

Capítulo 2

Transformaciones en el dominio espacial

2.1 Representación de una imagen digital: matrices vinculadas.

Vamos a considerar un modelo de imagen simple según el cual la imagen viene dada por una función de intensidad luminosa bidimensional $f(x,y)$, donde el valor $f(x,y)$ nos da la intensidad luminosa de la imagen en el punto (x,y) . Como la luz es una forma de energía, $f(x,y) \in [0, \infty)$.

La imagen percibida por una persona procede de la luz reflejada por los objetos que la componen. La intensidad luminosa $f(x,y)$ se puede caracterizar por dos componentes:

a) La cantidad de luz incidente en la escena visualizada (iluminación), que representaremos por $i(x,y) \in [0, \infty)$. Esta componente viene determinada por la fuente luminosa. Por ejemplo, el nivel de iluminación en una oficina comercial es de unas 100 candelas-pies.

b) La cantidad de luz reflejada por los objetos de la escena (capacidad reflectora), que representaremos por $r(x,y) \in [0,1]$, donde el valor 0 representa absorción total y el valor 1 reflexión total. Esta componente viene determinada por las características de los objetos de la escena. Por ejemplo, el acero tiene 0.65 y la nieve 0.93.

Aquí el término imagen se refiere a la función bidimensional $f(x,y)$ que representa la intensidad luminosa en el punto (x,y) , y viene dada por

$$f(x,y) = i(x,y) \times r(x,y)$$

Para procesar una imagen por ordenador se debe *discretizar* (digitalizar) tanto espacialmente como en intensidad luminosa. La digitalización de la coordenada espacial (x,y) se llama *muestreo de la imagen* y la digitalización de la intensidad luminosa se llama *cuantificación del tono de gris*.

Llamaremos tono o nivel de gris de una imagen monocromática en el punto (x,y) a la intensidad luminosa de la misma en el punto (x,y) . El nivel de gris estará comprendido entre un valor mínimo L_{\min} y un valor máximo L_{\max} . Al intervalo $[L_{\min}, L_{\max}]$ lo llamaremos la *escala de gris*.

Para obtener la digitalización de una imagen analógica se realiza un muestreo sobre ella utilizando una rejilla o plano cuadrículado, donde cada región es representada por un punto (píxel) cuyo nivel de gris es el valor, de un conjunto finito de valores, que mejor representa a los tonos de gris de dicha región. Si la imagen es rectangular y la rejilla cuadrículada se obtiene un conjunto de $M \times N$ valores discretos que corresponden a los tonos de gris de cada uno de los píxeles obtenidos por el proceso de muestreo, expresados por la siguiente matriz:

$$\begin{pmatrix} f(1,N) & \dots & f(m,N) & \dots & f(M,N) \\ \dots & \dots & \dots & \dots & \dots \\ f(1,n) & \dots & f(m,n) & \dots & f(M,n) \\ \dots & \dots & \dots & \dots & \dots \\ f(1,1) & \dots & f(m,1) & \dots & f(M,1) \end{pmatrix}$$

en la que hemos utilizado una representación cartesiana en lugar de matricial, pues m representa la columna y n la fila.

Por lo tanto, una **imagen digital** es una función cuyo dominio es un conjunto discreto

$$D = \{(i,j) : i=1,2,\dots,M, j=1,2,\dots,N\}$$

y cuyo rango es el conjunto

$$R = \{0, 1, 2, \dots, L-1\}.$$

Los valores de M y N y el número L de niveles de gris suelen ser potencias de 2 para facilitar su representación binaria.

Por lo tanto, hemos pasado a una nueva imagen, la imagen digital, donde sus tonos de gris son valores que pertenecen al conjunto discreto $\{0, 1, 2, \dots, L-1\}$, que es una *discretización* del intervalo original $[L_{\min}, L_{\max}]$ de tonos de gris de la imagen analógica. Para representar los L tonos de gris necesitamos B bits, siendo $L=2^B$. Por ejemplo, para 256 tonos de gris necesitamos 8 bits de manera que el vector binario (0,0,0,0,0,0,0,0) representa al negro y (1,1,1,1,1,1,1,1) al blanco. De esta forma la imagen puede venir dada por un vector binario (de ceros y unos), donde los L primeros valores binarios corresponden al tono de gris del primer píxel de la imagen y así sucesivamente. El número de bits requeridos para almacenar una imagen digitalizada es $M \times N \times B$. Como la representación matricial anterior es una aproximación a una imagen continua, conforme mayores sean los valores de los parámetros M, N y L , se conseguirá una mejor resolución (grado de detalle en la percepción de la imagen).

Como una imagen suele ocupar bastante memoria y en muchas aplicaciones sólo vamos a analizar zonas concretas de la imagen, que llamaremos zonas o regiones de interés, **ROI** (Region of Interest), que serán submatrices, necesitamos una representación matricial especial que especifique la posición de la submatriz de interés en la imagen completa. Así vamos a asociar a cada imagen una *matriz vinculada*. Una matriz $M \times N$ cuyos elementos son números reales o asteriscos (cuando no se conoce el tono de gris del píxel correspondiente), de la forma

$$\begin{pmatrix} a_{11} & \dots & a_{1j} & \dots & a_{1s} \\ \dots & \dots & \dots & \dots & \dots \\ a_{i1} & \dots & a_{ij} & \dots & a_{is} \\ \dots & \dots & \dots & \dots & \dots \\ a_{r1} & \dots & a_{rj} & \dots & a_{rs} \end{pmatrix}_{p,q}$$

donde los indicadores p y q son números enteros que especifican la posición de a_{11} en la cuadrícula de la imagen (abscisa q y ordenada p), se dice que es una *matriz vinculada* asociada a una imagen. El símbolo asterisco se le asigna a cada píxel que tenga un tono de gris desconocido, debido, por ejemplo, a que el sensor que recoge la imagen no nos da información de él.

Por lo tanto, una **matriz vinculada** es:

- Una matriz cuyos elementos $a_{ij} \in R \cup \{*\}$
- Un indicador (p,q) que especifica la posición del primer elemento de la matriz en la rejilla utilizada en la digitalización de la imagen.

Una imagen puede venir representada de la forma siguiente:

3											
2		3	5		2						
1			3	3	3	1					
0			1	0	2	7					
-1											
-2											
...	-1	0	1	2	3	4	5	6	7	8	...

y una matriz vinculada asociada a ella, puede ser:

$$\begin{pmatrix} 3 & 5 & * & 2 & * \\ * & 3 & 3 & 3 & 1 \\ * & 1 & 0 & 2 & 7 \\ * & * & * & * & * \end{pmatrix}_{0,2}$$

donde utilizamos el indicador $p=0$ y $q=2$, para expresar que el primer elemento, $a_{11}=3$, ocupa la posición $(0,2)$ en la cuadrícula (considerada en el plano cartesiano).

Obsérvese que la imagen anterior también se puede representar por una matriz vinculada de tamaño 3×5 , si prescindimos de la cuarta fila, es decir, por una matriz que tiene una fila menos. Así,

dada una imagen cualquiera, siempre podemos encontrar una representación minimal para una matriz vinculada, es decir, por aquella que tenga el mínimo número de filas y columnas.

Una medida de la cantidad de información de una matriz vinculada $M \times N$, es la *densidad de información*, que representaremos por d , y se define como el número de elementos de la matriz, que son números reales, es decir, que no son *, dividido por el número total de elementos. Así, $d=1$ significa que no hay asteriscos, y la imagen, por tanto, ocupa toda la matriz, en cuyo caso diremos que la matriz está *saturada*.

2.2 Relaciones básicas entre píxeles.

A continuación vamos a estudiar algunas características topológicas de los píxeles. El entorno vertical y horizontal de tamaño 5 del píxel p , de coordenadas (i,j) , viene dado por el conjunto de 5 píxeles de coordenadas,

$$(i+1,j), (i-1,j), (i,j), (i,j+1), (i,j-1),$$

y lo representaremos por $N_5(p)$.

El entorno diagonal de tamaño 5 del píxel p , de coordenadas (i,j) , viene dado por el conjunto de píxeles de coordenadas,

$$(i+1,j+1), (i-1,j-1), (i,j), (i-1,j+1), (i+1,j-1),$$

y lo representaremos por $N_D(p)$; el conjunto de píxeles $N_5(p) \cup N_D(p)$ es un entorno de tamaño 9 del píxel p y lo representaremos por $N_9(p)$.

Para estudiar la relación entre píxeles vamos a tener en cuenta su proximidad espacial y su similitud en los niveles de gris. Por ello para establecer si dos píxeles están conectados tenemos que ver si son adyacentes en algún sentido y si sus niveles de gris son similares (con respecto a algún criterio de similaridad). Sea V el conjunto de valores de los tonos de gris utilizado para definir la conectividad (V va a ser un conjunto de números "próximos" y correlativos). Por ejemplo, V puede ser el conjunto de números del 16 al 22. Así podemos definir tres tipos de conectividad (figura 2.2):

- a) Conectividad de tipo 5: los píxeles p y q con valores en V están conectados si $q \in N_5(p)$
- b) Conectividad de tipo 9: los píxeles p y q con valores en V están conectados si $q \in N_D(p)$
- c) Conectividad de tipo mixto: los píxeles p y q con valores en V están conectados si:
 - i) $q \in N_5(p)$
 - ó
 - ii) $q \in N_D(p)$ y el conjunto $N_5(p) \cap N_D(q)$ no tiene píxeles con niveles de gris que pertenezcan a V .

Obsérvese que la conectividad de tipo mixto es una modificación de la conectividad de tipo 9, que se introduce para eliminar las conexiones con caminos múltiples que se pueden obtener en la conectividad de tipo 9.

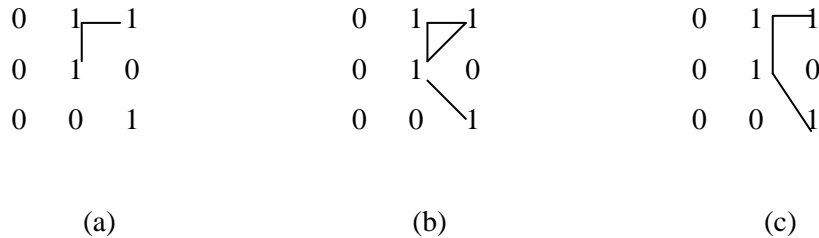


Figura 2.1. (a) Conectividad de tipo 5; (b) conectividad de tipo 9 y (c) conectividad mixta.

