumenta en función de su distancia al eje de intensidad. La saturación en los puntos del eje de intensidad es cero (los puntos del eje son grises).

Para determinar el tono, considérese el plano definido por los vértices cian, negro y blanco. Todos los puntos contenidos en el plano tendrán el mismo tono (recuérdese que todos los colores generados por 3 colores están en un triángulo definido por esos 3 colores).

Rotando el plano en el eje de intensidad, obtendremos diferentes tonos.

La conclusión es que los valores de tono, saturación e intensidad para formar el espacio HSI se pueden obtener a partir del cubo de color RGB.

La clave para entender el espacio HSI es el eje de intensidad vertical y el conjunto de colores en los planos perpendiculares a este eje. Al moverse, estos planos tendrán forma triangular o hexagonal. Esto puede visualizarse mucho mejor rotando nuestro cubo de manera que el eje de intensidad quede en la misma dirección en que mira un observador, como en la figura.

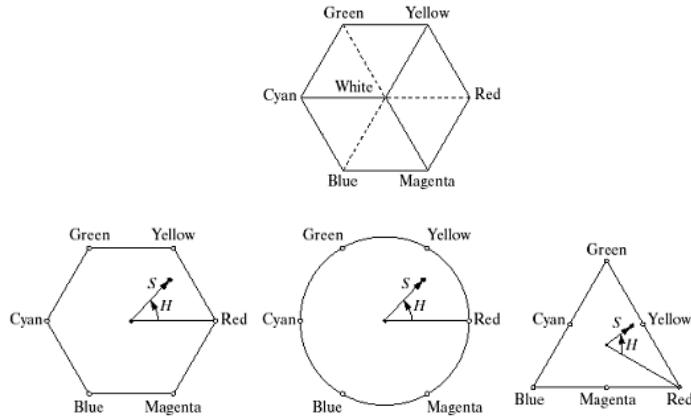


Fig 5.8 Componentes de tono y saturación del HSI

Ahora vemos un plano en donde los colores primarios están separados por 120 grados, y los secundarios a 60° de los primarios. Se muestra también un punto de color arbitrario. Su tono está determinado por un ángulo desde la línea perpendicular al eje de intensidad al vértice rojo. La saturación es la longitud del vector al punto de color.

Esta representación también se puede encontrar en forma de círculo, triángulo, etc.

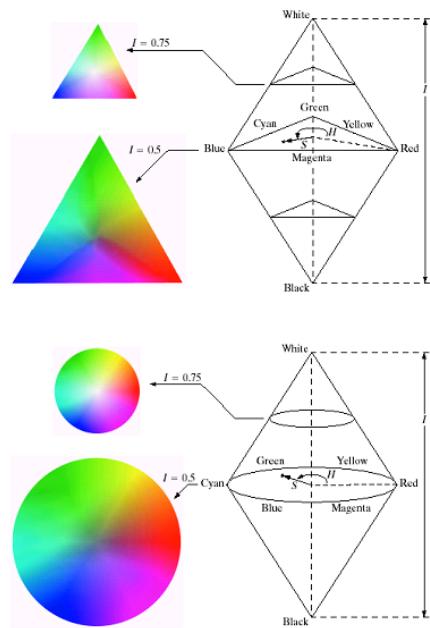


Fig 5.9 El modelo HSI en planos circular y triangular

Conversión de RGB a HSI

Las siguientes fórmulas pueden deducirse a partir de los razonamientos geométricos que acabamos de ver.

El componente H de tono está dado por:

$$H = \begin{cases} \Theta & \text{si } B \leq G \\ 360 - \Theta & \text{si } B > G \end{cases} \quad (5.7)$$

donde

$$\Theta = \cos^{-1} \frac{\frac{1}{2}[(R - G) + (R - B)]}{[(R - G)^2 + (R - B)(G - B)]^{1/2}}$$

El componente S de saturación:

$$S = 1 - (3 / (R + G + B)) [\min(R, G, B)] \quad (5.8)$$

Y el de intensidad:

$$I = (R + G + B) / 3 \quad (5.9)$$

Se asume que los valores RGB han sido normalizados al rango [0, 1] y que el ángulo Θ se mide desde el eje rojo del espacio HSI.

El tono puede normalizarse al rango [0, 1] dividiendo entre 360. Los otros componentes ya se obtienen normalizados si los valores de RGB estaban normalizados.

Conversión de HSI a RGB

Dados los valores de HSI en el intervalo [0, 1], para obtener los valores RGB, nuestras ecuaciones dependerán de los valores de H. Tendremos 3 sectores de interés correspondiendo a los intervalos de 120° que separan los primarios.

Comenzamos multiplicando H por 360, para regresar al rango [0, 360].

En el sector RG ($0 \leq H < 120^\circ$):

$$B = I(1 - S) \quad (5.10)$$

$$R = I[1 + S \cos H / \cos(60^\circ - H)] \quad (5.11)$$

$$G = 3I - (R + B) \quad (5.12)$$

En el sector GB ($120^\circ \leq H < 240^\circ$):

Se deben sustraer 120 grados de H:

$$H = H - 120$$

$$R = I(1 - S) \quad (5.13)$$

$$G = I[1 + S \cos H / \cos(60^\circ - H)] \quad (5.14)$$

$$B = 3I - (R + G) \quad (5.15)$$

En el sector RG ($240^\circ \leq H < 360^\circ$):

Se debe sustraer primero 240 grados de H:

$$H = H - 240^\circ$$

$$G = I(1 - S) \quad (5.16)$$

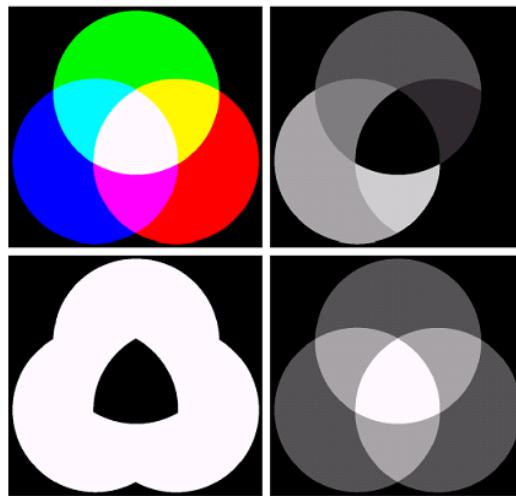
$$B = I[1 + S \cos H / \cos(60^\circ - H)] \quad (5.17)$$

$$R = 3I - (G + B) \quad (5.18)$$

Manipulación de los componentes HSI

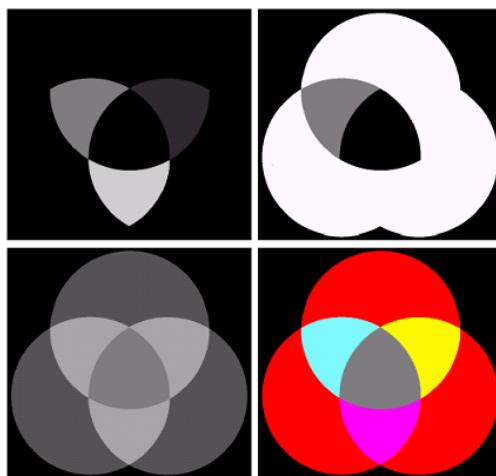
Para cambiar un color individual en cualquier región de una imagen RGB, cambiamos los valores de la región correspondiente en la imagen de tono, después convertimos la nueva imagen H, sin cambiar las imágenes S e I a RGB.

Par cambiar la saturación, hacemos lo mismo con la imagen S.



5.10 Imagen RGB y las imágenes de los componentes de tono, saturación e intensidad

Podemos observar que aunque estos cambios son simples, el modelo HSI es poderoso en el sentido en que nos permite control independiente sobre los componentes de tono, saturación e intensidad.



5.11 Manipulación de los componentes HSI de la imagen anterior y la imagen resultante

5.3 Procesamiento del falso color

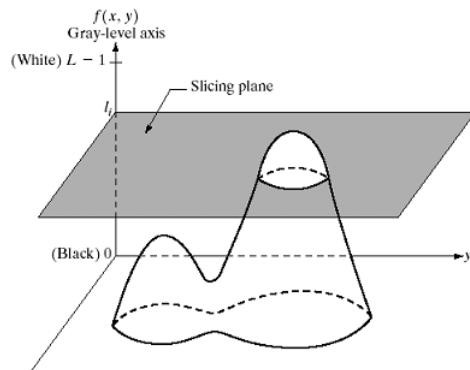
El procesamiento del falso color (a veces llamado seudocolor) consiste en asignar colores a valores de niveles de gris según un criterio especificado.

El uso primordial del seudocolor es para visualización humana e interpretación de eventos en escala de gris en una imagen o una secuencia de imágenes.

Una de las principales motivaciones en el uso de color es el hecho de que los seres humanos podemos discernir miles de tonalidades e intensidades de color, comparado con cerca de 2 docenas de sombras de gris.

5.3.1 División del color

Si una imagen se interpreta como una función 3D (coordenadas espaciales contra intensidad), el método puede ser visto como asignar planos paralelos al plano de coordenadas de la imagen, así cada plano divide la función en el área de intersección. La figura muestra un ejemplo al utilizar un plano en $f(x, y) = l_i$ para dividir la función de imagen en 2 niveles.



