

01170
8



UNIVERSIDAD NACIONAL AUTONOMA
DE MEXICO

FACULTAD DE INGENIERIA
DIVISION DE ESTUDIOS DE POSGRADO

COMPRESION DE IMAGENES BASADA EN
MODELOS GAUSSIANOS DE PERCEPCION
VISUAL

T E S I S

Que para obtener el grado de
MAESTRIA EN INGENIERIA ELECTRICA
CON ESPECIALIDAD EN PROCESAMIENTO
DIGITAL DE IMAGENES

presenta

JOSE LUIS SILVAN CARDENAS

DIRECTOR DE TESIS: DR. BORIS ESCALANTE RAMIREZ



México, D. F.

2002

TESIS CON
FALLA DE ORIGEN



Universidad Nacional
Autónoma de México



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.

TESIS CON
FALLA DE ORIGEN

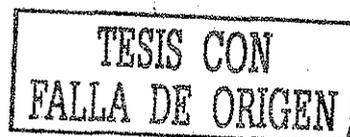
TESIS CON
FALLA DE ORIGEN

A mi familia

TESIS CON
FALLA DE ORIGEN

Tabla de Contenido

Tabla de Contenido	v
Lista de Figuras	viii
Resumen	x
Agradecimientos	xiii
1 Introducción	1
1.1 Imagen digital	2
1.2 Clasificación de las técnicas de compresión	4
1.3 Teoría taza-distorsión	6
1.4 Compresión con pérdidas	8
1.5 Compresión por el sistema visual	11
1.6 Imágenes de prueba	13
2 Transformada de Hermite multiescala	15
2.1 Introducción	16
2.2 Caso 1-D: Transformada de Hermite Multiescala	17
2.2.1 Descomposición multiescala	17
2.2.2 Análisis estructural	18
2.2.3 Definición de la THM	19
2.2.4 Acotamiento del rango de escalas	20
2.2.5 Esquema piramidal continuo	21
2.2.6 Predicción entre niveles	23
2.3 Caso 2-D: Transformada de Hermite multiescala-multidireccional	24
2.3.1 Descomposición multiescala de señales 2-D	24
2.3.2 Análisis estructural de señales 2-D	25
2.3.3 Definición de la THMM	26



2.3.4	Filtrado direccional	27
2.3.5	Aplicaciones	28
2.4	Caso discreto: THDM y THDMM	31
2.4.1	Espacio-escala discreto	31
2.4.2	Derivadas discretas	33
2.4.3	Definición de la THDM	33
2.4.4	Un algoritmo rápido	34
2.4.5	Definición de la THDMM	37
2.5	Conclusión	39
3	Clasificación de estructuras en el dominio de la THDR	42
3.1	Introducción	43
3.2	Descomposición de la imagen	44
3.3	Reconstrucción de la imagen	46
3.4	Latís de muestreo	47
3.5	Rotación local	48
3.6	Clasificación	50
3.7	Conclusión	54
4	Compresión basada en la descomposición THD-Quadtree	57
4.1	Introducción	58
4.2	Descomposición THD-Q	59
4.2.1	Umbral de homogeneidad	60
4.2.2	Reconstrucción THD-Q	61
4.2.3	Información lateral	61
4.3	Cuantización vectorial	61
4.3.1	Inicialización del libro de códigos	63
4.4	Resultados	64
4.5	Conclusión	66
5	Discusión	69
A	Caso 1-D	76
B	Caso 2-D	78
C	Caso discreto	81
D	Rotación de las DG	84

Índice de Figuras

1.1	Curva taza-distorsión típica para una fuente discreta de alfabeto finito.	7
1.2	Estructura general de los esquemas de compresión con pérdidas. . . .	9
1.3	Imágenes de prueba.	14
2.1	Esquema piramidal de análisis y síntesis de la THM.	22
2.2	Reconstrucción de un borde gaussiano.	30
2.3	Descomposición piramidal THDM de la imagen LENA.	35
2.4	Reconstrucción piramidal THDM de la imagen LENA y error de re- construcción	36
2.5	Descomposición piramidal THDM de la imagen HOUSE.	38
2.6	Descomposición piramidal THDMM de la imagen HOUSE.	39
2.7	Imágenes reconstruidas y errores para los casos de las Figuras 2.5 y 2.6 respectivamente.	40
3.1	Distribución de los operadores sobre rejillas de muestreo cuadradas. .	48
3.2	Umbral de detección de contraste (a) y umbral de enmascaramiento de contraste (b).	51
3.3	Resultados de la clasificación basada en el modelo de percepción. . . .	55
3.4	Imagen HOUSE codificada y error de reconstrucción para: la THD estándar (arriba), la THD con la latís quincunx (en medio) y la THD con clasificación perceptual sobre la latís quincunx (abajo).	56
4.1	Ejemplo de un quadtree de profundidad 3 y 13 nodos hoja.	62

4.2	Curvas de desempeño del codificador DHT-Q para las imágenes PEPPER, HOUSE y CAMERA.	65
4.3	Segmentación quadtree de las imágenes PEPPER, HOUSE, CAMERA Y BARBARA.	67
4.4	Resultados de compresión con la THD-Q (a)-(b) y con la TCD (c)-(d)	68
D.1	Mariposa del filtro binomial generalizado.	85

Resumen

El objetivo del presente trabajo consiste en aplicar ciertos modelos de procesamiento y percepción de estímulos visuales al problema de la compresión de imágenes naturales. Las imágenes reconstruidas deberían permitir al lector evaluar la calidad subjetiva obtenida con las diferentes técnicas de compresión presentadas. Sin embargo, debe tomarse en cuenta que las limitaciones del proceso de impresión ocultan detalles que sólo son evidentes en otros sistemas de despliegue.

En el Capítulo 1 se presenta un resumen sobre compresión de imágenes digitales, incluyendo definiciones elementales relacionadas con la teoría de la información, así como la clasificación de las técnicas de compresión y una descripción de los procesos que lleva a cabo el sistema visual humano (SVH) para la compresión de los datos visuales. Las ideas expuestas en esta última parte establecen el punto de partida del presente trabajo, ya que los esquemas de compresión propuestos se basan en modelos de ciertas etapas de procesamiento del SVH.

En el Capítulo 2 se deriva un modelo multicanal para la representación de imágenes naturales basado en la teoría espacio-escala. El modelo tiene una inspiración biológica de tal forma que sus parámetros pueden definirse de acuerdo a ciertas propiedades del sistema visual. Primero se estudia el caso unidimensional (1-D) y posteriormente se extienden los resultados a dos dimensiones. Los operadores que se obtienen tanto en la fase de análisis como en la fase de síntesis son derivadas de Gaussiana, de tal forma que, para el caso bidimensional (2-D), se tiene la posibilidad de representar la imagen en función de la respuesta de filtros direccionales. Al sustituir la normalización de coordenadas por un muestreo se obtiene un análisis de la señal similar al que lleva a

cabo la transformada de Hermite jerárquica (THJ). También se define la versión discreta de las transformadas multiescala y multidireccional a partir de la representación espacio-escala de señales discretas en la cual se mantienen muchas de las propiedades del caso continuo. En esta parte también se discuten las consideraciones prácticas para su implantación en un sistema digital .

En el Capítulo 3 se presenta un esquema de representación de imágenes basado en una clasificación perceptual de los coeficientes de la transformada de Hermite discreta rotada (THDR) sobre una latís de muestreo aproximadamente hexagonal. Las funciones de análisis de la THD corresponden a funciones binomiales que aproximan a las derivadas de Gaussiana con la ventaja de que pueden ser calculadas mediante un algoritmo rápido. El muestreo hexagonal es muy eficiente en la transferencia de información espacial y además tiene gran relevancia en el SVH. Los coeficientes de un mismo orden se mapean mediante una transformación unitaria especificada localmente. La transformación se basa en funciones binomiales generalizadas, y como resultado se obtienen los coeficientes de la THD referidos a un sistema de coordenadas rotado de tal forma que la energía de los coeficientes se concentra a lo largo de una coordenada. Dicha representación permite un proceso de clasificación perceptual en el cual se define un umbral de error máximo tolerado para la aproximación del bloque por una constante (patrón 0-D) o por una estructura orientada (patrón 1-D). La compresión se consigue, entonces, eliminando los coeficientes que son visualmente irrelevantes.

En el Capítulo 4 se define un algoritmo de descomposición multiescala no redundante basado en una estructura de árbol cuaternario (*quadtree*). Cada nodo del árbol es representado en el dominio de la transformada de Hermite discreta (THD) por un conjunto de coeficientes espectrales. La descomposición involucra la localización de la señal por ventanas binomiales de diferentes tamaños traslapadas entre sí, así como la expansión local de dichas señales en un conjunto de los polinomios de Krawtchouck (que son la versión discreta de los polinomios de Hermite). El algoritmo emplea el método de subdivisión sucesiva de bloques de la resolución más baja a la más alta.

Los bloques que no satisfacen una condición de homogeneidad son subdivididos en bloques más pequeños y analizados en escalas menores (mayor resolución). Dicha condición involucra un error perceptual definido en el dominio de la transformada. La reconstrucción de los datos de la imagen se lleva a cabo mediante la integración de los bloques ponderados por la ventana y divididos entre una función de peso asociada al quadtree. El quadtree debe ser tal que la imagen reconstruida sea visualmente sin pérdidas. La descomposición, denominada THD-Quadtree, se insertó en un esquema de compresión con un cuantizador vectorial. El diseño del cuantizador se llevó a cabo mediante el algoritmo estándar LBG. Los resultados comparativos respecto a la TCD se presentan al final del capítulo.

Finalmente, en el Capítulo 5 se presenta una discusión sobre los modelos propuestos en esta tesis desde la perspectiva de los requerimientos de un sistema de compresión de imágenes. Se mencionan las ventajas, desventajas y las líneas potenciales de investigación.

Agradecimientos

Quiero agradecer al Dr. Boris Escalante Ramírez, mi asesor, por sus varias sugerencias y apoyo constante durante esta investigación. Le agradezco especialmente la oportunidad que me brindó de participar en el proyecto *Advanced Modelling of Visual Information Processing* (AMOVIP) de donde surgieron los objetivos primordiales del presente trabajo.

Debo mencionar también, que durante mis estudios de maestría en la División de Estudios de Posgrado de la Facultad de Ingeniería, percibí la beca del Consejo Nacional de Ciencia y Tecnología (CONACyT) y que dicho apoyo fue crucial para la exitosa culminación de este trabajo.

Agradezco a mi tío José Luis y mi tía Elsa por facilitarme la vida brindándome su casa y sus atinados consejos. Por supuesto, estoy muy agradecido con mis padres por su paciencia y su amor. Sin ellos este trabajo nunca habría existido (literalmente hablando). A mi hermana Landy y a mi cuñado Chuchín por el apoyo que me brindaron en momentos difíciles de mi vida, sinceramente gracias.

Por último, y no por ello menos importantes, a Rossany por su valiosa ayuda en la revisión de la ortografía de este documento y por sus palabras de aliento que me motivaron a terminar finalmente; a Ramón y Analía por hacer mis tardes un poco menos aburridas mientras trabajaba en esta tesis; y a todos aquellos que en forma directa o indirecta hicieron posible la feliz culminación de este trabajo, gracias.

Y por todos ellos, doy gracias a Dios.

México, D. F.
Febrero 2002

José Luis Silván Cárdenas

Capítulo 1

Introducción

El objetivo del presente trabajo consiste en aplicar ciertos modelos de procesamiento y percepción de estímulos visuales al problema de la compresión de imágenes naturales. Las imágenes reconstruidas deberían permitir al lector evaluar la calidad subjetiva obtenida con las diferentes técnicas de compresión presentadas. Sin embargo, debe tomarse en cuenta que las limitaciones del proceso de impresión ocultan detalles que sólo son evidentes en otros sistemas de despliegue.

En este capítulo introductorio hemos querido presentar un resumen sobre compresión de imágenes digitales, incluyendo definiciones elementales relacionadas con la teoría de la información, así como la clasificación de las técnicas de compresión y una descripción de los procesos que lleva a cabo el sistema visual humano (SVH) para la compresión de los datos visuales. Las ideas expuestas en esta última parte establecen el punto de partida del presente trabajo, ya que los esquemas de compresión propuestos se basan en modelos de ciertas etapas de procesamiento del SVH.

1.1 Imagen digital

La demanda para el manejo de imágenes en forma digital se ha incrementado aceleradamente durante los últimos años. Gracias al mejoramiento del desempeño y a la significativa reducción en los costos de analizadores gráficos (scanners); cámaras digitales, impresoras y otros medios digitales, la imagen puede ser hoy en día fácilmente convertida al formato digital. Adicionalmente, existen distintas modalidades de captura de imágenes en medicina, tales como la tomografía computarizada (CT) o la resonancia magnética (MRI) las cuales generan directamente imágenes digitales. El uso de gráficos por computadora en las ramas de publicidad y entretenimiento está extendiéndose y su uso en la visualización científica y en aplicaciones de ingeniería continúa creciendo a pasos agigantados. La razón por el interés en imágenes digitales es clara: La representación de imágenes digitales permite que la información visual sea fácilmente manipulada en formas útiles y novedosas. Este hecho, combinado con el crecimiento exponencial de la capacidad de cómputo durante las últimas décadas, ha promovido el uso de sistemas de imágenes en diversos campos como la astronomía, percepción remota, medicina, foto-periodismo, artes gráficas, publicidad, manufactura, etc.

A pesar de las ventajas, existe un problema potencial con las imágenes digitales, la gran cantidad de bits que se requiere para almacenarlas. Afortunadamente, las imágenes digitales, en su representación canónica, contienen generalmente una gran cantidad de redundancia. La compresión de imágenes, intenta aprovechar esta redundancia para reducir el número de bits requeridos para representar una imagen. Esto puede producir un ahorro significativo en la memoria requerida para el almacenamiento o en la capacidad del canal que se requiere para la transmisión de la imagen. En la

literatura existen algunas revisiones bastante amplias sobre la materia de compresión de imágenes [27]- [24].

Ahora bien, porqué es necesario comprimir una imagen? Considere la cantidad de almacenamiento requerida para una escena LANDSAT Thematic Mapper (usada en percepción remota) que es de aproximadamente de 6000×6000 pixeles/banda espectral, 8 bits/pixel y 6 bandas espectrales no térmicas. El número de bits requeridos es aproximadamente 1.7×10^9 . Obviamente el almacenamiento de incluso unas cuantas imágenes podría significar un problema y de ahí la necesidad de compresión. Otro ejemplo claro de porqué es necesario comprimir una imagen es la transmisión de una imagen de video digital de baja resolución de 512×512 pixeles, 8 bits/pixel 3-colores sobre una línea telefónica. Usando un MODEM de 9600 baudios (bits/s) la transmisión tomaría aproximadamente 11 minutos, lo cual es inaceptable para la mayoría de las aplicaciones.

La redundancia presente en una imagen digital es altamente dependiente del sistema que se use para formar la imagen así como de los parámetros usados para representarla. En particular, la tasa de muestreo, el número de niveles de cuantización y la presencia de fuentes de ruido pueden afectar la compresión máxima posible. Aunque pueden existir variaciones de imagen a imagen, la siguiente tendencia es válida para la mayoría de los algoritmos de compresión basados en medidas estadísticas.

Mientras mayor es la tasa de muestreo, la correlación entre pixeles es mayor; lo cual permite mayores tasas de compresión. La tasa de compresión (abreviado CR por sus siglas en inglés: *Compression Rate*) se define como el cociente del el número de bits de la imagen original entre el número de bits de la imagen comprimida. Un incremento en los niveles de cuantización reduce, hasta cierto punto, la correlación de

pixel a pixel, reduciendo así la compresión máxima posible. La presencia de cualquier fuente de ruido (por ejemplo, el ruido granular de una impresión fotográfica) o el ruido introducido por el sensor, decrementa la correlación pixel a pixel y reduce la máxima compresión posible.

1.2 Clasificación de las técnicas de compresión

Los ejemplos de la sección previa ilustran claramente cómo el proceso de digitalización genera un gran número de bits por cada imagen. Sin embargo el número real de bits requeridos puede ser substancialmente menor debido a la redundancia. En general se identifican tres tipos de redundancia:

- Redundancia espacial, la cual se debe a la correlación (o dependencia) entre valores de pixeles vecinos.
- Redundancia espectral, la cual se debe a la correlación entre los planos de colores (*e.g.*, en una imagen de colores RGB) o bandas espectrales (*e.g.*, una fotografía aérea usada en percepción remota).
- Redundancia temporal, la cual se debe a la correlación entre tramas sucesivas de una secuencia de imágenes.

Las investigaciones en compresión de imágenes intentan reducir el número de bits requeridos para representar la imagen mediante la reducción de la redundancia. Adicionalmente, se intenta establecer los límites fundamentales en el desempeño de cualquier esquema de compresión para una clase de imágenes dadas. Esto se hace usando los conceptos de la teoría de información [8]. Pero más allá de estos conceptos

básicos, es necesario desarrollar una variedad de algoritmos adecuados para distintas aplicaciones.

Existen muchos enfoques para compresión de imágenes, pero todos ellos se han clasificado en dos grupos fundamentales: *sin pérdida* y *con pérdida*. En compresión sin pérdida (también denominada compresión reversible), la imagen reconstruida es numéricamente idéntica a la imagen original en una base pixel a pixel. Obviamente este tipo de compresión es ideal ya que nada de la información es comprometida. Sin embargo, sólo es posible alcanzar una modesta cantidad de compresión usando este tipo de algoritmos. En compresión con pérdida (o compresión irreversible), la imagen reconstruida contiene degradaciones relativas a la imagen original. Como resultado, se obtiene una mayor compresión comparada con la compresión sin pérdida. En general se obtiene cada vez más compresión a expensas de mayor distorsión. Es importante notar que estas degradaciones pueden o no ser visualmente aparentes. De hecho, a veces se usa el término *visualmente sin pérdidas* para caracterizar a los esquemas de compresión que introducen pérdidas invisibles bajo condiciones de visibilidad normales. Desafortunadamente la definición del término visualmente sin pérdidas es bastante subjetiva y se debe tener mucho cuidado en su interpretación. Es concebible entonces, que un algoritmo que es visualmente sin pérdidas bajo ciertas condiciones de visibilidad no lo sea para otras condiciones más estrictas.

La propuesta presentada en esta tesis cae dentro del grupo de técnicas de compresión con pérdidas y, desde luego, la intención es que las degradaciones introducidas sean invisibles y, para tasas de compresión muy bajas, sean más tolerables que las degradaciones introducidas por los codificadores estándares.

1.3 Teoría taza-distorsión

La taza límite de un esquema de codificación sin pérdida está determinada por la entropía de la fuente. En muchas situaciones prácticas se puede permitir hasta cierta cantidad de degradación irreversible en la imagen. Este tipo de degradación es generalmente controlado por el usuario ajustando un conjunto de parámetros como por ejemplo los intervalos de cuantización. La pregunta relevante aquí es: Cuál es la mínima taza de bits que se requiere para codificar una fuente si se desea que la degradación esté por debajo de cierto nivel?. Esta es la interrogante fundamental que intenta resolver la rama de la teoría de la información conocida como teoría taza-distorsión [7]. Dicha teoría establece los límites teóricos en el desempeño de los métodos de compresión con pérdida de acuerdo a un criterio de fidelidad. En general la teoría establece una función taza-distorsión $R(D)$ que cumple con las siguientes propiedades.

- Para cualquier nivel de distorsión D dado, es posible encontrar un esquema con taza arbitrariamente cercana a $R(D)$ y distorsión media arbitrariamente cercana a D .
- Es posible encontrar un código que logre la reproducción con distorsión D (o menor) a una taza por debajo de $R(D)$.

Se puede demostrar que $R(D)$ es una función convexa, continua y estrictamente decreciente

La Figura 1.1 muestra una función taza-distorsión típica para una fuente discreta de alfabeto finito. La taza mínima requerida para la compresión libre de distorsión

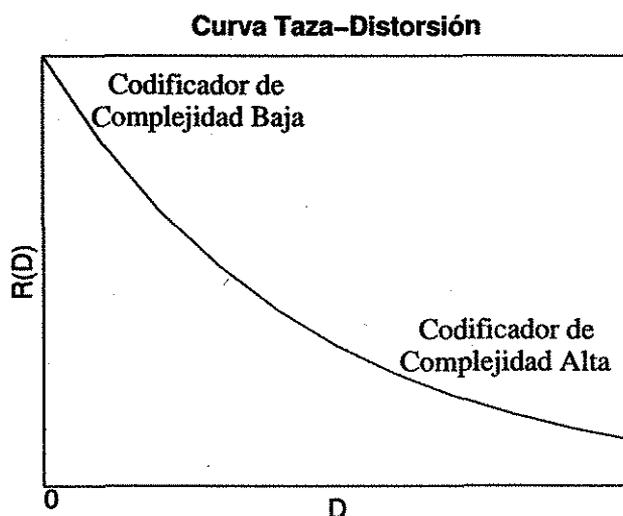


Figura 1.1: Curva taza-distorsión típica para una fuente discreta de alfabeto finito.

para una fuente dada es $R(0)$ y es menor o igual que la entropía de la fuente, dependiendo de la medida de distorsión. También se muestra en esta curva hipotética el desempeño de los codificadores de alta y baja complejidad. En general mientras más sofisticado es el esquema (es decir, mientras mejor modela la estadística de la fuente), más cercano del límite $R(D)$ es su desempeño. Para caracterizar la función $R(D)$, se requiere un modelo de fuente y un criterio de distorsión. El problema es matemáticamente tratable para modelos simples de *fuentes discretas sin memoria* (FDSM) y considerando medidas de distorsión libres de contexto. Una medida libre de contexto, también llamada distorsión de una sola letra, implica que la distorsión es función del símbolo original y su valor reproducido, y no depende de otros términos de la secuencia de símbolos de la fuente ni de sus reproducciones. Bajo estas suposiciones, se han encontrado soluciones en forma cerrada para fuentes con ciertas distribuciones y empleando los criterios de error cuadrático medio y error absoluto medio [54]. Para aquellas distribuciones donde la solución en forma cerrada no

existe, se puede encontrar una solución numérica usando el algoritmo propuesto por Blahut [8]. Desafortunadamente estos resultados son de poca utilidad en imágenes ya que las imágenes naturales son altamente correlacionadas y no son modeladas adecuadamente por una FDSM. Además, en la mayoría de las aplicaciones, la imagen comprimida es evaluada por sujetos humanos, y para cada pixel, la distorsión percibida es una función compleja de dicho pixel y de sus vecinos. Esto limita la utilidad de las medidas de distorsión libres de contexto. Más aún, incluso en los casos más simples, la realización física del esquema de codificación óptimo, el cual alcanza la cota teórica, sería extremadamente complicada.

La determinación de $R(D)$ para modelos de fuentes que describan más eficientemente a un conjunto de imágenes naturales y la determinación de medidas de distorsión que correlacionen bien con los criterios visuales permanecen como áreas activas de investigación hasta la fecha en que se escribe esta tesis.

