

01170
S



**UNIVERSIDAD NACIONAL AUTONOMA
DE MEXICO**

DIVISION DE ESTUDIOS DE POSGRADO
FACULTAD DE INGENIERIA

"COMPRESION PERCEPTIVA DE IMAGENES DIGITALES
UTILIZANDO MAPAS DE BRILLANTEZ, EN UN ESQUEMA
DE CUANTIZACION VECTORIAL POR TRANSFORMADA"

T E S I S

QUE PARA OBTENER EL TITULO DE
MAESTRO EN INGENIERIA

P R E S E N T A
MAURICIO MORENO GUTIERREZ



DIRECTOR DE TESIS: DR. BORIS ESCALANTE RAMIREZ

MEXICO, D.F.

OCTUBRE DE 2001

ESTA TESIS NO SALE
DE LA BIBLIOTECA



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.

A mi madre, Cecilia, por ser mi guía espiritual y por conjuntar el amor, la bondad, la sensibilidad y la sapiencia en un solo ser. Gracias por ser partícipe del día a día de mi vida. Te debo lo que soy y lo que seré. Siempre.

A mi padre, José, por ser mi ejemplo, mi maestro y mi consejo; por tener siempre el valor y la confianza de enfrentar la vida con atingencia y sabiduría. Gracias por ser, al mismo tiempo, el labrador de mis surcos y la brújula de mis derroteros. Tu cálida luz es la fortaleza de mi espíritu.

A mi hermano, Rodrigo, ejemplo de entereza, inteligencia, dedicación y solidaridad, de quien tanto aprendo y quien tanto me motiva. Gracias por estar siempre allí, donde te necesito, y por seguirme acompañando a lo largo del impredecible enramaje de la vida.

A mi novia, Samari, fusión inigualable de amor, belleza y bondad. Gracias por ser mi motivo, mi vía y mi meta. Brindo, hoy y siempre, por la cantidad de vivencias que compartiremos juntos. Ya lo sabes: tú me completas.

Agradecimientos

Al Dr. Boris Escalante Ramírez, por su dedicación, su apoyo y sabio consejo en la construcción de éste y de otros proyectos. Su dirección fue siempre atinada y útil.

A mis sinodales, Dr. Víctor García Garduño, Dr. Miguel Moctezuma Flores, Dr. Bohumil Psenicka, y Dr. Carlos Rivera Rivera, por sus acertadas intervenciones en la maquila de esta tesis y por las fructíferas horas compartidas en los salones de clase.

A todos mis profesores, tanto de licenciatura como de maestría, por su empeño y compromiso en la realización diaria de una vocación tan noble y valiosa, como es la docencia.

A la UNAM, generosa hacedora de hombres íntegros, estudiosos, inteligentes y comprometidos, cuna de ideales y feliz culpable de todo esto. La gratitud es inmensa para con ella. A su Facultad de Ingeniería, sin duda la mejor del país.

A CONACYT, por el soporte económico proporcionado para llevar a cabo tanto mis estudios de maestría, como éste trabajo de tesis.

ÍNDICE GENERAL

| | |
|--|------------|
| Índice General | i |
| Índice de Figuras | iii |
| Índice de Tablas | vii |
| I. Introducción | 1 |
| II. Percepción de Imágenes en el Sistema de Visión Humana | 4 |
| 2.1 La Imagen Digital | 4 |
| 2.1.1 Compresión de Señales | |
| 2.2 El Sistema de Visión Humana | 10 |
| 2.2.1 Medidas de Rendimiento en la Compresión de Señales | |
| 2.3 Codificación Perceptiva | 18 |
| 2.3.1 Señales Audiovisuales en la Percepción Humana | |
| 2.3.2 Modelos del Sistema Perceptivo Humano | |
| 2.3.3 Codificación Perceptiva de Imágenes | |
| III. Cuantización Vectorial | 26 |
| 3.1 Introducción a la Cuantización Vectorial | 26 |
| 3.2 Cuantización Escalar | 29 |
| 3.3 Cuantizadores Vectoriales sin Memoria | 32 |
| 3.3.1 Cuantización | |
| 3.3.2 Distorsión | |
| 3.3.3 Distorsión Media, Ergodicidad y Estacionariedad | |
| 3.3.4 Optimalidad | |
| 3.3.5 Algoritmo de Lloyd Generalizado | |
| 3.3.6 Tipos de Cuantizador Vectorial sin Memoria | |
| 3.4 Cuantización Vectorial con Memoria | 41 |
| 3.5 Cuantización Vectorial en Imágenes | 41 |
| 3.5.1 Cuantización Vectorial Adaptiva | |
| 3.5.2 Estructuras Quadtree | |
| IV. El Mapeo Luminancia – Brillantez | 47 |
| 4.1 Percepción de la Brillantez | 47 |

| | |
|---|------------|
| 4.2 Modelos de Brillantez | 50 |
| 4.3 Generación del Mapa de Brillantez | 54 |
| 4.4 La Aplicación del Mapeo al Mundo Bidimensional | 62 |
| 4.5 Algunos Experimentos con Mapas de Brillantez | 67 |
| V. La Transformada de Hermite | 73 |
| 5.1 Introducción a la Transformada de Hermite | 74 |
| 5.2 Transformadas Polinomiales Unidimensionales | 75 |
| 5.3 Transformada de Hermite Unidimensional | 78 |
| 5.4 Transformada de Hermite Bidimensional | 82 |
| 5.5 Transformada Discreta de Hermite | 83 |
| 5.6 La Rotación en la Transformada de Hermite | 85 |
| 5.7 La Clasificación de Estructuras con la Transformada de Hermite | 88 |
| VI. Descripción del Sistema de Compresión | 90 |
| 6.1 Vista General del Sistema | 90 |
| 6.2 El Codificador | 92 |
| 6.2.1 El Proceso de Transformación, Rotación y Clasificación Hermitiano | |
| 6.2.2 Cuantización Vectorial | |
| 6.2.3 La Escritura del Archivo | |
| 6.3 El Decodificador | 106 |
| 6.4 La Interfaz de Usuario | 108 |
| VII. Resultados y Conclusiones | 111 |
| 7.1 Diferencias Estructurales en Campos de Brillantez | 111 |
| 7.2 Decodificación de Imágenes y Calidad de Reproducción | 114 |
| 7.3 La Calidad de la Imagen en Función del Tamaño y Constitución del Libro de Códigos | 123 |
| 7.4 Baja Distorsión entre Imágenes Codificadas en Brillantez y su Mapa Correspondiente | 129 |
| 7.5 Posibles Mejoras y/o Modificaciones al Trabajo | 134 |
| 7.6 Conclusiones | 135 |
| VIII. Referencias Bibliográficas | 140 |

ÍNDICE DE FIGURAS

Capítulo 2

| | |
|---|----|
| 2.1 Pintura rupestre en Lascaux (izq); <i>Mujer</i> , de Joan Miró (der) | 4 |
| 2.2 El proceso básico de obtención de una imagen digital | 6 |
| 2.3 Estructura General del Ojo Humano | 11 |
| 2.4 Distribución de Conos y Bastones en la Retina con Respecto al Ángulo Foveal ... | 12 |
| 2.5 Cociente de Weber en Función de la Intensidad | 14 |
| 2.6 Amplitud de la sonorización en la palabra “sal” | 21 |
| 2.7 Sensibilidad al Contraste en el HVS | 23 |

Capítulo 3

| | |
|--|----|
| 3.1 Cuantizador escalar uniforme de 3 bits sobre una función de densidad laplaciana | 31 |
| 3.2 Regiones de Voronoi | 33 |
| 3.3 Funcionamiento de un codificador VQ | 36 |
| 3.4 Esquema de un algoritmo LGB básico | 38 |
| 3.5 Cuantizador vectorial sencillo para imágenes | 42 |
| 3.6 Ejemplo de una estructura quadtree básica en una imagen digital | 46 |

Capítulo 4

| | |
|---|----|
| 4.1 El fenómeno de inducción de la brillantez | 49 |
| 4.2 Proceso del mapeo luminancia-brillantez | 53 |
| 4.3 Forma de los filtros estructurales: central y periférico | 55 |
| 4.4 Imagen original “House” arriba. Debajo, dos señales escaladas: la de la izquierda corresponde al índice 3 de resolución, y la de la derecha al índice 7 (penúltimo) en resolución | 63 |

| | |
|--|----|
| 4.5 Mapa de brillantez de la imagen "House" | 66 |
| 4.6 Comparación de los perfiles de la ilusión óptica de la barra, en luminancia (en medio) y en brillantez (abajo). En la ilusión se muestra la cota a la que se llevó a cabo el corte del perfil | 67 |
| 4.7 Comparación de los perfiles de la ilusión óptica de los círculos, en luminancia (en medio) y en brillantez (abajo). En la ilusión se muestran la cotas a las que se llevaron a cabo ambos perfiles | 69 |

Capítulo 5

| | |
|--|----|
| 5.1 Estructura orientada unidimensional (borde) y textura (estructura 2D) | 73 |
| 5.2 Esquema de transformada y antitransformada polinomiales | 78 |
| 5.3 Funciones de Hermite con $\sigma=1$ | 81 |
| 5.4 Imagen con fuerte orientación unidimensional. La matriz de la izquierda contiene los coeficientes transformados sin rotar, la de la derecha rotados y concentrados | 87 |

Capítulo 6

| | |
|--|----|
| 6.1 Esquema tradicional de codificación – decodificación | 90 |
| 6.2 Esquema en bloques del proceso de codificación, desde la entrada de la imagen hasta la escritura del archivo comprimido en el medio de almacenamiento | 92 |
| 6.3 Muestreo del látice, simplificado usando sólo dos "juegos" de rejillas: los nones y los pares, que generarán, en conjunto, estructuras de tipo hexagonal – véase la línea falsa (punteada) | 94 |
| 6.4 La imagen de la izquierda es producto de una decodificación con estructura hexagonal. En la de la derecha fue aplicada la tradicional estructura cuadrangular de bloques adyacentes. Note la diferencia en calidad visual. Ambas imágenes fueron severamente cuantizadas | 94 |
| 6.5 Sub-proceso de transformación, rotación y determinación de estructura local | 97 |
| 6.6 Histograma comparativo de la energía en los coeficientes de dos secuencias de entrenamiento: en rojo, la secuencia basada en luminancias. En azul, la secuencia basada en brillantez. Nótese la mayor dispersión en la gráfica roja | 98 |
| 6.7 La imagen original (izquierda) es transformada, rotada y clasificada. En las tres imágenes de la derecha se sobreponen pequeños círculos de color en las zonas estructurales correspondientes | |

| | |
|--|-----|
| (0D figura de arriba, 1D en medio, 2D figura de abajo) | 99 |
| 6.8 Cuantizador escalar uniforme de 8 niveles para ángulos entre 0 y π [rad] | 102 |
| 6.9 Proceso de entrenamiento y de generación de los libros de código | 103 |
| 6.10 Estructura de la palabra en el archivo codificado. En la parte de abajo se hallan los tamaños, en bits, de cada sección. Nótese que el tamaño de las secciones “etiquetas” dependen del tamaño de los libros de código utilizados | 105 |
| 6.11 Esquema general del decodificador | 106 |
| 6.12 Vista general de la interfaz de usuario en funcionamiento | 108 |
| 6.13 Interfaz para decodificación y despliegue de archivos de imagen BVQ | 109 |
| 6.14 Vista general de la interfaz de entrenamiento para libros de código | 110 |

Capítulo 7

| | |
|---|-----|
| 7.1 Diferencias en la clasificación de patrones. Arriba, clasificación de la imagen original “Venice” en luminancias. Abajo, clasificación del mapa de brillantez de “Venice”. Los puntos rojos corresponden a estructuras 0-D, los azules a 1-D y los verdes a 2-D | 112 |
| 7.2 Obsérvese el mismo fenómeno que en la figura pasada, en esta ocasión para un detalle de la imagen “Goldhill”, tanto en luminancia (arriba) como en mapa brillantez (abajo) | 113 |
| 7.3 a) Imagen “Mountain” original (arriba), b) imagen comprimida con método en luminancias (abajo). Tasa de bits de 0.6 bpp, c) Imagen comprimida con método en cruz (arriba), d) imagen JPEG de misma compresión (abajo) | 115 |
| 7.4 a) Imagen “Clown” original (arriba), imagen comprimida con método en luminancias (abajo), b) imagen comprimida por método en cruz (arriba), imagen JPEG de misma compresión (abajo) | 118 |
| 7.5 a) Imagen “Boat” original (arriba), imagen comprimida con método en luminancias (abajo), b) imagen comprimida por método en cruz (arriba), imagen JPEG de misma compresión (abajo) | 121 |
| 7.6 La imagen “Baboon” fue comprimida utilizando libros de código de varias dimensiones, comenzando con uno de 128 elementos hasta uno de 4096 | 126 |
| 7.7 Confrontación de radio de compresión y SNR para las distintas imágenes “Baboon” obtenidas en la figura anterior | 127 |
| 7.8 Comparación del rendimiento bits por píxel – SNR para los métodos de luminancia, cruz y el estándar JPEG en la imagen “Lena”, de 256x256 pixeles .. | 128 |

| | |
|--|-----|
| 7.9 Imagen "Peppers" (arriba, izquierda) y su mapa de brillantez ecualizado para su representación visual (arriba, derecha). Sus histogramas se muestran debajo de ellos | 129 |
| 7.10 1) Imagen MIT con histograma, 2) reconstrucción del original (método en cruz), con tasa de 0.5 bpp, 3) mapa de brillantez de MIT con histograma, 4) reconstrucción, en brillantez, del mapa | 132 |

ÍNDICE DE TABLAS

Capítulo 6

| | |
|--|-----|
| 6.1 Ejemplos de concentración de energía en bloques transformados 0D, 1D y 2D | 100 |
|--|-----|

Capítulo 7

| | |
|---|-----|
| 7.1 Datos de la compresión de la imagen "Mount" | 117 |
| 7.2 Datos de la compresión de la imagen "Clown" | 120 |
| 7.3 Datos de la compresión de la imagen "Boat" | 120 |
| 7.4 Comparación de SNR para diversos experimentos de codificación | 131 |

Capítulo I

Introducción

El presente trabajo corresponde a una investigación realizada en la División de Estudios de Posgrado de la Facultad de Ingeniería de la Universidad Nacional Autónoma de México, dirigida y asesorada por el Dr. Boris Escalante Ramírez. El autor la propone como tesis para obtener el grado de Maestro en Ingeniería, en el área de Ingeniería Eléctrica y dentro del grupo de estudio del Procesamiento Digital de Señales.

Como se puede advertir en el título del trabajo, el campo de acción del mismo se halla limitado al ámbito de las señales bidimensionales (imágenes, particularmente) y tendrá como ejes principales la codificación y compresión de la señal vía cuantización vectorial, la transformada polinomial discreta de Hermite y el empleo de mapas de brillantez como herramienta perceptiva. A lo largo del estudio se ahondará en los temas mencionados y se planteará una propuesta para satisfacer los objetivos y el punto partida de la tesis, mismos que procedo a explicar.

El objetivo principal de esta tesis es el de diseñar un modelo de codificación y compresión de imágenes digitales valiéndose de esquemas de cuantización vectorial que actúen en dominios de brillantez con criterios de distancia/distorsión que tengan correspondencia con aspectos psico-visuales, es decir, que cuantifiquen la distorsión entre dos vectores (de imágenes digitales) o que sirvan de base para establecer un sistema de codificación, basándose en las características perceptivas principales del sistema de visión humana y en su respuesta a los estímulos luminosos. La codificación de la energía en el dominio de la brillantez, la determinación estructural local, el entrenamiento de los libros de código y el análisis de orientaciones locales en las imágenes se hará con la ayuda de la transformada de Hermite.

La tesis está dividida en capítulos temáticos. Después de esta breve introducción, se expondrán las bases teóricas de la imagen digital, su estudio como señal, algunos ejemplos de procesamiento y sus objetivos, y principalmente la importancia que encierra el diseño de algoritmos de compresión. Se relatarán brevemente los estándares

y métodos de compresión de imágenes más populares y las distintas vías por las que se han indagado soluciones a este proceso. Asimismo, dentro del mismo capítulo II, se explicarán, de forma sencilla, las características esenciales del sistema de visión humana, sus ventajas y desventajas y se tratarán algunos aspectos relacionados con el comportamiento de la visión ante distintos estímulos oculares.

Posteriormente, en el tercer capítulo, se hablará de un importante procedimiento de análisis de señales digitales: la cuantización vectorial. Habrá que remitirse a los fundamentos matemáticos del método y se enfocará su desempeño al tratamiento de imágenes. Se percibirán tanto la importancia como el alcance de la cuantización vectorial y se abrirá brecha para introducir su estudio en diversos dominios que son de particular interés. De este capítulo se derivará la propuesta de diseño de un cuantizador vectorial que se integrará al modelo de compresión de imágenes que esta tesis pretende estructurar, explicado ampliamente en el capítulo VI.

Como cuarto capítulo, se revisarán los trabajos más relevantes en materia de modelos de brillantez. Se tratará el fundamento fisiológico de estas investigaciones, sus alcances y consecuencias, y se evaluará la construcción de los llamados “mapas de brillantez” o “mapas luminancia-brillantez” (cuantificaciones de la percepción humana en la reflectancia de los objetos visibles), mismos que han llamado poderosamente la atención al interior del grupo de estudiosos de la imagen desde hace muchos años.

Se intentará, asimismo, ensamblar un modelo adecuado de brillantez (bidimensional) que fungirá como aliado en la búsqueda de un esquema robusto de compresión que trabaje, principalmente, en este dominio.

En este punto del trabajo (capítulo V) se hará un estudio exhaustivo de la transformada de Hermite dentro del marco de las transformadas polinomiales y se derivará brevemente su estructura discreta para ser aplicada a este proyecto. Se hablará de las ventajas que residen en su utilización sobre otras transformadas que gozan de mayor popularidad y se explicará el porqué es considerada una poderosa herramienta para sintetizar y clasificar imágenes con patrones, unidimensionales y bidimensionales, fuertemente orientados. Este aspecto resultará de gran utilidad al llevar a cabo la construcción del codificador, ya que la transformada polinomial hermitiana discreta será la encargada de decidir qué tipo de cuantización vectorial (o escalar) aplicar y

cómo codificar la energía contenida en la imagen eficientemente. Además, será el elemento definidor de las clases estructurales de la imagen. Se hablará, además, del látice sobre el que será aplicada la transformada y el porqué de la elección del mismo.

El sexto capítulo fusionará las propuestas derivadas del estudio de los cuatro principales temas (representados en los capítulos anteriores) para proponer un modelo-sistema de compresión de imágenes que se valga de cuantizadores vectoriales en el dominio transformado de la brillantez, con objeto de determinar las estructuras importantes, y determinando la distribución de la energía en el dominio Hermitiano. Se explicarán sus principales elementos constitutivos, tanto en codificación como en decodificación, se recapitulará el procedimiento que derivó en su desarrollo y se presentará la propuesta de una medida alterna de distorsión que haya sido elaborada con los resultados obtenidos en la investigación. Dicha construcción será sometida a prueba en el siguiente capítulo (VII), donde se evaluarán las condiciones del modelo global, su capacidad de compresión, la calidad psico-visual de sus productos y su efectividad. Se presentarán distintos exámenes aplicados al compresor y se mostrarán ejemplos de su desempeño en algunos tipos de imágenes digitales.

Finalmente, este séptimo y último capítulo contendrá las conclusiones a las que se llegó después de el estudio, diseño y evaluación de la investigación, se tratarán posibles mejoras futuras y se propondrán alternativas por las que puede encaminarse la energía de trabajos posteriores.

Las referencias bibliográficas (libros, publicaciones periódicas, publicaciones locales y artículos) utilizadas en esta tesis se ordenarán para su rápida revisión en la última sección del trabajo, así como los correspondientes índices de figuras y tablas.

Capítulo II

Percepción de Imágenes en el Sistema de Visión Humana

2.1 La Imagen Digital

La fascinación por las imágenes ha acompañado al hombre desde que éste existe en la faz de la tierra, en su afán de construir descripciones vívidas y gráficas de cualquier situación, cosa o pensamiento. El procesamiento digital de imágenes (la manipulación de imágenes a través de sistemas de cómputo) es un tópico reciente, si lo comparamos con la antiquísima inquietud del ser humano por el estímulo visual. Desde las cromáticas y frías paredes de Lascaux y Altamira hasta la genialidad, casi incomprensible, de Miró, la imagen pictórica ha acompañado al hombre a través de varios siglos como sinónimo de testimonio y arte, y como agente de comunicación y de permanencia (figura 2.1).

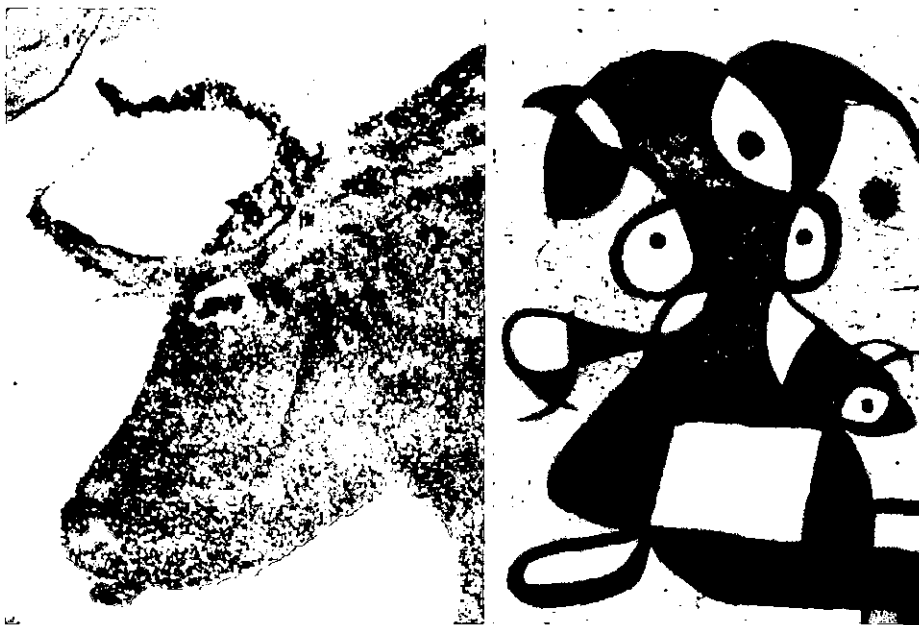


Figura 2.1 Pintura rupestre en Lascaux (izq); *Mujer*, de Joan Miró (der)

La imagen no ha sido únicamente limitada al lienzo; la escultura en sí puede ser considerada una imagen; vívida y sólida, distinta, pero imagen al fin. La aparición y evolución de la fotografía a finales del siglo XIX resultó un avance sin precedentes en la añeja historia de la imagen, porque representaba por fin el anhelo cumplido de atrapar indefinidamente la luz de una escena en un medio físico. Esta disciplina artística ha seguido evolucionando tremendamente en su técnica, en su teoría y en sus recursos tecnológicos. Las imágenes en movimiento fueron al fin domadas, como la fotografía, en el mismo siglo de la ilustración: el cine, propuesto al mundo por los hermanos Lumière, llegó para quedarse a formar parte del mundo visual, en un principio, y años después del audiovisual. La aparición a mediados del siglo XX de la televisión sació las ansias del hombre de enviar, a través del aire, escenas a lugares remotos y a enormes distancias. De pronto, a través de un extraño aparato infestado de cables y tubos, se podía ser testigo de acontecimientos que estaban ocurriendo muy lejos, pero en ese mismo instante; o de producciones masivas dedicadas al entretenimiento o a la información. El frenético avance de la electrónica en la segunda mitad del siglo XX permitió al ser humano diseñar y construir dispositivos cada vez más rápidos, baratos, eficientes y pequeños que llevaran a cabo complicadas tareas de cálculo. Así nació la computación electrónica: el hombre se ha valido de la física en su escala más ínfima para edificar pequeños sistemas que cuenten, sumen y memoricen eficaz y velozmente por él. Al mismo tiempo que se hacía más evidente la penetración del cómputo en el mundo moderno, se integraron centenas de disciplinas a su firme paso. Una de ellas es la de la imagen digital, que se ha visto muy favorecida por la exitosa aparición de equipos de procesamiento cada vez más poderosos y por el esfuerzo en investigación que a esta área tecnológica se le dedica en todo el orbe. El procesamiento digital de imágenes se ha constituido ya en la actualidad, como una disciplina formal que engloba múltiples aspectos de óptica, matemáticas, fotografía, tecnología computacional, algorítmica y electrónica.

Debido a que las computadoras trabajan con información numérica, no pictórica, la imagen digital encuentra de esta manera sus diferencias con la analógica¹. Para obtener una imagen digital a partir de una imagen analógica, o de una escena del mundo real, hay que

¹ Estrictamente, todas las imágenes que vemos son analógicas, porque las captamos a través de nuestra vista, que así funciona, pero la diferencia reside en la forma en que estén constituidas.

llevar a cabo una tarea, llamada *digitalización*, que consta fundamentalmente dos pasos: *muestreo* y *cuantización*. La digitalización es el proceso para convertir una imagen de su forma original a una representación digital. El término “convertir” es incluido en la definición como no destructivo, ya que la imagen original no es destruida, sino utilizada como guía para dicha conversión (fig. 2.2).

Muestreo, en el caso de imágenes digitales, es llevar a cabo la medición del nivel de luminancia, o de gris, en ciertas localidades de la imagen, igualmente distribuidas a lo largo del espacio bidimensional. Generalmente se lleva a cabo con un dispositivo sensor de imagen que produce un voltaje proporcional a la intensidad luminosa presente en la escena (Shapiro, 1992). Cuantización es la representación del valor medido con un número entero. Los sensores de muestreo de imágenes están generalmente seguidos de convertidores analógico-digital (ADC), un circuito electrónico que genera un número proporcional al voltaje de entrada que reciben. Los productos discretos básicos de este “enrejado”, provocado al muestrear la imagen, se llaman *pixeles*². Cada uno de ellos es el encargado de almacenar el valor cuantizado correspondiente a la región espacial que representa. Citemos ahora, una definición formal de imagen digital (Castleman, 1996): “Una imagen digital es una función bidimensional muestreada y cuantizada que ha sido generada por medios ópticos; muestreada en un arreglo o red rectangular igualmente espaciada, y cuantizada a intervalos de amplitud iguales”.