5.12 Plano divisor de los niveles de gris

Si se asigna un color diferente a cada lado del plano en la figura, cada pixel cuyo nivel de gris esté sobre el plano será codificado con un color, y cualquier pixel bajo el plano será codificado con otro. El resultado es una imagen en 2 colores cuya apariencia relativa puede ser controlada moviendo el plano divisor.

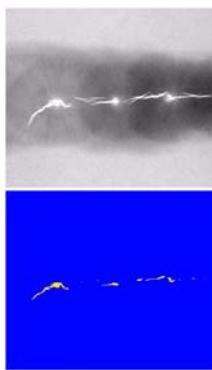
La técnica se puede resumir de la siguiente manera:

Si $[0, L-1]$ representa la escala de gris, el nivel l_0 representa el negro $[f(x,y) = 0]$ y en nivel l_{L-1} representa el blanco $[f(x, y) = L - 1]$, supóngase que P planos perpendiculares al eje de intensidad se definen en los niveles $l_1, l_2, l_3, \dots, l_p$, entonces, asumiendo que $0 < P < L-1$, los P planos dividen el nivel de gris en $P + 1$ intervalos V_1, V_2, \dots, V_{P+1} .

Las asignaciones de nivel de gris se hacen de acuerdo a

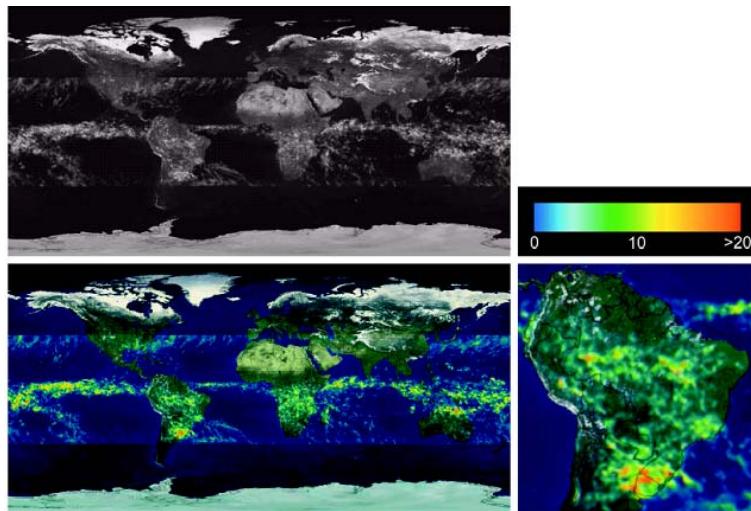
$$f(x, y) = c_k \quad \text{si } f(x, y) \in V_k \quad (5.19)$$

donde c_k es el color asociado a la intensidad k .



5.13 Imagen de rayos X de una soldadura y la imagen procesada

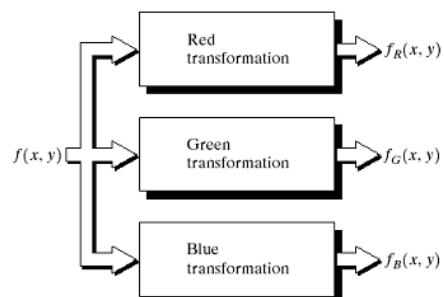
En la figura 5.13 vemos la imagen de rayos X de una soldadura con algunas fracturas y porosidades. Estas se pueden detectar fácilmente ya que, cuando no hay nada que impida el paso de los rayos X, estos toman el valor máximo (p.ej. 255). La otra imagen de la figura es una imagen procesada en donde al valor 255 se le da un color amarillo y los demás se presentan en azul, para facilitar la detección de fallas.



5.14 Imagen satelital monocromática a partir de varias bandas de satélite, la imagen con colores asignados

La figura 5.14 muestra una imagen en niveles de gris resultado de varias imágenes de sensores que miden los niveles de lluvia tropical. Debajo de esta se muestran los mismos resultados, pero asignando la gama de colores que se muestra a un costado a los niveles de gris correspondientes.

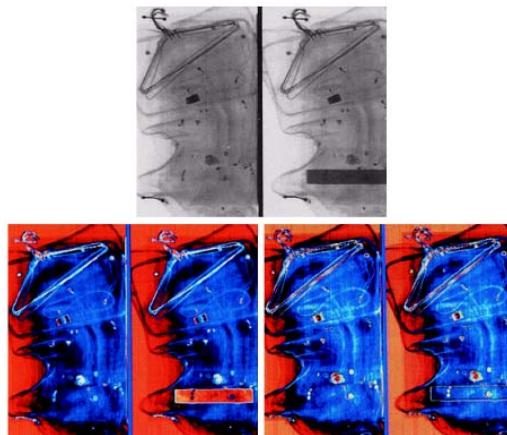
5.3.2 Transformaciones de nivel de gris a color



5.15 Diagrama de bloques para el procesamiento de pseudocolor

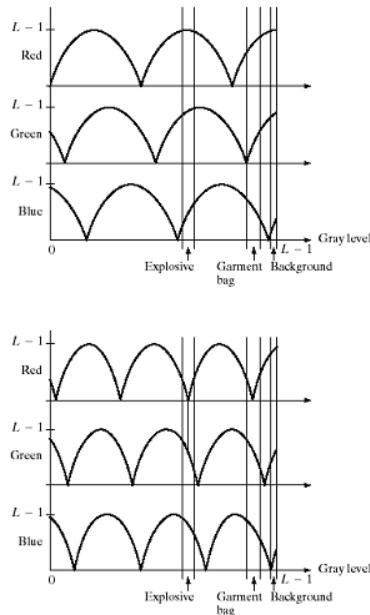
Se realizan 3 transformaciones independientes en el nivel de gris de cada pixel (una para cada canal R, G, B). Esto produce una imagen compuesta cuyo contenido de color es modulado por las funciones de transformación.

En la figura 5.16 se muestra un ejemplo. Las imágenes en niveles de gris son resultado de una máquina de rayos X en un aeropuerto. En la imagen en nivel de gris de la derecha se simula una carga de explosivos. Estas imágenes se procesan con las funciones de la figura 5.17 (en realidad es la misma función, pero desfasada para cada componente RGB con el fin de producir diferentes colores para diferentes niveles de



5.16 Mejora de pseudocolor usando transformación de grises a colores

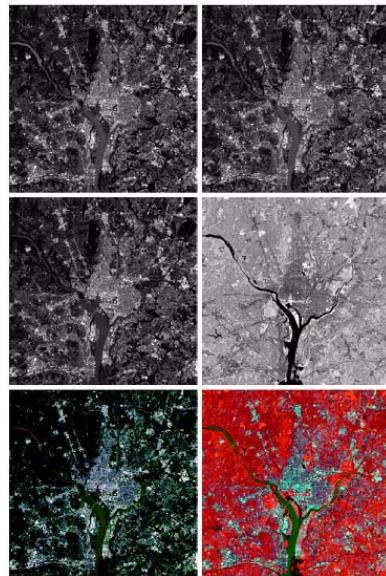
gris que muestren diferentes objetos) para obtener las imágenes a color que se muestran debajo.



5.17 Las funciones utilizadas. Se usa la misma función desfasada para cada componente

El ejemplo anterior utiliza una sola imagen para producir una imagen a color (tres imágenes monócromas). A veces lo que interesa es combinar varias figuras monócromas en una imagen compuesta de color (figura 5.12). Un uso frecuente de este enfoque se da en el procesamiento multiespectral de imágenes, donde diferentes sensores producen imágenes monócromas individuales, cada una en una banda espectral.

La figura muestra 4 imágenes monócromas dadas por 4 diferentes bandas de un satélite. La quinta figura muestra el resultado de utilizar las 3 primeras figuras como sus componentes de RGB y la sexta figura muestra el resultado de utilizar la cuarta figura (banda infrarroja) como componente de R de la imagen anterior.



5.18 Imágenes de las bandas 1-4 de satélite y combinación de ellas como componentes de RGB en la quinta y sexta imagen

5.4 Introducción al procesamiento de imágenes de color

Las técnicas desarrolladas en las secciones siguientes ilustran como pueden utilizarse las imágenes a color para una variedad de tareas de procesamiento de imágenes.

Hay 2 categorías. En la primera se procesa cada componente de imagen individualmente y después se forma una sola imagen compuesta a partir de todas. En la segunda categoría se trabaja directamente con pixeles de color.

Los pixeles de color se toman como vectores con componentes dados por los 3 (o más, en otros modelos) planos de color.

5.5 Transformaciones de color

Se transforma la imagen dentro de un mismo modelo de color.

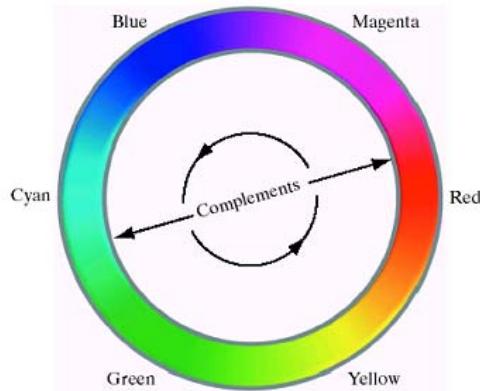
La formulación es la misma que para el nivel de gris

$$g(x, y) = T [f(x,y)] \quad (5.20)$$

$f(x, y)$ es una imagen de entrada a color, y $g(x, y)$ es la imagen de salida a color. No olvidemos que aquí el valor de cada pixel es una tupla de 3 o 4 elementos.

