



UNIVERSIDAD NACIONAL AUTONOMA
DE MEXICO

FACULTAD DE INGENIERIA

**CODIFICACION DE IMAGENES UTILIZANDO
UNA DESCOMPOSICION SUB-BANDA POR MEDIO
DE FILTROS ESPEJO EN CUADRATURA (QMF)
NO SEPARABLES,**

FALLA DE ORIGEN

TESIS

QUE PARA OBTENER EL TITULO DE:

INGENIERO MECANICO - ELECTRICISTA
(Area Eléctrica - Electrónica)

PRESENTA:

ALBERTO VELAZQUEZ OSNAYA

Director: Dr. Francisco García Ugalde



MEXICO, D. F.

1995



Universidad Nacional
Autónoma de México

Dirección General de Bibliotecas de la UNAM

Biblioteca Central



UNAM – Dirección General de Bibliotecas
Tesis Digitales
Restricciones de uso

DERECHOS RESERVADOS ©
PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL

Todo el material contenido en esta tesis esta protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México).

El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo titular de los Derechos de Autor.

AGRADECIMIENTOS

Quisiera agradecer muy especialmente al Dr. Francisco García Ugalde por su continuo interés y su valiosa orientación en el desarrollo de este trabajo.

A todos mis compañeros y profesores de la Facultad de Ingeniería y de la División de Estudios de Posgrado por todo su apoyo y por sus valiosos consejos.

Contenido

1	Introducción	
1.1	Formación de una Imagen Digital	2
1.2	¿Porque Comprimir una Imagen?	3
1.3	Clasificación de las Técnicas de Compresión	4
1.4	Efectos de los Parámetros de Digitalización sobre la Compresión	5
1.5	Organización de la Tesis	6
2	Conceptos Básicos	
2.1	Modelo Visual	7
2.2	Revisión del Proceso de Muestreo	9
2.2.1	El Muestreo	9
2.2.2	El Teorema del Muestreo	10
2.2.3	Conversión de la Tasa de Muestreo	10
2.3	Cuantización	12
2.4	Entropía	14
2.5	Código de Huffman	15
3	Técnicas de Compresión de Imágenes	
3.1	La Evaluación de la Calidad en las Imágenes	17
3.1.1	Escalas Subjetivas	17
3.1.2	Mediciones Numéricas	18
3.2	Técnicas de Codificación de Imágenes	20
3.2.1	Codificación por PCM	20
3.2.2	Codificación Estadística	21
3.2.3	Codificación predictiva	21
3.2.4	Codificación por Transformada	22
3.3	Transformadas Sub-banda	24
3.3.1	Propiedades de las Transformadas Sub-banda	25
3.4	Ejemplos de Transformadas Sub-banda	26
3.4.1	La Transformada de Gabor	26
3.4.2	La Transformada del Coseno Discreta (TCD)	27
3.4.3	La Pirámide Laplaciana	27
3.4.4	Filtros Espejo en Cuadratura	28
4	Pirámides Ortogonales Para la Codificación de Imágenes	
4.1	La Pirámide Separable QMF	35
4.2	La Pirámide Quinconce	38

5 Selección y Simulación de un Esquema de Descomposición	
5.1 Implementación de la Pirámide Quinconce (Etapa de Transmisión)	43
5.1.1 Análisis	44
5.1.2 Cuantización	48
5.1.3 Codificación	48
5.2 Implementación de la Pirámide Quinconce (Etapa de Recepción)	49
5.2.1 Decodificación	50
5.2.2 Descuantización	50
5.2.3 Síntesis	50
6 Análisis y Evaluación de Resultados	
6.1 Caso de estudio No. 1	56
6.2 Caso de estudio No. 2	60
6.3 Caso de estudio No. 3	64
7 Conclusiones	
Bibliografía	

Capítulo 1

Introducción

Gracias al perfeccionamiento y disminución en el costo de los equipos como: "scanners", cámaras fotográficas, impresoras de texto y todo aquello relacionado directamente con el proceso de digitalización de imágenes, se ha generado una creciente demanda en el manejo de las mismas, ya que ahora se pueden realizar con una mayor facilidad los procesos de conversión de una imagen analógica a un formato digital. Actualmente es posible la adquisición digital de imágenes en una forma directa, debido a que los sensores y desarrollos electrónicos asociados a la adquisición se hacen cada vez más comunes; un ejemplo de ello es la aparición de diversos tipos de cámaras electrónicas en el mercado. Otros ejemplos de adquisición de imágenes en un formato digital, se presentan en diferentes modalidades dentro del campo de la medicina, como en la *Tomografía Computarizada (TC)* o en las *Imágenes de Resonancia Magnética (IRM)*. En este sentido, la creación de imágenes por computadora (sintéticas) esta siendo también una importante fuente de generación de datos digitales. La razón por la cual se tiene un gran interés en la manipulación de imágenes digitales es claro: el representar una imagen digitalmente permite que la información visual sea trabajada con mayor facilidad de manera óptima y en nuevas formas. Además, el crecimiento prácticamente exponencial en el área de computo, con respecto a la década pasada, ha dado como resultado la creación y aplicación de diversos sistemas digitales para imágenes en diferentes áreas de trabajo, como es el caso de la Astronomía, el sensado remoto, la Medicina, las Artes Gráficas, la publicidad y en algunos procesos de manufactura.

A pesar de las ventajas que se tienen con la representación digital de una imagen, existe un problema ocasionado por la gran cantidad de datos (bits) que son necesarios para su representación. Afortunadamente una imagen

representada en su forma canónica contiene generalmente una cantidad significativa de redundancia. Por lo que la compresión, entendiéndola como el medio de llevar a cabo una codificación eficiente de los datos en una imagen, tiene como propósito aprovechar la redundancia que esta contiene y reducir con ello el número de bits requeridos para su representación. Cabe mencionar que los resultados obtenidos de una compresión pueden ser significativos para el almacenamiento o transmisión de una imagen.

1.1 Formación de una Imagen Digital

La fuente para generar una imagen digital puede ser una escena tridimensional del mundo real, o una imagen generada previamente en dos dimensiones como en el caso de una fotografía. En cualquiera de estos dos casos la fuente de información es analógica, es decir, continua tanto en amplitud como en espacio. Ahora bien, para generar una imagen digital, lo primero que se tiene que hacer es muestrear la fuente en posiciones discretas usando algún tipo de sensor. A las muestras obtenidas del proceso anterior, se les conoce con el término de pixels o pels (contracción de picture element). Los valores de los pixels producidos por el sensor son continuos sobre un rango finito y tienen una relación lineal con la intensidad de energía radiante de cada localidad muestreada.

Existen diferentes maneras de realizar el muestreo de las posiciones anteriormente mencionadas; sin embargo, la más común es utilizar una rejilla rectangular equidistante. Idealmente cada muestra corresponde a una región infinitesimalmente pequeña de la fuente, pero a consecuencia de la propia naturaleza física de los sensores y de su óptica asociada, cada muestra constituye el valor integrado de un área finita de la fuente. El número de localidades muestreadas por unidad de área define lo que se conoce como *tasa de muestreo* (sampling rate) del sistema, dicha tasa debe ser elegida tomando en cuenta el Teorema de Muestreo de Nyquist. Si el criterio establecido por el teorema no es considerado, la imagen muestreada presentará un efecto de "aliasing" generado cuando las frecuencias espaciales sean mayores a $1/2$ de la tasa de muestreo y estas aparezcan como frecuencias más pequeñas. Para prevenir el "aliasing", puede realizarse un prefiltrado en el sistema con la finalidad de limitar el ancho de banda de la fuente antes de ser muestreada (la integración realizada por el sensor actúa también como un prefiltrado). En un sistema de digitalización de imágenes la tasa de muestreo se da frecuentemente en

términos de la resolución del "scanner" utilizado, siendo esta el inverso de la tasa de muestreo. En general la resolución del "scanner" requerida depende directamente de la aplicación. Por ejemplo, para una radiografía de 14 x 17 pulgadas observada por un radiólogo a una distancia típica de 14 pulgadas, una resolución de scanner de $70\mu\text{m}/\text{pixel}$ que alcanza a capturar frecuencias espaciales tan grandes como de 7 ciclos/mm es adecuada (esto se hace tomando en cuenta el sistema visual humano SVH). Para un negativo de 35 mm de una fotografía, sujeto a una ampliación y a determinado procesamiento, una resolución de scanner de $12\mu\text{m}/\text{pixel}$ es más común.

Continuando con el proceso de digitalización, a cada punto muestreado de valor continuo se le aplica un proceso de cuantización, obteniéndose con ello un número de valores discretos que representan la imagen digital. Tomando en cuenta que las imágenes digitalizadas serán observadas por una o varias personas en la mayoría de las aplicaciones, es importante realizar la cuantización en un dominio que vaya de acuerdo con el SVH de percepción. Por lo que, en muchos casos los valores de los pixels son sujetos a no-linealidades (logarítmicas o funciones de raíz cúbica) antes de ser cuantizados, consiguiéndose un ajuste o aproximación a las no-linealidades del SVH. El número de niveles de cuantización requerido para una adecuada representación de la imagen, depende también de la aplicación. Para documentos de texto binario, se requieren únicamente dos niveles (1 bit/pixel), es decir, cada punto muestra puede tomar uno de dos valores; por ejemplo, negro o blanco. Para escenas naturales o fotografías de tono continuo, es común usar 8 bits/pixel (256 niveles). Sin embargo, dependiendo del rango dinámico de la fuente y del tipo de salida del sensor (logarítmica o lineal), puede ser necesario emplear 10 o 12 bits/pixel. Consecuentemente y con el propósito de corregir los errores que se pueden cometer al implementar analógicamente las no-linealidades, los digitalizadores de imágenes más sofisticados, primero utilizan de 12 a 14 bits para cuantizar los valores de los pixels en un espacio lineal, y aprovechando la no linealidad de los datos digitales pueden entonces recuantizar en 8 bits.

1.2 ¿Porque Comprimir una Imagen?

La necesidad de aplicar un proceso de compresión a una imagen se observa desde el momento en que se calcula el número de bits que se obtienen como resultado de una tasa de muestreo típica y de diferentes esquemas de cuantización. Como ejemplos, se presentan los siguientes casos en los que se puede

valorar en forma global la cantidad de bytes necesarios para representar una imagen:

- Una imagen de video a color: 512 x 512 pixels/color, 8 bits/pixel y 3 colores $\Rightarrow \approx 6 \times 10^6$ bits,
- El negativo de una fotografía de 24 x 36 mm (35mm) barrido a $12\mu\text{m}$: 3000 x 2000 pixels/color, 8 bits/pixel y 3 colores $\Rightarrow \approx 144 \times 10^6$ bits,
- Una radiografía de 14x17 pulgadas adquirida a $70\mu\text{m}$: 5000 x 6000 pixels, 12 bits/pixel $\Rightarrow \approx 360 \times 10^6$ bits.

Obviamente, el almacenamiento de unas cuantas imágenes resulta ser un verdadero problema. Como otro ejemplo de la necesidad de realizar una compresión, se puede considerar el tiempo de transmisión al enviar una imagen de video con las mismas características que la presentada anteriormente, sobre una línea telefónica, utilizando un modem de 9600 bauds (bits/s). La transmisión de una sola imagen podría tardar aproximadamente 11 minutos, lo cual no es aceptable en la mayoría de las aplicaciones.

1.3 Clasificación de las Técnicas de Compresión

Es claro que en el proceso de digitalización de una imagen se obtiene un número de bits muy grande. Sin embargo, el número de bits requeridos para representar la información contenida en una imagen, se puede reducir substancialmente por la redundancia. En general, se pueden identificar tres tipos diferentes de redundancia en imágenes digitales:

- Redundancia espacial, es la correlación (o dependencia) que existe entre valores de pixels vecinos.
- Redundancia espectral, se refiere a la correlación entre diferentes planos de color o de diferentes bandas espectrales.
- Redundancia temporal, se debe a la correlación entre diferentes estructuras dentro de una secuencia de imágenes.

El objetivo principal de la compresión es reducir el número de bits requerido para representar una imagen, removiendo la redundancia en los datos.

Existen muchas formas de clasificar las diferentes técnicas de compresión de imágenes, pero de entre estas se pueden generalizar dos grupos fundamentales: *compresión sin pérdidas* y *compresión con pérdidas*.

En la compresión sin pérdidas la imagen reconstruida es numéricamente igual a la imagen original pixel por pixel.

En la compresión con pérdidas, la imagen reconstruida contiene degradaciones con respecto a la imagen original. Pero en términos generales, se consigue una mayor compresión con este tipo de técnicas. Es importante subrayar que las degradaciones que se obtienen a consecuencia de una compresión pueden ser visualmente aparentes o no serlo. En consecuencia el término visualmente sin pérdidas se ha empleado frecuentemente para caracterizar esquemas de compresión con pérdidas, consiguiéndose con ello resultados en los cuales las pérdidas no son visibles bajo condiciones normales de visibilidad. Desafortunadamente, el término visualmente sin pérdidas es bastante subjetivo, por lo que hay que tener precaución en su interpretación.

1.4 Efectos de los Parámetros de Digitalización sobre la Compresión

La redundancia presente en una imagen digital es altamente dependiente del sistema empleado, de la forma de la imagen y de los parámetros utilizados en su representación. En particular la tasa de muestreo, el número de niveles de cuantización y la presencia de una fuente y/o sensor con ruido, pueden de alguna forma afectar la compresión. En este sentido se han encontrado tendencias, como las que se presentan a continuación, que suceden generalmente en algoritmos de compresión, aún y cuando existan variaciones de una imagen a otra.

- Como consecuencia de un incremento en la tasa de muestreo, se tiene un aumento en la correlación de pixel a pixel, lo que permite obtener tasas de compresión mayores. La tasa de compresión se define como:

$$CR = \frac{\text{No. bits imagen original}}{\text{No. bits imagen comprimida}}$$

- El incremento en el número de niveles de cuantización reduce hasta un cierto límite la correlación que existe entre pixels, disminuyendo con ello la compresión obtenida.

- La presencia de una imagen ruidosa o la introducción de ruido por un sensor disminuyen la correlación entre pixels, provocando una reducción en la compresión.

1.5 Organización de la Tesis

En el capítulo 2 se presenta una revisión de conceptos básicos sobre comunicaciones y de algunos tópicos que son necesarios para una mejor comprensión de este trabajo.

El capítulo 3 presenta algunas de las diferentes técnicas que existen para realizar una compresión de imágenes, se hace referencia tanto a técnicas tradicionales como a técnicas que utilizan una descomposición en sub-bandas. Se resaltan las características de la descomposición en sub-bandas ya que son la base de este trabajo y se presenta una breve revisión de las estructuras QMF (filtros espejo en cuadratura).

El capítulo 4 muestra la aplicación de una descomposición en sub-bandas utilizando estructuras QMF con filtros separables y no separables. Las descomposiciones que se realizan permiten obtener una estructura pirámidal.

Bajo este contexto, es posible elegir una estructura de codificación por lo que en el capítulo 5 se hace la elección y simulación algorítmica de una estructura de codificación.

En el capítulo 6 se presenta el análisis de resultados en base a la compresión lograda y a la calidad subjetiva de la imagen.

Por último, el capítulo 7 presenta las conclusiones a las cuales se llegaron con la simulación realizada.

Capítulo 2

Conceptos Básicos

2.1 Modelo Visual

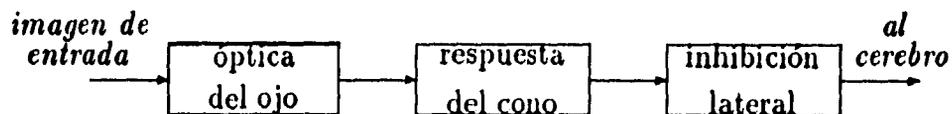
Como consecuencia de que en muchas aplicaciones el *sistema visual humano* (SVH) representa la etapa final de un sistema de codificación de imágenes, resulta de vital importancia conocer sus características principales para poder optimizar los sistemas de codificación.

Idealmente la distorsión en una imagen puede ser controlada de tal forma que no resulte visible aunque exista numéricamente. En la práctica, los sistemas de codificación son optimizados explotando las limitaciones del SVH buscando obtener una distorsión visible mínima. Para lograr dicho propósito, es necesario comprender el mecanismo de visión humano, o al menos aquellas características que nos permiten entender un modelo que describa el fenómeno visual. Las características o los fenómenos de mayor importancia a considerar dentro del SVH, para la codificación de imágenes son: la sensibilidad al contraste, la crominancia, la respuesta espacial y el efecto de enmascaramiento.