1.4 Compresión con pérdidas

En los esquemas de compresión con pérdidas se permiten degradaciones en la imagen reconstruida como intercambio para una tasa bits más reducida en comparación con los esquemas sin pérdidas. Estas degradaciones pueden ser o no visualmente aparentes, y se puede obtener una mayor compresión permitiendo mayor degradación. La estructura general de un esquema de compresión con pérdidas se muestra en la Figura 1.2. Esta incluye tres componentes:

- Descomposición o transformación de la imagen,
- Cuantización de los datos en un conjunto de símbolos

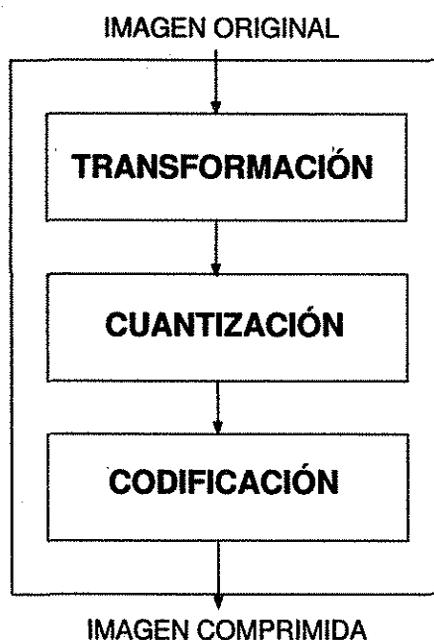


Figura 1.2: Estructura general de los esquemas de compresión con pérdidas.

- Codificación de los símbolos

La importancia relativa de cada componente varía de una técnica a otra, y no todos los componentes son necesarios para una técnica en particular. Como regla general, mientras más sofisticado es un esquema de compresión, mejor es la calidad que se puede alcanzar para una tasa de bits dada [38].

La descomposición o transformación de la imagen se lleva a cabo para reducir el rango dinámico de la señal, para eliminar información redundante o, en general, para lograr una representación que pueda ser codificada más eficientemente. La principal diferencia con los esquemas de compresión sin pérdidas es la inclusión de la etapa de cuantización. La cuantización de los datos reduce el número de posibles símbolos. El tipo y grado de cuantización tiene un gran impacto en la tasa de bits y la calidad de la imagen. También, es deseable realizar la cuantización de tal forma que la

secuencia de salida pueda ser subsecuentemente codificada de manera eficiente. El proceso de cuantización puede incluir técnicas tales como códigos de Huffman [23] o códigos aritméticos [40]. En general, cualquiera de los componentes de un esquema de compresión con pérdidas se puede realizar en modo adaptable o no adaptable. Un esquema de compresión es adaptable si la estructura de un componente o sus parámetros cambian localmente dentro de la imagen tomando así ventaja de las variaciones en la estadística local. La adaptabilidad ofrece el potencial para mejorar el desempeño en la compresión a costa de un incremento en la complejidad. La adaptabilidad se puede realizar en forma causal o no causal.

En un sistema con adaptabilidad causal, los parámetros del codificador se basan solamente en los valores de píxeles previamente reconstruidos y cualquier proceso de toma de decisión en el codificador debe ser duplicado en el decodificador. Este sistema tiene la ventaja de que no requiere información sobre la toma de decisiones del codificador, sin embargo tiene un par de desventajas asociadas. Primero, el codificador falla generalmente en la adaptación cuando ocurren cambios abruptos en la estadística de la señal de tal forma que no puede ser inferida a partir de valores previamente reconstruidos. Segundo, la adaptabilidad causal incrementa la complejidad del codificador y del decodificador en la misma proporción, ya que el decodificador debe duplicar el proceso de toma de decisiones del codificador.

En sistemas con adaptabilidad no causal, los parámetros del codificador se basan tanto en valores previos como en valores futuros de la señal. Dado que estos últimos valores no están disponibles en el decodificador, el codificador debe enviar bits adicionales para informar al decodificador sobre cualquier adaptación. Aunque esto produce una mayor tasa de bits, también incrementa el desempeño global del

sistema. Además, el incremento en la complejidad del decodificador debido a la adaptación es mínima ya que el decodificador no requiere repetir el proceso de selección del codificador.

1.5 Compresión por el sistema visual

El proceso de visión inicia con la absorción de la luz mediante un pigmento contenido en los fotorreceptores conocidos como conos y bastones, llamados así por su forma. Los primeros contienen uno de tres pigmentos visuales con máximos de sensibilidad espectral en $450nm$ (conos azules), $525nm$ (conos verdes) y $555nm$ (conos rojos). El pigmento visual asociado a los bastones se llama *rhodopsin* y presenta una curva de sensibilidad espectral más extendida que los conos con un máximo en $500nm$. Además de los fotorreceptores, la retina contiene otros cuatro grupos de células involucradas en el proceso visual. Estas son las células bipolar, ganglionar, horizontal y amácrina. Existen subtipos de cada una de éstas, de hecho, la retina contiene al menos 50 tipos de células neuronales morfológica y funcionalmente distintas. El impulso eléctrico generado por los fotorreceptores pasa a través de las células horizontal, bipolar y amácrina y llega finalmente a las células ganglionares las cuales representan el camino común de la salida neuronal del ojo. Los axones de las células ganglionares forman el nervio óptico. El hecho de que existen cerca de 120 millones de fotorreceptores y solamente cerca de un millón de células ganglionares implica una considerable convergencia de la salida de los fotorreceptores. Esta convergencia toma parte en el preprocesamiento del estímulo visual que se lleva a cabo en la retina. Los trabajos experimentales donde se ha registrado la salida de una sola célula ganglionar demuestran que éstas tienen un campo receptivo específico en el cual un estímulo evoca un

tren de impulsos neuronales. En la mayoría de los casos, los campos receptivos son de naturaleza concéntrica con la parte central excitatoria y el borde inhibitorio. El inicio de iluminación de la parte central del campo receptivo causa un incremento en la tasa de pulsos neuronales sobre el nivel de reposo. En contraste, cuando se ilumina el anillo concéntrico que rodea el campo receptivo, existe solo respuesta cuando la iluminación se apaga.

De esta descripción es evidente que el sistema visual humano lleva a cabo un significativo grado de procesamiento de imagen para alcanzar su notable desempeño. En el corazón del problema descansa la necesidad de minimizar la cantidad de información que requiere ser procesada por el sistema visual. Por ejemplo, como el ojo responde a frecuencias espaciales hasta cerca de 60 ciclos por grado a una tasa máxima de 60 ciclos por segundo, entonces se tienen un número máximo de muestras por segundo que deben ser enviadas a la corteza visual del orden de 5.5×10^{10} . Este valor excede el ancho de banda disponible en el nervio óptico, de ahí que el sistema visual emplee un número de estrategias para comprimir los datos. En lugar de simplemente enviar toda la información a la corteza, es suficiente con señalar solamente los cambios en la información. El tipo de receptores centro-encendido y borde-apagado actúan efectivamente como un filtro *pasobanda* que remueve las bajas frecuencias espaciales de la información de la escena y transmite únicamente la información sobre los cambios en los datos espaciales de la imagen de la retina. Un desafortunado resultado de esto podría ser que si la escena no cambia, entonces la imagen se desvanecería. Afortunadamente esto no sucede en la práctica porque, incluso cuando se observa fijamente un objeto en reposo, el ojo no se mantiene realmente estacionario. Durante la fijación hay pequeños movimientos continuos (desplazamiento y parpadeo) y un temblor

sobrepuesto de hasta 150 ciclos por segundo con una amplitud de aproximadamente medio diámetro de un cono ($2\mu m$). Esto asegura que la imagen de la retina siempre está cambiando incluso cuando la escena permanezca estática. Si la imagen de la retina se estabiliza artificialmente, entonces la imagen se desvanece.

Además de la compresión alcanzada por las células ganglionares, también están presentes células que responden a estructuras específicas en la imagen, tales como líneas con orientaciones particulares, esto reduce aún más la cantidad de información que debe ser transmitida a la corteza.

Por lo tanto, se puede ver que el ojo no es simplemente una cámara, sino un dispositivo complejo de captura y procesamiento, y que el objetivo del procesamiento no es solamente hacer un uso óptimo de los canales de dato disponibles en el sistema visual sino también poner los datos de imagen en un formato adecuado para su interpretación por el cerebro. Estos mecanismos presentes en el SVH son los que han inspirado un número de esquemas de compresión con pérdidas durante las últimas tres décadas ([37], [56], [59], [61]).

1.6 Imágenes de prueba

En la Figura 1.3 se muestran las imágenes empleadas para probar los algoritmos propuestos en esta tesis. De izquierda a derecha y de arriba hacia abajo son: LENA, HOUSE, PEPPERS, BARBARA, MIT y CAMERAMAN; todas en formato BMP de 256 por 256 píxeles en niveles de grises codificados en 8 bits. Estas imágenes son frecuentemente usadas para comparar el desempeño de los distintos modelos de procesamiento, ya que exhiben características que permiten estudiar un amplio espectro de estructuras que son visualmente importantes.

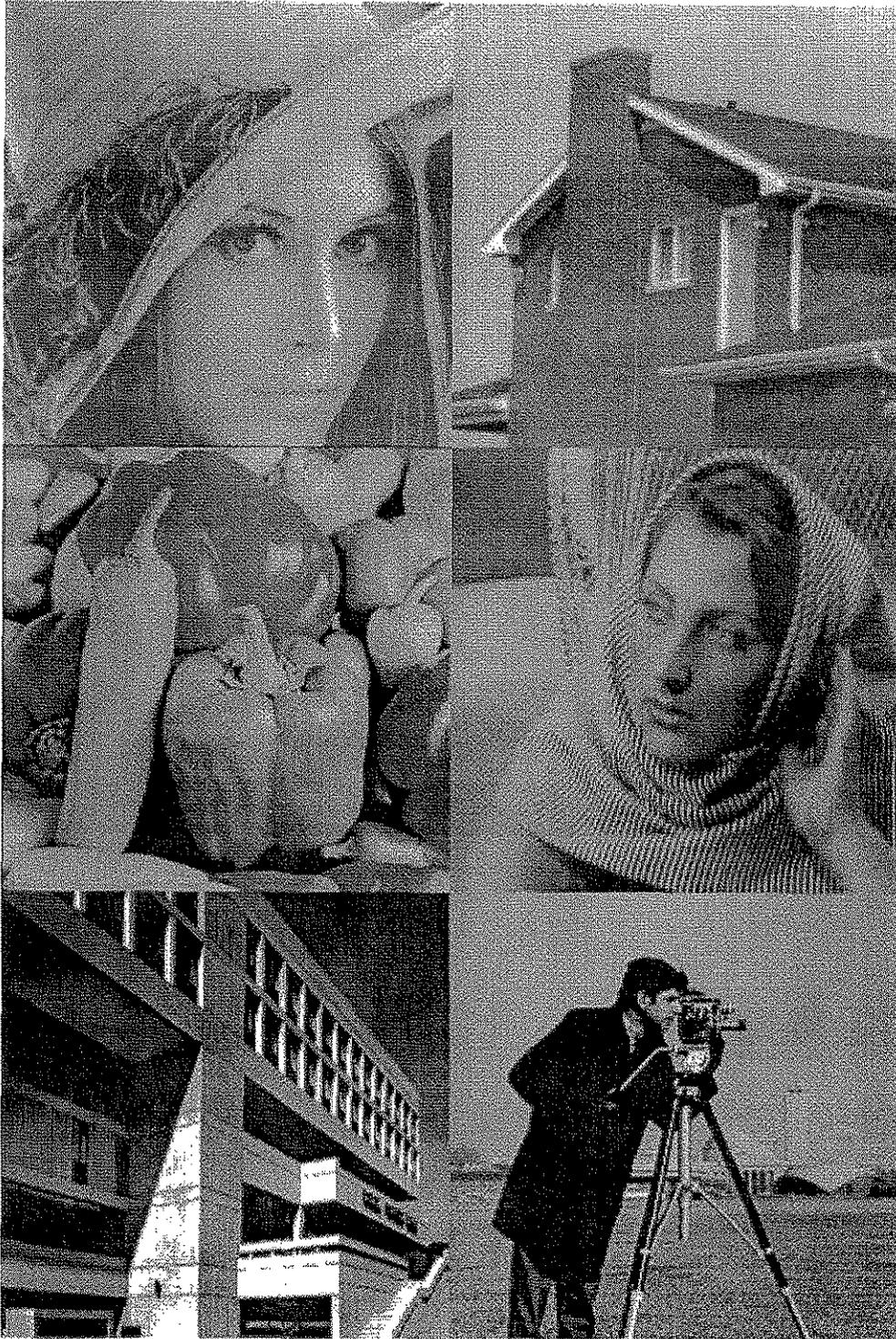


Figura 1.3: Imágenes de prueba.

TESIS CON
FALLA DE ORIGEN

Capítulo 2

Transformada de Hermite multiescala

En este capítulo se deriva un modelo multicanal para la representación de imágenes naturales basado en la teoría espacio-escala. El modelo tiene una inspiración biológica de tal forma que sus parámetros pueden definirse de acuerdo a ciertas propiedades del sistema visual. Primero se estudia el caso unidimensional (1-D) y posteriormente se extienden los resultados a dos dimensiones. Los operadores que se obtienen tanto en la fase de análisis como en la fase de síntesis son derivadas de Gaussiana, de tal forma que, para el caso bidimensional (2-D), se tiene la posibilidad de representar la imagen en función de la respuesta de filtros direccionales. Al sustituir la normalización de coordenadas por un muestreo se obtiene un análisis de la señal similar al que lleva a cabo la transformada de Hermite jerárquica (THJ). También se define la versión discreta de las transformadas multiescala y multidireccional a partir de la representación espacio-escala de señales discretas en la cual se mantienen muchas de las propiedades del caso continuo. En esta parte también se discuten las consideraciones prácticas para su implantación en un sistema digital .

2.1 Introducción

La calidad de los resultados en el procesamiento de imágenes es a menudo evaluada por humanos. En tal caso los errores son significativos solo si se perciben. Por lo tanto, con un profundo entendimiento del Sistema Visual Humano (SVH) se pueden explotar sus propiedades para obtener imágenes que se *vean* mejor.

Uno de los modelos más conocido del SVH a nivel de la corteza visual está basado en las funciones de Gabor [5], [37]. El diseño de filtros para la simulación del procesamiento visual ha sido también de interés en esta área, tal es caso de la transformada *Cortex* [56] cuyas funciones de análisis se diseñaron para aproximar el perfil de las funciones de Gabor, con la ventaja de poder ajustar los anchos de banda radial y angular. Un modelo alternativo propuesto por Young en 1987 se basa en la Gaussiana y sus derivadas [63]. Estas funciones, igual que las funciones de Gabor, son espacialmente locales y consisten de regiones de excitación e inhibición alternadamente en una envolvente que se desvanece gradualmente. Young mostró que las derivadas de Gaussiana (DG) modelan con mayor precisión los datos medidos en los campos receptivos que las funciones de Gabor, con la ventaja adicional de la ortogonalidad de las funciones situadas en un mismo punto de análisis. Más aún, se ha demostrado que las DG pueden modelar también los campos receptivos a nivel de las neuronas sensibles a la disparidad binocular [17].

Algunos modelos matemáticos basados en estos operadores a una sola escala se describen en [9],[18] y [35]. Este último se extiende al caso multiescala en [34] y en ambos casos, las DG se interpretan como el producto de los polinomios de Hermite por una ventana Gaussiana, donde la señal multiplicada por la ventana se expande sobre la base de los polinomios de Hermite. La transformada de imagen propuesta en

este capítulo presenta los mismos operadores a múltiples escalas pero con un esquema de síntesis que integra las respuestas de todos los canales a distintas escalas¹.

2.2 Caso 1-D: Transformada de Hermite Multiescala

En esta sección se construye una transformada multiescala basada en derivadas de Gaussiana la cual es referida como transformada de Hermite multiescala (THM). La derivación de los resultados presentados en esta sección se puede consultar en el Apéndice A.

2.2.1 Descomposición multiescala

Para comprender mejor el concepto de escala debemos ver la imagen como la representación de una escena física. Dicha representación estará afectada por cantidades físicas que pueden medirse o estimarse. Así por ejemplo, el efecto de la distancia de observación se puede traducir en un escalamiento de la distribución de intensidades de la representación, esto significa que al alejarnos del objeto observado se incrementa la escala de la representación y por lo tanto se pierden detalles, mientras que al acercarnos disminuimos la escala y percibimos más detalles.

La representación espacio-escala [60] emplea las versiones escaladas de la Gaussiana de integral unitaria $G(x) = \exp(-x^2)/\sqrt{\pi}$ para eliminar detalles de la señal de luminancia $L(x)$ mediante la operación de convolución, es decir que $L(x, s) = L(x) \star$

¹Una versión preliminar de este trabajo fue presentada en el *5th World Multiconference on Systemics, Cybernetics and Informatics* [50] y otra en el *SPIE's 46th Annual Meeting* [15]

$G(x, s)$ es la representación de la señal a la escala s , donde $G(x, s) = G(x/\sqrt{4s})/\sqrt{4s}$ es la Gaussiana a la escala $s > 0$. La variable normalizada $\xi = x/\sqrt{4s}$ es referida como coordenada natural.

Para poder arribar a un resultado que ponga en evidencia la relación existente entre las estructuras de la imagen percibidas a diferentes escalas, necesitamos tener una idea sobre la forma en la que el sistema de visión humana podría llevar a cabo dicho proceso. Los resultados psicofísicos evidencian la existencia de por lo menos cuatro canales en cada punto del campo visual a nivel de la retina (específicamente en las células ganglionares [59]). Dichos canales exhiben una respuesta aproximada a una diferencia de Gaussianas (o DoG del inglés *Difference of Gaussians*) cuya respuesta al impulso se escribe como $DoG(x, s_1, s_2) = G(x, s_1) - G(x, s_2)$ para $s_2 > s_1$.

La respuesta del filtro DoG contiene la información que debe ser agregada a la representación a la escala s_2 para obtener la representación a la escala menor s_1 . Así, la señal puede ser expresada como una suma de canales DoG:

$$L(x) = \sum_{k=-\infty}^{\infty} L(x) \star DoG(x, s_{k-1}, s_k) \quad (2.1)$$

con $s_k > s_{k-1}$.

Aquí, requeriremos que el parámetro de espaciamiento $\tau = (s_k - s_{k-1})/s_k$ sea constante para todo entero k , lo cual está en acuerdo con datos psicofísicos que soportan la existencia en el SVH de un número de canales cuyas frecuencias centrales mantienen una relación de aproximadamente una octava [59], es decir $\tau \approx 0.75$.

2.2.2 Análisis estructural

En la mayoría de las aplicaciones es deseable saber qué parte de la información contenida en cada canal, hablando en términos perceptivos, debe agregarse con mayor

intensidad (por ejemplo, bordes, líneas, esquinas, etc.) y qué parte de información debe ser atenuada porque resulta molesta a la vista (por ejemplo, texturas finas o ruido). En tal caso, resulta apropiado llevar a cabo un análisis derivativo para discernir entre las diferentes estructuras de interés [20]. Así, la estructura local de una señal se infiere a partir de las derivadas de las representaciones escaladas. Por otra parte, las derivadas de $L(x, s)$ se pueden obtener convolucionando la función $L(x)$ con los operadores derivadas de Gaussiana: $G_n(x, s) = \partial_x^n G(x, s)$.

Por tanto, el siguiente paso es expresar el filtro DoG en función de las DG, esto se consigue tomando la expansión en serie de Taylor de la Gaussiana a la escala s_{k-1} alrededor de la escala s_k , es decir,

$$DoG(x, s_{k-1}, s_k) = \sum_{n=1}^{\infty} \frac{(-\tau s_k)^n}{n!} G_{2n}(x, s_k) \quad (2.2)$$

En la práctica se debe limitar el número de términos de esta serie a un número pequeño. Estudios neurofisiológicos indican que el sistema visual trabaja con derivadas hasta de cuarto orden [61], [62], [63]. En nuestro estudio matemático, sin embargo, consideraremos todos los términos de la serie ya que las aproximaciones pueden ser estudiadas como casos particulares.

2.2.3 Definición de la THM

La expresión (2.2) muestra explícitamente que las derivadas de orden impar no contribuyen en la descripción de la imagen como se apunta en [6], donde se reportan experimentos de realce basado en las DG de orden par. Por otro lado, es deseable que, al menos en la etapa de análisis del modelo, aparezcan todas las derivadas con el fin de tener una descripción más rica de la estructura de la señal. Por lo tanto, es necesario distinguir la fase de análisis de la fase de síntesis en nuestro modelo. Esto

se logra factorizando las DG que aparecen en la serie (2.2) usando la propiedad de cerradura bajo convolución. Más aún, si interpretamos las operaciones de convolución como producto interno, podemos escribir las expresiones para la reconstrucción de la señal como:

$$L(x) = \sum_{k=-\infty}^{\infty} \sum_{n=1}^{\infty} \tau^n D_n^{(k)}(x) \quad (2.3)$$

donde las funciones

$$D_n^{(k)}(x) = \langle L_n^{(k)}(\xi), G_n^{(k)}(x, \xi) \rangle_{\xi} \quad (2.4)$$

para $n = 1, 2, \dots$ aportan los *detalles* de la señal a todas las escalas y están determinadas por las derivadas escaladas,

$$L_n^{(k)}(\xi) = \langle L(x), G_n^{(k)}(x, \xi) \rangle_x \quad (2.5)$$

las cuales se determinan proyectando la señal de entrada sobre las *funciones de análisis*

$$G_n^{(k)}(x, \xi) = \frac{1}{\sqrt{2s_k}} G_n^* \left(\frac{x}{\sqrt{2s_k}} - \xi \right) \quad (2.6)$$

para $n = 1, 2, \dots$ y $s_k = (1 - \tau)^{-k} s_0$ para todo k entero.

Al análisis de la señal expresado en la Ecuación (2.5) es lo que llamamos THM directa, mientras que la reconstrucción de la señal mediante la Ecuación (2.3) y las *funciones de detalle* (2.4) la llamamos THM inversa.

2.2.4 Acotamiento del rango de escalas

Consideremos un número finito K de canales DoG y supongamos que la escala menor es suficientemente pequeña. La descomposición multicanal se reduce a

$$L(x, s_0) = L(x, s_K) + \sum_{k=1}^K \sum_{n=1}^{\infty} \tau^n D_n^{(k)}(x) \quad (2.7)$$

donde s_0 y s_K determinan las resoluciones límite del análisis. El residuo pasobajas $L(x, s_K)$ no puede ser despreciado para valores de K en un rango práctico, pero su variación es, en general, menor que la de la señal original. En cambio, el residuo pasobajas $L(x) - L(x, s_0)$ es despreciado en virtud de que s_0 es suficientemente pequeño.

En los casos en los cuales el residuo pasobajas contiene información relevante, se puede agregar un canal definido entre las escalas 0 y s_0 ($\tau = 1$), ya que la señal es la representación a la escala cero. Por lo tanto, si generalizamos las definiciones de las ecuaciones (2.4), (2.5) y (2.6) para $n = 0$, podemos escribir la THM como

$$L(x) = D_0^{(K)}(x) + \sum_{n=1}^{\infty} D_n^{(0)}(x) + \sum_{k=1}^K \sum_{n=1}^{\infty} \tau^n D_n^{(k)}(x) \quad (2.8)$$

donde los dos primeros términos del miembro derecho son justamente el residuo pasobajas y el residuo pasobajas respectivamente. El caso cuando $K = 0$ se conoce como transformada de Hermite de una sola escala (o simplemente TH). En tal caso no aparece el tercer término de la Ecuación 2.8.

2.2.5 Esquema piramidal continuo

Al esquema de procesamiento en el cual la imagen se descompone en un número de subimágenes pasobajas o pasobanda, las cuales son submuestreadas en proporción a su resolución, se conoce como *esquema piramidal*. Cada nivel de la pirámide guarda una relación lineal con los niveles adyacentes y dicha relación es invariante con la escala. Esta definición está enclavada en el contexto de señales discretas, pero puede ser extrapolada al caso continuo si reemplazamos el remuestreo por un reescalamiento de la coordenada espacial. En tal caso, todos los niveles de la pirámide quedan definidos en un dominio continuo.