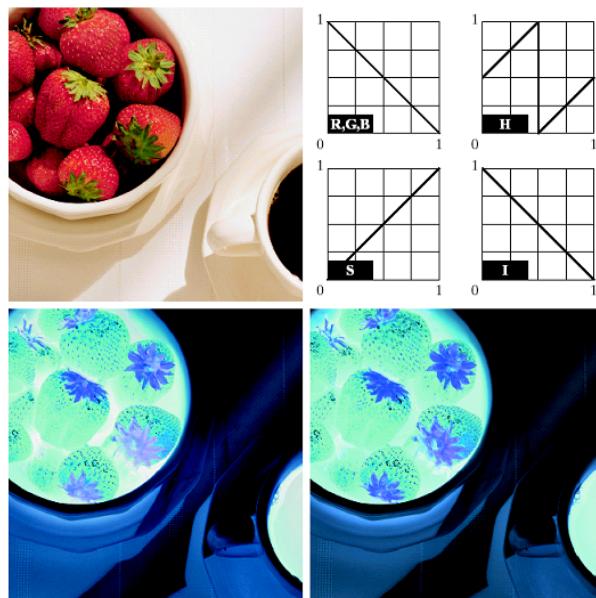
5.5.2 Complementos de color

Los tonos en posiciones opuestas del círculo de color son llamados complementarios o complementos. Nos interesan ya que los complementos son análogos a los negativos en las imágenes de niveles de gris.



5.19 El círculo de color y los complementos

Como en el caso de las imágenes de niveles de gris, son útiles para mostrar detalles claros en regiones oscuras de una imagen, particularmente cuando las regiones oscuras son dominantes en tamaño.



5.20 Transformación de complemento

5.5.3 Fraccionamiento del color

Subrayar un rango específico de colores en una imagen es útil para separar objetos de lo que los rodea. La idea básica es o mostrar los colores de interés para que sobresalgan del fondo, o usar la región definida por esos colores como máscara para procesamiento ulterior.

Se extienden las técnicas de fraccionamiento de nivel de gris.

Una de las maneras más simples de fraccionar o dividir una imagen de color es mapear los colores fuera del rango de interés con un color neutral. Si los colores de interés están dentro de un cubo de ancho W y centrados en un color prototípico (p. Ej. El color promedio) de componentes ($a_1, a_2, a_3, \dots, a_n$), la transformación necesaria es:

$$s_i = \begin{cases} 0.5 & \text{si } [|r_j - a_j| > W / 2] \text{ para todo } 1 \leq j \leq n \\ r_i & \text{en los otros casos} \end{cases} \quad i=1,2,\dots,n \quad (5.21)$$

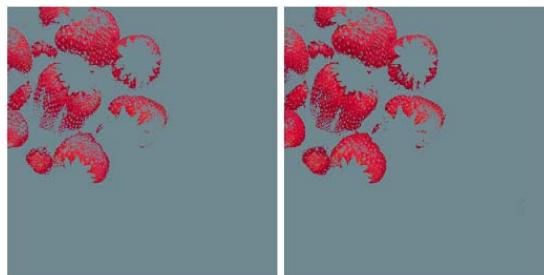
Esta transformación subraya los colores alrededor del color prototipo forzando los demás al punto medio del espacio de referencia de color (un punto neutral arbitrariamente escogido), por ejemplo un gris(0.5, 0.5, 0.5).

Si se usa una esfera:

$$s_i = \begin{cases} 0.5 & \text{si } \sum_{j=1}^n (r_j - a_j)^2 > R^2 \\ r_i & \text{en los otros casos} \end{cases} \quad i = 1, 2, \dots, n \quad (5.22)$$

Donde R es el radio de la esfera y a_1, a_2, \dots, a_n son los componentes en el centro (el color prototipo).

En la imagen 5.19 se muestra el resultado de aplicar las transformaciones con un cubo y una esfera a la imagen de la figura 5.20 con el fin de aislar el componente rojo, los valores para W y R_o se determinaron interactivamente (“prueba y error”).



5.21 Transformaciones de fraccionamiento de color con RGB con un cubo, y con una esfera.

5.5.4 Correcciones de color y tono

Se pueden realizar transformaciones de color en la mayoría de las computadoras de escritorio. Junto a scanners, cámaras digitales e impresoras, una computadora personal se convierte en un cuarto oscuro digital que nos permite ajustar y corregir colores y tonos, sin necesidad del procesamiento químico de un cuarto oscuro tradicional.

La efectividad de las transformaciones en el campo de la mejora de fotografías y la reproducción de color se juzga en última instancia en las impresiones.

Ya que estas transformaciones se desarrollan, refinan y evalúan en monitores, es necesario mantener un alto nivel de consistencia de color entre los monitores utilizados y los dispositivos de salida. Los colores en un monitor deberían representar lo más precisamente posible cualquier imagen fuente digitalizada, así como la salida final impresa. Esto se consigue con un *modelo de color independiente del dispositivo* que relaciona los gamuts de color de los monitores y los dispositivos de salida.

La elección de muchos *Sistemas de gestión de color* es el modelo CIE L*a*b*, también conocido como CIELAB. Los componentes de color L*a*b* están dados por las siguientes ecuaciones:

$$L^* = 116 \cdot h(Y / Y_w) - 16 \quad (5.23)$$

$$a^* = 500 [h(X/X_w) - h(Y/Y_w)] \quad (5.24)$$

$$b^* = 200 [h(Y/Y_w) - h(Z/Z_w)] \quad (5.25)$$

donde

$$h(q) = \begin{cases} q^{1/3} & q > 0.008856 \\ 7.787q + 16 / 116 & q \leq 0.008856 \end{cases} \quad (5.26)$$

donde X_w , Y_w , Z_w son los valores de tristímulo del blanco de referencia¹¹. El espacio formado por estos componentes es *colorimétrico* (los colores que se perciben idénticos se codifican idénticamente), *perceptualmente uniforme* (se perciben uniformemente las diferencias de color entre varios tonos), e *independiente del dispositivo*.

El modelo no se puede desplegar directamente, se necesita transformar a otro modelo, sin embargo su gamut engloba todo el espectro visible y puede representar de manera precisa los colores de cualquier dispositivo de despliegue, entrada o impresión.

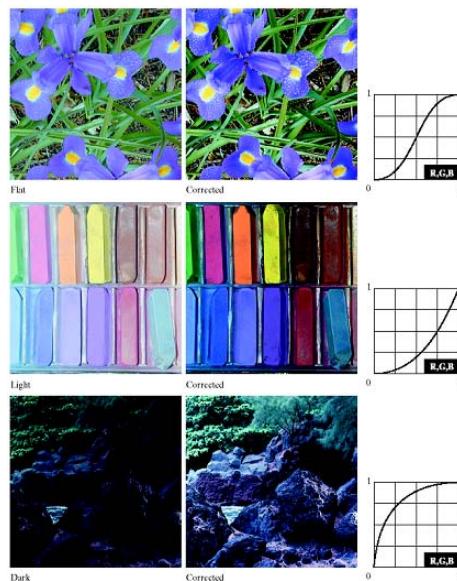
También es un excelente separador de la intensidad (la iluminación L^*) y el color (a^* para rojo menos verde y b^* para verde menos azul), lo que lo hace muy útil para manipulación de imágenes (edición de tono y contraste) y aplicaciones de compresión de imágenes.

El principal beneficio de los sistemas de calibración de imágenes es que permiten corregir desequilibrios de color y de tono de manera interactiva e independiente – en dos operaciones secuenciales. Primero se corrigen los problemas de rango tonal, y en seguida se corrigen las irregularidades de color.

Las transformaciones que modifican los tonos de una imagen son normalmente interactivas. La idea es ajustar experimentalmente los niveles de contraste y brillo de una imagen para conseguir la mejor presentación de detalles en un rango conveniente de intensidades. Los colores, en sí, no son cambiados, esto significa, en RGB y CMY (K), que se utilizará la misma función de transformación para cada uno de los componentes, en HSI, sólo se cambia el componente de intensidad.

En la figura 5.22 se muestran transformaciones típicas para corregir 3 tipos comunes de desequilibrio tonal (imagen clara, plana, y oscura). La primera curva es ideal para aumentar el contraste – su punto medio está anclado para iluminar los brillos y oscurecer las sombras (se puede utilizar la curva inversa para la operación inversa). La segunda y tercera transformaciones corren imágenes claras y oscuras y son parecidas a la transformación de potencia que vimos anteriormente.

¹¹ Típicamente el blanco de un difusor perfectamente reflejante en términos del estándar D65 de iluminación de CIE ($x = 0.3127$ y $y = 0.3290$ en el diagrama de cromaticidad de CIE).



5.22 Correcciones tonales afectando todos los componentes

Después de establecer correctamente las características tonales, se procede a corregir los desequilibrios de color (color por color). Esto puede automatizarse o, como en el caso anterior, realizarse de manera interactiva. La figura muestra la imagen corregida y diferentes desequilibrios de color con su correspondiente función correctora.