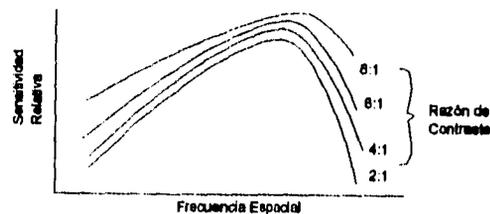
Se ha establecido ampliamente que la sensibilidad de un humano a los cambios de contraste en imágenes monocromáticas es de aproximadamente un 2%. Por otro lado, en cuanto a la sensibilidad a los cambios de crominancia, se ha encontrado que el ser humano es muy sensible a los cambios de verde, sensible en forma moderada a los cambios de rojo y menos sensible a los cambios de azul. Se sabe que los receptores en el ojo: la retina, bastones y conos no responden linealmente a una excitación. En este sentido, se han sugerido varios modelos de representación, sin embargo, los que han tenido una mayor aceptación son los que presentan una respuesta logarítmica y los que presentan una respuesta de raíz cúbica. Cabe mencionar que todos los conocimientos obtenidos sobre los fenómenos característicos del SVH, son

de mucha utilidad para el diseño de cuantizadores en los sistemas de codificación.

En la actualidad se han elaborado bastantes estudios sobre la respuesta en frecuencia espacial del SVH, concluyéndose en forma general, que el ojo humano actúa como un filtro pasabanda no-lineal, como se puede observar cualitativamente en la fig. 1. La atenuación de frecuencias espaciales altas, se atribuye a las limitaciones de respuesta espacial que tiene la óptica del ojo. Por otro lado, la interacción de los bastones y conos, conocida con el nombre de inhibición lateral, provoca un efecto de diferenciación espacial dando como resultado una atenuación de las frecuencias bajas. Por último, el efecto de enmascaramiento del sistema visual se refiere a la reducción de sensibilidad de contraste, en la vecindad límite de los gradientes de luminancia y color. En estas regiones existe una percepción reducida de la distorsión causada por la codificación.



(a)



(b)

FIG. 1. Modelado de la visión humana. (a) Modelo; (b) función de transferencia.

2.2 Revisión del Proceso de Muestreo

2.2.1 El Muestreo

Uno de los conceptos fundamentales en el procesamiento digital de señales es el muestreo. Este proceso permite obtener un conjunto de valores que son representativos del proceso en estudio. En otras palabras, el muestreo consiste en representar una señal $x_c(t)$ en valores de $x_c(nT)$, uniformemente espaciados en múltiplos enteros del intervalo de tiempo T , conocido como *período de muestreo*.

Una forma conveniente de interpretar el muestreo es como un proceso de modulación o multiplicación como el que se muestra en la fig. 2. La señal continua $x_c(t)$ es multiplicada (modulada) por un tren de impulsos periódico (función de muestreo) $s(t)$, obteniéndose una señal modulada $x_c(t)s(t)$. Esta

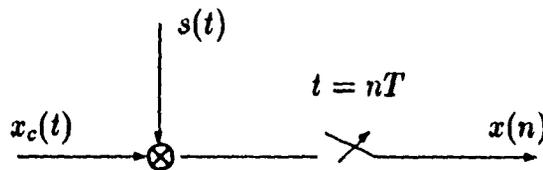


FIG. 2. Interpretación del Proceso de Muestreo

señal es discretizada en el tiempo resultando:

$$x(n) = \lim_{\epsilon \rightarrow \infty} \int_{t=nT-\epsilon}^{nT+\epsilon} x_c(t)s(t)dt$$

donde:

$$s(t) = \sum_{l=-\infty}^{\infty} u_0(t - lT)$$

y en donde $u_0(t)$ es una función impulso unitario ideal. En este contexto de interpretación, $x(n)$ representa el área bajo el impulso en un tiempo nT . Deduciéndose que:

$$x(n) = x_c(nT)$$

2.2.2 El Teorema de Muestreo

Este teorema establece las condiciones bajo las cuales un conjunto de muestras es una representación correcta de la señal muestreada. Se dice que una señal es representada correctamente por un conjunto de muestras con periodicidad T , cuando es posible a partir de estos valores reconstruir completamente la señal original.

A partir de una señal analógica $x_c(t)$ dada, siempre es posible obtener la señal digital $x(n)$. Sin embargo, el proceso inverso no siempre es posible; es decir, $x_c(t)$ especifica únicamente a $x(n)$, pero $x(n)$ no necesariamente especifica únicamente a $x_c(t)$. En este sentido, para lograr una reconstrucción completa de una señal a partir de sus muestras es necesario que el espectro de la señal original no contenga ninguna componente de frecuencia mayor o igual a $f_s/2$ ($f_s=1/T$, frecuencia de muestreo).

Si este no es el caso, las muestras se encimarán y se producirá un efecto de "aliasing" que provoca que la señal reconstruida sea diferente de la original. En estas circunstancias, el Teorema de Muestreo establece la siguiente consideración:

Una señal que no contiene ninguna componente de frecuencia igual o mayor a un valor f_m , puede ser completamente determinada por un conjunto de valores espaciados regularmente, en un intervalo de tiempo $T = 1/(2f_m)$.

Se puede decir entonces, que la frecuencia de muestreo de una señal se determina por el límite superior de su banda de frecuencia.

2.2.3 Conversión de la Tasa de Muestreo

En algunas aplicaciones la señal de entrada a un sistema puede ya estar muestreada con un periodo T predeterminado, y resulta más eficiente o conveniente realizar diferentes partes de un proceso algorítmico con diferentes tasas de muestreo; en tal caso es necesario convertir la tasa de muestreo de las señales en el sistema de un valor a otro.

El proceso de convertir digitalmente la tasa de muestreo $F = 1/T$ de valor determinado, en una tasa diferente $F' = 1/T'$ se conoce con el término de *conversión de la tasa de muestreo*. Cuando la tasa de muestreo resultante es

menor que la original, esto es:

$$F' < F$$

el proceso recibe el nombre de *decimación*. En contraparte, al proceso de conversión digital de la tasa de muestreo F en una tasa mayor F' , es decir:

$$F' > F$$

se le conoce como *interpolación*, debido a que se están creando muestras del proceso físico original a partir de un conjunto reducido de éstas.

La decimación e interpolación de las señales son procesos duales; por ejemplo, un sistema que implementa un decimador puede ser transformado en un sistema digital dual, que implemente un interpolador empleando técnicas de transformación [Crochiere/Rabiner,1983].

En otro contexto, un decimador es un sistema que a partir de una secuencia de entrada $x(n)$ produce una salida $y(n) = x(Mn)$, la cual representa una secuencia comprimida en el dominio del tiempo. En general, la salida $y(n)$ de un decimador se obtiene conservando de la secuencia de entrada únicamente las muestras que ocurren en tiempos que son múltiplos enteros de M , en donde M se le conoce como razón de decimación y es de valor entero. En la fig. 3 se muestra esquemáticamente la manera de representar a un decimador y enseguida se presenta la relación matemática con la cual se obtiene su salida.



$$Y(e^{j\omega}) = \frac{1}{M} \sum_{k=0}^{M-1} X(e^{j(\omega-2\pi k)/M})$$

o bien:

$$Y(z) = \frac{1}{M} \sum_{k=0}^{M-1} X(z^{1/M} W^k)$$

en donde:

$$W = e^{-2\pi j/M} \text{ y } z = e^{j\omega}$$

FIG. 3. Representación esquemática y matemática de un decimador

Un interpolador es un sistema que introduce $M - 1$ muestras de valor cero entre muestras adyacentes de la secuencia de entrada $x(n)$ y provoca una expansión en el dominio del tiempo. En la fig. 4 se presenta un interpolador y posteriormente su formulación matemática.

Los efectos de compresión y expansión en el dominio del tiempo producidos por los procesos de decimación e interpolación, corresponden en el dominio de la frecuencia a una expansión y compresión respectivamente.

$$x(n) \longrightarrow \boxed{\uparrow M} \longrightarrow y(n) = \begin{cases} x(\frac{n}{M}) & , n=\text{múltiplo de } M \\ 0 & , \text{cualquier otro caso} \end{cases}$$

$$Y(e^{j\omega}) = X(e^{j\omega M}) \quad \text{ó} \quad Y(z) = X(z^M)$$

FIG. 4. Representación de un Interpolador

2.3 Cuantización

Al proceso de asignación de un valor escalar continuo f (puede representar la intensidad de un pixel) a un nivel específico L , se le conoce como cuantización de amplitud. Si además, cada valor escalar es cuantizado independientemente, es común usar el término *cuantización escalar* para identificar al proceso de asignación. En este trabajo sólo se mencionará este tipo de cuantización, para una mayor información de éste y otros tipos se recomienda consultar [Lim,1990].

Si consideramos \hat{f} como el resultado de cuantizar f , podemos escribir:

$$\hat{f} = Q(f) = r_i, \quad d_{i-1} < f \leq d_i$$

donde Q representa la operación de cuantización, r_i para $1 \leq i \leq L$ representa los niveles de reconstrucción (cuantización) y d_i para $0 \leq i \leq L$ representa los $L + 1$ niveles de decisión. Si los niveles r_i y d_i se logran determinar de alguna forma, la cuantización de f es entonces un proceso determinístico, lo que permite escribir:

$$\hat{f} = Q(f) = f + e_Q$$

donde e_Q es el error de cuantización dado por:

$$e_Q = \hat{f} - f$$

Los criterios que generalmente se usan para determinar los niveles r_i y d_i , tienen como objetivo minimizar el error utilizando una medida de la distorsión.

Aquellos cuantizadores en los que dichos niveles se encuentran igualmente espaciados, reciben el nombre de *cuantizadores uniformes*. Específicamente para estos cuantizadores tenemos que:

$$d_i - d_{i-1} = \Delta, \quad 1 \leq i \leq L$$

$$r_i = \frac{d_i + d_{i+1}}{2}, \quad 1 \leq i \leq L$$

donde Δ es el tamaño del paso de cuantización. La fig.5 es un ejemplo de un cuantizador uniforme.

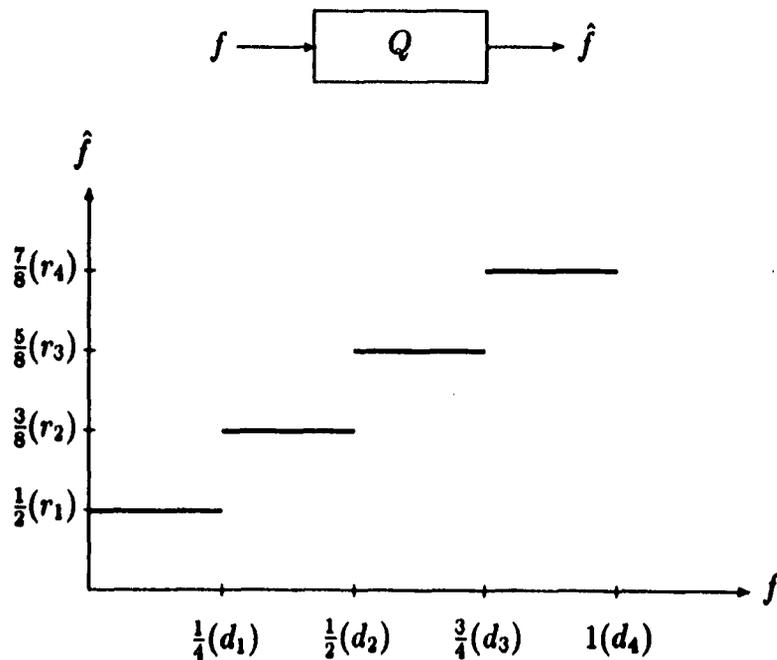


FIG. 5. Cuantizador uniforme, con 4 niveles de reconstrucción y con valores de f que varían de 0 y 1.

A pesar de que la cuantización uniforme se obtiene de forma directa y como una aproximación natural, no siempre es óptima. Existen aplicaciones en las que por las características de la señal que se va a cuantizar, es necesario que el número de niveles de cuantización sea mayor en algunas regiones que en otras. En otras palabras, se necesita realizar una *cuantización no uniforme*. El criterio que normalmente se utiliza para determinar los niveles r_i y d_i de un cuantizador no uniforme, es el de la *minimización del error cuadrático medio* (MMSE). El ejemplo de un cuantizador que utiliza una medida de la distorsión basado en el criterio anterior, es el que se conoce con el nombre de cuantizador *Lloyd-Max*. Para obtener los niveles de cuantización de este cuantizador, se realiza un proceso iterativo a partir de una función de densidad de probabilidad de entrada y se calculan los niveles para dicha señal con una densidad de probabilidad de amplitud normalmente distribuida.

2.4 Entropía

En términos generales, la imagen o la información que se transmite por un canal se forma por una secuencia de símbolos elegidos de un conjunto finito, alfabeto de la fuente. Al alfabeto se le asocia una distribución de probabilidad y a la emisión de símbolos sucesivos en el mensaje se le controla por medio de un mecanismo probabilístico. Los mensajes transmitidos llevan por lo regular cantidades de información diferente, por lo que es necesario definir una cantidad de información promedio o *Entropía*.

Para un sistema capaz de transmitir n niveles discretos en intervalos de tiempo λ , el número de combinaciones de diferentes señales que se presentan en T segundos es $n^{\frac{T}{\lambda}}$. Como la cantidad de información es proporcional al tiempo de transmisión, se puede obtener el logaritmo de $n^{\frac{T}{\lambda}}$ para determinar la información transmitida en T segundos como:

$$H = \frac{T}{\lambda} \log_2 n \quad [\text{bits}]$$

Por otro lado, a la máxima cantidad de información por segundo que es capaz de transmitir un sistema, se le define como *capacidad* y puede expresarse en bits por segundo:

$$C = \frac{H}{T} \quad [\text{bits/seg}]$$

A la frecuencia relativa de ocurrencia de algún evento o combinación, se le define como probabilidad de ocurrencia y se le puede representar simbólicamente como P , donde:

$$P = \frac{\text{número de veces que ocurre un evento}}{\text{número total de posibilidades}}$$

Considerando el caso en el que la probabilidad de ocurrencia de los niveles de una señal o evento no sean iguales, se puede decir que la frecuencia de ocurrencia de cada nivel de la señal será representada por P_i , donde: $P_1 + P_2 + \dots + P_n = 1$. Entonces, para cada intervalo la información contenida será de $-\log_2 P_i$. Ahora, si sumamos la información en bits promedio por cada símbolo que aparece $t * P_i$ veces sobre un intervalo t , se puede escribir:

$$H = -t * \sum_{i=1}^n P_i \log_2 P_i \text{ bits en } t \text{ periodos}$$

Para un intervalo T , se puede escribir:

$$H = -\frac{T}{\lambda} * \sum_{i=1}^n P_i \log_2 P_i \text{ bits en } T \text{ segundos}$$

Para un mensaje con n símbolos posibles y con una probabilidad de ocurrencia de P_i a P_n , la información promedio por símbolo en un intervalo de λ segundos es:

$$H_{prom} = -\sum_{i=1}^n P_i \log_2 P_i \text{ bits/intervalo del simbolo}$$

Esta última ecuación representa la definición matemática de *Entropía*. El término entropía es usado para denotar el número promedio de bits que se requieren para la representación de cada símbolo en un alfabeto fuente.

2.5 Código de Huffman

La codificación por Huffman es una técnica empleada para reducir la longitud promedio de un código que representa los símbolos de un alfabeto.

El código de Huffman se puede generar a través de una estructura de árbol como se ilustra en la figura 6. En la figura los símbolos se ordenaron en

forma descendente con respecto a su frecuencia de aparición. Los símbolos con las frecuencias más pequeñas (X_3 y X_4) son combinados en un sólo nodo, obteniéndose una nueva probabilidad de ocurrencia (0.25). A continuación el nodo es mezclado con la siguiente probabilidad de ocurrencia más pequeña; en caso de que la probabilidad resultante de la primera combinación sea mayor que otras, se vuelven a elegir las dos más pequeñas para combinarlas. El proceso anterior se realiza en forma iterativa hasta llegar a un nodo cuya probabilidad sea uno; una vez que se término con este proceso, se prosigue con la asignación binaria de ceros y unos a cada rama o segmento que salen de un nodo, lo que permite obtener el código para cada caracter. El código se obtiene siguiendo la trayectoria que se marca a partir del nodo con probabilidad uno, hasta el símbolo inicial y anotando los ceros y unos encontrados en el recorrido.

El número promedio de bits por símbolo obtenidos se puede calcular multiplicando la longitud del código de Huffman por su probabilidad de ocurrencia.

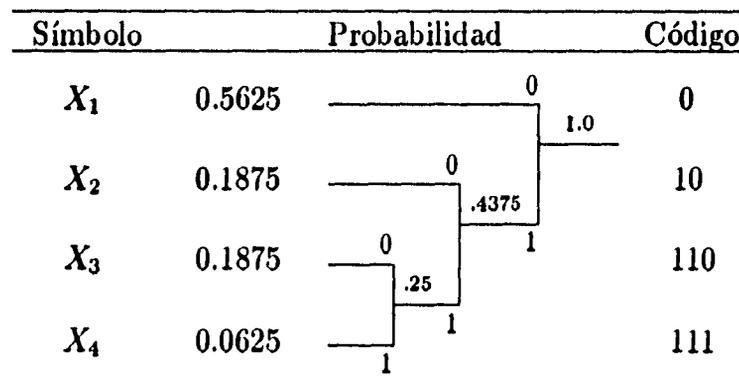


FIG. 6. Diagrama de árbol para generar un Código de Huffman.