En la parte (a) de la figura 2.1 aparecen conectados del pixel del centro los píxeles de un entorno de tipo 5, tomando $V=\{1\}$; en la parte (b) de la figura 2.1 aparecen conectados del pixel del centro los píxeles de un entorno de tipo 9, tomando $V=\{1\}$ y en la parte (c) aparecen conectados los píxeles de un entorno de tipo mixto, tomando $V=\{1\}$.

Se dice que el píxel p es **adyacente** al píxel q , si los dos están conectados. El tipo de adyacencia se corresponderá con el tipo de conectividad. Dos subconjuntos de píxeles (imágenes) S_1 y S_2 se dice que son adyacentes si algún píxel de S_1 es adyacente con algún píxel de S_2 .

Un **camino** desde el píxel p , con coordenadas (x_0, y_0) , al píxel q , con coordenadas (x_n, y_n) , es una sucesión de diferentes píxeles con coordenadas

$$(x_0, y_0), (x_1, y_1), \dots, (x_{n-1}, y_{n-1}), (x_n, y_n),$$

donde (x_i, y_i) es adyacente a (x_{i-1}, y_{i-1}) , es decir, adyacentes consecutivamente. Diremos que la *longitud* de este camino es n .

Si p y q son píxeles de un subconjunto S , entonces diremos que p está conectado con q en S , si existe un camino de p a q formado sólo por píxeles de S . Para cualquier píxel p de S , el conjunto de píxeles de S que están conectados con p se dice que es una **componente conexa** de S .

La capacidad de asignar diferentes niveles a varias componentes conectadas, disjuntas entre sí, de una imagen es de vital importancia en el análisis automático de imágenes.

2.3 Operaciones aritméticas, geométricas, lógicas y vectoriales.

Las operaciones aritméticas que se pueden realizar con imágenes son las mismas que las que se realizan con números reales, vectores y matrices.

La suma de dos imágenes, f y g , se realiza sumando los valores de los píxeles que tienen la misma posición en la rejilla, es decir:

$$[SUM(f, g)](i, j) = \begin{cases} f(i, j) + g(i, j) & \text{si } f(i, j), g(i, j) \in R \\ * & \text{si } f(i, j) \text{ ó } g(i, j) \text{ es } * \end{cases}$$

La multiplicación de dos imágenes, f y g, se realiza multiplicando los valores de los píxeles que tienen la misma posición en la rejilla, es decir:

$$[MULT(f, g)](i, j) = \begin{cases} f(i, j) \cdot g(i, j) & \text{si } f(i, j), g(i, j) \in R \\ * & \text{si } f(i, j) \text{ ó } g(i, j) \text{ es } * \end{cases}$$

El máximo de dos imágenes f y g, MAX(f,g), se obtiene calculando el máximo de los valores de los píxeles que tienen la misma posición. Análogamente para el mínimo, MIN(f,g). El máximo de dos imágenes será una imagen formada por los píxeles más claros y el mínimo por los más oscuros.

La multiplicación de una imagen f por un escalar t se realiza multiplicando el valor de cada píxel por el escalar:

$$[ESCALAR(t, f)](i, j) = \begin{cases} t \cdot f(i, j) & \text{si } f(i, j) \neq * \\ * & \text{si } f(i, j) = * \end{cases}$$

donde t es un número real.

La opuesta de la imagen f se define como:

$$[-(f)](i, j) = \begin{cases} -f(i, j) & \text{si } f(i, j) \neq * \\ * & \text{si } f(i, j) = * \end{cases}$$

Obsérvese que es un caso particular de ESCALAR para t = -1.

La inversa de una imagen se define como:

$$[INV(f)](i, j) = \begin{cases} 1/f(i, j) & \text{si } f(i, j) \neq * \text{ y } f(i, j) \neq 0 \\ * & \text{en otro caso} \end{cases}$$

La traslación de una imagen f, i píxeles hacia la derecha y j píxeles hacia arriba, se define como:

$$f = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix}_{r,t} \quad \text{TRAS}(f; i, j) = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix}_{r+i, t+j}$$

$$[TRAS(f; i, j)](u, v) = f(u-i, v-j)$$

Como se puede apreciar, los valores a_{pq} no cambian, sólo cambia la localización de cada elemento en la rejilla:

$$\text{TRAS}(f;i,j) = (a_{pq})_{r+i, t+j}$$

La rotación o giro de una imagen, sobre el punto de la rejilla (0,0), puede ser de 90, 180, 270 y 360 grados. El resultado de la rotación de 90 grados (en el sentido de las agujas del reloj) de una imagen f cuya matriz vinculada correspondiente es una matriz $m \times n$, y cuyo indicador de posición en la rejilla es (r,t) , viene dado por una matriz vinculada $n \times m$ cuyo indicador de posición en la rejilla es $(-t, r+m-1)$. Al operador rotación de 90 grados lo vamos a llamar $\text{NOVENTA}(f)$. Este operador, definido punto a punto quedaría:

$$[\text{NOVENTA}(f)](i,j) = f(-j,i)$$

si la rotación es en el sentido de las agujas del reloj, y

$$[\text{NOVENTA}(f)](i,j) = f(j,-i)$$

si el giro es en el sentido contrario a las agujas del reloj.

Para realizar una rotación de 180^0 basta con aplicar primero una rotación de 90^0 , y sobre ella aplicar otra nueva rotación de 90^0 ; es decir, aplicar dos veces este operador y, por tanto, lo representaremos por NOVENTA^2 . Aplicando otra rotación de 90^0 sobre esta última, obtenemos una rotación final de 270^0 de la imagen original.

A partir de una imagen también se le puede obtener su imagen simétrica con respecto a un eje (imagen especular), por ejemplo, de 45^0 , 135^0 ó de 180^0 , que es como si la hubiéramos desplazado pero dándole la vuelta, de forma que queda como la imagen formada en un espejo. Por ejemplo, una vuelta de 135^0 , viene definida por la siguiente transformación:

$$[\text{FLIP}(f)](i,j) = f(-j, -i)$$

y en términos de las matrices vinculadas, sería

$$\begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix}_{r,t} \longrightarrow \boxed{\text{FLIP}} \longrightarrow \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix}_{-t, -r}$$

Vamos a estudiar algunas operaciones del tipo base de datos como la operación de SELECCIÓN, que consiste en seleccionar la parte de la imagen que está dentro de una determinada ventana W ,

$$W=\{ (i,j): h \leq i < h+r, \quad k < j \leq k+s \}$$

El operador SELECCION depende de la imagen y de cuatro parámetros, que son: el tamaño de la ventana, $r \times s$ y la localización (h,k) de la primera columna y de la primera fila de la imagen de salida, es decir:

$$[SELECT(f; m, n, h, k)](i, j) = \begin{cases} f(i, j) & \text{para } (i, j) \in W \\ * & \text{en otro caso} \end{cases}$$

Veamos ahora la operación de EXTENSION. Dadas dos imágenes f y g , la operación de EXTENSION de la imagen f , mediante la imagen g , viene definida de la siguiente forma:

$$[EXTENSION(f, g)](i, j) = \begin{cases} f(i, j) & \text{si } f(i, j) \neq * \\ g(i, j) & \text{en otro caso} \end{cases}$$

Mediante la composición de estos operadores básicos podemos construir nuevos operadores que llamaremos de alto nivel. Por ejemplo, el operador EXTSUMA(f, g), se puede definir como:

$$EXTSUMA(f, g) = \begin{cases} SUMA(f, g) & \text{en } D_f \cap D_g \\ f & \text{en } D_f - D_g \\ g & \text{en } D_g - D_f \end{cases}$$

donde D_f es el dominio de la imagen f (conjunto de puntos (x,y) , cuyo nivel de gris es un número real) y D_g es el dominio de la imagen g . También se puede expresar como la siguiente composición:

$$EXTSUMA(f, g) = EXTENSION [EXTENSION (SUMA(f, g), g), f]$$

Obsérvese como el dominio del resultado de aplicar el operador EXTSUMA a las imágenes f y g , es la unión de los dominios de f y g , mientras que el dominio de aplicar el operador SUMA a las imágenes f y g , es la intersección de los dominios de f y g .

Hasta ahora, las transformaciones que hemos utilizado, tenían como entrada una imagen y como salida otra, es decir, eran de imagen a imagen. Además, una imagen venía dada por una matriz vinculada. Ahora, vamos a considerar la imagen como un vector tridimensional definido por las localizaciones absolutas (i,j) y por los niveles de gris $f(i,j)$ de los píxeles correspondientes, de manera que vamos a considerar la imagen constituida por el DOMINIO y por el RANGO. El DOMINIO viene constituido por h pares ordenados:

$$D = [(i_1, j_1), (i_2, j_2), \dots, (i_h, j_h)]$$

y el RANGO viene dado por el conjunto de los h valores de gris correspondientes:

$$R = [y_1, y_2, \dots, y_h] \text{ donde } y_k = f(i_k, j_k).$$

Veamos ahora algunas transformaciones estructurales que nos permitirán crear nuevas imágenes y obtener imágenes binarias. Se define el operador estructural CREAR de la siguiente forma:

$$[CREAR(D, R)](i, j) = \begin{cases} y_k & \text{si } (i, j) = (i_k, j_k) \\ * & \text{en otro caso} \end{cases}$$

El operador UMBRAL se define para producir una imagen binaria con sólo dos niveles de gris, 1 (blanco) y 0 (negro). Este operador se aplica a la imagen y depende de un parámetro t , llamado umbral, y viene definido de la siguiente forma:

$$[UMBRALE(f; t)](i, j) = \begin{cases} 1 & \text{si } f(i, j) \geq t \\ 0 & \text{si } f(i, j) < t \\ * & \text{si } f(i, j) = * \end{cases}$$

Uno de los métodos más usuales de extraer una figura o una característica de particular interés de una imagen (como puede ser un borde), es aplicando el método de la umbralización. El operador UMBRAL produce una imagen en blanco y negro, en la que el objeto de interés puede ser negro y el fondo blanco. Así pues, este operador transforma una imagen digital en una imagen binaria.