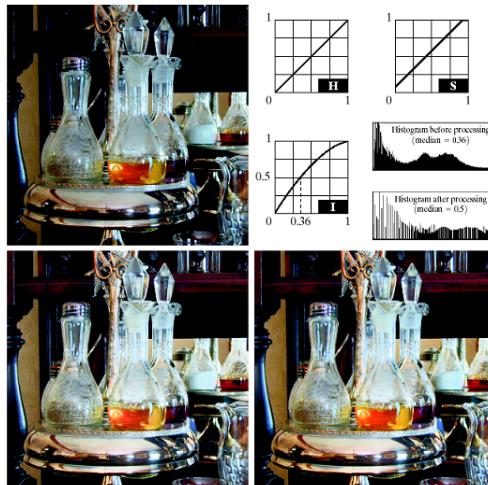
5.23 Balanceo de color

Al ajustar los componentes de color de una imagen, es importante recordar que cada acción afecta el equilibrio de color total de la imagen.

5.5.5 Procesamiento del histograma

Las transformaciones de procesamiento de nivel de gris de histograma anteriormente vistos pueden aplicarse directamente a imágenes de color de manera automática. Sin

embargo, no es muy conveniente aplicar ecualización de histograma directamente a los componentes de color de una imagen, ya que esto llevará a colores erróneos. Un enfoque más lógico es ecualizar las intensidades de color de manera uniforme, dejando los tonos sin cambiar.



5.24 Ecualización de histograma de componente intensidad y ajuste de saturación

Con este enfoque en general no se alteran los valores de tono y saturación de una imagen, aunque puede que la percepción de color total se altere un poco. Esto puede corregirse aumentando ligeramente el componente de saturación de la imagen. En la figura vemos la imagen original, la imagen ecualizada, y la imagen obtenida a partir de la saturación de la imagen ecualizada, como se ve en las funciones que se muestran.

5.6 Suavizado y realce

El siguiente paso en la transformación es el tomar en cuenta los vecinos del pixel a transformar.

5.6.1 Suavizado de imágenes a color

Se puede extender fácilmente el concepto de la máscara de promediado a imágenes de color. Como se ha visto, en lugar de valores escalares de nivel de gris utilizaremos los vectores componentes anteriormente mencionados.

Si S_{xy} denota el conjunto de coordenadas que define una vecindad con centro en (x, y) en una imagen RGB, la media de los vectores componentes RGB es

$$\bar{c}(x, y) = \frac{1}{K} \sum_{(x,y) \in S_{xy}} c(x,y) \quad (5.28)$$

de donde

$$\bar{c}(x, y) = \begin{cases} \frac{1}{K} \sum_{(x,y) \in S_{xy}} R(x,y) \\ \frac{1}{K} \sum_{(x,y) \in S_{xy}} G(x,y) \\ \frac{1}{K} \sum_{(x,y) \in S_{xy}} B(x,y) \end{cases} \quad (5.29)$$

Lo que es equivalente a suavizar cada plano RGB de manera independiente de la misma manera que lo hacíamos con el nivel de gris.

Se obtiene un resultado similar aplicando el filtro de suavizado al componente de intensidad de la imagen en HSI. Esto es aplicable también para otras técnicas de transformación de color.



5.25 Suavizado de imagen con cada componente RGB, con el componente de intensidad, y la diferencia entre ambos

5.6.2 Realzado de imágenes a color

En el sistema de color RGB, el laplaciano del vector c es

$$\nabla^2[c(x,y)] = \begin{cases} \nabla^2R(x,y) \\ \nabla^2G(x,y) \\ \nabla^2B(x,y) \end{cases} \quad (5.30)$$

De donde se deduce que podemos calcular el Laplaciano de una imagen de color calculando el Laplaciano de cada componente de color .

Se obtiene un resultado similar aplicando el laplaciano al componente de intensidad de la imagen en HSI.



5.26 Realzado de imagen con cada componente RGB, con el componente de intensidad, y la diferencia entre ambos

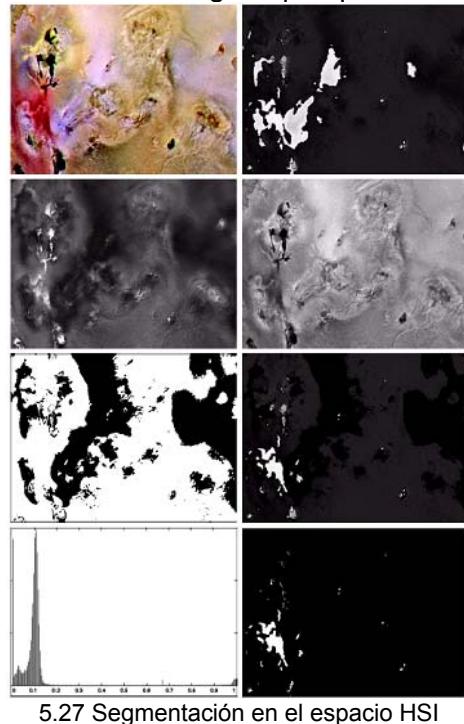
5.7 Segmentación de color

La segmentación es un proceso que partitiona la imagen en regiones. Aunque en este curso no se contempla, la veremos brevemente en esta sección.

5.7.1 Segmentación en HSI

Si queremos segmentar una imagen basados en los colores, es natural pensar en el modelo HSI porque el tono se representa individualmente en el componente H. Típicamente se utiliza el componente de saturación como máscara para aislar las regiones de interés de la imagen de tono.

Para segmentar la región de interés roja en la parte inferior izquierda de la primera imagen (figura 5.27), se separa la imagen en sus componentes HSI (imagen 2, 3 y 4), se obtiene una máscara binaria a partir de la imagen de saturación con un umbral del 10% (ya que el rojo que nos interesa está en ese rango). La sexta imagen muestra el resultado de usar esa máscara con la imagen de tono, y la octava imagen muestra el resultado de la misma operación obteniendo una máscara con el umbral del 9%. Esta última da una mejor aproximación de la región que queremos obtener.



5.7.2 Segmentación en el espacio vectorial RGB

La segmentación es una de las áreas donde se obtienen mejores resultados al utilizar vectores RGB.

Supóngase que el objetivo es segmentar objetos de un rango específico de color RGB. Dado un conjunto de puntos de color de muestra representativos de los colores que nos interesan, obtenemos una estimación del color “promedio” que queremos segmentar.

Representaremos este color promedio por el vector \mathbf{a} en RGB.

El objetivo de la segmentación es clasificar cada pixel RGB en una imagen dada como parte del rango de color especificado o no. Es necesario entonces medir de alguna manera la similitud. Utilizaremos la distancia euclídea: dada \mathbf{z} que denota un punto arbitrario en el espacio RGB, decimos que \mathbf{z} es similar a \mathbf{a} si la distancia entre ellos es menor a un umbral especificado D_0 . La distancia euclídea está dada por

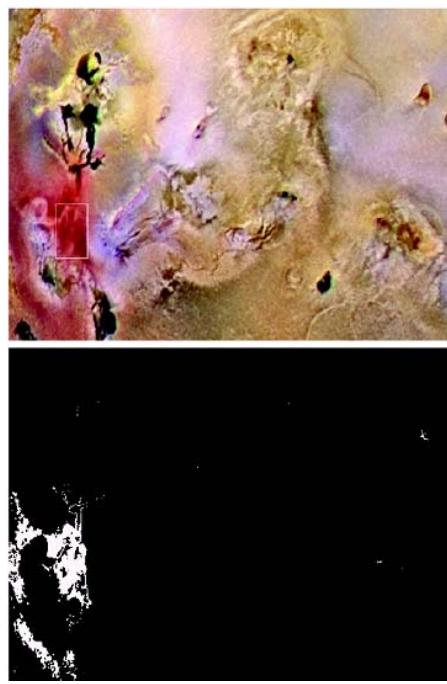
$$D(\mathbf{z}, \mathbf{a}) = \|\mathbf{z} - \mathbf{a}\| \quad (5.31)$$

$$= [(\mathbf{z} - \mathbf{a})^T (\mathbf{z} - \mathbf{a})]^{1/2} \quad (5.32)$$

$$= [(z_R - a_R)^2 + (z_G - a_G)^2 + (z_B - a_B)^2]^{1/2} \quad (5.33)$$

donde los subíndices R, G y B denotan los componentes RGB de los vectores \mathbf{a} y \mathbf{z} . El volumen comprendido dado $D(\mathbf{z}, \mathbf{a}) \leq D_0$ es una esfera sólida de radio D_0 .

Se puede trabajar con los cuadrados para evitar cálculo de raíces, sin embargo aún así es computacionalmente costoso para imágenes de tamaño práctico. Una buena aproximación es utilizar una caja centrada en \mathbf{a} con dimensiones a lo largo de cada eje de color proporcionales a la desviación estándar de las muestras alrededor de cada eje.



5.28 Segmentación en el espacio RGB

El rectángulo de la figura 5.28 contiene muestras de los colores rojizos que queremos segmentar. Se obtiene el vector \mathbf{a} a partir de la media de estas muestras y después se calculan las desviaciones estándar de los valores rojo verde y azul de estas muestras. Se centra una caja en \mathbf{a} y se le dan dimensiones correspondientes a un valor de 1.25 la desviación estándar en cada eje (R, G, B). De esta manera se obtiene la imagen binaria que se muestra en la figura.

6 Compresión de imágenes

En nuestros días, una enorme cantidad de información se maneja por medios digitales. La compresión es necesaria en aplicaciones de almacenamiento, transmisión, tiempo real. En este capítulo veremos los fundamentos de la compresión en la teoría de la información, la terminología utilizada en el campo de la compresión, y algunos modelos que se aplican comúnmente a las imágenes.

