

ANEXO I. Transformada Wavelet

I.1 Introducción

La transformada wavelet discreta es una herramienta que se ha popularizado enormemente en el mundo científico. Centrándonos en el campo de procesamiento de teoría de señal, las wavelet unifican en una misma teoría diferentes técnicas, como son el procesamiento de señal multiresolución, usado en visión por computador, la codificación en subbandas, utilizada para la compresión de voz e imagen, o las expansiones en series de wavelet útiles en el campo de la matemática aplicada.

La idea de observar una señal en varias escalas y analizarla usando varias resoluciones se ha venido aplicando en diferentes campos de las matemáticas e ingeniería. No obstante, la normalización del concepto de multiresolución es relativamente reciente. La llevaron a cabo a mediados de los ochenta un matemático, un físico teórico y un geofísico (Meyer, Grossman y Morlet) y marcó el comienzo de la teoría de las wavelet. La estructura multiresolución de las wavelet permite tanto un estudio local de la señal, para un análisis de sus detalles, así como un estudio global de la misma, similar al realizado por el análisis de Fourier [Trenas00].

La transformada de una señal $x(t)$ es una operación matemática que da como resultado la representación de la misma señal en un dominio diferente. Lógicamente, en la elección de una transformada discreta intervienen factores como las propiedades de la misma y su grado de adecuación a la aplicación en concreto, así como la carga computacional. Por ejemplo, la transformada rápida de Fourier permite realizar cálculos de forma más eficiente en el dominio de la frecuencia que en el temporal, en ciertos casos.

Además, la transformación de una señal, al ofrecer una representación alternativa, revela a menudo características esenciales de la misma que en el dominio original son muy difíciles, o imposibles, de detectar. Así, la existencia de múltiples periodicidades, patrones de fase o espectrales, son características del dominio de la frecuencia útiles en el campo de la clasificación y detección.

Otra importante aplicación de las transformadas es la compresión de datos. Si se trata de imágenes, por ejemplo, lo que buscamos es una matriz de transformación que dé lugar a una imagen transformada dispersa y en la que los elementos de mayor magnitud (los de mayor energía, por tanto) se concentren mayoritariamente en una pequeña región.

Las wavelets dan lugar a representaciones multigrad, puesto que mantienen la ortogonalidad respecto a sus dilataciones y traslaciones. Sus características de localidad tiempo-frecuencia resultan útiles en el análisis y procesado de señales no estacionarias. Debido a la propiedad de reconstrucción perfecta y la ausencia de degradación perceptual en los bordes, resultan adecuadas en aplicaciones de codificación de vídeo.

Por otro lado, los coeficientes wavelet presentan un decaimiento en su magnitud que es el más rápido de entre todas las bases ortonormales, lo que los hace muy dispersos. Por esta razón las bases wavelet resultan ser óptimas para la compresión de datos, reducción de ruido, y estimación estadística. También son numerosas las aplicaciones existentes en el área de las telecomunicaciones y el procesado de la señal, debido a la naturaleza multifrecuencia de la transformada.

Son una excelente alternativa a los sistemas de bases ortogonales tradicionales.

I.2 Limitaciones de la transformada de Fourier

La transformada ortonormal de Fourier de una señal proporciona una descripción de la distribución de la energía de la señal con respecto a la frecuencia, es decir, su espectro [Trenas00].



FIGURA I.1. Transformada de Fourier.

La transformada nos informa del contenido de la señal en términos de unas funciones base del dominio transformado. La representación de la señal en el dominio de Fourier es única, y consiste en una suma continua de sinusoides de diferentes amplitudes, frecuencias y fases, las funciones base ortonormales, de acuerdo con la siguiente expresión:

$$X(\omega) = \int_{-\infty}^{+\infty} x(t) e^{-j\omega t} dt \quad (I.1)$$

El valor transformado $X(\omega)$ de la ecuación (I.1) representa la media en el tiempo del contenido de energía de la señal a una cierta frecuencia ω . Otra forma de ver esta ecuación es que el coeficiente $X(\omega)$ informa de la cantidad de $e^{j\omega t}$ que $x(t)$ contiene. Por lo tanto, se puede obtener como la correlación cruzada de $x(t)$ con $e^{j\omega t}$, es decir, la proyección de $x(t)$ sobre las funciones base.

La transformada inversa de Fourier sintetiza $x(t)$ a partir de las funciones base $e^{j\omega t}$, con amplitud compleja $X(\omega)$:

$$x(t) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} X(\omega) e^{j\omega t} d\omega \quad (I.2)$$

Sin embargo, numerosas señales son no-estacionarias. La potencia y/o el espectro de la señal pueden variar con el tiempo. Así por ejemplo, en la voz, las frecuencias de los formantes son diferentes en diferentes segmentos de una cierta palabra. Del mismo modo, en el caso de señales transitorias con una duración finita, el proceso de promediado puede dar lugar a que se genere un espectro complicado en el que se oscurezca el hecho de que la señal sólo ha estado presente durante un intervalo de tiempo finito.