Para estudiar las propiedades del operador UMBRAL, se puede establecer una relación de orden sobre imágenes que tienen el mismo dominio. Se dice que una imagen g es menor o igual que la imagen h , si g y h tienen el mismo dominio D y para cada píxel $(i, j) \in D$, el nivel de gris de g es menor o igual que el de h . Es decir,

$$g(i, j) \leq h(i, j), \forall (i, j) \in D$$

Así,

- 1) Si $g \leq h \Rightarrow UMBRAL(g; t) \leq UMBRAL(h; t)$.
- 2) Si $t \leq s \Rightarrow UMBRAL(g; s) \leq UMBRAL(g; t)$.

La propiedad 1) expresa que el UMBRAL de una imagen más luminosa conduce a una imagen más luminosa.

La propiedad 2) expresa que la umbralización de la misma imagen con un umbral más alto, conduce a una imagen menos luminosa.

Otra operación parecida a la anterior es la TRUNCACION, que se define como:

$$[TRUNCAR(f;t)](i,j) = \begin{cases} f(i,j) & \text{si } f(i,j) \geq t \\ 0 & \text{si } f(i,j) < t \\ * & \text{si } f(i,j) = * \end{cases}$$

Obsérvese que $TRUNCAR(f;t) = MULT[f, UMBRAL(f;t)]$.

Análogamente, se definen los siguientes operadores:

$$[IGUAL(f;t)](i,j) = \begin{cases} 1 & \text{si } f(i,j) = t \\ 0 & \text{si } f(i,j) \neq t \text{ y } f(i,j) \neq * \\ * & \text{si } f(i,j) = * \end{cases}$$

$$[MAYOR(f;t)](i,j) = \begin{cases} 1 & \text{si } f(i,j) > t \\ 0 & \text{si } f(i,j) \leq t \\ * & \text{si } f(i,j) = * \end{cases}$$

Veamos ahora algunas operaciones vectoriales. Se define el operador PIXSUM como

$$PIXSUM(f) = \sum_{(i,j) \in D_f} f(i,j).$$

El operador MEDIA se define de la siguiente forma:

$$MEDIA(f) = \frac{PIXSUM(f)}{CARD(F)}$$

donde $CARD(f)$ es el tamaño de la imagen (número de píxeles).

Se define el operador PRODUCTO ESCALAR de las imágenes f y g como:

$$PRESC(f,g) = \begin{cases} \sum_{(i,j) \in D} f(i,j) \cdot g(i,j) & \text{si } f \text{ y } g \text{ tienen el mismo dominio} \\ * & \text{si } f \text{ y } g \text{ tienen distinto dominio} \end{cases}$$

Se define la NORMA Euclídea de una imagen de la siguiente forma:

$$NORMA(f) = \|f\| = \sqrt{PRESC(f,f)}$$

Un **vector de imágenes** es una colección ordenada de n imágenes, que denotaremos por:

$$f = (f_1, f_2, \dots, f_n)$$

donde f_k es la imagen k , y

$$f(i,j) = (f_1(i,j), f_2(i,j), \dots, f_n(i,j)),$$

siendo $f_k(i,j)$ el nivel de gris del píxel (i, j) de la imagen k , $k=1,2,...,n$.

Sobre este vector se pueden realizar las mismas operaciones que con vectores, pero teniendo en cuenta que algunas de sus componentes pueden ser $*$. Se puede definir una *norma* (magnitud del vector) de la misma manera que en los espacios vectoriales. Así, el operador $MAG0(f)$ corresponde a la norma L_∞ (norma de Tchebycheff), que viene dado por

$$\|f(i, j)\|_\infty = \max\{|f_1(i, j)|, |f_2(i, j)|, \dots, |f_n(i, j)|\}$$

Análogamente, se define el operador $MAG1(f)$, que corresponde a la norma L_1 (norma rectangular) de la siguiente forma:

$$\|f(i, j)\|_1 = |f_1(i, j)| + |f_2(i, j)| + \dots + |f_n(i, j)|$$

y el operador $MAG2(f)$, que corresponde a la norma L_2 (norma euclídea):

$$\|f(i, j)\|_2 = \left[|f_1(i, j)|^2 + |f_2(i, j)|^2 + \dots + |f_n(i, j)|^2 \right]^{1/2}$$

2.4. Filtros lineales de promedio (máscaras o plantillas)

Un *entorno* del píxel (i,j) , de tamaño $(2m+1) \times (2n+1)$, es el conjunto de píxeles que viene definido como:

$$N_{(2m+1) \times (2n+1)}(i, j) = \{ (r, s) \in \mathbb{Z} \times \mathbb{Z} : i-m \leq r \leq i+m, j-n \leq s \leq j+n \}$$

Una *máscara* (o plantilla) g de tamaño $(2m+1) \times (2n+1)$ es una matriz de tamaño $(2m+1) \times (2n+1)$ cuyo elemento central corresponde al origen de coordenadas (representado por $g(0,0)$) y todos sus elementos son números reales. Es decir,

$$g \equiv \begin{pmatrix} g(-m, n) & \dots & g(0, n) & \dots & g(m, n) \\ \dots & \dots & \dots & \dots & \dots \\ g(-m, 0) & \dots & g(0, 0) & \dots & g(m, 0) \\ \dots & \dots & \dots & \dots & \dots \\ g(-m, -n) & \dots & g(0, -n) & \dots & g(m, -n) \end{pmatrix}$$

Dado un entorno $N(i, j)$ del píxel (i, j) , de tamaño $(2m+1) \times (2n+1)$, podemos superponer sobre él una plantilla (máscara) del mismo tamaño (se hace coincidir el centro de la máscara con el píxel (i, j)) y sustituir el nivel de gris $f(i, j)$ de dicho píxel por una suma ponderada de los niveles de gris de los píxeles del entorno con los valores correspondientes de la máscara, es decir, por la expresión:

$$f^*(i, j) = \sum_{(r,s) \in N} g(r, s) \cdot f(i+r, j+s)$$

Si hacemos $h=i+r$ y $k=j+s$ entonces también se puede expresar como

$$f * (i, j) = \sum_{(h,k) \in \mathbb{Z} \times \mathbb{Z}} g(h-i, k-j) \cdot f(h, k),$$

pues los valores de la máscara $g(r,s)$ son iguales a cero para $(r,s) \notin N(i,j)$ y los valores de la imagen $f(h, k)$ son también iguales a cero para $(h,k) \notin D_f$. Obsérvese que la expresión anterior es la operación de *convolución* entre las funciones f y g . Esta transformación se conoce con el nombre de **filtro lineal de promedio**. En otras palabras, dada una máscara g y una imagen f , se define el operador FILTRO de la siguiente forma:

$$[FILTRO(f; g)](i, j) = \sum_{(r,s) \in N} g(r, s) \cdot [TRAS(f; r, s)](i, j)$$

También se puede decir que este operador es un *filtro lineal de promedio móvil*, pues realiza un promedio con los valores de su entorno (ponderados con los coeficientes correspondientes de la máscara) para cada uno de los píxeles de la imagen. Obsérvese que el operador FILTRO tiene como entrada dos imágenes (la imagen original y la máscara), y como salida una sola imagen (la imagen filtrada).

Cuando todos los valores de la máscara son no negativos y además suman uno ($PIXSUM(M)=1$) decimos que se trata de un **filtro de paso baja**. Si se aplica el operador FILTRO de *paso baja*, la imagen obtenida es una imagen más suavizada (se han reducido las fluctuaciones en los niveles de gris), pues en la imagen filtrada, la intensidad de gris de cada píxel es una media ponderada de las intensidades de los píxeles de su entorno. Por ello se dice que se ha utilizado un filtro de paso baja, ya que atenúa o elimina las altas frecuencias, es decir, las grandes variaciones en los tonos de gris de los píxeles, y deja pasar a través del filtro sólo las bajas frecuencias, es decir, no modifica (o apenas modifica) las pequeñas variaciones. Como consecuencia la imagen filtrada es más borrosa ya que se ha atenuado el contraste o diferencia entre los niveles de gris de píxeles adyacentes. Además, como $PIXSUM(M) = 1$, la imagen filtrada tiene la misma luminosidad que la imagen original. Por lo tanto una ventaja de este tipo de filtro es que atenúa los ruidos de alta frecuencia a cambio de introducir borrosidad en la imagen (falta de contraste).

Cuando una máscara tiene valores negativos y positivos y la suma total es cero ($PIXSUM(M)=0$) decimos que se trata de una **máscara o filtro de paso alta**. Estos filtros destacan y acentúan las grandes variaciones entre los tonos de gris de los píxeles (dejan pasar las componentes de alta frecuencia) y atenúan o eliminan las pequeñas diferencias (componentes de baja frecuencia). Es decir, el resultado neto de un filtro de paso alta es destacar sólo aquellos píxeles en cuyo entorno hay grandes diferencias en los tonos de gris y que corresponde, por tanto, a los ejes, los bordes, los contornos y otros detalles que configuran la forma de los objetos de la imagen, a costa de incrementar los ruidos. Como $PIXUM(M)=0$, los niveles de gris de la nueva imagen serán positivos

o negativos y próximos a cero en los entornos con poca variación en los niveles de gris. La intensidad luminosa media de la imagen será igual a cero. Si aplicamos a la nueva obtenida el operador valor absoluto y después el operador umbral obtenemos una imagen binaria donde los píxeles que corresponden a ejes aparecen con valor 1 y el resto con valor cero.