6.1 Fundamentos

Compresión y redundancia de datos

Se requiere distinguir entre los términos *datos* e *información*. Para nosotros, los datos son el medio por el que se transporta la información. Esto quiere decir que se pueden utilizar distintas cantidades de datos para representar la misma cantidad de información.

El término *compresión de datos* se refiere al proceso de reducir la cantidad de datos requeridos para representar una cantidad dada de información.

Cuando un conjunto de datos contiene más del mínimo necesario para transmitir la información, decimos que existe *redundancia de datos*.

La redundancia de datos se puede representar matemáticamente de la siguiente forma: Si n_1 y n_2 denotan el número de unidades que cargan información (por ejemplo, bits) en 2 conjuntos de datos representando la misma información, la *redundancia relativa de datos* R_D del primer conjunto de datos se define como:

$$R_D = 1 - 1/C_R \quad (6.1)$$

Donde C_R representa la *relación de compresión* y es

$$C_R = n_1/n_2 \quad (6.2)$$

Por ejemplo, si $n_2 = n_1$, $C_R = 1$ y $R_D = 0$, quiere decir que la primera representación no contiene datos redundantes (relativo a la segunda).

Cuando $n_2 \ll n_1$, $C_R \rightarrow \infty$ y $R_D \rightarrow 1$, quiere decir que hay una compresión considerable y datos altamente redundantes.

Cuando $n_2 \gg n_1$, $C_R \rightarrow 0$ y $R_D \rightarrow -\infty$, quiere decir que el segundo conjunto contiene mucho más datos que la representación original (el indeseable caso de *expansión de datos*).

C_R y R_D están generalmente en los intervalos abiertos $(0, \infty)$ y $(-\infty, 1)$, respectivamente.

Una relación de compresión de 10 (o 10:1) significa que el primer conjunto tiene 10 unidades por cada unidad del segundo conjunto. La redundancia correspondiente de 0.9 implica que 90% de los datos en el primer conjunto es redundante.

En el caso de la compresión de imágenes, se identifican y explotan 3 tipos de redundancia de datos: *redundancia de código*, *redundancia entre pixeles* y *redundancia psicovisual*. Al reducir o eliminar una o más de estas conseguimos comprimir los datos.

6.1.1 Redundancia de código

Utilizaremos el ya familiar concepto de histograma de niveles de gris de una imagen para mostrar cómo éste puede ayudar en gran medida en la construcción de códigos que reducen la cantidad de datos utilizados para representar la imagen.

Se asume que la variable aleatoria discreta r_k en el intervalo $[0, 1]$ representa los niveles de gris de una imagen y que cada r_k ocurre con una probabilidad $p_r(r_k)$:

$$p_r(r_k) = n_k / n \quad k = 0, 1, 2, \dots, L - 1 \quad (6.3)$$

donde L es el número de niveles de gris, n_k es el número de veces que el nivel k aparece en la imagen, y n es el número total de pixeles en la imagen.

Si el número de bits utilizado para representar cada valor de r_k es $I(r_k)$, entonces el número promedio de bits utilizados para representar cada pixel será

$$L_{\text{prom}} = \sum_{k=0}^{L-1} I(r_k) p_r(r_k) \quad (6.4)$$

Es decir que la longitud promedio de las palabras código asignadas a los valores de los niveles de gris se halla sumando el producto del número de bits empleados para representar cada nivel de gris y la probabilidad de que aparezca ese nivel. Así el número total de bits requeridos para codificar una imagen $M \times N$ es MNL_{prom} .

Si se representan los niveles de gris con un código binario natural de m bits,

$$L_{\text{prom}} = m \quad (6.5)$$

r_k	$p_r(r_k)$	Code 1	$I_1(r_k)$	Code 2	$I_2(r_k)$
$r_0 = 0$	0.19	000	3	11	2
$r_1 = 1/7$	0.25	001	3	01	2
$r_2 = 2/7$	0.21	010	3	10	2
$r_3 = 3/7$	0.16	011	3	001	3
$r_4 = 4/7$	0.08	100	3	0001	4
$r_5 = 5/7$	0.06	101	3	00001	5
$r_6 = 6/7$	0.03	110	3	000001	6
$r_7 = 1$	0.02	111	3	000000	6

Por ejemplo, en la tabla se muestra la distribución de niveles de gris de una imagen de 8 niveles. Si se utiliza un código natural de 3 bits (Code 1) para representar los 8 niveles, $L_{\text{avg}} = 3$ porque $I_1(r_k) = 3$ para todo r_k . Si se utiliza el código 2, el número promedio de bits es

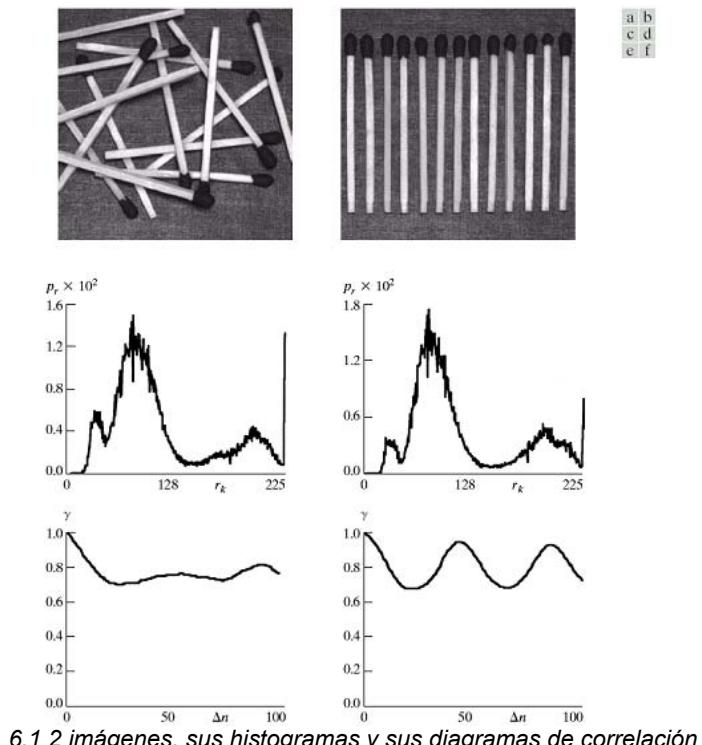
$$\begin{aligned}
 L_{\text{avg}} &= \sum_{k=0}^7 I(r_k) p_r(r_k) \\
 &= 2(0.19) + 2(0.25) + 2(0.21) + 3(0.16) + 4(0.08) + 5(0.06) + 6(0.03) + (5)(0.02) \\
 &= 2.7 \text{ bits}
 \end{aligned}$$

En el ejemplo anterior, asignar menos bits a los niveles de gris más probables y más bits a los menos probables se tradujo en una compresión de los datos. A este proceso se le denomina *codificación de longitud variable*.

La redundancia de código aparece cuando los códigos asignados a un conjunto de sucesos (cada nivel de gris) no han sido seleccionados de manera que es obtenga el mayor rendimiento posible de las probabilidades de que ocurran estos sucesos.

6.1.2 Redundancia entre pixeles

Consideremos las imágenes de la figura 6.2. Sus histogramas son prácticamente idénticos, con 3 rangos de nivel de gris dominantes. Se puede usar un código del tipo que acabamos de ver para reducir la redundancia de datos. Sin embargo, este proceso de codificación no alteraría el nivel de correlación entre los pixeles de las imágenes.



6.1.2 imágenes, sus histogramas y sus diagramas de correlación

Es decir que *los códigos empleados para representar los niveles de gris de cada imagen no tienen nada que ver con la correlación¹² entre pixeles*, estas correlaciones resultan de relaciones estructurales o geométricas entre los objetos de la imagen.

¹² La correlación no se verá en este curso. Para profundizar, consulte Gonzales & Woods (2002) o Jahne (2003).

Las figuras e y f muestran los coeficientes de autocorrelación calculados a lo largo de una línea de cada imagen. Estos perfiles se pueden relacionar cualitativamente con la estructura de las imágenes de las imágenes correspondientes, especialmente en la segunda figura donde la elevada correlación de los píxeles separados se puede relacionar directamente con el espacio entre cada cerillo.

Además, los píxeles adyacentes de ambas imágenes están muy correlacionados.

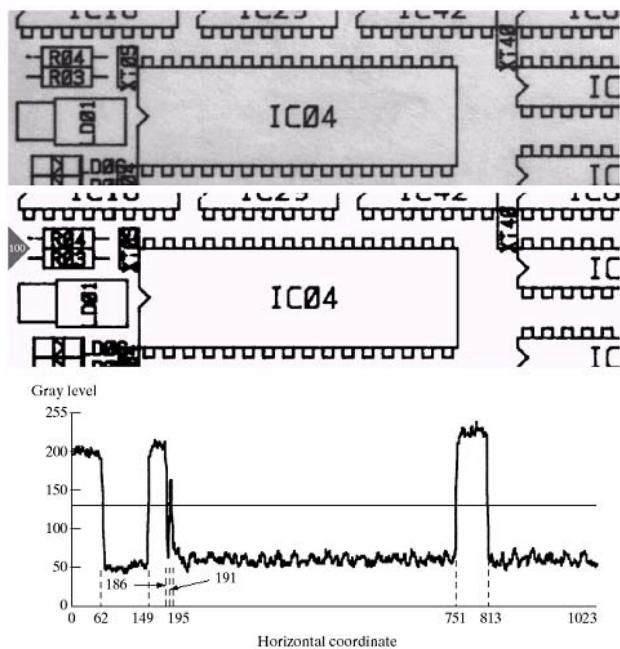
Estas ilustraciones reflejan otra forma de redundancia de los datos. *Ya que es posible predecir el valor de un pixel a partir de sus vecinos, la información que aporta individualmente un pixel es relativamente pequeña.*

Es decir, la mayor parte de la contribución visual de un pixel se puede inferir de los valores de sus vecinos (es redundante).