Por lo tanto, la transformada de Fourier no es la herramienta idónea para el análisis de señales no estacionarias. Para su completa caracterización en el dominio de la frecuencia se hace necesario incluir como parámetro al tiempo, es decir, se debe realizar un análisis tiempo-frecuencia de la señal.

I.2.1 Análisis de Fourier de señales no estacionarias

Para el análisis de señales no-estacionarias resulta ser más adecuada la utilización de la transformada de *Fourier localizada* (STFT, Short Time Fourier Transform) de *Gabor*, en la cual se aplica una función de enventanado $v(t)$ antes de realizar la transformación. Dicha transformación viene dada por la expresión:

$$X_{STFT}(\omega, \tau) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} x(t) v(t - \tau) e^{-j\omega t} dt \quad (1.3)$$

La ventana acorta el intervalo temporal sobre el que se está promediando, de forma que el análisis de Fourier se está realizando sobre un intervalo temporal pequeño centrado en el instante de tiempo τ .

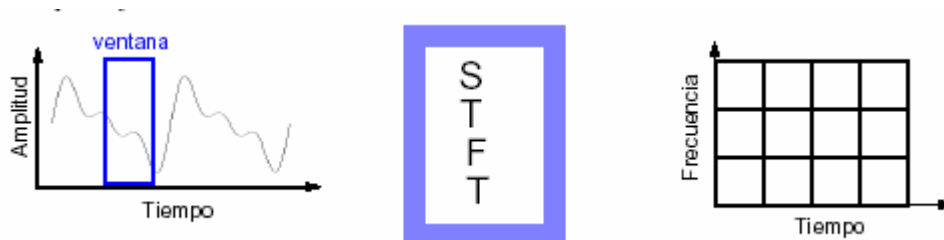


FIGURA I.2. Transformada STFT para análisis de señales no estacionarias.

La STFT también puede ser considerada como la implementación de un banco de filtros. A una cierta frecuencia, la expresión vista para la STFT está realizando el filtrado de la señal a lo largo de todo el tiempo, con un filtro paso banda cuya respuesta impulso es la función ventana $v(t)$, modulada a esa frecuencia.

En la mayoría de las implementaciones prácticas, $v(t)$ es una señal de duración finita que tiene la mayor parte de su energía concentrada en las bajas frecuencias. La duración y ancho de banda de $v(t)$ limitan tanto la resolución temporal, $\delta(t)$, como la resolución en frecuencia, $\delta(\omega)$ de la transformada, puesto que cualquier función de ventana debe cumplir la relación $\delta(t)\delta(\omega) \leq 0,5$ para ser realizable. Esta inecuación, que es equivalente al principio de incertidumbre de Heisenberg de la física cuántica, indica que la resolución temporal/espectral disponible va a estar necesariamente limitada para algunas señales, independientemente de la ventana elegida. Por un lado el ancho (T) de la

ventana debe ser suficientemente grande como para proporcionar la resolución espectral deseada, pero también debe ser suficientemente estrecha como para que no se dejen de observar los eventos dependientes del tiempo. Si la señal contiene dos impulsos a una distancia de d segundos, T debe ser inferior a d segundos.

Pero aunque una ventana muy estrecha, idealmente un impulso, proporciona una resolución (localización) temporal perfecta, su resolución espectral sería muy baja puesto que su ancho de banda sería infinito. Por otro lado, un filtro muy estrecho proporciona una buena localización espectral, pero una localización temporal muy pobre, puesto que su respuesta impulsiva no decae rápidamente con el tiempo. Las sinusoides son perfectamente locales en el dominio de la frecuencia, pero tienen una duración infinita (un soporte no compacto).

Su empleo como funciones base en el análisis de Fourier se fundamenta en la utilización de las cancelaciones para la representación (síntesis) de las discontinuidades en el tiempo. Por lo tanto, las sinusoides no van a resultar tan eficientes para la representación de funciones con soporte compacto en comparación con otras funciones base que sí tengan esta propiedad. Con el término de eficiencia nos estamos refiriendo al número de coeficientes requerido en el dominio transformado para representar una función dada.

La STFT ha sido ampliamente utilizada en el ámbito del procesado de la voz, particularmente para la generación de espectrogramas. El espectrograma representa las magnitudes del espectro utilizando como ejes los valores de frecuencia/tiempo para los que se han evaluado. Existen espectrogramas de banda ancha y banda estrecha, según se quiera tener una mayor resolución temporal o espectral.

Otros ámbitos de aplicación del análisis tiempo-frecuencia van desde la geofísica a la biomedicina. En cualquiera de los casos, nos enfrentamos a la imposibilidad de mejorar simultáneamente las resoluciones temporal y frecuencial, puesto que si se quiere alcanzar una cierta resolución en el tiempo o la frecuencia, hay que utilizar diferentes anchos de la ventana. Por otro lado, la mayor anchura posible para la ventana viene, muy a menudo, impuesta por el mayor intervalo durante el cual puede suponerse que la señal es estacionaria.

Uno de los problemas fundamentales de la STFT, es que esta carece de algunas de las propiedades que hacen de la transformada de Fourier una herramienta de análisis tan atractiva: **unicidad y ortonormalidad**.