Cuando se desea detectar los ejes o bordes de una imagen se utilizará un filtro de paso alta, como por ejemplo:

$$\begin{pmatrix} 0 & -1 & 0 \\ -1 & 4 & -1 \\ 0 & -1 & 0 \end{pmatrix}$$

Con ello se consigue que cuando el entorno de un píxel tenga los mismos niveles de gris (entorno constante) o valores próximos de niveles de gris (entorno con pequeñas fluctuaciones), entonces al aplicarle una máscara con $\text{PIXSUM} = 0$, el valor de salida del píxel será 0 ó próximo a 0, pues no hay contraste significativo entre los píxeles del entorno al que se ha aplicado la máscara M . Por el contrario, si en el entorno hay diferencia significativa entre los niveles de gris, el valor de salida del píxel podrá no ser próximo a cero, marcando así las diferencias. Por lo tanto, cuando se utiliza una máscara con coeficientes positivos y negativos, y $\text{PIXSUM} = 0$, se resaltan las grandes diferencias entre los niveles de gris de los píxeles del entorno (altas frecuencia) y se eliminan las pequeñas fluctuaciones en los niveles de gris (bajas frecuencias). Así, una máscara con coeficientes negativos y positivos y con $\text{PIXSUM} = 0$ es un filtro de *paso alta*.

La imagen obtenida por una máscara de paso alta se puede obtener también como la diferencia entre la imagen original y la imagen filtrada con una máscara de paso baja.

Los filtros de promedio móvil pueden ser:

- a) *Invariantes en el espacio*, en cuyo caso, la imagen filtrada se obtiene de la aplicación de la misma máscara a cada uno de los píxeles la imagen, o
- b) *Variables en el espacio*, cuando el filtro se realiza mediante la aplicación de una colección de máscaras, de manera que a subconjuntos diferentes de píxeles se le aplican máscaras diferentes.

2.5. Filtros no lineales

Dado un entorno $N(i, j)$ del píxel (i, j) , de tamaño $(2m+1) \times (2n+1)$, podemos sustituir su tono del gris por el valor mínimo de los tonos de gris de los píxeles de su entorno. Esto hace el operador *mínimo* (MIN) que viene definido por la expresión:

$$[\text{MIN}(f, N)](i, j) = \text{Min} \{ f(r, s) : (r, s) \in N(i, j) \}.$$

La imagen resultante tiene menor luminosidad y suele ser más homogénea que la imagen original. Su utilización nos puede ayudar, por ejemplo, a corregir la desigual iluminación de una imagen.

Análogamente podemos definir el operador *máximo* (MAX) mediante la expresión:

$$[\text{MAX}(f, N)](i, j) = \text{Max} \{ f(r, s) : (r, s) \in N(i, j) \}$$

Otro operador no lineal es el filtro *MEDIANA* que viene definido como:

$$[\text{MEDIANA}(f, N)](i, j) = \text{Mediana} \{ f(r, s) : (r, s) \in N(i, j) \}$$

Este filtro nos permitirá eliminar o atenuar los ruidos puntuales o granulares de las imágenes, siempre que el porcentaje de píxeles con ruido en el entorno sea menor que la mitad de los elementos del entorno.

2.6. Filtros de tipo Gradiente

Si $f(x, y)$ es una función de valores reales de dos variables reales, las derivadas parciales $\partial f / \partial x$ y $\partial f / \partial y$ nos dan la tasa de cambio instantánea en las direcciones del eje de abscisas y del eje de ordenadas respectivamente. Con el fin de estudiar las tasas de cambio de los niveles de gris vamos a definir los operadores basados en diferencias. Definiremos el operador DIFERENCIA en la dirección de x , DX , como:

$$[DX(f)](i, j) = f(i, j) - f(i - 1, j)$$

y análogamente, el operador diferencia en la dirección del eje de ordenadas es:

$$[DY(f)](i, j) = f(i, j) - f(i, j - 1).$$

El operador DX mide la diferencia entre el valor de gris de un píxel y el de su adyacente sobre el eje X . Así, en una imagen podremos apreciar si la iluminación de la misma va de claro a oscuro (o de oscuro a claro) en la dirección horizontal, según que los valores de los elementos de DX sean positivos o negativos, y cuantificar dicha variación. Como los ejes verticales vienen caracterizados por una variación de la iluminación en la dirección del eje X , podemos detectarlos mediante dicho operador. Utilizando el operador DX con los operadores ABS y UMBRAL se podrán obtener los ejes horizontales en una imagen binaria (blanco y negro).

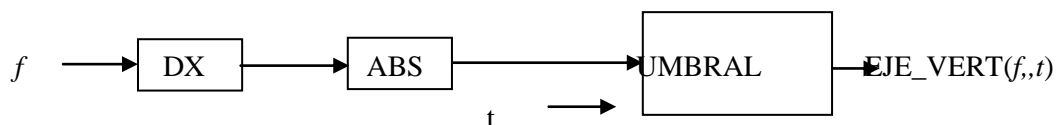


Figura 2.6.1. Diagrama del operador EJE_VER(f, t).

De manera análoga, mediante la aplicación del operador DY, se puede apreciar la variación de arriba a abajo, o de abajo a arriba, de la iluminación de la imagen y se pueden detectar ejes horizontales, usando el operador compuesto EJE_HOR(f,t).

Es importante resaltar que estos operadores, DX y DY, se pueden contemplar como filtros de promedio móvil. En efecto, consideremos las máscaras

$$D1 = \begin{pmatrix} -1 & 1 \end{pmatrix} \quad \text{y} \quad D2 = \begin{pmatrix} 1 \\ -1 \end{pmatrix},$$

en las que subrayamos el valor de la máscara que corresponde al píxel (i,j). Así,

$$\begin{aligned} [\text{FILTRO}(f;D1)](i,j) &= [(-1) \times f(i-1,j)] + [1 \times f(i,j)] \\ &= f(i,j) - f(i-1,j) \\ &= [\text{DX}(f)](i,j), \end{aligned}$$

y análogamente si tomamos la máscara D2, se obtiene DY.

Para obtener los bordes de una figura elíptica se usa conjuntamente DX y DY, de forma que el operador resultante sea independiente de la orientación (*isotrópico*). Se define el GRADIENTE de una imagen digital como el vector imagen:

$$\text{GRAD}(f) = (\text{DX}(f), \text{DY}(f))$$

y el **operador gradiente** viene dado por la expresión:

$$[\text{GRAD}(f)](i,j) = ([\text{DX}(f)](i,j), [\text{DY}(f)](i,j))$$

Obsérvese que el gradiente tiene una única imagen de entrada y dos imágenes de salida. Como el gradiente es un vector se le puede aplicar una norma, que nos mide la magnitud del gradiente. Así, los operadores MAG0(GRAD(f)), MAG1(GRAD(f)) y MAG2(GRAD(f)) son el resultado de aplicarle a una imagen el operador GRAD(f) y luego MAG0(f), MAG1(f) y MAG2(f), respectivamente. El operador MAG2(GRAD(f)) es *isotrópico* (es decir, invariante a giros).

Hasta ahora, hemos observado la variación sólo en un sentido, es decir, con respecto al píxel de la izquierda. Si tenemos en cuenta el píxel de la izquierda y el de la derecha obtenemos el operador DIFERENCIA SIMÉTRICA en la dirección del eje X

$$\begin{aligned} [\text{SIMDX}(f)](i,j) &= ([\text{DX}(f)](i,j) + [\text{DY}(f)](i,j)) / 2 \\ &= [f(i+1,j) - f(i-1,j)] / 2 \end{aligned}$$

y análogamente el operador DIFERENCIA SIMÉTRICA en la dirección del eje Y,

$$[\text{SIMDY}(f)](i,j) = [f(i,j+1) - f(i,j-1)] / 2.$$

Obsérvese que estos operadores pueden ser vistos como filtros, considerando las máscaras:

$$S1 = \begin{pmatrix} 1/2 & 0 & 1/2 \end{pmatrix} \quad y \quad S2 = \begin{pmatrix} 1/2 \\ 0 \\ -1/2 \end{pmatrix}$$

Definimos ahora el GRADIENTE SIMETRICO como el vector imagen

$$\mathbf{SIMGRAD}(f) = (\text{SIMDX}(f), \text{SIMDY}(f)),$$

y el operador gradiente simétrico viene dado por el operador compuesto **MAG0(SIMGRAD(f))**.

Estas expresiones se puede generalizar y obtener el **operador diferencia de Prewitt**, que contempla los cambios en seis píxeles:

$$\begin{aligned} [\text{PREWDX}(f)](i,j) &= ([\text{DX}(f)](i+1,j+1) + [\text{DX}(f)](i,j+1) + [\text{DX}(f)](i+1,j) + \\ &\quad [\text{DX}(f)](i,j) + [\text{DX}(f)](i+1,j-1) + [\text{DX}(f)](i,j-1)) / 6 \\ &= [f(i+1,j+1) + f(i+1,j) + f(i+1,j-1) - f(i-1,j+1) - f(i-1,j) - f(i-1,j-1)] / 6. \end{aligned}$$

También se puede expresar este operador en términos de la **máscara de Prewitt**:

$$\text{PREW1} = \begin{pmatrix} -1 & 0 & 1 \\ -1 & 0 & 1 \\ -1 & 0 & 1 \end{pmatrix}$$

y por tanto,

$$\text{PREWDX}(f) = \text{ESCALAR}[1/6; \text{FILTRO}(f; \text{PREW1})]$$

Este filtro permite detectar ejes verticales. De forma análoga, se define este operador en la dirección del eje de ordenadas mediante la máscara:

$$\text{PREW2} = \begin{pmatrix} 1 & 1 & 1 \\ 0 & 0 & 0 \\ -1 & -1 & -1 \end{pmatrix}$$

y por tanto,

$$\text{PREWDY}(f) = \text{ESCALAR}[1/6; \text{FILTRO}(f; \text{PREW2})]$$

que detecta ejes horizontales.

Utilizando el operador diferencia de Prewitt, definimos el **operador gradiente de Prewitt**:

$$\mathbf{PREWGRAD}(f) = (\text{PREWDX}(f), \text{PREWDY}(f))$$

Para cuantificar el vector, aplicaremos una norma, y con ello, obtenemos los operadores **PREWMAG0(f)**, **PREWMAG1(f)** y **PREWMAG2(f)**.