Para reducir estas redundancias entre píxeles, la distribución bidimensional se transforma a un formato más eficaz, usualmente no visualizable.

Por ejemplo, se podría utilizar las diferencias entre pixels adyacentes para representar una imagen.

A este tipo de transformaciones se les conoce como *correspondencias o mapeos*. Se dice que son *reversibles* si se pueden reconstruir los elementos originales a partir del conjunto de datos transformados.



Line 100: (1, 63) (0, 87) (1, 37) (0, 5) (1, 4) (0, 556) (1, 62) (0, 210)

6.2 Un mapa de bits, su imagen binaria, el diagrama de niveles de gris de la línea 100 y su codificación por longitud de códigos variable

En la figura se muestra un procedimiento de mapeo sencillo. La figura muestra un diagrama electrónico muestreado a 330 dpi, y su versión binaria. Ya que la imagen binaria contiene muchas regiones de intensidad constante, una representación más

eficiente puede ser construída mapeando los pixeles en cada línea de rastreo $f(x,0)$, $f(x,y)$, ..., $f(x, N-1)$ en una secuencia de pares $(g_1, w_1), (g_2, w_2)$, ..., donde g_i denota el i -ésimo nivel de gris en la línea y w_i el código de longitud de la i -ésima corrida. En la figura se muestra el código de longitud variable de la línea 100. Sólo 88 bits se necesitan para representar los 1024 bits de la línea.

6.1.3 Redundancia psicovisual

En el capítulo 2 vimos cómo la iluminación de una región tal como la percibe el ojo humano depende de muchos factores. El ojo no responde con la misma sensibilidad a toda información visual. Cierta información tiene menos importancia relativa que otra en nuestro proceso visual normal. Se dice que esta información es psicovisualmente redundante y se puede eliminar sin que se altere significativamente la calidad de la percepción de la imagen.

En general un observador busca características diferenciadoras, tales como bordes o regiones de diferentes texturas, y las combina mentalmente en grupos reconocibles. El cerebro relaciona estos grupos con el fin de completar la interpretación de la imagen.

Al contrario que las redundancias ya vistas, la redundancia psicovisual está asociada a la información visual real o cuantificable. Su eliminación es posible porque la propia información no es esencial para el procesamiento visual normal.

Como la eliminación de este tipo de datos se traduce en una pérdida de información cuantitativa, a menudo se denomina con el nombre de *cuantificación*.

Ya que es una operación irreversible, la cuantificación conduce a una compresión con pérdida de datos.



6.3 Imagen original de 8 bits, su cuantificación a 4 bits y la cuantificación mejorada a 4 bits

La figura muestra la imagen original con 256 niveles de gris y una imagen cuantificada a 16 niveles de gris. Como lo vimos en el capítulo 2, la cuantificación conduce a la presencia de falso contorno. La tercera imagen muestra la imagen cuantificada a 16 niveles de gris con un método que toma en cuenta las peculiaridades del sistema visual

humano, llamado *cuantificación de escala de grises mejorada* (IGS). Consiste en añadir a cada pixel un componente pseudoaleatorio con el fin de eliminar los cambios bruscos tales como los del falso contorno.

La tabla muestra el método. Se toman los 4 bits menos significativos del nivel de gris pixel anterior y se le suman al actual. Los 4 bits más significativos de la suma forman el código IGS. Si los 4 bits más significativos del valor actual son 1111, se suma 0000.

Pixel	Gray Level	Sum	IGS Code
$i - 1$	N/A	0000 0000	N/A
i	0110 1100	0110 1100	0110
$i + 1$	1000 1011	1001 0111	1001
$i + 2$	1000 0111	1000 1110	1000
$i + 3$	1111 0100	1111 0100	1111

6.1.4 Criterios de fidelidad

El quitar datos psicovisualmente redundantes tiene como resultado una pérdida de información real o cuantitativa. Existen 2 clases de criterios utilizados para cuantificar la pérdida de información: El criterio de fidelidad objetivo y el criterio de fidelidad subjetivo.

Cuando el nivel de pérdida de información puede expresarse como una función de la imagen original y la imagen comprimida (y posteriormente descomprimida), se dice que se basa en un *criterio de fidelidad objetiva*.

Uno de estos criterios es el error raíz media cuadrada (e_{rms}) entre una imagen de entrada y una de salida.

Si $f(x,y)$ es la imagen de entrada y $g(x,y)$ es la aproximación de $f(x,y)$ resultante de comprimir y después descomprimir la entrada, para cada valor de x y y , el *error* entre las dos se define como

$$e(x,y) = g(x,y) - f(x,y) \quad (6.6)$$

y el *error total* entre las dos imágenes es

$$\sum_{x=0}^{M-1} \sum_{y=0}^{N-1} [g(x,y) - f(x,y)] \quad (6.7)$$

Donde las imágenes son de tamaño $M \times N$. El e_{rms} entre $f(x,y)$ y $g(x,y)$ es la raíz cuadrada del error al cuadrado en promedio en el arreglo $M \times N$

$$e_{rms} = [1/MN \sum_{x=0}^{M-1} \sum_{y=0}^{N-1} [g(x,y) - f(x,y)]^2]^{1/2} \quad (6.8)$$

Otro criterio de fidelidad relacionado es la relación media cuadrada señal - ruido de la imagen comprimida-descomprimida. Si $g(x,y)$ se considera como la suma de la imagen original $f(x,y)$ y la señal de ruido $e(x,y)$, la relación media cuadrada señal - ruido de la imagen de salida es

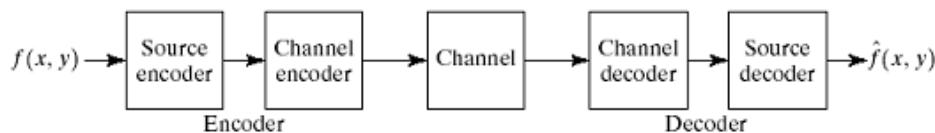
$$SNR_{ms} = \sum_{x=0}^{M-1} \sum_{y=0}^{N-1} g(x,y)^2 / \sum_{x=0}^{M-1} \sum_{y=0}^{N-1} [g(x,y) - f(x,y)]^2 \quad (6.9)$$

El valor rms de la razón señal-ruido, denotado SNR_{rms} se obtiene de la raíz cuadrada de SNR_{ms} .

Los criterios de fidelidad subjetiva se aplican mostrando una imagen a un grupo de observadores y promediando sus evaluaciones. Usualmente se utilizan valores como *mucho peor, peor, ligeramente peor, igual, ligeramente mejor, mejor, mucho mejor* para evaluar.

6.2 Modelos de compresión de imágenes

Las técnicas de compresión de datos vistas en la sección anterior usualmente son combinadas para formar sistemas prácticos de compresión de imágenes. En esta sección examinaremos las características generales de un sistema de este tipo y desarrollaremos un modelo general para representarlo.



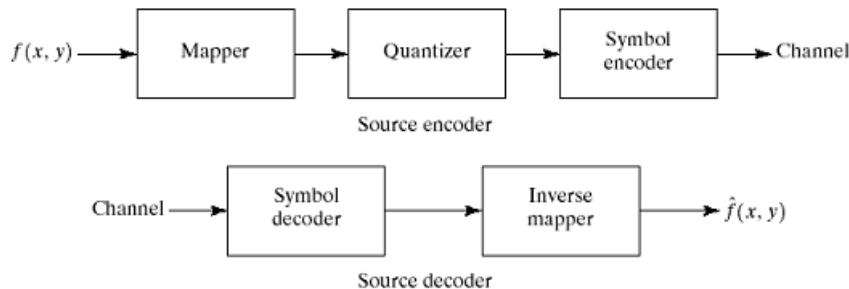
6.4 Modelo general de sistema de compresión

Una imagen $f(x,y)$ se introduce en un *codificador*, el cuál crea un conjunto de símbolos para representarla. Este conjunto se transmite por medio de un *canal* a un *decodificador*, donde se reconstruye una imagen $g(x,y)$. $g(x,y)$ puede o no ser una réplica exacta de $f(x,y)$, si lo es, se dice que el sistema es libre de error o que preserva la información, si no, hay algún nivel de distorsión en la imagen de salida.

El codificador está formado por un *codificador de fuente*, que remueve las redundancias de la entrada, y un *codificador de canal*, que aumenta la inmunidad al ruido de la salida del codificador de fuente. El *decodificador de canal* y *decodificador de fuente* que componen el decodificador realizan las funciones inversas. Si el canal está libre de ruido, se puede prescindir del codificador y decodificador de canal.

6.2.1 El codificador y el decodificador de fuente

Normalmente, se modela un codificador de fuente con una serie de 3 operaciones independientes diseñadas para reducir una de las 3 redundancias vistas anteriormente.