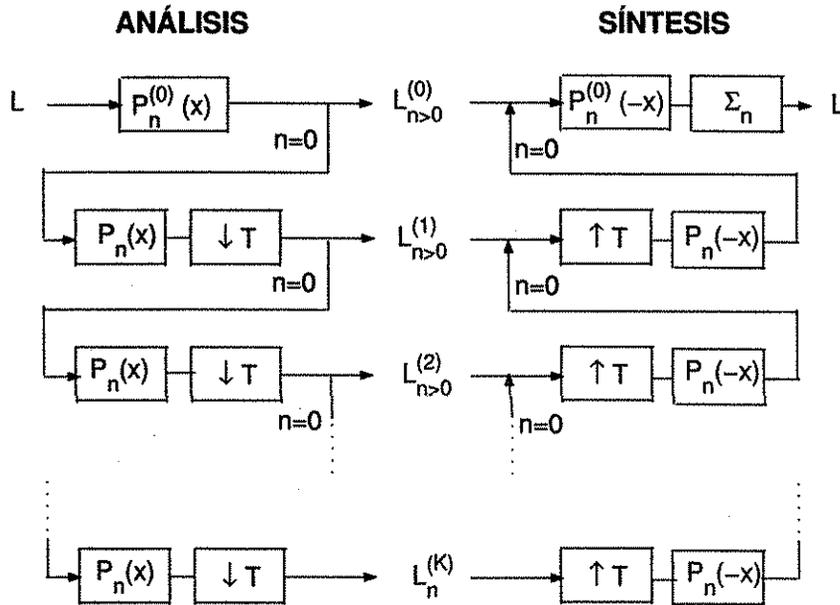


Figura 2.1: Esquema piramidal de análisis y síntesis de la THM.

La relación piramidal de la THM se deriva del hecho de que las funciones de análisis de un orden dado forman una pirámide Gaussiana dado por la relación lineal

$$\sqrt{\tau^n} G_n^{(k)}(x, \xi) = \int_{-\infty}^{\infty} G_0^{(k-1)}(x, \eta) P_n(T\xi - \eta) d\eta \quad (2.9)$$

donde los coeficientes de la pirámide $P_n(x) = G_n^*(-x/\sqrt{4\sigma})/\sqrt{4\sigma}$ son derivadas de Gaussiana a la escala $\sigma = \tau T^2/4$, la cual se mantiene constante para todos los niveles. El factor $T = 1/\sqrt{1-\tau}$ determina la proporción entre las coordenadas de un nivel dado y del nivel anterior. Para $\tau = 0.75$ se tiene $T = 2$.

Aplicando sucesivamente la propiedad (2.9) en la expresión (2.4) se obtiene la relación piramidal *ascendente*:

$$\tau^n D_n^{(k)} = \{[(\sqrt{\tau^n} L_n^{(k)\uparrow T} \star P_n)^{\uparrow T} \star \underbrace{P_0]^{\uparrow T} \dots \star P_0}_{k-1 \text{ factores}} \star P_0^{(0)} \quad (2.10)$$

para $k = 1, \dots, K - 1$ y $P_n^{(0)}(x) = G_n^*(-x/\sqrt{4s_0})/\sqrt{4s_0}$.

Por otra parte, si se aplica la misma propiedad a la fase de análisis de la Ecuación (2.5), se obtiene la relación piramidal *descendente*:

$$\sqrt{\tau^n} L_n^{(k)} = \{[(L \star P_0^{(0)}) \star \underbrace{P_0^{\downarrow T} \dots \star P_0^{\downarrow T}}_{k-1 \text{ factores}} \star (-1)^n P_n]\}^{\downarrow T} \quad (2.11)$$

Los símbolos $\uparrow T$ y $\downarrow T$ significan que la coordenada se multiplica por T y T^{-1} respectivamente. De las ecuaciones (2.10) y (2.11) se obtiene el esquema piramidal mostrado en la Figura 2.1 donde, por simplicidad, se escribió $L_n^{(k)}$ en lugar de $\sqrt{\tau^n} L_n^{(k)}$. Como se puede observar en este diagrama, el esquema piramidal tiene la característica de que la reconstrucción se realiza por etapas agregando sucesivamente las contribuciones de cada nivel de la pirámide comenzando con el residuo pasobajas.

2.2.6 Predicción entre niveles

Una desventaja de la THM es que incrementa las dimensiones del espacio de representación de la señal (de posición a posición-escala-orden de derivación) que en la práctica se traduce en un incremento en la redundancia. Parte de esa redundancia se debe a la dependencia lineal entre las funciones de análisis de distintas escalas. En particular, se puede demostrar que las funciones del nivel $k - 1$ se relacionan con las del nivel k mediante

$$\sqrt{\tau^n} G_n^{(k-1)}(x, \xi) = \sum_{m=n}^{\infty} c_{n,m} \int_{-\infty}^{\infty} \sqrt{\tau^m} G_m^{(k)}(x, \eta) P_{m-n}(\eta - T^{-1}\xi) d\eta \quad (2.12)$$

con las constantes $c_{n,m} = \sqrt{C_m^n} T^{-m} [(1 + T^2)/2]^{m-n}$.

Esta dependencia lineal entre las funciones de análisis permite insertar la THM en un esquema de predicción en el cual cada nivel se estima a partir del siguiente nivel,

de tal forma que en un sistema de compresión sólo se requiere codificar los residuos de predicción. De esta forma se reduce significativamente la entropía de los datos. Este hecho es análogo a sobremuestrear una señal y codificar las diferencias entre muestras consecutivas en lugar de codificar las muestras como tal, ya que el incremento en la tasa de muestreo reduce la entropía global de los datos.

2.3 Caso 2-D: Transformada de Hermite multiescala-multidireccional

En esta sección generalizamos las THM para señales bidimensionales. Adicionalmente, se inserta un esquema multidireccional basado en derivadas direccionales. En este caso, la transformada que se obtiene es referida como THM multidireccional (THMM). Se demuestra cómo la THMM puede ser obtenida mediante la THM 2-D. Para los detalles consulte el Apéndice B.

2.3.1 Descomposición multiescala de señales 2-D

Para el caso de señales multidimensionales la representación multiescala se lleva a cabo filtrando la señal en cascada a lo largo de cada coordenada ya que la Gaussiana multidimensional se expresa como un producto de Gaussianas 1-D de cada coordenada. En particular, para el caso bidimensional se definen los operadores de derivadas escaladas como $G_{m,n-m}(x, y, s) = G_m(x, s)G_{n-m}(y, s)$ para $m = 0, 1, \dots, n$ y $n = 0, 1, \dots$, donde n es el orden de derivación total. Por lo tanto, la descomposición de señales bidimensionales en canales DoG es análoga al caso de señales 1-D.

2.3.2 Análisis estructural de señales 2-D

Otro factor físico que influye en la representación de los patrones visuales es la inclinación relativa del sistema de visión con respecto a un eje de referencia en la escena. Dicha inclinación hará que las estructuras de la imagen se perciban con una orientación determinada, la cual sólo puede ser *vista* por operadores sensibles a la orientación.

La ventaja de usar derivadas de Gaussiana se hace evidente cuando se requiere representar estructuras orientadas ya que la función DoG es isotrópica. Se puede demostrar que el conjunto formado por las $n + 1$ derivadas bidimensionales de orden n permite representar patrones de hasta n orientaciones simultáneas [44].

Para propósitos de análisis de orientación es conveniente trabajar con las versiones rotadas de estos operadores como se sugiere en [36]. Sean los operadores $G_{m,n-m}(x, y, s, \theta) = G_{m,n-m}(x \cos \theta + y \sin \theta, -x \sin \theta + y \cos \theta, s)$, donde θ es el ángulo de rotación, se puede demostrar que estas funciones están relacionadas linealmente con las funciones referidas al sistema original (ver Apéndice D).

Las derivadas direccionales de Gaussiana (DDG), denotadas por $G_{n,0}(x, y, s, \theta)$, se han empleado eficazmente en la detección y discriminación de orientaciones [44]. Estos operadores son ideales para la descripción de estructuras orientadas porque satisfacen la llamada propiedad de direccionamiento (*steering*), a partir de la cual se demuestra que sólo se requiere conocer la respuesta de la DDG de orden n en $n + 1$ direcciones para poder conocer la respuesta en cualquier dirección [16].

La expansión en serie de Taylor de la DoG bidimensional puede escribirse ya sea en función de las DG bidimensionales o bien en función de las DDG. La primera expansión produce la THM en dos dimensiones, que no es otra cosa que la transformación

definida en la sección anterior pero aplicada sobre ambas coordenadas espaciales (ver Apéndice B).

El segundo caso es más interesante y vale la pena describirlo con mayor detalle. En este caso la expansión de la DoG se escribe como

$$DoG(x, y, s_{k-1}, s_k) = \sum_{n=1}^{\infty} \sum_{j=0}^n \frac{c_n (-\tau s_k)^n}{n!} G_{2n,0}(x, y, s_k, \theta_j) \quad (2.13)$$

donde $\theta_j = \theta_0 + j\pi/(n+1)$ para $j = 0, \dots, n$ y c_n son las constantes definidas en (B.4). Note que el orden de derivación n determina la resolución angular y que el ángulo θ_0 es un parámetro libre que puede ser elegido según convenga.

2.3.3 Definición de la THMM

Partiendo de la expansión (2.13) y procediendo como en el caso 1-D, se construye una transformada que resulta más apropiada para la descripción de imágenes, ya que las funciones de análisis que se obtienen incluyen las tres transformaciones geométricas fundamentales: traslación, rotación y escalamiento.

El proceso de análisis

$$L_n^{(k,j)}(\xi, \eta) = \langle L(x, y) G_n^{(k,j)}(x, y, \xi, \eta) \rangle_{(x,y)} \quad (2.14)$$

mediante las *funciones de análisis*

$$G_n^{(k,j)}(x, y, \xi, \eta) = \frac{1}{2s_k} G_{n,0}^* \left(\frac{x \cos \theta_j + y \sin \theta_j}{\sqrt{2s_k}} - \xi, \frac{-x \sin \theta_j + y \cos \theta_j}{\sqrt{2s_k}} - \eta \right) \quad (2.15)$$

con $\theta_j = \theta_0 + j/(n+1)$ para $j = 0, \dots, n$; $n = 1, 2, \dots$ y $s_k = (1 - \tau)^{-k} s_0$ para todo k entero, es referido como transformada de Hermite multiescala-multidireccional (THMM) porque emplea las DDG a múltiples escalas y orientaciones. Y por tanto, al

proceso de reconstrucción de la señal a partir de esta representación se le denomina THMM inversa.

En este caso, la reconstrucción de la señal

$$L(x, y) = \sum_{k=-\infty}^{\infty} \sum_{n=1}^{\infty} \sum_{j=0}^n c_n \tau^n D_n^{(k,j)}(x, y) \quad (2.16)$$

se lleva a cabo sumando las contribuciones de las *funciones de detalle direccional*

$$D_n^{(k,j)}(x, y) = \langle L_n^{(k,j)}(\xi, \eta) G_n^{(k,j)}(x, y, \xi, \eta) \rangle_{(\xi, \eta)} \quad (2.17)$$

sobre el conjunto de escalas y orientaciones previamente definido.

2.3.4 Filtrado direccional

La representación de imágenes mediante la THMM tiene la desventaja de que las funciones de análisis de la Ecuación (2.15) son no-separables y, por lo tanto, su implementación puede resultar costosa en cuanto a recursos computacionales. Sin embargo, ese costo puede ser reducido si se calcula la transformada del caso separable y posteriormente se transforman los coeficientes del mismo orden total a través de las *funciones de ángulo*

$$\alpha_{m,n-m}(\theta) = \sqrt{C_n^m} \cos^m \theta \sin^{n-m} \theta \quad (2.18)$$

para $m = 0, 1, \dots, n$, ya que las funciones de análisis en el punto $(\xi, \eta) = (0, 0)$ están relacionadas linealmente como

$$G_n^{(k,j)}(x, y) = \sum_{m=0}^n \alpha_{m,n-m}(\theta_j) G_m^{(k)}(x) G_{n-m}^{(k)}(y) \quad (2.19)$$

para $n = 1, 2, \dots$

La proyección de la señal sobre las funciones de análisis equivale a una convolución con los filtros direccionales que se obtienen reflejando las funciones de la Ecuación (2.19) respecto a x y y , más la normalización de coordenadas respecto a escala y orientación.

El filtrado direccional es ampliamente citado en la literatura como mecanismo para hacer explícita la información relevante de una imagen [16], [25], [36]. Por ejemplo, considérese una estructura 1-D orientada con un ángulo ϕ medido respecto al eje x . La THMM que se obtiene para esta estructura tiene la forma $L_n^{(k,j)} = \cos^n(\theta_j - \phi) H_n^{(k)}$ donde $H_n^{(k)}$ es la THM del perfil 1-D de la estructura. Estos coeficientes 1-D contienen información sobre contraste y nivel de luminancia, mientras que los coeficientes $L_n^{(k,j)}$ contienen además información sobre la localización y orientación de la estructura a distintas escalas [34].

En trabajos previos sobre codificación de imágenes [46] y estimación del flujo óptico en secuencias de imágenes [14] se ha demostrado la relación que existe entre el filtrado direccional y el análisis de proyecciones locales.

2.3.5 Aplicaciones

La literatura sobre aplicaciones de las DG a los problemas de visión automática y procesamiento de imágenes predice la utilidad de la THM en problemas de realce [6], detección de orientación [44], reducción de ruido [13], estimación del desenfoque para el cálculo de profundidad [64], estimación de movimiento [32],[47], codificación [18], etc. Sin embargo, note que esta representación no parece ser muy apropiada para la compresión de imágenes ya que, desde la perspectiva de compresión, se requiere que la imagen sea representada por el mínimo número de parámetros posible a partir de

los cuales se pueda reconstruir una imagen *visualmente idéntica* a la original.

Evidentemente nuestra representación, lejos de eliminar redundancia, hace crecer la dimensionalidad de los datos de la imagen al insertar los parámetros de escala y orientación. A esto se suma el hecho de que la condición de ortogonalidad de la Ecuación (A.5) en el Apéndice A, sólo se satisface respecto al orden de derivación entre funciones con la misma posición, escala y orientación. Toda esa redundancia, sin embargo, no debe ser considerada como una limitación, sino como un mecanismo para incrementar la tolerancia a la degradación por truncamiento o cuantización severa de los datos. En una descomposición como la que lleva a cabo la transformada *Wavelet*, que satisface la condición de ortogonalidad entre funciones de análisis con distinta posición y escala, se suelen introducir errores significativos en términos perceptivos bajo condiciones de truncamiento o cuantización severa.

Para los propósitos de codificación sería deseable conocer cuál es el mínimo número de componentes de la descomposición que se pueden tomar sin afectar significativamente la calidad subjetiva de la imagen. Desafortunadamente, no se conoce una expresión matemática para medir la calidad subjetiva de una imagen, pero sí podemos darnos una idea cualitativa del efecto visual resultante en la imagen reconstruida para una aproximación en particular.

Aquí consideramos la aproximación de primer orden del filtro DoG y un número finito de canales (Ec. 2.7). El modelo de señal que se obtiene es

$$\hat{L}(x) = L(x, Ts_K) + T \sum_{k=0}^{K-1} \tau D_1^{(k)}(x) \quad (2.20)$$

Si s_0 es *suficientemente* pequeño, entonces la contribución de error más significativa es introducida por la aproximación del filtro DoG. El cálculo de este error es de poca utilidad para el caso de imágenes. En lugar de realizar el cálculo numérico

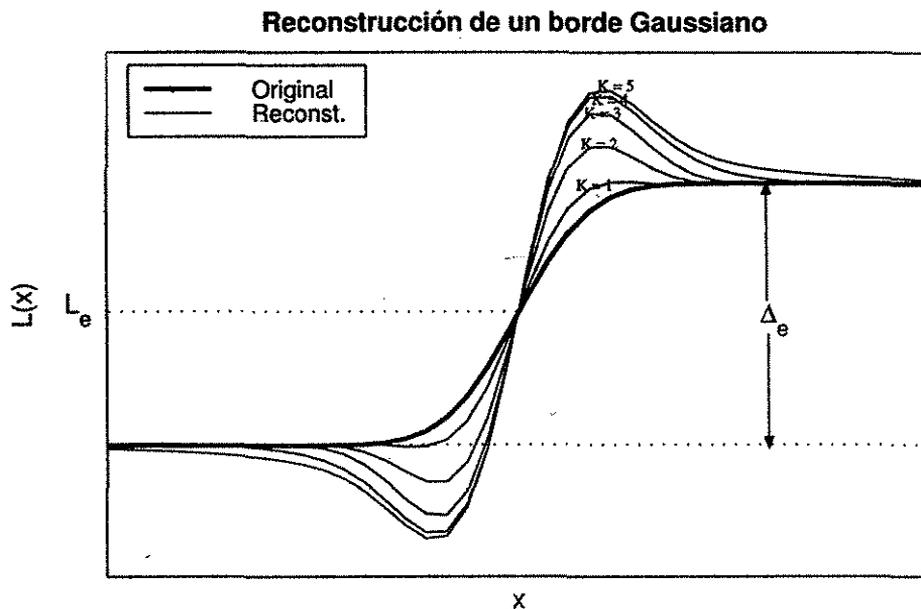


Figura 2.2: Reconstrucción de un borde gaussiano.

podemos darnos una idea de los efectos visuales que puede producir analizando las distorsiones en las estructuras primitivas de la imagen. La primitiva 1-D más importante es, quizá, el *borde*, que no es otra cosa que un cambio substancial en el nivel de luminancia.

El modelo gaussiano de un borde con luminancia media L_e , contraste Δ_e y escala s_e se describe matemáticamente por la función $L(x) = L_e + 0.5\Delta_e \text{erf}(x/\sqrt{4s_e})$. La Figura 2.2 muestra la reconstrucción del un borde gaussiano usando la aproximación de primer orden con $\tau = 0.75$ y $K = 1, \dots, 5$. Se observa que la distorsión que se obtiene es un sobrepaso cerca de la región de transición que aumenta con K pero parece converger a un límite. Este efecto de sobrepaso, lejos de degradar la estructura, resulta visualmente benéfico porque mejora el contraste del borde. De hecho, la evidencia psicofísica demuestra que el SVH lleva a cabo un tipo de realce similar [39].

Esto se puede ver claramente en bordes contrastados con alta pendiente de transición (s_e pequeño), en los cuales aparecen dos bandas cerca del punto de transición, una más oscura que el nivel bajo del borde y una más clara que el nivel alto [42]. Este efecto es conocido como *bandas de Mach* en honor a Ernst Mach (1896-1916) quien las describió por primera vez en 1865.

Con base en lo anterior, podemos predecir que la imagen reconstruida aparecerá mejor contrastada que la imagen original, lo cual es útil en aquellas aplicaciones donde la calidad subjetiva de la imagen es muy importante.

2.4 Caso discreto: THDM y THDMM

Dado que en la práctica las imágenes obtenidas son representadas digitalmente para su manipulación y almacenamiento, es necesario una formulación discreta del modelo. En este capítulo definimos tanto la THM discreta (THDM), como la THMM discreta (THDMM) a partir de la representación espacio-escala en el dominio discreto. Esta representación mantiene muchas de las propiedades del caso continuo. También se discuten las consideraciones prácticas para su implantación en un programa de computadora.

2.4.1 Espacio-escala discreto

Un problema concerniente a la representación espacio-escala de señales discretas es cómo discretizar la teoría espacio-escala y al mismo tiempo mantener las propiedades del caso continuo. Para señales unidimensionales se puede formular una teoría discreta completa siguiendo el mismo axioma de causalidad del caso continuo:

$H(\cdot, s)$ es un kernel discreto de la representación espacio-escala si, y solo si, el número de extremos locales de $L \star H$ no excede el número de extremos de L .

El único kernel de soporte compacto que satisface este axioma es la función binomial generalizada. Y si la definición se combina con el requerimiento de que la familia de transformaciones de suavizado debe obedecer la propiedad de semi-grupo: $H(\cdot, s_2) \star H(\cdot, s_1) = H(\cdot, s_2 + s_1)$, entonces el kernel más conveniente de la representación espacio-escala discreta es la secuencia binomial simétrica $B[x, N] = 2^{-2N} C_{2N}^{x+N}$ para $N \in \mathbb{Z}^+$, $x = -N, \dots, N$ y $B[x, N] = 0$ para cualquier otro valor de x . Este kernel satisface varias propiedades en el dominio discreto que son similares a las del kernel Gaussiano en el dominio continuo; por ejemplo, tiende a la función delta discreta cuando $N \rightarrow 0$, mientras que para valores grandes de N se aproxima a la Gaussiana.

A pesar de que esta formulación de la teoría espacio-escala es completa y análoga al caso continuo, no puede ser extrapolada a espacios de más de una dimensión. Se ha demostrado que usando argumentos similares al caso continuo no existe ningún kernel trivial en dos o más dimensiones que garantice no introducir nuevos extremos locales. Sin embargo, la causalidad de los extremos locales puede ser expresada alternativamente como [31]:

Si para un nivel de escala s_0 , un punto x_0 es un máximo (mínimo) local a ese nivel, entonces su valor no debe incrementar (decrementar) al incrementar la escala.

Una descripción intuitiva de este requerimiento es que previene que los extremos locales sean realizados o aparezcan de la nada. Basado en este axioma, se puede generalizar el kernel binomial a dos dimensiones como $B[x, y, N] = B[x, N]B[y, N]$. En este caso, aunque la escala es la misma en las orientaciones vertical y horizontal, no es así para las orientaciones oblicuas; lo cual supone una limitación para el análisis

de orientaciones.

2.4.2 Derivadas discretas

El equivalente discreto de las derivadas escaladas se escribe en términos de diferencias finitas del kernel binomial, es decir, $B_n[x, N] = \Delta_x^n B[x - n, N - n/2]$ para $n = 0, \dots, 2N$. En lo sucesivo nos referiremos a estas funciones como *Diferencias de Binomial* (DB). Se puede demostrar que las DB tienden asintóticamente a las DG al crecer la escala [53], específicamente, $B_n[x, N] \sim G_n(x, N/4), N \rightarrow \infty$. De ahí que la mayoría de sus propiedades pueden predecirse con bastante precisión a partir de las propiedades correspondientes del caso continuo haciendo $N = 4s$.

La generalización de estos operadores a 2-D es inmediata por la separabilidad del kernel binomial, es decir, $B_{n,m}[x, y, N] = B_n[x, N]B_m[y, N]$. Por lo que el filtro de derivada direccional se puede definir convenientemente como

$$B_n[x, y, N, \theta_j] = \sum_{m=0}^n C_m^n \cos^m(\theta_j) \sin^{n-m}(\theta_j) B_m[x, N] B_{n-m}[y, N] \quad (2.21)$$

donde el parámetro de orientación está discretizado como en el caso continuo: $\theta_j = \theta_0 + j\pi/(n+1)$, para $j = 0, \dots, n$ y un valor θ_0 arbitrario.