Para el caso de la figura se obtuvieron 1.63 *bits/símbolo*. Hay que subrayar que la longitud promedio de bits obtenida por el código de Huffman debe ser aproximado al valor de entropía, en caso de que esta se hubiese calculado.

Capítulo 3

Técnicas de Compresión de Imágenes

3.1 La Evaluación de la Calidad en las Imágenes

El poder realizar una medición cuantitativa de una imagen procesada es de gran importancia para el diseño y la evaluación de las diferentes técnicas de compresión. Por tal razón, el tipo de medición elegida debe permitir realizar una comparación con alguna cantidad o valor estandar, además, si esta cantidad de medición es una función analítica, debe ser posible emplearla con la finalidad de optimizar los parámetros del sistema de transmisión. A continuación se presentarán algunos de los métodos más comunes en la medición de la calidad de las imágenes.

3.1.1 Escalas Subjetivas

Una medición aceptable de la calidad en una imagen debe tener una buena correlación con la formulación que llevará a cabo la medición; como la selección de escalas subjetivas. En aplicaciones para codificación de imágenes existen dos escalas subjetivas relevantes: *la de calidad absoluta y la de degradaciones percibidas*. En las escalas de calidad absoluta se cuenta comunmente con 5 o 7 pasos que son etiquetados desde inaceptable hasta excelente. Cada etiqueta o nivel de la escala intenta evaluar la calidad considerando alguna otra imagen como referencia. En las escalas de degradaciones percibidas se tiene simplemente una escala cuyos niveles indican el grado de degradación en una imagen codificada con respecto a una representación perfecta. La tabla No.1 muestra dos ejemplos de los posibles rangos o niveles en escalas subjetivas, estos rangos se utilizan comunmente en

Escala de Calidad en Imágenes	
Escala absoluta	
5.	Excelente
4.	Buena
3.	Regular
2.	Mala
1.	Indeseable
Escala de Degradaciones	
1.	No perceptibles
2.	Muy poco perceptibles
3.	Perceptibles, pero con degradaciones muy pequeñas en la imagen
4.	Las imperfecciones en la imagen son aceptables
5.	Poco aceptables
6.	Inaceptables
7.	Extremadamente inaceptables

TABLA 1

el área de transmisión de imágenes [Pratt,1979].

3.1.2 Mediciones Numéricas

Las mediciones de calidad en imágenes numéricas se dividen básicamente en dos clases: *univariantes* y *bivariantes*. Las univariantes están formadas por una sola imagen y su escala de medición es calibrada evaluando la desviación de la imagen examinada con respecto a una estandar. Las bivariantes son indicadores diferenciales de la calidad de una imagen, obtenidos utilizando dicha imagen antes y después de procesar.

Las mediciones univariantes se basan comunmente en el espectro de frecuencia espacial de una imagen. En el dominio continuo, la frecuencia espacial de una imagen con una distribución de luminancia espacial $F(x, y)$ se define en términos de la transformada de Fourier de dos dimensiones como:

$$\mathfrak{F}(\omega_x, \omega_y) = \int_{-L}^L \int_{-L}^L F(x, y) \exp[-j(\omega_x x + \omega_y y)] dx dy$$

donde L representa los límites espaciales de la imagen, ω_x y ω_y denotan las frecuencias espaciales. La transformada discreta de Fourier

$$\mathfrak{F}(u, v) = \frac{1}{(JK)^{1/2}} \sum_{j=0}^{J-1} \sum_{k=0}^{K-1} F(j, k) \exp\{-2\pi i(\frac{uj}{J} + \frac{vk}{K})\}$$

puede ser utilizada para un arreglo de imagen discreta $F(x, y)$, donde se asume que la imagen esta uniformemente muestreada sobre una rejilla de dimensiones $J \times K$ y donde (u, v) representa las componentes de frecuencia espacial. Las mediciones típicas incluyen en todas las frecuencias la energía integrada o el valor medio de energía, dichas mediciones han sido de mucha utilidad cuando se evalua el funcionamiento de sistemas de transmisión de imágenes con una respuesta lineal. Para el caso de sistemas de codificación con una respuesta no lineal, las mediciones realizadas son frecuentemente erroneas y contradictorias.

Las mediciones de calidad bivalentes son extremadamente útiles para aplicaciones de transmisión de imágenes, ya que estas proporcionan un indicador de la degradación entre las imágenes de entrada y salida de un sistema de codificación. La medida bivalente más comun es el *error cuadrático medio* (MSE). El MSE comunmente es normalizado por la energía o el pico de magnitud de la imagen de entrada. En el dominio discreto el MSE normalizado entre una imagen de entrada $F(j, k)$ y una imagen reconstruida de la codificación $\hat{F}(j, k)$ está definido como:

$$\epsilon_N = \frac{\sum_{j=0}^{J-1} \sum_{k=0}^{K-1} [F(j, k) - \hat{F}(j, k)]^2}{\sum_{j=0}^{J-1} \sum_{k=0}^{K-1} [F(j, k)]^2}$$

El error de amplitud normalizado se define como:

$$\epsilon_A = \frac{\sum_{j=0}^{J-1} \sum_{k=0}^{K-1} [F(j, k) - \hat{F}(j, k)]^2}{JKA^2}$$

donde A denota la amplitud máxima de $F(j, k)$.

Es importante mencionar que aunque existen muchas formas de evaluar la calidad de una imagen procesada, estas no son completamente aceptables. Por tal motivo en la actualidad se siguen realizando investigaciones sobre mediciones de calidad en imágenes. Por otro lado, a pesar del error que se puede cometer al emplear alguna forma de medición, el diseño de los sistemas de codificación se realiza con los procedimientos de medición existentes.

3.2 Técnicas de Codificación de Imágenes

Como se mencionó anteriormente el objetivo de la compresión es reducir el número de elementos con los que se representa una imagen, lo que permite optimizar los sistemas de codificación a través de una reducción de la tasa de transmisión. También se habló de la importancia que tiene el considerar las limitaciones del sistema de visión humano en el diseño de un codificador y del papel que juega la redundancia en la compresión de imágenes. Por lo que, en las siguientes secciones se presentarán algunas de las más importantes técnicas de codificación de imágenes, las cuales pretenden explotar uno o ambos puntos anteriormente mencionados para lograr su propósito de compresión, sin olvidar el mantener una calidad aceptable en la imagen. En las siguientes secciones no se hará un estudio exhaustivo de los diferentes métodos de codificación, por lo que es recomendable revisar algunas de las referencias presentadas para lograr una mayor profundidad de los temas.

3.2.1 Codificación por PCM

El método que utiliza una Modulación por Codificación de Pulso (PCM) es la forma más simple de llevar a cabo una codificación de una imagen. En este sistema la señal de imagen es muestreada, y cada muestra es cuantizada y codificada para su transmisión. Para el caso de una transmisión por facsimil, la señal de imagen es cuantizada en dos niveles únicamente, negro o blanco, y codificada en un bit por muestra, cero o uno. Las imágenes monocromáticas normalmente son cuantizadas usando de 64 a 256 niveles por muestra, es decir, con 6 u 8 bits por muestra. La codificación por PCM puede ser considerada como la base de un sistema de codificación, ya que las reducciones en la tasa de transmisión son comúnmente calculadas con respecto a una codificación con PCM.

La reducción en el número de niveles de cuantización para una imagen codificada con PCM produce un efecto de contornos falsos, provocando que los saltos discretos entre los niveles de cuantización sean observados en regiones donde los cambios de luminancia son suaves. El efecto de contornos falsos se puede reducir en gran medida agregando una pequeña cantidad de ruido pseudoaleatorio [Pratt,1979] a la señal de la imagen antes de la cuantización.

3.2.2 Codificación Estadística

La codificación por PCM es deficiente debido a que no considera la dependencia que existe entre los pixels vecinos, además trata a todos los niveles de amplitud cuantizados como si fueran equivalentes. Una reducción sustancial en la tasa de transmisión es posible teóricamente codificando grupos de pixels, y asignando palabras de código cuya longitud sea inversamente proporcional a la probabilidad de ocurrencia del grupo de amplitud patrón. El código de Huffman puede usarse para desarrollar un código de longitud promedio mínima para un conjunto de probabilidades patrón dado [Held,1983].

La codificación estadística constituye la base para la mayoría de las técnicas de codificación binaria por facsimile. Para estas fuentes, el número posible de patrones binarios es de tamaño razonable. Pero, en el caso de imágenes con múltiples niveles de gris, al número de patrones puede ser demasiado grande. Por ejemplo, en un grupo de 4 x 4 pixels, con cuantización en 256 niveles, corresponde aproximadamente a 10^{40} patrones posibles. Consecuentemente, la aplicación de la codificación estadística para imágenes multinivel se ha limitado a regiones tan pequeñas como de aproximadamente dos o tres pixels.

La efectividad de la codificación estadística se ve limitada por la determinación de las probabilidades de la fuente. Para la mayoría de las imágenes, las probabilidades varían drásticamente de una imagen a otra, y aun dentro de una misma imagen. Los resultados de una buena codificación pueden ser obtenidos únicamente por una estimación adaptativa de las probabilidades de la fuente y de una revisión apropiada del código de la misma.

A consecuencia de que no se obtiene mucha ganancia en términos de compresión en comparación con la complejidad del codificador, la técnica estadística por sí sola, no se ha utilizado con frecuencia en aplicaciones de codificación de imágenes monocromáticas y de color. Sin embargo, estos conceptos han sido de gran utilidad en combinación con otros métodos de codificación.

3.2.3 Codificación predictiva

En la fig. 7 se presenta un diagrama a bloques de un sistema general de codificación predictiva para imágenes. En este sistema, la amplitud de cada pixel adquirido es precedida en base a la historia de los pixels previamente adquiri-

dos. En seguida, la predicción estimada $\tilde{F}(j, k, t)$ es sustraída de la amplitud del pixel actual $F(j, k, t)$, y la señal de diferencia $D(j, k, t)$ es cuantizada, codificada y transmitida. En el receptor, la señal de diferencia cuantizada $\hat{D}(j, k, t)$

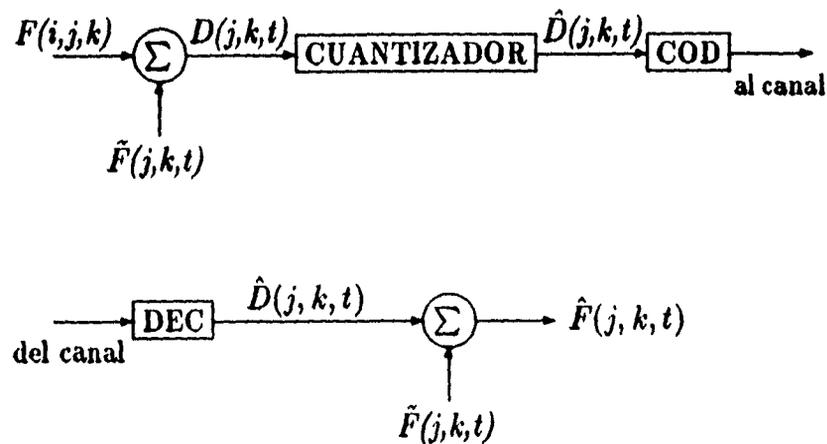


FIG. 7. Sistema de Codificación Predictivo

es sumada con la señal de predicción $\tilde{F}(j, k, t)$ para generar una imagen de reconstrucción $\hat{F}(j, k, t)$.

El codificador predictivo de imágenes más simple es un sistema de modulación delta, en el cual la predicción se forma a partir de predicciones previas de los valores cuantizados únicamente con dos niveles.

3.2.4 Codificación por Transformada

En un sistema de codificación por transformada, como el que se ilustra en la fig. 8, se realiza una transformación en dos dimensiones sobre una imagen o sobre subsecciones de la misma (bloques). Las transformadas de Fourier, seno, coseno, Hadamard, Haar, Slant y de Karhunen-Loève se han utilizado ampliamente en la codificación de imágenes. La transformación produce un arreglo relativamente descorrelacionado y permite que los coeficientes obtenidos de dicha transformación, sean cuantizados y codificados eficientemente para ser transmitidos.

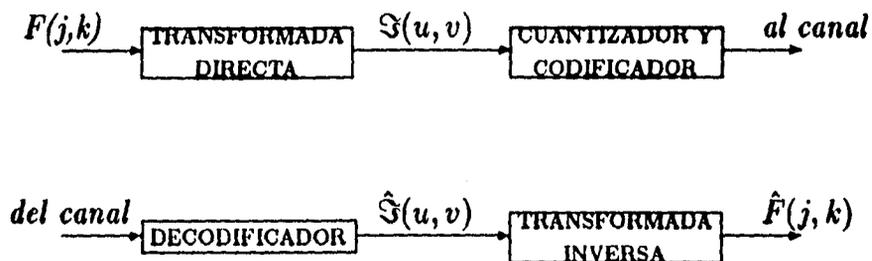


FIG. 8. Sistema de codificación por Transformada

Después de realizar una transformación, las muestras transformadas son seleccionadas para ser transmitidas de acuerdo con un criterio de zona, o utilizando un umbral resultado de la estrategia de muestreo. En la selección por zona se eligen los coeficientes que se encuentran localizados en una región específica; usualmente se eligen aquellos coeficientes de frecuencia espacial baja para ser transmitidos. Con el criterio de umbral, el coeficiente que es seleccionado para la transmisión es aquel cuya magnitud es mayor a un valor límite. Una vez que se realizó la selección de las muestras o coeficientes que serán transmitidos se procede a cuantizar y a codificar dichas muestras para la transmisión.

En imágenes monocromáticas y utilizando alguna de las técnicas de compresión por transformada se han logrado tasas de bit de 1.0 a 1.5 *bits/pixel* con un error cuadrático medio menor del 0.5% [Pratt,1979].

En la siguiente sección se presentará un caso especial de transformación que representa la base sobre la cual se desarrolló esta tesis.

3.3 Transformadas Sub-banda

Las transformaciones lineales son la base de una gran cantidad de técnicas utilizadas en el procesamiento, análisis y codificación de imágenes. Una subclase de transformadas lineales son las *Transformadas Sub-banda*, las cuales ofrecen propiedades muy utilizadas para el tratamiento de imágenes. Tradicionalmente los codificadores que se basan en transformadas lineales son divididos en dos categorías :

- Codificadores por Transformada
- Codificadores Sub-banda

Esta distinción se debe en parte a la naturaleza de los métodos de programación utilizados para los dos tipos de representación.

Las técnicas de codificación por transformada se apoyan principalmente en transformadas lineales ortogonales, de las cuales ya se mencionaron anteriormente algunos ejemplos como la transformada de Fourier, la transformada del coseno discreta y la de Karhunen-Loève, entre otras.

El cálculo de las transformadas sub-banda se realiza generalmente por una convolución de la señal de entrada con un conjunto de filtros pasabanda y decimando los resultados. Cada una de las señales subbanda se sitúa en una porción particular del espectro de frecuencia que corresponde a la información que se está presentando en una escala espacial particular. Para la reconstrucción de la señal, las señales subbanda son sobremuestreadas, filtradas y posteriormente combinadas aditivamente. Para propósitos de codificación, las transformadas sub-banda pueden ser utilizadas para controlar el error de cantidades relativas en diferentes partes del espectro de frecuencia [Woods,1991]. La mayoría de los filtros diseñados para codificadores sub-banda tratan de minimizar el "aliasing" que se da como resultado del proceso de submuestreo. En el dominio espacial, el "aliasing" aparece en la imagen de salida como evidencia de la estructura de muestreo. En un sistema sub-banda ideal se incorporan filtros pasabanda que evitan que se presente el "aliasing". Sin embargo, los filtros producen resonancia (fenómeno de Gibbs) en el dominio espacial lo cual es indeseable perceptualmente.

3.3.1 Propiedades de las Transformadas Sub-banda

El objetivo de esta sección, es presentar un conjunto de propiedades relevantes para el problema de la codificación de imágenes, que puedan ser usadas en la selección de una transformación lineal.

Orientación y Escala

Debido a que las imágenes contienen objetos y rasgos de muy diferentes tamaños pueden ser vistas sobre un amplio rango de distancias, y de esta forma, una transformación puede analizar la imagen simultáneamente (e independientemente) en diferentes escalas. Varios autores argumentan que una adecuada partición en términos de escala es aquella en la que las escalas son relacionadas por una constante fija de proporcionalidad. En el dominio de la frecuencia, esto corresponde a una descomposición en sub-bandas localizadas, del mismo tamaño sobre una escala logarítmica.

