UNIDAD I

INTRODUCCIÓN AL PROCESAMIENTO DIGITAL DE IMÁGENES

1.1 Fundamentos y representación digital de imágenes.

Interés por PDI

Mejora de la información pictórica para la interpretación humana.
Procesar una imagen para la percepción autónoma de una máquina.

Una de las primeras aplicaciones en PDI era mejorar las fotografías digitalizadas de periódico enviadas por cable submarino entre Londres y Nueva York. Se utilizó entonces el sistema Bartlane que inicialmente contemplaba imágenes monocromas con 5 tonos de gris. Hacia 1929 se logró mejorar la calidad de las imágenes con un sistema Bartlane de 15 tonos.

El PDI tuvo un gran impulso con:

- El surgimiento de las PC' s digitales de gran potencia.
- Los programas espaciales.

En 1964 se mejoraron imágenes ruidosas de la Luna transmitida por la sonda espacial Ranger 7. De entonces a la fecha el PDI ha crecido enormemente.

Aplicaciones

Medicina: realzan contraste o codifican los niveles de intensidad de colores para facilitar la interpretación de imágenes de rayos X y otras imágenes biomédicas.

Ambientalista: estudian los patrones de polución a partir de imágenes aéreas o satelitales.

Arqueología: para restaurar imágenes borrosas que eran los únicos registros existentes de piezas extrañas perdidas o dañadas después de fotografiarlas.

Física y áreas afines: realzan imágenes de experimentos difíciles de repetir (por costo u operación).

En cuanto al PDI para la percepción automatizada se tiene:

- Reconocimiento automático de caracteres.
- Visión industrial para ensamblado e inspección de productos.
- Reconocimientos militares.
- Reconocimiento de huellas digitales.
- Análisis de muestras de sangre.
- Análisis de imágenes aéreas y satelitales para predicción del clima.

Representación de una imagen.

Imagen: es una función bidimensional de intensidad de luz *f(x,y),* donde *x* y *y* son las coordenadas espaciales y el valor de *f* en un punto *(x,y)* cualquiera es proporcional al nivel de brillo (o nivel de gris) de la imagen en ese punto.

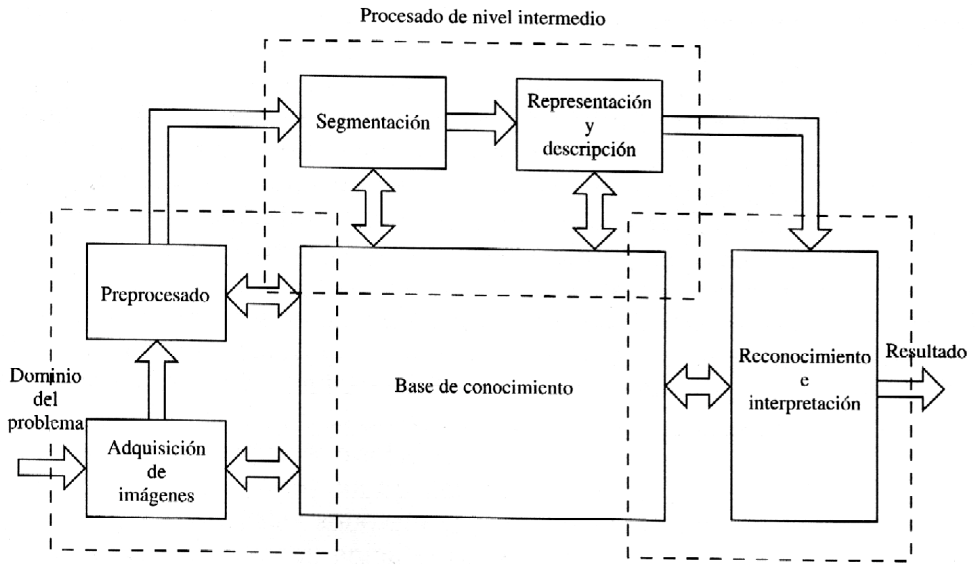
Imagen digital: Imagen *f(x,y)* que se ha discretizado tanto en coordenadas espaciales como en brillo. Se puede conceptualizar como una matriz cuyos índices de fila y columna indican un punto de la imagen y el valor del elemento de la matriz indica el nivel de gris. Los elementos de la imagen se denominan "pixels".

El tamaño de la imagen digital depende de la aplicación. Se entiende por tamaño de imagen digital la resolución total de los pixels que componen la imagen. Una imagen de 512x512 pixels y 128 niveles de gris corresponde a la calidad de imágenes de la TV monocromática.

1.2 Etapas fundamentales de procesamiento digital de imágenes

En la figura I.1 se muestra un diagrama a bloques de las etapas en el procesamiento de imágenes y la secuencia en que comúnmente se aplican. Mediante estas etapas se debe producir un resultado a partir de un determinado problema mediante PDI.

Ejemplo: Se puede utilizar PDI para leer de forma automática la dirección de paquetes postales.



Etapas del Procesamiento Digital de Imagenes

Figura I.1

Dominio del problema: paquetes postales
Resultado: dirección y secuencia de caracteres.

Adquisición de imágenes: Puede ser un escáner o una cámara y una tarjeta A/D como interfase. Para el ejemplo concreto se utilizaría una cámara de gran resolución.

Preprocesamiento: Mejora la imagen para que los procesos posteriores tengan éxito. Aquí se aplican técnicas para mejorar contraste, eliminar ruido, realzar contornos, etc.

Segmentación: Separa una imagen en sus partes constitutivas u objetos. En el reconocimiento de caracteres, la segmentación se usa para aislar caracteres y palabras del fondo.

Representación y Descripción : Considera cómo representar las partes u objetos segmentados en una imagen; es decir, se pueden representar en función de su contorno o de las características de los pixels que forman la región.
Para representar partes de una imagen es necesario utilizar la descripción, que no es más que usar propiedades y atributos de los objetos para clasificarlos.

Reconocimiento: Proceso que asigna una etiqueta a un objeto basándose en la información proporcionada por sus descriptores.

Interpretación: Implica asignar significado a un conjunto de objetos reconocidos. Por ejemplo, una secuencia de caracteres reconocidos se puede interpretar como un nombre.

Base de conocimiento: Generalmente es una base de datos con información que tiene que ver con el fenómeno que se estudia o el cual se desea analizar mediante PDI. Además de guiar la operación de cada módulo de procesamiento, la base de conocimiento controla también la interacción entre módulos.

En algunos casos no son necesarias todas las etapas de PDI. Por ejemplo, en el realce de imágenes para la percepción humana rara vez se usan algoritmos más allá del preprocesamiento.

1.3 Elementos de un sistema de PDI.

Los elementos son:

- Adquisición.
- Almacenamiento.
- Procesamiento.
- Presentación.

Adquisición

Para la adquisición se requieren dos elementos:

- Un dispositivo sensible a energía electromagnética (luz) que produzca una señal eléctrica proporcional a la intensidad luminosa.

- Un dispositivo para convertir la señal eléctrica del sensor a una forma digital. Un convertidor A/D o digitalizador. Algunos de los dispositivos más usados e n la adquisición son:

- Microdensitómetro
- Tubo vidicon
- Sensores CCD o de estado sólido.

Los microdensitómetros tienen la desventaja de que requieren que la imagen esté en forma de transparencia, sin embargo, presentan una gran exactitud.

Las cámaras de vidicón y sensores CCD detectan el nivel de intensidad luminosa reflejada por la imagen.

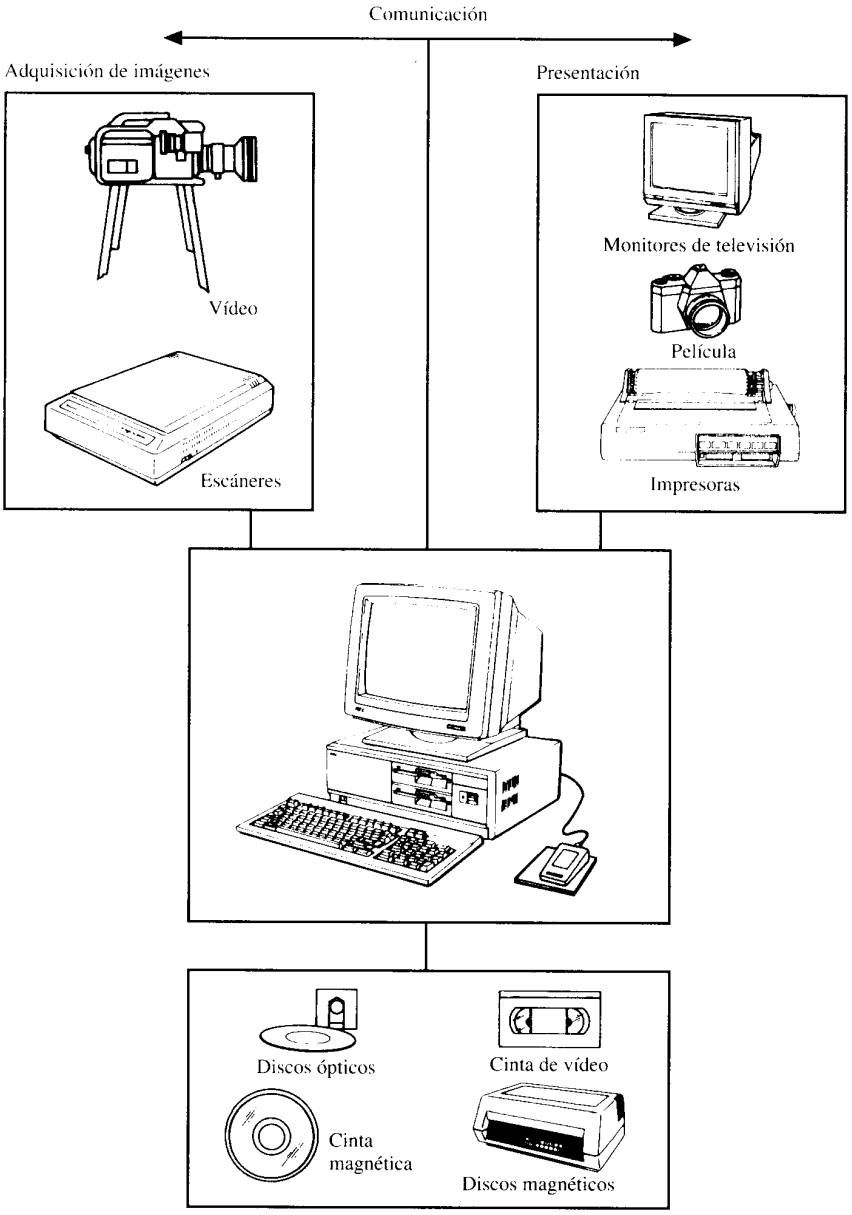


Figura I.2
Elementos funcionales básicos de un sistema de procesamiento de imágenes: adquisición, almacenamiento, procesamiento, comunicación y presentación

Las matrices de elementos de estado sólido están compuestos por elementos discretos de silicio, fotodetectores, que tienen un voltaje de salida proporcional a la intensidad luminosa. Existen arreglos lineales de CCD y arreglos matriciales (de área). Un sensor de área puede operar a velocidades muy altas (1/10,000 segundos) y son ideales cuando se necesita congelar el movimiento.

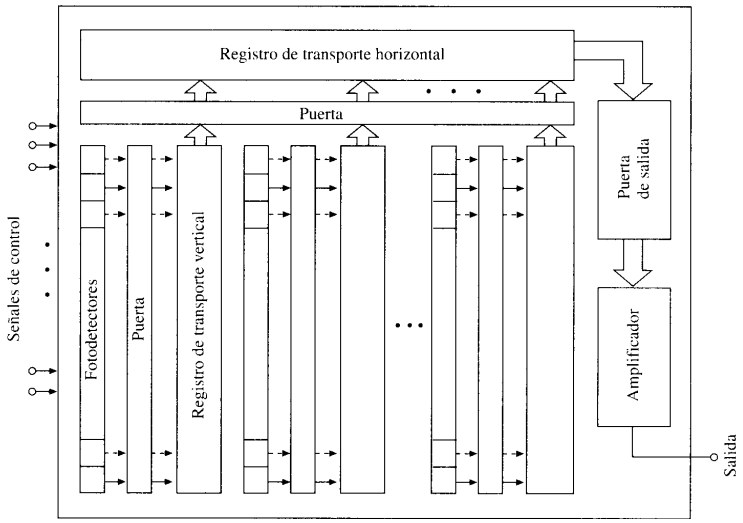
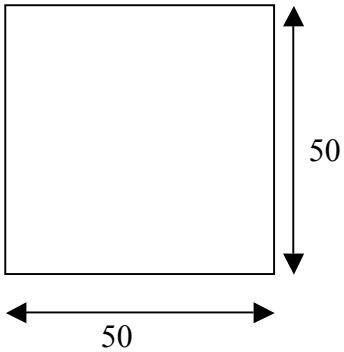


Figura I.3
Arquitectura interna de un sensor CCD de área



Considerando una matriz de sensores de 50X50, el tiempo que toma en leer de un sensor (píxel), considerando una velocidad 30 cuadros/seg, sería:

$$\frac{1/30}{(50)^2} = 13.2\ \mu\text{seg}$$

Los sensores de barrido tienen resoluciones que van desde 256 a 4096

Sensores de área
- 32x32
- 256x256
- 512x512
- 640x480
- 1280x1024

DPI = dots per inch (puntos por pulgada)

iso dpi	⊗	600 dpi óptica
2400 dpi	⊗	19200 dpi digital

A precios mayores existen sensores de gran resolución que usan un movimiento mecánico de un CCD para alcanzar 2048x2048 o 4096x4096. La digitalización de la imagen se logra al enviar la señal de salida de video de estas cámaras a un circuito A/D. Los más comunes son los de 8 bits.