2.4.3 Definición de la THDM

El filtro DoG puede ser aproximado por una *Diferencia de Binomiales* (DoB) y dado que las funciones Binomiales son de soporte compacto, su representación mediante las DB se expresa como una suma finita (Apéndice C). A partir de esta aproximación se construye la versión discreta de la THM a la que denominamos THDM. La descomposición de una señal discreta se representa como una suma finita de las *funciones de*

detalle discreta $D_n^{(k)} = (L_n^{(k)})^{\uparrow T_k} \star T_k B_n^{(k)}$, esto es,

$$L[x] = D_0^{(K)}[x] + \sum_{n=1}^{N_0} D_n^{(0)}[x] + \sum_{k=1}^K \sum_{n=1}^{\tau N_k} \Gamma_n^{(k)}(\tau) D_n^{(k)}[x] \quad (2.22)$$

donde, como en el caso continuo, la fase de análisis $L_n^{(k)} = (L \star (-1)^n B_n^{(k)})^{\downarrow T_k}$ puede verse como la proyección de la señal de entrada sobre las *funciones de análisis*

$$B_n^{(k)}[x, \xi] = B_n^*[x - T_k \xi, N_k/2] \quad (2.23)$$

Note la similitud de estas expresiones con las obtenidas para el caso continuo con número finito de canales DoG. De hecho, los coeficientes $\Gamma_n^{(k)}$ tienden a τ^n a medida que el índice de escala k crece ($N_k \rightarrow \infty$). El primer término de la Ecuación (2.22) corresponde al residuo pasobajas, el segundo, al canal DoB entre las escalas cero y N_0 , y el tercero, a los canales espaciados logarítmicamente con parámetro de espaciamiento $\tau = (N_k - N_{k-1})/N_k$. El factor de normalización $T_k = \sqrt{N_k/2}$ debe ser entero. Evidentemente, el valor $T_k = 2^k$ genera una secuencia de escalas $N_k = 2^{2k+1}$ cuya constante de espaciamiento $\tau = 0.75$.

2.4.4 Un algoritmo rápido

Note que, tanto en la fase de análisis como en la fase de síntesis, se requiere convolucionar la señal con los filtros $B_n^*[x, N_k/2]$, $n = 1, 2, \dots$. Estos filtros satisfacen un número de propiedades de recursividad que facilitan su cálculo para valores de N_k grandes. Algunas de sus propiedades se pueden explicar mejor a través de su función de transferencia

$$b_n^{(k)}(z) = \sqrt{C_{N_k}^n} z^{N_k/2} \left(\frac{1 - z^{-1}}{2} \right)^n \left(\frac{1 + z^{-1}}{2} \right)^{N_k - n} \quad (2.24)$$

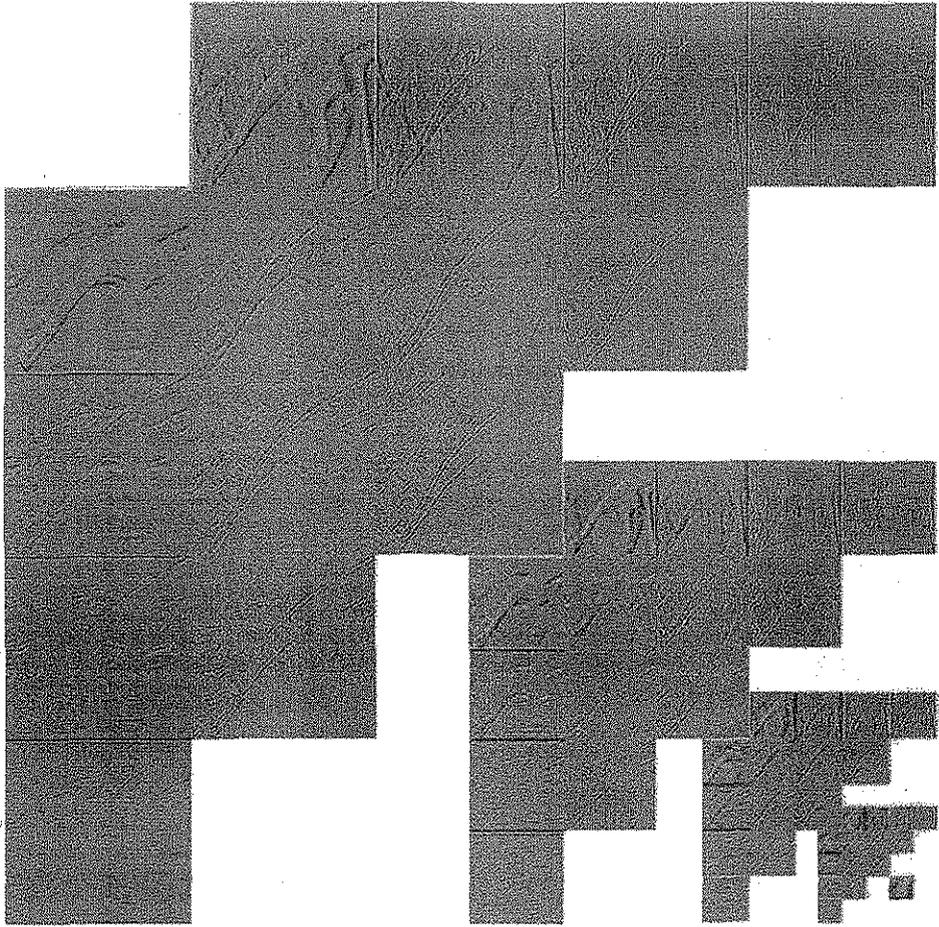


Figura 2.3: Descomposición piramidal THDM de la imagen LENA.

Observe que la máscara de convolución del filtro de orden n puede ser generado por $N - n$ convoluciones sucesivas de la secuencia $\{1, 1\}$ y n convoluciones sucesivas con la secuencia $\{1, -1\}$ más un adelanto de $N_k/2$ muestras, ya que la transformada \mathcal{Z} de estas secuencias corresponden a los binomios $(1 + z^{-1})$ y $(1 - z^{-1})$ respectivamente. Por lo tanto se puede emplear un esquema en cascada usando únicamente (salvo por la normalización) sumas enteras y retardos. Para más detalles consulte la referencia [19].

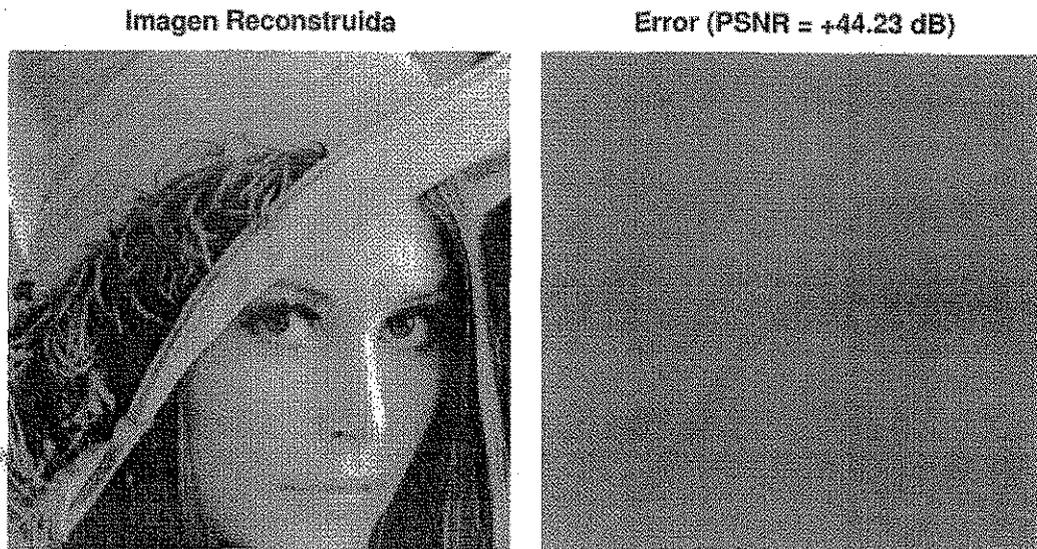


Figura 2.4: Reconstrucción piramidal THDM de la imagen LENA y error de reconstrucción

En algunas aplicaciones se puede emplear el esquema piramidal del caso continuo trasladado al caso discreto a partir de la aproximación de las DG mediante las DB. Por ejemplo, para pasar del nivel k al $k + r$ se convoluciona el coeficiente de orden cero con los filtros $P_n^{(r)}[x] = B_n^*[-x, 4^r - 1]$ seguido de un submuestreo 2^r , mientras que para ir en sentido inverso, primero intercalamos $2^r - 1$ ceros entre las muestras de los coeficientes y luego convolucionamos cada uno con $2^r P_n^{(r)}[-x]$. Lo ideal sería poder escoger $r = 1$, pero desafortunadamente esto no permite calcular coeficientes de orden mayor a 6 requeridos en los niveles superiores. Por lo tanto, se deben emplear los filtros con $r = 2$ para calcular los coeficientes de orden $n = 7, \dots, 30$, con $r = 3$, para $n = 31, \dots, 2(4^3 - 1)$, etc., según se requiera. En la práctica no es necesario obtener los τN_k coeficientes para el nivel k . Experimentalmente se observó que los coeficientes superior a cuarto orden pueden ser ignorados porque, en general, no contienen información relevante. Esto está de acuerdo con datos neurofisiológicos del

sistema visual que establecen que el SVH trabajan con derivadas de hasta cuarto orden [61], [62], [63].

La Figura 2.3 muestra un ejemplo de la THDM usando la aproximación del esquema piramidal en la imagen LENA. La THDM inversa de estos coeficientes reconstruye la imagen de la Figura 2.4 con la imagen de error mostrada a la derecha. Observe que el error se concentra en los contornos de la imagen. De hecho, se trata de un sobre paso en los bordes, el cual los hace ver más contrastados. Este tipo de error podría ser benéfico en imágenes de bajo contraste.

2.4.5 Definición de la THDMM

La THDM se generaliza directamente a dos o más dimensiones en forma similar al caso continuo, mientras que la versión discreta de la THMM se calcula partir de ésta mediante la definición de la Ecuación 2.21. Es decir, para un punto genérico se tiene que

$$L_n^{(k,j)} = \sum_{m=0}^n \alpha_{m,n-m}(\theta_j) L_{m,n-m}^{(k)} \quad (2.25)$$

para el análisis de la señal, mientras que para la síntesis de la señal se efectúa la THDM inversa con los coeficientes

$$H_{m,n-m}^{(k)} = c_n \sum_{j=0}^n \alpha_{m,n-m}(\theta_j) L_n^{(k,j)} \quad (2.26)$$

Es de esperarse que, debido al submuestreo, los coeficientes $L_{m,n-m}^{(k)}$ y $H_{m,n-m}^{(k)}$ reproduzcan imágenes ligeramente diferentes. En las Figuras 2.5 y 2.6 se ilustra un ejemplo con la imagen HOUSE. Primero se calculó la THDM (Figura 2.5) y posteriormente se mapearon estos coeficientes mediante la Ecuación (2.25). Para ello se

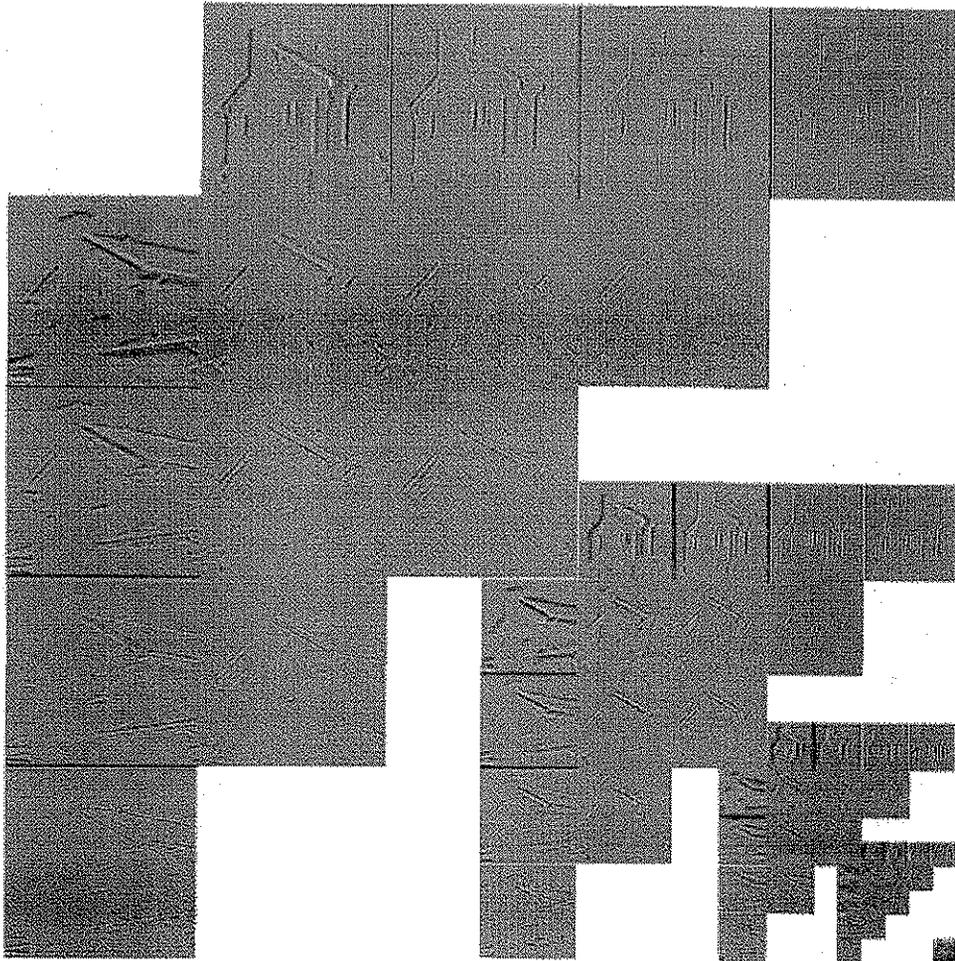


Figura 2.5: Descomposición piramidal THDM de la imagen HOUSE.

eligió $\theta_0 = \arctan(L_{0,1}^{(k)}/L_{1,0}^{(k)})$ en todos los niveles. La elección de este ángulo hace $L_1^{(k,0)} = 0$, por lo que en su lugar se muestra θ_0 (Figura 2.6).

Observe que la energía de los coeficientes de la THDMM se concentra en la dirección horizontal. Las imágenes reconstruidas para estos dos casos se muestran en la Figura 2.7. Se observa que el error debido al truncamiento de la expansión es parecido en ambos casos, pero la ventaja de la THDMM es que tiende a compactar más la energía a lo largo de una coordenada.

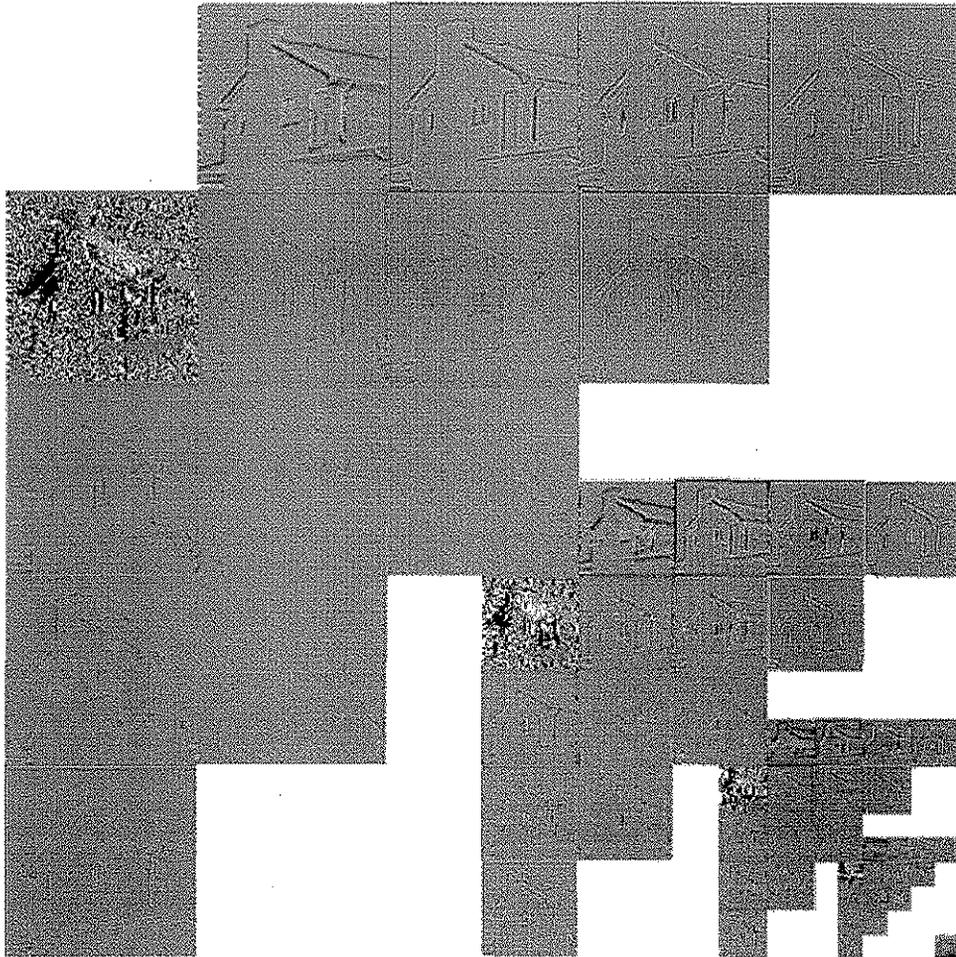


Figura 2.6: Descomposición piramidal THDMM de la imagen HOUSE.

2.5 Conclusión

En este capítulo se introdujo un modelo multicanal, en el que cada canal tiene la forma del perfil de los campos receptivos de las células ganglionares. Se mostró como al representar estos canales mediante los operadores derivadas de Gaussiana se lleva a cabo un análisis similar al de la transformada de Hermite jerárquica [35], [34], y opcionalmente, una descomposición multidireccional en 2-D. La diferencia con la THJ radica en que los coeficientes de transformación son reescalados a coordenadas

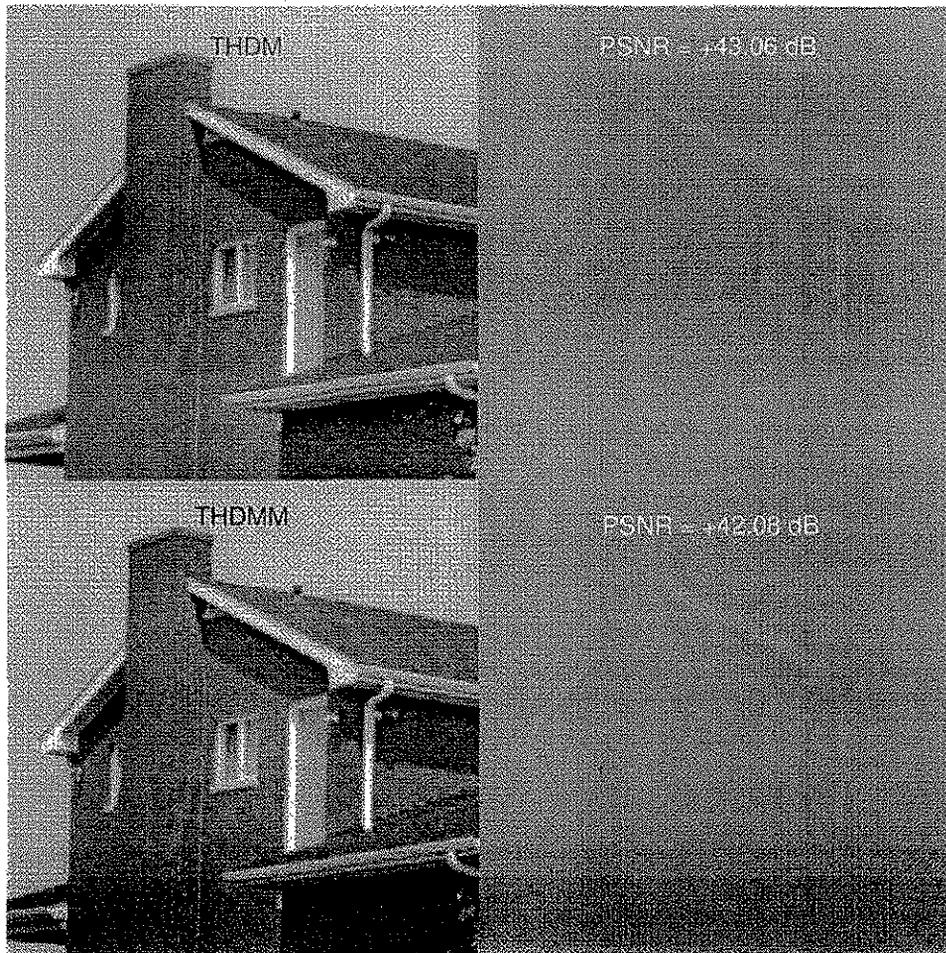


Figura 2.7: Imágenes reconstruidas y errores para los casos de las Figuras 2.5 y 2.6 respectivamente.

normalizadas en lugar de ser submuestreados (sin embargo, en el dominio discreto la normalización de coordenadas se aproxima por el remuestreo). Otra diferencia se tiene en la reconstrucción de la señal. Mientras que la THJ predice la respuesta entre niveles sucesivos, la THM suma las contribuciones de todos los niveles. Este proceso de integración es similar al de la transformada *Cortex* y surge naturalmente porque la descomposición de la señal es creada a partir de un proceso de substracción con los filtros DoG.

TESIS CON
FALLA DE ORIGEN

Se mostró cómo una aproximación de primer orden produce una descomposición más apropiada para compresión, con la particularidad de mejorar el contraste en los bordes. Sin embargo, en aplicaciones donde la reconstrucción de la señal debe ser numéricamente idéntica a la señal original, deberá considerarse una aproximación de orden mayor.

Se definió la versión discreta de la THM en función de filtros binomiales, los cuales pueden ser eficientemente calculados. Las simulaciones hechas en computadora demostraron que si se considera un número de términos no menor a 4, la reconstrucción es visualmente sin pérdida. Desde un punto de vista fisiológico, se puede concluir que el SVH presenta limitaciones en la discriminación de variaciones de orden superior. Desde el punto de vista psicofísico el resultado puede ser interpretado de la siguiente manera: como el error introducido crece gradualmente con la escala porque el número de términos ignorados aumenta con ésta (es decir, que tiende a concentrarse en bandas cercanas a la frecuencia cero), se produce un efecto que es aproximado a variar la iluminación global de la escena, lo cual no influye significativamente en el brillo percibido de acuerdo a la ley de constancia de brillo que establece que: "el brillo percibido en un objeto es aproximadamente independiente de la iluminanci3n" [12].

Se definió también la THDMM empleando una aproximación de las derivadas direccionales de Gaussiana. Se observó que al elegir la orientación del gradiente como orientación principal, la energía de la señal tiende a concentrarse en los coeficientes de una de las coordenadas.

Capítulo 3

Clasificación de estructuras en el dominio de la THDR

En este capítulo se presenta un esquema de representación de imágenes basado en una clasificación perceptual de los coeficientes de la transformada de Hermite discreta rotada (THDR) sobre una latís de muestreo aproximadamente hexagonal. Las funciones de análisis de la THD corresponden a funciones binomiales que aproximan a las derivadas de Gaussiana con la ventaja de que pueden ser calculadas mediante un algoritmo rápido. El muestreo hexagonal es muy eficiente en la transferencia de información espacial y además tiene gran relevancia en el SVH. Los coeficientes de un mismo orden se mapean mediante una transformación unitaria especificada localmente. La transformación se basa en funciones binomiales generalizadas, y como resultado se obtienen los coeficientes de la THD referidos a un sistema de coordenadas rotado de tal forma que la energía de los coeficientes se concentra a lo largo de una coordenada. Dicha representación permite un proceso de clasificación perceptual en el cual se define un umbral de error máximo tolerado para la aproximación del bloque por una constante (patrón 0-D) o por una estructura orientada (patrón 1-D). La compresión se consigue, entonces, eliminando los coeficientes que son visualmente

irrelevantes.