Otro operador que promedia los valores de gris de los mismos píxeles pero con pesos diferentes viene dado por las **máscaras de Sobel**:

$$SOB1 = \begin{pmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{pmatrix}, \quad SOB2 = \begin{pmatrix} 1 & 2 & 1 \\ 0 & 0 & 0 \\ -1 & -2 & -1 \end{pmatrix}$$

y el **operador diferencia de Sobel** en cada dirección viene dado por:

$$SOBDX(f) = \text{ESCALAR}[1/8; \text{FILTRO}(f; SOB1)]$$

$$SOBDY(f) = \text{ESCALAR}[1/8; \text{FILTRO}(f; SOB2)]$$

El **gradiente de Sobel** viene dado por el vector:

$$SOBGRAD(f) = (SOBDX(f), SOBDY(f)).$$

Como puede observarse, el **gradiente de Sobel** utiliza el factor 2 para ponderar dos píxeles. Se puede generalizar este operador utilizando el factor λ y en lugar de dividir por 8, se dividirá por, $(4+2\lambda)$.

Otro operador gradiente de la imagen comúnmente empleado es el **gradiente de Roberts**. Este operador pondera los píxeles que están en la dirección de 45° y -45° . Así, las **máscaras de Roberts** son:

$$ROB1 = \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix}, \quad ROB2 = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}.$$

El **operador gradiente cruzado de Roberts** viene dado por la expresión:

$$[\mathbf{ROBGRAD}(f)](i, j) = \left(\frac{\sqrt{2}}{2} [f(i, j) - f(i-1, j+1)], \frac{\sqrt{2}}{2} [f(i, j+1) - f(i-1, j)] \right)$$

El factor $\sqrt{2}/2$ resulta del hecho de que la distancia que hay entre estos píxeles es $\sqrt{2}$.

En muchas imágenes aparecen figuras o patrones definidos por alguna forma de homogeneidad con respecto al resto de los píxeles, bien por los niveles de gris o por las propiedades de textura de la imagen. Así, una figura puede aparecer destacada por sus niveles de gris más blancos frente a un fondo más oscuro. Utilizando el operador UMBRAL podemos destacar la figura. Sin embargo, hay imágenes más complejas, en las que el ojo distingue perfectamente los objetos y con el operador UMBRAL se confunden. Esto ocurre si una de las figuras es más clara que el fondo y la otra más oscura.

Vamos a analizar las diferencias en el contraste local mediante la detección de ejes. Un eje es una zona de demarcación o transición entre dos regiones que difieren según alguna medida de homogeneidad (como el nivel de gris). Una arista se puede presentar de dos formas:

- 1) Como el borde entre dos regiones diferentes, siendo cada una de estas regiones homogéneas con respecto a algún criterio de homogeneidad.
- 2) Como un arco delgado que destaque sobre el fondo.

Para la detección de aristas tomaremos como entrada al proceso una imagen y como salida una imagen de aristas. La manera más adecuada de producir una imagen de aristas es basándose en los cambios o variaciones de la homogeneidad en los niveles de gris. Esto se puede realizar aplicando algún operador GRADIENTE a la imagen original, después se le aplica el operador MAGNITUD a las imágenes de salida, y, finalmente, se aplica el operador UMBRAL a la imagen resultante, con un valor umbral apropiado, para separar aquellos píxeles cuya magnitud del gradiente es mayor. La bondad del resultado obtenido dependerá de los operadores GRADIENTE y UMBRAL utilizados. La metodología a seguir en la detección de los ejes o aristas depende sólo de las magnitudes relativas de los gradientes dentro de la imagen, y, por lo tanto, la multiplicación escalar por factores (1/6 en el gradiente de Prewitt y un 1/8 en el gradiente de Sobel) no juega ningún papel relevante. Así, la metodología a seguir consiste en aplicar algún operador de magnitud al resultado del operador gradiente:

$$(\text{FILTRO}(f;M1), \text{FILTRO}(f;M2))$$

donde la expresión M1 puede ser DX, PREW1, SOB1 o ROB1, y después se aplica a la imagen resultante el operador UMBRAL.

Para la obtención de una imagen binaria en la que poder detectar claramente las aristas o bordes de los objetos utilizaremos operadores del tipo: PREWEJE0(f; t), PREWEJE1(f; t), PREWEJE2(f; t), ROBEJE0(f; t), ROBEJE1(f; t), ROBEJE2(f; t), donde, por ejemplo:

$$\text{PREWEJE0}(f;t) = \text{UMBRALE} [\text{MAG0} [\text{PREWGRAD}(f)] ; t]$$

Además de utilizar las máscaras que detectan los cambios en las direcciones verticales y horizontales, o, como en el caso del gradiente de Robert, en las direcciones de 45° y -45° , podemos utilizar máscaras que tengan en cuenta los cambios en todas las direcciones que son múltiplos de 45° . Como hay ocho direcciones diferentes, en cada filtrado se utilizarán ocho máscaras. Así, por ejemplo, las ocho máscaras de Prewitt serían:

$$\begin{pmatrix} 1 & 1 & 1 \\ 1 & -2 & -1 \\ 1 & -1 & -1 \end{pmatrix} \quad \begin{pmatrix} 1 & 1 & 1 \\ 1 & -2 & 1 \\ -1 & -1 & -1 \end{pmatrix} \quad \begin{pmatrix} 1 & 1 & 1 \\ -1 & -2 & 1 \\ -1 & -1 & 1 \end{pmatrix}$$

$$\begin{pmatrix} -1 & 1 & 1 \\ -1 & -2 & 1 \\ -1 & 1 & 1 \end{pmatrix} \begin{pmatrix} NO & N & NE \\ E & - & O \\ SO & S & SE \end{pmatrix} \begin{pmatrix} 1 & 1 & -1 \\ 1 & -2 & -1 \\ 1 & 1 & -1 \end{pmatrix}$$

$$\begin{pmatrix} 1 & -1 & -1 \\ 1 & -2 & -1 \\ 1 & 1 & 1 \end{pmatrix} \begin{pmatrix} -1 & -1 & -1 \\ 1 & -2 & 1 \\ 1 & 1 & 1 \end{pmatrix} \begin{pmatrix} -1 & -1 & 1 \\ -1 & -2 & 1 \\ 1 & 1 & 1 \end{pmatrix}$$

2.7. Matrices de relación espacial

Se trata ahora de estudiar las relaciones espaciales entre los píxeles con tonos de gris parecidos. Una relación espacial de parámetros (r, s) viene dada por la siguiente aplicación entre píxeles:

$$R_{(r,s)}: (i, j) \longrightarrow (i+r, j+s),$$

donde r y s son dos números enteros.

Si consideramos el entorno $N_8(i, j)$, podemos establecer 8 relaciones diferentes, según el píxel del entorno que le asociemos al píxel (i, j) . Así, representaremos por $R_{(1,0)}$ la relación

$$(i, j) \longrightarrow (i+1, j)$$

y continuando en el sentido contrario de las agujas del reloj, la relación $R_{(1,-1)}$ es:

$$(i, j) \longrightarrow (i+1, j-1)$$

Dada una relación R , representaremos por $h_R(p, q)$ el número de pares de píxeles (i, j) y (i', j') tales que:

1. $(i, j)R(i', j')$, es decir, (i, j) está relacionado con (i', j')
2. $f(i, j) = p$ y $f(i', j') = q$

Se define la **matriz de relación espacial** asociada a la relación R entre píxeles como la matriz $L \times L$ cuyo elemento (p, q) viene dado por $h_R(p, q)$, es decir la matriz

$$h(f; R) = \begin{pmatrix} h_R(0,0) & \dots & h_R(0,s) & \dots & h_R(0,L-1) \\ \dots & \dots & \dots & \dots & \dots \\ h_R(r,0) & \dots & h_R(r,s) & \dots & h_R(r,L-1) \\ \dots & \dots & \dots & \dots & \dots \\ h_R(L-1,0) & \dots & h_R(L-1,s) & \dots & h_R(L-1,L-1) \end{pmatrix}$$

donde se supone que el rango de los niveles de gris viene dado por el conjunto $\{0, 1, 2, \dots, L-1\}$.

El análisis de un conjunto de estas matrices, correspondientes a un conjunto de relaciones espaciales, nos dará, generalmente, información sobre las aristas y sobre la *textura* de la imagen original. Una textura es una región de la imagen que presentan una cierta regularidad estadística, de manera que está construida a partir de una misma ventana, llamada *téxel*, cuya repetición sucesiva, sujeta a pequeñas variaciones estadísticas, genera la imagen texturada.

Para evitar problemas en el tiempo de computación y de almacenaje, se suelen considerar sólo conjuntos reducidos de relaciones de adyacencia.

EJEMPLOS:

a) Consideremos la siguiente imagen cuyo rango de niveles de gris es el conjunto $\{0,1\}$:

$$\begin{pmatrix} 1 & 0 & 1 & 0 & 1 \\ 0 & 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 1 \\ 0 & 1 & 0 & 1 & 0 \end{pmatrix}$$

Si utilizamos la relación espacial $R_{(2,0)}$ obtenemos la siguiente matriz de relación espacial:

$$\begin{pmatrix} 6 & 0 \\ 0 & 6 \end{pmatrix}$$

Como los elementos de la matriz principal acaparan todos los valores positivos (no nulos) ya que los demás son cero, es decir, los tonos de gris de los pares de píxeles de la relación son iguales, lo que nos indica la existencia de una textura generada por dicha relación. Si deseamos buscar una textura mas gruesa utilizaríamos una relación con mayor separación como, por ejemplo, la relación R_{20} , que nos conduce a la siguiente matriz de relación espacial:

$$\begin{pmatrix} 0 & 4 \\ 4 & 0 \end{pmatrix}$$

que nos indica que no hay textura según dicha relación, ya que los valores no nulos no están mayoritariamente sobre la diagonal principal.