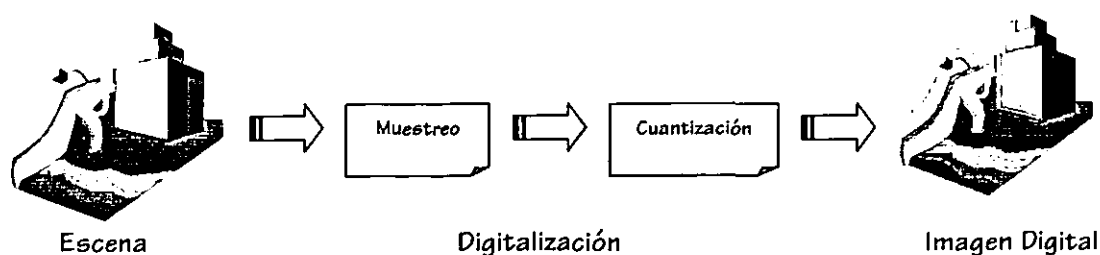


Figura 2.2: El proceso básico de obtención de una imagen digital

Una vez que estos pasos han concluido, se puede llevar a cabo cualquier tipo de procesamiento digital con dicha imagen. Ésta se ha convertido ahora en un simple arreglo matricial de valores enteros, correspondientes a la intensidad luminosa de la escena. Esto abre un panorama inmenso al tratamiento de la información, ya que se puede aplicar a dicha

² Abreviatura, en inglés, de “picture elements”

matriz un sinnúmero de transformaciones que produzcan *otra* imagen digital con características distintas. Algunas de estas operaciones, por mencionar ejemplos sencillos, (Castleman, 1996) son:

- Ecuación del histograma: Se adecua, a través de una transformación lineal, la distribución de niveles de intensidad³ en una imagen digital para repartirlos de forma más equitativa y lograr así un resultado de mejor calidad visual. Se usa también para fines de codificación.
- Operaciones puntuales: Son aquellas que llevan de una imagen de entrada a una de salida de manera que el valor de *cada* píxel de salida depende únicamente de la transformación aplicada a su contraparte de entrada. Incluyen tareas como calibración fotométrica y calibración de dispositivos de despliegue.
- Operaciones algebraicas: Producen una imagen de salida que resulta de la suma, diferencia, producto o cociente de dos o más imágenes de entrada. Aquí se incluyen aplicaciones importantes como adición y reducción de ruido, detección de movimiento simple, etc.
- Operaciones geométricas: Cambian las relaciones espaciales entre los objetos que pueblan una imagen. Se conocen también como transformaciones espaciales; reglas que definen el movimiento de cada píxel al trasladarlo de una posición a otra dentro de la imagen. Incluyen aplicaciones como la interpolación, magnificación, reducción, deformación, movimiento, rotación, rectificación y proyección cartográfica de escenas.

Sin duda, el acervo más rico de investigación en imágenes digitales no se halla ligado únicamente a los tópicos mencionados. La teoría de los sistemas lineales ha encajado perfectamente con el procesamiento digital de la imagen al verla no solo como un motivo de transformación, sino como una *señal*.

³ Un histograma es la gráfica que enfrenta niveles de intensidad de una imagen contra cantidad de píxeles que se hallan en cada nivel

2.1.1 Compresión de Señales

El mundo de las señales digitales es amplio y complejo y se ha constituido como una de las principales (y más activas) áreas de la ingeniería moderna.

Como actividad, el procesamiento digital de señales generalmente crea grandes cantidades de información, y la necesidad de transmitirla o almacenarla ha obligado a encauzar esfuerzos hacia la obtención de técnicas de compresión cada vez más eficientes. La compresión de datos es la acción de llevar un mapeo de un conjunto de datos a una cadena de bits, con el fin único de disminuir el número de bits requeridos para representar al conjunto de datos dado (Cosman, 1993). Dicho en otras palabras, estamos reduciendo la cantidad de datos en una señal, tratando de perder lo mínimo (o nada) de información. La clave consiste en eliminar la redundancia o los datos considerados irrelevantes. Los dos tipos de compresión que hay son los siguientes:

1. Compresión sin pérdidas (*lossless*): Es aquella en la que a partir de la señal comprimida se puede reconstruir *exactamente* la señal original. Se intenta eliminar redundancia entrópica. Los algoritmos de compresión de datos sin pérdidas caen en dos grandes categorías: las técnicas basadas en diccionario y los métodos estadísticos. Las técnicas basadas en diccionario generan un archivo comprimido conteniendo códigos de longitud revisada (de usualmente 12 a 16 bits), cada uno de los cuales representa una secuencia particular de bytes en el archivo original. Un ejemplo que cae dentro de esta clasificación es la codificación por longitud de la corrida (RLE). Los métodos estadísticos implementan la compresión de datos representando caracteres frecuentemente ocurrentes en el archivo con menos bits de los que realmente tienen. Este es el tipo de análisis que Samuel Morse utilizó cuando definió el código internacional del telégrafo. El código de Huffman (Huffman, 1952), introducido en 1952, es un método estadístico sin pérdidas que siempre encuentra un código de longitud variable con la menor redundancia posible. Usa un árbol de codificación binaria para representar los valores que tienen mayor ocurrencia en pocos bits y los valores con menor ocurrencia con más bits.
2. Compresión con pérdidas (*lossy*): Ocurre cuando es imposible regresar a la señal original a partir de la comprimida, misma que ha sufrido cambios irreversibles en el proceso de

compresión. En esta técnica, información considerada “innecesaria” es eliminada. Técnicas comunes de compresión con pérdidas son las de codificación por transformada combinadas con cuantizaciones escalares o vectoriales. Es sabido que una de las aplicaciones más utilizadas de las transformadas, o transformaciones discretas, es la compresión de imágenes. Combinadas con otras técnicas, permiten la transmisión, almacenamiento y despliegue de imágenes y secuencias de video que de otra forma serían poco prácticas. Supongamos que tenemos un grupo de imágenes que serán codificadas en una representación compacta de datos. Podemos transformar las imágenes (lo que implica trasladar sus valores a otro dominio más propicio para la compactación), descartar los coeficientes que sean cercanos a cero, y cuantizar severamente aquellos que son pequeños, para concentrar nuestra transmisión de datos y nuestros recursos de almacenamiento solo en los coeficientes que contienen mas información sobre la imagen. Cuando la imagen sea reconstruida más tarde, muy poco contenido importante se habrá perdido. Esta técnica explicada a grandes rasgos se llama codificación de imágenes por transformada .

En varios poderosos sistemas de codificación de imágenes, los dos tipos de compresión mencionados son mezclados para obtener un radio de compresión más alto. Se lleva a cabo, por ejemplo, una codificación por transformada y posteriormente se hace un análisis entrópico para optimizar la distribución de los signos a transmitir.

Las aplicaciones de la compresión de señales son enormes. A ella debemos que existan eficientes sistemas de tratamiento de imágenes, voz, audio, video y muchas otras. Podemos mencionar algunos ejemplos:

- Compresión de voz en telefonía digital
- Correo de voz
- Video-telefonía
- Telefonía celular
- Audio y video-conferencias
- Transmisión de medios (audio, video e imágenes) vía internet
- Almacenamiento y procesamiento de imágenes biomédicas, geológicas, astronómicas y científicas en general

-
- Distribución de música, revisión ágil de acervos musicales
 - Transmisión facsímil de alta fidelidad
 - Películas en CD
 - Televisión digital
 - HDTV

La estandarización de una tecnología es vital para la intercomunicación de aplicaciones y equipo, y el mundo de la imagen digital no es la excepción. La industria digital está basada en una tecnología hasta cierto punto nueva, pero que nació inmersa en el contexto de grandes industrias con tecnologías razonablemente maduras. La imagen digital liga y acerca a varios campos y con ellos a sus estándares; ejemplos tan disímolos entre sí son la fotografía, el diseño gráfico, el arte visual, la astronomía, la geología, el despliegue de imágenes en dispositivos, la industria de la impresión y la edición, el entretenimiento, etc. En cuanto a compresión, exitosos estándares desarrollados en la década pasada han contribuido fuertemente al acuerdo entre usuarios, aplicaciones y hardware. Por mencionar algunos: CCITT⁴ (viejo estándar con varias modificaciones para la transmisión de facsímil vía telefónica), JBIG⁵ (estándar para codificación progresiva de imágenes binarias), JPEG⁶ (famoso estándar de compresión de imágenes, mundialmente utilizado en sus múltiples versiones), TIFF, GIF (Graphics Interchange Format), etc.

2.2 El Sistema de Visión Humana

El ser humano recibe los estímulos que la naturaleza y su entorno le envían a través de los sentidos, mismos que envían las señales al cerebro para su proceso, evaluación y reconocimiento. De los cinco sentidos, la vista ocupa un lugar preponderante al ser el encargado de interpretar la luz, de recibirla y convertirla en percepción. Como sabemos, nuestros ojos son capaces de captar solo una pequeña banda del amplio espectro de frecuencias en las radiaciones electromagnéticas, la llamada “espectro visible”, que va de

⁴ Sus siglas corresponden a: “Consultative Committee on Telephone and Telegraph”

⁵ Sus siglas corresponden a: “Joint Bi-level Image Experts Group”

⁶ Sus siglas corresponden a: “Joint Pictures Experts Group”

los 7.5×10^{14} [Hz] a los 4.3×10^{14} [Hz], y que se halla justo por arriba (en unidades frecuenciales) de las ondas infrarrojas y debajo de la zona ultravioleta (Zemansky, 1988).

El estudio de la visión humana es indudablemente un tema que encierra muchos secretos al investigador de hoy. Su relación con múltiples áreas de la ciencia lo hacen un tópico por demás interesante y rico, como lo evidencia el tremendo volumen bibliográfico dedicado a este tema. Las complicadas tareas que lleva a cabo el sentido de la vista resultan difíciles de entender, aunque para nosotros el acto de ver objetos parezca un proceso sencillo, limpio y libre de esfuerzo. En esta tesis nos referiremos a algunas características del sistema de visión humana (HVS⁷, por sus siglas en inglés) que pueden ser amalgamadas con aspectos del procesamiento digital de imágenes.

Como sugiere Zetzsche (Zetsche, 1989), los aspectos esenciales del procesamiento visual de imágenes fijas en un observador humano son: la adaptividad a la iluminación de la escena, la descomposición escalar y el enmascarado de patrones específicos.

El ojo humano es casi esférico, con un diámetro medio aproximado de 20 mm. Está rodeado por tres membranas: la córnea y la esclerótica, que constituyen la cubierta exterior, la coroides y la retina.

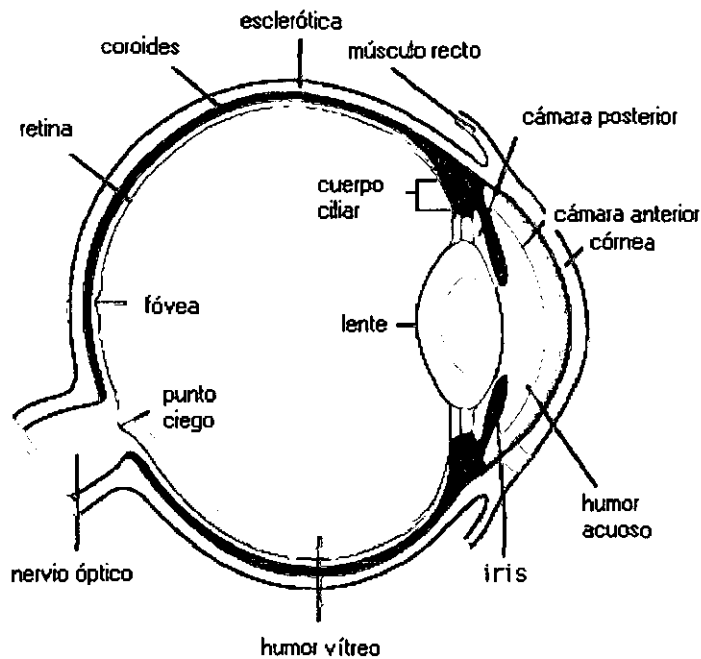


Figura 2.3: Estructura general del ojo humano

⁷ Human Visual System

La córnea es un tejido resistente y transparente que cubre la superficie anterior del ojo. En prolongación de la córnea, la esclerótica es una membrana opaca que encierra el resto del globo ocular (figura 2.3).

La coroides, debajo de la esclerótica, contiene una red de venas que constituyen la “alimentación” del ojo, y en su extremo anterior está dividida en cuerpo ciliar y diafragma o iris. El iris es el que se abre o cierra para controlar la cantidad de luz que entra al ojo. El cristalino, justo debajo de la córnea, absorbe aproximadamente el 8% del espectro luminoso visible, con una absorción ligeramente superior en las longitudes de onda más cortas. Tanto la luz infrarroja como la ultravioleta son absorbidas de forma apreciable, por lo que en cantidades excesivas pueden dañar al ojo.

La membrana más interna del ojo es la retina, que recubre la totalidad de la pared posterior.

Resulta de suma importancia conocer, al menos superficialmente, el funcionamiento de la retina, que es la primera estructura neuronal involucrada en la percepción visual y que ha sido ampliamente estudiado por neurobiólogos y por neurólogos. La retina es una extensión del cerebro que cubre la zona posterior del ojo y que recibe las imágenes de los objetos.

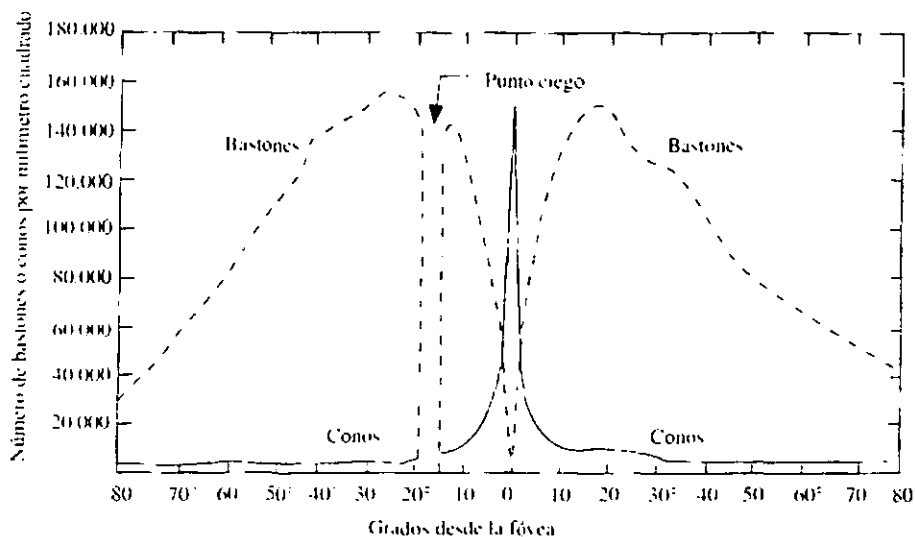


Figura 2.4: Distribución de Conos y Bastones en la Retina con Respecto al Ángulo Foveal

De forma más precisa, podemos explicar a la retina como una delicada membrana formada por finas capas regulares de células especializadas en distintas tareas (Szentaghotai, 1974). Los dos grupos de células más importantes son los *conos* y los *bastones*. Estas unidades fisiológicas actúan como transductores: convierten la energía luminosa incidente en energía electro-química. Contrario a lo que podría pensarse, la distribución de estas células a lo largo de la retina no es uniforme; se puede observar en la figura anterior (2.4) que la mayor concentración de conos se halla en el eje visual (excentricidad cero), área a la que se llama *fovea* (Heucke, 2000).

La fovea está delimitada por una depresión con un diámetro de aproximadamente 1.5 mm (equivalentes a 5.2° de ángulo visual) en la superficie de la retina. En el centro de la fovea, cubriendo un área de 0.3 mm de diámetro (aprox. 1° de ángulo visual), no hay nada más que conos; los bastones comienzan a aparecer paulatinamente al alejarnos de esta singular área. Al aumentar la excentricidad, los conos disminuyen de forma dramática, mientras que la densidad de bastones aumenta progresivamente y alcanza su máximo alrededor de los 20° de excentricidad.

Como es de esperarse, los conos y los bastones están destinados a tareas distintas. Aparentemente, los conos están involucrados en la visión diurna y los bastones en la nocturna: sus canales asociados son muy sensibles a formas, mientras que los conos se encargan de distinguir detalles. Estos últimos tienen un tiempo de respuesta corto, a diferencia de los bastones (Dinet, 1997).

La capacidad del ojo humano de discriminar entre cambios de iluminación para cada nivel específico de adaptación⁸ también es de considerable interés. Un experimento clásico utilizado para determinar la capacidad del sistema visual humano de discriminar la iluminación, como el descrito por Schouten (Schouten, 1993), consiste en colocar a un sujeto observando un área plana, uniformemente iluminada, lo suficientemente grande para que ocupe todo el campo visual. Esta área es habitualmente un simple difusor, tal como un vidrio esmerilado, que se ilumina desde atrás con una fuente de luz, cuya intensidad, I , puede variarse. A este campo se añade un incremento de iluminación, ΔI , en forma de un destello de corta duración que aparece como un círculo en el centro del campo

⁸ El HSV no puede operar en todo el rango dinámico de luminosidad simultáneamente, por lo que lleva a cabo cambios en la sensibilidad global. A este fenómeno se le conoce como adaptación a la iluminación.

uniformemente iluminado. Si el ΔI no es lo suficientemente brillante, el sujeto deberá decir “No” implicando que no percibe el cambio. Conforme el ΔI aumenta, el sujeto puede dar una respuesta positiva “Sí” indicando que ha percibido un cambio. Finalmente, cuando el ΔI sea suficientemente intenso acabará diciendo “Sí” todas las veces. La cantidad

$$\frac{\Delta I_C}{I} \quad (2.1)$$

donde ΔI_C es el incremento de iluminación discernible el 50% de las veces con la iluminación de fondo I , se conoce como coeficiente de Weber (González, 1992). Un pequeño valor en este cociente indica que se puede discriminar un pequeño cambio de intensidad. Esto representa una buena discriminación de iluminación. De forma inversa, un valor alto en esa razón significa que se necesita un gran cambio de intensidad (pobre discriminación de la iluminación).

Una representación logarítmica del cociente de Weber en función de la intensidad se muestra en la siguiente figura (2.5). Se observa que la discriminación de la iluminación es pobre para niveles de iluminación bajos y mejora significativamente conforme crece la iluminación de fondo. Las dos ramas de la curva reflejan el hecho de que para niveles de iluminación bajos la visión se realiza a través de la actividad de los bastones, mientras que a niveles elevados (mostrando mejor discriminación) la visión es función de los conos.

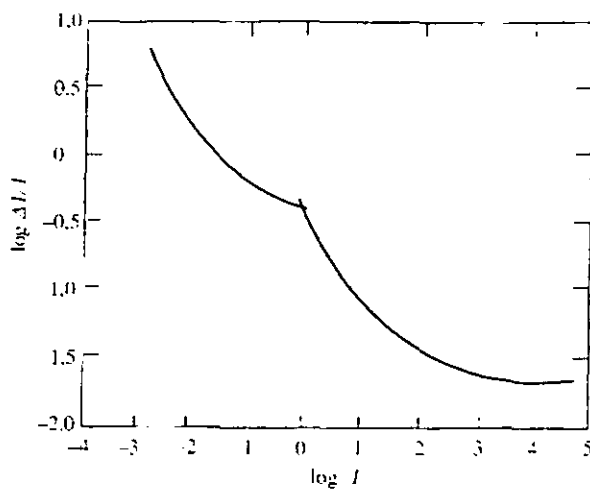


Figura 2.5 Cociente de Weber en Función de la Intensidad

Algunos métodos de procesamiento digital de imágenes orientados a objetos (Heucke, 2000) modelan la fisiología de la retina con tres círculos concéntricos, que cumplen con distintos niveles de percepción de brillantez. Ya se hablará con detalle de este importante concepto en capítulos posteriores. Dichos modelos plantean que el círculo interior se encargue de representar la fovea central, con su correspondiente importancia al fijar detalles, el siguiente círculo contendrá el *fondo* del objeto y el círculo más externo el fondo de los alrededores. Hay interesantes consecuencias de la utilización de esquemas como éste. La estructura concéntrica presentada en tres niveles puede ser aprovechada, por ejemplo, en imágenes de bajo ancho de banda, para fines de segmentación.

El poder resolutivo del ojo (es decir, su habilidad de separar señales espacialmente o temporalmente), depende fuertemente de la calidad óptica de la imagen proveída por el sistema ocular y del potencial anatómico y fisiológico de la retina de tratar dicha imagen. Consecuentemente, el tamaño de los receptores en el arreglo retinal es un parámetro importante que tiene un efecto definitivo en el poder de resolución del ojo humano, que puede ser considerado en términos de agudeza visual. La agudeza visual es una medida de la capacidad de sistema visual de detectar detalles pequeños; convencionalmente se define como una función inversa del ángulo mínimo de resolución expresado en minutos de arco (Dinet, 1997). La máxima agudeza visual se halla concentrada en el área foveal, donde la densidad de conos es la más alta. Después de esta zona, la agudeza desciende dramáticamente para finalmente caer más lentamente en los extremos. Podemos concluir que la región de la retina que aloja a la imagen juega un papel importantísimo en la calidad de la agudeza visual, por lo que las áreas periféricas otorgan bajos rendimientos; y que los elementos involucrados en el funcionamiento correcto de esta medida son los conos. Nuestros ojos no pueden analizar una escena completa con el mismo detalle, tendremos que “barrerla” con movimientos oculares hasta generar en nuestro cerebro un mapa fiel de la misma, con el detalle que hayamos deseado dedicarle. La visión periférica se limita a determinar cuáles áreas del espacio ocular podrían ser de importancia, para después fijar la sección de interés en el ángulo visual del centro de la retina (fovea), y así observar sus detalles.

El conducto de comunicación entre la zona retinal y nuestro cerebro es el nervio óptico, que manda la información electro-química al tálamo, quien lo distribuye al córtex

visual primario (llamado por varios científicos V1). En este punto se analizan, primeramente, aspectos simples de la escena visual. Posteriormente, la información es proyectada hacia las áreas corticales que se hallan alrededor del córtex visual primario, y se lleva a cabo un análisis más profundo y elaborado de la imagen. Aquí es en donde se realizan las funciones cognitivas más complicadas. Al parecer, los monos son los animales que poseen un sistema visual-cerebral más parecido al nuestro, por lo que son los elementos favoritos para la experimentación en este nivel. En el córtex las señales son transferidas vía sinapis a otras células nerviosas, por lo que si se hace un “sembrado” de electrodos en distintas zonas del mismo, se pueden determinar las áreas físicas de procesamientos visuales muy específicos que el cerebro usa para reconstruir la realidad. Según Richard Andersen (Andersen, 1998), el cerebro usa cinco estrategias básicas en su procesamiento visual, a saber: codificación poblacional (significa que una percepción entera es almacenada con la ayuda de muchísimas unidades pequeñas), localización funcional (se refiere a que diferentes partes del córtex se especializan en procesos de visión particulares, pero que mantienen su comunicación): se han hallado ya 35 distintas áreas corticales con diferentes tareas en la visión de los monos, por lo que se presume que los seres humanos tengamos aún más. La tercera estrategia es el procesamiento paralelo: el cerebro divide la vasta información visual y la procesa en paralelo a lo largo de líneas separadas, esto para que un posterior procesamiento jerárquico transfiera los datos de cierto nivel a otro más complicado a lo largo que se avanza dentro del sistema. La última de las estrategias es la asociación (una vez que se desmenuzó la imagen y se analizó en niveles de trabajo paralelos y jerárquicos, se tiene que recombinar para que ésta sea encadenada en una sola percepción).

De todas estas características visuales se desprenden interesantes hipótesis a lo largo de la bibliografía, con el objeto de explotarlas para generar algoritmos, propuestas e investigaciones en el campo de la computación visual. La que esta tesis pretende presentar es una de ellas. A continuación se hablará brevemente de la compresión de señales y se ligarán las cualidades del sistema de visión humana con los paradigmas de diseño de medidas de percepción adecuadas y con la teoría de la codificación.

2.2.1 Medidas de rendimiento en la compresión de señales

Como sabemos, la batalla en la compresión de señales consiste en mantener una tasa de bits baja a cambio de una calidad de reproducción aceptable. Si deseamos cuantificar el rendimiento de un compresor, podríamos atenernos, digamos, a los siguientes parámetros:

- **Calidad de la señal:** Algunos autores definen un número pequeño de *grados* subjetivos de calidad, asociados con un adjetivo (excelente, buena, regular, etc.) que serán aplicados al resultado final por una gran cantidad de usuarios con el fin de “calificar” la señal decodificada (Wang, 1992), o bien, se diseña la evaluación de la señal en función de su perceptibilidad⁹. En procesos de evaluación de señales de voz es habitual que la medida se haga atendiendo a la *inteligibilidad* de la trama. Evidentemente, el sentido de subjetividad emerge de manera importante en estos parámetros de rendimiento.
- **Tasa de bits:** Sus dimensiones son el número de bits entre la unidad de contexto en la que la señal es expresada, como bits por píxel (imágenes), bits por segundo (tasa de transmisión, velocidad), bits por muestra (más general), etc. El Teorema de Nyquist nos recuerda que hay que mantener la frecuencia de muestreo de la señal al menos al doble de la frecuencia fundamental (o ancho de banda) original (Oppenheim, 1989). Existen numerosos estándares que restringen la tasa de bits de un sistema de codificación como parámetro principal a seguir.
- **Complejidad:** Llamamos complejidad de un algoritmo de compresión al esfuerzo computacional empleado en codificar y decodificar la señal, una medida ampliamente utilizada es el tiempo de procesamiento (o número de instrucciones requeridas) y la memoria utilizada en el dispositivo encargado de realizarla, aunque hay otros criterios dignos de mención, sobre todo en sistemas portables, como el costo, el consumo de energía y el tamaño físico del dispositivo. El vertiginoso avance en la tecnología de los DSP hace posible que millones de instrucciones se lleven a cabo en un segundo en un solo dispositivo, pero la complejidad de la tarea que realicen siempre será una variable que desearemos controlar muy de cerca.

⁹ Facilidad con la que ésta es percibida por los sentidos.

- Retrasos en comunicación: Un algoritmo que resulte significativamente complejo demandará mayor tiempo en su ejecución, tanto en la codificación como en la decodificación. Dicha carga computacional acarreará retrasos inherentes a la señal en el canal de transmisión. Este parámetro es muy importante sobre todo en sistemas de comunicación bidireccional (telefonía, videoconferencia, videofonía, etc.).

Dependiendo de la aplicación que se evalúe, será natural que algunos de estos parámetros tendrán más importancia que otros.

2.3 Codificación perceptiva

La compresión de señales generalmente conlleva pérdidas irreversibles en la información que éstas portan. Sabemos que, inevitablemente, en el desempeño de un sistema codificador, su calidad será inversamente proporcional a la tasa de compresión que alcance. Al hablar de señales digitales podemos mencionar algunos ejemplos de las más comunes: audio, voz, imágenes y video (imágenes en movimiento). Los criterios de calidad con los que se evalúa una señal comprimida son muy diversos, pero dado que el ser humano, a través de sus sentidos, es el receptor y destino de dichas señales, la evaluación más efectiva siempre será la subjetiva. Muchos esfuerzos en el campo de la investigación siguen encaminando al diseño de criterios de calidad que sean lo más fieles posible al sistema sensorial humano. Esto implica complicaciones tremendas, ya que un individuo reaccionará, prácticamente en todos los casos, distinto que otro al comparar y evaluar estímulos procedentes de señales. Existen algoritmos clásicos para calcular medidas de distorsión, como el MSE¹⁰, que son rápidos y sencillos, pero que han demostrado funcionar pobremente en situaciones en las que la peculiar percepción humana reacciona de forma inesperada.