6.5 Codificador y decodificador de fuente

En la primera parte, el *mapeador* transforma los datos de entrada a un formato diseñado para reducir la redundancia entre pixeles. Esta operación es usualmente reversible y puede o no resultar en compresión de datos. La codificación por reducción de longitud de código es un ejemplo de mapeado que resulta en compresión.

La segunda parte, o el *cuantificador*, reduce la precisión de la salida del mapeador de acuerdo a un criterio de fidelidad preestablecido (esta etapa reduce la redundancia psicovisual). Esta operación es irreversible (debe ser omitida si se quiere compresión sin errores).

En la etapa final el *codificador de símbolos* crea un código de longitud variable que representa la salida del cuantificador y mapea la salida de acuerdo al código, asignando las palabras clave más cortas a los valores más frecuentes, reduciendo la redundancia de código.

Esta operación es reversible.

El decodificador de fuente de la figura contiene sólo 2 componentes: el *decodificador de símbolos* y el *mapeador inverso*. Estos bloques realizan en orden invertido las operaciones inversas del codificador de símbolos de la fuente y el mapeador.

6.2.2 El codificador y decodificador de canal

Estos están diseñados para reducir el impacto de ruido en el canal insertando redundancia controlada en los datos codificados por la fuente.

Una de las técnicas más utilizadas de codificación de canal es el código de Hamming (R.W. Hamming [1950]), basado en añadir suficientes bits a los datos codificados para asegurarse de que un número mínimo de bits debe cambiar entre palabras de código válidas.

Hamming mostró que si 3 bits de redundancia se añaden a una palabra de 4 bits, de modo que la *distancia*¹³ entre 2 palabras válidas es 3, se puede detectar y corregir cualquier error de *un solo bit*.

La palabra de código Hamming de 7 bits (7, 4) $h_1h_2\dots h_5h_6h_7$ asociada a un número binario de 4 bits $b_4b_3b_2b_1$ es

$$\begin{array}{ll} h_1 = b_3 \oplus b_2 \oplus b_0 & h_3 = b_3 \\ h_2 = b_3 \oplus b_1 \oplus b_0 & h_5 = b_2 \\ h_4 = b_2 \oplus b_1 \oplus b_0 & h_6 = b_1 \\ & h_7 = b_0 \end{array}$$

donde + en un círculo denota la operación OR exclusiva.

Nótese que los bits h_1 , h_2 y h_4 son bits de paridad par¹⁴ para los campos de bits $b_3b_2b_0$, $b_3b_1b_0$ y $b_2b_1b_0$, respectivamente.

¹³ La distancia entre 2 palabras de código se define como el número mínimo de dígitos que debe cambiar en una palabra para obtener la otra palabra. Por ejemplo, la distancia entre 101101 y 100111 es 2.

Para decodificar, se revisa que si valor codificado tenga paridad impar en los campos de bits en que se estableció paridad par. Si la palabra de paridad $c_4c_2c_1$ es diferente de cero, hay un error

$$\begin{aligned}c_1 &= h_1 \oplus h_3 \oplus h_5 \oplus h_7 \\c_2 &= h_2 \oplus h_3 \oplus h_6 \oplus h_7 \\c_4 &= h_4 \oplus h_5 \oplus h_6 \oplus h_7\end{aligned}$$

y el decodificador simplemente complementa el bit en la posición indicada por la palabra de paridad. Después se extrae el valor binario decodificado de la palabra corregida como $h_3h_5h_6h_7$

6.3 Elementos de la teoría de la información*

Después de ver varias maneras de reducir la cantidad de datos necesaria para representar una imagen, es natural preguntarse qué tantos datos son los mínimos necesarios para representarla... la teoría de la información provee las bases matemáticas para responder este tipo de preguntas.

6.3.1 Midiendo la información

La premisa fundamental de la teoría de la información es que la generación de información puede ser modelada como un proceso probabilístico que puede ser medido de una manera que coincide con lo que nos dice la intuición.

Se dice que un evento aleatorio E que ocurre con probabilidad $P(E)$ contiene

$$I(E) = \log 1 / P(E) = -\log P(E) \quad (6.10)$$

unidades de información. La cantidad $I(E)$ se llama *información propia* de E. En general, la cantidad de información propia atribuída a E se relaciona inversamente a la probabilidad de E. Si $P(E) = 1$ (El evento siempre ocurre), $I(E) = 0$ y no se le atribuye información¹⁴.

La base del logaritmo en la ecuación determina la unidad utilizada para medir la información. Si se escoge la base 2, la unidad será el bit. Si $P(E) = 1/2$, $I(E) = -\log_2(1/2)$, o un bit. Es decir, cuando ocurre uno de 2 posibles eventos igualmente posibles, la cantidad de información es 1 bit.

6.3.2 El canal de información

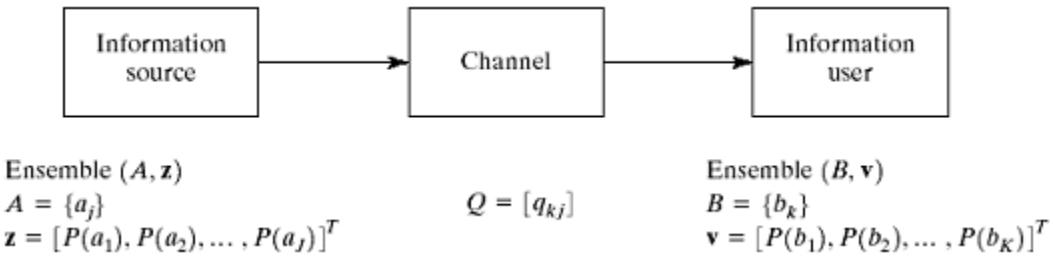
Cuando se transmite información propia entre una fuente de información y el usuario de la información, se dice que la fuente está conectada al usuario por medio de un *canal de información*. El canal es el medio físico que une la fuente con el usuario.

¹⁴ Una cadena de bits binarios tiene paridad par si el número de bits con valor 1 es par.

* Este tema se verá sólo si queda tiempo.

¹⁵ Como no hay incertidumbre asociada al evento, no hay transferencia de información al comunicar que el evento ha ocurrido. Si $P(E) = 0.99$, el comunicar que el evento ha ocurrido significa dar una pequeña cantidad de información. El comunicar que NO ha ocurrido significa más información, porque es mucho menos probable que esto ocurra.

Nos interesa particularmente la *capacidad del sistema*, definida como su habilidad para transferir información.



La fuente de información de la figura genera una secuencia aleatoria de símbolos de un conjunto finito (o infinito pero contable) de símbolos posibles. El conjunto de símbolos de la fuente $\{a_1, a_2, \dots, a_J\}$ es el *Alfabeto de la fuente A*, y los elementos del conjunto son llamados *símbolos*. La probabilidad de que el evento de la fuente resulte en un símbolo a_j es $P(a_j)$ y

$$\sum_{j=1}^J P(a_j) = 1 \quad (6.10)$$

Un vector $J \times 1 \mathbf{z} = [P(a_1), P(a_2), \dots, P(a_J)]^T$ se usa para representar el conjunto de todas las probabilidades de los símbolos de la fuente.

(A, \mathbf{z}) describe la fuente de información completamente.

La probabilidad de que la fuente discreta emita un símbolo a_j es $P(a_j)$, así que la información propia generada por la emisión de un solo símbolo de la fuente es $I(a_j) = -\log P(a_j)$. Si se generan k símbolos, el símbolo a_j se generará, en promedio, $kP(a_j)$ veces.

Por lo tanto la información propia promedio de k salidas será

$$-k \text{ SUMATORIA } (j = 1 \text{ a } J) P(a_j) \log P(a_j)$$

y la información promedio por salida de la fuente, $H(z)$ es

$$H(z) = - \text{SUMATORIA}(j = 1 \text{ a } J) P(a_j) \log P(a_j)$$

Esta cantidad es llamada la *incertidumbre* o *entropía* de la fuente. Define la cantidad promedio de información (en unidades m-arias por símbolo) obtenida al observar una salida de una sola fuente.

Si los símbolos de la fuente son igualmente probables, la entropía se maximiza y la fuente provee la información promedio más grande posible por símbolo de la fuente.

Después de modelar la fuente de información, podemos modelar fácilmente el canal de información. La información transferida a la salida del canal toma valores de un conjunto de símbolos $\{b_1, b_2, \dots, b_K\}$ llamado el *alfabeto de canal*, B . La probabilidad del evento de que el símbolo b_k se presente al *usuario de la información* es $P(b_k)$. (B, \mathbf{v}) donde $\mathbf{v} = [P(b_1), P(b_2), \dots, P(b_K)]^T$ describe completamente la salida de canal (y por tanto la información recibida por el usuario).

6.4 Compresión libre de errores

En algunas aplicaciones la compresión libre de errores es la única alternativa aceptable. En esta sección veremos las principales estrategias de compresión libre de errores en uso. Estas están generalmente compuestas de 2 operaciones relativamente independientes: el diseño de una representación por reducción de redundancia entre píxeles y codificación de la representación para eliminar la redundancia de código.

6.4.1 Codificación por longitud variable

El enfoque más simple de compresión sin errores es reducir solamente la redundancia de código. Para hacerlo construiremos un código de longitud variable que asigne las palabras más cortas a los niveles de gris más probables. Hay varias maneras de construir tal código. Recuérdese que, en la práctica, los símbolos de entrada pueden ser tanto los niveles de gris como el resultado del mapeo para reducir la redundancia entre píxeles.