I.3 La transformada Wavelet continua

La transformada wavelet continua (CWT) es una alternativa a la STFT. La CWT combina la aplicabilidad al análisis tiempo-frecuencia, propia de la STFT, con las propiedades matemáticas que eran tan interesantes en la transformada de Fourier. La transformada wavelet proyecta una función del tiempo (la señal) en una función dependiente de dos parámetros: a y τ (en vez de ω y τ). El parámetro a es denominado *escala*, mientras que τ es *traslación*. Suponemos que la señal $x(t)$ es tal que $x(t) \in L^2(\mathfrak{R})$, y por lo tanto:

$$\int x^2(t) dt < \infty \quad (1.4)$$

La CWT de $x(t)$ es:

$$X_{CWT}(a, \tau) = \frac{1}{\sqrt{a}} \int_{-\infty}^{+\infty} x(t) \psi\left(\frac{t-\tau}{a}\right) dt \quad (1.5)$$

Donde $\psi(t)$ es la **función wavelet madre** y las funciones-base vienen representadas por $\psi\left(\frac{t-\tau}{a}\right) \frac{1}{\sqrt{a}}$ y constituyen el resultado de las dilataciones y traslaciones a que es sometida $\psi(t)$, son conocidas como **familia de las wavelet**.



FIGURA I.3. Transformada Wavelet.

La función $\psi(t)$ puede ser real o compleja, con lo que la transformada wavelet resultante será real o compleja, respectivamente. Cuando $\psi(t)$ es compleja, hay que utilizar su complejo conjugado $\psi^*(t)$ en la ecuación (1.5). Realizando el cambio de variable $at' = t$, la ecuación se convierte en:

$$X_{CWT}(a, \tau) = \sqrt{a} \int_{-\infty}^{+\infty} x(at') \psi\left(t' - \frac{\tau}{a}\right) dt' \quad (1.6)$$

Esta ecuación nos muestra la equivalencia entre escalar $\psi(t)$ en (1.5) o escalar $x(t)$ en (1.6). La transformación puede interpretarse también como la implementación de un banco de filtros no uniforme (y continuo en el tiempo) en el que se mantiene una relación constante entre la resolución en frecuencia y en el tiempo, dependiendo de a . Valores grandes de a dan lugar a una mayor resolución espectral, mientras que los valores pequeños proporcionan mejor resolución temporal.

Puede demostrarse que la CWT, al contrario que la STFT, permite una reconstrucción perfecta de la señal, sin ninguna ambigüedad, a partir de los coeficientes de la transformada.

$$x(t) = \frac{1}{C_\psi} \int_{-\infty}^{+\infty} \int_{a>0} X_{CWT}(a, \tau) \cdot \frac{1}{\sqrt{a}} \psi\left(\frac{t-\tau}{a}\right) \frac{1}{a^2} d\tau da \quad (1.7)$$

Donde C_ψ es una constante que depende únicamente de $\psi(t)$ y a es positivo.

Para garantizar la reconstrucción perfecta, la constante debe cumplir:

$$C_\psi = \int_0^{+\infty} \frac{|\Psi(\omega)|^2}{\omega} d\omega < \infty \quad (1.8)$$

Para que $C_\psi < \infty$, la transformada de Fourier de $\psi(t)$ debe ser tal que $|\Psi(\omega)| < \infty$ para cualquier ω y $\Psi(0) = 0$, o lo que es igual:

$$\int \psi(t) dt = 0 \quad (1.9)$$

En la ecuación (I.7), tanto a como τ son variables continuas, con lo que la representación de $x(t)$ mediante la CWT resulta redundante. Además, en la práctica, para la CWT sólo se puede utilizar un número finito de valores de (a, τ) . Un algoritmo discreto para aproximar la CWT se obtiene muestreando a y τ , de tal forma que $a = a_0^j > 0$ y $\tau = a_0^j k \tau_0$ donde a_0 y τ_0 son los intervalos de muestreo y $j, k \in \mathbb{Z}$. Lo usual es que se utilicen los valores 2 y 1 para a_0 y τ_0 respectivamente. Sustituyendo estos valores de a y τ en la ecuación se obtiene la transformada Wavelet discreta (DWT):

$$X_{DWT}(j, k) = \int_{-\infty}^{+\infty} x(t) \cdot \psi_{j,k}(t) dt \quad (\text{I.10})$$

Donde $\psi_{j,k} = 2^{-j/2} \psi(2^{-j}t - k\tau_0)$, $\psi_{0,0}(t) = \psi(t)$. Obsérvese que tanto $x(t)$ como $\psi(t)$ siguen siendo funciones continuas.

Pero la DWT, a diferencia de la CWT, no siempre determina una reconstrucción perfecta de la señal bajo las condiciones que se especificaron con anterioridad. Sin embargo, la reconstrucción de la señal puede conseguirse con un error despreciable siempre que existan unas constantes A, B tales que $0 < A \leq B < \infty$ y para cualquier señal $x(t) \in L^2(\mathfrak{R})$ se verifique:

$$A \|x(t)\|^2 \leq \sum_{j,k} \left| \langle x(t) \psi_{j,k}(t) \rangle \right|^2 \leq B \|x(t)\|^2 \quad (\text{I.11})$$

La familia de funciones base se denomina estructura y las constantes A y B son las cotas de la estructura. En la práctica lo deseable es que las funciones base sean ortogonales, en cuyo caso $A=B=1$, asegurándose así la reconstrucción perfecta a partir de los coeficientes DWT.