Dicha medida de distorsión es un elemento fundamental en el diseño de un compresor de señales, y su diseño, que deberá derivarse de un delicado estudio de la

¹⁰ Error mínimo cuadrático, las siglas del inglés Minimum Square Error.

percepción sensorial, resulta digno de análisis. Es por eso que se ha dedicado parte de este capítulo a la revisión de las características básicas del sistema de visión humana, ya que las señales con las que trabajaremos en esta tesis serán imágenes fijas. En las líneas siguientes tocaremos algunos tópicos relacionados con el modelado de la percepción humana, el concepto de distorsión y el procesamiento digital de señales (de forma más precisa, imágenes).

Como sabemos, el objetivo principal de la compresión de señales es lograr una tasa baja de bits al mismo tiempo que se alcanza una pérdida *percibida* mínima en el destino final de la señal. En un caso deseable, el ruido (o distorsión) que es inevitablemente agregado en el proceso de codificación, puede ser “enmascarado” de forma parcial o total por la señal misma. Dicho enmascaramiento puede definirse de la siguiente forma: el ser humano, debido a sus características sensoriales, es incapaz, en algunas condiciones, de detectar componentes de señal y de ruido en la misma localidad temporal, espectral o espacial. El objetivo de la compresión perceptiva es detectar y explotar esas condiciones. Hay una gran ventaja inmersa en este concepto: la percepción de ruido puede llegar a ser nula si la SNR¹¹ local es baja (Jayant, 1993). Si se logra detectar que el ruido presenta una distribución tal que se enmascara en todas, o casi todas las localidades de la señal, la reconstrucción de la misma será perfecta o cuasi-perfecta.

Resulta sumamente interesante revisar algunos sencillos resultados experimentales obtenidos al poner a prueba los umbrales mínimos de percepción sensorial. Algunos curiosos ejemplos de percepción apenas distinguible (JNP) son (INT-1):

- Vista: La flama de una vela a una distancia de 45 [km], en una noche absolutamente oscura y despejada.
- Oído: El “tic-tac” de un reloj a 4 [m] de distancia en condiciones absolutamente silenciosas.
- Gusto: El sabor de azúcar al disolver una cucharada pequeña en 7 [l] de agua
- Aroma: Una gota de perfume concentrado dispersando su olor en tres recámaras contiguas.
- Tacto: El ala de una mosca cayendo en la mejilla, a una distancia de 1 [cm].

¹¹ Relación señal a ruido, de sus siglas en inglés (signal to noise ratio)

El concepto de “Distorsión Apenas Perceptible”, o JND por sus siglas en inglés¹², define el rango perceptivo que va desde el nulo hasta la sensación mínima de distorsión en los sentidos. Es decir, el límite define a una señal perfecta a la tasa mínima posible de bits. Ligado a este concepto aparece el de entropía perceptiva, asociada a la cantidad de información contenida en los datos tratados perceptivamente. La codificación perceptiva es importante para todas las señales, pero lo es aun más para aquellas que carecen de un buen modelo de fuente. De hecho, el paradigma de la codificación perceptiva es el de manipular el ruido en componentes de distorsión que sean perceptivamente indetectables (Jayant, 1993). El problema que presentan los algoritmos tradicionales basados en este paradigma es que no son localmente óptimos, sin embargo son robustos contra variaciones como la distancia de observación (en el caso de imágenes y video). Nótese que no se desea maximizar, como en los codificadores tradicionales, la relación señal a ruido.

Podemos proponer cuatro operaciones fundamentales en la codificación de señales a tasas bajas de bits: el prefiltrado (adecua la señal a las restricciones intrínsecas del codificador), la reducción de la redundancia, la eliminación de la información irrelevante (generalmente producto de alguna cuantización) y el post-filtrado (proceso que realza la señal ya comprimida). En definitiva, los dos pasos intermedios son los que más nos atañen al hablar de codificación perceptiva.

Casi todas las señales que se desean comprimir en el mundo real tienen redundancia. Dicha redundancia puede ser eliminada parcialmente trabajando desde varios dominios, como pueden ser el temporal, el frecuencial y el dominio transformado. Asimismo, es típico que las señales de entrada a los algoritmos de cuantización sean secuencias de predicción de errores o coeficientes producto de alguna transformada.

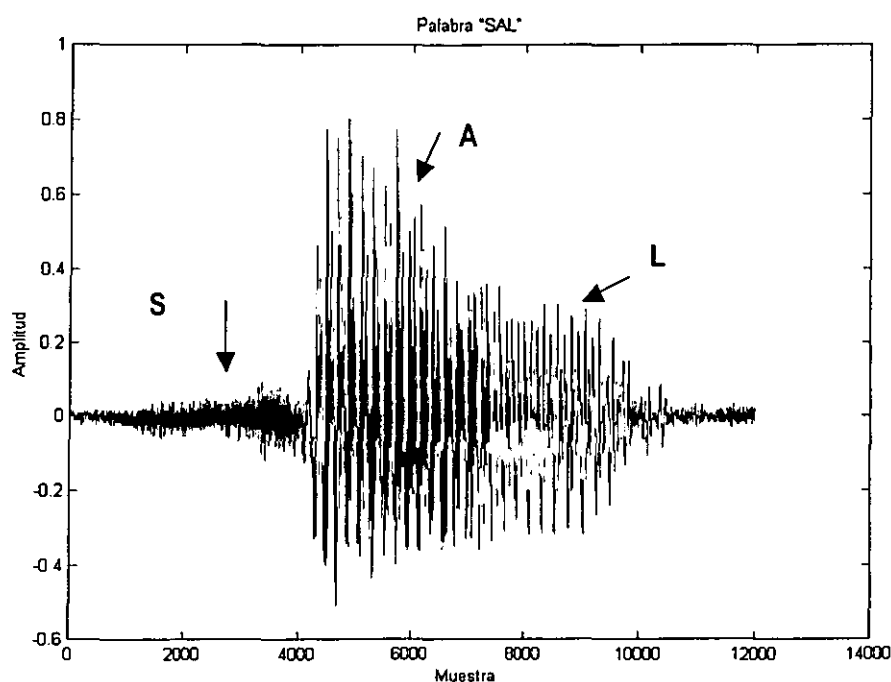
2.3.1 Señales audiovisuales en la percepción humana

Las señales audiovisuales son, de manera general, formas de onda: amplitud contra tiempo (audio y voz), intensidad contra espacio (imágenes) e intensidad contra espacio-tiempo (video). Atendiendo a su tratamiento perceptivo, y con el afán de puntualizar ciertos

¹² Just Noticeable Distortion

aspectos delicados en su naturaleza, podemos destacar cuatro importantes propiedades de dichas señales:

- 1) No estacionariedad: La no estacionariedad es común a las señales de audio, voz, video e imágenes. Por ejemplo, en el procesamiento de voz, las señales de tipo *unvoiced*, o sordas¹³, se caracterizan por una baja intensidad y un tiempo de acción corto. El sonido fricativo “s” en la palabra “sal” es bastante ilustrativo, como lo muestran la figura 2.5, donde cede su lugar a la explosión energética en la vocal “a” (sonido *voiced*, o sonoro¹⁴) para que ésta muera paulatinamente en su transición a la consonante “l”.



2.6 Amplitud de la sonorización en la palabra “sal”

En general, las señales de audio tienen sonidos sostenidos que conviven con filosos ataques y transiciones de baja energía. Algo similar ocurre con las imágenes, donde se da una simbiosis de regiones uniformes de baja frecuencia espacial y regiones texturizadas o ruidosas donde la varianza local es muy alta. La codificación perceptiva provee herramientas adicionales para manejar estas características, como se verá más adelante.

¹³ Aquellos sonidos que no hacen vibrar a las cuerdas vocales al paso del aire

¹⁴ Son los que, a diferencia de los sordos, se producen por la vibración de las cuerdas vocales (Rabiner, 1978)

- 2) Periodicidad: Hay algunas condiciones de periodicidad en señales de voz (el tono –o frecuencia fundamental- de las sonorizaciones *voiced*, o en algunos instrumentos musicales), y en imágenes (línea a línea o trama a trama en video). Tal vez la periodicidad no sea bella y exacta, pero se puede rastrear un comportamiento cuasi-periódico, a veces local, que involucre redundancia. Un resultado exitoso de esta idea es la compensación de movimiento en el video digital, tan utilizada en estándares de compresión ampliamente aceptados como el MPEG (Moreno, 1999).
- 3) Densidad Espectral de Potencia: Tópico que resulta por demás interesante e importante. En términos globales, o a largo plazo, las señales audiovisuales presentan espectros de frecuencia paso-bajas, aunque los análisis en tiempo corto (o en espacio limitado) revelan partes de ellas que pueden ser consideradas paso-banda o paso-todo (como un sonido *unvoiced* o la imagen de un tablero de ajedrez). En procesamiento de audio, sobre todo, la descripción frecuencial de la señal apoya fuertemente a la labor de eliminación de la redundancia. Para las imágenes, el conocimiento de las características en frecuencia de las mismas, suele ser de enorme utilidad, sobre todo si se combina con algún método eficiente de codificación o de cuantización.
- 4) Propiedades del color y de las señales estereofónicas: El uso de los famosos espacios de color, como YUV y YIQ¹⁵ en vez del clásico RGB (red-green-blue) en el procesamiento de imágenes es otro ejemplo de un paradigma que es a la vez eficiente estadísticamente y perceptivamente. La energía cromática de una imagen en colores es generalmente menor y posee un ancho de banda menor que la luminancia, hecho que es aprovechado y explotado en estándares de compresión para lograr muy bajas tasas de bits.

2.3.2 Modelos del sistema perceptivo humano

Una manera común y simple de modelo simple del HVS contiene un filtro paso-bajas, una no-linearidad logarítmica y un filtro paso altas multi-canal de “afilado” de la señal (Hall, 1977), pero por más complicado que se haga el modelo, no existe un

¹⁵ En ambos, como en otros no mencionados, la luminancia es representada con la letra Y, y el par de crominancias, o *chroma*, son dos componentes independientes comúnmente asociadas a las letras UV e IQ respectivamente.

codificador perceptivo óptimo, ya que ni siquiera poseemos un modelo biológico completo de los complicados fenómenos que intervienen en la percepción sensorial.

Una estrategia para construir modelos perceptivos, como ya se dijo, es aprovechando el fenómeno de enmascaramiento de la distorsión o del ruido. Formalmente, el enmascaramiento según Jayant (Jayant, 1993) es un complejo fenómeno producido en la transducción de los componentes neurológicos sensoriales. Es altamente adaptivo y se refiere a la perceptibilidad de una señal presencia de otra en su vecindad temporal, espacial o frecuencial.

Los analizadores básicos en tiempo-frecuencia en la cadena perceptiva humana son descritos por filtros paso-banda, tanto en audición como visión. Un ejemplo de este fenómeno son los filtros cocleares¹⁶, propuestos y revisados por Scharf (Scharf, 1970). Los 25 filtros críticos paso-banda van, sucesivamente, aumentando su ancho de banda conforme la frecuencia aumenta. Barren desde menos de 50 [Hz] hasta los 15.5 [KHz]. Es importante mencionar que el estudio de estas “bandas críticas” tanto en voz como (audición) y vista no son uniformes. Para incorporar esta característica al diseño de un codificador, es necesario usar modelos de enmascaramiento con un soporte frecuencial no uniforme. Observemos la siguiente gráfica, (fig. 2.7) que muestra la función de sensibilidad al contraste del sistema de visión humana (el dominio de la función es la frecuencia espacial):

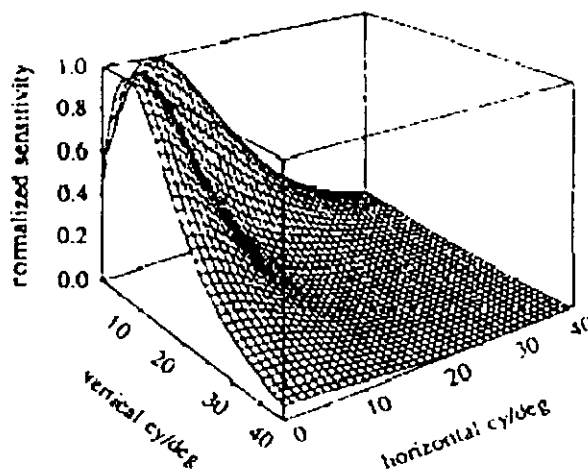


Figura 2.7 Sensibilidad al Contraste en el HVS

¹⁶ Relativos a la cóclea, en el oído humano.

Esta función describe las variaciones en sensibilidad visual con respecto a la frecuencia espacial. Obsérvese como decrece la sensibilidad al alejarse del área foveal.

2.3.3 Codificación perceptiva de imágenes

En los últimos años se han desarrollado múltiples esfuerzos, como el que en esta tesis se presenta, para vincular a la percepción visual con la compresión de imágenes digitales. La mayoría de estos trabajos se han enfocado en la sensibilidad frecuencial del sistema de visión humana, como la describe la expresión llamada “función de transferencia de la modulación” (MTF). Dicha función describe la sensibilidad del ojo humano a enrejados (retículas) de ondas senoidales en distintas frecuencias (Cornsweet, 1970). En proyectos posteriores se han intentado compatibilizar las cualidades de la MTF (que recibe entradas senoidales) y los codificadores basados en DCT (Westen, 1995), determinando un umbral de “distorsión apenas perceptiva” o JND estático para cada banda de frecuencia. Los modelos pueden sofisticarse incluyendo sensibilidad al contraste, que puede ser agregada implícitamente o desarrollando funciones de enmascaramiento contra brillantez local (que es el área en la que profundizaremos en capítulos posteriores).

Desgraciadamente, ningún modelo perceptivo utiliza a fondo las propiedades de enmascaramiento del sistema de visión humana.

Como se mencionó con antelación, la SNR no es un predictor acertado de calidad subjetiva en imágenes, especialmente en tasas altas de compresión. Una manera de atacar este problema es transformando los errores a un espacio en el que pueden usarse medidas como la MSE o similares. Westen (Westen, 1995) y Schouten (Schouten, 1993) son dos ejemplos de trabajos que transforman las imágenes a “espacios visuales” con ayuda de leyes como la de Weber, la MTF y los mapas de brillantez.

La descripción mas profunda de procedimientos en la construcción de familias de codificadores de varios tipos (basados en JND, del tipo JPEG, facsímiles, etc.) se halla en Jayant (Jayant, 1993).

En un futuro, la investigación dedicada a algoritmos de codificación perceptiva deberá tender a tornarse más robusta, más escalable y más portable. Se deberá lidiar con

señales de todo tipo, con codificadores de diversas morfologías y con señales inmersas en muchos ambientes.

Capítulo III

Cuantización Vectorial

La cuantización vectorial es el esqueleto del cual se sujetarán los elementos constituyentes de este trabajo de tesis. Es por eso que resulta de gran importancia revisar y redactar los conceptos, teoría, técnicas y aplicaciones más importantes de la CV en aras de solidificar las bases para la evolución de los siguientes capítulos. Se tratarán de forma breve y sencilla los aspectos clave en la teoría de la cuantización, escalar primeramente, y su extrapolación a vectores; su aplicación al mundo de las señales digitales, y la cabida que la cuantización vectorial ha hallado en el procesamiento de la imagen digital.

Se eligió la C.V. en esta tesis como estructura de codificación por tener la versatilidad de acomodar los elementos principales de estudio: el mapeo de la luminancia a la brillantez con campos visuales receptivos, y el dominio transformado polinomial, mismo que por su estructura de trabajo en lártice se presta para ser ubicado de esta forma. Las opciones al interior de la cuantización vectorial son infinitas. Sin embargo, y como se explicará en el capítulo 6, se optó por adecuar, probar y modificar varios métodos relativamente convencionales para lograr la consecución de los objetivos y determinar el mejor de los caminos.

3.1 Introducción a la Cuantización Vectorial

Como se apuntó en el capítulo pasado, la compresión de señales es un tópicó de suma importancia en el desenvolvimiento de la tecnología actual, debido al gran impacto que el procesamiento digital de señales en general ha causado en muchas áreas del desarrollo ingenieril. Recordemos que la compresión de señales busca, o se mueve fundamentalmente alrededor de tres objetivos:

1. La minimización de la capacidad requerida (en términos de comunicación) para transmitir señales de alta calidad, o en su defecto, obtener la fidelidad más alta posible con un canal de comunicaciones digitales dado.
2. La minimización de la capacidad de almacenamiento requerida para guardar dicha información en medios de rápido acceso y en bases de datos, o de forma equivalente, obtener la mejor calidad posible y la mayor cantidad de datos almacenados en un medio dado.
3. Obtener la representación más simple y precisa posible de una señal, así como buscar la minimización en la complejidad de algoritmos de procesamiento digital de señales, como son, por mencionar algunos: la clasificación, la transformación y el encriptamiento.

Resulta de vital importancia la formalización de ciertos términos con los que se trabajará gran parte de esta tesis. A continuación se mencionan importantes definiciones tradicionales que servirán de cimiento a la explicación de la teoría y las aplicaciones de la cuantización vectorial.

Definamos *señal* como una forma de onda con amplitud continua en tiempo continuo. La *fente* es el mecanismo o dispositivo que produce las señales. A las señales en tiempo discreto comúnmente se les llama *datos* (estrictamente se trata una secuencia de números, o vectores). Si una señal se halla acotada en tiempo discreto y amplitud discreta, se llamará *señal digital*. Un esquema de codificación específico, es decir, una regla o mapeo que especifique cómo se transformarán los símbolos de la fuente en un nuevo conjunto de símbolos se llama *código* o *sistema de codificación*.

En la literatura de la teoría de la información y las comunicaciones, la codificación de la señal es denominada *codificación de fuente*, término aportado por Shannon (Shannon, 1948) en su trabajo pionero de la teoría matemática de la comunicación. De la misma forma, la *codificación de canal*, o de control de errores, tratará de transformar la señal de forma que pueda navegar y permitir comunicación la confiable a través de canales ruidosos.

En realidad, Shannon es también el padre de la cuantización vectorial. Él propuso un sistema de codificación muy particular: un código fuente en "bloque" que mapeara segmentos consecutivos y no traslapados de la señal de entrada hacia bloques consecutivos

y no traslapados de símbolos de canal (usualmente vectores binarios). El mapeo no dependería de acciones previas del codificador o el decodificador (no tiene memoria, aunque se verá que puede haber) y sigue una regla específica. Lo que haría el decodificador es simplemente mapear símbolos (ahora binarios) en bloques o vectores-reproducciones del bloque (vector) de entrada original. Shannon supuso la existencia de una medida de distorsión que cuantificara la calidad entre un vector de entrada dado y su reproducción correspondiente. Su criterio de optimalidad era, entonces, la minimización de la distorsión promedio sujeta a una tasa constante en el código. En pocas palabras, el codificador tendría que elegir, al recibir un vector, la aproximación que más se pareciera a él (que tuviera la mínima distorsión con respecto al original) y asignarle la *palabra de código*¹ correspondiente. El nombre original que dio Shannon a este codificador fue “Código de fuente sujeto a un criterio de fidelidad”, pero poco después se popularizó el término de “Cuantización vectorial”, ya que es una generalización natural a vectores, como veremos adelante, de un cuantizador escalar simple: es el mapeo de vectores reales (un conjunto ordenado de muestras de señal) en vectores binarios utilizando una regla de mínima distorsión.

Técnicas como PCM², codificación por transformada, codificación predictiva y sus versiones adaptativas; han adquirido creciente popularidad debido a su simplicidad y a su buen desempeño. Estas técnicas comparten una propiedad fundamental: la cuantización, codificación o conversión de cantidades continuas en discretas se lleva a cabo en escalares, o sea, en muestras individuales reales de formas de onda (o píxeles de imágenes). PCM lo hace con un modelo libre de memoria, ya que cada entrada sucesiva es codificada utilizando una regla de correspondencia que no depende de ninguna entrada pasada o de ninguna salida obtenida por el codificador. La codificación por transformada lo hace tomando bloques transformados de un vector para después codificar escalarmente las coordenadas del vector transformado. La codificación predictiva cuantiza el error, que se obtiene de la diferencia entre la muestra nueva y la predicción de la misma basándose en salidas anteriormente codificadas.

¹ El índice que representa a ése y sólo a ése vector determinado.

² Siglas de “Pulse Code Modulation”; en español, “Modulación por Código de Pulsos”

Uno de los resultados más importantes del trabajo de Shannon, es la demostración de que el mejor rendimiento en un sistema de codificación siempre se obtendrá codificando vectores en vez de escalares, independientemente de si la fuente de datos tiene o no memoria (Shannon, 1948). Cualquier codificador que involucre escalares en su cuantización es, en teoría, *subóptimo*.

En el momento de su aparición, la teoría de Shannon no impactó fuertemente el desarrollo de innovaciones, debido en gran parte a que no proveía las herramientas de construcción y diseño de los vectores del código, además de que los tradicionales codificadores escalares daban buenos resultados con una adaptividad bien acoplada. El primer algoritmo que aparece en la bibliografía es el de Lloyd, a finales de los 70's, como producto de una extrapolación de un algoritmo iterativo de cuantización escalar (Lloyd, 1982). Variaciones de este algoritmo han demostrado ser de gran utilidad en la construcción de cuantizadores vectoriales con o sin memoria para una vectores de entrada de muchos tipos (formas de onda de voz, audio, imágenes y algunos otros procesos estocásticos).

3.2 Cuantización Escalar

La cuantización escalar no es un tema central de estudio en esta tesis, pero resulta de gran importancia comprender y analizar, al menos brevemente, sus fundamentos para preparar el despliegue del tópico principal, la cuantización vectorial. Además, será utilizada en algunas secciones particulares del sistema de codificación que se propondrá más adelante. Como ya se dijo, la cuantización vectorial puede ser vista como una extrapolación de la cuantización escalar, por lo que un breve apunte será expuesto en los siguientes párrafos.

La cuantización es la clave en el proceso de conversión analógico-digital. En su forma más simple, un cuantizador colecta un simple número y selecciona, dentro de un conjunto finito de valores numéricos permitidos, el valor más aproximado a él. Generalmente, el valor de entrada al proceso es *analógico* (puede tomar cualquier cantidad en un dominio continuo) y la salida es digital (siendo un entero especificado en el conjunto $\{ 1, 2, 3, \dots, N \}$, donde N es el tamaño del conjunto de valores de salida).

Formalmente (Gersho, 1992), se define un cuantizador escalar Q (unidimensional, por supuesto) de N puntos como el mapeo $Q : R \rightarrow C$ donde R es el dominio de los reales y

$$C \equiv \{y_1, y_2, y_3, \dots, y_N\} \subset R \quad (3.1)$$

es el conjunto de salida, o *libro de códigos*, de tamaño $|C| = N$. Los valores de salida, y_i , se conocen como niveles de salida, puntos de salida o valores de reproducción, dependiendo de la aplicación.

La resolución, o tasa de código, r , de un cuantizador escalar, mide el número de bits requerido para especificar únicamente, sin lugar a ambigüedad, al valor cuantizado:

$$r = \log_2 N \quad (3.2)$$

La resolución indica la exactitud con la que la amplitud analógica original queda descrita.

Hay una partición de la línea real R en N celdas o átomos R_i , asociada con cualquier cuantizador de N puntos, para $i = 1, 2, \dots, N$. La celda número i está dada por

$$R_i = \{x \in R : Q(x) = y_i\} \equiv Q^{-1}(y_i), \quad (3.3)$$

la imagen inversa de y_i bajo Q . Esta definición se complementa diciendo que $\bigcup_i R_i = R$ y $R_i \cap R_j = \emptyset$ para $i \neq j$. Una celda que carezca de límite se llama *sobrecargada*. Cada celda acotada es una celda *granular*. La siguiente figura (fig. 3.1) facilitará la comprensión del funcionamiento de un cuantizador escalar: en ella se muestra la gráfica de un cuantizador escalar que responde a una función de densidad probabilística de tipo Laplaciano. Se tomó como ejemplo este tipo de distribución por ser no lineal, no uniforme y por ser frecuentemente utilizada en el modelado de fenómenos probabilísticos³. El

³ Si construyéramos el cuantizador como *no uniforme*, podríamos buscar, por ejemplo, que el área bajo la curva de la función de densidad fuera la misma para todas las regiones de cuantización, y así garantizar una probabilidad uniforme en cada nivel del cuantizador.

cuantizador sí es uniforme, ya que la separación entre sus 8 niveles⁴ es siempre la misma. En el eje vertical aparecen valores propios de la función de densidad laplaciana, y en el eje horizontal sólo se anotan los posibles valores de representación del cuantizador. Las líneas punteadas indican las fronteras entre valores de reproducción, por lo que el dato que vaya a ser cuantizado será sometido a esa revisión, para determinar bajo qué región de cuantización se encuentra, y así ser asociado con su valor correcto de reproducción. Nótese que a mayor cantidad de valores de reproducción, se tendrá más precisión en la cuantización y habrá un menor error.