3.1 Introducción

Los métodos estándares de compresión basados en transformadas están orientados principalmente a reducir la *redundancia interpixel* mediante la decorrelación de los píxeles en un entorno definido por una región rectangular. Se han estudiado formas discretas de la transformadas de Fourier [41], Hadamard [10], Coseno [1], Karhunen-Loève [26] y Slant como vehículo para la reducción de la tasa de bit necesaria en la transmisión de señales. En particular, la transformada Karhunen-Loève (TKL) decorrelaciona óptimamente los datos. Sin embargo, dado que la transformación es dependiente de la imagen, los requerimientos de cálculo han desviado el interés hacia la transformada coseno discreta (TCD), la cual es independiente de la imagen y exhibe un nivel de decorrelación comparable a la TKL.

En el esquema estándar, los bloques transformados son cuantizados en forma independiente. Esto implica que la imagen comprimida a bajas tasas de bit presente un "efecto de bloque" que resulta molesto a la vista. Para reducir este tipo de distorsión se ha propuesto un post-procesamiento que consiste en suavizar los bordes de la cuadrícula de bloques [21], [2], [3]. Otro inconveniente de esta representación, es que en muchos sistemas de visualización y procesamiento, especialmente en aquellos que son interactivos, no se hace explícita la información visual relevante. En este trabajo empleamos la versión discreta de la transformada de Hermite [35] sobre una rejilla de muestreo aproximadamente hexagonal. Se ha demostrado que una versión modificada de la THD es tan eficiente como la TCD [18], con la ventaja adicional en la rapidez del algoritmo requerido y en la relevancia de las funciones de análisis para

la extracción de estructuras primitivas en una escena [34].

Otra redundancia que se intenta eliminar es la de tipo psicovisual, es decir, aquella información que no es relevante para la percepción de la escena. El mecanismo estándar empleado para la reducción de este tipo de redundancia es la cuantización basada en modelos de percepción visual (para el caso de la TCD vea por ejemplo [57] y [58]). Aquí proponemos la eliminación de coeficientes espectrales irrelevantes con base en la dimensionalidad de la estructura local de la imagen. Las estructuras son clasificadas en 0-D, 1-D y 2-D de acuerdo a ciertos umbrales de percepción¹.

3.2 Descomposición de la imagen

El primer paso consiste en el mapeo de los datos de la imagen (mediante la THD a una sola escala) en un conjunto de coeficientes espectrales a partir de los cuales es posible obtener una reconstrucción perfecta de la imagen.

La THD se basa en los *filtros binomiales*

$$B_{n,k} = (-1)^n 2^{-N} \sqrt{C_N^n} \Delta_k^n \{C_{N-n}^{k-n}\} \quad (3.1)$$

para $n, k = 0, \dots, N$, donde n es el orden de derivación y $N + 1$ es la longitud del filtro. En este trabajo se emplearon filtros con N igual a una potencia de 2.

Por otro lado, se definen las *funciones binomiales*

$$A_{n,k} = \sqrt{2^{-N}} \Delta_k^n \{C_{N-n}^{k-n}\} \quad (3.2)$$

para $n, k = 0, \dots, N$, las cuales están relacionadas con los filtros binomiales mediante

$B_{n,k} = W_k A_{n,N-k}$, donde

$$W_k = \sqrt{2^{-N} C_N^k} \quad (3.3)$$

¹Una versión preliminar de este capítulo fue presentada en el *SPIE's 46th Annual Meeting* [49]

es la *ventana binomial*.

Las funciones binomiales satisfacen un número de propiedades importantes las cuales pueden ser estudiadas usando la representación matricial $\mathbf{A} = [A_{n,k}]_{n,k=0,\dots,N}$.

Por ejemplo, para $N = 8$ la matriz se escribe explícitamente como

$$\frac{1}{\sqrt{2^8}} \begin{pmatrix} 1 & 8 & 28 & 56 & 70 & 56 & 28 & 8 & 1 \\ 1 & 6 & 14 & 14 & 0 & -14 & -14 & -6 & -1 \\ 1 & 4 & 4 & -4 & -10 & -4 & 4 & 4 & 1 \\ 1 & 2 & -2 & -6 & 0 & 6 & 2 & -2 & -1 \\ 1 & 0 & -4 & 0 & 6 & 0 & -4 & 0 & 1 \\ 1 & -2 & -2 & 6 & 0 & -6 & 2 & 2 & -1 \\ 1 & -4 & 4 & 4 & -10 & 4 & 4 & -4 & 1 \\ 1 & -6 & 14 & -14 & 0 & 14 & -14 & 6 & -1 \\ 1 & -8 & 28 & -56 & 70 & -56 & 28 & -8 & 1 \end{pmatrix} \quad (3.4)$$

donde cada renglón representa una máscara de análisis.

Las funciones binomiales satisfacen la propiedad de ortogonalidad $\mathbf{A}\mathbf{A} = \mathbf{I}$, donde \mathbf{I} es la matriz identidad. Por lo tanto, la transformación definida por la matriz \mathbf{A} lleva a cabo las mismas operaciones tanto para el análisis como para la síntesis. La salida de los filtros binomiales en una posición genérica se puede obtener mediante la transformación definida por la matriz $\mathbf{B} = \mathbf{W}\mathbf{A}$, donde $\mathbf{W} = \text{diag}(W_k)_{k=0,\dots,N}$ es una matriz diagonal que representa la ventana binomial. De la propiedad de ortogonalidad de la matriz \mathbf{A} se puede demostrar que $\mathbf{B}\mathbf{B}^T = \mathbf{W}^2$.

Si \mathbf{F} es una matriz cuadrada que contiene los datos de un bloque de $(N+1) \times (N+1)$ de la imagen de entrada, entonces $\mathbf{G} = \mathbf{B}\mathbf{F}\mathbf{B}^T$ son coeficientes de la THD. Se dice que la THD es una expansión local de los datos de la imagen ya que para cada punto sobre una latís de muestreo se puede obtener la representación local $\mathbf{W}^2\mathbf{F}\mathbf{W}^2 = \mathbf{B}^T\mathbf{G}\mathbf{B}$.

3.3 Reconstrucción de la imagen

La separación entre puntos vecinos de la rejilla de muestreo debe ser tal que exista traslape entre los bloques. El traslape es necesario por la localización de las funciones binomiales. Esto reduce el efecto de bloque a expensas de incrementar el número de datos necesarios para la reconstrucción de la imagen, lo que, a su vez, implica que existe más de una fórmula de reconstrucción. Para determinar la fórmula de reconstrucción óptima consideramos el siguiente análisis de error.

Sea \mathbf{f} el vector columna que contiene los datos de un segmento de la señal (por simplicidad consideramos el caso de una dimensión, la generalización a dos dimensiones es inmediata). Entonces, $\mathbf{g} = \mathbf{B}\mathbf{f}$ es el vector de coeficientes de la THD para un punto genérico de la rejilla de muestreo. Si estos coeficientes son perturbados por ruido blanco con media nula, digamos $\hat{\mathbf{g}} = \mathbf{g} + \mathbf{e}$, con $E\{\mathbf{e}\} = 0$ y $E\{\mathbf{e}\mathbf{e}^T\} = \sigma_e^2\mathbf{I}$, entonces el error de reconstrucción de un segmento (sin traslape) $\epsilon = \mathbf{f} - \mathbf{B}^{-1}\hat{\mathbf{g}}$ también es no correlacionado con media nula. Específicamente, $E\{\epsilon\epsilon^T\} = (\mathbf{W}^{-1})^2\sigma_e^2$, de donde se observa que la matriz \mathbf{W}^{-1} implicaría una mayor varianza en la orilla del segmento (en dos dimensiones esto se traduce en la acentuación del *efecto de bloque*).

Para reducir este efecto es necesario considerar segmentos traslapados y, en ese caso, la reconstrucción de la señal se lleva a cabo mediante un promedio ponderado de los segmentos reconstruidos, donde el peso de cada uno debe ser proporcional al inverso de su varianza, es decir,

$$f_k = \frac{\sum_p W_{k-p}^2 \hat{f}_{k-p}}{\sum_p W_{k-p}^2} \quad (3.5)$$

donde el numerador se expresa matricialmente como $\mathbf{W}^2\hat{\mathbf{f}} = \mathbf{B}^T\hat{\mathbf{g}}$ para todo punto p de la latís de muestreo. La expresión (3.5) no es otra cosa que la generalización del

teorema de fusión demostrado en [51], el cual establece que la función f_k así calculada es la mejor combinación lineal de los segmentos reconstruidos \hat{f}_{k-p} en el sentido del error cuadrático medio.

3.4 Latís de muestreo

El denominador de la Ecuación (3.5) es referido como *función de peso* y depende de la latís empleada. Note que el muestreo espacial debe tener una distribución tal que la división por la función de peso no haga demasiado sensible la reconstrucción de la imagen. Martens [35] sugiere un límite en el que la distancia entre muestras vecinas de una rejilla cuadrada es no mayor a $\sqrt{2N}$. Este límite define un radio de influencia $r = \sqrt{N}$ de las funciones de análisis tal como se ilustra en la Figura 3.1-(a). Sin embargo, para efectos de compresión encontramos más conveniente considerar $r = \sqrt{2N}$, y una latís cuadrada a 45° como la que se muestra en la Figura 3.1-(b). Esto reduce la redundancia intrínseca de la THD a la mitad sin introducir distorsiones significativas aún en condiciones de cuantización severa.

La latís de muestreo cuadrada a 45° (*quincunx grid*) se define matemáticamente como el conjunto de puntos de la forma

$$\mathbf{p} = T \begin{pmatrix} 1 & 0 \\ 0.5 & 0.5 \end{pmatrix} \begin{pmatrix} n \\ m \end{pmatrix} \quad (3.6)$$

donde n y m son enteros y T es un parámetro de distancia entre muestras. En nuestro caso elegimos $T = 2r$. En la práctica se puede calcular la THD estándar con espaciamiento entre muestras igual a $\sqrt{2N}$ seguido de un muestreo entrelazado con $T = 1$.

La distribución de estos puntos de muestreo es más parecida a la distribución de las

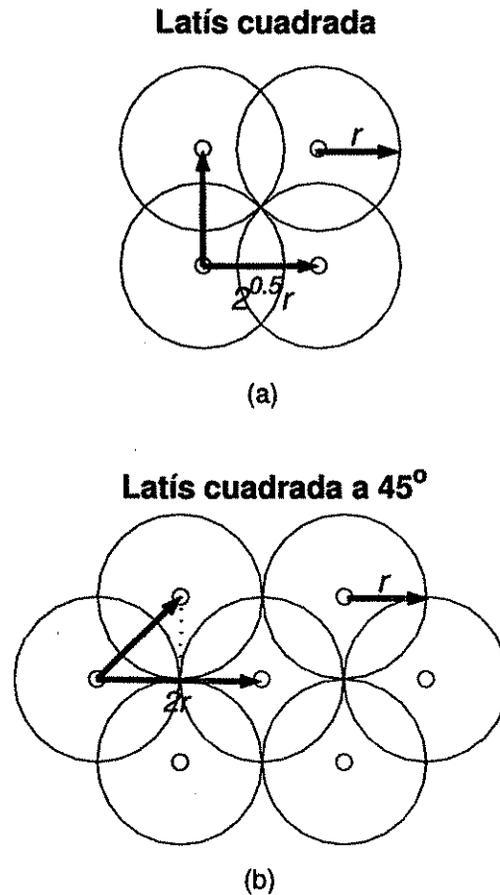


Figura 3.1: Distribución de los operadores sobre rejillas de muestreo cuadradas.

células de la retina (la cual exhibe una geometría hexagonal [22]), que la distribución de la rejilla empleada en la definición original de la transformada de Hermite.

3.5 Rotación local

La rotación local en el dominio de la transformada puede verse como un mapeo de los coeficientes espectrales (para cada punto en la rejilla de muestreo) a través de una transformación unitaria (Apéndice D). Esta se lleva a cabo sobre los coeficientes

de un mismo orden y puede ser representada matricialmente como se describe a continuación.

Sean \mathbf{g}_n para $n = 0, \dots, 2N$, los vectores columna formados por las antidiagonales de la matriz de coeficientes \mathbf{G} y sean las *funciones binomiales generalizadas* definidas como

$$A_{n,k}^{(\theta)} = c^{-k} s^k \Delta^n \{ C_{N-n}^{k-n} c^{2k-n} s^{N-2k+n} \} \quad (3.7)$$

con $c = \cos \theta$ y $s = \sin \theta$ para $0 \leq \theta < \pi$.

Se puede ver que si el parámetro de sesgo θ se hace igual $\pi/4$ se obtiene el caso simétrico, es decir, $A_{n,k} = A_{n,k}^{(\pi/4)}$. Usando la notación matricial denotamos la matriz de funciones binomiales generalizadas por $\mathbf{A}^{(\theta)}$. Esta matriz satisface la misma propiedad de ortogonalidad del caso simétrico. Las primeras tres matrices (quitando el caso $N = 0$, el cual se reduce a la unidad) tienen la siguiente expresión explícita

$$\begin{pmatrix} s & c \\ c & -s \end{pmatrix}, \begin{pmatrix} s^2 & 2sc & c^2 \\ sc & c^2 - s^2 & -sc \\ c^2 & -2sc & s^2 \end{pmatrix}, \begin{pmatrix} s^3 & 3s^2c & 3sc^2 & c^3 \\ s^2c & 2sc^2 - s^3 & c^3 - 2s^2c & -sc^2 \\ sc^2 & -2s^2c + c^3 & s^3 - 2sc^2 & s^2c \\ s^3 & -3s^2c & 3sc^2 & -c^3 \end{pmatrix} \quad (3.8)$$

La rotación de los vectores de coeficientes para un ángulo θ dado se obtienen a partir de la transformación definida por la matriz $\mathbf{R}^{(\theta)} = \mathbf{W}\mathbf{A}^{(\theta)}\mathbf{W}^{-1}$, es decir que el vector $\mathbf{g}_n^{(\theta)} = \mathbf{R}^{(\theta)}\mathbf{g}_n$ representa los coeficientes de orden n referido a los ejes coordenados rotados por un ángulo θ medido en sentido inverso a las manecillas del reloj, mientras que la rotación inversa se consigue aplicando la matriz de rotación transpuesta a los vectores de coeficientes rotados. En lo subsecuente haremos referencia a los coeficientes rotados a través de la matriz $\mathbf{G}^{(\theta)}$ cuyas antidiagonales son justamente los vectores $\mathbf{g}_n^{(\theta)}$.

En todos los ejemplos presentados aquí, se eligió el ángulo del gradiente: $\theta = \arctan(G_{0,1}/G_{1,0})$.

3.6 Clasificación

La idea de clasificar los bloques de coeficientes previamente rotados tiene la finalidad de reducir la redundancia aprovechando el hecho de que en muchos casos el valor de la imagen puede ser localmente aproximado por estructuras que son 0-D o 1-D.

Para bloques localizados en regiones homogéneas, todos los coeficientes, excepto $G_{0,0}$ (el coeficiente de DC), tienen poca relevancia en la reconstrucción de la señal. Mientras que los coeficientes para bloques localizados sobre los contornos presentan una concentración de energía importante sobre la primera columna de la matriz $\mathbf{G}^{(\theta)}$ (por simplicidad se omitirá el super índice de aquí en adelante) y, en tal caso, los coeficiente $G_{i,j}$ para $j > 0$ pueden ser despreciados. En el resto de los casos deberán codificarse todos los coeficientes.

Si las clases se etiquetan como 0-D, 1-D y 2-D respectivamente se tienen las siguiente hipótesis

- H_0 = el bloque es 0-D
- H_1 = el bloque es 1-D
- H_2 = el bloque es 2-D

La partición del espacio en las clases definidas arriba puede llevarse a cabo más fácilmente asumiendo que cualquiera de ellas puede ser perfectamente descrita por

las siguientes variables:

$$L = G_{0,0}, C = \left[\sum_{i=0}^N \sum_{j=0}^N G_{i,j}^2 - L^2 \right]^{1/2}, \Delta C = \left[C^2 - \sum_{i=1}^N G_{i,0}^2 \right]^{1/2} \quad (3.9)$$

las cuales proporcionan medidas de intensidad media, contraste, y contraste residual

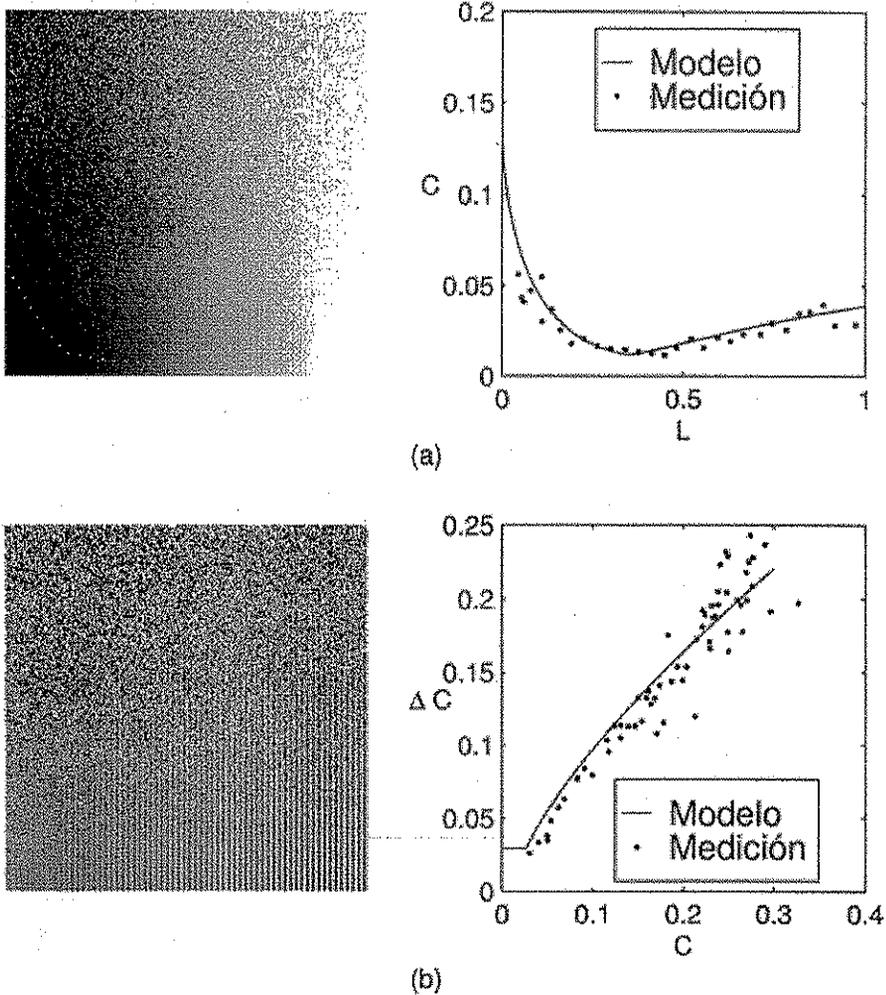


Figura 3.2: Umbral de detección de contraste (a) y umbral de enmascaramiento de contraste (b).

La clasificación se lleva a cabo en dos pasos. En primer lugar separamos los

bloques 0-D mediante la comparación

$$C \underset{H_1 \cup H_2}{\lesssim}^{H_0} k_0 C_{thr}(L) \quad (3.10)$$

donde la curva $C_{thr}(L)$ representa el *umbral de detección de contraste* [4]. Esta curva se infiere mediante la imagen de estímulo de la Figura 3.2-(a) (izquierda). La imagen presenta ruido blanco gaussiano cuya media varía linealmente en la dirección horizontal y cuya desviación estándar varía linealmente en la dirección vertical. La curva trazada sobre la imagen es una curva de detección típica que señala el umbral a partir del cual el contraste local es visualmente relevante.

El umbral se ajustó por el siguiente modelo

$$C_{thr} = C_{min} + \left| \frac{L^\alpha - L_{min}^\alpha}{L^\alpha + L_{min}^\alpha} \right|^{1/\alpha} \quad (3.11)$$

donde el exponente α está entre 0 y 1 (típicamente 0.6) y C_{min} es el contraste mínimo al nivel de intensidad L_{min} , el cual define a su vez el nivel de máxima sensibilidad al contraste. En la Figura 3.2-(a) (derecha) se muestran la curva teórica de la Ecuación (3.11) y las mediciones sobre el perfil mostrado en la imagen. Se observa que el modelo ajusta bastante bien los datos subjetivos.

En el segundo paso, separamos la clase 1-D del conjunto de bloques no clasificados mediante la comparación

$$\Delta C \underset{H_2}{\lesssim}^{H_1} k_1 \Delta C_{thr}(L, C) \quad (3.12)$$

donde ΔC_{thr} es el *umbral de enmascaramiento de contraste*.

Idealmente, si un bloque contienen una estructura orientada, toda la energía queda contenida completamente en la dirección del gradiente y el contraste residual es cero. En la práctica, sin embargo, lo más común es que exista un contraste residual diferente de cero aún en los casos en que visualmente se perciba una estructura orientada. Esto

se debe al fenómeno visual conocido como *enmascaramiento de contraste*, el cual se refiere justamente a la reducción en la visibilidad de una componente de la imagen debido a la presencia de otra.

El modelo de enmascaramiento visual que empleamos aquí está basado en los trabajos de Legge y Foley [29], [28]:

$$\Delta C_{thr} = \max(C_{thr}, C^\beta C_{thr}^{1-\beta}) \quad (3.13)$$

donde el exponente β está entre 0 y 1. Note que cuando $\beta = 0$ no existe enmascaramiento y el umbral es constante igual a C_{thr} . Cuando $\beta = 1$ se tiene lo que comúnmente se denomina *ley de Weber*. Un valor típico es $\beta = 0.7$. En la Figura 3.2-(b) muestra la curva teórica (derecha) y las mediciones obtenidas a lo largo de la curva mostrada en la imagen de estímulo (izquierda). La curva define un umbral subjetivo que separa las estructuras 1-D de las estructuras 2-D. En este caso la imagen de estímulo está compuesta de barras senoidales cuya amplitud varía linealmente en dirección horizontal más ruido blanco con desviación estándar variando linealmente en dirección vertical.

En la Figura 3.3 se muestran algunos ejemplos de clasificación basada en el modelo de percepción visual con las imágenes PEPPER, HOUSE, CAMERA y MIT respectivamente. La ventaja de usar una latís *quincunx* y la clasificación de los coeficientes se ilustra en la Figura 3.4 donde se muestran las imágenes reconstruidas y los errores de reconstrucción para tres casos. El primer codificador se basa en la THD estándar, el segundo incluye la latís *quincunx* y el tercero incluye además la clasificación perceptual. En los tres casos se consideraron los valores de $N = 8$ y $T = 4$, y se eliminaron todos los coeficientes de orden superior a N . La compresión en bits por pixel (bpp) se estimó con base en la entropía de los coeficientes. Para el

caso con clasificación se agregaron dos bits adicionales para codificar la clase a la que pertenece el bloque de coeficientes. Se observó una mejora substancial en la tasa de compresión sin introducir artefactos visualmente significativos.

3.7 Conclusión

Las principales aportaciones de este capítulo son: 1) La implantación de la THD sobre una latís aproximadamente hexagonal (conocida comúnmente como *quincunx grid*). Esta latís permite definir una distancia entre muestras vecinas que es mayor a la distancia sugerida en la definición original. El incremento de este parámetro reduce la redundancia inherente de la THD a la mitad. 2) Se propuso la generalización del algoritmo de Hashimoto-Sklansky basado en funciones binomiales generalizadas para el mapeo de los coeficientes espectrales referidos a un sistema de coordenadas que está orientado en dirección del gradiente. Como resultado de la rotación local, la energía se compacta en un número menor de coeficientes. 3) Para aprovechar la compactación de la energía se clasificaron los coeficientes mediante un modelo de percepción visual para estructuras 0-D, 1-D y 2-D representadas en el dominio de la transformada. Se observó que el modelo de percepción visual puede ser ajustado muy bien a los umbrales psicofísicos encontrados en el SVH.