Para señales de dos dimensiones, una región localizada en el plano de la frecuencia corresponde espacialmente a una escala y *orientación* particular. Una orientación específica permite a la transformada la extracción de un alto orden de estructuras orientadas que se encuentran típicamente en las imágenes, como bordes y líneas. Por tal razón, es muy utilizada la partición de una señal de entrada en paquetes localizados en el dominio de la frecuencia.

Localización Espacial

Además de la localización en frecuencia, es muy ventajoso que las funciones de base estén localizadas espacialmente, ya que la transformada debe codificar información posicional. La necesidad de la localización espacial es particularmente clara en sistemas de *vision por computadora*, en donde la información referente a la localización de rasgos en una imagen es crítica.

La forma más simple de realizar una localización tanto en el espacio como en la frecuencia espacial es calcular una transformada discreta de Fourier o una transformada del coseno discreta en un bloque pequeño. Pero la introducción arbitraria de bloques limitantes violan la continuidad de la imagen. Además, la localización en frecuencia de las funciones es pobre (cuando es considerada en el contexto de la imagen como un todo), debido a los bordes marcados que forman la ventana que define al bloque.

Gabor[1946] demostró que las funciones producidas como el producto de sinusoidales complejas y ventanas Gaussianas, presentan una localización óptima para la juntura *tiempo-frecuencia*, y desarrolló una transformada, usando dichas funciones como conjunto base.

Ortogonalidad

La justificación que usualmente se da para la condición de ortogonalidad es en términos de la descorrelación. Dada una señal con propiedades estadísticas de segundo orden, existe una transformada ortogonal que descorrelaciona la señal. En otras palabras, la correlación de segundo orden de los coeficientes transformados debe ser cero. La ortogonalidad no es usualmente discutida en el contexto de las transformaciones sub-banda, aunque muchas de estas sean ortogonales. Sin embargo, es importante tener en cuenta que una transformada que no es ortogonal en forma estricta, puede ser indeseable para la codificación.

3.4 Ejemplos de Transformadas Sub-banda

3.4.1 La Transformada de Gabor

Como se argumentó anteriormente, para obtener una descomposición óptima es necesario considerar una localización tanto espacial, como frecuencial. A este respecto, se presenta una solución para el problema de la localización en la descomposición en sub-bandas, propuesto por Dennis Gabor[1946]. Gabor intrduce una transformada unidimensional en la que las funciones base son sinusoidales ponderadas por ventanas Gaussianas, y obtuvo que dichas funciones base están localizadas tanto en el espacio como en la frecuencia, además, demostró que esta juntura (espacio-frecuencia) presenta una localización óptima con respecto a la medida que él eligió. Para el caso bidimensional, las funciones base de Gabor son sinusoidales direccionales ponderadas por ventanas gaussianas.

Una de las dificultades que se presenta con la transformada de Gabor es el hecho de que es estrictamente no ortogonal. Además en una aplicación para codificación, los errores introducidos por la cuantificación de los coeficientes, se distribuyen a través de todo el dominio espacial, así como también, en el frecuencial, aun y cuando los valores de los coeficientes sean calculados en

base a la información de la localización espacial y de las regiones de frecuencia.

3.4.2 La Transformada del Coseno Discreta (TCD)

El uso de la transformada del coseno discreta en los sistemas de codificación de imágenes, se justifica frecuentemente, por el hecho de que es una aproximación a la transformación óptima para una señal con características estadísticas Gauss-Markov de primer orden. En la práctica, la transformada usualmente no se calcula de manera global, pero es aplicada a sub-bloques de la imagen que no se traslapan. El bloque resultante de funciones base en la TCD constituye una transformada sub-banda, sin embargo, las sub-bandas no se encuentran muy bien localizadas. Si se considera en la estructura de un sistema A/S (análisis-síntesis), que las sub-bandas de la imagen submuestreada contienen una importante cantidad de "aliasing" y tomando en cuenta que la transformada es invertible, resulta claro esperar que el "aliasing" sea cancelado en la etapa de síntesis. Sin embargo, si los coeficientes transformados son cuantizados o descartados, el "aliasing" no es cancelado en su mayoría, produciéndose errores en la imagen reconstruida.

3.4.3 La Pirámide Laplaciana

Una de las primeras técnicas de descomposición en sub-bandas (por octavas), fue desarrollada por Burt y aplicada al problema de codificación de imágenes por Burt y Adelson. Ellos usaron una pirámide con pequeños filtros de tipo gaussiano en cascada, creando una representación sub-banda bastante completa, a la que le llamaron *Pirámide Laplaciana*. Un sistema que constituye un nivel de esta pirámide se muestra en la figura 9. La señal se pasa a través de un filtro pasobajas, $B(\omega)$, y entonces se submuestra para producir la sub-banda pasobajas $W_0(\omega)$. Una sub-banda pasoaltas, $W_1(\omega)$, es formada sobremuestreando $W_0(\omega)$, convolucionando con un filtro de interpolación $A(\omega)$ y sustrayendo el resultado de la señal original. La señal es reconstruida sobremuestreando y filtrando $W_0(\omega)$ con $A(\omega)$ y sumando el resultado con $W_1(\omega)$. Esta reconstrucción es exacta, a pesar de los filtros $B(\omega)$ y $A(\omega)$ elegidos. La pirámide completa se construye de una manera recursiva, es decir, volviendo a aplicar el sistema a la sub-banda pasobajas del primer nivel.

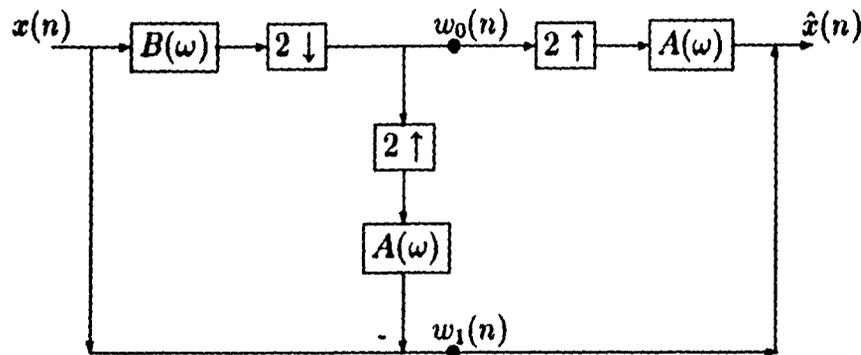


FIG.9. Sistema que genera un nivel de la Pirámide Laplaciana

Además de lo conveniente que resulta para la compresión de datos, la naturaleza multi-escala de la pirámide permite que esta sea utilizada en tópicos de transmisión progresiva.

Por otro lado, una de las desventajas más serias que presenta la pirámide laplaciana para la codificación de imágenes es que los errores de cuantización de las sub-bandas pasoaltas no permanecen en estas sub-bandas. En vez de esto, aparecen en la imagen reconstruida como ruido de banda ancha. Como con la transformada de Gabor, la fuente de esta dificultad es la no ortogonalidad de la transformada. Además, las funciones base bidimensionales no son orientadas, por lo que, no se extrae la redundancia de estructuras orientadas que se encuentran típicamente en las imágenes.

3.4.4 Filtros Espejo en Cuadratura

Los filtros espejo en cuadratura (QMF) fueron introducidos por Esteban y Galand [1977] como un método para descomponer una señal unidimensional en sub-bandas, esta descomposición logra una eficiente decimación de las sub-bandas y logra una reconstrucción de la señal original libre de "aliasing". Las funciones base de los QMF pueden ofrecer una buena localización tanto espacial como frecuencial, se pueden poner en cascada formando un conjunto ortonormal y escalable.

Vetterli [1984], Woods y O'Neil [1986] han extendido las técnicas QMF a señales bidimensionales. Varias versiones de pirámides QMF para descomposición han sido aplicadas para la codificación de imágenes por Gharavi,

Tabatabai [1987] y Tran et al [1987].

A continuación se presenta una breve revisión sobre estructuras QMF de dos canales que descomponen sólo señales monodimensionales, por lo que es recomendable revisar Crochiere y Rabiner [1983] y Vaindynathan [1987] para lograr una mayor profundidad en el tema.

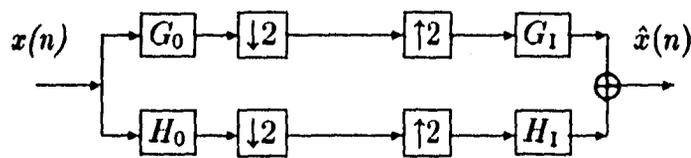


FIG.10. Sistema de análisis/síntesis de dos canales

Banco De Filtros Espejo En Cuadratura

Considerando un sistema de descomposición análisis/síntesis de dos sub-bandas y con un orden de decimación e interpolación de dos, como el mostrado en la fig 10. Se puede decir, en principio que un buen sistema de este tipo es aquel que garantice la conservación de la información procesada, es decir:

$$\hat{x}(n) = x(n)$$

Basados en las relaciones para la decimación e interpolación que se dieron en el capítulo 2 y en relación al sistema anterior de dos canales, es posible expresar a $\hat{X}(z)$ como:

$$\hat{X}(z) = \frac{1}{2}[H_1(z) \cdot H_0(z) + G_1(z) \cdot G_0]X(z) + \frac{1}{2}[H_1(z) \cdot H_0(-z) + G_1(z) \cdot G_0(-z)]X(-z)$$

en donde :

$$z = e^{2\pi jf}$$

$G_0(z), G_1(z), H_0(z)$ y $H_1(z)$: funciones de transferencia de los filtros cuyas respuestas a impulso son respectivamente $g_0(n), g_1(n), h_0(n), h_1(n)$

$X(z)$ y $\hat{X}(z)$: espectros de las señales de entrada $x(n)$ y de salida $\hat{x}(n)$ respectivamente.

El objetivo siguiente, es lograr que el sistema tenga una función de transferencia lineal $T(z)$.

$$T(z) = \frac{\hat{x}(z)}{x(z)}$$

y para conseguirlo, el término $X(-z)$ efecto de la decimación debe desaparecer. Así que, los filtros de síntesis deben tener una relación con los de análisis de la siguiente forma:

$$\begin{aligned} G_1(z) &= H_0(-z) \\ H_1(z) &= -G_0(-z) \end{aligned}$$

de modo que:

$$\begin{aligned} g_1(n) &= (-1)^n h_0(n) \\ h_1(n) &= (-1)^{n+1} g_0(n) \end{aligned}$$

Obteniéndose una función de transferencia que se puede escribir como:

$$T(z) = \frac{\hat{X}(z)}{X(z)} = \frac{1}{2} [G_0(z)H_0(-z) - H_0(z)G_0(-z)]$$

Ahora bien, sabiendo que la condición de reconstrucción perfecta está expresada a través de una función de transferencia, donde el módulo es constante (distorsión de amplitud) y la fase es lineal (distorsión de fase).

$$T(z) = Kz^{-N}$$

Se puede afirmar que la optimización del sistema depende de los filtros G_0 y H_0 . De hecho existen varias soluciones a este problema de optimización y en este trabajo se revisará una de las más frecuentemente utilizadas.

Solución QMF

Para resolver el problema de optimización Esteban y Galand consideraron que G_0 y H_0 son filtros del tipo FIR (respuesta al impulso finita), de fase lineal y de orden $N - 1$, tales como:

$$\begin{aligned} H_0(z) &= |H_0(z)|z^{-\left(\frac{N-1}{2}\right)} \\ G_0(z) &= |G_0(z)|z^{-\left(\frac{N-1}{2}\right)} \end{aligned}$$

además, si las funciones de transferencia de dichos filtros se encuentran en cuadratura; por ejemplo que una sea la imagen espejo de la otra con respecto a una frecuencia normalizada de $\frac{1}{4}$.

$$\begin{aligned} G_0(z) &= H_0(-z) \\ g_0(n) &= (-1)^n h_0(n) \end{aligned}$$

y tomando en cuenta la linealidad de la fase se puede escribir:

$$\begin{aligned} G_0(z) &= |H_0(-z)|(z)^{-\frac{N-1}{2}} \\ G_0(z) &= |G_0(z)|(-z)^{-\frac{N-1}{2}}(-1)^{\frac{N-1}{2}} \end{aligned}$$

por lo tanto, la función $T(z)$ se puede expresar de la siguiente manera:

$$T(z) = \frac{1}{2}[|H_0(z)|^2 - (-1)^N - 1|G_0(z)|^2]z^{-(N-1)}$$

consiguiéndose eliminar la distorsión de fase gracias el empleo de filtros con fase lineal. Ahora sólo resta optimizar la amplitud; para ello es importante mencionar que si N es impar la frecuencia $f = \frac{1}{4}$ es descartada, ya que el valor de $T(z)$ es nulo para dicha frecuencia y se produce una distorsión de amplitud severa. Por tal razón, el soporte de los filtros H_0 y G_0 debe ser par y $T(z)$ se puede reescribir como:

$$|T(z)| = \frac{1}{2}[|H_0(z)|^2 + |H_0(-z)|^2]$$

esta última ecuación representa la distorsión de amplitud residual y su valor depende en gran medida de la elección de H_0 .

En el siguiente capítulo se presentan dos estructuras piramidales como aplicaciones de la transformada sub-banda para señales bidimensionales; de las cuales se elegirá una para su simulación.

Capítulo 4

Pirámides Ortogonales Para la Codificación de Imágenes

En este capítulo se estudiarán las transformadas piramidales que permitan descomponer una imagen sobre un conjunto de funciones de base, dichas funciones deben presentar características tales como: orientación y escalabilidad, localización tanto en espacio como en frecuencia y además ortogonalidad.

Por otro lado, con la intención de lograr un mejor manejo de las transformaciones se presenta a continuación una revisión básica de ideas y de notación [Adelson/Simoncelli, 1987]; para simplificar las operaciones unicamente se considerará una imagen discreta unidimensional $e[n]$, siendo n el número de pixel. En estos términos, la transformada de una imagen expresa a $e[n]$ como una suma de funciones de base $f_i[n]$ ponderadas por los coeficientes p_i :

$$e[n] = \sum_i p_i f_i[n]$$

Los coeficientes p_i pueden obtenerse como el producto interno de la imagen con un conjunto de funciones de muestreo g_i :

$$p_i = \sum_n g_i[n] e[n]$$

Esto puede ser expresado en una forma matricial de la siguiente manera: Considerando que la señal es un vector columna e y definiendo la matriz de base F compuesta por las columnas f_i :

$$F = \begin{bmatrix} \cdot & \cdot & \cdot & & \\ \cdot & \cdot & \cdot & & \\ \cdot & \cdot & \cdot & & \\ f_0 & f_1 & f_2 & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & & & \\ \cdot & \cdot & \cdot & & & \\ \cdot & \cdot & \cdot & & & \end{bmatrix}$$

Se requiere entonces que

$$e = Fp$$

en donde p es el vector columna de coeficientes. Estos coeficientes se calculan por la multiplicación de la imagen con la matriz transpuesta de muestreo G , donde G esta compuesta por las columnas g_i . De modo que,

$$p = G^T e \implies e = FG^T e$$

por lo tanto, G debe ser la transpuesta de la inversa de F , esto es:

$$G = (F^{-1})^T$$

Cabe mencionar que para el caso de una transformada ortonormal las funciones de muestreo son idénticas a las funciones de base: $G = F$. Aunque, en forma general esto no es necesario.

El interés particular en este trabajo es sobre las funciones que puedan ser particionadas en otras clases de menor tamaño, las cuales sean versiones desplazadas una de la otra. Es decir, en matrices de base de la forma:

$$F = \begin{bmatrix} f_a & & f_b & & & \\ \cdot & f_a & \cdot & f_b & & \\ \cdot & \cdot & f_a & \cdot & \cdot & f_b & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & & & \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & & & \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & & & \end{bmatrix}$$

Con lo cual, tenemos que las correspondientes matrices de muestreo son de

la forma:

$$G = \begin{bmatrix} g_a & & g_b & & & & \\ \cdot & g_a & & \cdot & g_b & & \\ \cdot & \cdot & g_a & \cdot & \cdot & g_b & \cdot \cdot \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \\ & \cdot & \cdot & \cdot & \cdot & & \\ & & \cdot & & \cdot & & \\ & & & & \cdot & & \\ & & & & & & \cdot \end{bmatrix}$$

Una interpretación alternativa que describe esta transformación se muestra en la figura 11; este tipo de esquema se presentó en el capítulo anterior para un caso particular de dos canales, con el objeto de apoyar el desarrollo matemático, sin embargo, no se describió el proceso del sistema representado esquemáticamente. En la figura 11, la imagen es considerada como una secuencia $e[n]$, en donde los "kernels" g_a, g_b, \dots son convolucionados con la imagen y el resultado es submuestreado por los factores k_a, k_b, \dots para obtener las sub-imágenes $p_a[i], p_b[i], \dots$. Estas sub-imágenes constituyen la representación de la transformación. Para la reconstrucción las sub-imágenes son sobremuestreadas y convolucionadas con los "kernels" base f_a, f_b, \dots obteniéndose las sub-imágenes expandidas q_a, q_b, \dots . Las cuales se suman para obtener la imagen original $e[n]$.