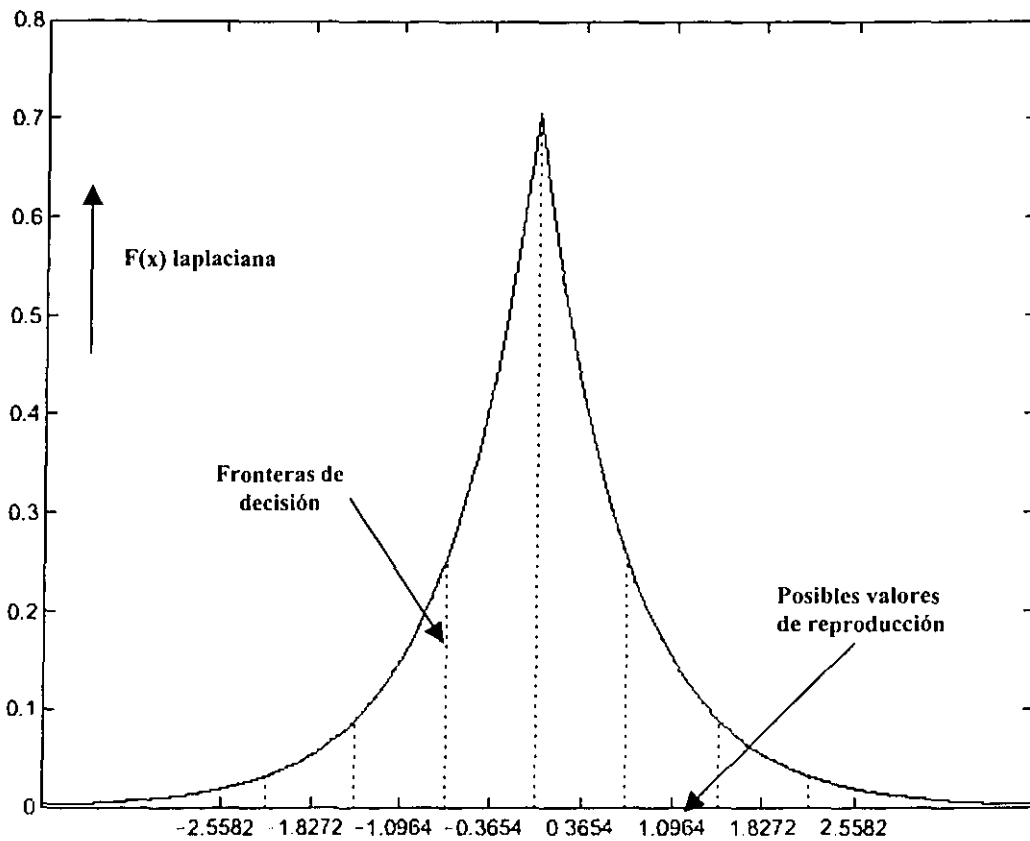


Fig. 3.1. Cuantizador escalar uniforme de 3 bits para una función de densidad laplaciana

En el contexto de un sistema de comunicaciones con formas de onda, el codificador transmite el índice i del nivel seleccionado, y_i , elegido para representar una muestra de entrada, más no el valor y_i . De esta forma, si la tasa R es un entero, se puede asignar a cada

⁴ Por lo que funciona con 3 bits solamente.

y_i una R -tupla binaria única (se codifican los y_i en vectores binarios de forma invertible). Esta R -tupla binaria puede ser transmitida o almacenada y posteriormente el decodificador reconstruiría el valor de reproducción correspondiente.

Nótese que el diseño del decodificador puede incluir una estructura de tabla de búsqueda, o libro de códigos, en el que esté el conjunto de salidas, almacenado con muy alta precisión sin afectar la tasa de transmisión R . Una secuencia de valores de reproducción obtenidos de esta forma dan una aproximación a la secuencia original de muestras y , por ende, con la ayuda de un filtro interpolador adecuado, una aproximación a la forma de onda original (Gersho, 1992).

3.3 Cuantizadores Vectoriales sin Memoria

3.3.1 Cuantización

Matemáticamente, un cuantizador vectorial k -dimensional sin memoria, o simplemente VQ, está formado por dos mapeos: un codificador γ que asigna a cada vector de entrada $\mathbf{x} = (x_0, x_1, \dots, x_{k-1})$ un símbolo de canal $\gamma(\mathbf{x})$ de algún conjunto de símbolos de canal \mathbf{M} , y un decodificador β que asigna a cada símbolo de canal v en \mathbf{M} un valor en un alfabeto de reproducciones posibles $\hat{\mathbf{A}}$. El conjunto de símbolos de canal es comúnmente visto como un espacio de vectores binarios (por conveniencia). Un caso frecuentemente utilizado para \mathbf{M} es que sea el conjunto de todos los vectores 2^R binarios R -dimensionales. El alfabeto de reproducción puede o no ser el mismo que el espacio de vectores de entrada.

Si \mathbf{M} tiene M elementos, la cantidad $R = \log_2 M$ es la tasa del cuantizador, en bits por vector, y $r = R/k$ es la razón en bits por símbolo o bits por muestra.

Al verlo como un sistema de compresión de datos, los vectores de entrada pueden ser las muestras consecutivas de una señal, vectores de parámetros, porciones de imágenes, o más aun, como se propondrá en capítulos posteriores, valores en un dominio visual transformado.

La siguiente figura (3.2), esquematiza las llamadas "regiones de Voronoi" que se forman al establecer un cuantizador vectorial sobre un dominio bidimensional. Cada

celdilla irregular tiene un punto (centroide) que es el encargado de representar a todos los elementos de ese dominio limitado.

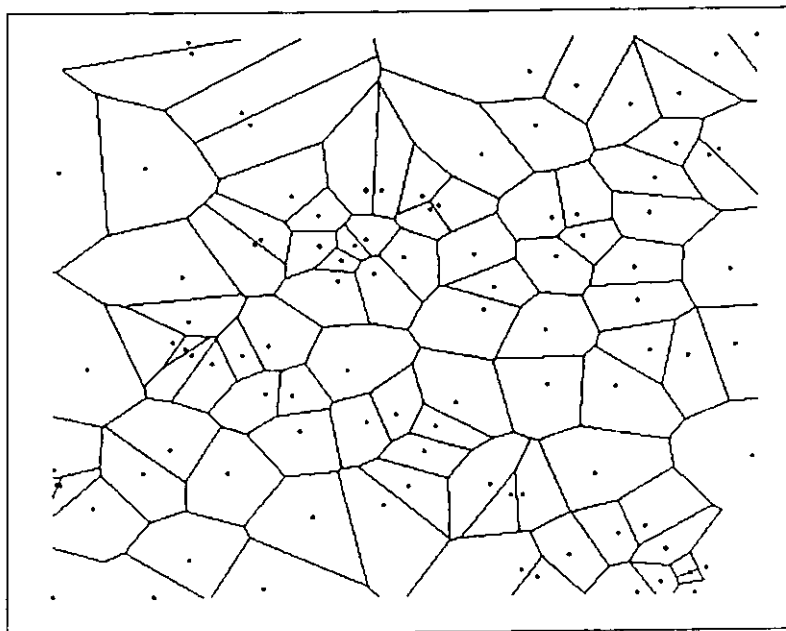


Fig. 3.2. Regiones de Voronoi

En este esquema, se supone que el canal se halla libre de ruido, es decir, que $U_n = \hat{U}_n$. Sabemos que esto en el mundo real es imposible, sin embargo no representa un obstáculo infranqueable, ya que con la incorporación de algún método de corrección de código, se puede sumergir el cuantizador en un ambiente ruidoso.

La meta de un sistema de cuantización como el expuesto es reproducir de la “mejor” forma al vector de entrada para una tasa fija R . Esta idea no estaría completa si pasamos por alto el concepto de *medida de distorsión*.

3.3.2 Distorsión

La medida de distorsión d es la asignación de un costo $d(x, \hat{x})$ de reproducir cualquier vector de entrada x como un vector de reproducción \hat{x} . Dada esta medida de distorsión, se puede cuantificar el rendimiento de un sistema al calcular la distorsión media $E d(X, \hat{X})$ entre la entrada y la reproducción final. Dicho esto, un codificador será bueno si

arroja una distorsión media pequeña. En la práctica, la medida más importante es el promedio temporal, o media de muestras a largo plazo:

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=0}^{n-1} d(X_i, \hat{X}_i) \quad (3.4)$$

mientras el límite sea calculable. Una medida de distorsión debe tener las siguientes cualidades: debe estar bien estructurada y debe ser comprensible para permitir su análisis y modificación, debe ser lo más rápidamente posible calculable, y, probablemente la más complicada; debe tener relación directa con la distorsión subjetiva entre vectores, es decir, debe ser coherente con la calidad subjetiva percibida al final del proceso. En este punto es en el que se ahondará posteriormente en esta tesis, al proponer una medida de distorsión adecuada para el caso en cuestión.

No podemos dejar de mencionar un par medidas de distorsión ampliamente utilizadas. Una de ellas es la del error cuadrático y la otra es la distancia Itakura-Saito. Para el primero de los casos, la distorsión se expresa así:

$$d(x, \hat{x}) = x - \hat{x}^2 = \sum_{i=0}^{k-1} (x_i - \hat{x}_i)^2 \quad (3.5)$$

Sus ventajas son varias: es rápida en su implementación y en su cómputo, y es simple en su funcionamiento. Sin embargo, se ha demostrado que en muchas ocasiones no va de acuerdo a la calidad subjetiva (o perceptual) de la señal decodificada en relación a la original (Li, 1997).

La segunda medida de distorsión es generalmente utilizada en trabajos que involucran procesamiento digital de voz y los vectores de reproducción se pueden ver como modelos de filtro "todo-polo" para sintetizar la reproducción en el receptor usando ruido o una fuente periódica generados localmente (Gray, 1984).

3.3.3 Distorsión media, ergodicidad y estacionariedad

Si el proceso estocástico en estudio es estacionario⁵ y ergódico⁶, el promedio temporal mencionado anteriormente en (3.4) resulta ser el mismo que el valor esperado $E\{X\}$ del proceso. Recordemos que se desea que dicho valor sea pequeño. El problema con el valor esperado es que es muy útil para obtener información sobre los límites teóricos de rendimiento, pero no es siempre calculable en la práctica. De hecho, es casi imposible determinarlo porque no se conocen las distribuciones de probabilidad requeridas en imágenes, voz, video y muchos otros tipos de señales comunes. La manera más práctica de resolver el problema es tomar secuencias muy largas de información de *entrenamiento*, estimar el valor esperado “correcto” de distorsión con la media muestral, y tratar de diseñar un código que minimice la distorsión muestral promedio para dicha secuencia de entrenamiento. Si la fuente de entrada fuera en realidad un proceso estacionario y ergódico, no habría problema, ya que la media muestral se hallaría cerca del valor esperado, y el mismo código usado en datos futuros arrojaría aproximadamente los mismos promedios (Gray, 1984). El secreto entonces, consiste en diseñar un código con la ayuda de una secuencia de entrenamiento suficientemente larga.

3.3.4 Optimalidad

Se sabe que un cuantizador vectorial será óptimo si minimiza una distorsión promedio $E d\{X, \hat{f}[Y(X)]\}$. Siguiendo la misma lógica que Lloyd (Lloyd, 1982), se pueden mencionar dos propiedades necesarias de optimalidad:

⁵ Un proceso estocástico $x(t)$ es estacionario en el *sentido estricto* (SSS) si sus propiedades estadísticas son invariantes con respecto a una traslación del origen, es decir, $x(t)$ y $x(t+c)$ tienen las mismas estadísticas para todo c .

Un proceso estocástico $x(t)$ es estacionario en el *sentido amplio* (WSS) si su media es constante, $E\{x(t)\} = \eta$ y su autocorrelación depende sólo de una diferencia de tiempos, $t = t_1 - t_2$: (Papoulis, 1991)

$$E\{x(t+\tau)x^*(t)\} = R(\tau)$$

⁶ Un proceso $x(t)$ es ergódico con respecto a su media, si su promedio temporal $\eta_T = \frac{1}{2T} \int_{-T}^T x(t) dt$ tiende al valor esperado, $\eta = E\{x(t)\}$, cuando T tiende a infinito. (Papoulis, 1991)

1) Optimización del codificador dado el decodificador.

Si se sabe que la meta del VQ es minimizar la distorsión promedio, y dado un decodificador específico β , ningún cuantizador vectorial sin memoria puede dar mejores resultados que aquél que selecciona la palabra de código v en M que dará la mínima distorsión posible a la salida, esto es, tomar el símbolo de "canal" v que arroje el mínimo en

$$d\{x, \beta[y(x)]\} = \min_{v \in M} d[x, \beta(v)] = \min_{y \in C} d(x, y) \quad (3.5)$$

visto de otra forma, para un decodificador dado dentro de un esquema VQ, el mejor codificador posible es aquel que obtenga la menor distorsión posible (el vecino más cercano), dada por:

$$\gamma(x) = \min_{v \in M}^{-1} d[x, \beta(v)] \quad (3.6)$$

donde la notación del inverso significa que hemos seleccionado la v dado el mínimo en (3.5).

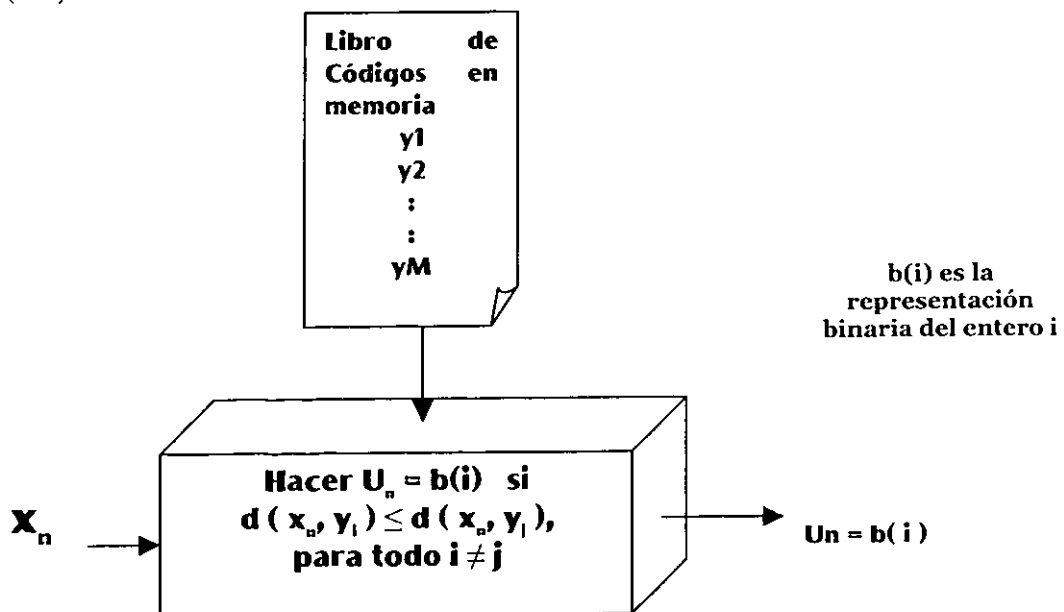


Figura 3.3 Funcionamiento de un codificador VQ

El esquema del funcionamiento de un codificador se muestra en la anterior figura (3.3).

2) Optimización del decodificador dado el codificador.

Si se tiene un codificador γ determinado, entonces ningún decodificador asociado puede dar mejores resultados que aquél que asigna a cada símbolo de canal v el centroide generalizado (centro de gravedad, baricentro) de todos los vectores fuente codificados en v , o lo que es igual:

$$\beta(v) = \text{cent}(v) = \min_{x \in \mathcal{A}}^{-1} E(d(X, x) | \gamma(X) = v) \quad (3.7)$$

donde $\beta(v)$ es el vector que arroja la distorsión promedio mínima condicional, dado que el vector de entrada fue mapeado hacia v .

3.3.5 Algoritmo de Lloyd Generalizado

Ya que hemos mencionado el algoritmo de Lloyd, tratemos de explicarlo con más detenimiento ahondando en la versión generalizada propuesta por Linde, Buzo y Gray (Linde, 1980), conocido con las siglas LGB. Su sustento principal son las dos condiciones de optimalidad tratadas en el inciso anterior. El diseño original (Lloyd) fue planteado para PCM con una variable escalar aleatoria que tuviera función de densidad de probabilidad conocida y una medida de distorsión cuadrática. Sin embargo, se encontró utilidad en la generalización de dicho algoritmo a vectores. El único requisito necesario para la medida de distorsión es que se puedan calcular los centroides. Los pasos del algoritmo son:

Paso 1. Se tiene una secuencia de entrenamiento y un decodificador iniciales.

Paso 2. Codificar la secuencia de entrenamiento a una secuencia de símbolos de canal usando la regla de mínima distorsión del decodificador. Si la distorsión promedio es pequeña, terminar.

Paso 3. Reemplazar las palabras anteriores de código del decodificador, para cada símbolo de canal v , por el centroide de todos los vectores de entrenamiento mapeados en v en el paso 2. Ir al paso 2.

El tope para detener el algoritmo generalmente es una distorsión relativa de umbral. La figura siguiente (3.4) esquematiza la estructura del algoritmo.

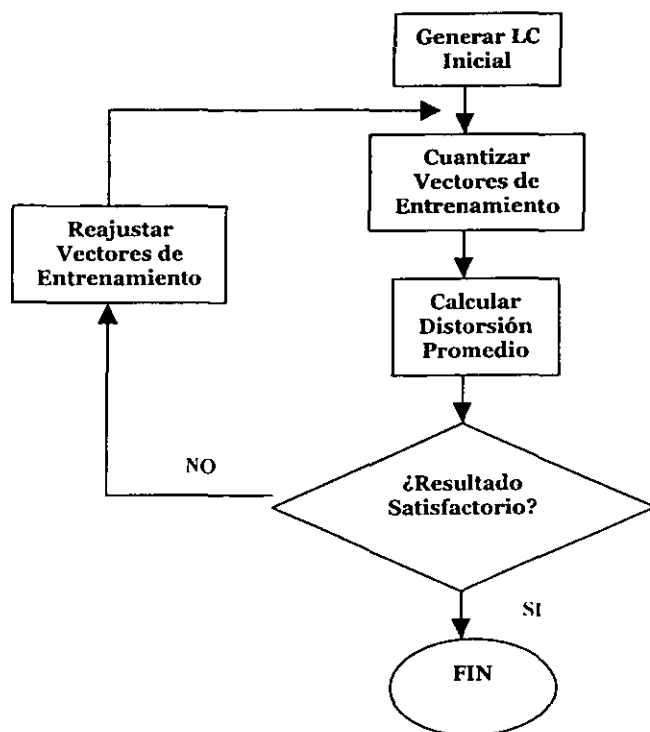


Figura 3.4. Algoritmo LGB básico. En el caso esquematizado, la generación del LC inicial crea el número definitivo de vectores, mientras en el caso alternativo, se duplica el tamaño del LC en cada iteración.

Libro de Códigos Inicial:

Como se vio, un algoritmo iterativo de esas características requiere un libro de códigos inicial para comenzar a funcionar. Principalmente, hay dos formas de atacar el problema: se puede empezar con un libro de código simple, del tamaño requerido, y afinarlo hasta lograr una versión satisfactoria; o se puede partir de un solo vector (o un grupo pequeño de ellos) e iterativamente alcanzar el número deseado de vectores.

En el primer caso, para construir un libro de tamaño n , se sugiere en la literatura (Gray, 1984), comenzar con n vectores aleatorios, o bien, proponer un código escalar, como en un cuantizador uniforme, k veces sucesivamente, y después “emparejar” el libro de códigos resultante hasta su tamaño correcto. A esto se le llama “códigos de producto”.

En el segundo caso, se utiliza la técnica de separación⁷, que consiste en iniciar con un solo vector para una dimensión fija, y separar sucesivamente el mismo hasta llegar, de forma arborescente, al tamaño de libro de código deseado. Primeramente se halla el centroide de toda la secuencia de entrenamiento (código óptimo de tasa cero). Este nuevo vector es levemente “perturbado” (o puede escogerse un vector muy lejano al mismo, intencionalmente) para tener ya sea dos vectores muy cercanos o muy lejanos, respectivamente. Con el nuevo grupo de vectores se recalculan los centroides. El proceso corre de forma iterativa, en donde el código final de cualquier etapa es dividido para ser el inicial de la siguiente. El algoritmo termina al obtener el número deseado de palabras.

3.3.6 Tipos de Cuantizador Vectorial sin Memoria

Las cargas computacionales en el VQ se pueden reducir, como lo sugieren algunos modelos reportados en la bibliografía, y que se mencionan a continuación:

- VQ con búsqueda en árbol: Son un producto natural del algoritmo de separación (párrafo anterior), en el que el codificador va guiándose a través del árbol navegando por las ramas donde encuentra menor distorsión. Al tomar cada decisión, además de avanzar hacia el mejor vector posible, construye la etiqueta de dicho vector del libro, que será utilizado para representar al de entrada.
- VQ Multipaso: El multipaso es un VQ con búsqueda en árbol, con la diferencia de que únicamente se almacena un libro de códigos pequeño para cada nivel del árbol, en vez de un libro diferente para cada nodo de cada nivel. La ventaja de esto es que se reduce la carga computacional y el espacio de almacenamiento. Un ejemplo de VQ multipaso binario es el propuesto por Ang y Durrani (Ang, 1986): funciona con dos codificadores vectoriales; se codifica el vector de entrada con el primer

⁷ Ampliamente denotada con el término inglés *splitting*.

codificador y se obtiene un vector de error. Dicho vector de error se codifica posteriormente con el otro codificador y los dos símbolos juntos forman el símbolo de canal, que será reconstruido por el decodificador sumando los vectores de reproducción correspondientes.

- VQ Ganancia/Forma: Es un cuantizador vectorial en el que dos códigos distintos, pero dependientes, se encargan de la codificación de la “forma” y la “ganancia” de la señal. La forma está definida como el vector original de entrada normalizado con la remoción del término de ganancia (energía).
- VQ de separación de media: En este sencillo cuantizador, se retira la media muestral en vez del término de energía. Recordemos que la media muestral, $\langle x \rangle$ de un vector k -dimensional, es $k^{-1} \sum_{i=0}^{k-1} x_i$. Para llevar a cabo esta tarea, se utiliza un cuantizador escalar para codificar la media muestral del vector, con objeto de que la media muestral codificada sea restada de todos los componentes del vector de entrada para obtener un vector con, aproximadamente, media muestral cero. Este nuevo vector es entonces cuantizado vectorialmente.
- VQ de Láttice: Esta estructura de VQ es capaz de llevar a cabo búsquedas eficientes y administración de memoria. Es una generalización k -dimensional del cuantizador escalar uniforme. Un láttice en un espacio k -dimensional es una colección de todos los vectores de la forma $y = \sum_{i=0}^{n-1} a_i e_i$, donde $n \leq k$ y donde e_0, \dots, e_{n-1} son un conjunto de vectores linealmente independientes en \mathbf{R}^k , y donde los a_i son enteros arbitrarios. Un cuantizador de láttice es aquel cuyas palabras de código son un subconjunto de un láttice. Desafortunadamente, no se prestan para ser construidos con el algoritmo de Lloyd, o variaciones de él, sin perder su estructura, y los buenos cuantizadores producidos con el algoritmo de Lloyd generalmente no pueden ser bien aproximados por medio de láttices (Gray, 1984).

3.4 Cuantización Vectorial con Memoria

Una manera sencilla de incorporar memoria en un cuantizador vectorial consiste en utilizar distintos libros de código para cada vector de entrada, y esta elección se hace basándose en vectores de entrada anteriores. El decodificador debe saber qué libro de códigos está siendo usado por el codificador con objeto de decodificar correctamente los símbolos. Esto puede ser llevado a cabo de dos formas distintas: 1) el codificador puede usar un procedimiento de selección de libro de código que dependa sólo de sus salidas anteriores y de esta forma garantizar que la secuencia de libros de código sea rastreada por el decodificador (cuantización vectorial retroalimentada), y 2) que el decodificador sea informado acerca del libro de código seleccionado vía un canal especial de baja tasa de bits.

Algunos exitosos modelos de cuantización vectorial predictiva fueron propuestos por Cuperman y Gersho (Cuperman, 1982), quienes generalizaron el cuantizador escalar predictivo DPCM, así como la primera técnica general de diseño para cuantizadores vectoriales de estado finito planteada por Foster y Gray (Foster, 1982).

3.5 Cuantización Vectorial en Imágenes

Las imágenes digitales son un caso especial de señal. La cuantización vectorial ha demostrado ser una eficiente herramienta para codificar y comprimir imágenes, y numerosos trabajos, que sería imposible citar, han dirigido sus esfuerzos a encontrar formas más eficientes de adaptar el funcionamiento de los cuantizadores vectoriales a un procesamiento más eficiente de las mismas.

El procesamiento digital de imágenes puede ser visto como agente de codificación en varios tipos de aplicaciones, entre las que se hallan la representación, el modelado, la restauración, el realce, el análisis y la reconstrucción. El esquema típico de un cuantizador vectorial sencillo para imágenes es el que se observa en la figura (fig. 3.5).

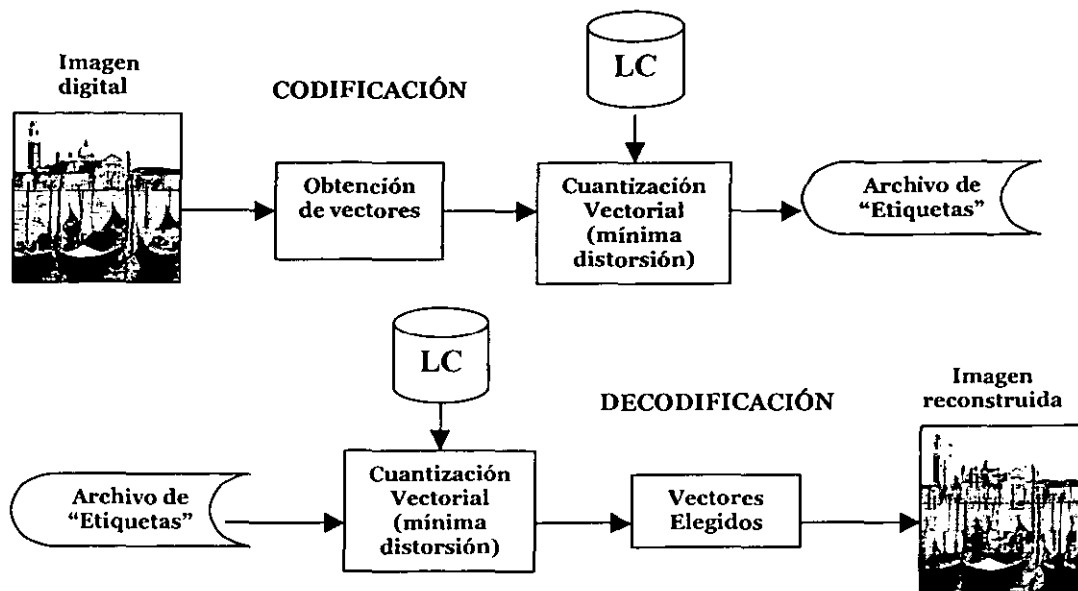


Figura 3.5. Cuantizador vectorial sencillo para imágenes. Nótese que la obtención de vectores puede involucrar cualquier dominio distinto al de la luminancia espacial

En este esquema, que es el más sencillo para procesar una imagen vectorialmente, los datos originales (imagen original) se comprimen y se almacenan como un arreglo de índices, que hacen referencia a ciertos vectores del libro de códigos. El decodificador simplemente revisa a qué vectores están atribuidos dichos índices y reconstruye la imagen aproximada.

En la literatura reciente han aparecido múltiples formas de combinar los pasos de codificación y decodificación. Otro caso sencillo para la rápida transmisión de imágenes, consiste en enviar los índices secuencialmente conforme éstos son obtenidos, sin necesidad de generar un archivo intermedio, y el decodificador irá reconstruyendo la imagen conforme va recibiendo esas etiquetas.

Los parámetros claves en el funcionamiento del VQ en imágenes son el tamaño de los bloques de píxeles operacionales, y si dichos bloques son procesados de forma independiente (sin traslape) o no. Además, la elección del tamaño del libro de códigos, su debida preparación y el diseño de una medida de distorsión adecuada son esenciales para que el proceso de codificación sea satisfactorio. Hay que hacer notar que una vez entrenado el cuantizador, la mayor parte del tiempo de cómputo estará dedicada a la elección de los

vectores representativos en la codificación, ya que la decodificación será un simple y rápido proceso de búsqueda en tabla.