Los trabajos futuros orientados sobre esta línea de investigación deberán considerar una evaluación cuantitativa de la eficiencia de esta representación para un esquema de codificación de imágenes, así como la integración de la etapa de clasificación de los coeficientes dentro de un cuantizador perceptivo.

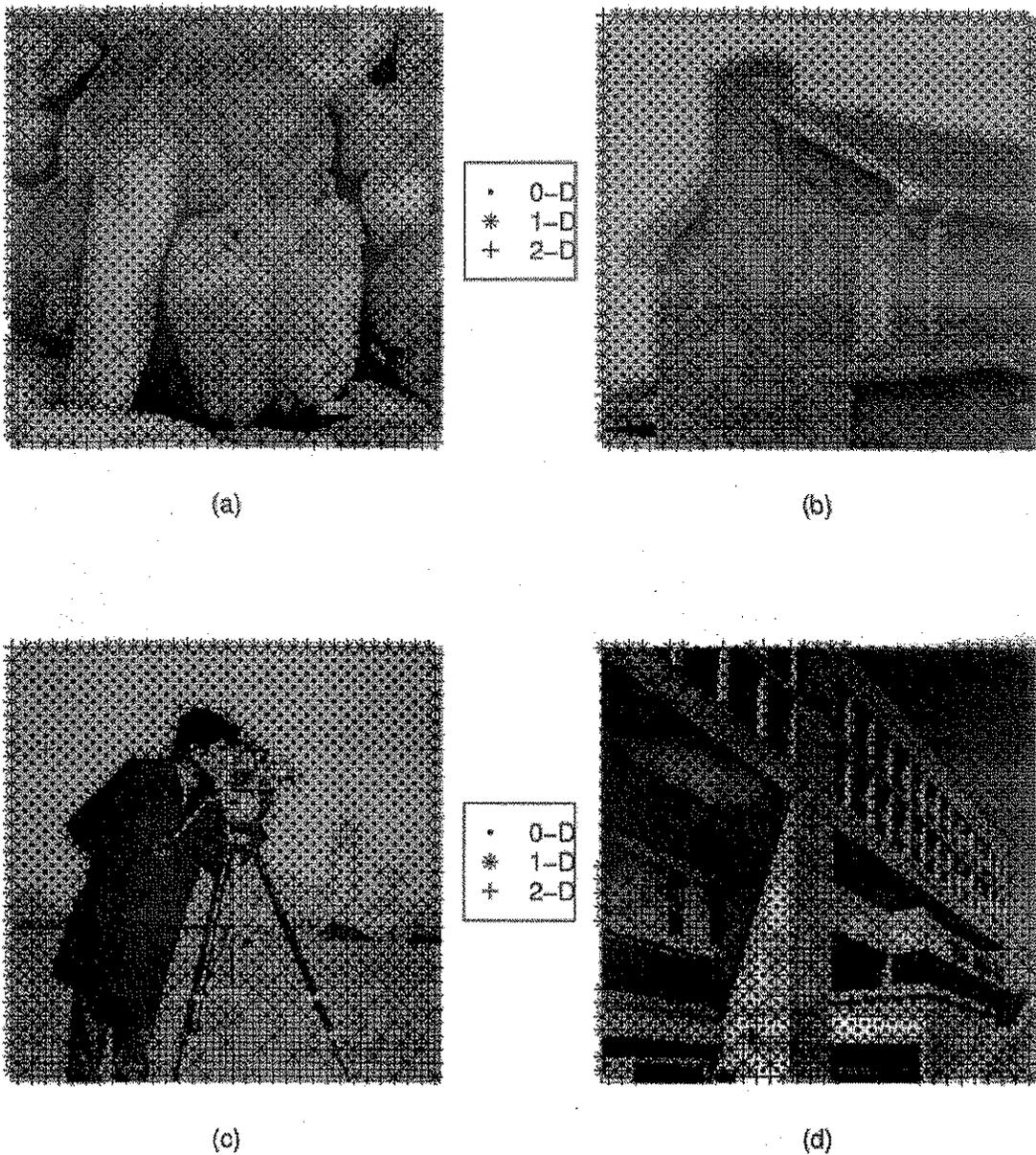


Figura 3.3: Resultados de la clasificación basada en el modelo de percepción.

TESIS CON
FALLA DE ORIGEN

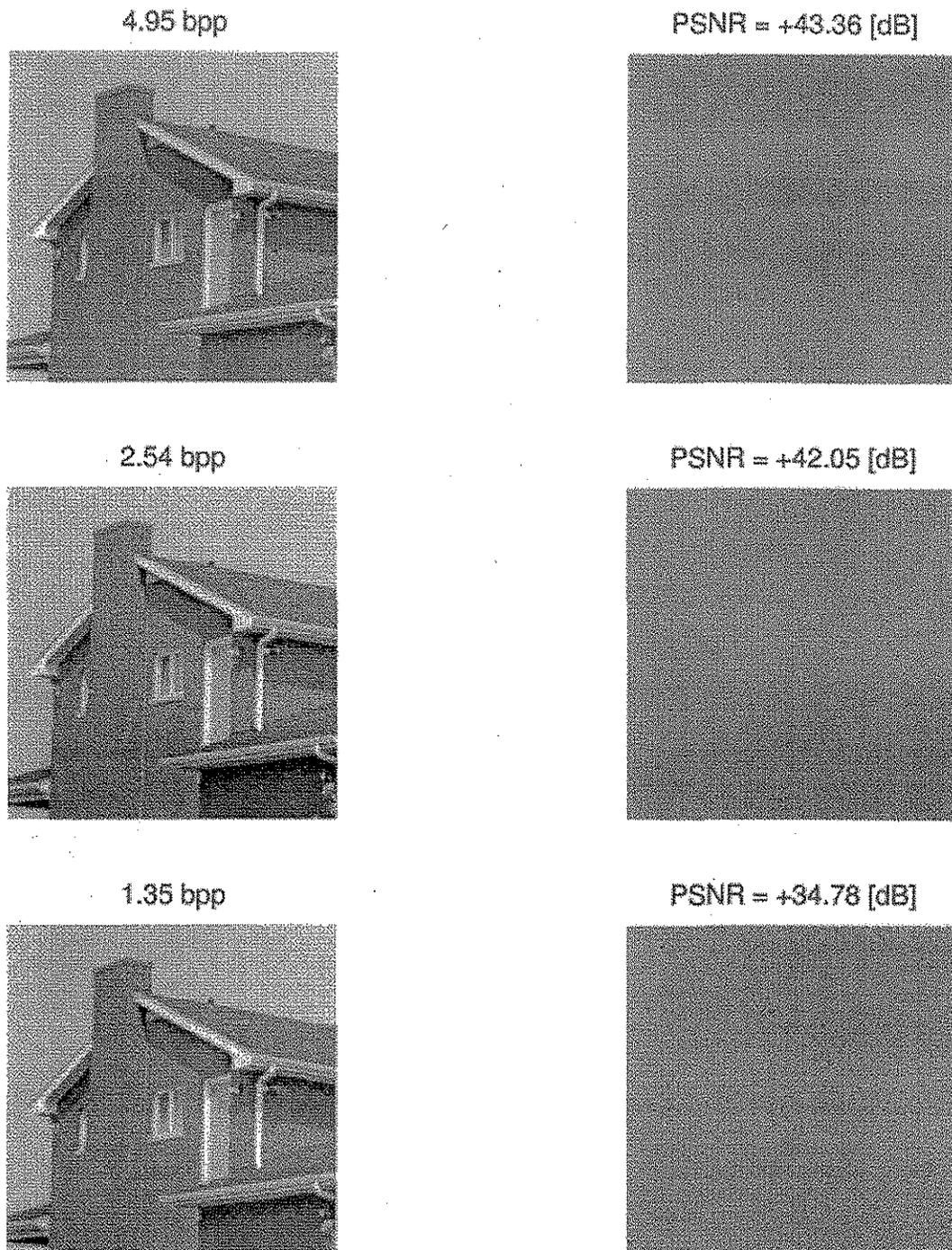


Figura 3.4: Imagen HOUSE codificada y error de reconstrucción para: la THD estándar (arriba), la THD con la latís quincunx (en medio) y la THD con clasificación perceptual sobre la latís quincunx (abajo).

Capítulo 4

Compresión basada en la descomposición THD-Quadtree

En este capítulo se define un algoritmo de descomposición multiescala no redundante basado en una estructura de árbol cuaternario (*quadtree*). Cada nodo del árbol es representado en el dominio de la transformada de Hermite discreta (THD) por un conjunto de coeficientes espectrales. La descomposición involucra la localización de la señal por ventanas binomiales de diferentes tamaños traslapadas entre sí, así como la expansión local de dichas señales en un conjunto de los polinomios de Krawtchouk (que son la versión discreta de los polinomios de Hermite). El algoritmo emplea el método de subdivisión sucesiva de bloques de la resolución más baja a la más alta. Los bloques que no satisfacen una condición de homogeneidad son subdivididos en bloques más pequeños y analizados en escalas menores (mayor resolución). Dicha condición involucra un error perceptual definido en el dominio de la transformada. La reconstrucción de los datos de la imagen se lleva a cabo mediante la integración de los bloques ponderados por la ventana y divididos entre una función de peso asociada al quadtree. El quadtree debe ser tal que la imagen reconstruida sea visualmente sin pérdidas. La descomposición, denominada THD-Quadtree, se insertó en un esquema

de compresión con un cuantizador vectorial. El diseño del cuantizador se llevó a cabo mediante el algoritmo estándar LBG. Los resultados comparativos respecto a la TCD se presentan al final del capítulo.

4.1 Introducción

Uno de los principales problemas en la codificación de imágenes a bajas tasas de bit es la no estacionaridad de la mayoría de las imágenes. Un algoritmo más eficiente de codificación de imágenes y video debe compensar la no estacionaridad mediante la adaptación espacial asignando una tasa extra para información adicional sobre la descripción de las distintas áreas predominantes en la imagen. Uno de los primeros tipos de representación multiescala de los datos de imagen fue introducido por Klinger (1971). Es una representación con estructura de árbol en la cual la imagen es subdividida recursivamente en regiones pequeñas. La idea básica es la siguiente: Considere, por simplicidad, una imagen L de tamaño $2^K \times 2^K$ para K entero positivo, y defina una medida de variación del nivel de gris Σ para cualquier región. Esta medida puede ser, por ejemplo, la desviación estándar de los niveles de gris. Sea $L^{(K)} = L$. Si $\Sigma(L^{(K)})$ es mayor que un umbral especificado α , entonces divida $L^{(K)}$ en subimágenes $L_j^{(K-1)}$ ($j = 1, \dots, p$) de acuerdo a una regla especificada. Aplique el mismo proceso recursivamente a todas las subimágenes hasta la convergencia. Este algoritmo genera un árbol de grado p en el que cada hoja es un bloque homogéneo con $\Sigma(L_j^{(k)}) \leq \alpha$.

El uso de la estructura quadtree para la codificación de imágenes ha sido considerado ampliamente ([11], [43], [45], [52], [55]), pero hasta ahora no se han empleado bloques traslapados con una medida de homogeneidad perceptual que produzca una

reconstrucción libre del efecto de bloque para altas tasas de compresión. En el presente capítulo se describe una descomposición quadtree basada en la transformada de Hermite discreta que denominamos THD-Quadtree¹ (o THD-Q). Las características son: el empleo de bloques traslapados, expansión local truncada y umbral perceptual para la decisión de subdivisión de bloques. El truncamiento es tal que el número total de coeficientes no supera el número de datos originales, obteniendo una ganancia en la entropía resultante. Evidentemente la imagen reconstruida no es numéricamente idéntica a la original, pero se intentó que fuera visualmente idéntica.

4.2 Descomposición THD-Q

En el capítulo previo se definió la matriz de transformación de la THD como $\mathbf{B} = [B_{n,k}]_{n,k=0,\dots,N}$ donde $B_{n,k}$ es una secuencia binomial en k de longitud $N + 1$ que aproxima la derivada de Gaussiana de orden n . Para un bloque \mathbf{F} de $(N + 1) \times (N + 1)$ de la imagen de entrada, $\mathbf{G} = \mathbf{B}\mathbf{F}\mathbf{B}^T$ son coeficientes de la THD.

Como la descomposición THD-Q considera bloques de entrada de tamaño igual a una potencia de 2, la distancia entre puntos vecinos de la latís debe ser $\sqrt{2}r_j = 2^{j+1}$ para $j = 1, \dots, K$ donde K es la profundidad máxima del árbol. Además, si consideramos que el radio de influencia de las funciones binomiales es $r_j = \sqrt{2N_j}$, como se vió en el capítulo previo, entonces las escalas empleadas deben ser $N_j = 4^j$. Sin embargo, dada la restricción impuesta en el tamaño de los bloques de ser una potencia de 2, empleamos $N_j = 4^j - 1$. Por otro lado, como se desea que el número de muestras resultantes no sea mayor al número de muestras de la imagen original, es

¹Un trabajo relacionado fue presentado en la *IEEE International Conference on Telecommunications* [48]

necesario eliminar los coeficientes de orden mayor a 2^j para el nivel j . Así, definimos las matrices de transformación $\mathbf{B}_j = [B_{n,k}]_{n=0,\dots,2^j-1; k=0,\dots,4^j-1}$. Note que para el nivel j los bloques de entrada deben ser de tamaño 4^j mientras que el bloque de salida es de tamaño 2^j . Esto implica que debe existir traslape entre las expansiones locales en puntos vecinos de la latís asociada al quadtree.

4.2.1 Umbral de homogeneidad

El umbral para detectar si un bloque es homogéneo se basa en la función de enmascaramiento de contraste ΔC_{thr} definida en el capítulo previo. En este caso se define el contraste residual en función de la energía contenida en los coeficientes que son despreciados, es decir,

$$\Delta C = \left[\sum_{n=2^j+1}^{N_j} \sum_{m=2^j+1}^{N_j} G_{n,m}^2 \right]^{1/2} \quad (4.1)$$

Por lo tanto si $\Delta C > k\Delta C_{thr}$ se lleva a cabo la subdivisión del bloque en 4 bloques iguales. Los cuales son subsecuentemente mapeados mediante la matriz \mathbf{B}_j correspondiente. El proceso se repite recursivamente (a menos que $j = 1$) para el siguiente nivel $j - 1$.

En la Figura 4.3 se muestran algunos ejemplos de segmentación quadtree obtenidos usando este algoritmo con una constante de proporcionalidad $k = 0.2$. Los bloques considerados son de 4, 8, 16 y 32 pixeles. La constante k se ajusta al nivel de compresión deseado ya que mientras mayor es el valor de k menor es el número de nodos hoja del árbol.

4.2.2 Reconstrucción THD-Q

Para el proceso de reconstrucción se suman las contribuciones traslapadas de cada expansión local. Una contribución en el nivel j se expresa matricialmente como $\mathbf{W}_j^2 \mathbf{F}_j \mathbf{W}_j^2 = \mathbf{B}_j^T \mathbf{G}_j \mathbf{B}_j$, donde $\mathbf{G}_j = \mathbf{B}_j \mathbf{F}_j \mathbf{B}_j^T$ son los coeficientes de la transformación y \mathbf{W}_j es la matriz diagonal definida por la ventana binomial en el nivel j . La suma de todas las contribuciones divididas por su correspondiente función de peso² da como resultado la imagen reconstruida.

4.2.3 Información lateral

Dado un quadtree con l nodos hoja, existen $i = (l - 1)/3$ nodos internos. Un método eficiente de transmisión del quadtree es codificar las subdivisiones en secuencia *depth first*. Esto requiere de $b = l + i$ bits. Si el bit 1 indica una subdivisión y 0 indica un nodo terminal, la Figura 4.1 se codificaría como 1 0 10000 (1 0 10000 0 0) 0, donde los espacios y paréntesis se introducen por claridad. En este caso $l = 13$, $i = 4$ y $b = 17$.

4.3 Cuantización vectorial

Para efectos de compresión se cuantizaron los bloques de coeficientes mediante cuantizadores vectoriales. Se diseñó un cuantizador vectorial para cada nivel del árbol mediante el algoritmo estándar LBG [30]. Este algoritmo está basado en el enfoque de Lloyd [33], no es una técnica variacional y no involucra diferenciación, por lo que trabaja bien incluso cuando la distribución tiene componentes discretas. Sin embargo,

²La función de peso es la suma de las ventanas traslapadas

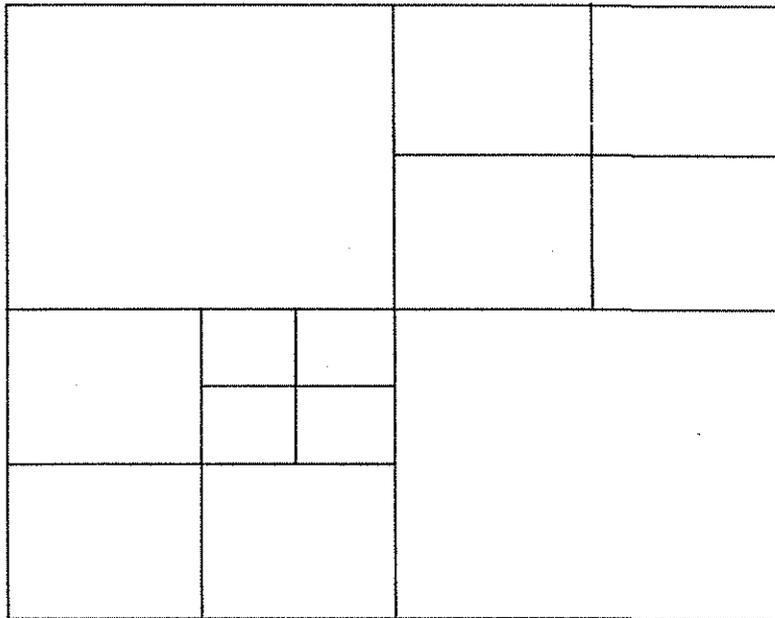


Figura 4.1: Ejemplo de un quadtree de profundidad 3 y 13 nodos hoja.

solo se puede garantizar la optimalidad local. El algoritmo consta de los siguientes pasos:

- (0) Inicialización: Dar el número de niveles de cuantización Q , el umbral de distorsión ϵ , un alfabeto de reproducción C_0 y una secuencia de entrenamiento $\{x_j; j = 0, \dots, n-1\}$. Hacer $m = 0$ y $D_{-1} = \infty$.
- (1) Dado $C_m = \{y_i; i = 1, \dots, Q\}$, encuentre la partición $P(C_m) = \{S_i; i = 1, \dots, Q\}$ con minima distorsión para la secuencia de entrenamiento: $x_j \in S_i$ si $d(x_j, y_i) \leq d(x_j, y_k)$ para toda k . Calcular la distorsión promedio³

$$D_m = \frac{1}{n} \sum_{j=0}^{n-1} \min_{y \in C_m} d(x_j, y) \quad (4.2)$$

³Se empleó la distancia euclidiana como medida de distorsión.

- (2) Si $(D_{m-1} - D_m)/D_m \leq \epsilon$, terminar con C_m como el alfabeto de reproducción; de otra forma, continuar.
- (3) Encontrar el alfabeto de reproducción óptimo $C_{m+1} = \{y_i; i = 1, \dots, Q\}$ con

$$y_i = \frac{1}{\#S_i} \sum_{x_j \in S_i} x_j \quad (4.3)$$

Reemplazar m por $m + 1$ y continuar con el paso (1).

Una vez obtenido el libro de código C_m se usa con datos fuera del conjunto de entrenamiento con la regla del vecino más cercano, la cual define una partición óptima del espacio euclidiano.

4.3.1 Inicialización del libro de códigos

La optimalidad global del quantizador está sujeta a la elección del libro de código inicial C_0 . Linde *et al* [30] proponen un método de "partición sucesiva" del libro de códigos comenzando con el centroide de la secuencia de entrenamiento. El método se resume como sigue:

- (0) Inicialización: Hacer $q = 1$ y defina $C_0 = \sum_{j=0}^{n-1} x_j/n$, el centroide de la secuencia de entrenamiento completa.
- (1) Dado $C_0 = \{y_i, i = 1, \dots, q\}$, "partir" cada vector y_i en dos vectores cercanos $y_i + \epsilon$ y $y_i - \epsilon$, donde ϵ es un vector fijo de perturbación. La nueva colección $C'_0 = \{y_i + \epsilon, y_i - \epsilon, i = 1, \dots, q\}$ tiene $2q$ vectores. Reemplazar q por $2q$
- (2) si $q < Q$, ejecutar el algoritmo para un quantizador de q niveles comenzando con C'_0 para generar un alfabeto C_0 más adecuado y regresar al paso (1). Si $q = Q$, hacer $C_0 = C'_0$ y terminar.

Este algoritmo es útil cuando se desea diseñar cuantizadores de tasas sucesivamente mayores hasta que se alcanza un nivel de distorsión aceptable.

4.4 Resultados

Las curvas de compresión en bits por pixel (bpp) contra el PSNR (*pseudo-signal to noise ratio*) mostradas en la Figura 4.2 demuestran un buen desempeño del codificador para altas tasas de compresión. Estas curvas fueron generadas variando únicamente la constante k (de 0.1 a 1.5 con incrementos de 0.1) para mover el umbral de homogeneidad, manteniendo constante el número de niveles de cuantización para cada escala de acuerdo a la Tabla 4.1.

j	N_j	bloque	Q_j
1	3	4×4	512
2	15	8×8	256
3	63	16×16	128
4	255	32×32	64

Tabla 4.1: Niveles de cuantización.

La evaluación del desempeño real del codificador requiere una comparación subjetiva entre imágenes codificadas con otros métodos estándares. Desafortunadamente este trabajo de comparación demanda mucho tiempo y experiencia en evaluación perceptual. Para los propósitos de esta tesis nos limitamos a señalar las diferencias que son más o menos evidentes entre dos esquemas de compresión. La Figura 4.4 muestra la imagen PEPPERS comprimida a 0.16 *bpp* con la TCD estándar y la descomposición THD-Q. En ambos casos se emplearon cuantizadores vectoriales diseñados mediante el algoritmo LBG con las mismas imágenes de entrenamiento.