En el esquema de la página siguiente podemos ver el contraste entre los cuatro dominios transformados de los que hemos hablado.

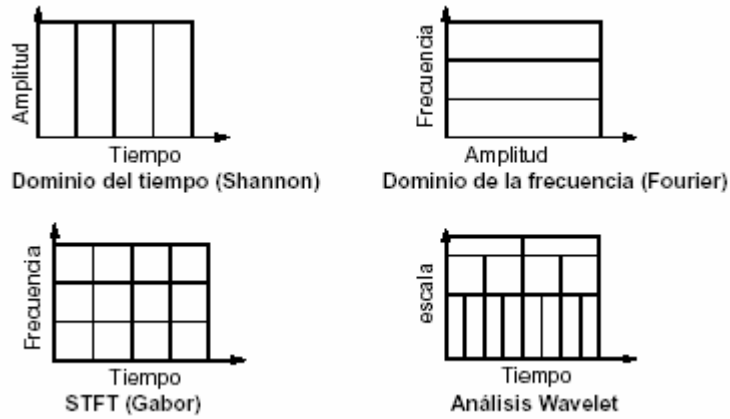


FIGURA I.4. Diferentes dominios de representación de señales comentados.

I.4 La transformada wavelet bidimensional

Como en el caso unidimensional, una base ortonormal separable para $L^2(\mathfrak{R})$ se construye utilizando productos separables de una función de escala ϕ y una wavelet ψ . La función de escala está asociada a un espacio multirresolución de aproximación $\{V_j\}_{j \in \mathbb{Z}}$.

Supongamos que $\{V_j\}_{j \in \mathbb{Z}}$ es el espacio multirresolución bidimensional separable determinado por $V_j^2 = V_j \otimes V_j$, donde \otimes indica el producto tensorial de los espacios.

Supongamos que W_j es el espacio de detalle que es complemento ortogonal del espacio de aproximación V_j^2 en el espacio de mayor resolución V_{j-1}^2 :

$$V_{j-1}^2 = V_j^2 \oplus W_j^2 \tag{I.12}$$

Con el objetivo de proporcionar una *base ortonormal de wavelets* para $L^2(\mathfrak{R})$, el siguiente teorema construye una *base de wavelets* para cada espacio de detalle W_j^2 .

Teorema. Sea ϕ una función de escala y ψ la función wavelet correspondiente, que genera una base ortonormal de $L^2(\mathfrak{R})$. Se definen tres wavelets:

$$\psi^1(x, y) = \phi(x)\psi(y), \quad \psi^2(x, y) = \psi(x)\phi(y), \quad \psi^3(x, y) = \psi(x)\psi(y) \quad (I.13)$$

Y para $1 \leq k \leq 3$ denominamos:

$$\psi_{j,n,m}^k(x, y) = 2^{-j/2} \psi^k \left(\begin{matrix} x - 2^j n & x - 2^j m \\ 2^j & 2^j \end{matrix} \right) \quad (I.14)$$

Siendo la familia de wavelets:

$$\left\{ \psi_{j,n,m}^1(x, y), \psi_{j,n,m}^2(x, y), \psi_{j,n,m}^3(x, y) \right\}_{(n,m) \in \mathbb{Z}^2} \quad (I.15)$$

una base ortonormal de W_j^2 . Y (I.16) una base ortonormal de $L^2(\mathfrak{R})$.

$$\left\{ \psi_{j,n,m}^1(x, y), \psi_{j,n,m}^2(x, y), \psi_{j,n,m}^3(x, y) \right\}_{(j,n,m) \in \mathbb{Z}^3} \quad (I.16)$$

El algoritmo rápido para el cálculo de la DTWT bidimensional consiste en la aplicación independiente de filtros paso-bajo a filas y columnas de la imagen, como se puede ver en el siguiente esquema.

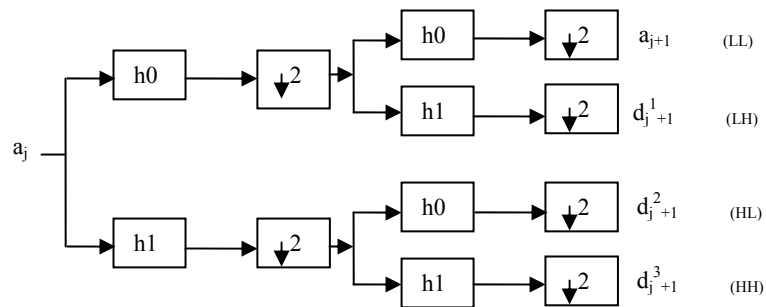


FIGURA I.5. Representa la implementación usando bancos de filtros de la transformada wavelet bidimensional estándar directa.