Almacenamiento

Una imagen de 8 bits y 1024x1024 pixels ocupa 1MB de memoria. El almacenamiento en PDI se divide en tres:

- Almacenamiento a corto plazo, para ser empleado durante el procesamiento.
- Almacenamiento en línea, para una utilización relativamente rápida.
- Almacenamiento en archivo, caracterizado por un acceso poco frecuente.

Para el *almacenamiento a corto plazo* se utiliza en gran medida la memoria RAM de la PC. También existen tarjetas especializadas que almacenan una o más imágenes a las que puede accederse con rapidez. Su capacidad de almacenamiento está limitada por el tamaño físico de la tarjeta y la densidad de almacenamiento de los elementos de memoria utilizados. Algunos tipos de tarjetas comerciales son : Memory stick, Compact Flash, Smart Media (8MB, 16MB, 32MB, 64MB, 96MB,128MB, 256 MB, 512 MB), etc.

Para el *almacenamiento en línea* se tiene discos magnéticos, discos duros, discos ópticos, cuando se requiere acceso frecuente a los datos. Existen cajas de discos (jukebox) con capacidad entre 30 y 100 discos ópticos que son RW.

Para el *almacenamiento en archivo* se tienen cintas magnéticas, discos ópticos (cuando se requiere un almacenamiento masivo). Las cintas magnéticas de alta densidad (6400 b/plg) almacenan una imagen de 1 MB en 4 metros. Los discos ópticos pueden almacenar hasta 800 MB de información.

Los discos DVD pueden almacenar hasta 8 GB. Se puede alcanzar una capacidad de 1TB con discos ópticos DVD, utilizando una jukebox; en ella se puede guardar un millón de imágenes de 8 bits de 1024x1024.

Procesamiento

Procedimiento en forma de algoritmo implementados en software. Sólo se emplea hardware cuando se requiere velocidad.
Ejemplo: Microscopía con baja iluminación, se reduce ruido de las imágenes por promediación a la misma velocidad que se adquirieron (30 cuadros / segundo).

Ahora se tienen tarjetas estándar con las conexiones de una computadora. Aunque todavía se siguen empleando grandes sistemas de procesamiento de imágenes para aplicaciones masivas (como el PDI satelital) la tendencia es hacia la miniaturización y adición de hardware de PDI a las PC's de propósito general.

Presentación
- Visualización de los resultados.
- Imagen impresa
- Monitor

UNIDAD II

ADQUISICIÓN Y REPRESENTACIÓN DE IMAGEN.

2.1 Modelo de imagen

Puesto que la luz es una forma de energía se tiene que la función que representa una imagen cumple:

$0<f(x,y)<\infty$; es mayor que cero y finita.

La naturaleza básica de $f(x,y)$ puede estar caracterizada por:

La cantidad de luz incidente procedente de la fuente sobre la escena contemplada.
La cantidad de luz reflejada por los objetos de la escena.

El primero se define por un componente de iluminación $i(x, y)$. El segundo se caracteriza por un componente de reflectancia $r(x,y)$.

De esta forma, una imagen se puede representar por:

$$f(x, y) = i(x, y)\, r(x, y) \qquad [2.1]$$

Donde : $0 < i(x,y) < \infty$;
$0 < r(x,y) < 1$

Por ejemplo: en un día soleado, se tiene un componente de iluminación de ≅9000 candelas/ pie^2. Cuando está nublado ≅1000 cd/ pie^2. En una noche clara, la luna llena produce una iluminación de 0.01 candela/ pie^2.

La iluminación en una oficina típica es ≅100 cd/ pie^2.

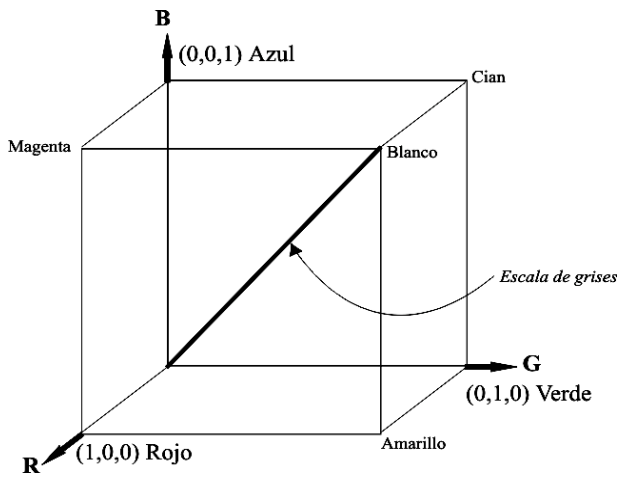
Algunos valores típicos de $r(x, y)$ son:

0.01	Para terciopelo negro
0.65	para acero inoxidable
0.8	muro con pintura blanca
0.9	lámina metálica plateada
0.93	nieve

Para una imagen monocromática se define la función de imagen $f(x,y)$, tomando valores de gris en:

$Lmin \leq l \leq Lmax$ (nivel de gris = *valor de f (x,y)*)

Al intervalo [Lmin, Lmax] se le denomina **escala de grises** y depende de la cantidad de bits con que se representa cada pixel.



cubo para representar los colores básicos de una imagen y sus combinaciones

Muestreo y cuantización.

Para PDI, la imagen se debe digitalizar en espacio y amplitud. La digitalización de las coordenadas espaciales (*x,y*) se denomina muestreo y la de amplitud se conoce como cuantificación del nivel de gris.

Una imagen continua se representa de forma aproximada por una serie de muestras igualmente espaciadas en forma de matriz de NxM.

$$f(x,y) \approx \begin{pmatrix} f(0,0) & f(0,1) \ldots\ldots\ldots\ldots & f(0, M-1) \\ f(1,0) & f(1,1) \ldots\ldots\ldots\ldots & f(1, M-1) \\ \vdots & & \\ f(N-1, 0) & f(N-1, 1) \ldots\ldots\ldots & f(N-1, M-1) \end{pmatrix}$$

En PDI, generalmente, tanto N como M y G (niveles de gris) son potencias de 2.

$$N=2^n; \qquad M=2^m \; ; \qquad G= 2^k$$

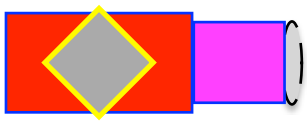
El número de bits necesarios para almacenar una imagen digitalizada es:

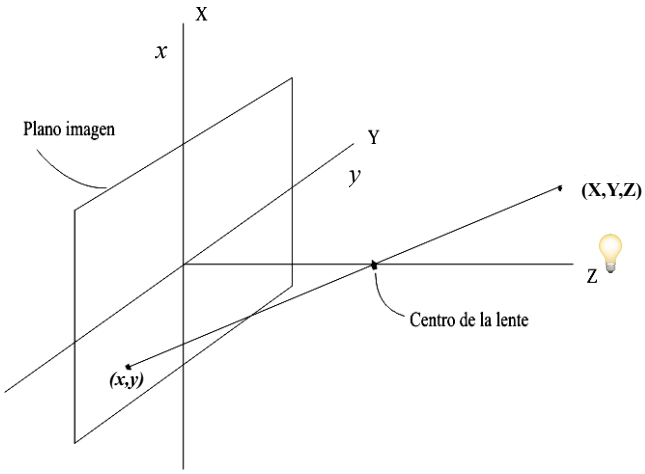
$$b= N \times M \times k \qquad \text{o} \qquad b= N^2 \times k \qquad \text{si } N = M$$

Por ejemplo: una imagen de 256 x 256 con 128 niveles de gris necesita 458, 752 bits o 57,394 bytes.
El grado de detalle (y aproximación) de una imagen digitalizada depende de la cuantización y el muestreo. Cuanto más se aproxime, más se acerca la imagen digitalizada a la imagen original. Pero el almacenamiento y los recursos para el PDI crecen rápidamente.

2.2 Geometría de imagen.

1. La siguiente figura muestra un modelo del proceso de formación de una imagen.





El sistema de coordenadas de la cámara (*x, y, z*) tiene el plano imagen coincidente con el plano XY, y el eje óptico (centro de la lente) coincide con el eje Z. Por tanto, el centro del plano imagen está en (0, 0, 0) y el centro de la lente en (0, 0, λ).

Se supone que el sistema de coordenadas de la cámara está alineado con el sistema de coordenadas global (X, Y, Z).

Sean (X, Y, Z) las coordenadas de cualquier punto en un escenario tridimensional (figura anterior), se considera que Z > λ (todos los puntos de interés quedan por delante de la lente). Primero se obtiene una relación que proporcione las coordenadas (*x,y*) donde se proyecta (X, Y, Z) sobre el plano imagen.

Por semejanza de triángulos se tiene:

$$\frac{x}{\lambda} = \frac{-X}{Z-\lambda} = \frac{X}{\lambda - Z} \qquad \ldots (1)$$

$$\frac{y}{\lambda} = \frac{-Y}{Z-\lambda} = \frac{Y}{\lambda - Z} \qquad \ldots (2)$$

Por lo tanto las coordenadas *x* y *y* del punto 3-D proyectado en el plano imagen son:

$$x = \frac{\lambda X}{\lambda - Z}; \qquad \ldots (3)$$

$$y = \frac{\lambda Y}{\lambda - Z}; \qquad \ldots (4)$$

Estas ecuaciones son no lineales debido a la división por Z.

Se pueden representar estas ecuaciones mediante matrices usando coordenadas homogéneas. Las coordenadas homogéneas de un punto con coordenadas cartesianas (X, Y, Z) se definen como (kX, kY, kZ, k) siendo k una constante arbitraria no nula.

En el sistema de coordenadas cartesianas globales, un punto en forma vectorial se representa por:

$$W = \begin{bmatrix} X \\ Y \\ Z \end{bmatrix} \qquad \ldots (5)$$

Y en coordenadas homogéneas

$$Wh = \begin{bmatrix} kX \\ kY \\ kZ \\ k \end{bmatrix} \qquad \ldots (6)$$

Se define la matriz de **transformación de perspectiva** como:

$$P = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & -1/\lambda & 1 \end{bmatrix} \qquad \ldots (7)$$

El producto $P \bullet W_h$ origina el vector C_h de las coordenadas de la cámara en forma homogénea:

$$Ch = \begin{bmatrix} kX \\ kY \\ kZ \\ -kZ/\lambda + k \end{bmatrix} \qquad \ldots (8)$$

Para convertir a forma cartesiana, se dividen los tres primeros elementos de este vector por el cuarto. Se tiene entonces, el vector de coordenadas cartesianas de la cámara:

$$C = \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} \dfrac{\lambda X}{\lambda - Z} \\ \dfrac{\lambda Y}{\lambda - Z} \\ \dfrac{\lambda Z}{\lambda - Z} \end{bmatrix} \qquad \ldots (9)$$

Los dos primeros elementos del vector C son las coordenadas (*x, y*) en el plano imagen de la proyección de un punto 3-D (X, Y, Z). El tercer elemento no tiene significado en el plano imagen.

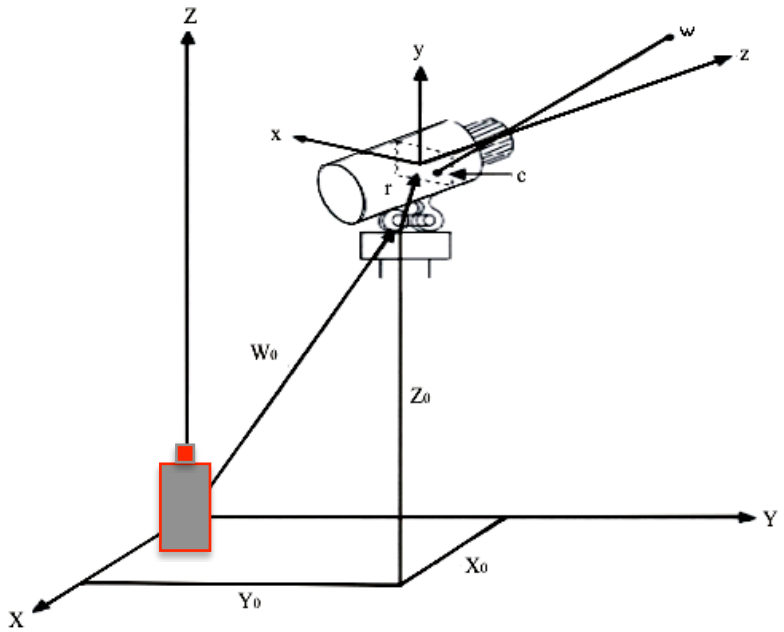
Nota: A menos que se conozca algo más sobre el punto tridimensional que ha generado un punto de la imagen (por ejemplo: su coordenada Z), no es posible recuperar completamente el punto 3-D a partir de su imagen. Se usan las ecuaciones:

$$X = \frac{x}{\lambda}(\lambda - Z)\,; \qquad Y = \frac{y}{\lambda}(\lambda - Z)$$

2.3 Modelo de cámara

Las ecuaciones (8) y (9) describen la formación de una imagen por la proyección de los puntos 3-D sobre un plano imagen. Estas ecuaciones son un modelo matemático básico para una cámara de imagen.

