Convolución

Extracto de "Visión por Computador. Fundamentos y métodos". Arturo de la Escalera. Prentice Hall, 2001

Copia para el alumno con fines didácticos

4.1. Procesamiento espacial

El procesamiento espacial lo constituyen aquellas técnicas que operan directamente sobre los valores de los pixeles de la imagen. Serán transformaciones del tipo:

$$S(x,y)=F(I(x,y))$$

siendo I(x,y) la imagen original, S(x,y) la resultante y F la transformación. Estas transformaciones suelen aplicarse a un entorno cuadrado del píxel.

4.1.1. Convoluciones.

Para el caso unidimensional, se define como convolución de la función $f(x)$ respecto a la función $h(x)$ una nueva función $g(x)$ tal que:

$$g(x) = h(x) * f(x) = \int_{i=-\infty}^{i=\infty} f(i)\, h(x-i)\, di$$

Si se tratara del caso discreto con dos secuencias f(x) y h(x) se obtendría una secuencia de salida g(x):

$$g(x) = h(x) * f(x) = \sum_{i=-\infty}^{i=\infty} f(i)\, h(x-i)$$

En la figura 4.1 puede observarse la convolución entre dos funciones:

$$f(x) = \begin{cases} 1 & 0 < x < 4 \\ 0 & \text{en cualquier otro caso} \end{cases}$$

$$h(x) = \begin{cases} x & -1 < x < 1 \\ 0 & \text{en cualquier otro caso} \end{cases}$$

Lo primero es obtener la función $h(-x)$ simétrica respecto al eje x. Esta señal se irá desplazando por el eje x, y el valor de la convolución para cada valor de x será el producto del área común a las dos señales. Así:

$-1<x<1$

$$g(x) = \int_{0}^{x+1} h(x-i)di = \int_{0}^{x+1} (x-i)di = \frac{1}{2}(x^2-1)$$

$1<x<3$

$$g(x) = \int_{x-1}^{x+1} (x-i)di = 0$$

$3<x<5$

$$g(x) = \int_{x-1}^{4} (x-i)di = -\frac{x^2}{2} + 4x - \frac{15}{2}$$

La señal $g(x)$ irá decreciendo conforme f(x) y h(x) se vayan solapando, hasta alcanzar un mínimo cuando toda la parte negativa haya entrado. El efecto de la parte positiva de h(x) implica que el valor de g(x) aumenta hasta hacerse cero. Conforme su parte negativa abandona a f(x), los valores serán positivos hasta alcanzar un máximo, tendiendo de nuevo a cero.

Para el caso de imágenes digitales se usan convoluciones bidimensionales discretas, cuya fórmula es:

$$g(x,y) = h(x,y) * f(x,y) = \sum_{i=-\infty}^{i=-\infty} \sum_{j=-\infty}^{j=\infty} f(i,j)\, h(x-i, y-j)$$

Lo más normal es usar convoluciones de 3x3 elementos. Entonces la expresión anterior puede concretarse en:

$$g(x,y) = h(x,y) * f(x,y) = \sum_{i=0}^{i=2} \sum_{j=0}^{j=2} f(i,j)\, h(x-i, y-j)$$

que por ejemplo para g(2,2) será:

$$\begin{aligned} g(2,2) = \ & \sum_{i=0}^{i=2} \sum_{j=0}^{j=2} f(i,j) h(2-i, 2-j) = \\ & f(0,0)h(2,2) + f(0,1)h(2,1) + f(0,2)h(2,0) + \\ & f(1,0)h(1,2) + f(1,1)h(1,1) + f(1,2)h(1,0) + \\ & f(2,0)h(0,2) + f(2,1)h(0,1) + f(2,2)h(0,0) \end{aligned}$$

En el caso de la visión por computador, en lugar de hablar de convolucionar la imagen f(x,y) con la transformación h(x,y), se habla de convolucionar con una máscara que tiene la forma:

h(0,0)	h(0,1)	h(0,2)
h(1,0)	h(1,1)	h(1,2)
h(2,0)	h(2,1)	h(2,2)