A continuación enlisto una serie de aplicaciones en las que ha hallado cabida el uso de cuantizadores vectoriales en imágenes:

- **Realce.** Su meta es acentuar ciertas características de la imagen para despliegue o análisis posterior. Una interesante sugerencia para su utilización en imágenes médicas se presenta en (Olhsen, 1991).
- **Clasificación y Análisis.** Incluyen análisis de escenas, interpretación de imágenes científicas, reconocimiento de caracteres, cartografía, etc. Un cuantizador vectorial puede ser de mucha ayuda sobre todo en los primeros niveles de la algorítmica requerida.
- **Visualización.** Se entiende como el uso de técnicas de graficación por computadora para desplegar datos experimentales o simulados en 2 o 3 dimensiones. La VQ ha demostrado ser un aliado al resolver el problema del *rendering*.
- **Detección de Bordes.** Los cuantizadores vectoriales de tasa variable tienen algunas propiedades inherentes de detección de bordes que se pueden aprovechar. Al detectar actividad, asignan mayor número de bits a esas zonas que a las regiones homogéneas.
- **Binarización.** Consiste en convertir imágenes en alta resolución a imágenes binarias (blanco o negro) para efectos de transmisión por facsímil, impresión láser, etc. En este caso la biblioteca de vectores de código sólo tendrá dos elementos, blanco y negro. La dificultad reside en la elección del tamaño, forma y asignación en los vectores de entrada.

Poderosos sistemas de procesamiento de imágenes, con diversos fines, y en diversos dominios, han sido reportados en la literatura. Muchos autores han propuesto recientemente la utilización de VQ en dominios distintos al espacial.

Ang (Ang, 1997), por ejemplo, presentó un interesante trabajo en el que explica como pueden ser eficientemente codificados vectorialmente los coeficientes resultantes de la

expansión de una imagen digital en *wavelets*⁸. Resulta interesante la utilización de varios libros de código, uno para cada escala wavelet. El resultado es de alta compresión con muy buena calidad perceptiva en las imágenes reconstruidas.

Otro tipo de estudios (Ching, 1997) han intentado incorporar la utilización de estructuras no regulares, tales como el *quadtree*, para optimizar el tamaño de los vectores a codificar, que en este caso es variable. La estructura *quadtree*, que se explica al final de este capítulo, se adapta perfectamente a la segmentación de imágenes, y da muy buenos resultados en esquemas de wavelets al agrupar conjuntos de coeficientes en áreas uniformes.

3.5.1 Cuantización Vectorial Adaptiva

Resulta de suma importancia mencionar al menos este tema debido a que una considerable cantidad de investigaciones en los últimos años se han enfocado a estudiar, diseñar y mejorar técnicas de cuantización vectorial adaptiva que den lugar a algoritmos de compresión y codificación de imágenes más eficientes.

En la cuantización vectorial adaptiva, un libro de códigos individual es compilado para cada imagen de entrada (Constantinescu, 1994). Como resultado, se obtiene un libro de códigos más pequeño y de mejor calidad. Si se logra superar el gran problema, que consiste en transmitir, para cada imagen, las palabras de código recién moldeadas; el cuantizador vectorial adaptivo superará al convencional por lo menos en cuanto a calidad de reproducción se supone. En el esquema de cuantización de imágenes adaptiva con longitud de código variable (IAVQ-VLTC) propuesto por Wang et. al. (Wang, 1992), la transmisión de palabras de código y etiquetas se intercala, de manera que el problema primordial se alivia un poco. A pesar de esto, dicho algoritmo se enfrenta al reto de diseñar un libro de códigos complejo: aspectos como el establecimiento de un libro óptimo inicial, búsquedas largas, carga computacional intensiva, regeneración iterativa de libros de código,

⁸ Las *wavelets*, u *ondeletas*, por su derivación del francés *ondelette*, suponen la descomposición de la imagen, que es una señal no estacionaria, a un conjunto de sub-imágenes escaladas (o sub-bandas en frecuencia), que son cuasi-estacionarias y más fáciles de tratar. (Ang, 1997).

uso de memoria y la complejidad inherente del algoritmo adaptivo impiden, en la mayor parte de los casos, un desempeño adecuado en aplicaciones de tiempo real.

3.5.2 Estructuras Quadtree

La segmentación quadtree es una técnica de descomposición jerárquica que divide en particiones una imagen, para dar como resultado bloques de tamaño variable ordenados en una estructura del mismo nombre. La estructura quadtree es arborescente⁹, y en cada nodo interno cuenta con ninguno, o cuatro hijos (ramas). Cuando ha terminado la segmentación, cada nodo en el quadtree corresponde a un sub-bloque de la imagen de entrada, con tamaño y ubicación conocidos. Es ampliamente utilizada por su simplicidad y por la variedad de aplicaciones que pueden derivarse de su utilización. Observe (Yu-Chen, 1999; Banister, 1999). Generalmente se asocia a esquemas de cuantización vectorial en imágenes.

Hay varias estrategias que se pueden seguir para codificar bloques de distinta importancia perceptiva. Algunas de ellas optan por clasificar los elementos de la estructura quadtree en “bajo detalle” y “alto detalle”, y codificar en consecuencia. Otras asocian los niveles de descomposición dentro del quadtree a libros de código distintos, y después llevan a cabo una cuantización vectorial de tasa variable.

Es indudable que en la elección del criterio de división (*splitting*) para cada nodo, residirá en gran parte el éxito o el fracaso de la segmentación deseada. El diseño de esta medida deberá estar en concordancia con los patrones a los que se desea dar mayor importancia.

A continuación se presenta un esquema típico de división quadtree (fig. 3.6). Para llevarla a cabo, se subdivide iterativamente la imagen en cuatro secciones, o ninguna, que pueden ser vistas como las ramas de un árbol. La división iterativa se detendrá en cada nivel de profundidad cuando la actividad del bloque analizado esté por debajo de un umbral (de varianza, en este caso) preestablecido.

⁹ Es decir, con estructura de árbol, en la que se parte de un padre único (imagen completa) que engendra a toda la descendencia.

Capítulo IV

El Mapeo Luminancia - Brillantez

En el presente capítulo se abordarán los conceptos teóricos esenciales del mapeo luminancia-brillantez. La interesante justificación perceptiva de dicho mapeo ocupa la primera sección del capítulo. Se definen ciertos conceptos formales y se introduce un importante efecto que se hace evidente en el mundo de la brillantez: la *inducción*. Posteriormente, el texto encadena las ideas con la revisión y análisis de algunos modelos de brillantez previamente diseñados, con muy distintas inclinaciones y objetivos. Se será muy claro en la puntualización de las cualidades que cada uno persigue, o de las facetas de imagen que se desea destacar al *mapear*. La formalización de una propuesta matemática interesante, y su derivación ocupa gran parte del capítulo. Además del desarrollo matemático, la última sección de la discusión expondrá la particularización del esquema al caso bidimensional, para poder experimentarlo en imágenes digitales. Se acompañará el “esqueleto” de ecuaciones con unos cuantos resultados obtenidos en el laboratorio, con base en la programación del mapeo. La teoría derivada del mapeo a la brillantez nos abre muchas posibilidades que indagar en la construcción del codificador. Para empezar, el simple hecho de tener la información de una imagen digital en un dominio distinto (isotópico pero distinto) al de la luminancia; y, en teoría, con una compresión inherente en el mismo, es materia susceptible de ser explotada con los temas estudiados en el resto de esta tesis.

4.1 Percepción de la Brillantez

Luminancia y brillantez son dos importantes términos en el tópico de la visión humana que tienden a ser confundidos. Como veremos a lo largo de este capítulo, son conceptos ligados a fenómenos similares pero se refieren a percepciones distintas. Su

comportamiento y su relación son importantes para entender el sistema de visión humana y su reacción a estímulos diversos. Además, trataremos de demostrar que dicha relación puede ser benéfica para fines de procesamiento digital de imágenes, en términos de compresión, visualización y codificación.

Formalizando un poco estos conceptos, y de acuerdo con el vocabulario del CIE¹, podemos decir que la brillantez es “un atributo de las sensaciones visuales de acuerdo al cual un área parece emitir más o menos luz” (CIE, 1987). Resulta más sencillo hallar definiciones formales de luminancia en la bibliografía. Una de ellas, que se apega al contexto, sería la siguiente: “La luminancia, medida en Lúmenes (lm), proporciona una medida de la cantidad de energía que llega al observador, procedente de una fuente luminosa” (González, 1992). Habrá que entender, entonces, que para fines dimensionales, la luminancia es una medida de *densidad de luz*. Estrictamente, acotada en relación de una cantidad de intensidad luminosa y una de superficie: en MKS² propondríamos cd/m^2 .

Algunas propiedades que suelen ser incorporadas en modelos HVS que, como el nuestro, tengan como objetivo la calidad perceptual de la imagen, son: la sensibilidad a la luz, la sensibilidad a frecuencias espaciales y los efectos de enmascaramiento en la imagen. La sensibilidad a la luz es la dependencia que existe entre la sensibilidad y la luminancia local. En general, el modelo HVS es más sensible en áreas oscuras que en áreas brillantes. La sensibilidad espacial se describe generalmente con la ayuda de un filtro paso-bajas o de un filtro paso-banda. De cualquier forma, la sensibilidad frecuencial es fuertemente dependiente de la luminancia del fondo. Dicha sensibilidad en frecuencia aumenta y su máximo se localiza en altas frecuencias cuando la luminancia se incrementa. Los efectos de enmascaramiento describen la influencia de los contenidos de la imagen en la visibilidad de distorsiones. Estos efectos ocurren principalmente en la vecindad de bordes y en regiones plagadas de textura en las imágenes (Westen, 1995).

En los últimos 15 años se han desarrollado varios trabajos en percepción visual del HVS. Algunos de ellos (Zetsche, 1989) proponen modelar la sensibilidad a la luz con un conjunto jerarquizado de filtros paso-bandas espacialmente orientados. La idea de utilizar los bancos de filtros ha surgido de la gran cantidad de experimentos visuales que se han

¹ Commission Internationale de l'Eclairage

² Metro-Kilogramo-Segundo

llevado a cabo con enrejados, o arreglos en busca de enmascaramientos, con formas regulares. El efecto de enmascaramiento en dichos modelos se explica por la elevación del umbral en las bandas espaciales en frecuencia. Se ha probado que el máximo enmascaramiento en bordes llega justamente en la posición del mismo.

Es un hecho que la relación que guardan las distribuciones de intensidad objetiva y subjetiva en una escena, no es trivial. Un buen ejemplo de este fenómeno es la “ilusión óptica” que se presenta a continuación (fig. 4.1):

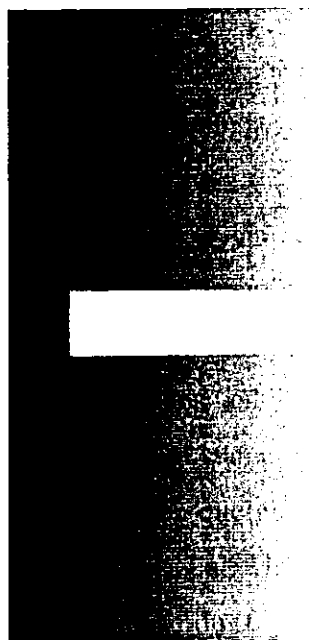


Figura 4.1 El fenómeno de inducción de la brillantez

En la figura se observa una barra de luminancia constante sobre un gradiente vertical que barre linealmente desde el tono más oscuro al claro del espectro de gris. A pesar de que la *reflectancia* de la barra es uniforme, tendemos a observarla más clara del lado izquierdo (donde está inmersa en tonos oscuros) que del derecho (donde se halla entre tonos claros). Varios modelos ópticos de este tipo existen, y pueden comprobar distintas paradojas del sistema de visión.

Una elegante demostración de la diferencia fundamental entre brillantez y luminancia fue dada por (Schirillo et al., 1990), quienes mostraron que la luminosidad es afectada profundamente por la profundidad, mientras que la brillantez no.

Una idea que es importante hacer notar es que si la luminancia de una escena aumenta, también lo hará su brillantez percibida; pero si se deja constante la luminancia y se altera la distribución espacial de dicha escena, la brillantez cambiará notablemente.

La estructura del mundo exterior es reflejada literalmente en un patrón de luminancia proyectado en la retina. Dicho patrón contiene una mezcla de información acerca de la fuente luminosa, las propiedades reflectivas, la forma y disposición de los objetos y la posición del observador. Sin duda alguna, hay una relación muy cercana entre la distribución de luminancias y la representación de brillantez, ya que las variaciones locales de brillantez y las variaciones de la luminancia están asociadas. Son correspondientes topológicamente. Como se profundizará más adelante, La brillantez es robusta ante variaciones en la distancia de observación y ante rotación. Estas consideraciones sugieren que las representaciones de brillantez son insensibles a las propiedades de la fuente luminosa y a las condiciones de observación.

4.2 Modelos de Brillantez

Desde hace varios años (más de cien para ser precisos), se afirmaba ya que la visión de brillantez era un fenómeno fuertemente orientado a objetos, como lo explicaban los trabajos de varios estudiosos de la visión (Mach, von Helmholtz), de forma que si se intenta llegar a un modelo explicativo, éste necesariamente estará ligado a patrones espaciales. Un excelente ejemplo de ello es el importante desarrollo de Blommaert y Martens (Blommaert, 1990) que buscó plantear las características de un mapeo robusto hacia la brillantez basándose en innumerables trabajos anteriores. Dicho esquema será la base para el desarrollo de nuestra función de brillantez.

La idea principal del modelo es, a grandes rasgos, que el procesamiento que permite la representación de la brillantez en una escena deberá ser robusto ante las propiedades de la fuente luminosa y las propiedades del observador; lo que supone, por ejemplo, que la

brillantez de una superficie y su contraste con respecto a su vecindad espacial son poco dependientes de la distancia a la que se observa la escena y poco dependientes de los cambios en intensidad que arroja la fuente luminosa que la baña. Es entonces cuando adquiere especial importancia práctica la suposición de que el fenómeno perceptivo está orientado a los objetos.

La elaboración de mapas de brillantez deben procurar la adquisición de una visión de la imagen en varias escalas. Esto puede ser logrado con filtrados sucesivos o con el uso de wavelets. Lo anterior es muy necesario debido a que, como ya se ha mencionado (Zetsche, 1989), el HVS es un aparato que trabaja a distintos niveles o “canales” de resolución. De esta forma intentaremos también que el resultado final sea dependiente del nivel luminoso absoluto (libre de *offset*). Además, una cualidad importante que debe presentar el conjunto de funciones del mapa, es que intente derivar el mapa de brillantez procurando que éste sea insensible a los cambios de distancia del observador.

Como ya se mencionó anteriormente, la representación en brillantez no es una simple copia de la distribución de luminancias que se fija en la retina, el fenómeno presentado en la figura anterior (fig. 4.1), como dice al calce, es llamado “fenómeno de inducción de brillantez” y su cuantificación es una de las metas del mapeo. Hay que hacer notar, de la misma figura, que efectivamente ésta puede ser descrita sin un conocimiento previo de la fuente luminosa, y que el efecto observado no varía casi nada con respecto a la posición del observador (mientras estas variaciones no lleguen a límites absurdos). Más aun, la rotación o el desplazamiento de toda la figura no representaría mayor problema para efectos de inducción. Estos hechos refuerzan la premisa de la robustez³, brevemente mencionada anteriormente.

Bloomaert y Martens (Bloomaert, 1990) llevan a cabo una interesante revisión de argumentos que resultan de suma importancia en la construcción del modelo. A saber,

- La brillantez de un objeto permanece constante si se varía la intensidad de la fuente luminosa.
- Si la forma o posición de la fuente luminosa cambia, la distribución de luminancias de la escena cambiará de varias maneras: en la forma y tamaño de las sombras y en

³ Dicha robustez será, entonces, vista como una insensibilidad de la representación de la brillantez a los cambios en las variables mencionadas.

la reflectancia de las superficies, además de que los gradientes de iluminación suaves que se extienden más allá de las fronteras del objeto (y que se deben sólo a las propiedades de la fuente luminosa), aparecerán o bien se alterarán.

- La distancia de observación determina la escala a la que los objetos son proyectados en la retina. De cualquier forma, esto no tiene mayor influencia en la brillantez del estímulo. Aparentemente esto no resulta problemático debido a que el HVS analiza la imagen retinal en un esquema multi-escala.
- Hay que tratar de preservar, en la representación de brillantez, las propiedades de los objetos en efectos causados por desplazamiento tangencial del observador con respecto a la escena. Esto se refleja en ligeras deformaciones a la geometría 2D de la distribución de brillantez, mismas que hay que atender.
- Como se dijo, la rotación en objetos tiene efectos indistinguibles en la percepción de la brillantez.
- Cambios en el punto en el que se fija la vista dan lugar a desplazamientos en la distribución de luminancias a lo largo de la retina, sin embargo, la fracción de Weber⁴ se mantiene constante bajo esta variación (siempre y cuando el estímulo no sea muy pequeño).

En el proceso de percepción de brillantez hay una compresión inherente de la información contenida en las luminancias. Para dilucidar el proceso de análisis de este fenómeno, se plantean aquí los tres aspectos de compresión más estudiados, y que se tratarán de explotar en el mapeo que se pretende lograr (Schouten, 1993):

- 1) Invariancia de la brillantez (conocido en la literatura como *brightness constancy*)
- 2) Compresión del contraste en el dominio de la brillantez
- 3) Fenómenos de inducción en brillantez (*brightness indentation*)

Los pasos que se seguirán en la obtención del mapeo luminancia-brillantez son tres, también: (fig. 4.2)

⁴ La fracción, o coeficiente de Weber, fue brevemente explicada en el capítulo 2. Ver también (González, 1992)

- 1) La obtención de un modelo de representación de luminancias multi-dimensional. Tratará de ser un esquema asociado a cierto nivel promedio, es decir, será una medida de contraste en varias escalas de resolución. Así lograremos que cada señal sea insensible al nivel de iluminación y refleje el comportamiento deseado de invariancia. Los “campos receptivos” que compondrán cada nivel resolutivo pueden ser, como lo proponen Blommaert y Martens (Blommaert, 1990), funciones de diferencias de gaussianas (DOG's).
- 2) Se procederá a armar una versión “ensamblada” de la imagen, que sea insensible a la iluminación e invariante ante los cambios en la distancia de observación. Esto se puede lograr sintetizando dicha señal ensamblada a partir de las señales escaladas, con una suma ponderada o con la integral de las señales a lo largo del dominio de las escalas.
- 3) Un ajuste local de la escala de brillantez para dar lugar al fenómeno de inducción. La inducción puede ser vista como una *deflexión* del mapa ensamblado, que genera un rango dinámico mas reducido en el mapa de brillantez, procurando que la información local del contraste no se vea seriamente deteriorada.

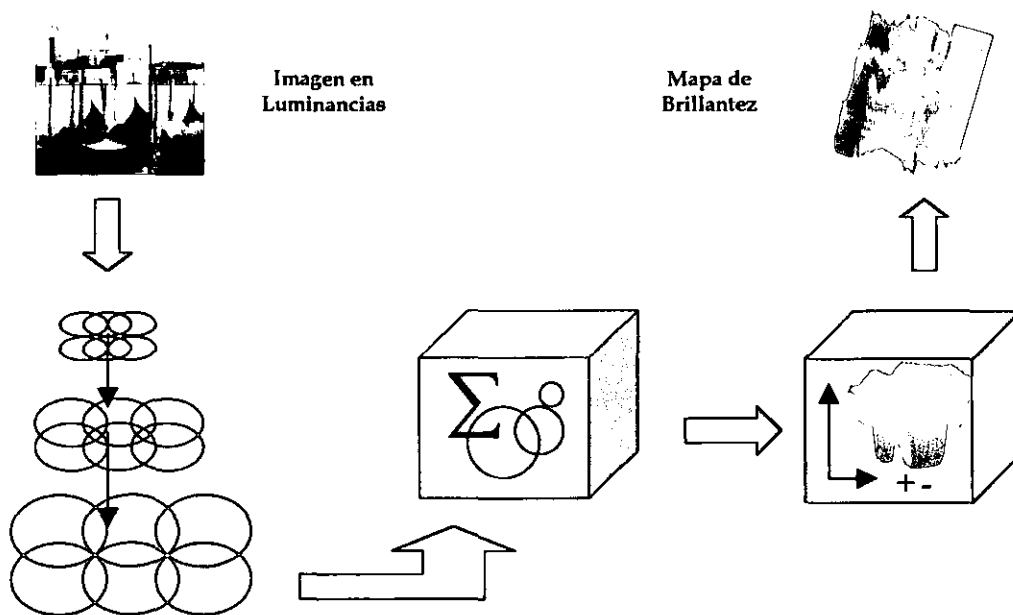


Figura 4.2 Proceso del Mapeo L-B: Primeramente, análisis multiescalar con campos receptivos, posteriormente, integración con suma ponderada de imágenes escaladas y finalmente ajuste a la escala local.

4.3 Generación del Mapa de Brillantez

Vayamos paso a paso planteando la construcción del mapeo:

a. Trabajo multi-escala

Funciona agregando una variable extra a la distribución de luminancias de entrada: la escala. El objetivo de trabajar en esquemas multi-escala es el de asemejar el procesamiento visual del ser humano, asociando campos receptivos de distinto tamaño a cada posición del espacio visual. El resultado de esta etapa será la de transformar la señal de entrada (distribución de luminancias) $L(\bar{x})$ en una señal escalada $h_A(\bar{x}, s)$, donde s denota la escala y el vector \bar{x} la posición.

Los campos receptivos encargados de llevar a cabo el análisis local son de dos tipos: el central y el periférico (fig. 4.3). Como ya se mencionó, la elección de funciones gaussianas isotrópicas en la morfología de los campos se debe a que han sido trabajadas anteriormente en la bibliografía (Watt, 1988) y a su fácil inclusión en el esquema. A saber:

$$R_c(\bar{x}, s) = \frac{1}{\pi^2 s^n} \exp\left(-\frac{\bar{x}^T \bar{x}}{s^2}\right), \quad (4.1)$$

$$R_p(\bar{x}, s) = \frac{1}{\pi^2 (\mu s)^n} \exp\left(-\frac{\bar{x}^T \bar{x}}{(\mu s)^2}\right), \quad (4.2)$$

con $\mu > 1$. El número entero n es la dimensión de \bar{x} .

Deseamos que una señal escalada cualquiera sea el resultado de la interacción entre los mecanismos lineales de centro y periferia de los campos receptivos, por lo que tendrá la siguiente forma:

$$h_A(\bar{x}, s) = f[V_c(\bar{x}, s), V_p(\bar{x}, s)], \quad (4.3)$$

en donde V_c es la respuesta central y V_s es la respuesta periférica de la estructura.

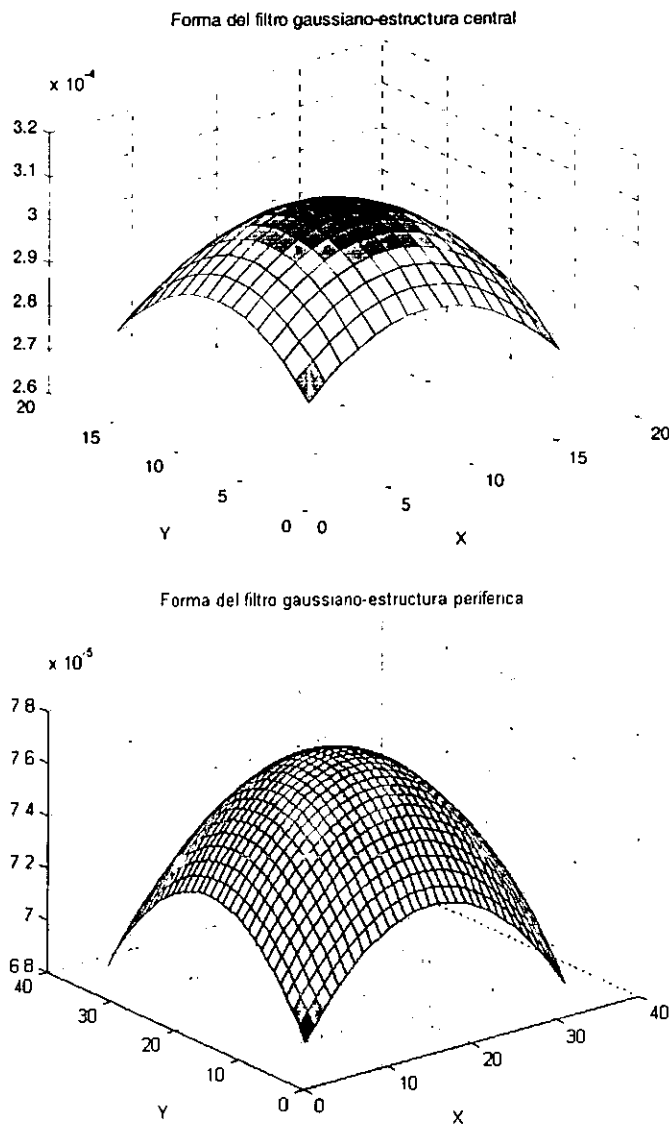


Fig. 4.3 Forma de los filtros estructurales: central y periférico (nótese la diferencia en escala)

Se aceptarán como soluciones respectivas a estas respuestas las convoluciones de la distribución de luminancias $L(\bar{x})$ con los campos receptivos $R_c(\bar{x},s)$ y $R_s(\bar{x},s)$; de esta manera:

$$V_c(\bar{x}, s) = \int L(\bar{u}) R_c(\bar{x} - \bar{u}, s) d\bar{u} , \quad (4.4)$$

$$V_s(\bar{x}, s) = \int L(\bar{u}) R_s(\bar{x} - \bar{u}, s) d\bar{u} , \quad (4.5)$$

Recordemos que el parámetro s regula el tamaño del campo receptivo.

El siguiente problema es lograr que la función planteada en (4.3) funcione como una medida de contraste independiente de la luminancia. Esto se puede lograr de la siguiente forma:

$$h_A(\bar{x}, s) = \alpha \cdot \frac{V_c(\bar{x}, s) - V_s(\bar{x}, s)}{V_c(\bar{x}, s)} , \quad (4.6)$$

Hay una dependencia necesaria a la luminancia, y se especifica por el factor α . Los experimentos llevados a cabo por Schouten (Schouten, 1993), lo motivaron a proponer una función lineal del logaritmo de V_c , $\log V_c$, misma que retomamos:

$$\alpha = \alpha[\log V_c(\bar{x}, s)] = \beta(\log V_c(\bar{x}, s) - \delta) , \quad (4.7)$$

con β y δ constantes.

b. Ensamblaje de las señales escaladas.