La siguiente figura muestra un sistema de coordenadas global (X,Y,Z) para localizar tanto los puntos en la cámara como los puntos 3-D (indicados por W).



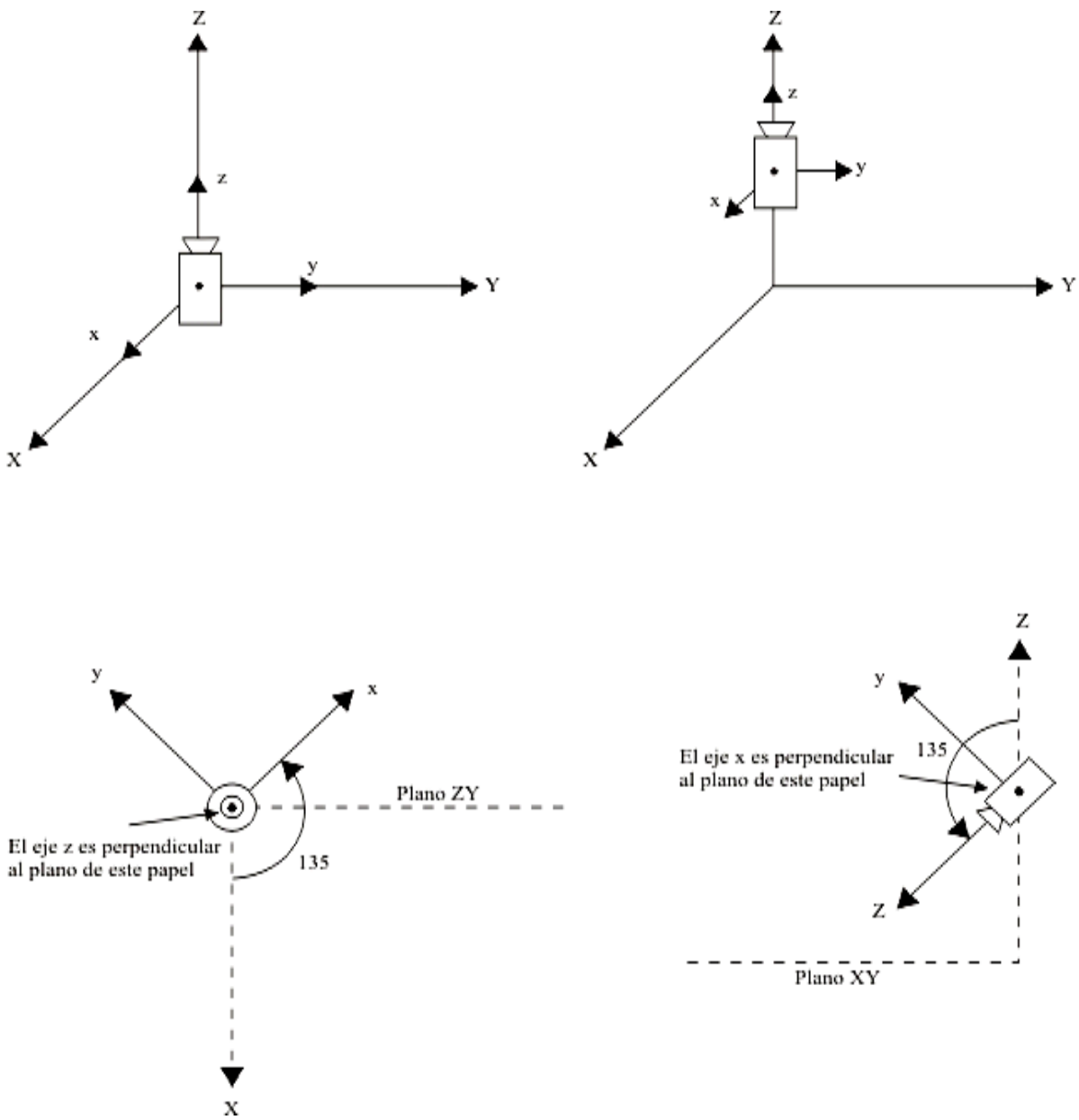
C= punto imagen
W= punto 3-D
(x,y,z)= sistema de coordenadas de la cámara
θ= ángulo de giro en el plano horizontal entre los ejes x y X
α=ángulo de inclinación entre eje z y Z
W_o= separación del centro del trípode al origen del sistema global
r= separación del centro del plano imagen respecto al centro del trípode

Figura 2
Geometría de la imagen con 2 sistemas de coordenadas

Supóngase que, inicialmente, la cámara estaba en la posición "normal" (el centro y el origen del plano imagen coincidía con el origen de coordenadas globales y todos los ejes alineados).

Se supone entonces, que la configuración de la figura 2 se obtiene con estos 4 pasos:

1. Desplazamiento del centro del pivote desde el origen, W_0
2. Giro en horizontal del eje x.
3. Inclinación del eje z.
4. Desplazamiento del plano imagen con respecto al centro del trípode.



La traslación del origen del sistema de coordenadas global a la posición del centro del pivote se consigue aplicando la matriz de transformación:

$$G = \begin{bmatrix} 1 & 0 & 0 & -X_0 \\ 0 & 1 & 0 & -Y_0 \\ 0 & 0 & 1 & -Z_0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \qquad \ldots (10)$$

Es decir, un punto homogéneo W_h que estaba en las coordenadas X_0, Y_0, Z_0, estará en el origen del nuevo sistema de coordenadas después de la transformación G W_h.

Como se dijo, el ángulo horizontal (rotación) se mide entre los ejes *x* y X. En posición normal estos ejes están alineados. La rotación puede ser un ángulo θ con el eje Z como pivote.

La inclinación se mide entre *z* y Z; se inclina la cámara rotando el eje *z* un ángulo α usando X como pivote.

La matriz siguiente se usa para describir estas dos rotaciones:

$$R = \begin{bmatrix} \cos\theta & sen\theta & 0 & 0 \\ -sen\theta\cos\alpha & \cos\theta\cos\alpha & sen\alpha & 0 \\ sen\theta sen\alpha & -\cos\theta sen\alpha & \cos\alpha & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \qquad \ldots (11)$$

Finalmente, el desplazamiento del origen del plano imagen un vector **r** del centro del pivote se consigue mediante la matriz de transformación:

$$C = \begin{bmatrix} 1 & 0 & 0 & -r_1 \\ 0 & 1 & 0 & -r_2 \\ 0 & 0 & 1 & -r_3 \\ 0 & 0 & 0 & 1 \end{bmatrix} \qquad \ldots (12)$$

Así, aplicando a W_h la serie de transformaciones $CRGW_h$, se logra que coincidan los sistemas de coordenadas globales y de la cámara.

Las coordenadas en el plano imagen de un punto W_h se obtienen entonces aplicando la ecuación (8):

$$Ch = PWh$$

Por lo tanto, un punto del espacio 3-D homogéneo visto por una cámara con las coordenadas de la figura 2 tiene las siguientes coordenadas homogéneas en el plano imagen:

$$Ch = PCRGWh \qquad (13)$$

$$Ch = \begin{bmatrix} kX \\ kY \\ kZ \\ k \end{bmatrix}$$

Se obtienen las coordenadas cartesianas (x, y) del punto observado dividiendo el primer y segundo elemento de C_h por el cuarto, se obtienen:

$$x = \lambda \frac{(X - X_0)\cos\theta + (Y - Y_0)sen\theta - r_1}{-(X - X_0)sen\theta sen\alpha + (Y - Y_0)\cos\theta sen\alpha - (Z - Z_0)\cos\alpha + r_3 + \lambda} \qquad \ldots (14)$$

$$y = \lambda \frac{(X - X_0)sen\theta\cos\alpha + (Y - Y_0)\cos\theta\cos\alpha + (Z - Z_0)sen\alpha - r_2}{-(X - X_0)sen\theta sen\alpha + (Y - Y_0)\cos\theta sen\alpha - (Z - Z_0)\cos\alpha + r_3 + \lambda}$$

Son las coordenadas en el plano imagen de un punto 3-D con coordenadas globales (X, Y, Z). Estas ecuaciones se reducen a las ecuaciones (3) y (4) cuando:

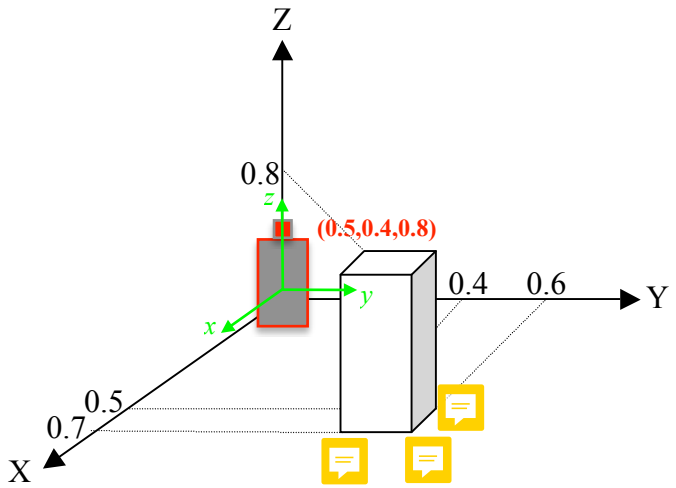
$$X_0 = Y_0 = Z_0 = 0 \;; \qquad r_1 = r_2 = r_3 = 0 \;; \qquad \alpha = \theta = 0^o$$

$$x = \frac{\lambda X}{\lambda - Z}$$

Ejemplo:

Dibuje la forma de la imagen que aparece en el plano de la cámara del objeto que se muestra, considerando los siguientes parámetros para la cámara.

Xo= Yo = 0.5; Zo= 2.5 m
r1= r2 = r3 = 0.01 m
α = 220º , θ = 50º.
λ = 0.035.



Xo=0.5, Yo=1, Zo=4

200o, 30o

UNIDAD III

MEJORAMIENTO DE IMAGEN

3.1 Dominio Espacial

El objetivo principal en el mejoramiento de imagen es procesar una imagen tal que el resultado sea mas conveniente, para una aplicación especifica.
La mayoría de las técnicas de mejoramiento son orientadas al problema.
Ejemplo: Las técnicas para el mejoramiento de imagen en rayos X, difieren de las técnicas aplicadas en el mejoramiento de imágenes de sondas espaciales.
Existen dos categorías de mejoramiento de imágenes:

- En el ***dominio espacial***: que se refieren al plano de la imagen misma, las técnicas se basan en la manipulación directa de pixels en la imagen.
- En el ***dominio frecuencial***: se basan en la modificación de la transformada de Fourier de una imagen. A veces se usan técnicas combinando ambos dominios.

Métodos en el dominio frecuencial

Su fundamento es el Teorema de Convolución. Sea *g(x,y)* una imagen formada por la convolución de una imagen *f(x,y)* y un operador *h(x,y)* lineal e invariante con la posición

$$g(x,y) = h(x,y) * f(x,y)$$

por lo tanto en el dominio de frecuencia:

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

G, *H* y *F* son las transformaciones de Fourier de *g*, *h* y *f*.

En la terminología de la teoría de sistemas lineales, *H(u,v)* se conoce como función de transferencia del proceso.
En una aplicación típica de mejoramiento de imagen, se conoce *f(x,y)* y el objetivo, después de obtener *F(u,v)*, es seleccionar *H(u,v)* tal que la imagen deseada

$$g(x,y) = \Im^{-1}\left[H(u,v)F(u,v)\right]$$

exhiba alguna característica realzada de *f(x,y)*.

En la figura, *h(x,y)* caracteriza un sistema cuyo fin es producir una imagen de salida *g(x, y)* a partir de una imagen de entrada *f(x,y)*.