El código de Huffman

La técnica más popular para remover redundancia de código se le debe a Huffman. Al codificar individualmente los símbolos de una fuente de información, el código de Huffman obtiene el menor número posible de símbolos de código por símbolo de la fuente.

El primer paso es crear una serie de reducciones de fuente ordenando las probabilidades de los símbolos y combinando los símbolos de menor probabilidad en un símbolo único que los reemplaza en la siguiente reducción de fuente. En la figura se muestra el proceso.

Original source		Source reduction			
Symbol	Probability	1	2	3	4
a_2	0.4	0.4	0.4	0.4	0.6
a_6	0.3	0.3	0.3	0.3	0.4
a_1	0.1	0.1	0.2	0.3	
a_4	0.1	0.1	0.1		
a_3	0.06	0.1			
a_5	0.04				

Reducciones de fuente de Huffman

A la izquierda, se muestran los símbolos ordenados de mayor a menor probabilidad. Para la primera reducción de código, se unen 0.06 y 0.04 y se combinan para formar un símbolo compuesto de probabilidad 0.1. Se repite el proceso con la columna resultante hasta que obtenemos una fuente reducida de 2 símbolos.

El segundo paso del procedimiento de Huffman es codificar cada fuente reducida, empezando con la menor hasta llegar a la original.

Original source			Source reduction							
Sym.	Prob.	Code	1	2	3	4				
a_2	0.4	1	0.4 1	0.4 1	0.4 1	0.6 0				
a_6	0.3	00	0.3 00	0.3 00	0.3 00	0.4 1				
a_1	0.1	011	0.1 011	0.2 010	0.3 01					
a_4	0.1	0100	0.1 0100	0.1 011						
a_3	0.06	01010	0.1 0101							
a_5	0.04	01011								

Asignación de códigos

El código binario mínimo para una fuente de 2 símbolos está formado por los símbolos 0 y 1. Como se muestra en la figura, estos símbolos se asignan a los 2 símbolos de la derecha (no importa a cuál). Ya que el símbolo de probabilidad 0.6 se generó combinando 2 símbolos en la fuente reducida a su izquierda, el 0 utilizado para codificarlo se usa ahora para los 2 símbolos que lo formaron, y un 0 y 1 se asignan arbitrariamente después de este para distinguirlos. La operación se repite hasta alcanzar la fuente original. El código final se muestra en la figura.

Podemos ver que la longitud promedio del código es

$$\begin{aligned} L_{\text{prom}} &= (0.4)(1) + (0.3)(2) + (0.1)(3) + (0.1)(4) + (0.06)(5) + (0.04)(5) \\ &= 2.2 \text{ bits /símbolo} \end{aligned}$$

El procedimiento de Huffman crea el código óptimo para un conjunto de símbolos y probabilidades si los símbolos se codifican de uno por uno.

La codificación y decodificación se hace por medio de una tabla. El código es un código de bloque que puede ser decodificado instantáneamente de manera única.

(ejemplo, decodificar 010100111100)

Otros códigos de longitud variable casi óptimos

Cuando se tiene una gran cantidad de símbolos a codificar, la construcción de un código binario de Huffman no es una tarea trivial. Para el caso general de J símbolos de la fuente, se requieren $J - 2$ reducciones de fuente y $J - 2$ asignaciones.

Dada tal complejidad computacional, a veces se requiere sacrificar eficiencia de código por simplicidad en la construcción de código.

En la tabla se ilustran 4 alternativas. En el último renglón observamos que la longitud promedio del código de Huffman es menor que la de los demás. El código binario natural tiene el promedio más grande.

Source symbol	Probability	Binary Code	Huffman	Truncated Huffman	B ₂ -Code	Binary Shift	Huffman Shift
<i>Block 1</i>							
a_1	0.2	00000	10	11	C00	000	10
a_2	0.1	00001	110	011	C01	001	11
a_3	0.1	00010	111	0000	C10	010	110
a_4	0.06	00011	0101	0101	C11	011	100
a_5	0.05	00100	00000	00010	C00C00	100	101
a_6	0.05	00101	00001	00011	C00C01	101	1110
a_7	0.05	00110	00010	00100	C00C10	110	1111
<i>Block 2</i>							
a_8	0.04	00111	00011	00101	C00C11	111 000	00 10
a_9	0.04	01000	00110	00110	C01C00	111 001	00 11
a_{10}	0.04	01001	00111	00111	C01C01	111 010	00 110
a_{11}	0.04	01010	00100	01000	C01C10	111 011	00 100
a_{12}	0.03	01011	01001	01001	C01C11	111 100	00 101
a_{13}	0.03	01100	01110	100000	C10C00	111 101	00 1110
a_{14}	0.03	01101	01111	100001	C10C01	111 110	00 1111
<i>Block 3</i>							
a_{15}	0.03	01110	01100	100010	C10C10	111 111 000	00 00 10
a_{16}	0.02	01111	010000	100011	C10C11	111 111 001	00 00 11
a_{17}	0.02	10000	010001	100100	C11C00	111 111 010	00 00 110
a_{18}	0.02	10001	001010	100101	C11C01	111 111 011	00 00 100
a_{19}	0.02	10010	001011	100110	C11C10	111 111 100	00 00 101
a_{20}	0.02	10011	011010	100111	C11C11	111 111 101	00 00 1110
a_{21}	0.01	10100	011011	101000	C00C00C00	111 111 110	00 00 1111
<i>Entropy</i>	4.0						
<i>Average length</i>	5.0	4.05	4.24	4.65	4.59	4.13	

Huffman truncado

La columna 5 ilustra una modificación simple del código Huffman conocida como *código Huffman truncado*. Un código de Huffman truncado se genera codificando sólo los PSI símbolos más probables de la fuente, siendo PSI algún entero positivo menor a J. Un código de prefijo seguido por un código de longitud variable se usa para representar los demás símbolos de la fuente. En la tabla PSI = 12 y el código de prefijo se genera como la 13ava palabra de código de Huffman (es decir, un símbolo de prefijo cuya probabilidad es la suma de las probabilidades de los símbolos a_{13} a a_{21} se incluye como el 13avo símbolo durante la codificación de Huffman de los 12 símbolos más probables, los 9 restantes se codifican usando el código de prefijo, en este caso 10, y un valor binario de 4 bits igual al subíndice del símbolo menos 13).

Código B

Cada palabra de código se hace a partir de bits de *continuación*, denotados C, y bits de información. El único propósito de los bits de continuación es separar palabras de código individuales, así que estos simplemente alternan entre 0 y 1 para cada palabra de código en una cadena. El código que se muestra es llamado código B₂, ya que se usan 2 bits de información por bit de continuación. La secuencia de códigos B₂ que

corresponde a la cadena de símbolos de la fuente $a_{11}a_2a_7$ es 001 010 101 000 010 o 101 110 100 110, dependiendo de si el primer bit de continuación se toma como 0 o 1.

Shift codes

Un *shift code* se genera

- 1) Ordenando los símbolos de la fuente de manera que sus probabilidades decrezcan (de manera monótona).
- 2) Dividiendo el número total de símbolos en bloques de símbolos del mismo tamaño.
- 3) Codificando los elementos individuales en cada bloque de manera idéntica.
- 4) Añadiendo símbolos especiales de *shift up* y / o *shift down* para identificar cada bloque.

Cuando se reconoce un símbolo *shift up* o *shift down* en el decodificador, se mueve un bloque arriba o abajo respecto al bloque de referencia predefinido.

Para generar el shift code de 3 bits de la columna 7, los símbolos de la fuente se ordenan por probabilidad de ocurrencia y se dividen en 3 bloques de 7 símbolos. Los símbolos individuales (a_1 a a_7) del bloque superior (el bloque de referencia) se codifican con los símbolos binarios 000 a 110. El octavo código (111) no se incluye en el bloque de referencia, pero se usa como un control de *shift up* que identifica a los bloques restantes (en este caso no se utiliza un símbolo *shift down*).

Los símbolos en los bloques restantes se codifican con uno o dos símbolos *shiftup* en combinación con los códigos binarios que representan el bloque de referencia.

Por ejemplo, el símbolo de la fuente a_{19} se codifica como 111 111 100.

El shift code de Huffman en la columna 8 se genera de manera similar. La diferencia es en la asignación de la probabilidad al símbolo *shift* antes de codificar con Huffman el bloque de referencia. Normalmente esta asignación se hace sumando las probabilidades de todos los símbolos de fuente fuera del bloque de referencia (igual que en el código de Huffman truncado). Aquí se suman los símbolos a_8 a a_{21} y es 0.39. El símbolo *shift* es entonces el símbolo más probable y se le asigna una de las palabras Huffman más cortas (00).

Codificación aritmética

Codificación LZW

6.4.3 Codificación por planos de bits

6.6 Estándares de compresión de imágenes

Bibliografía y artículos consultados

- [1] Gonzalez y Woods. **Digital Image Processing**. 2^a edición. Prentice Hall.
- [2] Gonzalez y Woods. **Procesamiento Digital de Imágenes**. Prentice Hall.
- [3] Jähne, Bernd. **Digital Image Processing**. 5^a edición. Springer 2002.
- [4] Foley, Van Dam, Feiner y Hughes. **Computer Graphics, principles and practice**. 2^a edición en C. Addison Wesley.
- [5] Hearn y Baker. **Gráficas por computadora**. 2^a edición. Prentice Hall.