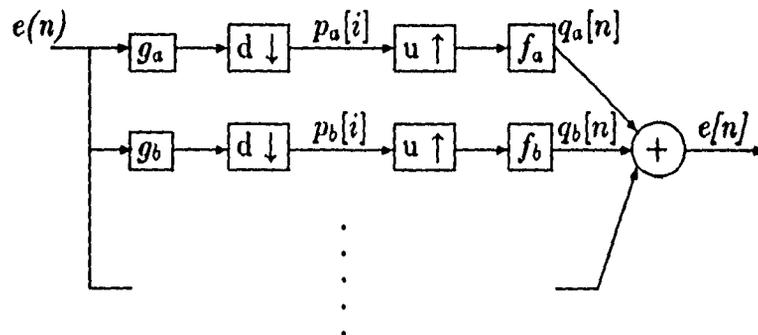
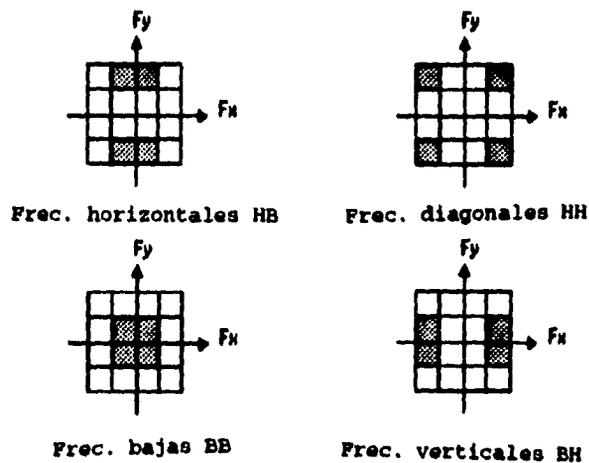


FIG. 11. Esquema de transformación

4.1 La Pirámide Separable QMF

La manera más directa de generar una transformación QMF bidimensional es aplicando dos transformaciones monodimensionales en forma separada. A partir de esta transformación se obtiene una descomposición como la que se muestra en la figura 12, en donde para un nivel determinado la imagen es descompuesta en una sub-imagen paso-bajas, una vertical, una horizontal y una diagonal; en la grafica también se muestra una rejilla que representa dos niveles piramidales. En general, una vez obtenida la descomposición en las



BH1		HH1
BH2	HW2	HB1
BB2	HB2	

Representación de la pirámide QMF (2 niveles)

FIG. 12. Partición frecuencial realizada por la pirámide QMF separable

diferentes bandas, la banda de frecuencias bajas o residuales se vuelve a procesar con la finalidad de obtener una descomposición piramidal, la cual pueda ser codificada posteriormente para propósitos de transmisión o de almacenamiento.

Para llevar a cabo la implementación de la pirámide podemos referirnos a la figura 13, la cual representa la etapa de análisis del sistema. En la figura, los filtros paso bajas H_0 y paso altas G_0 son aplicados de manera separada so-

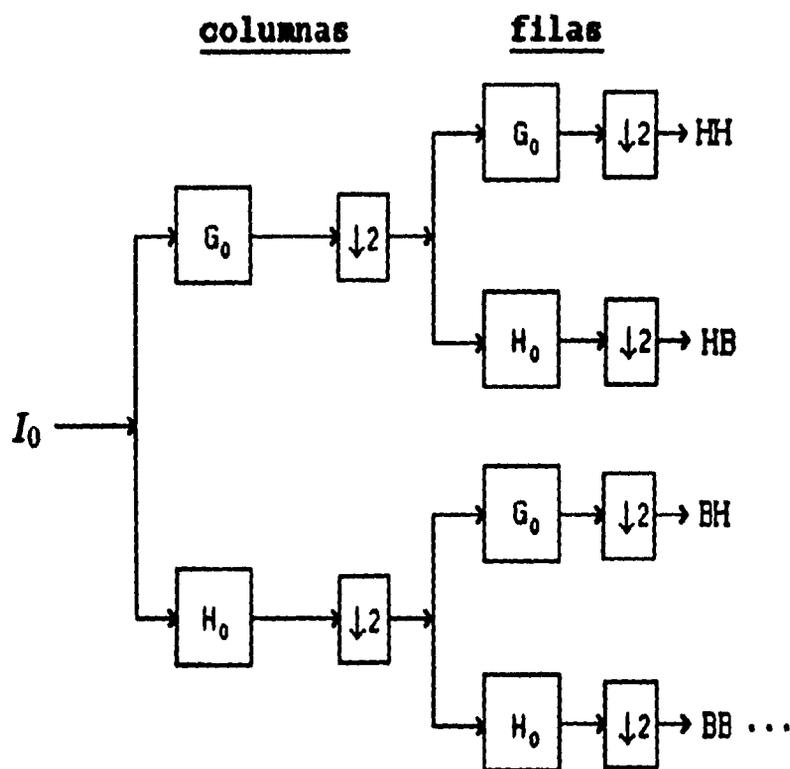


FIG. 13. Diagrama de Descomposición de la Pirámide QMF separable

bre los dos ejes coordenados de la imagen original I_0 y cada etapa de filtrado va acompañada de una decimación con factor 2, como se observa también en la figura.

Al final de una etapa de análisis las sub-imágenes obtenidas reducen su área en un cuarto con respecto al nivel anterior.

Como se mencionó anteriormente el proceso se aplica en forma iterativa sobre

la banda de bajas frecuencias, lo que permite construir los diferentes niveles de la pirámide.

Por otro lado, en la etapa de reconstrucción de la imagen se realiza un proceso inverso. A partir del último nivel de la pirámide que se obtuvo del proceso de análisis, se genera una banda residual con las cuatro bandas como se muestra en la figura 14. En este proceso de reconstrucción es necesario realizar una interpolación que compense los efectos de la decimación introducidos por la etapa de análisis. Por analogía con el proceso de descomposición, la reconstrucción que se realiza es iterativa y el tratamiento que se le da a las dos dimensiones de la imagen es en forma separada, como se ilustra también en

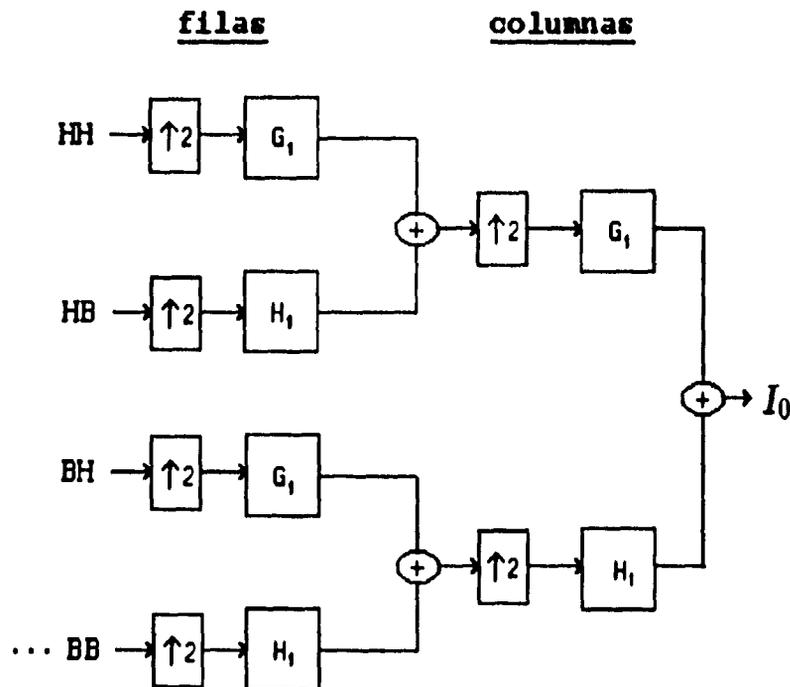


FIG. 14. Diagrama de Reconstrucción de una Señal Original a Partir de la Pirámide QMF

la figura 14. Hay que mencionar que los filtros H_0 y G_0 de la etapa de análisis y los filtros H_1 y G_1 de la etapa de síntesis, guardan una relación que permite obtener una reconstrucción perfecta de la imagen original.

Se recomienda para una mayor comprensión de éste y otros tipos de algoritmos de transformación revisar [Baaziz/Labit, 1990].

4.2 La Pirámide Quinconce

La pirámide Quinconce con filtros QMF no separables puede construirse por divisiones sucesivas del espectro de la imagen en diamantes anidados o contenidos en cuadros [Adelson/Simoncelli, 1987], tal como se observa en la figura 15. Esta muestra además el patrón de decimación de una estructura en Quinconce.

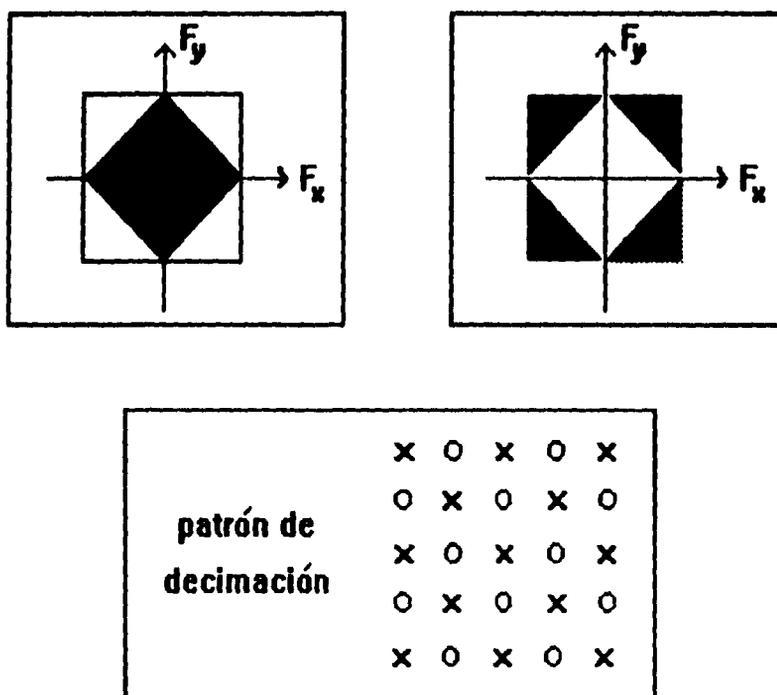


FIG. 15. Partición frecuencial realizada por la pirámide Quinconce

Por otro lado, en cada estado de la pirámide la escala lineal del "kernel" se incrementa por un factor de $\sqrt{2}$ y el tratamiento que se le da a la imagen I_0 es en dos dimensiones, obviamente por las características de los filtros. Esta transformación requiere de sucesivas decimaciones sobre una estructura quinconce [Vetterli, 1984]. Con la decimación quinconce el número de muestras de dos dimensiones es afectado por el valor de $1/2$ en cada estado. En la fig 16(a) se observa la manera de generar la pirámide Quinconce y en particular se muestra la construcción de 3 niveles de la pirámide, los cuales pueden ser codificados para transmisión o almacenamiento al igual que en la pirámide

QMF separable. En términos generales, se aplican los mismos procesos para obtener un sistema de análisis/síntesis, tanto en la pirámide Quinconce como en la pirámide QMF separable, por tal razón, en la fig. 16(b) se presenta la reconstrucción de una imagen por medio de la pirámide Quinconce y no se profundiza más sobre la función de cada uno de los bloques en los diagramas.

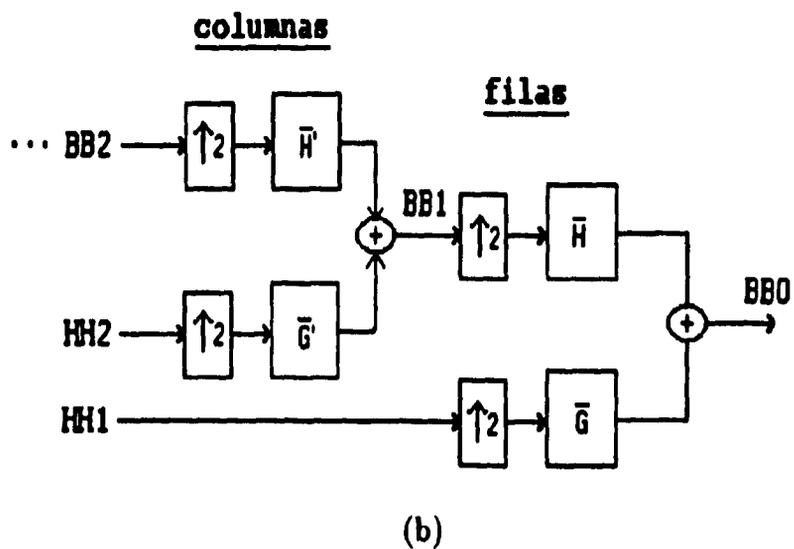
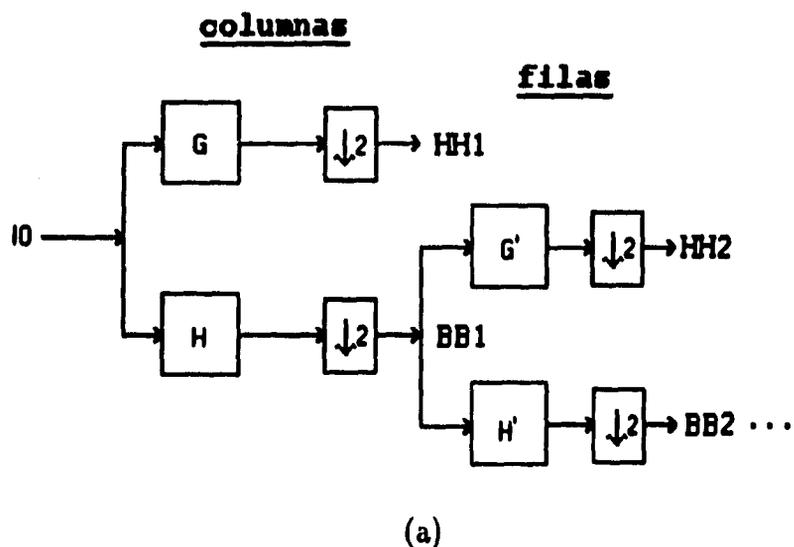


FIG. 16. Pirámide Quinconce (3 niveles). (a): análisis (b): síntesis

Capítulo 5

Selección y Simulación de un Esquema de Descomposición

La codificación de imágenes por transformaciones se realiza comunmente por una subdivisión de una imagen en bloques de 8 x 8 pixels [Karllson at el, 1988]. Cada uno de los bloques es transformado y codificado en forma independiente en la imagen. Sin embargo, el procedimiento no remueve la redundancia que existe entre bloques vecinos y su estructura puede llegar a ser visible en la imagen reconstruida cuando la compresión es alta. Una solución a este problema es la codificación sub-banda bidimensional.

Gracias a la investigación sobre filtros cuya estructura es no separable y a su aplicación en sistemas de análisis/síntesis, se ha encontrado que este tipo de filtros cuando se emplean en la codificación de imágenes, presentan propiedades direccionales que no se encuentran limitadas en las direcciones vertical y horizontal, como sucede cuando se utilizan filtros separables. Además, los filtros no separables pueden presentar una mejor selección de frecuencia, en relación a los filtros separables [Karllson at el, 1988].

Por las razones anteriormente mencionadas, se eligió simular una codificación sub-banda que utilice filtros no separables, en particular una estructura *Quin-conce* que permita cuantizar y codificar las sub-bandas con la finalidad de obtener la mayor compresión posible, sin generar una degradación severa en la imagen reconstruida.

En este caso de estudio se utilizaron dos imágenes: lena de 256 x 256 e interview de 536 x 674, ambas en 8 bits/pixel. Interview proviene de una secuencia de imágenes reales proporcionada por el CCETT (Centre Commun d'Etudes en Télévision et Télécommunications) de Cesson-Sévigné, Francia. En las figs. 17 y 18 se presentan las imágenes originales y sus respectivos histogramas.

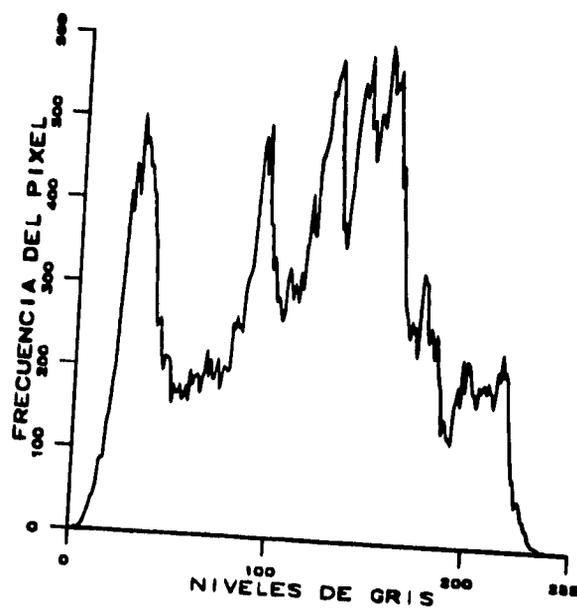


Fig. 17: Imagen original lena e histograma.