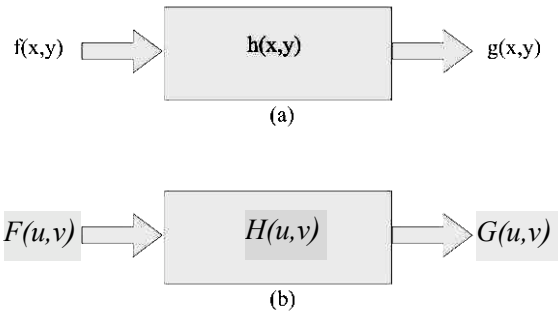


figura
Funcionamiento de un sistema lineal.
En (a), la salida del sistema es la convolución de h(x,y) con la señal de entrada.
En (b), la salida es el producto de H(u,v) por la señal de entrada

Técnicas en el Dominio Espacial:

El término Dominio Espacial, se refiere al conjunto de pixels que componen una imagen, y los métodos en el dominio espacial son procedimientos que operan directamente en éstos pixels.
Una función de procesamiento de imagen en el dominio espacial se puede representar como:

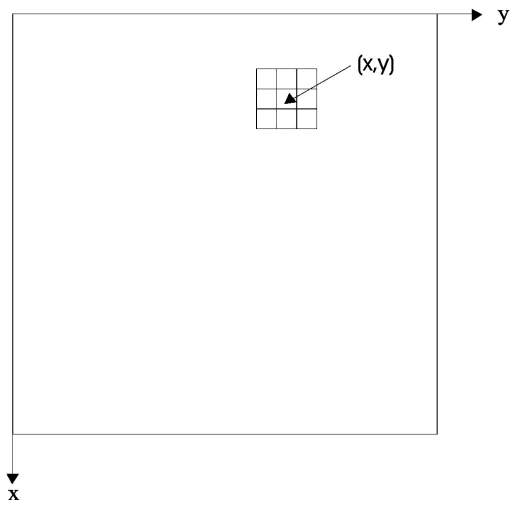
$g(x,y) = T[f(x, y)]$

$f(x,y) =$ Imagen original

$g(x,y) =$ Imagen procesada

T= Es un operador sobre f, definido en una vecindad de (x, y)

Se define una vecindad de (x,y) como un área de sub-imagen cuadrada o rectangular centrada en (x,y) como se muestra en la figura:

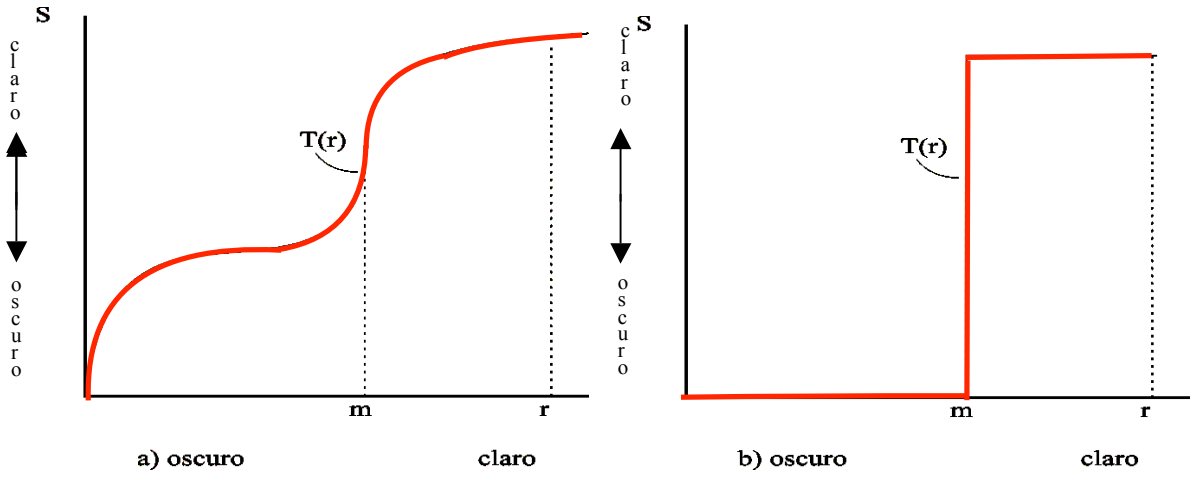


Vecindad de 3x3 alrededor de un punto (x,y) en una imagen.

El centro de la sub-imagen se mueve de pixel a pixel, empezando por ejemplo, en la esquina superior izquierda, y aplicando el operador en cada punto *(x,y)* para obtener *g(x,y)* en ese punto.
La forma más simple del operador *T* es cuando la vecindad es de 1X1; en este caso *g(x,y)* depende solo del valor de *f* en *(x,y)* y *T* es un operador de transformación del nivel de gris, es decir:

$$s = T(r)$$

s y *r* son variables que denotan el nivel de gris de *f* y *g* en un punto *(x,y)*, respectivamente.
Considere, por ejemplo, las siguientes funciones de transformación:



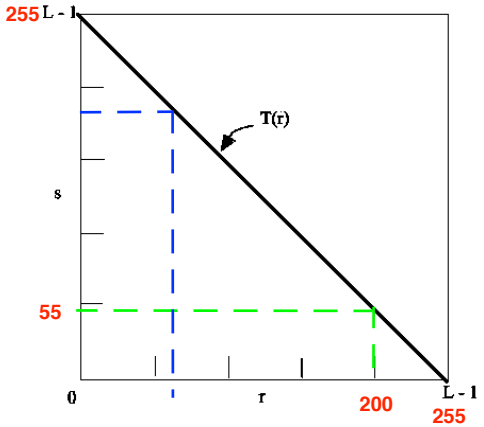
La figura del inciso a) produce una imagen de mayor contraste que la original al oscurecer los niveles de gris bajo **m** y aclarar los niveles arriba de **m** en la imagen original.
El inciso b) presenta el caso límite en que se produce una imagen de dos niveles de gris (binaria).
Sin embargo, el procedimiento general consiste en determinar el valor de *g(x, y)*, mediante los valores de *f(x, y)* en una vecindad definida de *(x, y)*. Estas vecindades alrededor de *(x, y)* (incluyendo el mismo punto) se conocen como máscaras (o también plantillas, ventanas o filtros).

Máscaras: Arreglo 2-D pequeño en el cual los valores de los coeficientes determinan la naturaleza del proceso.

3.2 TRANSFORMACIONES DE INTENSIDAD SIMPLE

Negativo de una imagen

Comúnmente se realiza esta transformación para realzar algunas características de imágenes médicas.
El negativo de una imagen se obtiene usando la función de transformación $s = T(r)$ de la siguiente figura, donde L es el número de niveles de gris. La idea es invertir el orden de negro a blanco tal que la intensidad de la imagen de salida, decrece cuando la intensidad de la imagen de entrada aumenta.



y = m x + b

s = T(r) = m r + b = 255 - r

m = -1
b = 255

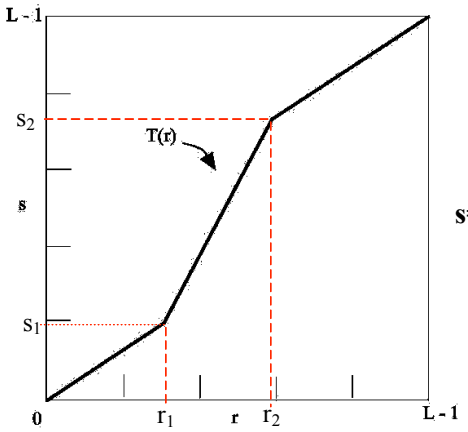
Obtención del negativo de una imagen, función de transformación de los niveles de gris

Ensanchamiento de contraste

Las imágenes con poco contraste pueden ser debidas a diversas causas, como iluminación deficiente, falta de rango dinámico en el sensor o incluso incorrecta selección de la apertura de la lente durante la captación de la imagen. La técnica de ensanchamiento de contraste consiste precisamente en aumentar el rango dinámico de los niveles de gris de la imagen; tiene una función de la forma:

r1=s1
r2=s2
r2>r1

r1=r2
s1=0
s2=255



s=T(r)=
- **m1*r 0<=r<r1**
- **m2*r+b2 r1<=r<r2**
- **m3*r+b3 r2<=r<=255**

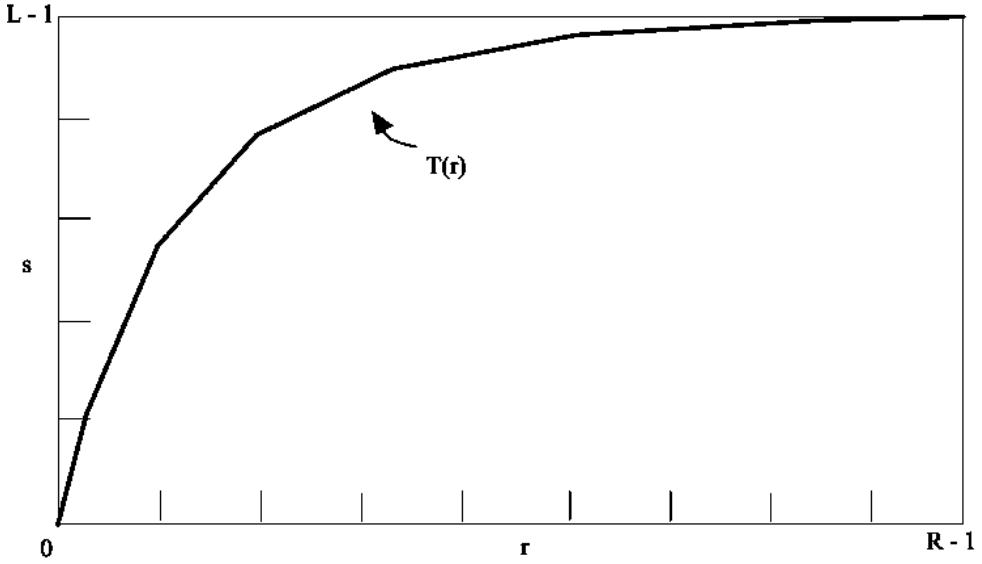
m2 = (s2-s1)/(r2-r1)

Aumento de contraste, perfil de la función de transformación

Compresión del intervalo dinámico

Algunas veces el intervalo dinámico de una imagen procesada excede por mucho la capacidad del dispositivo de despliegue, en cuyo caso solo las partes mas brillantes de la imagen son visibles en la pantalla.
Una forma efectiva para comprimir el intervalo dinámico de los valores de pixels es efectuar la siguiente transformación de intensidad

$$s = C \cdot \log(1 + |r|) \; ; \qquad C = \text{cte. de escala}$$



Compresión del rango dinámico, función de transformación logarítmica

Ejemplo:

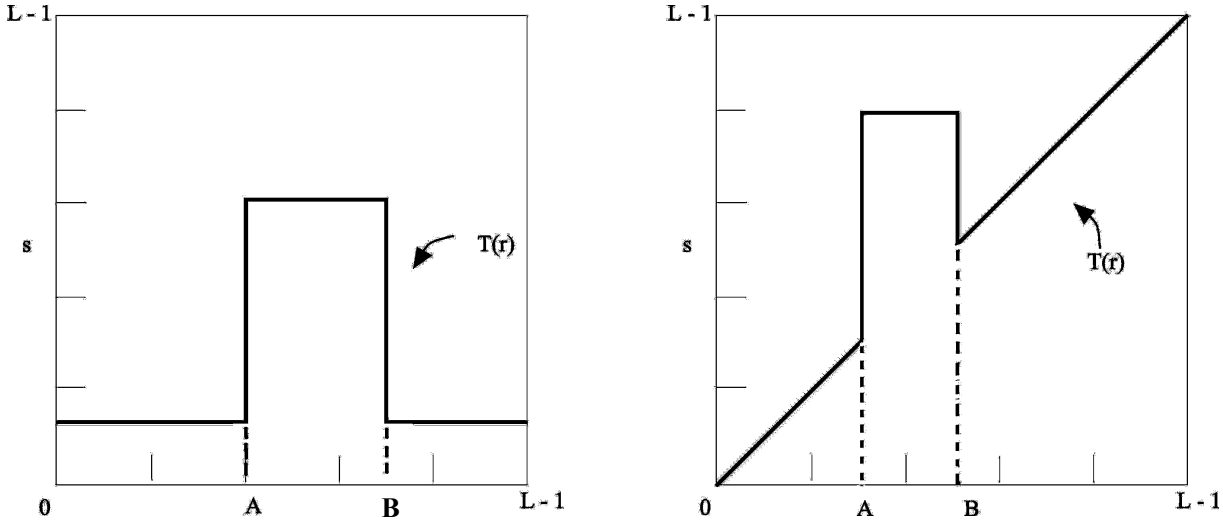
Una imagen que tiene un intervalo dinámico (después del procesamiento) de [0, R] = [0, 2.5 x 10^6] no podría desplegarse en un sistema 8 bits por tanto se debe comprimir su intervalo dinámico.