El objetivo de este segundo paso es transformar la señal escalada $h_1(\bar{x}, s)$ en el mapa ensamblado $A(\bar{x})$ gracias a una sumatoria lineal a través de las distintas escalas. El mapa ensamblado $A(\bar{x})$ puede verse como una síntesis intermedia de la distribución de luminancias que derivará en el mapa de brillantez.

Uno de los aspectos en los que Blommaert y Martens (Blommaert, 1990) trabajaron más específicamente fue la invariancia ante distancia de observación. Si requerimos que la señal escalada, o en términos más fisiológicos, su actividad neuronal derivada $V(k\bar{x}, ks)$ dé

lugar a un mapa coherente con el arreglo de escalas, la brillantez *escalada* deberá ser igual para el patrón escalado:

$$B'(\bar{x}) = \int f[V(k\bar{x}, ks), s] \cdot ds \quad (4.8)$$

$$= \int f \left[V(k\bar{x}, s), \frac{s}{k} \right] \cdot \frac{ds}{k} \quad (4.9)$$

y para el patrón original:

$$B(k\bar{x}) = \int g[V(\bar{x}, s)] \cdot \frac{ds}{s} \quad (4.10)$$

sin importar la escala k y el patrón de actividad neuronal $V(\bar{x}, s)$. Lo anterior implica que la expresión de ensamblaje adquiere la siguiente forma:

$$A(\bar{x}) = \int h_A(\bar{x}, s) \frac{ds}{s} \quad (4.11)$$

Dentro del esquema de un algoritmo, la integral debe estar definida. Blommaert y Martens (Blommaert 1990) sugieren también que el límite menor de la integral (s^-) corresponda, aunque sea burdamente, con el tamaño de los fotorreceptores o con las unidades mínimas de percepción. El límite superior (s^+) deberá ser de un orden de magnitud mayor que el tamaño del campo visual.

Por conveniencia matemática, podemos llevar a cabo la sustitución $s = \exp \sigma$. Al incorporar los límites mencionados, la integral queda de la siguiente manera:

$$A(\bar{x}) = \int_{\sigma^-}^{\sigma^+} h_A(\bar{x}, \sigma) d\sigma \quad (4.12)$$

Como se verá mas adelante, resultan de especial interés los valores esperados mínimo $\hat{A}_{\min}(\bar{x})$ y máximo $\hat{A}_{\max}(\bar{x})$ de la señal ensamblada, correspondientes a la posición \bar{x} . Si estas cantidades las expresamos en función de la señal escalada, $h_A(\bar{x}, \sigma)$, entonces:

$$\hat{A}_{\min}(\bar{x}) = \min_{\sigma} \left[\int_{\sigma}^{\sigma^*} h_A(\bar{x}, \sigma') d\sigma' \right] + \int_{\sigma^*}^{\sigma^+} h_A(\bar{x}, \sigma') d\sigma' , \quad (4.13)$$

$$\hat{A}_{\max}(\bar{x}) = \max_{\sigma} \left[\int_{\sigma}^{\sigma^*} h_A(\bar{x}, \sigma') d\sigma' \right] + \int_{\sigma^*}^{\sigma^+} h_A(\bar{x}, \sigma') d\sigma' , \quad (4.14)$$

los prefijos \min_{σ} y \max_{σ} significan *minimizar* y *maximizar* con respecto a σ , respectivamente. σ^* se define como el valor de σ más grande para el que el efecto del nivel interno de referencia es inapreciable. Estos parámetros serán utilizados en la siguiente etapa.

c. Ajuste local a la escala de brillantez.

La tarea principal de este último paso es transferir el mapa ensamblado a una escala de brillantez *compresible* con la ayuda de una transformación lineal. Generalmente, la parte más sensible de la escala de brillantez se aloja en el valor $B=0$, con lo que aparecerían valores negativos permisibles en la brillantez.

Considérese, por ejemplo, un número determinado de valores de ensamblaje en una vecindad. Es obvio que esta información local de ensamblaje se desplegaría de forma óptima en la escala de la brillantez si las muestras consideradas se proyectan alrededor de $B=0$. Este ajuste puede ser resuelto si recorremos las muestras de forma que algún valor representativo del ensamblaje se convierta en el punto más sensible de la escala de la brillantez. Vienen a la mente valores importantes como la media, la mediana, el promedio de los valores extremos, etc. Sin embargo, si el mapa de ensamblaje completo fuese transformado de esta manera, serían introducidas variaciones espaciales de brillantez que no precisamente corresponden a cambios reales en la señal de ensamblaje. Este defecto puede ser aliviado utilizando una proyección global, sin ajustar la transferencia de la escala

de ensamblaje a la escala de brillantez. De esta forma, por supuesto, las diferencias locales de ensamblaje no se desplegarían de forma óptima.

Lo importante, repito, es lograr un despliegue óptimo de información de contraste con un mínimo de variaciones espaciales introducidas en la brillantez. Schouten (Schouten, 1993) plantea un interesante desarrollo para obtener un ajuste óptimo que de resultados satisfactorios y que prediga eficientemente el efecto de la inducción en el dominio de la brillantez. El método de optimización utilizado es de la familia de los multiplicadores de Lagrange.

El *Teorema de Lagrange* (Larson, Hostetler; 1989) dice lo siguiente:

Sean f y g funciones con derivadas parciales primeras continuas tal que f tiene un extremo en el punto (x_0, y_0) de la curva de la ligadura $g(x, y) = c$. Si $\nabla g(x_0, y_0) \neq 0$, entonces existe un número real λ tal que

$$\nabla f(x_0, y_0) = \lambda \nabla g(x_0, y_0) \quad (4.15)$$

El método de los multiplicadores de Lagrange (Larson, Hostetler, 1989) puede abordarse de la siguiente forma:

Si f y g satisfacen las hipótesis del teorema de Lagrange y f tiene un máximo o mínimo sujeto a la ligadura $g(x, y) = 0$, entonces dicho extremo se producirá en uno de los puntos críticos de la función F dada por

$$F(x, y, \lambda) = f(x, y) - \lambda g(x, y) \quad (4.16)$$

La restricción implícita, en el caso que nos ocupa, se construye como sigue:

Debido a que deben ser introducidos el menor número posible de artefactos en el contraste, presentemos una señal escalada $h_B(\bar{x}, \sigma)$ que será una "copia perturbada" de $h_A(\bar{x}, \sigma)$. La diferencia entre ambas señales deberá ser mínima. De aquí parte la restricción medida en

$$\int_{\sigma}^{\sigma^+} \{h_B(\bar{x}, \sigma) - h_A(\bar{x}, \sigma)\}^2 d\sigma \quad (4.17)$$

misma que debe ser mínima. No olvidar que la señal $h_B(\bar{x}, \sigma)$ es la que potencialmente construirá el mapa de brillantez.

Como se mencionó, para hacer una representación óptima en la escala de la brillantez, en cada posición \bar{x} la brillantez debería corresponder al punto más sensible de la escala ($B=0$), y se utilizará como valor representativo el promedio del máximo y mínimo valores de brillantez esperados, $\hat{B}_{\min}(x)$ y $\hat{B}_{\max}(x)$; de forma que el multiplicador de Lagrange tendrá como factor

$$\left\{ \frac{\hat{B}_{\min}(\bar{x}) + \hat{B}_{\max}(\bar{x})}{2} \right\}^2 \quad (4.18)$$

habiendo especificado ambas partes, se procede a construir la función de costo:

$$\int_{\sigma^-}^{\sigma^+} \{h_B(\bar{x}, \sigma) - h_A(\bar{x}, \sigma)\}^2 d\sigma + \lambda \left\{ \frac{\hat{B}_{\min}(\bar{x}) + \hat{B}_{\max}(\bar{x})}{2} \right\}^2 \quad (4.19)$$

Para introducir la invariancia ante la distancia de observación, se emplea un error local, $\varepsilon(\bar{x})$, que es independiente de σ :

$$\varepsilon(\bar{x}) = h_B(\bar{x}, \sigma) - h_A(\bar{x}, \sigma) \quad (4.20)$$

Necesitamos conocer la estimación de $\hat{B}_{\min}(x)$ y de $\hat{B}_{\max}(x)$, y una manera de hacerlo independiente de la escala es así:

$$\hat{B}_{\min}(\bar{x}) = \hat{A}_{\min}(\bar{x}) + \varepsilon(\bar{x})(\sigma_+ - \sigma_-), \quad (4.21)$$

$$\hat{B}_{\max}(\bar{x}) = \hat{A}_{\max}(\bar{x}) + \varepsilon(\bar{x})(\sigma_+ - \sigma_-)$$

Sustituyendo las ecuaciones anteriores y la función (4.20) en el multiplicador de Lagrange, la función de costo queda de la siguiente manera:

$$\varepsilon^2(\bar{x})(\sigma_+ - \sigma_-) + \lambda \left\{ \frac{\hat{A}_{\min}(\bar{x}) + A_{\max}(\bar{x})}{2} + \varepsilon(\bar{x})(\sigma_+ - \sigma_-) \right\}^2 \quad (4.22)$$

Con el objeto de minimizar la expresión anterior, derivamos la función con respecto al error local, $\varepsilon(\bar{x})$, e igualamos a cero. Una vez despejados los términos, resulta lo siguiente:

$$\varepsilon(\bar{x}) = - \frac{\lambda}{1 + \lambda(\sigma_+ - \sigma_-)} \cdot \frac{\hat{A}_{\min}(\bar{x}) + \hat{A}_{\max}(\bar{x})}{2} \quad (4.23)$$

Este resultado no es muy útil aún, ya que lo que se desea es el mapeo de la brillantez. Por lo tanto, podemos definir dicha transformación con la ayuda de la función de error anterior.

Sabemos, por definición, que:

$$B(\bar{x}) = \int_{\sigma}^{\sigma_+} h_B(\bar{x}, \sigma) d\sigma \quad (4.24)$$

con la relación establecida en (4.20), simplificaremos la integral a:

$$B(\bar{x}) = A(\bar{x}) + \int_{\sigma_-}^{\sigma_+} \varepsilon(\bar{x}) d\sigma \quad (4.25)$$

Asociando el resultado obtenido en (4.23), la expresión final tiene la forma siguiente:

$$B(\bar{x}) = A(\bar{x}) - \frac{\lambda'}{1 + \lambda'} \cdot \frac{\hat{A}_{\min}(\bar{x}) + \hat{A}_{\max}(\bar{x})}{2} \quad \text{con } \lambda' = \lambda(\sigma_+ - \sigma_-) \quad (4.26)$$

Este resultado no es mas que una versión modificada del mapa de ensamblaje, con la cualidad de modelar el fenómeno de inducción de brillantez. Con la resta del término de ajuste, no se pierde la invariancia ante la distancia de observación.

4.4 La Aplicación del Mapeo al Mundo Bidimensional

Una vez que se ha desarrollado la teoría correspondiente al mapeo, se presentará la particularización de esta estructura matemática a señales bidimensionales. Lo que se busca con este procedimiento es la predicción, con comportamientos similares que en los casos unidimensionales (Blommaert, 1990), de fenómenos de inducción y de percepción generalizada de brillantez, posteriores al mapeo, con la comprensión inherente que pretendemos explotar. Proyectado en estructuras 2D. Esta sección del trabajo reviste gran importancia, ya que trata la incorporación de un mapeo luminancia-brillantez en un esquema formal de codificación de imágenes. La amplia bibliografía consultada no nos ha remitido a trabajos que hayan experimentado esta simbiosis, que probablemente nos regalará interesantes resultados al llevar a cabo comparaciones de eficiencia en la cuantización vectorial, sobre todo en lo que a creación de libros de código se refiere.

Si se desea incorporar la teoría analógica previa en un modelo que será programado en una computadora, tendremos que generar una “versión discreta” del mapeo de los tres pasos. Así las cosas, consideraremos primeramente el muestreo en la representación multi-escala.

Las imágenes digitales que se someterán al mapeo deberán pasar primeramente por un pre-procesamiento que garantice su normalización, es decir, que fuerze a que sus valores estén entre 0 y 1. Este pre-procesamiento también procurará “rodear” a la imagen con una región de intensidad uniforme, L_o , que haría las veces del nivel de referencia. Se intentará simular ese nivel con la media aritmética de los valores de la imagen. Esto hará que generalmente fluctúe dicho valor de referencia cerca del 0.5. No hay mayores restricciones en cuanto a las dimensiones de la imagen, pero se procurará que no sean imágenes demasiado pequeñas, ya que se verían rebasadas por el filtrado en las escalas más amplias.

El muestreo a lo largo de la escala dimensional s , será llevado a cabo a distancias exponencialmente crecientes. Esto significa que la integral (4.12) se sustituirá con una simple suma de Riemann de términos $h_A(x, y; \sigma_i)$ rescatados a lugares equidistantes en el eje σ , (donde σ equivale al logaritmo natural del parámetro de escala, s).

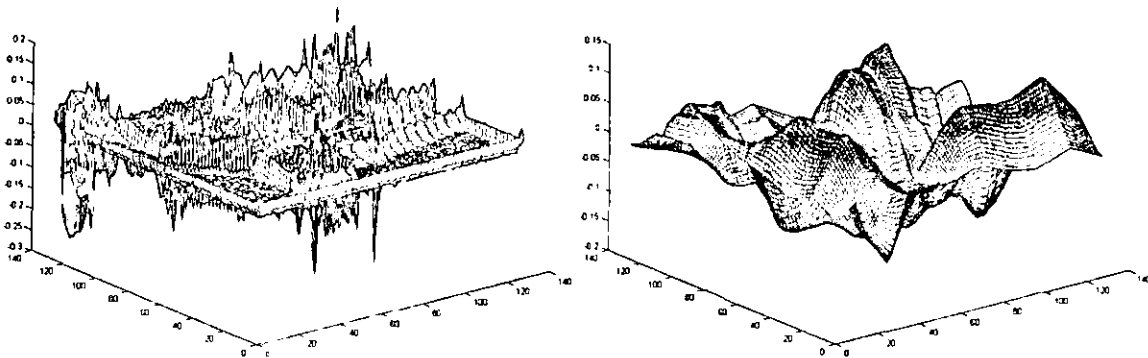


Figura 4.4 La imagen original "House" se muestra arriba. Debajo, dos señales escaladas: la de la izquierda corresponde al índice 3 de resolución, y la de la derecha al índice 7 (penúltimo) en resolución. Nótese la diferencia de detalle y suavidad en ambas.

Gerard Schouten (Schouten, 1990) sugiere que las imágenes convencionales sumergidas en un discreto marco periférico homogéneo, como las que se trabajarán, pueden ser adecuadamente "capturadas" con un número limitado de escalas. Hemos llevado a cabo esta experimentación y se han obtenido resultados afortunados para 7, 8 y 9 escalas. Desafortunadamente, al crecer exponencialmente los tamaños de los filtros, el tiempo de

cómputo empleado en llevar a cabo el filtrado siguiente, aumenta también en esta proporción. Hemos decidido limitar a 8 escalas el procedimiento con el fin de consumir menos tiempo en el filtrado. Además, el aumento de información que se logra al utilizar 9 y no 8 escalas, no es muy considerable (ver fig. 4.4).

La señal escalada que contiene la resolución más fina será $h_{\lambda}^{(1)}(x, y)$, mientras que la señal escalada con la resolución más burda será $h_{\lambda}^{(8)}(x, y)$. De esta forma, el índice de escala es i y la señal escalada en su conjunto es $h_{\lambda}(x, y; s_i)$, para $i \in \{1, 2, \dots, 8\}$, misma que para ser calculada requiere del cómputo individual de todas las escalas por separado. Las respuestas central y periférica, referidas a las ecuaciones (4.4) y (4.5) serán calculadas con un conjunto de filtros binomiales, que son aproximaciones discretas de funciones gaussianas.

Como se especificó en dichas ecuaciones, se requiere una convolución para hallar las respuestas. Los filtros $R_C(x, y; s_i)$ y $R_S(x, y; s_i)$ tendrán una forma muy similar a los de la figura 4.3, cubriendo las áreas locales receptoras de toda la imagen en esa escala. Schouten sugiere fijar el valor de μ , el radio de relación entre los campos receptoras central y periférico, en 2. Esto es por simplicidad, ya que es demostrable que la respuesta del campo periférico, para cierta escala, equivale a la respuesta del campo central en la siguiente:

$$V_S(x, y; s_i) = V_C(x, y; s_{i+1}) \quad (4.27)$$

esto acorta tiempos y ahorra carga computacional, además de hacer más entendible y analizable el proceso.

El mapa de ensamblaje $A(x, y)$ se obtendrá calculando una suma puntual entre señales, de la primera a la última, además de una constante A_G de *offset* que demostró ser necesaria para la coherencia del mapa:

$$A(x, y) = A_G + \ln 2 \sum_{i=1}^8 h_{\lambda}^{(i)}(x, y) \quad (4.28)$$

¿Por qué es necesaria la incorporación de la constante A_G ? Porque ajusta el “piso” del mapa de ensamblaje de manera que éste no sea afectado, por una constante, en el cálculo de las señales independientes de la iluminación, $((V_c - V_s)/V_c)$.

Recordemos la ecuación (4.7) en la que se explica la constitución del parámetro α . Poniendo a α en función de L_0 , el valor de referencia, nos arroja un término parametrizado:

$$\alpha(L_0) = \beta(L_0 - \delta) \quad (4.29)$$

con β y δ constantes. (Schouten, 1990) sugiere la utilización de $\beta=0.1$ y $\delta=-5$ con el tipo de imágenes trabajadas. El mismo autor explica cómo aproximar matemáticamente A_G , al menos con una ecuación de orden unitario, y con una constante derivada de la construcción de un nivel de ensamblaje normalizado de referencia (debido a la observación experimental):

$$A_G = \alpha(L_0) \cdot (1.22) \quad (4.30)$$

En este punto retomaremos los valores extremos definidos en las ecuaciones (4.13) y (4.14). Estos valores no son nada más que valores *esperados* de ensamblaje, mínimos y máximos:

$$\hat{A}_{\min}(x, y) = \min_k \left[\sum_{j=0}^k [(\ln 2) \cdot h_A^{(8-j)}(x, y)] + A_G \right] \quad \text{para } k = 0, 1, \dots, 7 \quad (4.31)$$

$$\hat{A}_{\max}(x, y) = \max_k \left[\sum_{j=0}^k [(\ln 2) \cdot h_A^{(8-j)}(x, y)] + A_G \right] \quad \text{para } k = 0, 1, \dots, 7 \quad (4.32)$$

ya que podemos considerar que en la escala superior la señal es prácticamente constante y tiende a A_G ,

$$(\ln 2) \cdot h_A^{(0)}(x, y) \approx A_G \quad (4.33)$$

El último paso para tener un buen mapa de brillantez consiste en el ajuste local, donde el efecto de inducción de brillantez sale a relucir, como se explica en la evolución de la sección 4.3. Recordando la ecuación (4.26), el mapa final de brillantez, $B(x,y)$ se genera así (con el parámetro laplaciano λ fijo en 2): Ver figura (4.5).

$$B(x,y) = A(x,y) - \frac{1}{3} \{ \hat{A}_{\min}(x,y) + \hat{A}_{\max}(x,y) \} \quad (4.34)$$

MAPA DE BRILLANTEZ

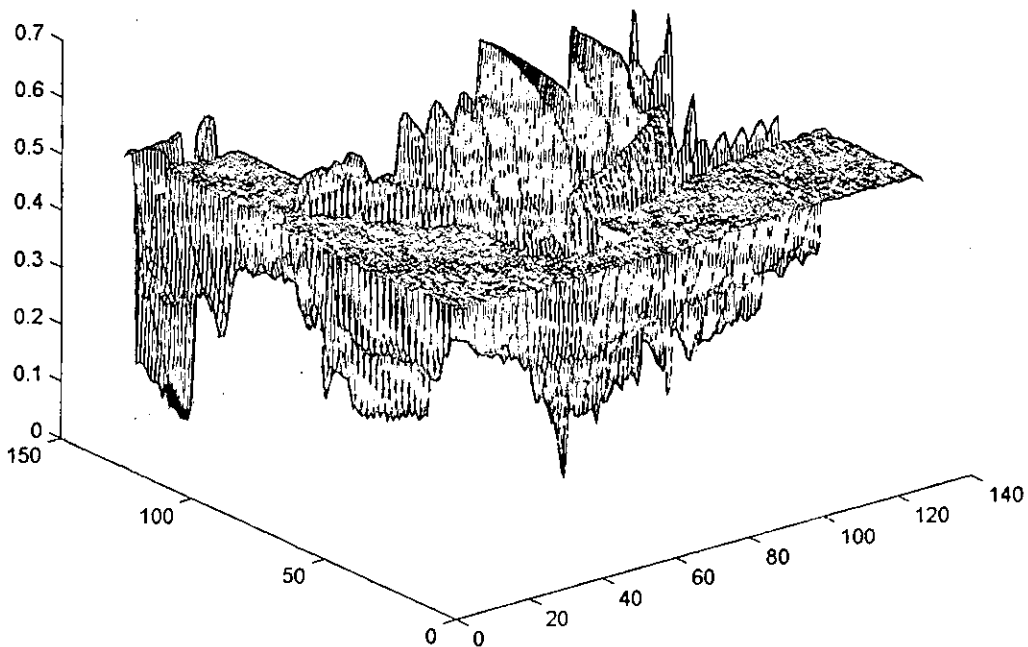


Figura 4.5. Mapa de brillantez de la imagen "House"

4.5 Algunos experimentos con mapas de brillantez

A continuación se presentarán brevemente ejemplos de mapas de brillantez, o de perfiles de los mismos, obtenidos a partir de imágenes de escenas sintéticas (generadas en computadora). Se pondrá en práctica el mapeo propuesto en los párrafos anteriores para observar su desempeño en la computadora. En estos ejemplos se intentará evidenciar las cualidades del mismo, así como las características favorables para la compresión y la codificación de imágenes, que aparecen en el dominio de la brillantez.

Como sugiere Schouten (Schouten, 1993), las llamadas “ilusiones” ópticas son excelentes oportunidades para generar situaciones extremas en el ojo humano y para poner a prueba el desempeño de los mapas de brillantez, estableciendo una comparación con los valores en luminancia. Retomemos la ilusión óptica de la barra uniforme sobre un gradiente, figura presentada en los inicios del presente capítulo (fig. 4.1).

Aquí se muestran los perfiles obtenidos a lo largo de la barra, tanto en luminancia como en su mapa de brillantez. A continuación se comentarán las diferencias. (fig.4.6).

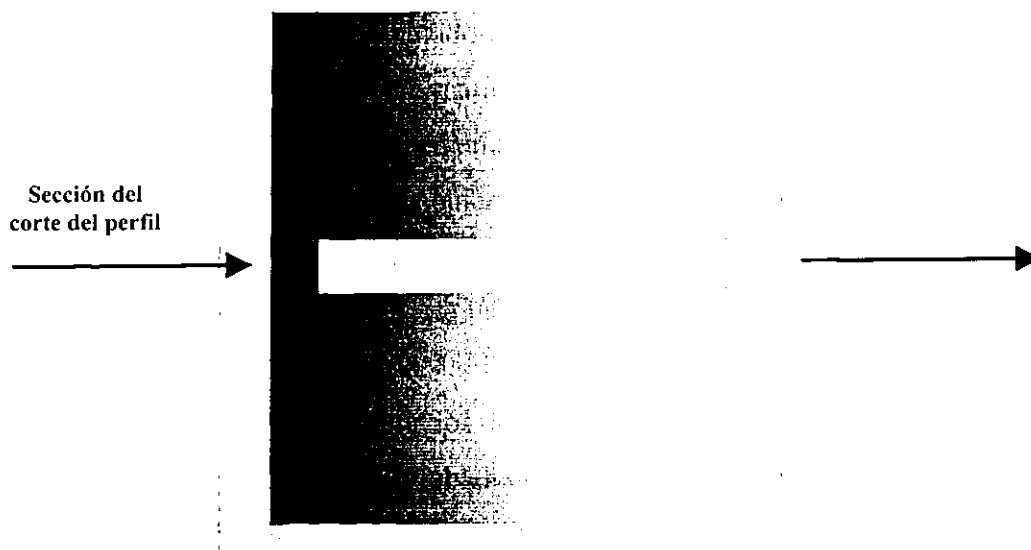


Figura 4.6 a) Ilusión óptica de la barra. Se muestra, con una flecha entrante y saliente, la cota a la que se llevó a cabo el corte del perfil.

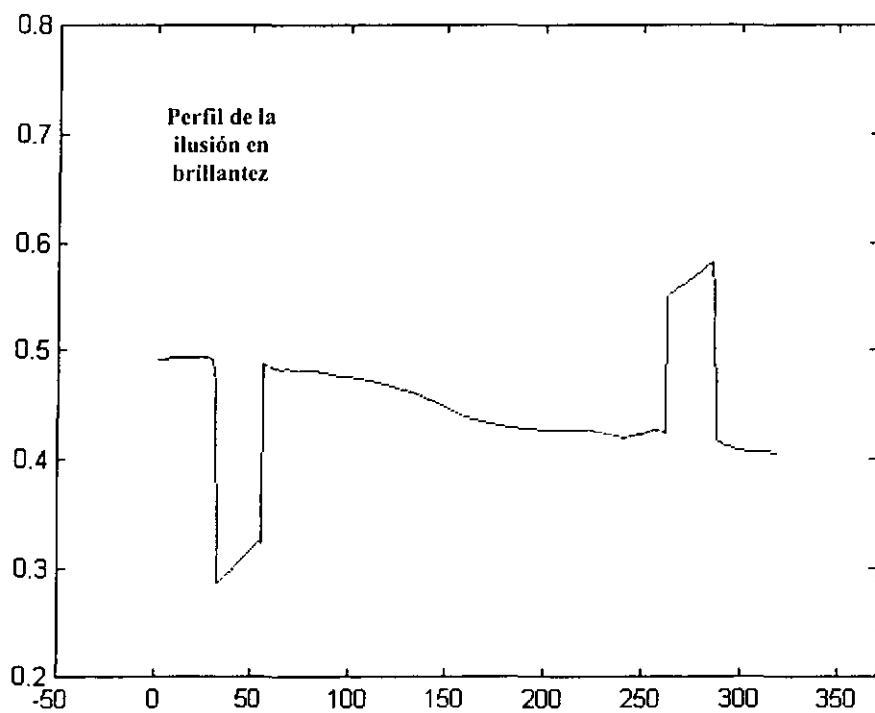
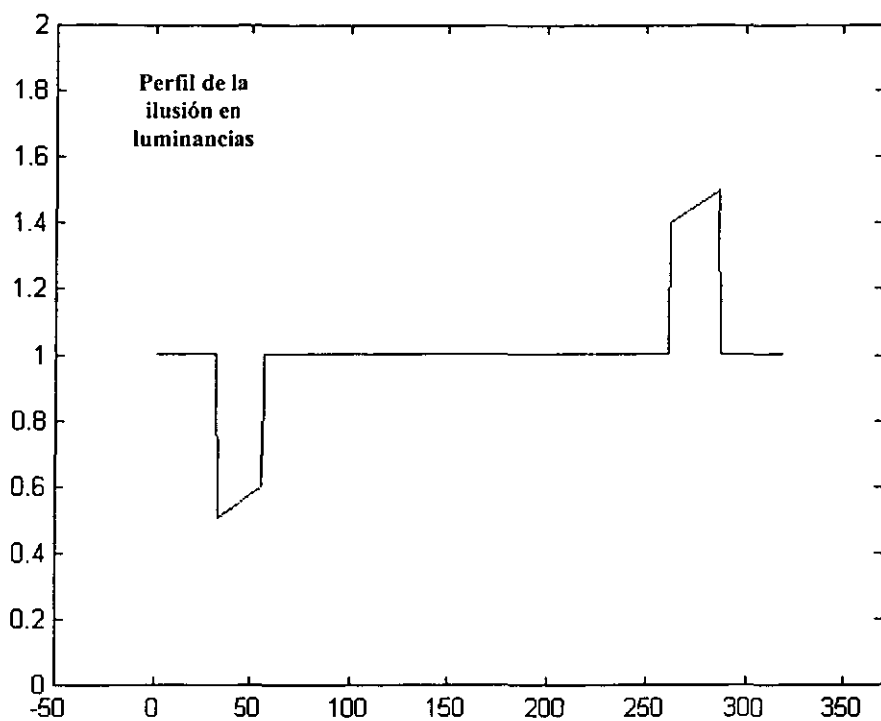


Figura 4.6. b) Comparación de los perfiles, en luminancia (en medio) y en brillantez (abajo) de la ilusión óptica de la barra.