Para el ejemplo Figura 4.4 y para la mayoría de los casos se observó que a pesar de

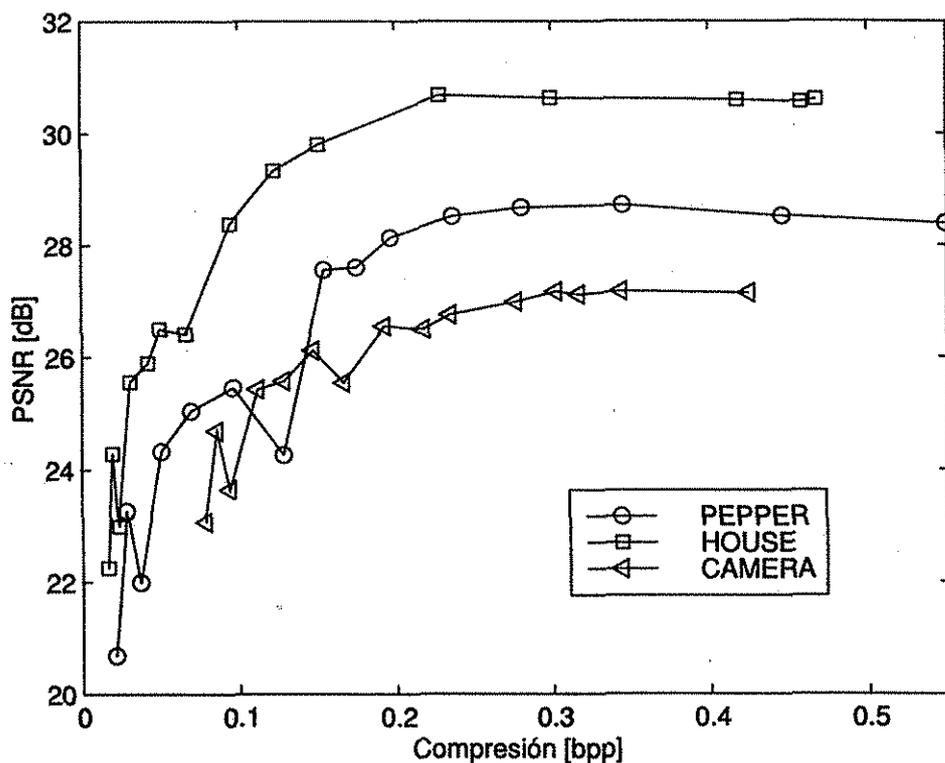


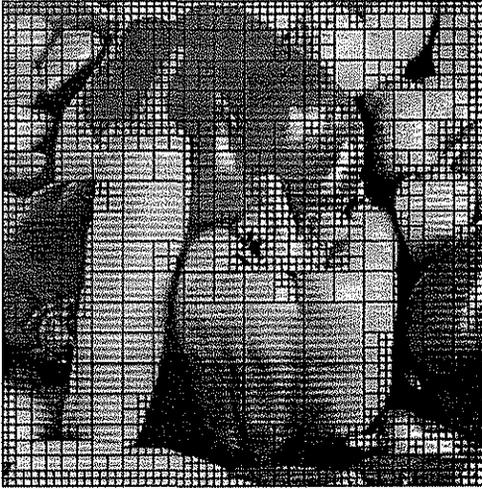
Figura 4.2: Curvas de desempeño del codificador DHT-Q para las imágenes PEPPER, HOUSE y CAMERA.

que la TCD presenta un mayor PSNR, ésta produce un efecto de bloque muy marcado sobre superficies homogéneas que hace ver la imagen más degradada que la comprimida con la THD-Q. En cambio la principal degradación producida por la THD-Q consiste en un efecto de sierra a lo largo de los bordes debido a la cuantización. Este efecto, sin embargo, podría ser disminuido si se usa un esquema de clasificación de estructuras, (como en el capítulo previo) a expensas de un incremento en la complejidad del cuantizador. Se observa también que las regiones homogéneas presentan un suavizado que resulta ser más agradable a la vista que la "bloquificación" introducida con el codificador de la TCD.

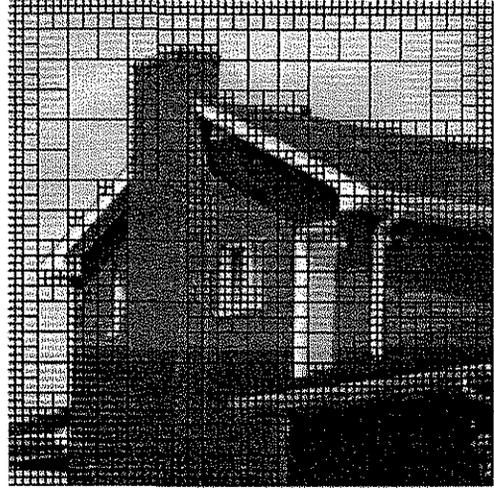
TESIS CON
FALLA DE ORIGEN

4.5 Conclusión

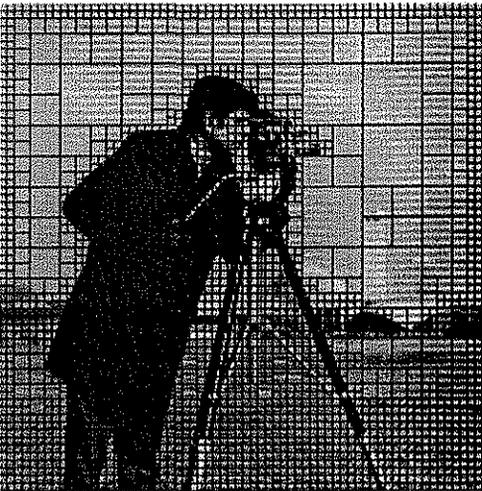
Se presentó un esquema de compresión de imágenes basado en una descomposición multiescala denominada THD-Q. Dicha representación emplea versiones discretas de derivadas de Gaussiana con diferentes amplitudes (escalas) en posiciones definidas por un quadtree. La expansión local de la imagen en función de las derivadas de Gaussiana queda limitada por el número de datos que son representados. De esa manera no se introduce redundancia en la representación. La reconstrucción a partir de los coeficientes espectrales (sin cuantización) no es numéricamente exacta, pero si la segmentación quadtree es correcta, las pérdidas por truncamiento no son visualmente aparentes. Por otra parte, se demostró cómo las degradaciones debido a la cuantización son menos significativas, hablando en términos visuales, a las producidas por la TCD para una misma tasa de compresión, no obstante que esta última presentó, en general, una mejor relación señal a ruido.



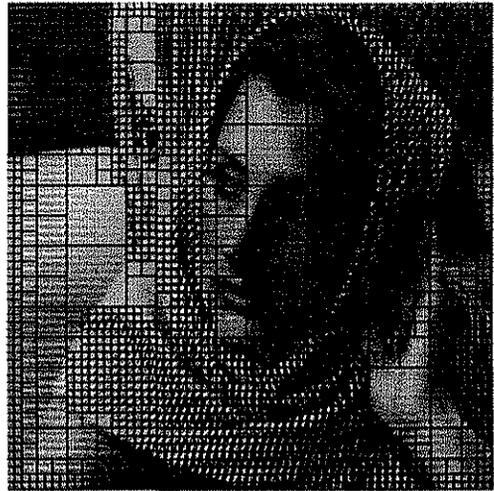
(a)



(b)



(c)



(d)

Figura 4.3: Segmentación quadtree de las imágenes PEPPER, HOUSE, CAMERA Y BARBARA.

TESIS CON
FALLA DE ORIGEN

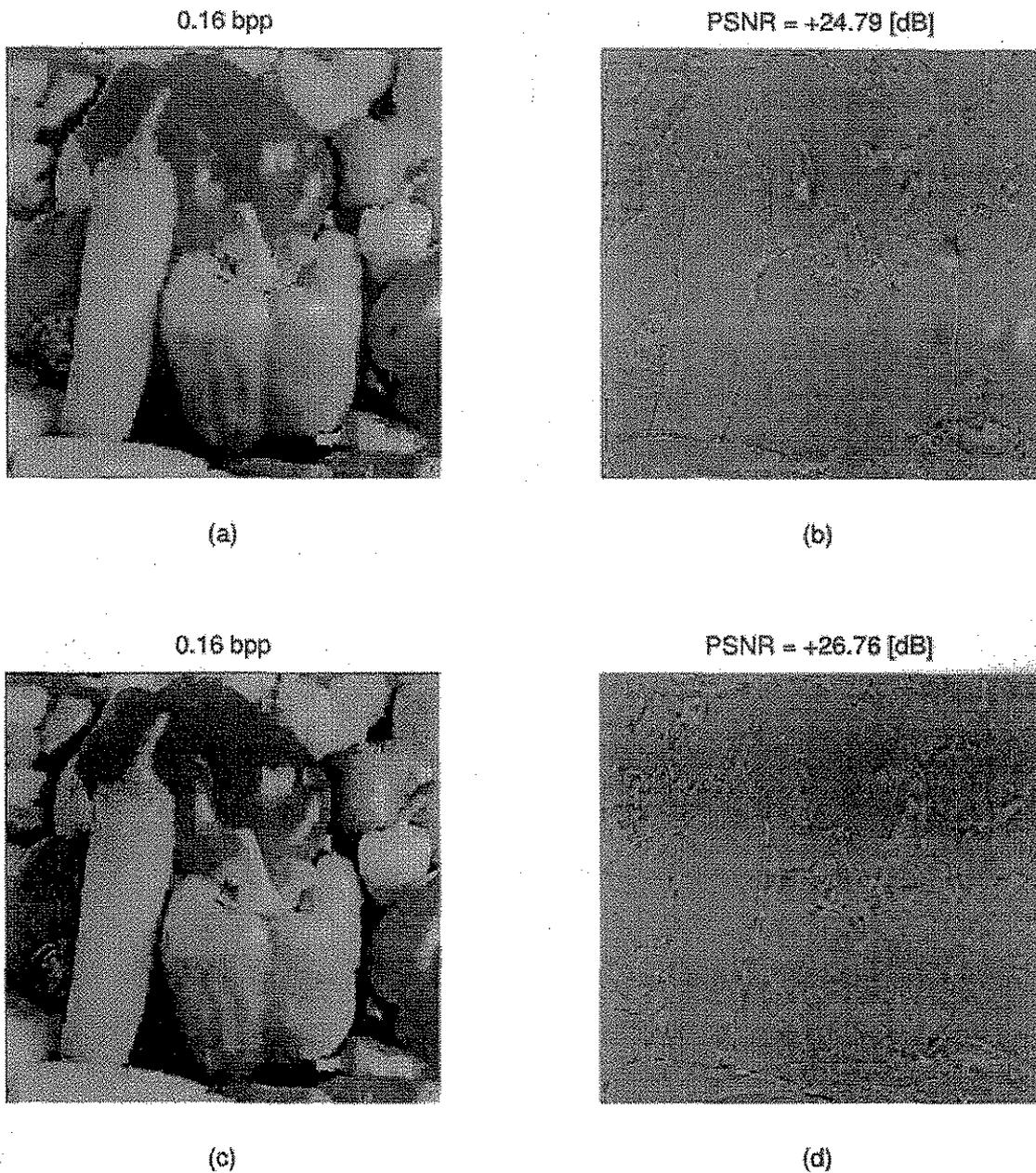


Figura 4.4: Resultados de compresión con la THD-Q (a)-(b) y con la TCD (c)-(d)

Capítulo 5

Discusión

En este trabajo hemos descrito varios esquemas de representación de imágenes para la compresión. Dichos esquemas están basados en modelos Gaussianos de las etapas primarias de procesamiento del Sistema Visual Humano (SVH). Las características que se intentaron emular incluyen procesamiento multiresolución, análisis local de orientaciones, clasificación de patrones visuales y la distribución espacial de los operadores.

Pero, cuál es el mejor algoritmo de compresión con pérdida?. Desafortunadamente no hay una respuesta absoluta a esta pregunta. La elección de un algoritmo particular para una aplicación dada depende de muchos factores. Por ejemplo cuando la compresión se usa en una aplicación de transmisión de imágenes, el codificador y decodificador deben trabajar en tiempo real, y los temas sobre complejidad de implantación, susceptibilidad a errores del canal, y requerimientos de memoria intermedia para igualar la tasa de salida del codificador a la tasa de transmisión del canal se vuelven importantes. En contraste cuando la compresión se usa para reducir los requerimientos de almacenamiento, la codificación no se requiere que sea en tiempo real. El codificador puede ser bastante complejo ya que es usado sólomente una vez

para una imagen dada, mientras que es deseable un decodificador simple porque es usado repetidamente. Además, los porcentajes de error encontrados en el almacenamiento y en aplicaciones de recuperación son típicamente de muchos órdenes de magnitud menores que los porcentajes de error para canales de comunicación. Esto permite crear algoritmos más sofisticados con mayor énfasis en la calidad final de la imagen.

La siguiente es una lista de factores que pueden influir en la elección de un algoritmo de compresión. En general, el peso de cada factor en una decisión depende mucho de la aplicación. Esta lista no es exhaustiva e intenta servir únicamente como guía general.

- **Sensibilidad a las imágenes de entrada.** Dado que en este trabajo nos enfocamos en los algoritmos de compresión de imágenes naturales en tonos de grises continuos, no discutiremos la sensibilidad para imágenes con texto o imágenes generadas por computadora. Dentro de la clase de imágenes de tonalidades continuas, las características de la imagen de entrada tales como rango dinámico, ruido de imagen, contenido frecuencial, correlación pixel a pixel, y resolución de la imagen pueden afectar todo el desempeño de un algoritmo. Por ejemplo, la descomposición de una imagen en quadtree, es altamente dependiente del contenido frecuencial y de la correlación pixel a pixel, lo cual impacta directamente la tasa de compresión. Además, el esquema de detección de homogeneidad requiere ajustar ciertos parámetros (umbrales) para alcanzar un buen desempeño con una clase dada de imágenes, pero el desempeño puede ser afectado notablemente si se introducen imágenes de otra clase como por ejemplo de menor contraste y mayor contenido frecuencial.

- **Taza de bit operacional.** En algunas aplicaciones, la prioridad es alcanzar un alto grado de compresión incluso a expensas de una imagen de baja calidad. En contraste, otras aplicaciones podrían requerir un alto grado de calidad de imagen que únicamente puede ser alcanzado a tazas modestas de compresión. Dichos requerimientos pueden limitar severamente la elección del algoritmo de compresión. En general, para la mayoría de los esquemas de compresión, existe un rango de la tasa de salida para la cual el algoritmo es más eficiente, es decir, el rango donde opera muy cerca de la curva $R(D)$.
- **Taza de bit constante vs calidad constante.** Los algoritmos que operan con una tasa de bit constante (tales como VQ o DPCM, sin códigos entrópicos) son más apropiados para aplicaciones de transmisión cuando se usa un canal de tasa fija sin memoria intermedia o para almacenamiento cuando se especifica el espacio de almacenamiento disponible. Desafortunadamente, debido a la amplia variación en el contenido de información de diferentes imágenes, dichos esquemas no proporcionan una calidad de reconstrucción constante. La tasa de compresión especificada puede ser innecesariamente alta para ciertas imágenes, mientras que otras pueden resultar de calidad insatisfactoria. En contraste, algunos esquemas pueden mantener una calidad visual constante a expensas de una tasa de compresión variable. Lo cual demanda un adecuado mecanismo de almacenamiento intermedio en aplicaciones de transmisión. En nuestro caso, esta característica depende en un cien por ciento del cuantizador. Así por ejemplo, para un cuantizador vectorial se tiene una tasa de compresión constante.

- **Implantación.** Esto se refiere a la naturaleza y complejidad del algoritmo relativo al *hardware* y *software* del ambiente en el que se desea construir. Se deben considerar tres aspectos del algoritmo: (1) complejidad computacional; es decir, número de corrimientos, sumas, multiplicaciones, comparaciones u otras operaciones requeridas por pixel, (2) requerimientos de memoria, y (3) factibilidad de paralelización o cualquier otra estructura de procesamiento eficiente. Dado que el modelo gaussiano se lleva a cabo mediante funciones binomiales de complejidad del orden de $N \log_2 N$ adiciones más N multiplicaciones de normalización para N datos, resulta una implantación de bajo costo y de tiempos de operación muy cortos. Además el algoritmo es altamente paralelizable.
- **Simetría codificador/decodificador.** Los codificadores y decodificadores de igual complejidad pueden ser aceptables en muchas aplicaciones de transmisión, sin embargo en aplicaciones donde es usado recurrentemente, tal como en un sistema de almacenamiento y consulta, es más deseable un decodificador simple. Tanto en la descomposición THDR como la THD-Q se lleva a cabo un análisis local de los datos para extraer información adicional sobre clasificación de patrones (THDR) o de segmentación (DTH-Q), lo que simplifica considerablemente el proceso de reconstrucción. Como consecuencia inmediata se tiene que el codificador es más complejo que el decodificador. En general, la inclusión de cualquier mecanismo de adaptabilidad en el sistema altera el balance entre la complejidad del codificador y del decodificador.
- **Tolerancia a errores del canal.** Desafortunadamente, uno de los precios que se paga por la compresión de los datos es el incremento en la susceptibilidad de los datos a los errores del canal, y el grado de susceptibilidad puede variar

mucho entre diferentes esquemas. En el caso de los algoritmos de procesamiento por bloques (por ejemplo, TCD), el efecto de un bit erróneo queda confinado a un solo bloque de la imagen, mientras que en otros resulta en una degradación severa en toda la imagen (DPCM). Una de las ventajas de la THD, en cualquiera de sus presentaciones, es que la redundancia es efectivamente empleada para reducir los errores externos. A diferencia de los algoritmos convencionales de procesamiento por bloque, la THD emplea bloques traslapados con una ventana de análisis que pone mayor énfasis al centro del bloque y menos en los bordes. La relativamente menor importancia de las orillas del bloque para reconstruir un pixel central se compensa mediante el traslape de los bloques lo que da como resultado que un error de bit en un bloque se reparta en los bloques vecinos, de tal forma que el error es menos notable.

- **Artefactos.** Los artefactos introducidos por un algoritmo dependen de su forma de operación. Algunos pueden incluso introducir más de un tipo de artefacto dependiendo de la tasa de compresión a la cual opere. Ya hemos enfatizados que ciertos artefactos tales como el efecto de bloque y bordes con forma de sierra son más molestos a la vista que incluso ruido aleatorio o suavizado global de los bordes. Pero además la visibilidad de estos artefactos depende de las condiciones particulares de visibilidad. Respecto a esta característica observamos que tanto la THDR como la THD-Q presentan resultados favorables. En particular, las comparaciones de la THD-Q contra la TCD demostraron que los artefactos debido a cuantización son menos notables en el primero que en el segundo.
- **Efecto de múltiples codificaciones.** Algunas aplicaciones podrían requerir que la imagen pase a través del compresor-descompresor muchas veces. Por

ejemplo un imagen puede ser comprimida y transmitida a un destino donde se descomprime y se despliega. Se altera una pequeña porción de la imagen y se envía a otro destino donde este proceso se repite otra vez. En tales aplicaciones, es esencialmente importante que la porción no alterada de la imagen no se degrade más comparada con la imagen obtenida en la primera compresión. Algunos esquemas de compresión no crean mayor pérdida en la calidad de la imagen cuando se usan sucesivamente en la misma imagen, mientras que otros causan degradaciones adicionales. Afortunadamente en el caso de la THD, como en la mayoría de los casos, las pérdidas subsecuentes son significativamente menor que las producidas en la primera etapa.

- **Capacidad de transmisión progresiva.** La transmisión progresiva permite enviar una imagen aproximada a baja tasa de bit para un rápido reconocimiento, y después se envían los detalles incrementalmente si el usuario los solicita. Por supuesto, cualquier esquema puede hacerse artificialmente progresivo usándolo para codificar una versión de baja resolución de la imagen original como aproximación y, posteriormente, codificando las diferencias entre esta y la original o posiblemente otro estado intermedio para crear tantos niveles de detalles como se deseen. Como se describió en el Capítulo 2 La THD multiescala es inherentemente realizable para su transmisión progresiva debido a la estructura multiresolución engranada, y puede operar en ese modo sin necesidad de requerir mayor complejidad. En cambio otras técnicas pueden operar en un modo progresivo pero a expensas de agregar mayor complejidad al sistema.

- **Compatibilidad del Sistema.** Si el sistema requiere compatibilidad con productos de otros fabricantes, entonces la elección del esquema de compresión puede dictarse por la existencia de estándares que ya han sido propuestos y adoptados como por ejemplo el estándar de facsímil de CCITT o el algoritmo JPEG de compresión para imágenes en color.

Apéndice A

Caso 1-D

Para el caso 1-D, comenzamos con la serie de Taylor:

$$G(x, s_{k-1}) = \sum_{n=0}^{\infty} \frac{(s_{k-1} - s_k)^n}{n!} \left[\frac{\partial^n}{\partial s^n} G(x, s) \right]_{s=s_k} \quad (\text{A.1})$$

y dado que la Gaussiana es una solución de la ecuación de difusión: $\partial_s G(x, s) = \partial_x^2 G(x, s)$ [25], al derivar sucesivamente respecto a s se obtiene

$$\frac{\partial^n G(x, s)}{\partial s^n} = G_{2n}(x, s) \quad (\text{A.2})$$

para $n = 1, 2, \dots$. Sustituyendo (A.2) en (A.1) y restando el primer término de la serie en ambos miembros, se llega directamente a (2.2).

En el siguiente paso aplicamos la factorización (derivada de la propiedad de *cerradura bajo convolución*): $(-1)^n G_{2n}(x, s_k) = G_n(-x, s_k/2) \star G_n(x, s_k/2)$, y al mismo tiempo, aplicamos la relación

$$\sqrt{(4s)^n} G_n(x, s) = \frac{1}{\sqrt{4s}} G_n\left(\frac{x}{\sqrt{4s}}\right) \quad (\text{A.3})$$

donde, por definición, $G_n(x) = G_n(x, 1/4)$ para $n = 0, 1, 2, \dots$ son las DG en coordenadas naturales. El resultado al que se llega es

$$DoG(x, s_{k-1}, s_k) = \sum_{n=1}^{\infty} \tau^n \frac{1}{\sqrt{2s_k}} G_n^* \left(\frac{-x}{\sqrt{2s_k}} \right) \star \frac{1}{\sqrt{2s_k}} G_n^* \left(\frac{x}{\sqrt{2s_k}} \right) \quad (\text{A.4})$$

donde $G_n^*(x)$ denota la normalización $G_n(x, 1/4)/\sqrt{2^n n!}$ respecto al producto interno

$$\int_{-\infty}^{\infty} \frac{G_n^*(\xi) G_m^*(\xi)}{G(\xi)} d\xi = \delta_{n,m} \quad (\text{A.5})$$

donde $\delta_{n,m}$ es la delta de Kronecker.

Considerando la expansión (A.4) en la Ecuación (2.1), se tiene

$$L(x) = \sum_{k=-\infty}^{\infty} \sum_{n=1}^{\infty} \tau^n \left\{ L(x) \star \frac{1}{\sqrt{2s_k}} G_n^* \left(\frac{-x}{\sqrt{2s_k}} \right) \right\} \star \frac{1}{\sqrt{2s_k}} G_n^* \left(\frac{x}{\sqrt{2s_k}} \right) \quad (\text{A.6})$$

pero si definimos

$$L_n^{(k)}(\xi) = \left[L(x) \star \frac{1}{\sqrt{2s_k}} G_n^* \left(\frac{-x}{\sqrt{2s_k}} \right) \right]_{x=\sqrt{2s_k}\xi} \quad (\text{A.7})$$

como la *fase de análisis* de la señal, entonces las definiciones

$$D_n^{(k)}(x) = L_n^{(k)} \left(\frac{x}{\sqrt{2s_k}} \right) \star \frac{1}{\sqrt{2s_k}} G_n^* \left(\frac{x}{\sqrt{2s_k}} \right) \quad (\text{A.8})$$

$$L(x) = \sum_{k=-\infty}^{\infty} \sum_{n=1}^{\infty} \tau^n D_n^{(k)}(x) \quad (\text{A.9})$$

corresponden a la *fase de síntesis* de la señal.