Cuando se aplica log (1+ $|r|$), el intervalo va de 0 a 6.4; para acomodarla aprovechando los 8 bits del sistema

$$C = 255/6.4$$

Rebanado [seccionado] de niveles de gris

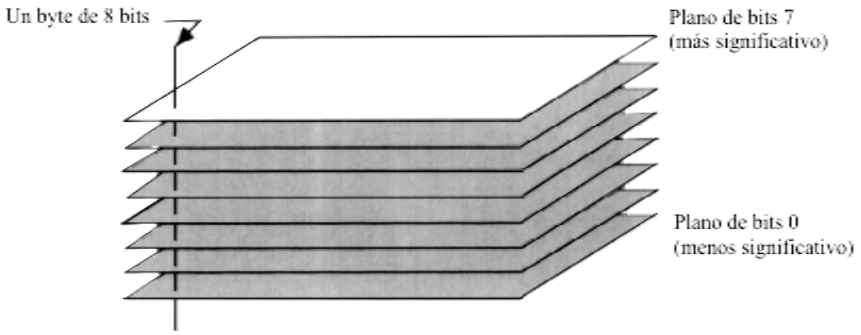
Comúnmente se desea resaltar un intervalo especifico de niveles de gris en una imagen. Algunas aplicaciones de este algoritmo son resaltar masas de agua en imágenes satelitales o realzar fracturas en imágenes de rayos X. Hay varias formas para lograr el rebanado de niveles, pero la mayoría de ellas son variaciones de dos formas básicas. Una es desplegar un valor alto para todos los niveles de gris en el intervalo de interés y un valor bajo para todos los demás; esta transformación produce una imagen binaria. La segunda forma básica, basada en la transformación mostrada en el inciso b) presenta un mayor brillo para el intervalo deseado de niveles pero preserva el fondo y las tonalidades de gris en la imagen.



Aquí tenemos dos funciones de transformación para resaltar el rango de intensidad [A,B].
La primera resalta el rango [A,B] reduciendo las restantes a una constante de tonalidad oscura.
La segunda igualmente resalta el rango [A,B], pero preservando la intensidad de lo demás.

Rebanado de planos de Bits

En lugar de destacar ciertos rangos de intensidades, a veces puede desearse destacar la contribución que realizan a la imagen determinados bits específicos. Supongamos que cada píxel de una imagen viene representado por 8 bits. Imaginemos también que la imagen está compuesta por ocho planos de un bit, que van desde el plano 0 para el bit menos significativo hasta el plano 7 para el bit más significativo. En términos de bytes de 8 bits, el plano 0 contiene todos los bits de orden inferior de los bytes que forman los pixels de la imagen, mientras que el plano 7 contiene todos los bits de orden más superior.



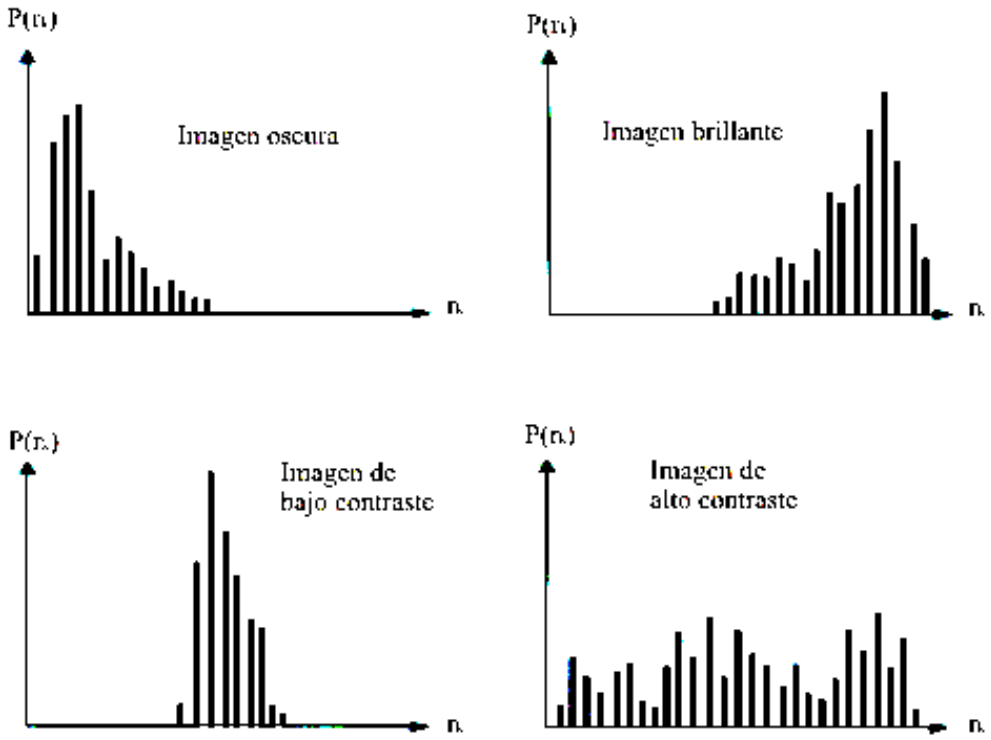
Procesamiento de histograma

El histograma de una imagen digital con niveles de gris en [0, L-1] es una función discreta de la forma:

$$p(r_k) = n_k / n$$

r_k = k_ésimo nivel de gris
n_k = número de pixels en la imagen que tiene ese nivel de gris
n = número total de pixels en la imagen

A grandes rasgos $p(r_k)$ da una estimación de la probabilidad de ocurrencia del nivel de gris r_k. La gráfica de esta función para todos los valores de k proporciona una descripción global de la apariencia de una imagen.



La forma del histograma de una imagen da información útil acerca de posibilidad de mejoramiento de contraste.

Ecualización de Histograma

Una gráfica de p (r_k) vs. r_k se conoce como histograma, y la técnica usada para obtener un histograma uniforme se conoce como <u>ecualización del histograma o linealización del histograma</u>.

La ecualización del histograma implica un incremento en el intervalo dinámico de los pixels lo cual tiene un efecto considerable en la apariencia de una imagen.

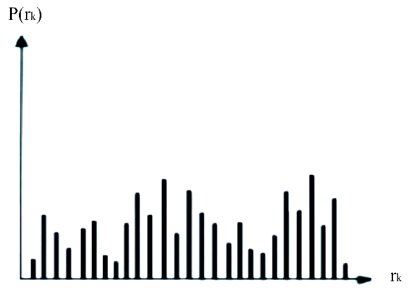
La forma discreta para ecualizar un Histograma es:

$$S_k = T(r_k) = \left(\frac{1}{n} \sum_{j=0}^{k} n_j \right) \cdot 255$$

$$k = 0,1,\ldots,L-1$$

$$0 \le r_k \le 255$$

$$0 \le Sk \le 255$$



Mejoramiento Local

La ecualización del histograma presentada es un algoritmo global, puesto que los pixels son modificados mediante una transformación basada en la distribución del nivel de gris sobre toda la imagen.
Algunas veces es necesario realzar detalles sobre áreas pequeñas. Se puede aplicar la ecualización de histograma de forma local en la vecindad de cada pixel en la imagen.
El procedimiento consiste en definir una vecindad cuadrada o rectangular y mover el centro de esta área de pixel a pixel. En cada posición, se calcula el histograma de los puntos en la vecindad y se obtiene una función de ecualización de histograma.

Esta función se usa para mapear el nivel de gris del centro de la vecindad. Este centro se mueve entonces aun pixel adyacente y se repite el procedimiento.

$$S_k = T(r_k) = \frac{1}{n}\sum_{j=0}^{k} n_j$$

Promediado de Imágenes

Considerar una imagen ruidosa g (x, y) formada por la adición de ruido η (x, y) a una imagen original, es decir:

$$g(x,y) = f(x,y) + \eta(x,y) \qquad \ldots (1)$$

donde se supone que en cada punto (x, y) el ruido no esta correlacionado y tiene un valor promedio de cero. El objetivo es reducir los efectos de ruido al sumar una serie de imágenes ruidosas, {g (x, y)}

Si el ruido satisface esos condiciones, una imagen g (x, y) formada por el promedio de M imágenes ruidosas diferentes:

$$\bar{g}(x,y) = \frac{1}{M}\sum_{i=1}^{M} g_i(x,y) \qquad \ldots (2)$$

origina que:

$$E\{\bar{g}(x,y)\} = f(x,y) \qquad \ldots (3)$$

y

$$\sigma^2_{\bar{g}(x,y)} = \frac{1}{M}\sigma^2_{\eta(x,y)} \qquad \ldots (4)$$

E { $\bar{g}$(x, y) } es el valor esperado de g (x, y); $\sigma^2_{\bar{g}(x,y)}$ y $\sigma^2_{\eta(x,y)}$ son las varianzas de g y η en (x, y). La desviación estándar en cualquier punto de la imagen promedio es:

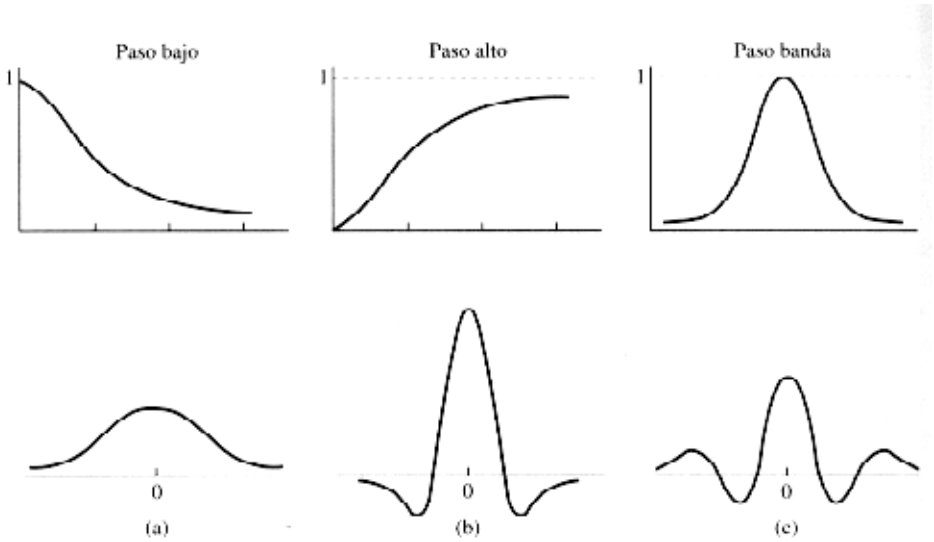
$$\sigma_{\bar{g}(x,y)} = \frac{1}{\sqrt{M}}\sigma_{\eta(x,y)} \qquad \ldots (5)$$

Las ecuaciones (4) y (5) indican que, a medida que *M* crece, la variación de valores de los pixels en cada punto *(x,y)* decrece. Esto significa que *g(x, y)* se aproxima a *f(x, y)* a medida que el número de imágenes ruidosas usadas en el proceso de promediado crece.

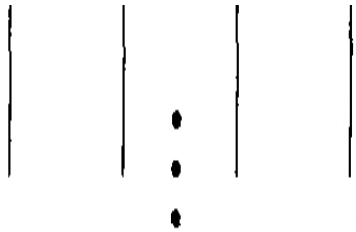
Filtrado espacial

El uso de máscaras espaciales para el procesamiento de imágenes se conoce usualmente como filtrado espacial y las máscaras como filtros espaciales. Los filtros pasa bajos atenúan o eliminan componentes de alta frecuencia (bordes, contornos, etc.), mientras que dejan intactos los elementos de bajas frecuencias (cambios suaves de gris, fondo, etc). Los componentes de altas frecuencias caracterizan bordes y otros detalles agudos en una imagen, tal que el efecto neto de un filtro pasa bajos es hacer borrosa la imagen. Similarmente, los filtros pasa altos atenúan o eliminan los componentes de baja frecuencia, responsables de la variación gradual o algunas características tal como el contraste global e intensidad promedio; el resultado neto es una reducción de estas características y un aparente realce de bordes y otros detalles finos.

La siguiente figura muestra las secciones transversales de los filtros pasa bajos, pasa altos y pasa bandas, en el dominio de la frecuencia y en el dominio espacial.



Independientemente del filtro usado, el procedimiento consiste en sumar los productos entre los coeficientes de la máscara del filtro y las intensidades de los pixels bajo la máscara en un punto específico de la imagen. La siguiente figura muestra una máscara general de 3x3.



Denotando los niveles de gris de los pixels (de la figura) bajo la máscara $z_1, z_2, \ldots, z_9$, la respuesta de una máscara lineal es:

$$R = w_1 z_1 + w_2 z_2 + \ldots + w_9 z_9$$

Filtros pasa bajos: Se usan para reducir ruidos y para puentear pequeños huecos en líneas o curvas.

En la siguiente figura se muestra una máscara de 3x3 (orden *n*=3) para implementar un filtraje pasa bajos.

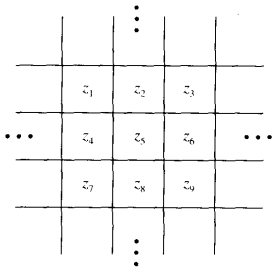
$\frac{1}{9} \times$

1	**1**	**1**
1	**1**	**1**
1	**1**	**1**

Filtro de mediana

Es un tipo de filtro que se utiliza para eliminar ruido de imágenes. A diferencia del promediado de imágenes utilizando también para eliminar el ruido, este tipo de filtro utiliza máscaras cuadradas o rectangulares de ordenamiento y no requiere tener varias imágenes ruidosas.

Considere la siguiente imagen con una máscara del filtro de mediana de 3x3



Supongamos que los niveles de gris bajo la máscara son como se muestra.