Las gráficas hablan por sí solas. Un análisis del perfil en luminancias deja ver la simple estructura de la ilusión: dos pequeños saltos de luminosidad (antes y después de entrar a la barra –uniforme-) y el regreso finalmente al nivel de referencia. Esto es lo que los números indican, sin embargo nuestros ojos perciben algo totalmente distinto. El mapa de brillantez acierta al predecir las estructuras “perceptivas” del perfil: un inicio relativamente uniforme, después la caída (a la zona oscura) antes de la barra y un tránsito sobre la barra con un cambio sensible de intensidad, sobre todo cerca de la mitad. El perfil no resulta simétrico, como su contraparte en luminancias, y esto se debe a la distinta percepción que el ojo humano tiene frente a las zonas oscuras y claras: efecto parcialmente explicado en la ley de Weber. En zonas de muy baja intensidad nos resulta más difícil identificar contrastes que en las zonas de alta luminosidad.

Trabajemos ahora con una ilusión un poco más elaborada. La siguiente ilusión consiste en un arreglo de círculos de misma intensidad repartidos sobre barras de intensidad decreciente (en la primera mitad de la figura) y sobre un gradiente similar al de la ilusión pasada (en la mitad inferior de la figura). Posteriormente se incluyen un par de perfiles sobre la ilusión en luminancias, y sus correspondientes en el mapa de brillantez, tomados a lo largo de los dos cortes interesantes (transversal a los dos arreglos de círculos).

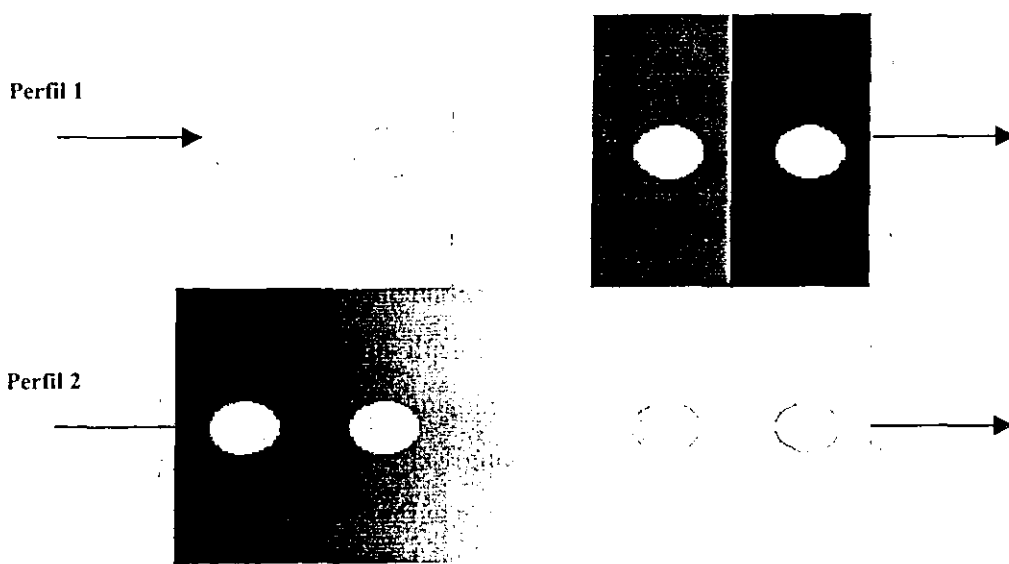


Figura 4.7 a) Ilusión óptica, con la indicación de los perfiles estudiados en este caso

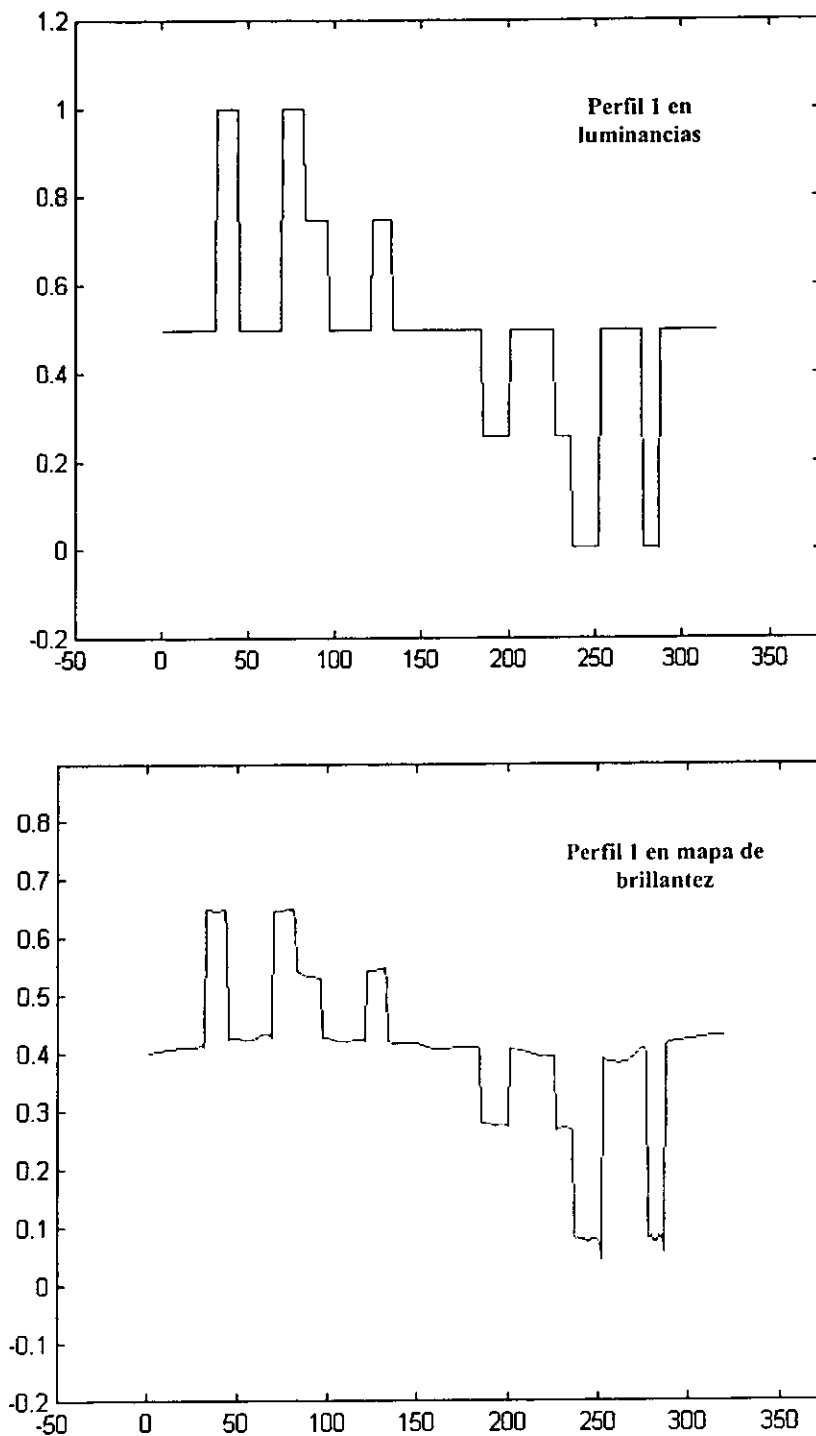


Figura 4.7 b) Perfil número 1, tanto en luminancia como en brillantez, del a ilusión óptica mostrada en la sección a) de la figura

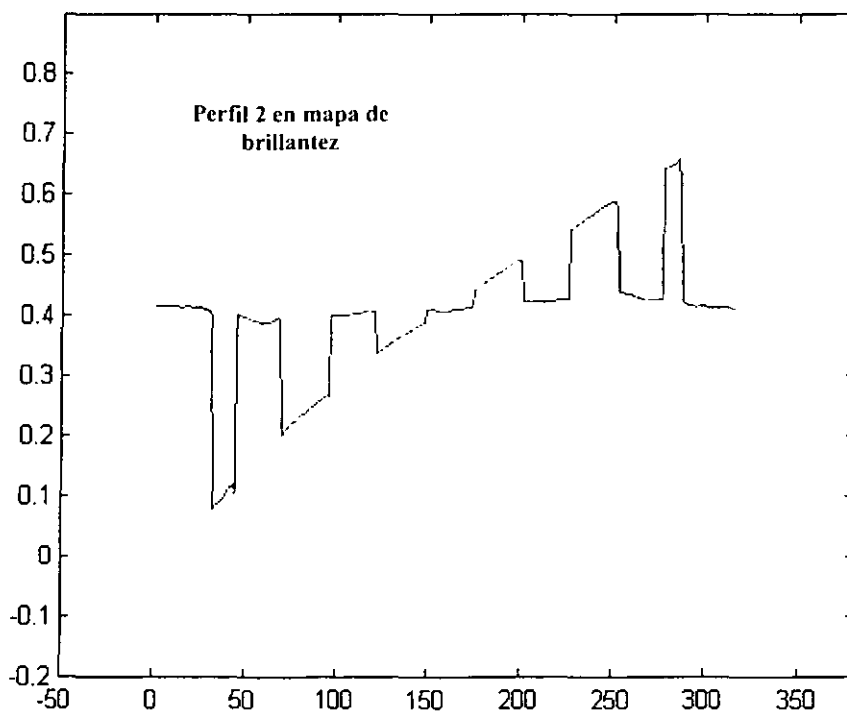
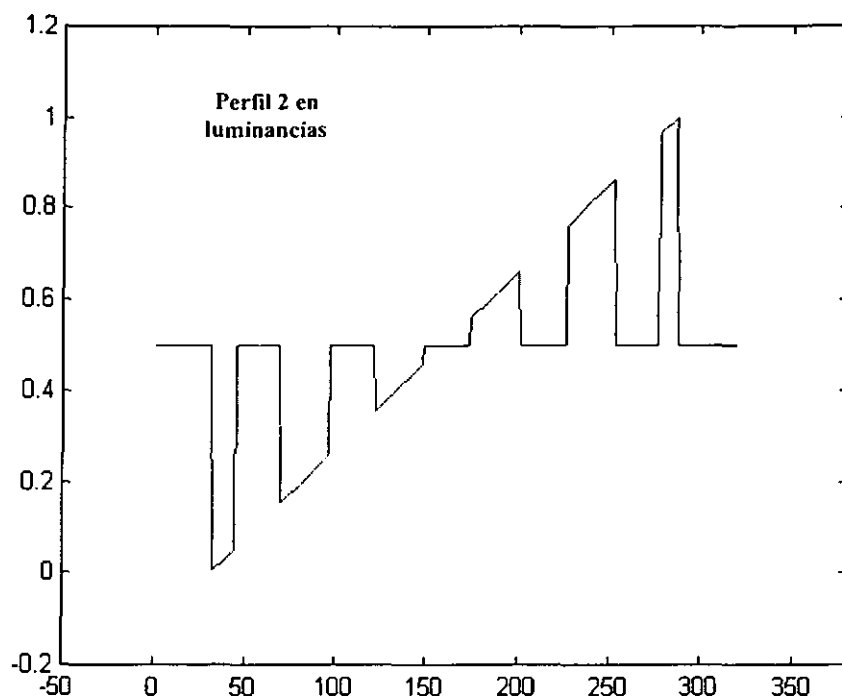


Figura 4.7 c) Perfil número 2, tanto en luminancia como en brillantez, del a ilusão óptica mostrada en la sección a) de la figura

El efecto de inducción, así como otras peculiares manifestaciones visuales aparecen también en esta ilusión óptica. Se dividió la imagen en dos secciones con objeto de observar las características del mapa de brillantez tanto en una disminución discreta de luminosidad como en un gradiente continuo.

Nótese en los perfiles, además de las características y la forma de las líneas, la escala. El dominio de la brillantez se caracteriza por llevar a cabo una compresión inherente, en la que la escala de representación de los valores es mucho más pequeña que en la imágenes en luminancia. Observe, también, la ligera curvatura que se forma en las regiones de los perfiles de brillantez correspondientes a los círculos.

Capítulo V

La Transformada de Hermite

Una imagen digital, como sabemos, está representada por un arreglo de intensidades. La clave para llevar a cabo un buen procesamiento de dicha imagen es identificar y hacer explícita la información relevante contenida en ella, a través de un análisis local.

Los codificadores de imágenes de segunda generación en adelante intentan rescatar las propiedades visuales de las señales que son relevantes al sistema de visión humano. Se ha verificado exhaustivamente que hay estructuras dentro de las imágenes a las que el HVS es sumamente sensible: los bordes, y en general, las orientaciones unidimensionales de energía. Otra estructura importante, pero que resulta de menor influencia al ojo son las texturas, que pueden ser modeladas como patrones bidimensionales dependientes de la escala de referencia (fig. 5.1). Una degradación textural de mediana escala será menos agresiva, perceptivamente, que la corrupción de un borde o un patrón orientado.



Figura 5.1 Estructura orientada unidimensional (borde) y textura (estructura 2D)

Este capítulo del trabajo tiene como objetivos principales un estudio rápido de la transformada de Hermite y la introducción de la transformada binomial como elemento de

compactación energética y como vía de reconocimiento estructural de localidades de la imagen. De igual forma, se explicará su incorporación al esquema general del codificador y las modificaciones que, en el caso discreto, demanda su implantación.

5.1 Introducción a la transformada de Hermite

La transformada binomial de Hermite ha demostrado ser una herramienta poderosa en la codificación de imágenes digitales debido a su alta concordancia con el modelo del campo receptivo visual humano (Ruiz, 1995), ya que está asociada a modelos gaussianos derivativos.

Para llevar a cabo el procesamiento local de una imagen, ésta debe ser multiplicada por una función “ventana”. El tamaño de dicha ventana establecerá el conjunto de puntos de la imagen que contribuirán a cada paso básico de procesamiento. La forma de la ventana determinará el peso relativo de cada punto contribuyente en ese conjunto, y para describir a la imagen en plenitud, este procesamiento local tiene que repetirse un número suficiente de posiciones de ventana. De esta manera, los parámetros a seleccionar son la forma, tamaño y espaciado de la función ventana. Una vez hecho esto, el segundo paso es llevar a cabo pasos específicos de procesamiento al interior de cada posición de *ventaneo*. El diseño de estos pasos de procesamiento obliga a tomar en cuenta *a priori* las características de los patrones relevantes a nuestro estudio, como lo sugiere Martens (Martens, 1990).

En la elección de escala en la ventana, se puede optar por dos caminos distintos: uno de ellos consiste en fijar el tamaño de la misma realizando el análisis local exhaustivamente, para determinar todos los patrones visuales de interés. La otra forma de atacar el problema es limitando la complejidad del análisis en la ventana, pero determinando subsecuentemente el tamaño ideal que deberá tener la ventana para describir la imagen localmente con la suficiente precisión. Como ya mencionamos en capítulos anteriores, el sistema de visión humana utiliza un principio similar de escala-espacio, lo cual hace aún más atractivo el estudio de la transformada.

La transformada de Hermite trabaja fundamentalmente con derivadas de gaussianas, que han demostrado (Young, 1998) modelar eficientemente operaciones de filtrado en

visión humana con la misma precisión que los multicitados filtros de Gabor y con menos parámetros. El alma de una transformada polinomial, como lo es la de Hermite, es la aproximación polinómica al interior de la ventana local.

Un aspecto que ha limitado la aplicación masiva de la transformada de Hermite es, como se verá, la necesidad de utilizar ventanas traslapadas en su cálculo, hecho que motiva la aparición de más coeficientes transformados que los que se tenían previamente en el dominio espacial.

5.2 Transformadas polinomiales unidimensionales

En la técnica de descomposición de señales llamada transformada polinomial, se aproxima localmente una señal con la ayuda de polinomiales. Este análisis consta de dos pasos. Primeramente, la señal original $L(x)$ se localiza multiplicándola por una señal ventana, $V(x)$. Nótese que es necesario repetir este proceso varias veces en distintas posiciones de ventana para obtener una buena aproximación. Consideraremos el caso en el que el espacio entre ventanas es uniforme (Martens, 1990):

A partir de la función de ventana, $V(x)$, se puede construir una función de ponderación:

$$W(x) = \sum_k V(x - kT) \quad (5.1)$$

con repeticiones periódicas en T . De esta forma, y garantizando que $W(x)$ es diferente de cero para toda x ,

$$L(x) = \frac{1}{W(x)} \sum_k L(x) \cdot V(x - kT) \quad (5.2)$$

de forma que el término interior de la sumatoria tenga siempre información suficiente acerca de la señal original.

El segundo paso consiste en aproximar la fracción de señal "ventaneada" con una polinomial. Para la expansión consideramos, como funciones básicas, las polinomiales $G_n(x)$, de grado n , que son ortonormales con respecto a $V_2(x)$.

Dichas polinomiales se hallan descritas únicamente por $V_2(x)$. Las polinomiales ortonormales para una función de ventana arbitraria $V_2(x)$, están dadas por:

$$G_n(x) = \frac{1}{\sqrt{M_{n-1}M_n}} \begin{pmatrix} c_0 & c_1 & \cdots & c_n \\ c_1 & c_2 & \cdots & c_{n+1} \\ \vdots & \vdots & & \vdots \\ c_{n-1} & c_n & \cdots & c_{2n-1} \\ 1 & x & \cdots & x^n \end{pmatrix} \quad (5.3)$$

donde el determinante M_n está definido así:

$$M_n = c_{i+j}, \quad i, j=0, \dots, n, \quad M_{-1} = 1 \quad (5.4)$$

y

$$c_n = \int_{-\infty}^{+\infty} x^n V^2(x) dx \quad (5.5)$$

es el momento de orden *enésimo*.

Si la función $V(x)$ es par, se pueden desarrollar las siguientes expresiones explícitas para polinomiales ortonormales hasta de orden 3:

$$\begin{aligned} G_0(x) &= \frac{1}{c_0} \\ G_1(x) &= \frac{x}{c_2} \\ G_2(x) &= \frac{(c_0 x^2 - c_2)}{c_0(c_0 c_4 - c_2^2)} \\ G_3(x) &= \frac{(c_2 x^3 - c_4 x)}{c_2(c_2 c_6 - c_4^2)} \end{aligned} \quad (5.6)$$

Es interesante hacer notar que, bajo condiciones generales para la señal original $L(x)$,

$$V(x-kT) \left[L(x) - \sum_{n=0}^{\infty} L_n(kT) \cdot G_n(x-kT) \right] = 0 \quad (5.7)$$

con

$$L_n(kT) = \int_{-\infty}^{+\infty} L(x) \cdot G_n(x-kT) V^2(x-kT) dx \quad (5.8)$$

Las condiciones de que $L(x)$ sea analítica y finita para todo x , son suficientes para garantizar la convergencia de la expansión (5.7) en la mayoría de las funciones ventana convencionales. Además, como es de esperarse, entre más alto sea el grado de expansión de las polinomiales, el error entre la polinomial y la señal original será cada vez más pequeño. De esta forma, la descripción de la señal local $L(x) \cdot V(x-kT)$ puede ser reducido especificando únicamente un conjunto finito de coeficientes polinomiales $L_n(kT)$.

La energía de la señal dentro de la ventana puede ser expresada en términos de los coeficientes de la expansión, de la siguiente forma:

$$\int_{-\infty}^{+\infty} L^2(x) V^2(x-kT) dx = \sum_{n=0}^{\infty} L_n^2(kT) \quad (5.9)$$

La expresión completa, que define la expansión para la señal completa, se deriva de la combinación de (5.2) y (5.7):

$$L(x) = \sum_{n=0}^{\infty} \sum_k L_n(kT) \cdot P_n(x-kT) \quad (5.10)$$

donde

$$P_n(x) = \frac{G_n(x)V(x)}{W(x)} \quad (5.11)$$

Lo que la ecuación (5.8) nos quiere decir es que el conjunto de coeficientes $L_n(kT)$ pueden derivarse de la convolución de la señal $L(x)$ con los *filtros*

$$D_n(x) = G_n(-x)V^2(-x) \quad (5.12)$$

seguido por un muestreo hecho a múltiplos de T . Este mapeo de la señal original $L(x)$ a los coeficientes polinomiales $L_n(kT)$ se llama *transformada polinomial hacia adelante*. El

proceso inverso, es decir, la síntesis de la señal a partir del conjunto de coeficientes, se hace a partir de (5.10) y se llama *transformada polinomial inversa*, o *antittransformada* (fig. 5.2). Dicha antittransformada se reduce a interpolar los coeficientes con la función *patrón*, $P_n(x)$, y llevar a cabo la sumatoria sobre todos los niveles n .

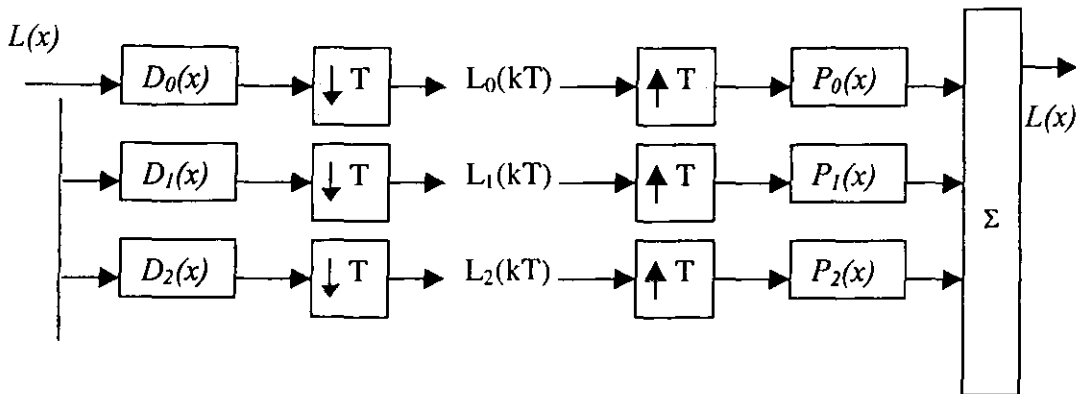


Figura 5.2 Esquema de transformada y antittransformada polinomiales

5.3 Transformada de Hermite unidimensional

A grandes rasgos, la peculiaridad que distingue a la transformada de Hermite de la familia global de las transformadas polinomiales es que la función de ventana local es gaussiana, algo así como esto:

$$V(x) = \frac{1}{\sqrt{x\sigma}} \exp\left(\frac{-x^2}{2\sigma^2}\right) \quad (5.13)$$

donde el factor de normalización es tal que $V^2(x)$ tiene energía unitaria. Las polinomiales ortogonales asociadas a $V^2(x)$ se llaman polinomiales de Hermite. El proceso completo de análisis y descomposición local, por ende, se llama transformada de Hermite.

Hay varios aspectos que hacen a la transformada de Hermite atractiva. Primeramente, la manipulación de ventanas gaussianas provoca que el desarrollo matemático sea asequible y verificable; además, el paso al dominio discreto genera una muy buena aproximación del mundo analógico de Hermite y posee cualidades muy

similares a éste. Otro punto interesante fue verificado en otra sección de esta tesis (cap. 4): cuando las ventanas gaussianas son separadas una de otra por el doble de la desviación estándar, σ , se obtiene un buen modelo de campos receptivos traslapados (Schouten, 1993), ampliamente utilizado en experimentos de percepción y aplicado en distintas facetas de este trabajo. Otra idea favorable es que las funciones de filtrado que están involucradas en la transformada de Hermite son, asimismo, derivadas de gaussianas; ampliamente citadas en trabajos de visión por computadora y de modelado del HVS (Westen, 1995 y otros más).

¿Cómo debe ser la función de ponderación $W(x)$ en la transformada unidimensional de Hermite?. Como dicha función es periódica (periodo T), puede ser expandida en una serie de Fourier (desarrollo en: Martens, 1990):

$$W(x) = \frac{2 \cdot \pi \sigma}{T} w(x) \quad (5.14)$$

con

$$w(x) = 1 + 2 \sum_{k=1}^{\infty} \exp\left[-\frac{1}{2} \left(k \frac{2\pi}{T}\right)^2\right] \cdot \cos\left(k \frac{2\pi x}{T}\right) \quad (5.15)$$

El parámetro de muestreo $\tau = \frac{T}{\sigma}$ es el que determina el contraste de la función de ponderación. Como se desea limitar el número de descomposiciones locales, sobre todo si de codificación se trata, conviene hacer τ lo más grande posible, pero por otro lado, y de acuerdo con (5.2), coincidimos en que $W(x)$ debe tender a ser aproximadamente constante. Se recomienda el uso de valores de τ cercanos a 2, con lo que la función ventana sería aproximadamente constante.