Para obtener h(-i,-j) hay que realizar dos giros. Así, se obtiene h(i,-j):

h(2,0)	h(2,1)	h(2,2)
h(1,0)	h(1,1)	h(1,2)
h(0,0)	h(0,1)	h(0,2)

Y h(-i,-j)

h(2,2)	h(2,1)	h(2,0)
h(1,2)	h(1,1)	h(1,0)
h(0,2)	h(0,1)	h(0,0)

La porción inicial de la imagen correspondiente es:

f(0,0)	f(0,1)	f(0,2)
f(1,0)	f(1,1)	f(1,2)
f(2,0)	f(2,1)	f(2,2)

y el valor del píxel de la imagen transformada será el mismo que se obtuvo antes.

Gráficamente se trata de poner las dos matrices, una encima de la otra, multiplicar cada celda por la inferior y sumar los productos. Para obtener el siguiente valor de g(x,y) la máscara h(x,y) se desplaza un elemento a la derecha y se repite la operación. Al acabar las columnas se vuelve al principio a una fila inferior, y así sucesivamente hasta llegar al último píxel.

De la figura 4.1 puede deducirse que el rango de la señal g(x) será siempre más grande que el de f(x). En el ejemplo f(x) toma valores distintos de cero entre 1 y 4 mientras que g(x) está entre -1 y 5. En el caso de imágenes no es así porque, por convenio, la operación se realiza para todos los pixeles de la imagen menos para los pixeles del borde, en los que no es posible aplicar la convolución al no solaparse totalmente la máscara y la imagen. Se tiene así un marco de anchura dos pixeles cuyo valor se suele tomar igual a cero. En el caso general, siempre que se realice una convolución se obtendrá una imagen más pequeña de dimensiones (N-(n-1))x(M-(m-1), donde N y M son las dimensiones de la imagen original y n y m las de la máscara.

En el caso de la señal de salida g(x) en 4.1 se observa que presenta dos picos, uno positivo y otro negativo, en los dos extremos de la señal f(x). Ha servido por lo tanto para detectar los bordes de esa señal, o cuándo la señal f(x) pasa de tomar un valor constante a otro. Es de esperar que si se aplica una convolución parecida en el caso bidimensional dé resultados análogos. En la figura 4.2 se ve el caso de aplicar la siguiente máscara:

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

Queda ahora ver como se representan los píxeles con valores negativos. Como en la pantalla todos los píxeles deben ser positivos, primero se normalizan los resultados entre –128 y 127, para después desplazar el origen 128 posiciones a la derecha. Así en el caso de la figura 4.2 derecha, pixeles oscuros corresponden a valores negativos, los claros a positivos, y los intermedios a valores en torno al cero.