De esta forma, en el primer nivel obtenemos cuatro nuevas imágenes, correspondientes a los coeficientes $\{a_1\}, \{d_1^1\}, \{d_1^2\}, \{d_1^3\}$ (dependiendo del par de filtros que se empleó en su generación), que son cuatro veces más pequeñas que la inicial. Si se continua procesando, y se van permutando los resultados tras cada etapa de filtrado, separando las salidas del filtrado paso-alto de las del

paso-bajo, llegaríamos a una distribución de puntos como la que se muestra en la figura I.6.

En ella, $HL_j = \{d_j^2\}$, por ejemplo, es el resultado de un filtrado paso-alto de las filas, un paso-bajo de las columnas, en el nivel j de la pirámide. Puede observarse que, por tratarse de la DTWT estándar, sólo se continúan aplicando nuevas etapas de filtrado a los LL_j .

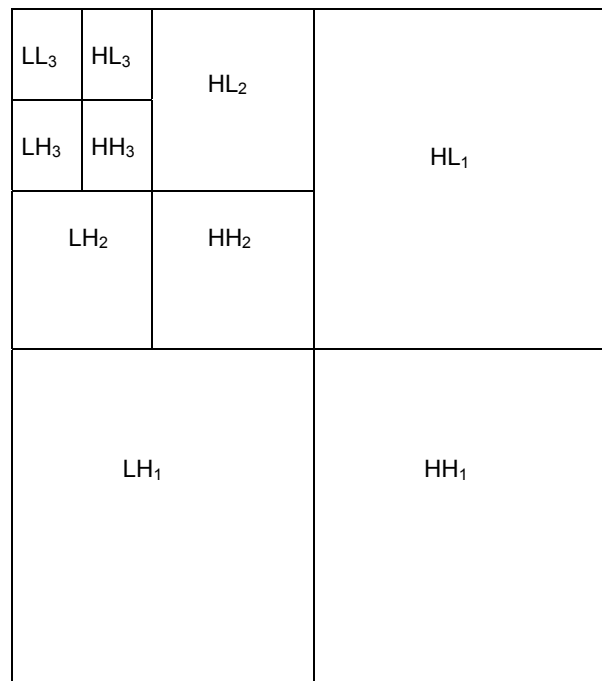


FIGURA I.6. Representación de las octavas de una DWT de tres niveles.

ANEXO II. Transformada discreta del Coseno (DCT)

II.1 Introducción

La aplicación de las transformadas al procesamiento digital de señal e imágenes ha aumentado considerablemente en los últimos años. Entre las más comúnmente utilizadas destaca la transformada discreta del coseno (DCT, *Discrete Cosine Transform*), que encuentra un uso muy extendido en aplicaciones de compresión de imágenes y de vídeo. Las propiedades únicas de la DCT permiten obtener altos índices de compresión a muy bajo coste. Por tanto, la DCT es la transformada utilizada en los actuales estándares de compresión de imágenes como JPEG.

El desarrollo de sistemas multimedia capaces de manejar video, voz, texto y gráficos hace que sea muy importante el uso de herramientas que permitan la manipulación eficiente de los datos multimedia.

La teoría de las transformadas ha supuesto un papel clave en el procesamiento de imágenes durante muchos años y continua siendo un tema de interés. La codificación por transformada desarrollada hace más de dos décadas, constituye un esquema de compresión de imágenes efectivo y la base de todos los estándares para compresión con pérdidas de información.

II.2 Transformada DCT de una dimensión

Vemos la transformada DCT en una dimensión, que resulta más simple en cuanto a notación y nos permite conocer aspectos relacionados con la representación en el dominio DCT, en general [Lim90].

Sea $x[n]$ una secuencia de N puntos, que es cero fuera del intervalo $0 \leq n \leq N-1$. Entre diferentes variaciones consideramos la conocida como DCT simétrica par, a menudo usada en aplicaciones de codificación de señal. Para obtener la DCT de la secuencia es conveniente convertir $x[n]$ de longitud N , en una secuencia $y[n]$ de $2N$ puntos a la que se le calcula la DTFT obteniéndose $Y(k)$, que se relaciona con $C_x(k)$, que es la DCT de N puntos de $x[n]$. Es decir:

$$x[n]_{N \text{ puntos}} \rightarrow y[n]_{2N \text{ puntos}} \rightarrow Y(k)_{2N \text{ puntos}} \rightarrow C_x(k)_{N \text{ puntos}}$$

Donde:

$$y[n] = \begin{cases} x[n] & 0 \leq n \leq N-1 \\ x[2N-1-n] & N \leq n \leq 2N-1 \end{cases} \quad (\text{II.1})$$

Siendo $y[n]$ es simétrica respecto al punto medio.

Si repetimos la secuencia $x[n]$ de forma periódica cada N puntos vemos que aparece una discontinuidad artificial al enlazar las partes inicial y final de la secuencia, sin embargo, si repitiéramos de la misma forma la secuencia $y[n]$ obtenida cada $2N$ puntos esa discontinuidad no aparecería, esta característica de eliminación de discontinuidades de borde artificiales contribuye a conseguir la propiedad de compactación de la energía característica de la DCT.