100	150	80
10	180	100
80	25	50

Para obtener el valor del pixel z_5 de la imagen filtrada, se ordenan los valores de gris de la imagen incluídos por la máscara en orden ascendente o descendente y se selecciona el 5º. nivel de gris como el nuevo valor para el pixel donde se encuentra centrada la máscara.
En este caso:

10, 25, 50, 80, 80, 100, 100, 150, 180
valor de gris para la imagen filtrada en la posición de z_5

Filtros agudizantes:

El objetivo principal de un agudizado es resaltar detalles finos en una imagen o realzar detalles que han sido borrados ya sea por error o como efecto natural de un método particular de adquisición de imagen.

Filtro espacial pasa altos básico:
se usa una máscara de la forma:

$\frac{1}{9} \times$

-1	**-1**	**-1**
-1	**-8**	**-1**
-1	**-1**	**-1**

filtro pasa altas

Filtro High-boost (alto realce)

Una imagen filtrada con un filtro PA se puede obtener como la diferencia entre la imagen original y la versión filtrada PB:

PA= original – PB … (1)

Multiplicando la imagen original por un factor de amplificación, denotado por A, se obtiene lo que se conoce como filtro high-boost o de énfasis de las frecuencias altas. Es decir:

HB= A (original) – PB
= (A- 1) original + original – PB
= (A- 1) original + PA

Si A = 1, se tiene un filtro PA estándar.

Cuando A > 1, parte de la imagen original se agrega nuevamente al resultado del filtro PA, lo cual restaura parcialmente los componentes de baja frecuencia perdidos en la operación de filtrado PA.

El resultado es que la imagen se parece más a la original, con un grado relativo de realce de bordes que dependen del valor de A para implementarlo, se puede usar una máscara como:

$$\frac{1}{9} \times \begin{bmatrix} -1 & -1 & -1 \\ -1 & w & -1 \\ -1 & -1 & -1 \end{bmatrix}$$

donde: $w = 9A - 1$ con $A \geq 1$

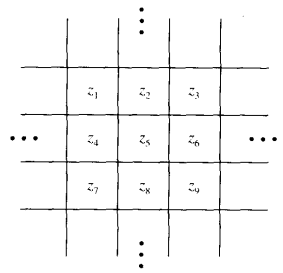
Filtros derivativos

El promedio de pixels de cierta región en una imagen, tiende a ser borrosa la imagen. Como el promedio es análogo a la integración, se espera que la diferenciación tenga el efecto opuesto y entonces agudizar una imagen.

La magnitud de este vector:

$$|\Delta f| = \left[\left(\frac{df}{dx} \right)^2 + \left(\frac{df}{dy} \right)^2 \right]^{1/2} \qquad \ldots (1)$$

Es la base de varios métodos de aproximación o diferenciación. Considere la región de una imagen de la figura donde las Z's denotan los valores de gris.



La ecuación (1) se puede aproximar en el punto Z_2 de diversas formas.

La más simple consiste en usar la diferencia ($z_5 - z_8$) en la dirección (x) y ($z_5 - z_6$) en (y), combinadas como:

$$|\Delta f| = \left[(z_5 - z_8)^2 + (z_5 - z_6)^2 \right]^{1/2} \qquad \ldots (2)$$

Se puede obtener resultados similares como:

$$|\Delta f| = \left[|z_5 - z_8| + |z_5 - z_6| \right] \qquad \ldots (3)$$

La ecuación anterior se puede implementar usando máscaras de 2x2

1	0
-1	0

1	-1
0	0

Operadores de Roberts

Se pueden usar máscaras de 3x3 para calcular para calcular Z_5 cumpliendo con la ecuación:

$$|\Delta f| = \left[(z_7 + z_8 + z_9) - (z_1 + z_2 + z_3) \right] + \left[(z_3 + z_6 + z_9) - (z_1 + z_4 + z_7) \right]$$

Las máscaras utilizadas son:

-1	-1	-1
0	0	0
1	1	1

-1	0	1
-1	0	1
-1	0	1

Operadores de Prewitt

Otras máscaras:

-1	-2	-1
0	0	0
1	2	1

-1	0	1
-2	0	2
-1	0	1

Operadores de Sobel

Segmentación de Imagen

Hasta ahora se han desarrollado algoritmos que se clasifican dentro de los métodos de pre-procesamiento. Empezaremos a tratar los métodos de análisis de imágenes.

El primer paso en el análisis de imágenes generalmente consiste en segmentar las imagen. La segmentación divide una imagen en sus partes constitutivas u objetos. El proceso de segmentación termina cuando se han aislado los objetos de interés en una imagen dada (es dependiente del problema).

Por ejemplo, cuando se desea detectar un objetivo a partir de imágenes aéreas, en el caso particular de vehículos en un camino.

Nota: En general, la segmentación autónoma es de las tareas más difíciles en PDI, este paso de procesamiento determina el éxito eventual del análisis.

Los algoritmos de segmentación para imágenes monocromáticas generalmente se usan en una de dos propiedades básicas de los valores de gris: discontinuidad y similitud.

Discontinuidad: El mecanismo es particionar una imagen basándose en los cambios considerables de nivel de gris.

Las principales áreas de interés son: Detección de puntos aislados y detección de líneas y bordes en una región de la imagen.

Similitud: Se basa en el umbralizado, crecimiento de región y en la división y unión de regiones.

Hay 3 tipos básicos de discontinuidad: puntos, líneas y bordes.

Detección de puntos: La detección de puntos aislados en una imagen es un proceso directo usando la mascara:

-1	**-1**	**-1**
-1	**-8**	**-1**
-1	**-1**	**-1**

Se puede decir que se ha detectado un punto en la posición donde se centra la máscara si: /R/=T

Donde:

$$R = w_1 z_1 \ . \ . \ . \ w_i z_n \qquad \ldots \text{(I)}$$

El método se basa en la idea de que el nivel de gris de un punto aislado será muy diferente de los niveles de gris de sus vecinos. Solo diferencias que son lo bastante grandes (dependiendo del valor de umbral: T) para ser considerados puntos aislados, son de interés.

Detección de líneas

El siguiente nivel de complejidad involucra la detección de líneas en una imagen. Considera las máscaras de la siguiente figura:

-1	-1	-1
2	2	2
-1	-1	-1

Horizontal

-1	-1	2
-1	2	-1
2	-1	-1

$+45^0$

-1	2	-1
-1	2	-1
-1	2	-1

Vertical

2	-1	-1
-1	2	-1
-1	-1	2

-45^0

Si la primera máscara se moviera por toda una imagen, respondería más fuertemente a líneas orientadas horizontalmente (con un pixel de espesor)
La segunda máscara, respondería más fuertemente a líneas orientadas a 45º (con un pixel de espesor)
sean R_1,R_2,R_3, y R_4 las respuestas de las máscaras de izquierda a derecha, donde las R's están dadas por:

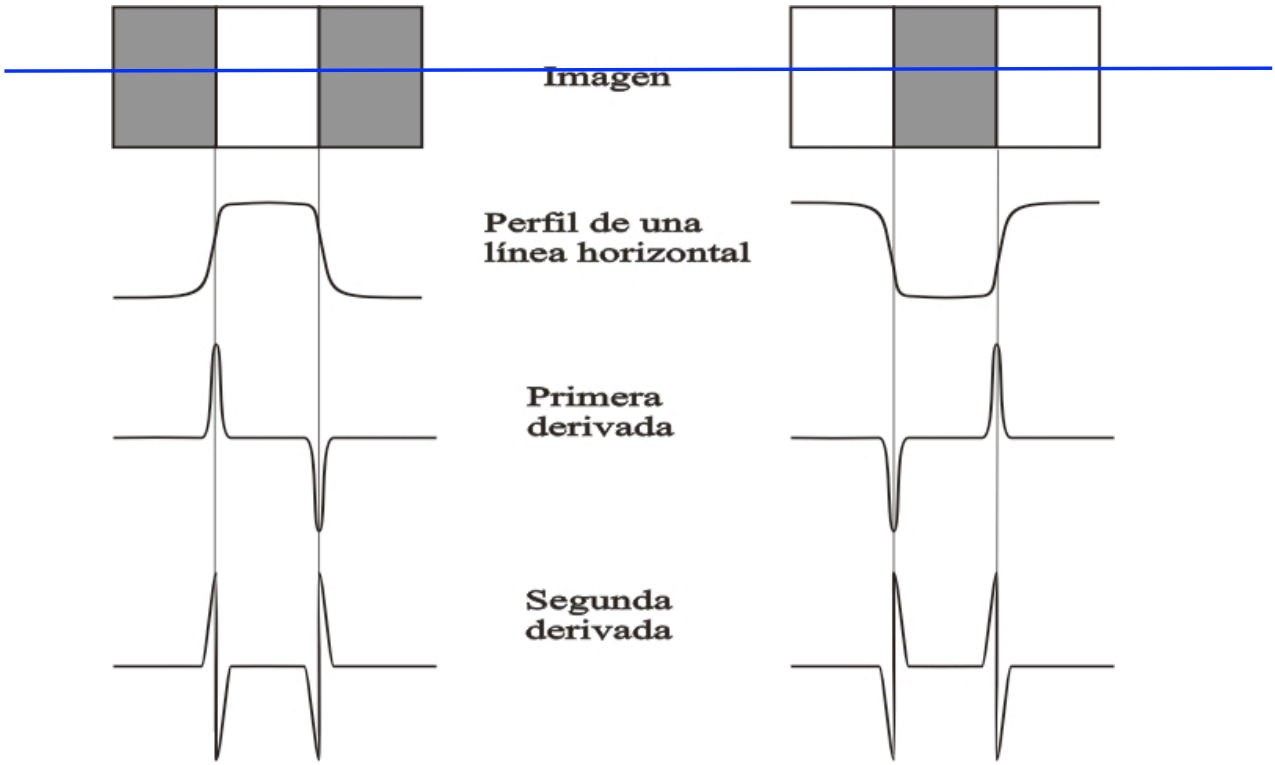
$$R = w_1 z_1 + w_2 z_2 + \ . \ . \ . \ + w_i z_n$$

Suponer que todas las máscaras recorren la imagen. Si en cierto punto de la imagen, $|R_i| > |R_j|$, está para toda $i \neq j$, se dice que ese punto está más asociado con una línea en la dirección de la máscara i. Por ejemplo, si en un punto de la imagen $|R_i| > |R_j|, j = 2,3,4$. Se dice que ese punto particular está más probablemente asociado con una línea horizontal.

Detección de bordes

Un borde es la frontera entre dos regiones con propiedades de nivel de gris relativamente distintas. Suponemos que las regiones son suficientemente homogéneas, tal que la transición entre dos regiones se puede determinar solo basándose en las discontinuidades del nivel de gris. Básicamente, la mayoría de las técnicas de detección de bordes se basan en el cálculo de un operador derivatorio local.

Considere el inciso a) de la siguiente figura que muestra la imagen de una cinta clara en un fondo oscuro, su perfil a lo largo de una línea horizontal de muestreo y la primera y segunda derivada de ese perfil.



Nótese del perfil que un borde (transición de oscuro a claro) se modela como un cambio de gris suave, en lugar de uno abrupto. Este modelo refleja el hecho de que los bordes en imágenes digitales generalmente se hacen borrosos como resultado del muestreo.

Observaciones: la primera derivada es cero en áreas de nivele de gris constante. La segunda derivada es positiva para la parte de la transición asociada con el lado oscuro del borde, y es negativa para la parte de la transición del lado claro.
Así, la magnitud de la primera derivada se puede usar para detectar la presencia del borde en una imagen y el signo de la segunda derivada se puede usar para determinar si un pixel de borde cae del lado oscuro o del lado claro. Nótese que la segunda derivada tiene un cruce por cero en el punto medio de la transición; este cruce por cero es una herramienta poderosa para localizar bordes en una imagen.

Operadores de gradiente

El gradiente de una imagen f(x,y) en (x,y) es el vector:

$$\nabla f = \begin{bmatrix} Gx \\ Gy \end{bmatrix} = \begin{bmatrix} \dfrac{\partial f}{\partial x} \\ \\ \dfrac{\partial f}{\partial y} \end{bmatrix} \qquad \ldots (A)$$

En la detección de bordes, la cantidad importante es la magnitud de ese vector.

$$\nabla f = mag(\nabla f) = \left[Gx^2 + Gy^2\right]^{1/2} \qquad \ldots (B)$$

Una práctica común es aproximar el gradiente con valores absolutos.

$$\nabla f = |Gx| + |Gy| \qquad \ldots (C)$$

Como se dijo antes las derivadas pueden implementarse en forma digital de varias formas. Sin embargo los operadores de **Sobel** tienen la ventaja de proporcionar dos efectos convenientes:
Diferenciación y suavizado debido a que los filtros derivativos enfatizan el ruido el efecto de suavizado es una caracteristica particularmente atractiva de los operadores **Sobel**

-1	-2	-1
0	0	0
1	2	1

-1	0	1
-2	0	2
-1	0	1

Operadores de Sobel

Aplicar a una imagen los operadores Sobel:

1. Primero solo el operador (1) para resaltar solo los bordes con dirección horizontal
2. Después el operador (2) para resaltar los bordes verticales
3. Los dos para detectar los bordes con cualquier orientación.