Si utilizamos la relación espacial $R_{(1,1)}$ obtenemos la siguiente matriz de relación espacial:

$$\begin{pmatrix} 4 & 0 \\ 0 & 4 \end{pmatrix}$$

que nos indica que no hay aristas de -45° ya que todos los elementos de fuera de la diagonal principal son nulos.

b) Consideremos ahora la siguiente imagen cuyo rango de niveles de gris es el conjunto $\{0,1,2,3,4\}$:

$$f = \begin{pmatrix} 4 & 0 & 4 & 0 & 4 & 1 \\ 0 & 3 & 0 & 4 & 1 & 3 \\ 4 & 1 & 4 & 1 & 3 & 0 \\ 0 & 4 & 0 & 3 & 0 & 3 \end{pmatrix}$$

Si consideramos la relación $R_{(1,1)}$ obtenemos la siguiente matriz de relación espacial:

$$\begin{pmatrix} 3 & 2 & 0 & 0 & 0 \\ 1 & 2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 2 & 1 \\ 0 & 0 & 0 & 1 & 3 \end{pmatrix}$$

donde se puede observar como los valores positivos están entorno a la diagonal principal, lo que refleja el hecho de que hay muy pocas fluctuaciones en los tonos de gris en la dirección de 45° , es decir, ausencia de ejes de -45° .

Análogamente, para la relación espacial $R_{(1,0)}$ se obtiene la matriz de relación espacial,

$$\begin{pmatrix} 0 & 0 & 0 & 3 & 4 \\ 0 & 0 & 0 & 2 & 1 \\ 0 & 0 & 0 & 0 & 0 \\ 3 & 0 & 0 & 0 & 0 \\ 3 & 4 & 0 & 0 & 0 \end{pmatrix}$$

que indica muchas altas variaciones (horizontales) entre los tonos de gris de un píxel y su vecino de la derecha, y eso ocurre cuando hay aristas verticales, pero en este caso sólo podemos hablar de pequeñas aristas verticales ya que no tienen continuidad y se van alternando.

Por lo tanto, para saber si hay aristas verticales podemos utilizar la relación $R_{(1,0)}$. Para saber si hay aristas horizontales la relación $R_{(0,1)}$ y para las aristas de 45° la relación $R_{(1,1)}$. Bastará con observar el valor de los elementos de la matriz de relación espacial cuyo índice (p, q) superan un cierto valor umbral. Finalmente, para saber si puede haber una textura tenemos que utilizar la relación espacial adecuada y observar los elementos que están entorno a la diagonal principal que deberán corresponder a valores altos con respecto a los demás términos.

donde \bar{r} es el valor medio de la distribución,

$$\bar{r} = \frac{1}{MN} \sum_{k=1}^L r_k n_k$$

Cuando todos los píxeles presentan el mismo nivel de gris su varianza es cero, y cuando mayor diferencia hay entre todos los posibles pares de píxeles entonces la varianza es mayor, pues si sustituimos la media por su valor y desarrollamos la expresión anterior de la varianza se llega a la expresión siguiente:

$$\sigma^2 = \frac{1}{MN(MN-1)} \sum_{(i,j) \neq (r,s)} ((f(i,j) - \bar{r}))^2$$

y así podemos ver una clara interpretación de la varianza como medida del contraste de una imagen.

Por otra parte, como los valores $\{p_k = n_k/MN; \quad k=1,2,...,L\}$ constituyen una distribución de probabilidad sobre el conjunto de los tonos de gris $\{r_1, r_2,...,r_L\}$, pues

$$\sum_{k=1}^L p_k = 1,$$

podemos utilizar la **entropía** de la variable aleatoria que toma los valores $\{r_1, r_2,...,r_L\}$, con probabilidades respectivas $\{p_1, p_2,...,p_L\}$, como una medida de contraste:

$$E = -\sum_{k=1}^L p_k \ln p_k.$$

Si $p_1=1$ y $p_k=0, k=2,3,...,L$, entonces todos los píxeles presentan el mismo nivel de gris, con valor r_1 , y la entropía vale 0, mientras que cuando la distribución de los tonos de gris de los píxeles es uniforme, es decir, $p_k = 1/L$, entonces la entropía alcanza su valor máximo que es $\ln(L)$, que corresponde a una imagen con máximo contraste. Esto sugiere que una medida normalizada en el intervalo $[0,1]$ del contraste de una imagen sea $E/\ln(L)$. Obsérvese que la entropía es también una medida de incertidumbre que cuando vale cero corresponde a máxima información y al mismo tiempo mínimo contraste, mientras que para una imagen con distribución uniforme, que corresponde al máximo contraste, la incertidumbre o falta de información es también máxima.

2.8 Técnicas de manipulación y modelado del contraste de la imagen: Modificación de los tonos de gris

Muchas veces nos encontramos con imágenes obtenidas bajo condiciones de escasa luminosidad o iluminación no uniforme que tienen poco contraste. Para el realzado de una imagen hay que realizar

un conjunto de transformaciones sobre los tonos de gris de los pixeles de la imagen; por ello, vamos a estudiar primeramente varias transformaciones sobre los tonos de gris que nos van a permitir el modelado de los tonos de gris para conseguir los fines deseados.

Sea T una función (transformación puntual) con rango (valores) en el conjunto $\{r_1, r_2, \dots, r_L\}$. Esta función transforma el nivel de gris pixel a pixel. La imagen transformada es, por tanto:

$$g(i, j) = T(f(i, j)), \quad i=1,2,\dots,M, j=1,2,\dots,N,$$

o lo que es lo mismo,

$$s_k = T(r_k), \quad k=1,2,\dots,L.$$

La función más sencilla es la transformación lineal, que viene dada por la expresión general:

$$g(i, j) = a + b \cdot f(i, j)$$

Si $r_p \leq f(i, j) \leq r_q$ para $i=1,2,\dots,M, j=1,2,\dots,N$, entonces la transformación lineal más adecuada viene determinada por los valores a y b siguientes:

$$b = \frac{r_L - r_1}{r_q - r_p}, \quad a = r_1 - \frac{r_L - r_1}{r_q - r_p} r_p$$

que pasa por el punto (r_p, r_1) y tiene de pendiente b , como puede verse en la figura 4.1. Esta transformación aumenta el contraste de la imagen, pues estira los valores de los tonos de gris de la imagen hasta ocupar el rango completo, $[r_1, r_2]$.

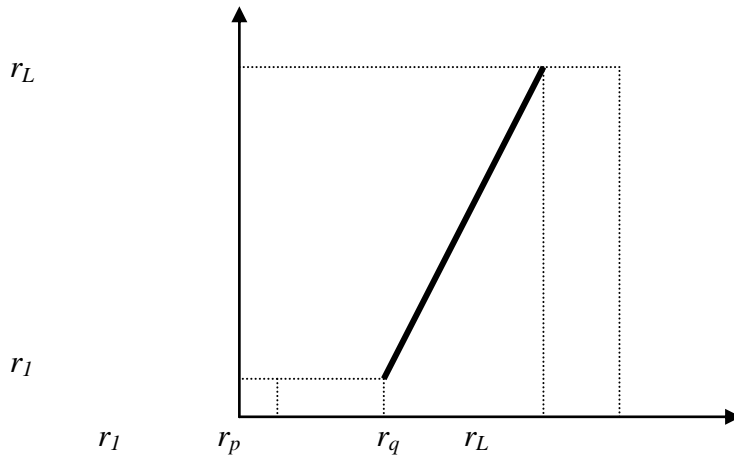


Figura 4.1. Transformación lineal.

Obsérvese que si $f(i, j) = r_k$ entonces el nuevo valor del nivel de gris del pixel (i, j) , que representaremos por s_k , viene determinado por la expresión:

$$s_k = \frac{r_L - r_1}{r_q - r_p} (r_k - r_p) + r_1$$

El término a desplaza los tonos de gris en un valor constante, actuando por tanto sobre el brillo de la imagen; según sea $a > 0$ ó $a < 0$ se incrementa o se disminuye el brillo de la imagen de cada pixel en la misma cantidad, a . La pendiente b de la recta es el parámetro que controla el contraste de la imagen;

si $b > 1$ se produce un incremento del contraste, para $0 < b < 1$ se produce una disminución del contraste, para $-1 < b < 0$ se obtiene una imagen inversa con menos contraste y para $b < -1$ se obtiene la imagen inversa con más contraste.

4.2.1 Transformaciones lineales a tramos: Extensión del contraste

Consideremos la imagen digitalizada con L tonos de gris, que por simplicidad van a ser los valores $0, 1, 2, \dots, L-1$, es decir,

$$u = f(x, y) \in \{0, 1, \dots, L-1\}$$

Un tipo de transformación un poco más general que la anterior para mejorar el contraste es la siguiente:

$$v = \begin{cases} \alpha u & \text{si } 0 \leq u \leq a \\ \beta(u - a) + v_a & \text{si } a \leq u < b \\ \gamma(u - b) + v_b & \text{si } b \leq u \leq L-1 \end{cases}$$

donde u representa el tono o nivel de gris de la imagen original y v el nivel de gris de la imagen transformada .

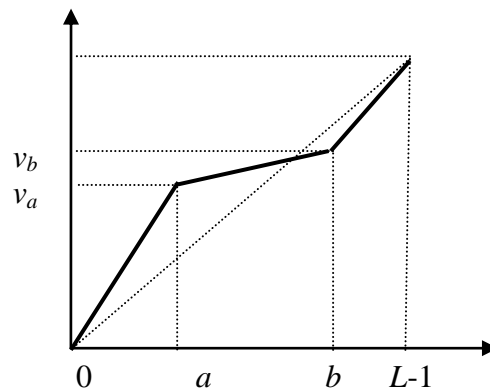


Figure 4.2. Transformación de tramos lineales

Con esta transformación realizamos una modificación de los tonos de gris de los píxeles de la imagen en función de los parámetros α , β y γ . La transformación de la figura 4.2 mejora el contraste en el tramo $[0, a]$, pues $\alpha > 1$, y también mejora el contraste en el tramo $[b, L-1]$ pues $\gamma > 1$, a costa de reducir el contraste en el tramo $[a, b]$, por ser $\beta < 1$. Por lo tanto, dicha transformación mejora el contraste de los píxeles más oscuros y, también, el de los píxeles más claros.

Un caso especial de transformación de tramos lineales es cuando alguno de los parámetros, α , β ó γ , es nulo, entonces se consigue un *recorte* de la imagen. Por ejemplo, cuando se sabe que los valores de la imagen están en el intervalo $[a, b]$ es conveniente utilizar dicha transformación con $\alpha = \gamma = 0$ pues se consigue el aumento del contraste y la reducción de ruido.