La DFT de $y[n]$ será:

$$Y(k) = \sum_{n=0}^{2N-1} y(n) \cdot W_{2N}^{kn}, \quad 0 \leq k \leq 2N-1 \quad (\text{II.2})$$

$$\text{Donde } W_{2N} = e^{-j\frac{2\pi}{2N}}$$

A partir de la que se obtiene la transformada DCT de N puntos de $x[n]$ como:

$$C_x(k) = \begin{cases} W_{2N}^{k/2} \cdot Y(k) & 0 \leq k \leq N-1 \\ 0 & e.o.c \end{cases} \quad (\text{II.3})$$

$$C_x(k) = \begin{cases} \sum_{n=0}^{N-1} 2x[n] \cos \frac{\pi}{2N} k(2n+1) & 0 \leq k \leq N-1 \\ 0 & e.o.c \end{cases} \quad (\text{II.4})$$

La ecuación (II.4) es la definición de DCT de $x[n]$. Para derivar la expresión de la DCT inversa procederíamos de forma similar al realizado para determinar la DCT directa, obteniéndose que:

$$x[n] = \begin{cases} \frac{1}{N} \sum_{k=0}^{N-1} w(k) \cdot C_x(k) \cos \frac{\pi}{2N} k(2n+1) & 0 \leq n \leq N-1 \\ 0 & e.o.c \end{cases} \quad (\text{II.5})$$

II.3 Transformada DCT en dos dimensiones

A partir del par de transformadas obtenido para el caso de la DCT en una dimensión podemos obtener su equivalente para 2 dimensiones sin más que realizar los mismos pasos ya vistos.

Resultando el siguiente par de transformadas:

$$C_x(k_1, k_2) = \left\{ \begin{array}{ll} \sum_{n_1=0}^{N_1-1} \sum_{n_2=0}^{N_2-1} 4x[n_1, n_2] \cos \frac{\Pi}{2N_1} k_1 (2n_1 + 1) \cdot \cos \frac{\Pi}{2N_2} k_2 (2n_2 + 1) & \begin{array}{l} 0 \leq k_1 \leq N_1 - 1 \\ 0 \leq k_2 \leq N_2 - 1 \end{array} \\ 0 & e.o.c \end{array} \right\}$$

(II.6)

$$x[n_1, n_2] = \left\{ \begin{array}{ll} \frac{1}{N_1 N_2} \sum_{n_1=0}^{N_1-1} \sum_{n_2=0}^{N_2-1} w_1(k_1) w_2(k_2) C_x(k_1, k_2) \cos \frac{\Pi}{2N_1} k_1 (2n_1 + 1) \cdot \cos \frac{\Pi}{2N_2} k_2 (2n_2 + 1) & \begin{array}{l} 0 \leq n_1 \leq N_1 - 1 \\ 0 \leq n_2 \leq N_2 - 1 \end{array} \\ 0 & e.o.c \end{array} \right\}$$

(II.7)

Algunas propiedades útiles de la transformada discreta del coseno son las de linealidad, de simetría y el hecho de que si la secuencia $x(n_1, n_2)$ se puede dividir en secuencias separables su transformada DCT también.

II.4 Conclusiones

Aparte de estas propiedades de manejo de la transformada indicadas, lo que realmente nos interesa para su aplicación en el campo de la codificación de imágenes es el hecho de que esta transformada permite la concentración de la mayor parte de la energía de la imagen en unos cuantos coeficientes transformados, aquellos correspondientes a la zona de baja frecuencia (aproximadamente un 16% del total de coeficientes) lo que es bueno para aplicaciones como compresión de imágenes así como para aquellas que requieren de la separación de las componentes más significativas de la señal del resto (aspecto en el que nos basamos para utilizar esta transformada en watermarking). Además (y relacionado con lo anterior) elimina la correlación existente entre filas y columnas, cualidad también importante en el tema objeto de estudio del presente documento.

ANEXO III. Métricas de calidad visual

Para aplicaciones en las que las imágenes son, en última instancia vistas por observadores humanos, el único método adecuado para cuantificar la calidad visual de la imagen es por medio de medidas subjetivas, en la práctica este tipo de medida no presenta ventajas, puesto que consume mucho tiempo y es costosa. El objetivo en este campo consiste en desarrollar una medida cuantitativa que pueda predecir de forma automática la calidad de la imagen percibida. Las métricas de calidad objetivas se pueden clasificar en función de la disponibilidad o no de la imagen original sin distorsionar con la que es comparada la imagen procesada, en nuestro caso mediante la inserción de una marca de agua. Así tendremos métricas de calidad con referencia completa, sin referencia o ciegas o bien con referencias reducidas, en las que no se dispone de la imagen original para comparar pero sí de algunas características de la misma utilizadas como información adicional para poder realizar la medida. En este aspecto, veremos solo métricas de calidad con referencia completa, es decir aquellas en las que la imagen original está disponible completamente [Kutter99], [Wang04].

La robustez de la marca de agua encajada en la imagen depende directamente de la capacidad de la técnica de inserción utilizada o medida en que la marca influye en la degradación de la imagen. Para un estudio más riguroso del rendimiento de las técnicas de sellado digital de imágenes con marcas de agua robustas, además de la respuesta de las mismas frente a diferentes ataques debe considerarse la degradación que sufre la imagen al ser marcada. En este apartado se revisan los criterios de distorsión basados en píxeles más usuales.