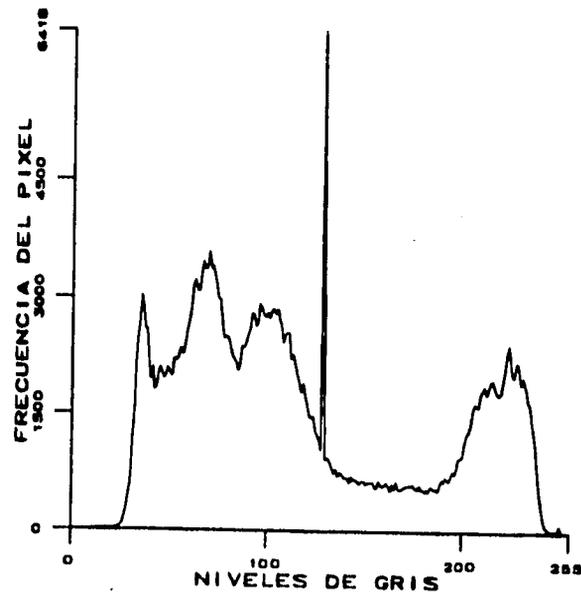


Fig. 18: Imagen original interview e histograma.

5.1 Implementación de la Pirámide Quinconce (Etapa de Transmisión)

La descomposición que se realizó sobre las imágenes para fines de transmisión fue llevada a cabo por tres etapas o bloques de simulación:

- Bloque de Análisis
- Bloque de Cuantización
- Bloque de Codificación

El esquema de codificación que se utilizó se muestra en un diagrama a bloques en la figura 19.

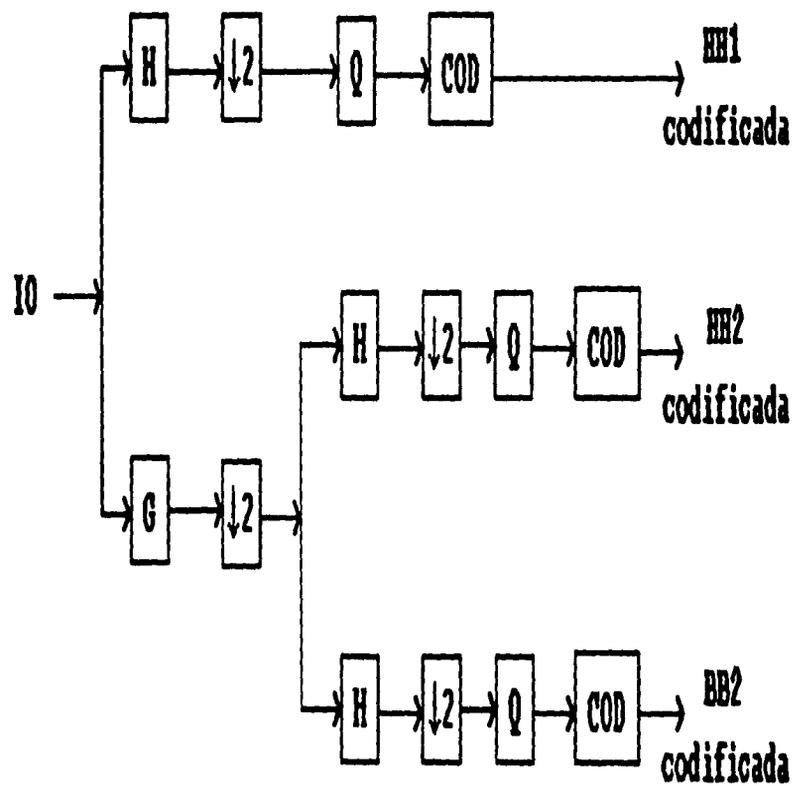


FIG. 19. Diagrama de Codificación

5.1.1 Análisis

El filtro paso bajas H (no separable) con el cual se genera la pirámide, debe tener propiedades que permitan deducir el filtro paso altas G y los filtros interpoladores \tilde{G} y \tilde{H} . El esquema de descomposición en sub-bandas empleando estos filtros debe tener una característica de reconstrucción perfecta.

Las propiedades básicas de los filtros son [Adelson/Simoncelli, 1987]:

- H es un filtro paso bajas no separable con respuesta al impulso finita.

$$h(n, m) = 0 \text{ para } |n| > N \text{ o } |m| > M$$

en donde $h(n, m)$ tiene un soporte finito de $(2N + 1)$ coeficientes horizontalmente y $(2M + 1)$ verticalmente.

- $h(n, m)$ normalizado presenta una simetría en un plano por octavas garantizando la nulidad de la fase:

$$\sum_n \sum_m h(n, m) = 1$$

$$h(|n|, |m|) = h(n, m) \quad \forall n, \forall m.$$

- El "kernel" asociado a $h(n, m)$ esta dado por:

e	a = 0.65549
f c f	b = 0.12930
f d b d f	c = -0.01294
e c b a b c e	donde d = -0.02594
f d b d f	e = -0.00118
f c f	f = -0.00156
e	

donde $h(0, 0) = a$.

- La respuesta al impulso g del filtro paso altas G se obtiene a partir de h . G y H forman un par QMF (Quadrature Mirror Filters) no separable.

$$g(n, m) = (-1)^{n+m} h(n, m)$$

$$\sum_n \sum_m g(n, m) = 0$$

El "kernel" correspondiente a este filtro es:

$$\begin{array}{cccccc}
 & & & & & -e \\
 & & & & -f & c & -f \\
 & & -f & d & -b & d & -f \\
 -e & c & -b & a & -b & c & -e \\
 & & -f & d & -b & d & -f \\
 & & & & -f & c & -f \\
 & & & & & & -e
 \end{array}$$

Construcción de niveles impares

Con la finalidad de normalizar la nomenclatura, es conveniente considerar a $BB_0(c_0, l_0)$ como la imagen original y a c_0 y l_0 como el largo y alto de la imagen respectivamente, dados en número de pixels.

Los dos primeros niveles de la pirámide se obtienen al aplicar en la imagen original BB_0 los filtros H y G seguidos de una decimación en quince, produciéndose con ello HH_1 y BB_1 . La rejilla de decimación asociada al filtro G es complementaria a la rejilla del filtro H . Consecuentemente existe un desfaseamiento entre H y G . Nótese que las sub-imágenes HH_1 y BB_1 son las altas y bajas frecuencias respectivamente.

Una vez que se realiza la operación anterior, cada una de las sub-bandas es trasladada sobre una rejilla rectangular para optimizar el espacio de memoria.

Este paso de las sub-bandas sobre una rejilla puede realizarse en sentido vertical u horizontal. Es decir, que si uno elige arbitrariamente por ejemplo el sentido horizontal, se obtienen sub-bandas como $BB_1(\frac{c_0}{2}, l_0)$ y $HH_1(\frac{c_0}{2}, l_0)$. Para el caso particular de este trabajo se eligió poner el resultado de los niveles impares sobre una rejilla rectangular vertical, en donde para la primera iteración tendremos que las bandas son $BB_1(c_0, \frac{l_0}{2})$ y $HH_1(c_0, \frac{l_0}{2})$.

El conjunto de esta manipulación sobre la imagen equivale frecuencialmente a la partición del espectro original, suponiendo que este sea cuadrado, en dos espectros triangulares que corresponden a las altas y bajas frecuencias.

La partición del espectro se ilustra en los incisos (a) y (b) de la figura 20.

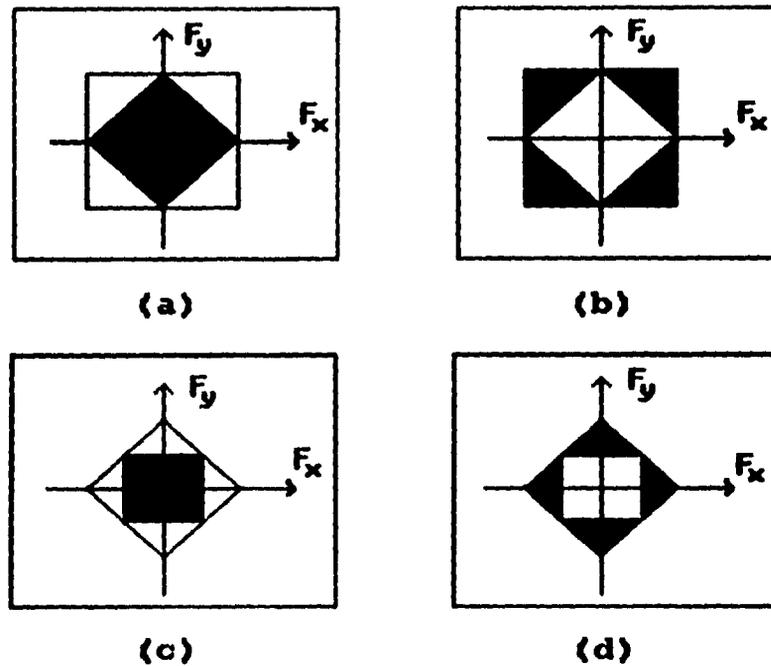


FIG. 20. Partición Frecuencial Correspondiente a una Pirámide Quinconce de Tres Niveles.

Las ecuaciones que se utilizaron para generar este proceso son las siguientes:

Para j impar:

$$\begin{aligned}
 BB_j(x, 2y) &= \sum_n \sum_m h(n, m) BB_{j-1}(2x - n, 2y - m) \\
 BB_j(x, 2y + 1) &= \sum_n \sum_m h(n, m) BB_{j-1}(2x + 1 - n, 2y + 1 - m) \\
 HH_j(x, 2y) &= \sum_n \sum_m g(n, m) BB_{j-1}(2x + 1 - n, 2y - m) \\
 HH_j(x, 2y + 1) &= \sum_n \sum_m g(n, m) BB_{j-1}(2x - n, 2y + 1 - m)
 \end{aligned}$$

donde:

$$\begin{aligned}
 0 &\leq x \leq c_j - 1 \\
 0 &\leq y \leq l_j - 1 \\
 c_j &= c_{j-1}/2 \\
 l_j &= l_{j-1}
 \end{aligned}$$

Construcción de niveles pares

La segunda iteración permite generar el segundo nivel HH_2 de la pirámide Quinconce.

Para generar los niveles pares es conveniente hacer las siguientes consideraciones:

- La descomposición del siguiente nivel se realiza siempre sobre las frecuencias bajas, en este caso BB_1 .
- La partición frecuencial que se realiza se puede observar en los incisos (c) y (d) de la figura 20.
- Los filtros que se utilizan en esta partición G' y H' se generan aplicando una rotación de 45 grados a los filtros G y H .
- Para poder aplicar los filtros H' y G' a la imagen BB_1 que se encuentra sobre una rejilla ortogonal, hay que acondicionar los filtros en un formato similar para poderlos aplicar directamente.

Una vez hechas las consideraciones anteriores, se aplican los filtros seguidos de un submuestreo vertical con un factor de dos, lo que permite generar a partir de BB_1 las dos sub-imágenes BB_2 y HH_2 .

Las ecuaciones que permiten generar los niveles pares son:

Para j par:

$$BB_j(x, y) = \sum_n \sum_m h(n, m) BB_{j-1}(x - n, 2y - m)$$

$$HH_j(x, y) = \sum_n \sum_m g(n, m) BB_{j-1}(2x - n, 2y + 1 - m)$$

donde:

$$0 \leq x \leq c_j - 1$$

$$0 \leq y \leq l_j - 1$$

$$c_j = c_{j-1}$$

$$l_j = l_{j-1}/2$$

5.1.2 Cuantización

En esta sección del sistema se elige el número de niveles del cuantizador (generalmente de 4 a 256) con los cuales se va a cuantizar cada una de las sub-bandas que serán transmitidas; también se elige el tipo de distribución dependiendo de los datos (Laplaciana o Gaussiana), a partir de la cual se generan los niveles de dicho cuantizador.

Para este efecto se eligió el algoritmo desarrollado por Lloyd-Max, del cual se puede conseguir una versión hecha en lenguaje FORTRAN 77, en Rao and Yip [1990].

En términos generales, la distribución con la cual se generan los niveles es seleccionada a partir del histograma de la sub-banda que será cuantizada, este procedimiento permite fijar la distribución; ya que para las sub-bandas de frecuencias altas ésta se aproxima a una de tipo laplaciano, mientras que para las sub-bandas de frecuencias bajas la óptima se aproxima a una gaussiana.

La entrada del cuantizador recibe muestras de valor real que son cuantizadas, de acuerdo al número de niveles que se hayan elegido. La salida son muestras de valor entero que permiten ser codificadas para su transmisión o almacenamiento.

5.1.3 Codificación

La salida del cuantizador se toma y acondiciona para efectos de transmisión. El código que se utilizó en esta parte del trabajo es un Código de Huffman con la misma estructura de árbol que se presentó en el capítulo 2, generado de la misma forma. Esta manera de generar el código de palabras (de longitud variable) no permite que ninguna palabra del código sea el prefijo de alguna otra, garantizando así una codificación única.

Una vez que se realizó la codificación se puede entonces medir la compresión que se obtuvo a partir del número de bytes necesarios para transmitir la imagen, con respecto de la original.

5.2 Implementación de la Pirámide Quinconce (Etapa de Recepción)

El sistema permite reconstruir la imagen original, obviamente con su correspondiente degradación causada por la etapa de codificación. Para lograr este objetivo se simularon los siguientes bloques:

- Bloque de Decodificación
- Bloque de Descuantización
- Bloque de Síntesis

El diagrama a bloques que se muestra en la figura 21 es el que se utilizó para la etapa de reconstrucción.

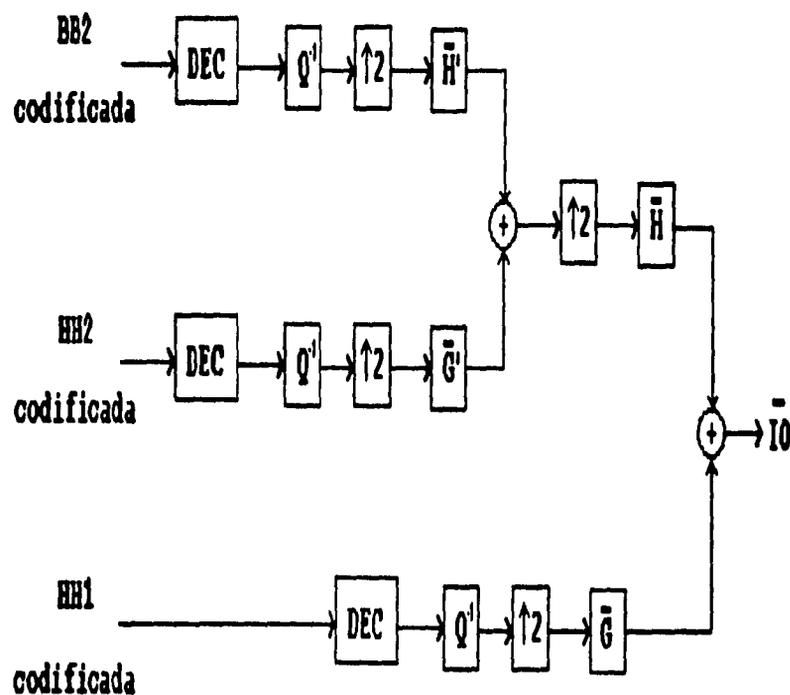


FIG. 21. Esquema de Reconstrucción.

5.2.1 Decodificación

Para llevar a cabo el proceso de decodificación, se reciben las muestras codificadas de las diferentes sub-bandas y una tabla, que se generó en el proceso de asignación del código. Empleando esta tabla se pueden recuperar los niveles de valor entero que representan la salida del decodificador.

5.2.2 Descuantización

La entrada de muestras de valor entero generadas por la etapa de decodificación son convertidas a un valor real, a través de una operación inversa a la de cuantización Q^{-1} , con lo que se obtienen las sub-bandas que permitirán generar la imagen original en el proceso de síntesis. Cabe mencionar que la información que se perdió como consecuencia de haber cuantizado las muestras, ya no se recupera por ninguno de los bloques de la etapa de recepción.

5.2.3 Síntesis

Este es el último bloque del sistema de recepción y permite mezclar las diferentes sub-bandas de la pirámide para generar la reconstrucción de la imagen original.

Por analogía con el algoritmo de descomposición, los niveles inferiores de la pirámide se generan en forma iterativa a partir de los niveles superiores. Hay que tomar en cuenta que antes de mezclar las bandas de un nivel superior, se tiene que efectuar un proceso de interpolación para ajustar los tamaños y realizar las conversiones de estructuras necesarias.