Apéndice B

Caso 2-D

Para el caso 2-D partimos de la siguiente expansión:

$$DoG(x, y, s_{k-1}, s_k) = \sum_{n=1}^{\infty} \frac{(-\tau s_k)^n}{n!} \left[\frac{\partial^n}{\partial s^n} G(x, y, s) \right]_{s=s_k} \quad (\text{B.1})$$

donde ahora las derivadas se obtiene a partir de la ecuación de difusión isotrópica en 2-D: $\partial_s G(x, y, s) = \partial_x^2 G(x, y, s) + \partial_y^2 G(x, y, s)$, esto es,

$$\frac{\partial^n}{\partial s^n} G(x, y, s) = \sum_{m=0}^n C_n^m G_{2m, 2n-2m}(x, y, s) \quad (\text{B.2})$$

o, alternativamente, usando las DDG,

$$\frac{\partial^n}{\partial s^n} G(x, y, s) = c_n \sum_{j=0}^n G_{2n, 0}(x, y, s, \theta_j) \quad (\text{B.3})$$

para las orientaciones $\theta_j = \theta_0 + j/(n+1)$, $j = 0, \dots, n$ y las constantes

$$c_n = \frac{1}{\sum_{j=1}^n \text{sen}^{2n} \left(\frac{j\pi}{n+1} \right)} = \frac{1}{n+1} \sum_{m=0}^n \frac{(C_n^m)^2}{C_{2n}^{2m}} \quad (\text{B.4})$$

para $n = 1, 2, \dots$

La sustitución de (B.2) en (B.1) da como resultado la siguiente expresión para la DoG bidimensional:

$$DoG(x, y, s_{k-1}, s_k) = \sum_{n=1}^{\infty} \tau^n \sum_{m=0}^n \frac{1}{2s_k} G_{m,n-m}^* \left(\frac{-x}{\sqrt{2s_k}}, \frac{-y}{\sqrt{2s_k}} \right) * \frac{1}{2s_k} G_{m,n-m}^* \left(\frac{x}{\sqrt{2s_k}}, \frac{y}{\sqrt{2s_k}} \right) \quad (B.5)$$

Mientras que la expresión alternativa (B.3) da como resultado:

$$DoG(\mathbf{r}, s_{k-1}, s_k) = \sum_{n=1}^{\infty} c_n \tau^n \sum_{j=0}^n \frac{1}{2s_k} G_{n,0}^* \left(\frac{-\mathbf{r}\mathbf{R}_j}{\sqrt{2s_k}} \right) * \frac{1}{2s_k} G_{n,0}^* \left(\frac{\mathbf{r}\mathbf{R}_j}{\sqrt{2s_k}} \right) \quad (B.6)$$

con los vectores $\mathbf{r} = (x, y)$, $\rho = (\xi, \eta)$ y la matriz de rotación

$$\mathbf{R}_j = \begin{pmatrix} \cos \theta_j & -\text{sen} \theta_j \\ \text{sen} \theta_j & \cos \theta_j \end{pmatrix} \quad (B.7)$$

De la Ecuación (B.5) se construye la THM-2D cuyas etapas están dadas por:

Análisis

$$L_{m,n-m}^{(k)}(\xi, \eta) = \left[L(x, y) * \frac{1}{2s_k} G_{m,n-m}^* \left(\frac{-x}{\sqrt{2s_k}}, \frac{-y}{\sqrt{2s_k}} \right) \right]_{(x,y)=\sqrt{2s_k}(\xi,\eta)} \quad (B.8)$$

Síntesis

$$D_{m,n-m}^{(k)}(x, y) = L_{m,n-m}^{(k)} \left(\frac{x}{\sqrt{2s_k}}, \frac{y}{\sqrt{2s_k}} \right) * \frac{1}{2s_k} G_{m,n-m}^* \left(\frac{x}{\sqrt{2s_k}}, \frac{y}{\sqrt{2s_k}} \right) \quad (B.9)$$

$$L(x, y) = \sum_{k=-\infty}^{\infty} \sum_{n=1}^{\infty} \sum_{m=0}^n \tau^n D_{m,n-m}^{(k)}(x, y) \quad (B.10)$$

Y para el caso de la Ecuación (B.6) se construye la THMM definida como:

Análisis

$$L_n^{(k,j)}(\rho) = \left[L(\mathbf{r}) * \frac{1}{2s_k} G_{n,0}^* \left(\frac{-\mathbf{r}\mathbf{R}_j}{\sqrt{2s_k}} \right) \right]_{\mathbf{r}=\rho\mathbf{R}_j^T\sqrt{2s_k}} \quad (B.11)$$

Síntesis

$$D_n^{(k,j)}(\mathbf{r}) = L_n^{(k,j)} \left(\frac{\mathbf{r}\mathbf{R}_j}{\sqrt{2s_k}} \right) * \frac{1}{2s_k} G_{n,0}^* \left(\frac{\mathbf{r}\mathbf{R}_j}{\sqrt{2s_k}} \right) \quad (B.12)$$

$$L(\mathbf{r}) = \sum_{k=-\infty}^{\infty} \sum_{n=1}^{\infty} \sum_{j=0}^n c_n \tau^n D_n^{(k,j)}(\mathbf{r}) \quad (\text{B.13})$$

Apéndice C

Caso discreto

En el primer paso, la señal se descompone en un número finito de canales DoB (*Difference of Binomials*), donde $DoB[x, N_1, N_2] = B[x, N_1] - B[x, N_2]$ para $N_1 < N_2$ es la contraparte discreta del filtro DoG. Por lo tanto, la descomposición multiescala finita de la señal toma la forma

$$L[x] = L[x, N_K] + \sum_{k=0}^K L[x] \star DoB[x, N_{k-1}, N_k] \quad (C.1)$$

con las escalas $N_{-1} = 0, N_0 = 4$ y $N_k = N_{k-1}/(1 - \tau)$ para $k = 1, \dots, K$. El operador \star denota convolución discreta.

En el siguiente paso se expresan los canales DoB en función de los operadores DB. Para ello, empleamos la versión discreta de la ecuación de difusión:

$$\Delta_N B[x, N] = 2^{-2} \Delta_x^2 B[x - 1, N] \quad (C.2)$$

donde

$$\Delta_x^n L[x] = \sum_{k=0}^n (-1)^{n-k} C_n^k L[x + k] \quad (C.3)$$

denota el operador *diferencia hacia adelante* de orden n (el subíndice se introduce para indicar respecto a qué variable discreta se hace la operación, pero puede omitirse para funciones de una sola variable). La ecuación de difusión discreta se demuestra por inducción matemática, usando la propiedad de cerradura bajo convolución del kernel binomial en la hipótesis de inducción.

De la ecuación de difusión se puede escribir la DB de orden n como

$$B_n[x, N] = B_n[x, N + 1] - 2^{-2} B_{n+2}[x, N + 1] \quad (\text{C.4})$$

Sustituyendo recurrentemente ($M - 1$ veces) los términos del miembro derecho se obtiene la siguiente expresión para la binomial

$$B[x, N] = \sum_{m=0}^M (-1)^m C_M^m 2^{-2m} B_{2m}[x, N + M] \quad (\text{C.5})$$

Esta expresión es el equivalente discreto de la serie de Taylor de la Gaussiana. Sustituyendo N por N_{k-1} y $N + M$ por N_k y restando el primer término de la suma se obtiene la expresión de la DoB en función de las DB de orden par:

$$DoB[x, N_{k-1}, N_k] = \sum_{n=1}^{\tau N_k} (-1)^n C_{\tau N_k}^n 2^{-2n} B_{2n}[x, N_k] \quad (\text{C.6})$$

Por otro lado se requiere que la factorización de los DB incorpore coordenadas normalizadas discretas con el factor de normalización $T_k = \sqrt{2s_k} = \sqrt{N_k/2}$ entero. Por lo tanto, empleamos la siguiente expresión en las operaciones de convolución de (C.1):

$$DoB[x - y, N_{k-1}, N_k] = \sum_{n=1}^{\tau N_k} \Gamma_n^{(k)}(\tau) \sum_{\xi} T_k B_n^*[y - T_k \xi, N_k/2] B_n^*[x - T_k \xi, N_k/2] \quad (\text{C.7})$$

con los coeficientes $\Gamma_n^{(0)}(\tau) = \Gamma_1^{(k)}(\tau) = 1$ y

$$\Gamma_n^{(k)}(\tau) = \tau^n \prod_{i=1}^{n-1} \frac{1 - i/\tau N_k}{1 - i/N_k} \quad (\text{C.8})$$

para $k = 1, \dots, K$ y $n = 2, 3, \dots, \tau N_k$. La notación $B_n^*[x, N]$ se emplea para las funciones normalizadas $2^{-n} \sqrt{C_{2N}^n} B_n[x, N]$, las cuales satisfacen la siguiente propiedad de ortogonalidad

$$\sum_{x=-N}^N \frac{B_n^*[x, N] B_m^*[x, N]}{B[x, N]} = \delta_{n,m} \quad (\text{C.9})$$

Por lo tanto,

$$L[x] \star DoB[x, N_{k-1}, N_k] = \sum_{n=1}^{\tau N_k} \Gamma_n^{(k)}(\tau) D_n^{(k)}[x] \quad (\text{C.10})$$

donde

$$D_n^{(k)}[x] = \sum_{\xi} L_n^{(k)}[\xi] T_k B_n^*[x - T_k \xi, N_k/2] \quad (\text{C.11})$$

y

$$L_n^{(k)}[\xi] = \sum_y L[y] B_n^*[y - T_k \xi, N_k/2] \quad (\text{C.12})$$

para $n = 1, 2, \dots, \tau N_k$

Por otra parte, el residuo pasobajas puede codificarse usando la aproximación

$$B[x - y, N_K] \simeq \sum_{\xi} T_K B[y - T_K \xi, N_K/2] B[x - T_K \xi, N_K/2] \quad (\text{C.13})$$

Por lo tanto, $L[x, N_K] \simeq D_0^{(K)}[x]$ donde $D_0^{(K)}[x]$ está dado por la generalización de (C.11) y (C.12) para $n = 0$ y $k = K$. Se puede demostrar que a medida que N_K crece, el error de la aproximación (C.13) tiende a cero.

Apéndice D

Rotación de las DG

Sean las DG en el dominio de Fourier

$$g_{n-m,m}(\omega_x, \omega_y) = (j\omega_x)^{n-m} (j\omega_y)^m g(\omega_x, \omega_y) \quad (\text{D.1})$$

donde $g(\omega_x, \omega_y)$ es la transformada de fourier de $G(x, y)$, y sean las constantes $c = \cos \theta$ y $s = \text{sen} \theta$. La versión rotada de las DG se expresa como:

$$\begin{aligned} g_{n-m,m}(\omega_x, \omega_y, \theta) &= (cj\omega_x + sj\omega_y)^{n-m} (-sj\omega_x + cj\omega_y)^m g(\omega_x, \omega_y) \\ &= \sum_{k=0}^{n-m} \sum_{i=0}^m (-1)^i C_m^i C_{n-m}^k c^{m-i+k} s^{n-m+i-k} g_{i+k, n-i-k}(\omega_x, \omega_y) \end{aligned}$$

haciendo los cambios de índices $k + i$ por k y $m - i$ por i , se obtiene:

$$\begin{aligned} g_{n-m,m}(\omega_x, \omega_y, \theta) &= \sum_{k=0}^n g_{k, n-k}(\omega_x, \omega_y) c^{-k} s^k \sum_{i=0}^m (-1)^{m-i} C_m^i C_{n-m}^{k+i-m} c^{2k+2i-m} s^{n+m-2k-2i} \\ &= \sum_{k=0}^n g_{k, n-k}(\omega_x, \omega_y) c^{-k} s^k \Delta^m \{ C_{n-m}^{k-m} c^{2k-m} s^{n+m-2k} \} \\ &= \sum_{k=0}^n A_{m,k}^{(n,\theta)} g_{k, n-k}(\omega_x, \omega_y) \end{aligned}$$

o bien, en el dominio espacial,

$$G_{n-m,m}(x, y, \theta) = \sum_{k=0}^n A_{m,k}^{(n,\theta)} G_{k,n-k}(x, y) \quad (\text{D.2})$$

donde

$$A_{m,k}^{(n,\theta)} = c^{-k} s^k \Delta^m \{ C_{n-m}^{k-m} c^{2k-m} s^{n+m-2k} \} \quad (\text{D.3})$$

son funciones de ángulo que definen una familia de funciones binomiales generalizadas en la variable discreta k . La razón por la que se denominan "binomiales" es porque su transformada Z , la cual se escribe como

$$a^{(n,\theta)}(z) = (s + z^{-1}c)^{n-m} (c - z^{-1}s)^m \quad (\text{D.4})$$

pue

{c,

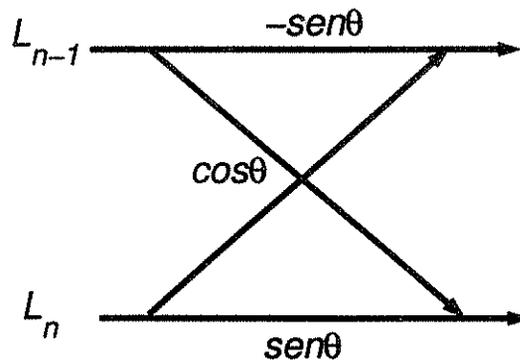


Figura D.1: Mariposa del filtro binomial generalizado.

Por lo tanto, la rotación local en el dominio de la THD puede hacerse empleando el algoritmo de Hashimoto-Sklansky [19], usando la mariposa de la Figura D.1 en lugar de la mariposa simétrica. El número de operaciones requeridas para la rotación de todos los coeficientes de orden n es tres veces el número requerido para el caso simétrico, o sea, $3[(n+1) \log_2(n+1)]$, más $n+1$ multiplicaciones para la normalización.

Bibliografía

- [1] N. Ahmed, T. Natarajan, and K. R. Rao, *Discrete cosine transform*, IEEE Trans. Comput. **C-23** (1974), 90–93.
- [2] A. J. Ahumada and R. Horng, *Smoothing DCT compression artifacts*, SID Digest **25** (1994), 708–711.
- [3] A. J. Ahumada, R. Horng, B. Rogowitz, and J. Allebach, *De-blocking DCT compressed images*, Human Vision, Visual Processing, and Digital Display V, SPIE Proc. **2179** (1994), 109–116.
- [4] H. B. Barlow, *Handbook of sensory physiology*, L. Hurvich and D. Jameson, New York, 1972.
- [5] M. Bastiaans, *Gabor's signal expansion and degrees of freedom of a signal*, Optic Acta **29** (1982), 1223–1229.
- [6] M. Basu and L. M. Kennedy, *An experiment with gaussian derivatives for image enhancement*, Systems, Man and Cybernetics **4** (1995), 3778 –3783.
- [7] T. Berger, *Rate distortion theory*, Engelwood Cliffs, NJ, 1971.
- [8] R. E. Blahut, *Principles and practice of information theory*, Addison-Wesley Reading, NA, 1987.

- [9] J. A. Bloom and T. R. Reed, *A gaussian derivative-based transform*, IEEE Transactions on Image Processing **5** (1996), no. 3, 551–553.
- [10] S. J. Campanella and G. S. Robinson, *A comparison of orthogonal transformations for digital speech processing*, IEEE Trans. Commun. Technol. (1971), 1045–1050.
- [11] Cheng-Tie Chen, *Adaptive transform coding via quadtree-based variable block size DCT*, Proc. ICASSP (1989), 1854–1857.
- [12] T. N. Cornsweet, *Visual perception*, Academic Press, New York, 1970.
- [13] B. Escalante-Ramírez and J. B. Martens, *Noise reduction in computerized tomography images by means of polynomial transforms*, Visual Communication and Image Representation **3** (1992), 272–285.
- [14] B. Escalante-Ramírez and J. L. Silván-Cárdenas, *Optical-flow estimation by means of local projection analysis with the Radón-Hermite transform*, Mathematical Modeling, Bayesian Estimation, and Inverse Problems (Denver, CO) (F. Petreux, A. Mohammad-Djatari, and E. Dougherty, eds.), vol. 3816, SPIE, 1999, pp. 121–129.
- [15] ———, *A multiresolution directional-oriented image transform based on gaussian derivatives*, Wavelets: Applications in Signal and Image Processing IX (San Diego, CA) (Laine, Unser, and Aldroubi, eds.), vol. 4478, SPIE, 2001.
- [16] W. T. Freeman and E. H. Adelson, *The design and use of steerable filters*, IEEE Transactions on Pattern Analysis and Machine Intelligence **13** (1991), no. 9, 891–906.

- [17] R. E. Frye and R. S. Ledley, *Derivative of gaussian functions as receptive field models for disparity sensitive neurons of the visual cortex*, Biomedical Engineering Conference (1996), 270–273, Proceedings of the 1996 Fifteenth Southern , 1996.
- [18] R. A. Haddad and A. N. Akansu, *A new orthogonal transform for signal coding*, IEEE Transactions on Acoustics, Speech and Signal Processing **36** (1988), no. 9, 1404–1411.
- [19] M. Hashimoto and J. Sklansky, *Multiple-order derivatives for detecting local image characteristics*, Computer Vision, Graphics and Image Processing **39** (1987), 28–55.
- [20] W. R. Hendee and P. N. T. Wells, *The perception of visual information*, Springer-Verlag New York Inc., NY, 1993.
- [21] R. Horng and A. J. Ahumada, *A fast DCT block smoothing algorithm*, Visual Communication and Image Processing Annual Meeting Proceedings, SPIE **2501** (1995), no. 5.
- [22] F. O. Huck, *Information theory analysis of sensor array imaging system for computer vision*, Proceeding of the SPIE **397** (1983), 82.
- [23] D. A. Huffman, *A method for the construction of minimum redundancy codes*, Proc. IRE **40** (1952), 1098–1101.
- [24] A. K. Jain, *Image data compression: A review*, Proc. IEE **69** (1981), no. 3, 349–389.
- [25] J. J. Koenderink and A. J. Van Doorn, *Generic neighborhood operators*, IEEE Transactions on Pattern Analysis and Machine Intelligence **14** (1992), no. 6, 597–605.

- [26] H. P. Kramer and M. V. Mathews, *A linear coding for transmitting a set of correlated signals*, IRE Trans. Inform. Theory **IT-2** (1956), 42–46.
- [27] M. Kunt, A. Ikonomopoulos, and Mokoher, *Second-generation image-coding techniques*, Proc. IEEE **73** (1985), no. 4, 549–574.
- [28] G. E. Legge, *A power law for contrast discrimination*, Vision Research **21** (1981), 457–467.
- [29] G. E. Legge and J. M. Foley, *Contrast masking in human vision*, Journal of the Optical Society of America **70** (1980), no. 12, 1458–1471.
- [30] Y. Linde, A. Buzo, and R. M. Gray, *An algorithm for vector quantization design*, IEEE Transactions on Communications **Com-28** (1980), 84–95.
- [31] T. Lindenbergh, *Sacale-space theory in computer vision*, Kluwer, Boston, MA., 1994.
- [32] Hongche Liu, Tsai-Hong Hong, M. Herman, and R. Chellappa, *A generalized motion model for estimating optical flow using 3-d hermite polynomials*, Pattern Recognition **1** (1994), 361–366.
- [33] S. P. Lloyd, *Least square quantization in PCM's*, Tech. report, Bell Telephone Laboratories, Murray Hill, NJ, 1957.
- [34] J. B. Martens, *The hermite transform-applications*, IEEE Transactions on Acoustics, Speech and Signal Processing **38** (1990), no. 9, 1607–1618.
- [35] ———, *The hermite transform-theory*, IEEE Transactions on Acoustics, Speech and Signal Processing **38** (1990), no. 9, 1595–1606.
- [36] ———, *Local orientation analysis in images by means of the hermite transform*, IEEE Transactions on Image Processing **6** (1997), no. 8, 1103–1116.

- [37] M. Porat and Zeevi, *The generalized gabor scheme of image representation in biological and machine vision*, IEEE Transactions on Pattern Analysis Machine Intelligence **10** (1988), 452–467.
- [38] M. Rabbani and P. W. Jones, *Digital image compression techniques*, SPIE Optical Engineering Pres., WA, 1991.
- [39] F. Ratliff, *Mach bands: Quantitative studies on neural networks in the retina*, Holden Day, Oakland, 1965.
- [40] J. Rissanen and G. G. Langdon, *Arithmetic coding*, IBM J. Res. Dev. **23** (1979), no. 2, 149–162.
- [41] G. Robinson and R. L. Granger, *FFT speech compression*, Proc. IEEE Int. Conf. Commun. **26** (June 1970), no. 5.
- [42] J. Ross, C. Morrone, and D. Burr, *The conditions under which mach bands are visible*, Vision Research **6** (1989), 699–715.
- [43] C. A. Shaffer and H. Samet, *Optimal quadtree construction algorithms*, Comput. Vis., Graph., Image Process **37** (1987), 402–419.
- [44] M. Shizawa and T. Iso, *Direct representation and detecting of multi-scale, multi-orientation fields using local differentiation filters*, IEEE Computer Society Conference on Computer Vision and Pattern Recognition (1993), 508–514, Proceedings CVPR '93, 1993.
- [45] E. Shusterman and M. Feder, *Image compression via improved quadtree decomposition algorithms*, IEEE Trans. on Image Processing **3** (1994), 207–215.
- [46] J. L. Silván-Cárdenas, M. P. Ángeles-Meza, and B. Escalante-Ramírez, *Radon-Hermite analysis applied to image coding*, Proceedings of the IEEE International

- Symposium on Industrial Electronics (Slovenia), vol. 3, ISIE, 1999, pp. 1204–1207.
- [47] J. L. Silván-Cárdenas and B. Escalante-Ramírez, *Application of the polynomial transforms to motion estimation*, Mathematical Modeling and Estimation Techniques in Computer Vision (San Diego, CA) (F. Petreux, J. L. Davison, and E. Dougherty, eds.), vol. 3457, SPIE, 1998, pp. 216–227.
- [48] ———, *Quadtree segmentation based on the modified discrete Hermite transform*, Proceedings of the International Conference on Telecommunications (Acapulco, Gro.), ICT-2000, 2000, pp. 1158–1161.
- [49] ———, *Image coding with a directional-oriented Hermite transform on a hexagonal lattice*, Applications of Digital Image Processing XXIV (San Diego, CA) (Tescher, ed.), vol. 4472, SPIE, 2001.
- [50] ———, *A multi-scale, multi-orientation image transform*, 5th World Multiconference on Systemics, Cybernetics and Informatics (Orlando, FL) (International Institute of Informatics and Systemics, eds.), SCI, 2001.
- [51] A. Singh, *Optic flow computation: A unified perspective*, IEEE Computer society Press., Los Alamitos, CA, 1991.
- [52] G. J. Sullivan and R. L. Baker, *Efficient quadtree coding of images and video*, IEEE Trans. Image Processing **3** (1994), 327–331.
- [53] G. Szegő, *Orthogonal polynomials*, Colloquium Publications, American Mathematical Society, 1959.
- [54] H. Tan and K. Yao, *Evaluation of rate distortion functions for a class of independent identically distributed sources under an absolute magnitude criterion*, IEEE Transactions on Information Theory **IT-21** (1975), 59–63.

- [55] J. Vaisey and A. Gersho, *Image compression with variable block size segmentation*, IEEE Trans. on Signal Processing **40** (1992), 2040–2060.
- [56] A. B. Watson, *The cortex transform: Rapid computation of simulated neural images*, Computer Vision, Graphics and Image Processing **39** (1987), 311–327.
- [57] ———, *DCT quantization matrices visually optimized for individual images*, Proceedings (Bellingham, WA) (SPIE, ed.), SPIE, 1993, pp. 202–216.
- [58] ———, *DCTune: A technique for visual optimization of DCT quantization matrices for individual images*, Society for Information Display Digest of Technical Papers XXIV (1993), 946–949.
- [59] H. R. Wilson and J. R. Bergen, *A four mechanism model for spatial vision*, Vision Research **19** (1979), 19–32.
- [60] A. Witkin, *Scale-space filtering: A new approach to multiscale description*, Image Understanding **3** (1984), 79–95.
- [61] R. A. Young, *The gaussian derivative theory of spatial vision: Analysis of cortical receptive field line-weighting profiles*, Tech. Report GRM-4920, Gen. Motors Res, 1985.
- [62] ———, *Simulation of human retinal function with the gaussian derivative model*, Proceedings IEEE CCVP (Miami, FL.), 1986, pp. 564–569.
- [63] ———, *The gaussian derivative model for spatial vision: I. retinal mechanisms*, Spatial Vision **2** (1987), 273–293.
- [64] D. Zion, S. Wang, and J. Vaillancourt, *Depth from defocus using the hermite transform*, International Conference on Image Processing (1998), 958–962, Proceedings ICIP98, vol. 2.