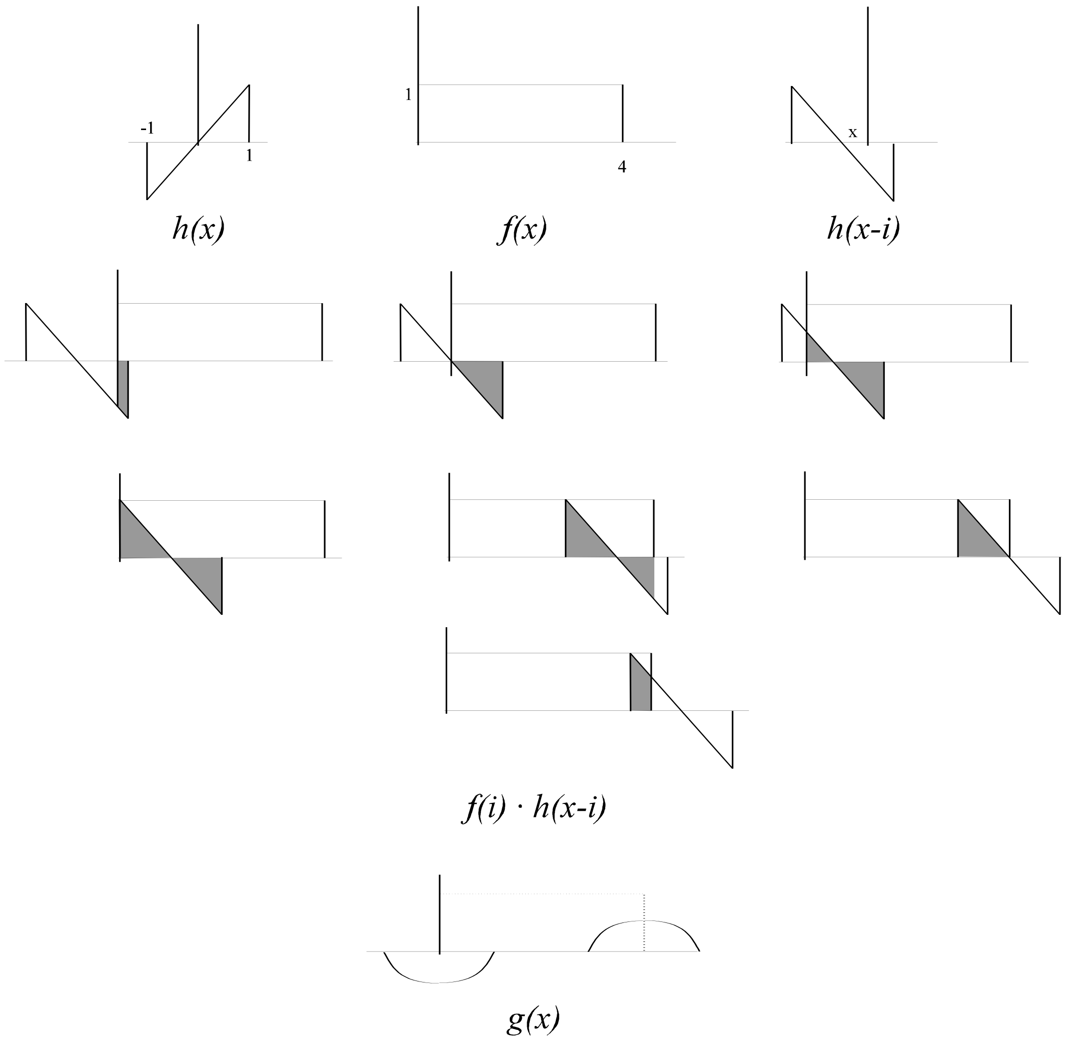

Figura 4.1. Convolución unidimensional.

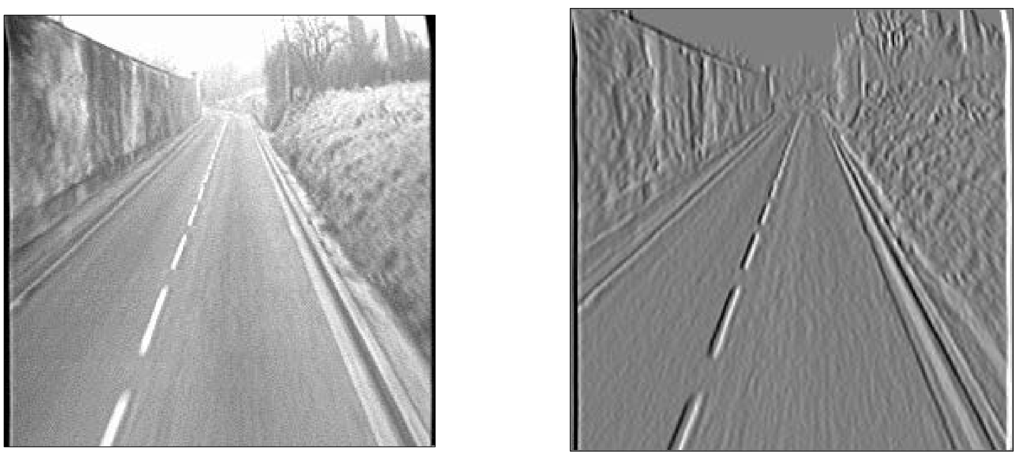

Figura 4.2. Convolución de una imagen digital.