Laplaciano

El laplaciano de una función 2-D, f(x,y), es una derivada de 2º orden definida como:

$$\nabla^2 f = \frac{\partial^2 f}{\partial x^2} + \frac{\partial^2 f}{\partial y^2} \qquad \ldots (1)$$

Como en el caso del gradiente, se puede implementar en forma digital de varias formas. Para una región de 3x3, la forma más comúnmente usada es:

$$\nabla^2 f = 4zs - (z_2 + z_4 + z_6 + z_8) \qquad \ldots(2)$$

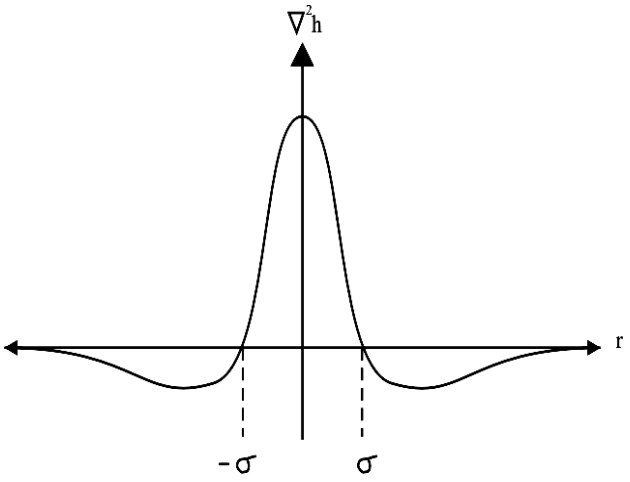
Aunque como se indicó antes, el laplaciano responde a transiciones en intensidad, rara vez en la práctica se utiliza para detectar bordes debido a varias razones. Como una derivada de 2º orden, el laplaciano típicamente es muy sensible al ruido. Más aún, el laplaciano produce bordes dobles y es incapaz de detectar la dirección del borde.
El uso más general del laplaciano es determinar la localización de bordes usando su propiedad de cruce por cero. Este concepto se basa en la convolución de una imagen con el laplaciano de una función 2-D estándar de una función gaussiana de la forma:

$$h(x,y) = -\exp\left(-\frac{x^2 + y^2}{2\sigma^2}\right) \qquad \ldots(3)$$

Haciendo $r^2 = x^2 + y^2$, entonces el laplaciano de la ecuación 3 (la 2ª derivada respecto a r es):

$$\nabla^2 h = -\left(\frac{r^2 - \sigma^2}{\sigma^4}\right)\exp\left(-\frac{r^2}{2\sigma^2}\right) \qquad \ldots(4)$$

La siguiente figura muestra una sección transversal de esta función:



nótese la suavidad de la función, su cruce por cero en $r = \pm\sigma$, y el centro (+) y sus alrededores (-)

1. aplicar la máscara obtenida con $\nabla^2 h$ y el valor de σ
2. se tiene una imagen con valores de gris (+) y (-) se escala al intervalo dinámico 0-255 (0-1), se presenta la imagen
3. obtener una imagen binaria:

- los valores (+) de la imagen anterior se hacen 255
- los valores (-) y 0 de la imagen anterior se hacen cero

4. determinar una imagen donde aparezcan los bordes que delimitan las regiones en la imagen

Enlace de bordes y detección de fronteras

Las técnicas presentadas hasta ahora detectan discontinuidades en intensidades de gris, idealmente, estas técnicas deberían producir pixels que caen solo en la frontera entre regiones. En la práctica, este conjunto de pixels rara vez caracteriza una frontera completamente, debido al ruido, rupturas en la frontera por una iluminación no uniforme, y otros efectos que hacen que las discontinuidades sean confusas. Considerando esto, los algoritmos de detección de bordes típicamente son seguidos por un enlazado de bordes y otros procedimientos de detección de fronteras diseñados para acomodar los pixels de bordes dentro de fronteras significativas.

Procesamiento local

Uno de los métodos más simples para enlazar puntos de un borde consiste en analizar las características de los pixels en una pequeña vecindad (3x3 ó 5x5) alrededor de cada punto (x,y) en una imagen sometida a detección de bordes. Todos los puntos similares se enlazan formando una frontera de pixels que comparten unas propiedades comunes. En este tipo de análisis se usan principalmente 2 propiedades para establecer la similitud de pixels de borde:

La intensidad de respuesta del operador gradiente utilizado para producir el pixel de borde
La dirección del gradiente

La primera propiedad está dada por el valor ECUACIÓN definido con las ecuaciones B o C anteriores. Así, un pixel de borde con coordenadas x',y' en la vecindad predefinida de (x,y), es similar en magnitud al pixel en (x,y) si:

$$\left|\nabla f(x,y) - \nabla f(x,y)\right| \leq T \qquad \ldots \text{(I)}$$

$$T = \text{ valor de umbral } \geq 0$$

La dirección del vector gradiente está dada por:

$$\alpha(x,y) = \tan^{-1}\left(\frac{Gy}{Gx}\right) \qquad \ldots \text{(II)}$$

el ángulo se mide respecto al eje x

Entonces un pixel de borde (x',y') en la vecindad predefinida de (x,y) tiene un ángulo similar al pixel en (x,y) si:

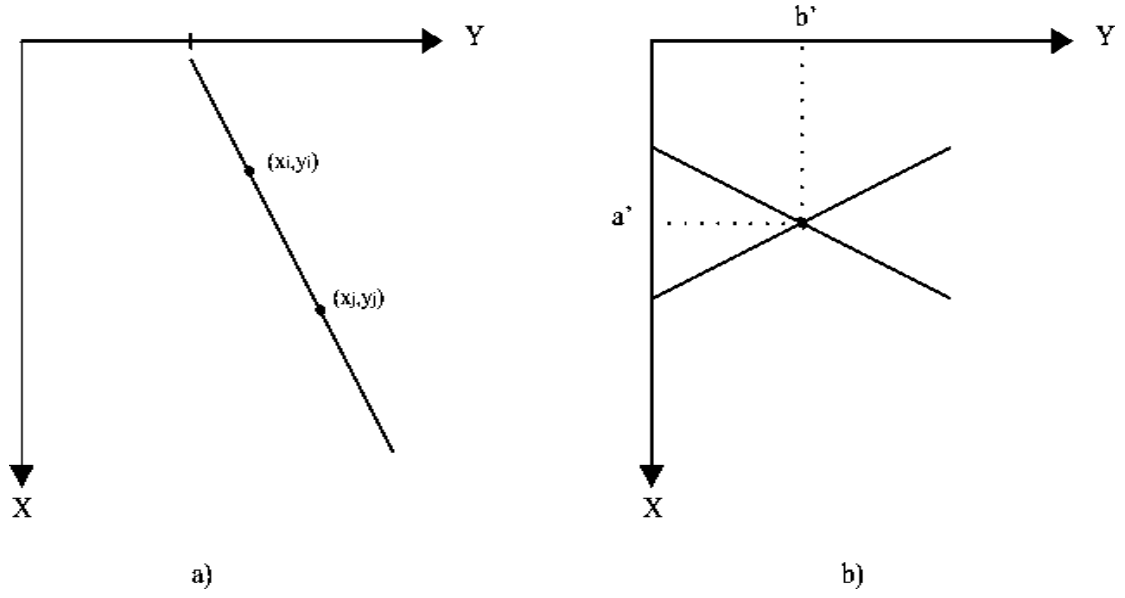
$$\left|\alpha(x,y) - \alpha(x',y')\right| < A \qquad \ldots \text{(III)}$$

Donde *A* es un ángulo de umbral.
Nótese que la dirección del vector gradiente es perpendicular al borde en el punto (x,y).
Un punto en la vecindad predefinida de (x,y) se enlaza al pixel en (x,y) si se satisfacen ambos criterios: de magnitud y dirección. Este procedimiento se repite en cada punto de la imagen. Se asigna un nivel de gris diferente a cada conjunto de pixels de borde enlazados, y generalmente se obtiene una imagen binaria.

PROCESAMIENTO GLOBAL

Transformada de Hough

Se considera que se desean detectar bordes con tendencia lineal.



(a) Plano XY; (b) espacio parámetro

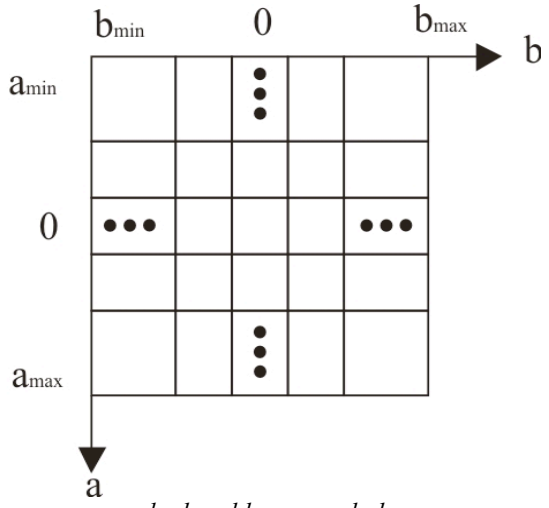
Considerando la ecuación de una recta:

$$y = ax + b$$

Donde:
a = pendiente
b = intersección con eje *y*

El método consiste en trabajar las líneas rectas que se encuentran en una imagen en otro plano auxiliar determinado por *a* y *b*.

Se debe llenar un arreglo de celdas acumuladores que determinan la cantidad de líneas que se encuentran en la imagen y sus propiedades.



arreglo de celdas acumuladores

Segmentación por regiones

Considere que R representa la región de una imagen completa. Se puede entender la segmentación como el proceso de particionar R en n regiones: R_1, R_2,....R_n, tal que:

a) $\bigcup_{i=1}^{n} Ri = R$,
b) Ri es una región conexa, para toda i= 1,2,...,n (compuesta de puntos conectados).
c) Ri $\cap$ Rj = $\varnothing$ (conjunto vacío) para toda i y j, con i $\neq$j
d) P(Ri)= VERDADERO para i=1, 2,...,n y
e) P(Ri $\cup$ Rj)= FALSO para i$\neq$ j

Donde P(Ri)es un predicado sobre los puntos en Ri.

El inciso d) tiene que ver con las propiedades que debe cumplir un pixel para que se incluya en la región Ri. (ejemplo P(Ri)= verdadero si todos los píxeles en Ri tienen la misma intensidad).

El inciso e) indica que las regiones Ri y Rj son diferentes en el sentido del predicado P.

Crecimiento de región (por agregación de píxel)

Es un procedimiento que agrupa píxels o subregiones en regiones mayores. El proceso más simple es la agregación de pixels; empieza con un conjunto de pixels "semilla" y de éstos crecen regiones al agregar a cada punto semilla aquellos pixels vecinos que tienen propiedades similares (nivel de gris, textura, color, etc.).

Para ilustrar este procedimiento, considere la sig. fig., donde los números dentro de las celdas representan niveles de gris. Los puntos cuyo valor se indica en rojo se usarán como

semillas. Si se usan dos puntos de inicio la segmentación consiste de, a lo mucho dos regiones.

0	0	5	6	7
1	1	5	8	7
0	1	6	7	7
2	0	7	6	6
0	1	5	6	5

Distribución de niveles de gris

$\xrightarrow{P}$

10	10	30	30	30
10	10	30	30	30
10	10	30	30	30
10	10	30	30	30
10	10	30	30	30

T=2

50	50	50	50	50
50	50	50	50	50
50	50	50	50	50
50	50	50	50	50
50	50	50	50	50

T=8

La condición "*P*" a usar para incluir un píxel en una de las dos regiones es que la diferencia absoluta entre los niveles de ese píxel y el nivel de la semilla sea menor que un valor de umbral.

En la aplicación de este método surgen dos problemas:

1)La selección de las semillas (puntos de inicio) que representan apropiadamente las regiones de interés,
2)La selección de propiedades adecuadas para incluir puntos en las diversas regiones durante el proceso de crecimiento.

Generalmente, la selección de conjuntos de uno o más puntos de inicio se basa en la naturaleza del problema. Por ejem, en aplicaciones militares con imágenes infrarrojas, los objetivos de interés generalmente son más calientes (y entonces aparecen más brillantes) que el fondo. En este caso se escogerían los pixels más brillantes como punto de inicio del algoritmo de crecimiento de región.
A veces cuando no se tiene la información previa, se aplican las propiedades que se usarán para el crecimientos de región y los pixels que mejor las cumplan se usan como semillas. En el ejem. anterior, si se obtiene el histograma de la imagen, mostraría que los pixels con intensidades 1 y 7 son los más predominantes.

También se deben tomar en cuenta los conceptos de conectividad o adyacencia. Agrupar pixels con la misma intensidad para formar una "región" sin poner atención a la conectividad daría lugar a una segmentación sin ningún significado.

Otro problema en el crecimiento de región es la formulación de un criterio de paro. Básicamente, el crecimiento de una región terminaría cuando no existan más pixels que satisfagan el criterio de inclusión en esa región.

Algunos criterios adicionales que incrementan la eficiencia del algoritmo de crecimiento pueden ser pareados entre un pixel candidato y los pixels crecidos hasta el momento y la forma de la región que está siendo crecida.