III.1 Métricas basadas en píxeles

La mayoría de las medidas de distorsión o métricas de calidad usadas en el procesamiento de información visual pertenecen al grupo de medidas de distorsión por diferencia. Todas estas medidas están basadas en la diferencia entre las señales original (no distorsionada) y la modificada. En la tabla I.1 se muestran algunas de las medidas más populares. La tabla I.2 muestra las medidas de distorsión basadas en la correlación entre las imágenes original y modificada [Kutter99].

Actualmente las medidas de distorsión más utilizadas en el campo de la codificación de imágenes y de video son, la relación señal-ruido (SNR) y la relación pico de señal-ruido (PSNR), comúnmente medidas en decibelios. Sin embargo, estas medidas de distorsión por diferencia no están relacionadas con las características del sistema de visión humana, lo que podría ser un problema a la hora de su aplicación para la evaluación de las características de calidad y robustez de las técnicas de watermarking empleadas, ya que estas técnicas a menudo explotan características del HVS para mejorar su robustez manteniendo su invisibilidad. Usar las métricas indicadas para cuantificar la distorsión causada por un determinado proceso de sellado con marcas de agua podría, por tanto, resultar en unas medidas cuantitativas de distorsión confusas y erróneas. Además estas técnicas se aplican, normalmente, a los canales de luminancia y crominancia de las imágenes. Si los métodos de sellado trabajan en el mismo espacio de color, por ejemplo la luminancia, no habría problema, pero si trabajan en un diferente espacio de color no sería adecuadas.

MÉTRICAS DE DISTORSIÓN POR DIFERENCIA	
Diferencia máxima	$MD = \max_{m,n} I(m,n) - \tilde{I}(m,n) $
Promedio del valor absoluto de la diferencia	$AD = \frac{1}{M \cdot N} \sum_{m,n} I(m,n) - \tilde{I}(m,n) $
Promedio normalizado del valor absoluto de la diferencia	$NAD = \frac{\sum_{m,n} I(m,n) - \tilde{I}(m,n) }{\sum_{m,n} I(m,n)}$
Error cuadrático medio	$MSE = \frac{1}{M \cdot N} \sum_{m,n} (I(m,n) - \tilde{I}(m,n))^2$
Error cuadrático medio normalizado	$NMSE = \frac{\sum_{m,n} (I(m,n) - \tilde{I}(m,n))^2}{\sum_{m,n} I(m,n)^2}$

Norma L^p	$\left(\frac{1}{M \cdot N} \sum_{m,n} I(m,n) - \tilde{I}(m,n) ^p \right)^{1/p}$
Error cuadrático medio laplaciano	$LMSE = \frac{\sum_{m,n} \left(\nabla^2 I(m,n) - \nabla^2 \tilde{I}(m,n) \right)^2}{\sum_{m,n} \left(\nabla^2 I(m,n) \right)^2}$
Relación señal - ruido	$SNR = \frac{\sum_{m,n} \left(I(m,n) \right)^2}{\sum_{m,n} \left(I(m,n) - \tilde{I}(m,n) \right)^2}$
Relación pico de señal - ruido	$PSNR = \frac{M \cdot N \max_{m,n} \left(I(m,n) \right)^2}{\sum_{m,n} \left(I(m,n) - \tilde{I}(m,n) \right)^2}$
Fidelidad de la imagen	$IF = \frac{1 - \sum_{m,n} \left(I(m,n) - \tilde{I}(m,n) \right)^2}{\sum_{m,n} \left(I(m,n) \right)^2}$

TABLA III.1. Métricas de distorsión visual basadas en la diferencia entre los píxeles de las imágenes original y distorsionada, donde I(m,n) representa un píxel de coordenadas (m, n) de la imagen original e $\tilde{I}(m,n)$ el equivalente de la imagen que contiene la marca de agua insertada [Kutter99].

MEDIDAS DE DISTORSIÓN POR CORRELACIÓN	
Correlación cruzada normalizada	$NC = \frac{\sum_{m,n} I(m,n) \cdot \tilde{I}(m,n)}{\sum_{m,n} \left(I(m,n) \right)^2}$
Calidad por correlación	$CQ = \frac{\sum_{m,n} I(m,n) \cdot \tilde{I}(m,n)}{\sum_{m,n} I(m,n)}$

TABLA III.2. Medidas de distorsión visual por correlación entre las señales original y marcada. $I(m,n)$ e $\tilde{I}(m,n)$ representan los píxeles (de coordenadas (m,n)) de las imágenes original y modificada respectivamente.