Al igual que en el proceso de descomposición, las ecuaciones para reconstruir los niveles pares e impares de la pirámide son diferentes, tal y como se muestra a continuación.

Reconstrucción de un nivel impar

Para j impar:

Hay que realizar la mezcla de HH_{j+1} y BB_{j+1} , de tamaños (c_{j+1}, l_{j+1}) y definidas sobre una rejilla ortogonal, para lograr la reconstrucción de la imagen BB_j , de tamaño $(c_j, l_j) = (2c_{j+1}, l_j)$ y definida sobre una rejilla quin-

conce.

Para obtener BB_j es necesario que los soportes de BB_{j+1} y HH_{j+1} sean dilatados a través de un ajuste con ceros.

$$\begin{aligned} BB'_{j+1}(2x, y) &= BB_{j+1}(x, y) \\ BB'_{j+1}(2x + 1, y) &= 0 \\ HH'_{j+1}(2x, y) &= 0 \\ HH'_{j+1}(2x + 1, y) &= HH_{j+1}(x, y) \end{aligned}$$

donde:

$$\begin{aligned} 0 &\leq x \leq c_{j+1} - 1 \\ 0 &\leq y \leq l_{j+1} - 1 \end{aligned}$$

En seguida se filtra y sobremuestra las imágenes que fueron dilatadas con \bar{H}' y \bar{B}' como se muestra:

$$\begin{aligned} \bar{g}'(n, m) &= g'(-n, -m) \\ \bar{h}'(n, m) &= h'(-n, -m) \\ IBB_{j+1}(x, y) &= 2 \sum_n \sum_m h'(n, m) BB'_{j+1}(x - n, y - m) \\ IHH_{j+1}(x, y) &= 2 \sum_n \sum_m g'(n, m) HH'_{j+1}(x - n, y - m) \end{aligned}$$

donde:

$$\begin{aligned} 0 &\leq x \leq 2c_{j+1} - 1 \\ 0 &\leq y \leq l_{j+1} - 1 \end{aligned}$$

Finalmente, las dos imágenes que fueron interpoladas se suman para obtener BB_j .

$$BB_j(x, y) = IBB_{j+1}(x, y) + IHH_{j+1}(x, y)$$

Reconstrucción de un nivel par

Para j par:

Se tiene que llevar a cabo una interpolación para después sumar las dos imágenes BB_{j+1} y HH_{j+1} , definidas cada una sobre una rejilla quinconce de tamaño (c_{j+1}, l_{j+1}) , para obtener una imagen BB_j , definida sobre una rejilla ortogonal y de tamaño $(c_j, l_j) = (c_{j+1}, 2l_{j+1})$.

Como en el caso anterior la interpolación se efectúa a través de una dilatación y de un filtrado adecuado.

- Dilatación:

$$\begin{aligned} BB'_{j+1}(x, 2y) &= BB_{j+1}(x, y) \\ BB'_{j+1}(x, 2y + 1) &= 0 \\ HH'_{j+1}(x, 2y) &= 0 \\ HH'_{j+1}(x, 2y + 1) &= HH_{j+1}(x, y) \end{aligned}$$

- Filtrado:

$$\begin{aligned} \bar{h}(n, m) &= h(-n, -m) \\ \bar{g}(n, m) &= g(-n, -m) \\ IBB_{j+1}(x, y) &= 2 \sum_n \sum_m \bar{h}(n, m) BB'_{j+1}(x - n, y - m) \\ IHH_{j+1}(x, y) &= 2 \sum_n \sum_m \bar{g}(n, m) HH'_{j+1}(x - n, y - m) \end{aligned}$$

donde:

$$\begin{aligned} 0 &\leq x \leq c_{j+1} - 1 \\ 0 &\leq y \leq 2l_{j+1} - 1 \end{aligned}$$

Por último, las dos imágenes que fueron interpoladas se suman para obtener BB_j .

$$BB_j(x, y) = IBB_{j+1}(x, y) + IHH_{j+1}(x, y)$$

La simulación del algoritmo de compresión anteriormente presentado, se desarrolló bajo el siguiente contexto de trabajo: estaciones de trabajo SUN y lenguaje de programación C.

Los resultados obtenidos se presentan en el siguiente capítulo; estos permitirán llevar a cabo una evaluación del algoritmo en términos de compresión y de la calidad subjetiva de la imagen.

Capítulo 6

Análisis y Evaluación de Resultados

Para poder presentar los resultados obtenidos en una forma más objetiva y fácil de entender, es necesario considerar lo siguiente:

- La descomposición que se obtiene al aplicar el algoritmo es como la que se presenta en la figura 22. En este caso se utilizó la imagen Interview para generar los tres primeros niveles de la pirámide Quinconce.
- Cada una de las sub-bandas obtenidas es cuantizada de acuerdo a su histograma:
 - Bandas de bajas frecuencias: Se utiliza un cuantizador que genera sus niveles a partir de una distribución gaussiana, dichos niveles se encuentran en un rango de entre 4 y 256.
 - Bandas de altas frecuencias: La distribución del cuantizador se eligió laplaciana y con niveles de 4 a 256.

La estrategia de cuantización elegida, la codificación por código de Huffman y las características particulares del sistema de A/S de la pirámide Quinconce, permitieron obtener los resultados que serán presentados en esta sección.

No se hace un análisis exhaustivo de todas las combinaciones posibles que se pueden realizar, sólo se utilizan las descomposiciones de 3 y 5 bandas con niveles de cuantización de entre 4 y 32 para las diferentes bandas.

Se eligieron 3 y 5 bandas haciendo la consideración de que para un caso de descomposición con filtros separables, esto sería cercano a obtener el primero y el segundo de los niveles de una pirámide con 4 y 7 bandas respectivamente.



FIG. 22. Descomposición Quince aplicada sobre la imagen interview

Por otro lado, la selección del número de niveles con los cuales se realiza la cuantización esta apoyada sobre una base de calidad visual de las imágenes utilizadas. Es decir, este rango en el número de niveles utilizados, permitió observar las transiciones de los efectos visuales causados por la cuantización de las diferentes bandas de frecuencia. Además, como la finalidad de este trabajo es obtener la mayor compresión posible conservando determinada calidad en la imagen reconstruida, no resulta conveniente utilizar un mayor número de niveles de cuantización, ya que a partir de 32 los efectos causados por esta desaparecen (visualmente) y la compresión lograda comienza a disminuir.

TERCER NIVEL PIRAMIDAL (LENA)					
NIVELES DEL CUANTIZADOR		TASA DE COMPRESION (b/p)	RAZON DE COMPRESION	NMSE (%)	NUMERO DE FIGURA
sub-bandas altas	sub-bandas bajas				
4	16	0.260437	30.7176	13.108010	23
8	16	0.273819	29.216383	1.192467	25
16	16	0.340317	23.50751	0.688879	-
32	16	0.458374	17.452996	0.626219	27
4	32	0.294830	27.134251	13.305424	29
8	32	0.30821	28.956136	0.747022	31
16	32	0.374710	21.349839	0.240774	-
32	32	0.492767	16.234842	0.181829	33

TABLA 2

6.1 Caso de estudio No. 1

En esta sección se presentan los resultados obtenidos al reconstruir la imagen Lena a partir de un tercer nivel piramidal, respetando lo que se estableció en relación a los niveles de cuantización y a las sub-bandas generadas.

Como se puede observar en la tabla 2 el NMSE (error medio cuadrático normalizado) que se originó al reconstruir la imagen, presenta un cambio considerable al llegar a niveles de cuantización de 4 y 16 ó de 4 y 32 para las sub-bandas de altas y bajas frecuencias respectivamente, con respecto

al comportamiento que éste presenta cuando se utilizaron otros niveles de cuantización. Dicho cambio en el NMSE se ve reflejado en la calidad de las imágenes reconstruidas, ya que estas pasaron de una calidad visual aceptable a una calidad inaceptable (ver figs. 23 y 29).

Las pruebas realizadas muestran que el proceso aplicado a las imágenes provoca dos efectos particulares :

- Falsos contornos
- Ruido de cuantización

Los falsos contornos fueron originados por la cuantización de las sub-bandas de bajas frecuencias y aparecen cuando el número de niveles del cuantizador es inferior a 32. Por otro lado, el ruido de cuantización se presenta cuando se cuantiza con menos de 16 niveles sobre las sub-bandas de altas frecuencias. La reconstrucción de Lena a partir de los primeros tres niveles, permite afirmar que la máxima compresión que se logro en este caso, conservando una muy buena calidad en la imagen reconstruida, es de 0.3747 bits/pixel.

Es importante aclarar que para esta prueba y las siguientes, la determinación de umbrales en donde aparecen los efectos ya mencionados y la máxima compresión, son obtenidos experimentalmente.

La presentación de las imágenes reconstruidas (figs. 23 a 34) muestra la evolución gráfica y visual de los resultados contenidos en la tabla 2.

Cada una de las imágenes contiene diferentes deformaciones. La fig. 27 representa los falsos contornos, las figs. 29 y 31 el ruido de cuantización siendo éste más notable en la fig. 29 y las figs. 23 y 25 muestran una combinación de ambos efectos, los cuales alteran en forma más severa a la imagen de la fig. 23.

La imagen en la cual no se observan los efectos mencionados y que representa una de las mejores reconstrucciones que se lograron con el algoritmo, se puede ver en la fig. 33.



FIG.23. Reconstrucción de Lena con NMSE=13.108.

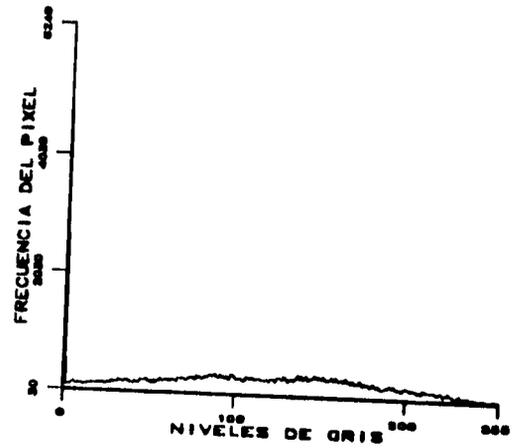


FIG.24: Histograma de Lena con NMSE=13.108.



FIG.25. Reconstrucción de Lena con NMSE=1.1924.

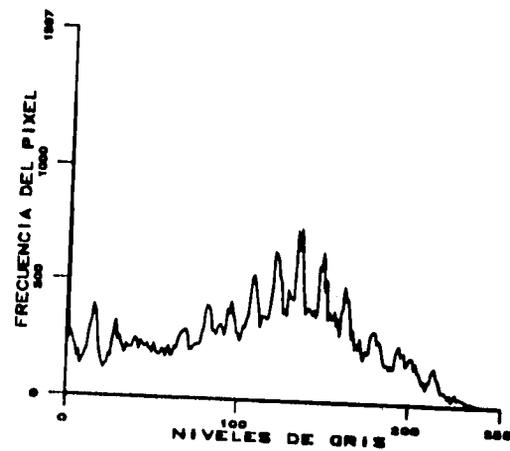


FIG.26: Histograma de Lena con NMSE=1.1924.



FIG.27. Reconstrucción de Lena con NMSE=0.6262.

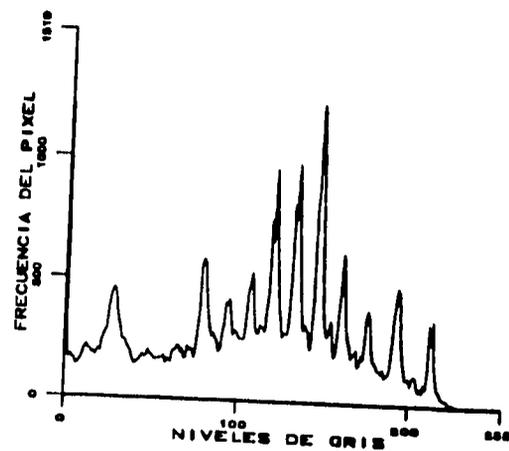


FIG.28: Histograma de Lena con NMSE=0.6262.



FIG.29. Reconstrucción de Lena con NMSE=13.305.

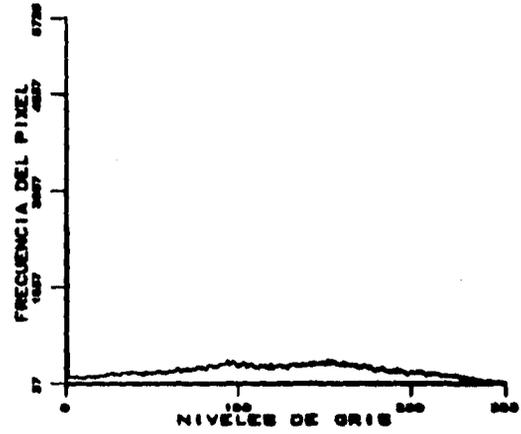


FIG.30: Histograma de Lena con NMSE=13.305.



FIG.31. Reconstrucción de Lena con NMSE=0.7470.

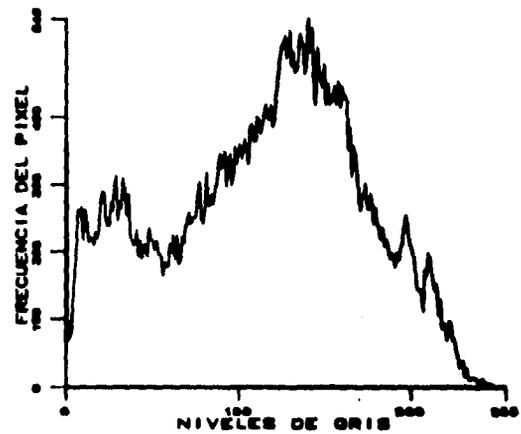


FIG.32: Histograma de Lena con NMSE=0.7470.



FIG.33. Reconstrucción de Lena con NMSE=0.1818.

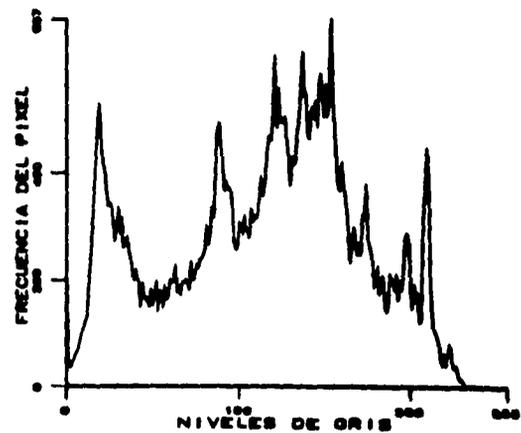


FIG.34: Histograma de Lena con NMSE=0.1818.

6.2 Caso de estudio No. 2

Las características del procesamiento aplicado sobre Interview son las mismas que se usaron para el caso de Lena. El propósito de esto es realizar una comparación en términos visuales y de compresión.

TERCER NIVEL PIRAMIDAL (INTERVIEW)					
NIVELES DEL CUANTIZADOR		TASA DE COMPRESION (b/p)	RAZON DE COMPRESION	NMSE (%)	NUMERO DE FIGURA
sub-bandas altas	sub-bandas bajas				
4	16	0.255138	31.35563	26.44556	35
8	16	0.260483	30.71220	1.578087	37
16	16	0.30885	25.90634	0.696927	-
32	16	0.42542	18.80493	0.622680	39
4	32	0.286743	27.89952	27.93229	41
8	32	0.292088	27.38897	1.293737	43
16	32	0.34041	23.50104	0.261590	-
32	32	0.45702	17.50447	0.185823	45

TABLA 3

Los resultados de esta simulación fueron muy parecidos a los que se obtuvieron en la anterior. Esto es, los efectos de falsos contornos y los de ruido de cuantización se presentaron al reconstruir las imágenes a partir de los mismos niveles de cuantización que para el caso de Lena, sin embargo, dichas alteraciones se aprecian con mayor claridad al reconstruir la imagen Interview que al reconstruir la imagen Lena. Esto se debe a la diferencia en

tamaño de una imagen con respecto a la otra.

Comparando las tablas 2 y 3 se puede apuntar que la magnitud de los errores y de la compresión son muy similares en ambas tablas, consecuentemente la calidad visual de las imágenes reconstruidas no difiere demasiado en los dos casos de estudio.

Por otro lado, la compresión máxima que se obtuvo al reconstruir la imagen Interview fue de 0.3404 bits/pixel con una calidad visual muy aceptable. Las figs. de la 35 a la 46 son los resultados de las imágenes reconstruidas y muestran la transición de los efectos provocados por el proceso.

Dados estos resultados, se decidió que la siguiente prueba sería aplicar el algoritmo en la reconstrucción de imágenes a partir de un quinto nivel pirámidal, conservando la cuantización con la cual se ha obtenido una mejor calidad visual en ambas imágenes. Por tal motivo, las diferencias y los efectos causados por el proceso sólo serán puntualizados, pero no se presentará ninguna gráfica o figura de la reconstrucción, a excepción de los casos ya mencionados de mejor calidad subjetiva.