Cuando $\alpha = \gamma = 0$ y $a = b$ entonces se obtiene la transformación umbral

$$v = \begin{cases} 0 & \text{si } u \leq a \\ L-1 & \text{si } u > a \end{cases}$$

que produce una imagen binaria.

Cuando se desea detectar o distinguir aquellas regiones de la imagen cuyos valores de gris están dentro del intervalo $[a,b]$ se utiliza la transformación (ver figura 4.3):

$$v = \begin{cases} L-1 & \text{si } a \leq u \leq b \\ 0 & \text{en otro caso} \end{cases}$$

donde el fondo de la imagen es negro y lo forman aquellos pixeles con tono de gris fuera del intervalo $[a,b]$. Se obtiene así una imagen segmentada. Si, por el contrario, se desea conservar la imagen fuera del intervalo $[a,b]$ entonces se debe utilizar la transformación (ver figura 4.4).

$$v = \begin{cases} L-1 & \text{si } a \leq u \leq b \\ u & \text{en otro caso} \end{cases}$$

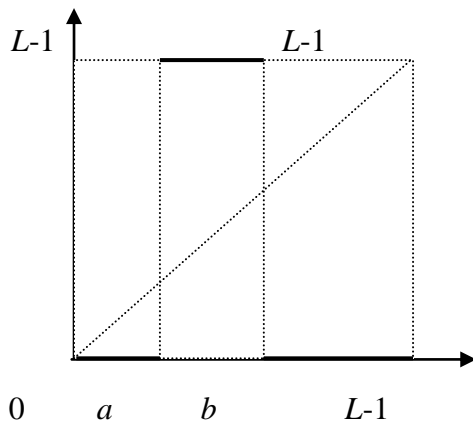


Figura 4.3. Segmentación sin fondo

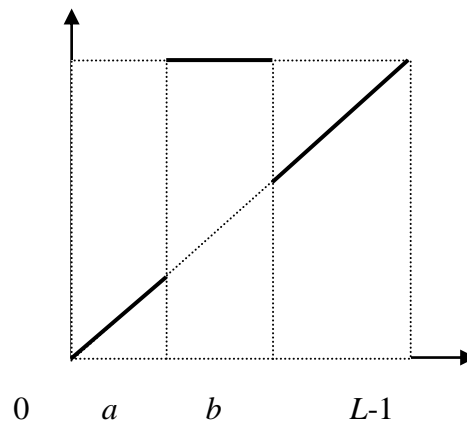


Figura 4.4. Segmentación con fondo

Cuando se desea mejorar el contraste de las partes más claras de la imagen a costa de las partes más oscuras, o al revés, se puede utilizar la transformación "goma de borrar" (ver la figura 4.5). La transformación lineal $v = (L-1) - u$ nos conduce al *negativo* de la imagen original.

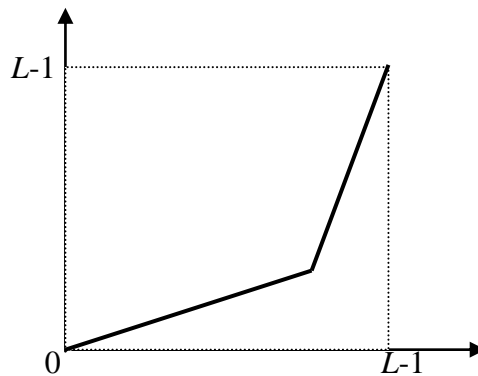


Figura 4.5. Transformación goma de borrar.

2.9. Transformaciones no lineales

Una de las transformaciones no lineales más utilizadas es la transformación gaussiana (ver figura 4.6) que viene dada por la función:

$$g(i, j) = \frac{\phi\left(\frac{f(i, j) - 0.5}{\sigma\sqrt{2}}\right) + \left\lceil \frac{0.5}{\sigma\sqrt{2}} \right\rceil}{\phi\left(\frac{0.5}{\sigma\sqrt{2}}\right)}$$

donde los corchetes en la expresión $[x]$ representan la parte entera de x , y

$$\phi(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-y^2} dy$$

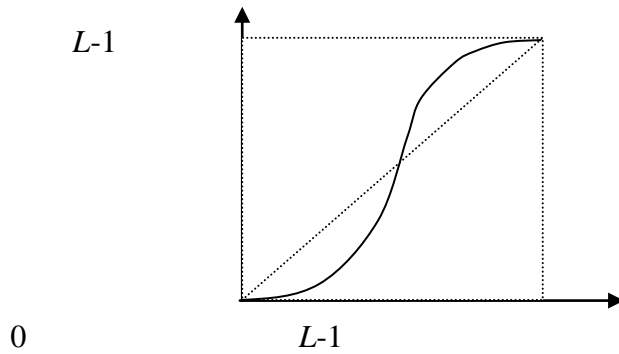


Figura 4.6. Transformación gaussiana.

Esta transformación aumenta el contraste de la imagen al hacer más oscuras las partes oscuras y más claras las partes claras (figura 4.6).

También podemos usar la función potencia:

$$g(i, j) = c[f(i, j)]^p,$$

donde podemos tomar $c=(L-1)^{1-p}$, para que asigne el valor $L-1$ al nivel $L-1$. Para valores de p menores que 1 esta transformación aumenta el contraste de las partes oscuras de la imagen, mientras que para valores mayores que 1 aumenta el contraste de las partes claras.

Otra transformación interesante es la *hiperbolización*, es decir, la transformación:

$$T^*(d_B) = \log(d_B + c)$$

donde

$$d_B = c \left(e^{\left[\log(1+1/c) \int_0^{d_A} p(d) dd \right]} - 1 \right)$$

y c es una constante que depende de la visión del ojo.

2.10 Igualación del histograma.

Hemos visto que los valores $\{p_k = n_k/MN; k=1,2,...,L\}$ constituyen una distribución de probabilidad sobre el conjunto de los tonos de gris $\{x_0, x_1, ..., x_{L-1}\}$, de manera que tenemos así la variable aleatoria X que toma los valores $\{x_0, x_1, ..., x_{L-1}\}$ con probabilidades respectivas $\{p_0, p_1, ..., p_L\}$. La representación gráfica de esta distribución se conoce con el nombre de **diagrama de barras** (ver figura 4.7), aunque en la bibliografía de procesamiento de imágenes se le llama histograma, pero el histograma es la representación gráfica de variables estadísticas continuas agrupadas en intervalos de clases y utiliza rectángulos cuya superficie es proporcional a las frecuencias. No obstante, nosotros también lo llamaremos histograma, para evitar confusión.

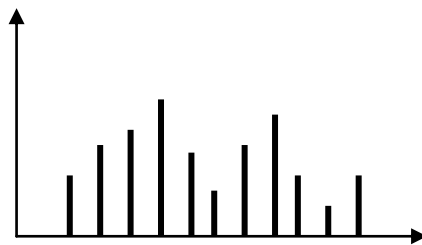


Figura 4.7. Diagrama de barras

Si utilizamos la entropía como medida de contraste hemos visto que la imagen con histograma uniforme es la que tiene mayor entropía (contraste). Por lo tanto, buscamos así la transformación T que nos lleve a una nueva imagen con histograma uniforme o casi uniforme. Sabemos, que si X fuera una variable aleatoria continua con función de densidad $p_X(x)$ entonces la variable aleatoria $U = F_X(X)$ tiene una distribución uniforme en el intervalo $[0,1]$, siendo $F_X(\cdot)$ la función de distribución de X , dada por la expresión:

$$F_X(x) = P(X \leq x) = \int_0^x p_X(s) ds,$$

En efecto,

$$P(U \leq u) = P(X \leq F_X^{-1}(u)) = F_X(F_X^{-1}(u)) = u.$$

Obsérvese que hemos utilizado que $F_X(x) \leq u \Leftrightarrow x \leq F_X^{-1}(u)$, que es cierto cuando X es continua. En el caso de que la variable aleatoria X sea discreta (nuestro caso) entonces se tiene que $P(F_X(X) \leq u) \leq u$ en lugar de $P(F_X(X) \leq u) = u$. En tal caso la distribución de U es aproximadamente uniforme. De hecho, para la variable aleatoria X que toma los valores $\{x_1, x_2, ..., x_L\}$ con probabilidades respectivas $\{p_1, p_2, ..., p_L\}$, y cuya función de distribución es

$$F(x) = \sum_{i \leq k} p_i, \quad \text{si } x_k \leq x < x_{k+1},$$

la transformación $U=F(X)$ es también una variable aleatoria discreta cuyos valores son $\{p_1, p_1 + p_2, p_1 + p_2 + p_3, \dots, 1\}$ con probabilidades respectivas $\{p_1, p_2, \dots, p_L\}$, que no tiene por qué ser uniforme. Además, se tiene que

$$P(U \leq u) = \sum_{i \leq k} p_i, \quad \text{para} \quad \sum_{i \leq k} p_i \leq u < \sum_{i \leq k+1} p_i,$$

en lugar de $P(U \leq u) = u$.

Por lo tanto, si suponemos que los valores de los tonos de gris son $0, 1, 2, \dots, L-1$, y deseamos conseguir una transformación cuyo histograma sea aproximadamente uniforme en los valores $0, 1, 2, \dots, L-1$, vamos a utilizar la transformación siguiente:

$$y = \frac{F_X(x) - F_X(x_0)}{1 - F_X(x_0)}(L-1),$$

siendo x_0 el menor valor que toma la variable aleatoria X con $p_0 > 0$. Esta transformación corresponde a la proyección de los valores $F(x_0), F(x_1), \dots, F(x_{L-1})$, del intervalo $[0, 1]$ sobre el intervalo $[0, L-1]$, como se muestra en la figura 4.3.7.