OTRAS MEDIDAS	
Contenido estructural	$SC = \frac{\sum_{m,n} (I(m,n))^2}{\sum_{m,n} (\tilde{I}(m,n))^2}$
Relación sigma señal-ruido global	$GSSNR = \frac{\sum_b \sigma_b^2}{\sum_b (\sigma_b - \tilde{\sigma}_b)^2}$ $\sigma_b = \sqrt{\frac{1}{P} \sum_{blockb} (I(m,n))^2 - \left(\frac{1}{P} \sum_{blockb} I(m,n)\right)^2}$
Relación sigma señal-ruido	$SSNR = \frac{1}{P} \sum_b SSNR_b \rightarrow SSNR_b = 10 \log_{10} \frac{\sigma_b^2}{(\sigma_b - \tilde{\sigma}_b)^2}$
Relación sigma-error	$SER_b = \frac{\sigma_b^2}{\frac{1}{P} \sum_{blockb} (I(m,n) - \tilde{I}(m,n))^2}$
Parecido de histogramas	$HS = \sum_0^{255} f_I(c) - f_{\tilde{I}}(c) $ <p>donde $f_I(c)$ es la frecuencia relativa del nivel c en imagen de 256 niveles</p>

TABLA III.3. Otras medidas del parecido entre las imágenes original y distorsionada.

ANEXO IV. Stirmark 3.1

En este apartado se indicarán nociones básicas referentes a la utilización de la versión 3.1 del banco de pruebas stirmark utilizado para analizar la robustez de las técnicas de sellado desarrolladas [Petitcolas99], [Petitcolas03].

IV.1 Aspectos generales

El proceso completo de ataque y detección puede implementarse mediante la utilización de ficheros por lotes.

Los formatos de imágenes con los que trabaja la utilidad son, JPEG, PPM Y PGM por lo que todos los demás formatos deberán ser convertidos a los mencionados cuando se quiera utilizar el banco de pruebas. En nuestro caso hemos trabajado con imágenes en formato JPEG.

IV.2 Versión 3.1

Las características de la versión 3.1 son:

- Posibilidad para seleccionar una categoría de pruebas única.
- Nuevas pruebas: distorsiones geométricas lineales generales, cambios en la relación de aspecto, filtrado de mediana de diferentes tamaños, borrado asimétrico de filas y columnas, nuevos valores por defecto para algunos tests ya desarrollados. Especialmente en la compresión JPEG.
- Opción de salvar y por tanto reutilizar parámetros de distorsiones geométricas.

IV.2.1 Opciones de línea de comandos

Las diferentes opciones para la línea de comandos son las que se indican a continuación.

-? Para conseguir ayuda

-b<float> El número de desplazamiento de píxel permitido en el centro de la imagen. Todos los píxeles son desplazados usando una función

de distribución. Los píxeles de las esquinas no se mueven mientras que los del centro son los que más se desplazan.

- d<float>** Valor máximo del byte por el que cualquiera de los valores RGB de cualquier píxel se permite que se desvíe del valor original (por defecto su valor es de 1.5).
- ll<n><float>** Esta opción es específica para el ataque de borrado Laplaciano que se aplica antes de una distorsión geométrica aleatoria. Se usa para especificar la fuerza LR para la escala 2^n .
- o<float>** Para indicar el número de píxeles distantes que se permite que la imagen objetivo esté fuera de la original (por defecto su valor es de 0.7). Los valores de muestra tomados desde fuera de la imagen original son extrapolados. Se puede especificar un porcentaje pero esto no es normalmente útil porque los valores superiores a la unidad provocan extrapolaciones poco útiles.
- PS<nombre_fichero>** Usado para salvar los parámetros de distorsión comunes en un fichero. Estos parámetros incluyen aquellos especificados desde la línea de comandos así como otros parámetros aleatorios o establecidos por defecto. Comando útil cuando queremos aplicar las mismas transformaciones a diferentes imágenes para poder comparar resultados
- PS+<nombre_fichero>** Salva los principales parámetros de distorsión y todos los números aleatorios usados en los procesos de distorsión geométrica. Crea ficheros de gran tamaño.
- PL<nombre_fichero>** Carga sólo los principales parámetros aleatorios salvados previamente con las opciones **-PS** o **-PS+**.
- PL+<nombre_fichero>** Carga todos los parámetros salvados previamente con la opción **-PS+**.

-
- q<int>** Para especificar el factor de calidad JPEG usado después de las distorsiones geométricas.
- NOJPEG** Para que stirmark no use compresión JPEG tras aplicar las distorsiones geométricas.
- r<float>** Indica la fracción de desplazamiento de píxel permitida por cualquier píxel. Durante el proceso de distorsión, cualquier píxel puede ser movido de forma aleatoria a lo largo de ambos ejes.
- s<int>** Valor de semilla para el generador de números aleatorios. Si se quieren hacer medidas sistemáticas con stirmark se deben repetir las cadenas con diferentes valores de semilla para conseguir resultados significativos en relación con la probabilidad de que stirmark elimine la posibilidad de recuperación de la marca de agua.
- PSNR<nombre>** Proporciona el valor de la PSNR del fichero de entrada con el fichero cuyo nombre se indica detrás de la opción PSNR.
- T<nombre>** Para realizar el test implementado por defecto en stirmark. Guarda los resultados en ficheros con el nombre indicado y un sufijo relacionado con el tipo de ataque implementado.
- S<int>** Se emplea únicamente con la opción -T. Se puede usar para seleccionar una determinada categoría de transformación, por ejemplo, rotación.