Separación y unión de regiones.

Otro método para la segmentación de regiones, consiste en dividir una imagen inicialmente en un conjunto de regiones arbitrarias y separados y entonces unir y/o separar las regiones, de forma que se satisfagan las condiciones del inciso a) al e) anteriores (véase segmentación por regiones).

Sea "R" la que representa la región completa de la imagen y seleccione un predicado "P" según inciso d) y e). Para una imagen cuadrada, una aproximación para segmentar "R" es dividirla sucesivamente en regiones cuadradas más y más pequeñas, tal que, para cualquier región Ri se cumpla el inciso d. Esto es, si P(Ri)= FALSO, dividir la imagen en cuadrantes. Si P es falso para cualquier cuadrante, dividir el cuadrante en subcuadrantes, y así sucesivamente. Esta técnica de división tiene una representación conveniente en la forma de árbol (en el cual cada nodo tiene 4 descendientes).

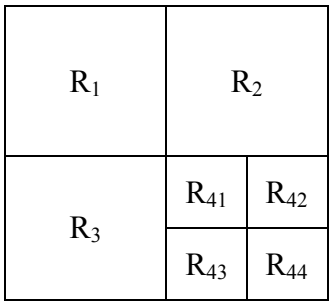
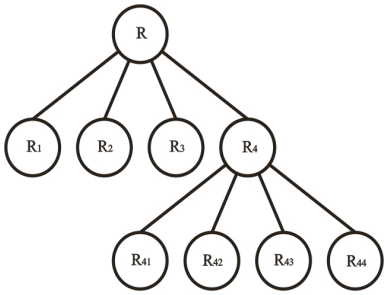


Imagen particionada



Árbol correspondiente

Con este procedimiento, la partición final probablemente contendrá regiones adyacentes con propiedades idénticas. Enseguida se aplica la unión o la fusión.

Para satisfacer los lineamientos del inciso a) al e) se requiere fusionar sólo las regiones adyacentes cuyos pixels combinados satisfacen el predicado b). Esto es, dos regiones adyacentes R_j y R_k se unen sólo si:

$$P(R_j \cup R_k) = \text{VERDADERO}.$$

Pasos del método

1. Separar en cuadrantes cualquier región R_i, donde $P(R_i)$= falso.
2. Unir cualquiera de las regiones adyacentes R_j y R_k para las cuales

$$P(R_j \cup R_k) = \text{verdadero}$$

3. Detenerse cuando no es posible separar o unir más.

Representación y descripción.

Después de segmentar una imagen, se necesita Representar y Describir de forma que sea adecuada para un procesamiento posterior por PC.
La representación consiste en asignar un modelo o identidad a una región (elemento) de una imagen que ha sido segmentada antes. La representación puede efectuarse utilizando características internas o externas de la región a considerar.

Representación externa: se utiliza generalmente el contorno o frontera de la región.

Representación interna: se utilizan las características internas de la región como niveles de gris de los pixels, textura, etc.

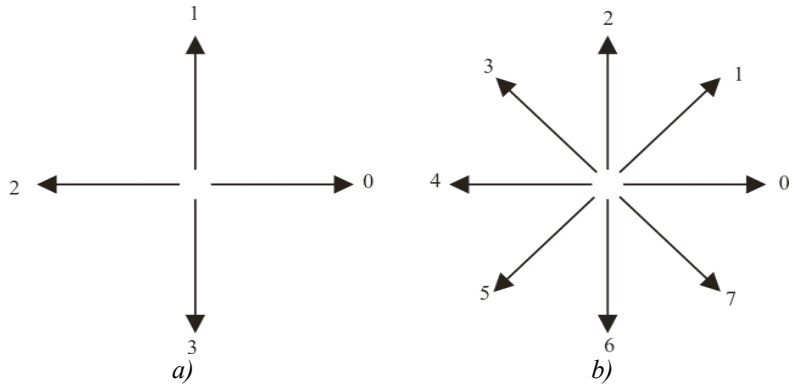
Para representar una región, ya sea con elementos internos o externos, se debe utilizar un modelo o característica para describir dicha región.

Por ejemplo, si se utiliza una representación externa de la región, se puede describir esa región mediante la longitud del contorno, la cantidad de líneas rectas o el número de curvas de cierto ángulo.

Sin importar la forma de representación, es importante considerar que los descriptores (elementos que describen una región representada por una de las dos formas) deben ser, en lo posible, independientes del tamaño, el desplazamiento y la rotación de las regiones a considerar.

Cadenas de código o Código Cadena

Las cadenas de código se utilizan para representar un contorno por medio de una sucesión conexa de segmentos de longitud y dirección especificadas. Normalmente esta representación se basa en segmentos de conectividad 4 u 8. la dirección de cada segmento se codifica utilizando un esquema de numeración como los que se indican a continuación:



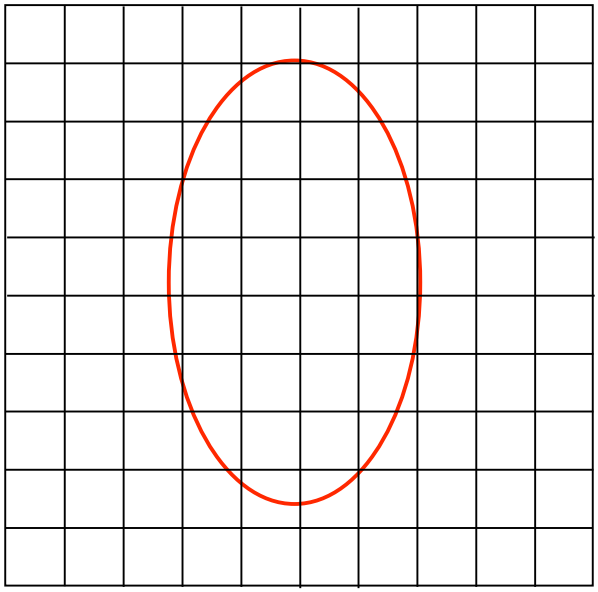
direcciones de código de cadena; (a) código de cadena de 4 direcciones y (b) código de cadena de 8 direcciones

Las imágenes digitales normalmente se adquieren y procesan en un formato de cuadrículas con igual espaciado en las direcciones x e y ; así pues, se podría generar un código de cadena siguiendo un contorno, por ejemplo, en sentido de las agujas del reloj y asignando una dirección a los segmentos que conectan cada par de píxeles.

Este método generalmente no se usa por 2 razones:

1) El código resultante generalmente es muy largo.
2) Cualquier disturbio a lo largo de la frontera debido a ruido o una segmentación imperfecta origina cambios en el código que puede no estar relacionado a la forma del borde.

Una forma de minimizar estos inconvenientes consiste en remuestrear el borde del objeto usando una malla, como se muestra enseguida. Así, a medida que se recorre el borde, cada píxel se asigna a un nodo de la malla dependiendo en la proximidad del borde original a ese nodo, (ver sig. figura). El borde remuestreado que se obtiene de esta manera puede ser representado por un código 4 o código 8.



El código de cadena de un contorno depende del punto de partida, pero dicho código se puede normalizar por un procedimiento directo: dado un código de cadena generado empezando en una posición arbitraria, se trata como una serie circular de números de dirección y se redefine el punto de partida de modo que la serie de números resultante forme un entero de módulo mínimo o máximo. También se puede normalizar por rotación utilizando la primera diferencia del código de cadena en vez del propio código. Por ejemplo, la primera diferencia del código de cadena de 4 direcciones 10103322 es 3133030. si decidimos tratar el código como una serie circular, entonces el primer elemento de la diferencia se calcula usando la transición entre los componentes último y primero de la cadena. Aquí el resultado es 3133030. se puede conseguir la normalización de tamaño alternado el del cuadriculado remuestreado.

Relaciones entre pixeles

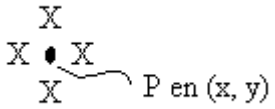
Cuando se considere una imagen digital, un pixel particular se denotará con letras minúsculas como *p* y *q*. Un subconjunto de pixels de *f (x,y)* se indicará con *S*.

Vecinos de un pixel

Un pixel *p* en la coordenada (*x,y*) tiene 4 vecinos verticales y horizontales con las siguientes coordenadas:

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

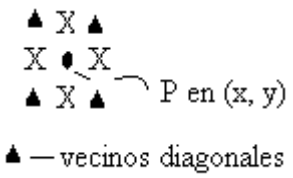
Estos se conocen como los 4 vecinos de *p* y se denotan como $N_4(p)$.

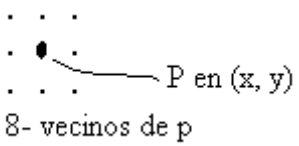


Los 4 vecinos diagonales de *p* tienen las coordenadas:

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

Se representan por $N_D(p)$; junto con los 4 vecinos de *p* se conocen como los 8 vecinos de *p* y se representan con $N_8(p)$





Conectividad

Es un concepto importante para conocer los límites de los objetos y los componentes de área de una imagen.

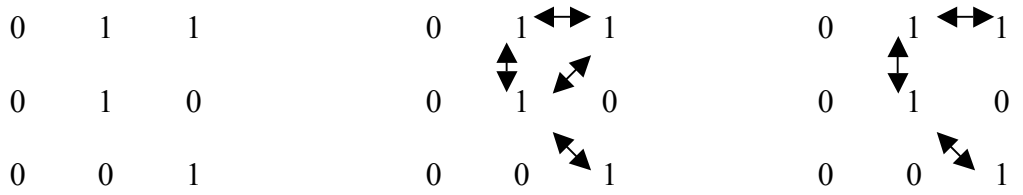
Para determinar si dos pixels son conexos (están conectados), debe determinarse si son adyacentes en algún sentido (vecinos) y si sus niveles de gris cumplen un criterio específico de similitud (ejemplo, si son iguales).

Sea "*V*" el conjunto de niveles de gris empleados para definir la conectividad (por ejemplo una imagen binaria puede ser V={1}. Se consideran 3 tipos de conectividad:

a) 4-conectividad: dos pixels *p* y *q* con valores dentro de "*V*" están 4-conectados si *q* pertenece a $N_4(p)$.
b) 8-conectividad: dos pixels *p* y *q* con valores dentro de *V* están 8-conectados si *q* pertenece a $N_8(p)$.
c) m-conectividad: dos pixels *p* y *q* están m-conectados si:

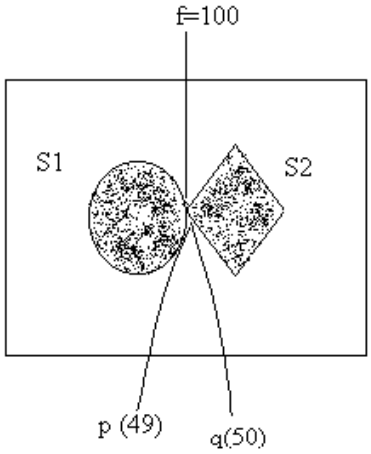
i) $q \in N_4(p)$ ó
ii) $q \in N_D(p)$ y además el conjunto $N_4(p) \cap N_4(q)$ está vacío

la m-conectividad se usa para eliminar la ambigüedad de la 8-conectividad.



Distribución de Conectividad. 8- conectividad m-conectividad.

Un pixel *p* es adyacente a un pixel *q* si están conectados; se define entonces 4- adyacente, 8- adyacente y/o m- adyacente.
Dos subconjuntos de imagen *S1* y *S2* son adyacentes si algún pixel de *S1* es adyacente a un pixel de *S2*.



Un *camino* desde *p* en coordenadas (*x, y*) hasta un pixel *q* en coordenadas (*s, t*) es una sucesión de diversos pixels de coordenadas:

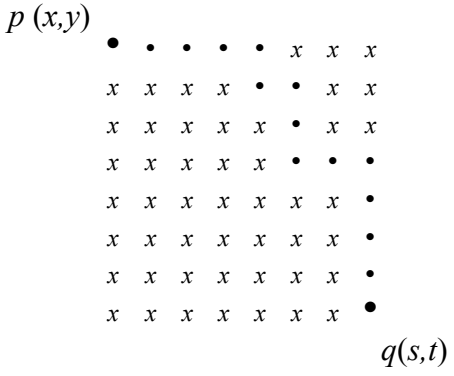
$(x_0, y_0);(x_1, y_1).....(x_n, y_n)$

donde:

$(x_0, y_0) = (x, y)$ en *p*
$(x_n, y_n) = (s, t)$ en *q*

(x_i, y_i) es adyacente a (x_{i-1}, y_{i-1})

$1 \leq i \leq n$ y *n* es la longitud del camino.



Si *p* y *q* son pixels en un subconjunto de imagen *S*, entonces *p* está conectado (conexo) a *q* en *S* si hay un camino desde *p* hasta *q* consistiendo totalmente por pixels en *S*.

Para cualquier pixel *p* dentro de *S*, el conjunto de pixels en *S* conectados con *p* se llama componente conexa de *S*. Por tanto, cualquier par de pixels de una misma componente conexa están conectados entre sí.