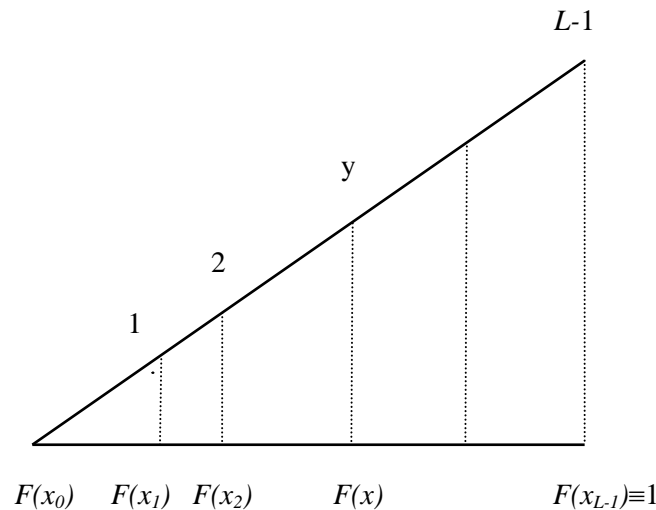


Figura 4.8. Proyección de los valores del intervalo $[0, 1]$ al intervalo $[0, L-1]$.

Como los valores de y de los niveles de gris deben de ser valores enteros hacemos la siguiente modificación de la transformación anterior para obtener una mejor aproximación:

$$y = \left\lceil \frac{F_X(x) - F_X(x_0)}{1 - F_X(x_0)}(L-1) + 0.5 \right\rceil, \quad (4.2.1)$$

Los corchetes $\lceil \cdot \rceil$ indican la parte entera.

Ejemplo 1: Consideremos una imagen digital de tamaño 64×64 con 8 tonos de gris ($L=8$), cuya distribución por tonos de gris es la siguiente:

X= Nivel de gris: 0	1	2	3	4	5	6	7	
Frecuencia absoluta n_i :	745	1320		940	753	40	0	122 176
Frecuencia relativa p_i :	0.18	0.32		0.23	0.19	0.01	0	0.03 0.04

Su función de distribución es:

$$F(x) = \begin{cases} 0 & \text{si } x < 0 \\ 0.18 & \text{si } 0 \leq x < 1 \\ 0.50 & \text{si } 1 \leq x < 2 \\ 0.73 & \text{si } 2 \leq x < 3 \\ 0.92 & \text{si } 3 \leq x < 4 \\ 0.93 & \text{si } 4 \leq x < 6 \\ 0.96 & \text{si } 6 \leq x < 7 \\ 1 & \text{si } x \geq 7 \end{cases}$$

Si utilizamos la transformación 4.2.1 para transformar los tonos de gris, con $p_0 = 0.18$ y $L=8$, tenemos:

$$s_0 = \left[\frac{F(0) - 0.18}{1 - 0.18} (8 - 1) + 0.5 \right] = 0,$$

y análogamente, $s_1=3$, $s_2=5$, $s_3=6$, $s_4=6$, $s_5=6$, $s_6=7$ y $s_7=7$, que conduce a la siguiente distribución de frecuencias:

X= Nivel de gris:	0	1	2	3	4	5	6	7		
Frecuencia absoluta n_i :	175		0		0	1320	0	940	793	298
Frecuencia relativa p_i :	0.18		0		0	0.32	0	0.23	0.20	0.07

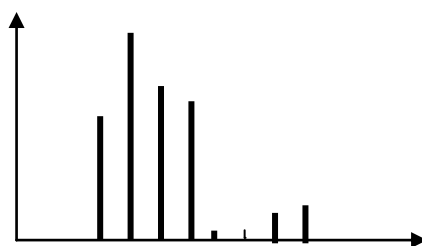
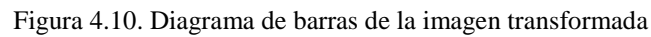


Figura 4.9. Diagrama de barras de la imagen original



Una manera diferente de modificar los tonos de gris para tratar de conseguir un histograma uniforme consiste en ordenar los píxeles de menor a mayor según su tono de gris. Se forman L grupos, siendo L en números de niveles o tonos de gris; el primer grupo estará constituido por los $[M \times N / L]$ píxeles con menor tono de gris, a los que se les asigna el nivel de gris (tono) 1; el segundo grupo estará constituido por los $[M \times N / L]$ píxeles con menor tono de gris del conjunto restante, y así sucesivamente. Cuando $M \times N / L$ no sea un número entero habrá que agrupar los píxeles en conjuntos de distinto tamaño, unos con $[M \times N / L]$ píxeles y otros con $[M \times N / L] + 1$ píxeles.

Consideremos la imagen siguiente que tiene 8 tonalidades:

Ordenamos los píxeles según sus tonalidades y le asignamos las nuevas tonalidades,

obteniéndose la imagen siguiente:

4	5	5	5	6
4	4	6	6	7
0	0	3	3	7
0	0	3	2	7
1	1	1	2	2

Cuando hay que elegir un píxel entre varios que tienen el mismo tono de gris se puede hacer siguiendo algún criterio de proximidad. Así, este método permite obtener muchas imágenes diferentes, todas ellas con un histograma uniforme o casi uniforme. Esto, aunque puede parecer un inconveniente, es una ventaja pues no conduce a la posibilidad de elegir la mejor de todas las obtenidas.

2.10 Realzado de una imagen en el dominio espacial

Ahora vamos a estudiar transformaciones de imágenes basadas en filtros para mejorar la calidad de la imagen acentuando los contornos de los objetos. Una imagen digital $f(m,n)$ se puede poner como una suma de dos imágenes; una imagen borrosa caracterizada porque sus bordes están poco perfilados (carencia de bordes) $f_B(m,n)$ y una imagen constituida por los bordes de los objetos $f_A(m,n)$. La imagen borrosa corresponde a una imagen filtrada con un filtro de paso-baja y la imagen de bordes a la imagen filtrada con un filtro de paso-alta, es decir,

$$f(m,n) = f_B(m,n) + f_A(m,n)$$

Por lo tanto, si a una imagen original le restamos su versión borrosa obtenida a partir de un filtro de paso-baja obtenemos una imagen donde sólo aparecen los bordes, es decir, la versión obtenida mediante un filtro de paso-alta. Así, si deseamos mejorar la calidad de una imagen, es decir, acentuar o enfatizar sus bordes, para que los objetos que la forman se vean mejor, basta con sumarle la imagen de bordes ponderada por un coeficiente λ :

$$f(m,n) + \lambda f_A(m,n)$$

donde el parámetro $\lambda > 0$ nos permite controlar el grado de acentuación de los bordes de la imagen.

Veamos dicha idea con más detalle. Si transformamos una imagen digital $f(m,n)$ mediante un operador de filtrado que utiliza una plantilla o máscara $g(m,n)$ obtenemos una nueva imagen dada por la expresión:

$$f^*(m,n) = \sum_{(i,j) \in N} f(m+i, n+j) h(i,j)$$

que nos proporciona como valor de la imagen en el píxel (m,n) una combinación lineal (media aritmética ponderada) de los valores de los píxeles de un entorno N ponderados con los valores correspondientes de la plantilla h . Por ejemplo, en el caso de utilizar una plantilla 3×3 , dada por la matriz

$$h = \begin{pmatrix} h(-1,1) & h(0,1) & h(1,1) \\ h(-1,0) & h(0,0) & h(1,0) \\ h(-1,-1) & h(0,-1) & h(1,-1) \end{pmatrix}$$

tenemos que

$$\begin{aligned}
f^*(m,n) = & f(m-1,n+1)h(-1,1) + f(m,n+1)h(0,1) + f(m+1,n+1)h(1,1) + \\
& f(m-1,n)h(-1,0) + f(m,n)h(0,0) + f(m+1,n)h(1,0) + \\
& f(m-1,n-1)h(-1,-1) + f(m,n-1)h(0,-1) + f(m+1,n-1)h(1,-1)
\end{aligned}$$

Si una imagen no tiene los contornos bien perfilados, una manera de aumentar su contraste es acentuando los mismos. Como acabamos de ver, basta con añadirle a la imagen una imagen de bordes obtenida a partir de la imagen original mediante un filtro de paso-alta. Es decir, la imagen

$$f^*(m,n) = f(m,n) + \lambda g(m,n)$$

donde λ es una cantidad positiva que regula el grado de acentuación de los bordes y $g(m,n)$ es una imagen de bordes de la imagen original, es decir, una imagen obtenida por un filtro de paso-alta.

Teniendo en cuenta que si filtramos una imagen con una plantilla con todos sus elementos cero salvo el elemento central que vale 1 obtenemos la misma imagen, y que la suma de dos imágenes que corresponden a filtraciones de una imagen con plantillas diferentes es igual que la imagen filtrada con la plantilla suma de las dos plantillas, es decir,

$$\begin{aligned}
f^*(m,n) &= \sum_{(m,n) \in N} f(m+i,n+i)h_1(i,j) + \sum_{(m,n) \in N} f(m+i,n+i)h_2(i,j) \\
&= \sum_{(m,n) \in N} f(m+i,n+i)[h_1(i,j) + h_2(i,j)]
\end{aligned}$$

resulta que si h es una plantilla de paso-alta entonces el filtrado con la plantilla

$$h = \begin{pmatrix} h(-1,1) & h(0,1) & h(1,1) \\ h(-1,0) & 1+h(0,0) & h(1,0) \\ h(-1,-1) & h(0,-1) & h(1,-1) \end{pmatrix}$$

produce directamente un realzado de la imagen original acentuando los bordes.

Un ejemplo de plantilla para el realzado de una imagen es la siguiente:

$$\begin{pmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix} + \begin{pmatrix} -\frac{1}{6} & -\frac{2}{3} & -\frac{1}{6} \\ -\frac{2}{3} & \frac{10}{3} & -\frac{2}{3} \\ -\frac{1}{6} & -\frac{2}{3} & -\frac{1}{6} \end{pmatrix} = \begin{pmatrix} -\frac{1}{6} & -\frac{2}{3} & -\frac{1}{6} \\ -\frac{2}{3} & \frac{13}{3} & -\frac{2}{3} \\ -\frac{1}{6} & -\frac{2}{3} & -\frac{1}{6} \end{pmatrix}$$

que suele llamarse plantilla para el enmascaramiento de la falta de perfilado en la imagen (*unsharp masking and crispening*).