FIG.35. Reconstrucción de Interview con NMSE=26.4455.

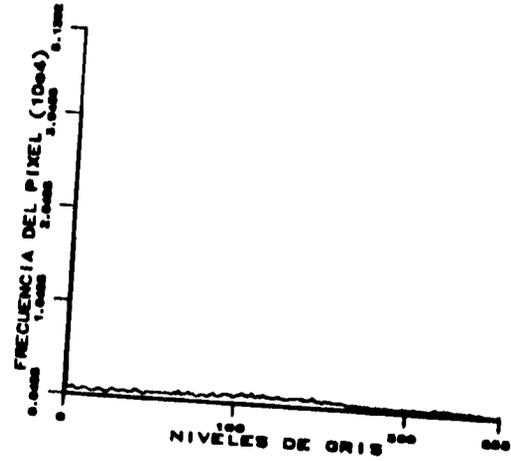


FIG.36: Histograma de Interview con NMSE=26.4455.



FIG.37. Reconstrucción de Interview con NMSE=1.5780.

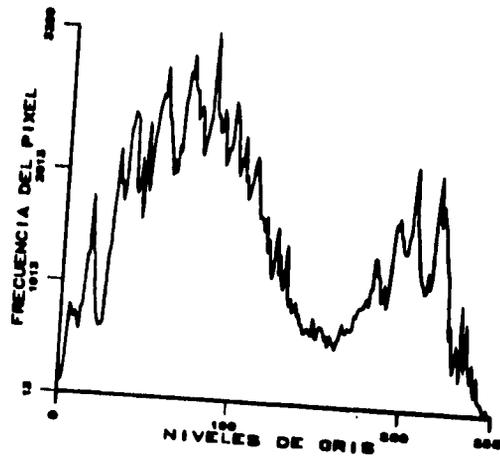


FIG.38: Histograma de Interview con NMSE=1.5780.



FIG.39. Reconstrucción de Interview con NMSE=0.6226.

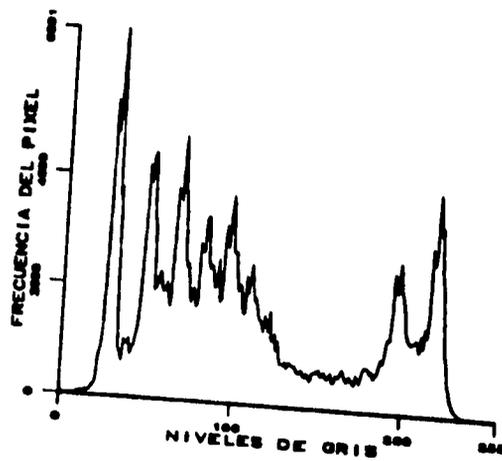


FIG.40: Histograma de Interview con NMSE=0.6226.



FIG.41. Reconstrucción de Interview con NMSE=27.9322.

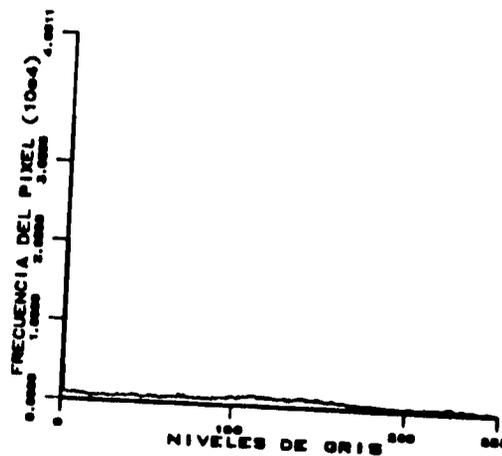


FIG.42: Histograma de Interview con NMSE=27.9322.



FIG.43. Reconstrucción de Interview con NMSE=1.2937.

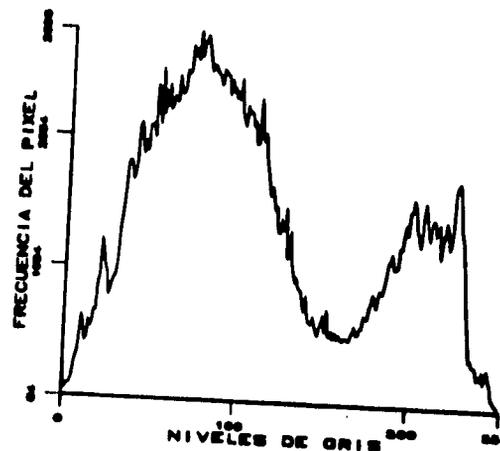


FIG.44: Histograma de Interview con NMSE=1.2937.



FIG.45. Reconstrucción de Interview con NMSE=0.1858.

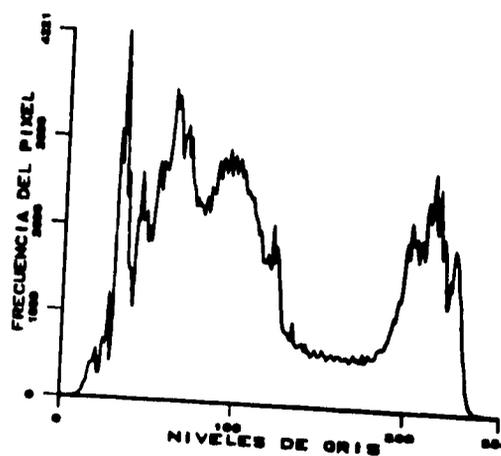


FIG.46: Histograma de Interview con NMSE=0.1858.

QUINTO NIVEL PIRAMIDAL				
	TASA DE COMPRESION (b/p)	RAZON DE COMPRESION	NMSE (%)	NUMERO DE FIGURA
LENA	0.329819	24.255748	0.351758	47
INTERVIEW	0.285824	27.989231	0.631520	48

TABLA 4

6.3 Caso de estudio No. 3

La evaluación visual que se realizó para ambas imágenes reconstruidas a partir de un quinto nivel piramidal, mostró que los efectos de falsos contornos y los de ruido de cuantización son más críticos conforme se aumenta el nivel de la piramide, como consecuencia las imágenes recostruidas tuvieron una calidad inferior con respecto a los dos primeros casos de estudio. Sin embargo, se puede afirmar que las imágenes reconstruidas a partir de niveles de cuantización de 16 para las sub-bandas de altas frecuencias y de 32 para las sub-bandas de bajas frecuencias, conservan una buena calidad.

Los resultados que se obtuvieron para los casos de máxima compresión con una buena calidad en las imágenes reconstruidas se pueden observar en las figs. 47 y 48. En estas se puede observar que la calidad es ligeramente inferior en relación a los casos presentados con anterioridad. Las imágenes de esta sección fueron cuantizadas con niveles de 16 para las sub-bandas de altas frecuencias y de 32 para las sub-bandas de bajas frecuencias. La compresión y el error se muestran en la tabla 4.

La ventaja de utilizar el quinto nivel piramidal, fué que en términos generales se logró una mayor compresión. Aunque para alguna aplicación práctica, se tendría que evaluar si el tiempo de procesamiento y la degradación de las imágenes reconstruidas justifican la pequeña compresión extra que se consigue usando este quinto nivel.

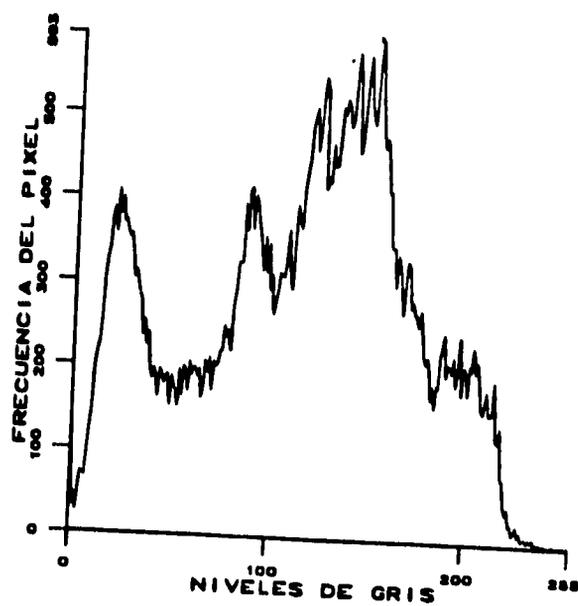


Fig. 47: Lena con $NMSE=0.3517$ e histograma.

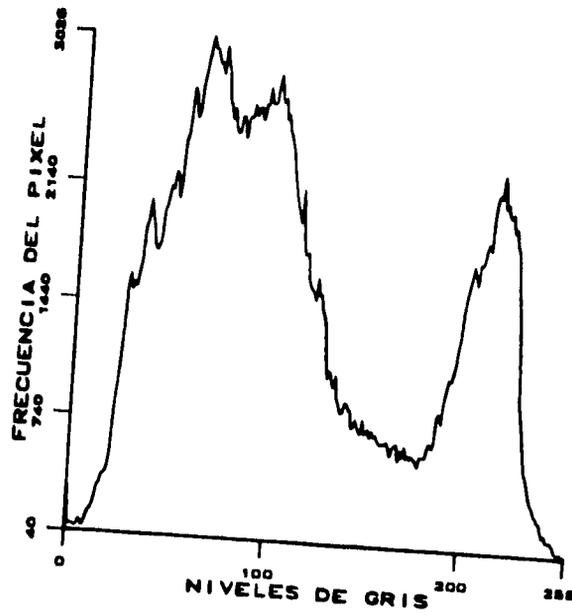


Fig. 48: Interview con $NMSE=0.6315$ e histograma.

Capítulo 7

Conclusiones

Las propiedades de la transformación sub-banda que se mencionaron en este trabajo: localización espacial, localización frecuencial, escalabilidad y ortogonalidad. Permiten aseverar que las transformaciones sub-banda que utilizan bancos de Filtros Espejo en Cuadratura (QMF) dentro de su esquema, son óptimas para la codificación de imágenes fijas, ya que estos bancos de filtros cumplen con las propiedades antes mencionadas. Es importante mencionar que la aplicación de filtros cuya estructura es separable proporciona una orientación específica en algunas de las sub-bandas, pero no en todas. En contraparte los filtros con estructura no separable basados en un muestreo Quinconce ofrecen una orientación específica en todas las sub-bandas, sin embargo, su implementación es más complicada.

Por otro lado, hay que resaltar que el sistema de Análisis/Síntesis que se utilizó en el presente trabajo como una parte fundamental del sistema de codificación, constituye un sistema de reconstrucción perfecta, es decir, sin pérdidas. Sin embargo, el sistema de codificación en su conjunto se clasifica en un sistema con pérdidas, ya que en la búsqueda de una mayor compresión de las imágenes procesadas, se tuvo que implementar una etapa de cuantización, la cual como ya se mencionó anteriormente ocasiona pérdida de información que no es posible recuperar por ninguna de las etapas del sistema.

Consecuentemente los efectos que se observaron en las imágenes reconstruidas son el resultado de tales pérdidas de información, estos efectos o defectos que se presentaron en las imágenes reconstruidas se pueden clasificar en: falsos contornos y un efecto de ruido de cuantización. Los falsos contornos y el ruido de cuantización aparecen cuando se realiza una cuantización en forma

severa sobre las sub-bandas de bajas frecuencias y las sub-bandas de altas frecuencias respectivamente.

Sobre el ruido de cuantización se puede afirmar que este es acumulativo, es decir, que el efecto que se ocasiona al obtener el primer nivel de la pirámide se suma al siguiente nivel de la descomposición, estos dos efectos al siguiente nivel y así sucesivamente.

Considerando que existen imágenes de muy diferente naturaleza es conveniente puntualizar que el sistema de codificación presentado, no podría ser utilizado en su totalidad por imágenes que requieran de técnicas libres de errores, un ejemplo de esto son las imágenes biomédicas en donde sería muy grave un error causado por el sistema de procesamiento empleado, sin embargo, sería interesante probar con la ayuda de un especialista en este tipo de imágenes hasta que punto es aceptable el sistema desarrollado en este trabajo.

Gracias a la representación multiescala que se obtiene como resultado de una descomposición en sub-bandas es posible una transmisión progresiva de la información, la cual tiene las siguientes características: es una transmisión en donde, en un primer paso, se transmite parte de la información, la correspondiente a las bajas frecuencias (zonas uniformes de las imágenes) y en pasos posteriores se transmite la información faltante, la correspondiente a las altas frecuencias (contornos o bordes de las imágenes).

Finalmente, otro aspecto de interés, es la evaluación de los niveles piramidales de una descomposición Quinconce o algunos de ellos, con la finalidad de determinar si estos son de utilidad para las otras áreas del procesamiento de imágenes, por ejemplo en procesos de segmentación.

ESTA TESIS NO DEBE
SALIR DE LA BIBLIOTECA

Bibliografía

- [Adelson / Simoncelli, 1987] Adelson E.H. and Simoncelli E., "Ortogonal pyramid transforms for image coding", Proc. SPIE Conf. Visual Commun. Image Proc., vol. 845, Cambridge, MA, Oct. (1987), pp 50-58
- [Baaziz / Labit, 1990] Baaziz et C. Labit, "Transformations pyramidales d'images numériques", IRISA, Campus de Beaulieu 35042 Rennes Cedex France, Publication Interne No. 526, Mars (1990).
- [Bellanger, 1984] Bellanger M. "Digital processing of signal", John Wiley & Sons, (1984)
- [Crochiere / Rabiner, 1993] Crochiere, R.E., and Rabiner, L.R., "Multirate digital signal processing", Englewood Cliffs, NJ: Prentice Hall, (1983)
- [Esteban / Galand, 1977] Esteban, D., and Galand, C. "Application of quadrature mirror filters to Split band voice coding schemes", Proc. ICASSP, pp 191-195 (1977)

- [Gabor, 1946] Gabor D., "Theory of communication", J. IEE, vol. 93, pp. 457-492, (1946)
- [Gharavi / Tabatabai, 1987] Gharavi, H., and Tabatabai A., "Application of quadrature mirror filtering to the coding of monochrome and color image", Proc. ICASSP, pp 2384-2387 (1987)
- [Gonzalez / Wintz, 1988] Gonzalez R. C. and Wintz P., "Digital Image Processing", Addison - Wesley, Reading, Mass. (1988)
- [Karlsson et al, 1988] Karlsson G. Vetterli M. and Kovacevic, "Non separable two-dimensional perfect reconstruction filter banks", Proc. SPIE, Visual Commun Image Proc., vol. 1001, pp. 187-199, (1988)
- [Lim, 1990] Lim, J.S. "Two dimensional signal and image processing", Englewood Cliffs, NJ: Prentice Hall (1990)
- [Lynch, 1985] Lynch, T. J. "Data compression techniques and application", NJ: Ban Nostran Reinhold Company, (1985)
- [Mallat, 1987] Mallat, S.G., "A theory for multiresolution signal decomposition: the wevelet representation", GRASP Lab Technical Memo MS-CIS-87-22, University Of Pennsylvania Dept. Of Computer and Information Science (1987)
- [Pratt, 1979] Pratt W. K., "Image transmission techniques", NJ:Academic , (1979)

- [Rabbani/Jones, 1991] Rabbani, M., Jones P., W. "Digital image compression techniques", Bellingham, Washington: SPIE, (1991) (Tutorial Texts Inc. Optical Engineering, vol. TT7)
- [Rao/Yip, 1990] Rao K. R. and Yip P. "Discrete Cosine transform algorithms, advantages, applications", Academic (1990)
- [Rosenfeld, 1982] Rosenfeld, A. "Digital picture processing", vol. 2, Academic Press Orlando, Second Edition.
- [Tran et al, 1987] Tran, A., Liu, K., Tzou, K., and Vogel, E., "An efficient pyramid image coding scheme", Proc. ICASSP, pp. 744-747, (1987)
- [Vaindynathan, 1987] Vaindynathan, P.P. "Quadrature mirror filter banks, M - band extensions and perfect - reconstruction techniques", IEEE, ASSP magazine, vol. 4, pp. 4-20, (1987) b.
- [Vaindynathan, 1993] Vaindynathan, P.P. "Multirate Systems and Filter banks", Englewood Cliffs, NJ: Prentice Hall (1993)
- [Vetterli, 1984] Vetterli, M. "Multidimensional sub-band coding: some theory and algorithms", Signal Processing, Vol. 6, pp. 97-112 (1984)
- [Vetterli, 1986] Vetterli, M. "Filter banks allowing perfect reconstruction", Signal Processing, vol. 10, pp. 219-244 (1986)

[Woods / O' Neil, 1986]

Woods, J.W. and O' Neil, S. O. "Subband coding of images", IEEE Trans, Acoust-Speech and Sig. Proc. V. ASSP-34, pp 1278-1288 (1986)

[Woods, 1991]

Woods, J.W. "Subband image coding", Kluwer Academic Publishers, Inc. (1991)