¿Cuáles son las propiedades de los filtros?. Como las funciones de filtro determinan que información se hace explícita en los coeficientes transformados, éstas influyen directamente en el desempeño de la transformada. A partir de (5.12), se llega a:

$$D_n(x) = \frac{(-1)^n}{2^n n!} \cdot \frac{1}{\sigma \pi} H_n\left(\frac{x}{\sigma}\right) e^{-\frac{x^2}{\sigma^2}} \quad (5.16)$$

dado que los polinomios de Hermite $\left\{ H_n\left(\frac{x}{\sigma}\right); n = 0, 1, \dots \right\}$ son ortogonales sobre la ventana gaussiana $V^2(x)$. Es posible demostrar que la función filtro $D_n(x)$ equivale a la derivada de orden enésimo de una gaussiana, de forma que:

$$D_n(x) = \frac{1}{2^n n!} \cdot \frac{d^n}{d\left(\frac{x}{\sigma}\right)^n} \left[\frac{1}{\sigma\sqrt{\pi}} e^{-\frac{x^2}{\sigma^2}} \right] \quad (5.17)$$

Con transformada de Fourier:

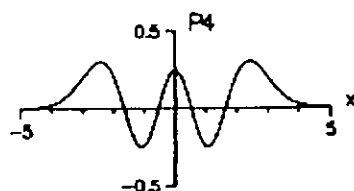
$$d_n(\omega) = \frac{1}{2^n n!} \cdot (j\omega\sigma)^n e^{-\frac{(\omega\sigma)^2}{4}} \quad (5.18)$$

De esta forma, los filtros de orden creciente analizan sucesivamente frecuencias más altas de la señal. Eventualmente, para un orden muy grande, los picos frecuenciales tienden a juntarse mucho, por lo que los filtros subsecuentes darán muy poca información adicional al proceso. Esta es la razón por la que, en la práctica, la transformada de Hermite siempre se limitará a unos cuantos coeficientes.

Las funciones patrón $P_n(x)$, como ya se dijo, son necesarias para resintetizar la señal original a partir de los coeficientes de la transformada de Hermite. Están dadas por las siguientes expresiones analíticas:

$$P_n(x) = \frac{T}{2^n n!} \cdot \frac{1}{\sigma} \frac{1}{2\pi} H_n\left(\frac{x}{\sigma}\right) e^{-\frac{x^2}{2\sigma^2}} w(x) \quad (5.19)$$

recordando que $w(x)$ es la función de ponderación dada en (5.15). A continuación se muestran las funciones de Hermite, con $\sigma=1$, desde P_0 hasta P_4 :



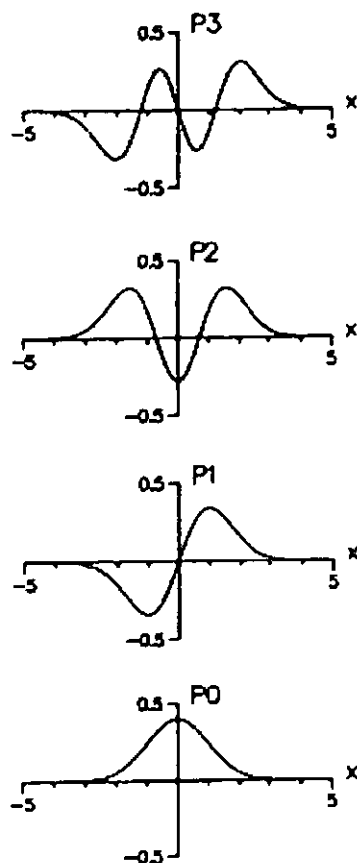


Figura 5.3 Funciones de Hermite con $\sigma=1$

Una propiedad importante de la transformada de Hermite es que es isomórfica con respecto a su transformada de Fourier (Martens, 1990):

$$p_n(\omega) = \frac{T}{2^n n!} \cdot (-j)^n H_n(\omega\sigma) e^{-\frac{(\omega\sigma)^2}{2}} \quad (5.20)$$

El hecho de que estas funciones se parezcan tanto a senos y cosenos “truncados” indica que la transformada de Hermite está fuertemente relacionada con un análisis de armónicas.

Como se postuló anteriormente, la transformada “práctica” de Hermite se limitará frecuentemente a los primeros términos, y como deseamos describir la señal adecuadamente, una selección correcta de σ es vital: por un lado, queremos que σ sea lo más grande posible, porque integrando extensas áreas mejorará nuestra relación señal-ruido y la eficiencia en la representación de la señal. Por otro lado, σ no debe ser muy grande

porque de esta forma la señal no podría ser descrita adecuadamente con los primeros términos de la expansión, necesitaríamos más de ellos, problema similar a la elección de s , el factor de escala, en los campos receptivos del mapeo luminancia-brillantez (cap. 4).

5.4 Transformada de Hermite bidimensional

Si la función ventana de una transformada polinomial bidimensional es separable, es decir,

$$V(x, y) = V(x)V(y)$$

y el látice de muestreo es cuadrado, entonces las funciones de filtro y patrón también son separables, lo que hace muy eficiente su implementación (Martens, 1990). Por ejemplo, los coeficientes polinomiales se obtienen convolucionando una imagen con las funciones de filtro $D_m(x)D_{n-m}(y)$. Donde $D_m(x)$ es la función de filtro para la ventana $V(x)$, seguido de un muestreo en las direcciones horizontal y vertical, en múltiplos de el espacio entre muestras, T .

La transformada de Hermite como tal, se presenta cuando la función ventana es gaussiana. Una gran ventaja de las ventanas gaussianas bidimensionales es que son tanto espacialmente separables como rotacionalmente simétricas. Correspondientemente a éstas, las funciones filtro son separables espacial y polarmente. La transformada de Fourier de $D_m(x)D_{n-m}(y)$, expresada en coordenadas polares $\omega_x = \omega \cos \theta$ y $\omega_y = \omega \sin \theta$, es

$$d_m(\omega_x)d_{n-m}(\omega_y) = g_{m,n-m}(\theta) \cdot d_n(\omega) \quad (5.21)$$

donde $d_n(\omega)$ es la transformada de Fourier de la función de Hermite unidimensional de filtro $D_n(r)$, con r como coordenada radial y

$$g_{m,n-m}(\theta) = \frac{n!}{m!(n-m)!} \cos^m \theta \cdot \sin^{n-m} \theta \quad (5.22)$$

expresa la selectividad direccional del filtro. Así, conforme ascendamos en el orden n del filtro, analizaremos sucesivamente frecuencias radiales más altas (resoluciones espaciales más grandes), análogo al caso unidimensional. Los filtros que sean del mismo orden n y que tengan distinto índice direccional m , harán distinciones entre diferentes orientaciones dentro de la imagen.

Se ha demostrado (Daugman, 1983) que sólo los filtros separables, como los que nos ocupan aquí, tienen la propiedad de dar curvas de ajuste de orientación para patrones unidimensionales distintos, como rejillas, líneas y bordes. Esto significa que estos filtros pueden detectar la orientación de un patrón unidimensional, independientemente de su estructura interna.

5.5 Transformada discreta de Hermite

El equivalente discreto de una ventana gaussiana, como se vio el capítulo anterior, es una ventana binomial. De ahí el nombre de este caso particular de transformada polinomial. Una ventana binomial se define así:

$$V^2(x) = \frac{1}{2^M} C_M^x \quad (5.23)$$

para $x=0, \dots, M$. Los polinomios asociados a esta ventana se llaman polinomios de Krawtchouk.

Como demuestra Martens (Martens, 1990) la transformada discreta de longitud M se aproxima a la transformada análoga de Hermite con $\sigma = M/2$. Es por esto que las propiedades de la transformada discreta pueden ser predichas adecuadamente a partir de sus correspondientes en la transformada análoga.

Si consideramos que M es par, se pueden centrar en el origen las funciones de filtro y de patrón moviendo la ventana binomial sobre $M/2$. Esto da lugar a la siguiente definición discreta de función de filtro de Hermite:

$$D_n(x) = G_n \left(\frac{M}{2} - x \right) \cdot V^2 \left(\frac{M}{2} - x \right) \quad (5.24)$$

para $x = -(M/2), \dots, M/2$. Estas funciones pueden describirse así:

$$D_n \left(\frac{M}{2} - x \right) = \frac{(-1)^n}{2^M \sqrt{C_M^n}} \Delta^n [C_M^x \cdot C_x^n] \quad (5.25)$$

donde

$$(-1)^n \Delta^n L(x) = \sum_{k=0}^n (-1)^k C_n^k L(x+k) \quad (5.26)$$

es el operador en diferencias enésimo. Obteniendo la transformada z de esta función de filtro, se llega a:

$$d_n(z) = \sum_{x=-M/2}^{M/2} D_n(x) z^{-x} = z^{-M/2} \cdot C_M^n \left(\frac{1-z}{2} \right)^n \left(\frac{1+z}{2} \right)^{M-n} \quad (5.27)$$

que, expresada en frecuencias angulares tiene la siguiente forma:

$$d_n(e^{-j\omega}) = C_M^n \left(j \sin \frac{\omega}{2} \right)^n \left(\cos \frac{\omega}{2} \right)^{M-n} \quad (5.28)$$

desde $n=0 \dots M$. Puede observarse que para ω pequeñas, el filtro se reduce a un operador derivativo de orden enésimo, al igual que en el caso análogo.

Los filtros apenas mencionados tienen la gran ventaja práctica de que pueden ser implementados en cascada con la ayuda de los sencillos filtros $z^{-1}(1+z)^2$, $z^{-1}(1-z)(1+z)$, $z^{-1}(1-z)^2$, que tienen kernels $[1 \ 2 \ 1]$, $[-1 \ 0 \ 1]$ y $[1 \ -2 \ 1]$, respectivamente. Así, con excepción del factor de amplificación C_M^n , estos filtros son realizables sin la ayuda de multiplicaciones generales.

5.6 La rotación en la transformada de Hermite

Como ya se explicó, los parámetros “libres” en la descomposición de Hermite a una señal son: la desviación estándar σ de la ventana gaussiana (que define el tamaño de las funciones básicas de Hermite) y la distancia d entre posiciones de ventana adyacentes. Otro aspecto mencionado es que la transformada de Hermite, en la práctica, se limitará a un número N de coeficientes (esto implica que una imagen no podrá ser reconstruida exactamente a partir de la transformación).

El término *steerable filter*¹, introducido por (Freeman, 1991), describe una clase de filtros en la cual un filtro con orientación arbitraria, es sintetizado como una combinación lineal de un conjunto de filtros básicos. Como todos los filtros de Hermite son polinomios multiplicados por una función ventana radialmente simétrica, se puede demostrar (Van Dijk, 1996), que $n+1$ filtros de Hermite de orden n forman una base “girable” para cada filtro individual de orden n . En otras palabras, pueden construirse versiones rotadas de un filtro de orden n con combinaciones lineales de los filtros del mismo orden. La anterior propiedad puede formalizarse así:

$$H_{n,k}^{\theta} = \sum_{m=0}^n c_{k,m}^{(n)}(\theta) H_{n,m} \quad (5.29)$$

donde $H_{n,k}^{\theta}$ ($k=0\dots n$) es el filtro k -ésimo de Hermite, de orden n , rotado θ , y $c_{k,m}^{(n)}$ los coeficientes de rotación. Si se hace adecuadamente el giro de los filtros de Hermite en cada posición, pueden obtenerse representaciones de la imagen adaptadas con respecto a la orientación local.

Para llevar a cabo tal efecto, se puede seguir la siguiente metodología: primeramente, aplicar una transformada cartesiana de Hermite a la imagen original. Basándose en los coeficientes $f_{n-m,m}$ de esta transformada, se genera una imagen de orientaciones en la que se indica la dirección de máxima energía para cada posición de

¹ Refiriéndose a “filtro girable” o “filtro rotable”.

ventana. Posteriormente, esta imagen de orientación sirve como base para "girar" la descomposición cartesiana de Hermite.

La transformada de Hermite entrega representaciones muy eficientes especialmente para patrones locales unidimensionales. Dicha representación consiste simplemente en un parámetro θ , que indica la orientación del patrón, y un pequeño número de coeficientes que indican el perfil del patrón perpendicular a su orientación. Por ejemplo, para un patrón unidimensional con orientación θ , tenemos:

$$f_{n-m,m}^{\theta} = 0 \quad (5.30)$$

para $m > 0$ (y $f_{n-m,m}^{\theta}$ denota el coeficiente m -ésimo de Hermite, de orden n , después de haber rotado θ). Para un patrón de este tipo, el girado sobre θ da lugar a una compactación de energía en los coeficientes $f_{n,0}^{\theta}$, mientras que en todos los demás hay ceros.

Las primeras tres matrices explícitas de rotación (exceptuando la de $N=0$, que da un escalar 1) tendrían la siguiente forma:

$$\begin{pmatrix} s & c \\ c & -s \end{pmatrix}, \begin{pmatrix} s^2 & 2sc & c^2 \\ sc & c^2 - s^2 & -sc \\ c^2 & -2sc & s^2 \end{pmatrix}, \begin{pmatrix} s^3 & 3s^2c & 3sc^2 & c^3 \\ s^2c & 2sc^2 - s^3 & c^3 - 2s^2c & -sc^2 \\ sc^2 & -2s^2c + c^3 & s^3 - 2sc^2 & s^2c \\ s^3 & -3s^2c & 3sc^2 & -c^3 \end{pmatrix} \quad (5.31)$$

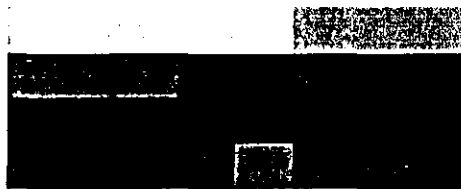
donde $c = \cos(\theta)$ y $s = \sin(\theta)$, para $0 \leq \theta \leq \pi$

Una descripción detallada y completa del proceso de rotación hermitiano, de la obtención de las matrices de rotación y su desenvolvimiento matemático a partir de las funciones binomiales generalizadas se halla en (Silván-Cárdenas, 2001).

A continuación se muestran los coeficientes de la transformación de Hermite discreta de la siguiente imagen. A la izquierda, el bloque transformado sin rotar. A la derecha, el bloque rotado (con la energía compactada sobre el eje vertical) y su ángulo de rotación θ .

Como se verá en el siguiente apartado, las propiedades energéticas de la transformada discreta de Hermite permitirán establecer una clasificación precisa de la

orientación local de cada sector de la imagen. Los posibles valores que arroje dicha clasificación serán de un patrón no-dimensional (región prácticamente uniforme), patrón unidimensional (región con orientación de clase similar al visto en la figura anterior) y patrones bidimensionales (texturas en general). Los patrones bidimensionales tienen que ser representados en ambas dimensiones, ya que los efectos visuales de cuantizaciones severas en este tipo de estructuras son fácilmente percibidos por el ojo humano.



| | | | | | | | |
|-------|-------|-------|-------|-------|-------|-------|-------|
| DC | 4.97 | -0.14 | 0.41 | 0.31 | -0.79 | -0.33 | 1.46 |
| 40.79 | 1.43 | 0.66 | -0.19 | -0.22 | 0.11 | 0.06 | 0.07 |
| -5.98 | 0.37 | -0.43 | -0.31 | -0.02 | -0.06 | 0.45 | -0.21 |
| -5.65 | 0.98 | -0.91 | 0.07 | -0.03 | -0.10 | 0.04 | -0.06 |
| 7.85 | -1.51 | -0.82 | 0.62 | -0.59 | -0.24 | -0.49 | 0.39 |
| -1.25 | -3.14 | 0.98 | 0.35 | -0.07 | 0.31 | 0 | -0.55 |
| -6.66 | 1.47 | 1.83 | -0.61 | -0.93 | 0.25 | 0.48 | -0.07 |
| 3.92 | 3.46 | 0.74 | -0.58 | -1.09 | -0.24 | 0.32 | 1.02 |

| | | | | | | | |
|-------|---|---|---|---|---|---|---|
| DC | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 41.09 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| -5.65 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| -5.43 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 7.84 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| -1.65 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| -7.32 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 4.25 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |

Ángulo de rotación, θ : 1.45 [rad] (aprox. 83°)

Figura 5.4 La imagen presenta una fuerte orientación unidimensional. La matriz de la izquierda contiene los coeficientes transformados sin rotar, la de la derecha rotados y concentrados.

5.7 La clasificación de estructuras con la transformada de Hermite

Como ya se ha venido tratando, el proceso de transformación y rotación hermitiano desemboca en la clasificación estructural de patrones. El tipo de estructura estará determinado principalmente por la cantidad de información necesaria (y mínima, en condiciones óptimas) para representar al bloque. Las regiones de la imagen que son evidentemente uniformes, son consideradas regiones 0-D o no-dimensionales, y para representarlas utilizaremos sólo el coeficiente de DC. Las regiones que contengan una estructura fuertemente orientada (líneas, bordes, etc.) necesitarán, posteriormente a la rotación de los coeficientes, únicamente del vector vertical que contiene la energía orientada. Estas estructuras serán 1-D o unidimensionales. El resto de los bloques (texturas, regiones heterogéneas, esquinas, ruido, etc.) serán clasificados como 2-D o bidimensionales.

Los tres descriptores empleados para construir el criterio de clasificación son los siguientes (Silván-Cárdenas, 2001). Note que G es la matriz de coeficientes transformados Hermitianos:

$$1. \text{ Luminancia media: } L = G_{0,0} \quad (5.32)$$

$$2. \text{ Contraste: } C = \left[\sum_{i=0}^N \sum_{j=0}^N G_{i,j}^2 - L^2 \right]^{1/2} \quad (5.33)$$

$$3. \text{ Contraste residual unidimensional: } \Delta C = \left[C^2 - \sum_{i=1}^N G_{i,0}^2 \right]^{1/2} \quad (5.34)$$

El primer paso en la clasificación consiste en dividir las estructuras 0-D de las otras dos. Esto puede lograrse analizando el contraste local. Para esto, aplicamos la siguiente regla de decisión:

$$C \begin{matrix} <^{0D} \\ >_{1D \cup 2D} \end{matrix} k_0 C_{thr}(L) \quad (5.35)$$

la curva $C_{thr}(L)$ es la función *light adaptation threshold*², propuesta por Barlow (Barlow, 1972), misma que es aproximada por (Silván-Cárdenas, 2001) de la siguiente forma:

$$C_{thr} = C_{min} + \frac{L^\alpha - L_{min}^\alpha}{L^\alpha + L_{min}^\alpha} \quad (5.36)$$

con $0 \leq \alpha \leq 1$ (generalmente 0.6) y C_{min} es el contraste mínimo en el nivel de luminancia L_{min} (el nivel donde el ojo tiene máxima sensibilidad al contraste)

Este umbral determina el límite a partir del cual el contraste local se vuelve visualmente relevante.

Una vez que hemos separado los patrones no-dimensionales, diferenciaremos las estructuras unidimensionales de las bidimensionales ayudados del contraste residual con la siguiente regla:

$$\Delta C \begin{cases} < H_1 \\ > H_2 \end{cases} k_1 \Delta C_{thr}(L, C) \quad (5.37)$$

La curva ΔC_{thr} es la función *contrast masking threshold*³, trabajada ampliamente por (Legge, 1980). En el caso ideal para un patrón unidimensional, toda la energía se halla concentrada a lo largo de la dirección del gradiente y el contraste residual es nulo, aunque esto se presenta en muy pocas ocasiones. Legge define la función de la siguiente forma:

$$\Delta C_{thr} = \max(C_{thr}, C^\beta C_{thr}^{1-\beta}) \quad (5.38)$$

donde $0 \leq \beta \leq 1$ (típicamente $\beta = 0.7$)

La experimentación realizada con la clasificación de estructuras explicada aquí, será presentada con amplitud en el séptimo capítulo, dedicado a la exposición de los resultados.

² Umbral de adaptación a la luz

³ Umbral de enmascaramiento del contraste, o sea, la reducción de visibilidad de un componente en una imagen debido a la presencia de otro.

Capítulo VI

Descripción del Sistema de Compresión

Los tópicos tratados a continuación son, propiamente, el ensamblaje de la teoría expuesta en los cuatro capítulos pasados. Como se mencionó brevemente en la introducción, el sexto capítulo explica detalladamente el “armado” del codificador, los procesos que intervienen en la comunicación entre sus módulos, las metodologías de programación empleadas, los lenguajes que sirvieron de plataforma a la construcción del modelo y las transformaciones que sufren los datos a lo largo de toda la línea de proceso. Es importante detallar las dificultades que se presentaron al elaborar las funciones, y las soluciones que fueron puestas en marcha para aliviarlas. El proyecto, en este capítulo, será analizado desde el punto de vista de codificación y decodificación a niveles de implementación.

6.1 Vista general del Sistema

El sistema de procesamiento de imágenes presentado en esta tesis cumple con las características tradicionales de codificación-decodificación, fácilmente representables así: (fig. 6.1):

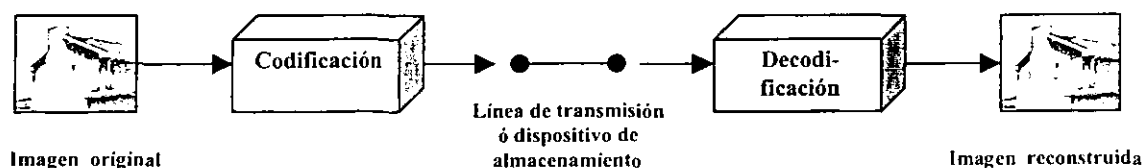


Fig. 6.1 Esquema tradicional de codificación - decodificación

De esta forma, y como es natural, limitamos nuestro sistema de codificación-decodificación a cierto tipo de señales bidimensionales: las imágenes digitales

La codificación y decodificación serán llevadas a cabo en un equipo de cómputo, y los módulos encargados de ambas tareas fueron diseñados e implementados en el lenguaje matemático MatLab (versión 5.3). Dicha elección fue tomada, principalmente, por las enormes ventajas que supone la modificación y adecuación, en ciertos puntos, de algunas rutinas empleadas en proyectos anteriores en la DEPMI, y por la facilidad que presenta el lenguaje al manejo matemático de señales discretas bidimensionales. Otro argumento importante a favor de MatLab es la amplia bibliografía que existe con proyectos experimentados en él, y la aparición regular de módulos de herramientas aplicables al procesamiento digital de señales y al tratamiento matemático matricial. Además, gran parte del proyecto desarrollado como tesis de licenciatura (Moreno, 1999), y que ha sido de mucha utilidad en algunas partes de este trabajo, fue escrito en MatLab. Algunas desventajas derivadas de esta elección fueron el alto consumo de tiempo que MatLab emplea en ciertos procesos (muchos de los cuales hubo que optimizar, para mejorar ese aspecto), la administración de la memoria y pequeñas dificultades en el manejo de datos en bajo nivel (a nivel de bits, en la escritura de los archivos, principalmente). Sin embargo, estas desventajas no fueron insalvables y no radicaban en el funcionamiento intrínseco del proyecto, por lo que continuamos desarrollándolo en este ambiente de programación.

El objetivo inicial proyectado para el sistema era la codificación de imágenes, con fines, esencialmente, de compresión. Había, sobre todo, que dar especial énfasis a la utilización de los campos de brillantez en el proceso. No fue tocado el tema de transmisión, por lo que en el marco del esquema tradicional (fig. 6.1) la línea de transmisión es reemplazada por un dispositivo de almacenamiento, en el que estará guardada la imagen comprimida, para después ser decodificada en el módulo correspondiente.

El grueso de la teoría aplicada al proyecto y de la complejidad intrínseca del sistema se halla en el proceso de codificación, del que hablaremos a detalle en el siguiente subcapítulo. Otro de ellos será dedicado a la decodificación. Se anotarán, en cada subsección, algunos detalles relativos a la programación de los módulos y las particularizaciones teóricas empleadas en su construcción. En la parte final del capítulo se incluye una breve descripción de la interfaz de usuario que se construyó, con la explicación de sus funciones

básicas. Dicha interfaz facilita la manipulación del sistema a nivel de usuario, evitando la interacción directa con las funciones en las líneas de comando del lenguaje y con las múltiples variables de trabajo utilizadas, haciendo más agradable y natural el uso del codificador y del decodificador.

6.2 El Codificador

El codificador también puede ser explicado en un esquema de bloques, como se muestra a continuación: (fig. 6.2)

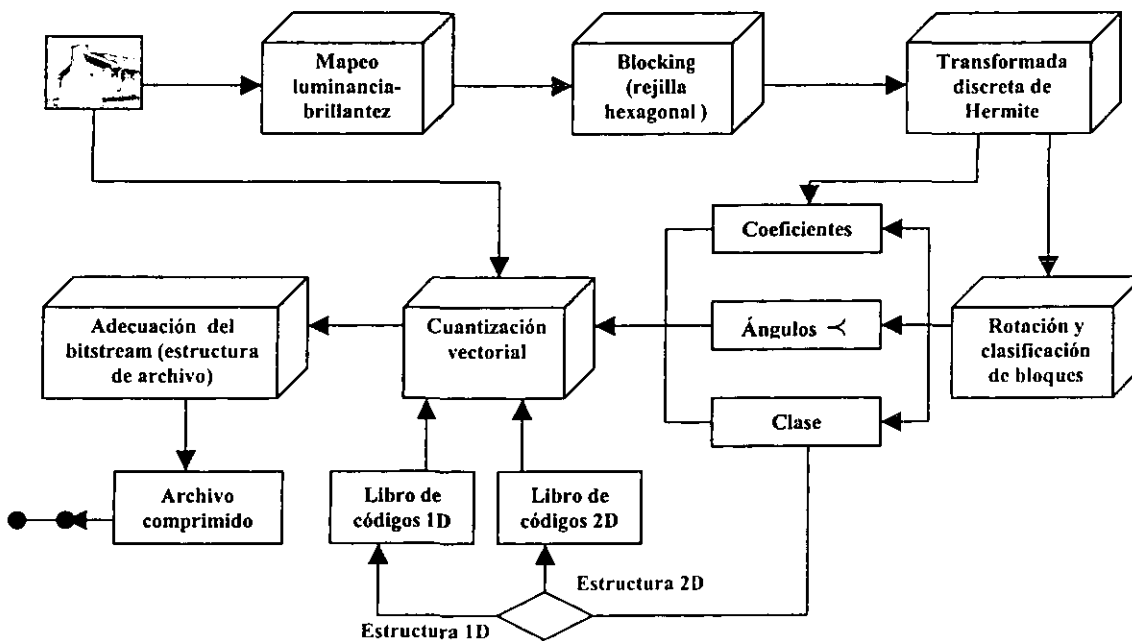


Fig. 6.2 Esquema en bloques del proceso de codificación, desde la entrada de la imagen hasta la escritura del archivo comprimido en el medio de almacenamiento

El sistema aceptará como entrada una imagen digital, a la que hará ciertas revisiones (y en su caso, modificaciones) pre-procesatorias¹ para evaluar si es factible su codificación.

¹ Estas evaluaciones comprenden procesos como la obtención de los datos de un archivo, normalización a uno (para imágenes en niveles enteros de luminancia), adecuación de tamaños (para matrices de dimensiones que

En caso que sea así, se lleva a cabo el mapeo de la imagen en luminancias al dominio de la brillantez. Hay que poner un énfasis especial en este proceso. El mapeo luminancia-brillantez se compone de varios módulos encargados de distintas tareas, referidas a la teoría escrita en el capítulo 4, en especial el desarrollo bidimensional, explicado en la sección 4.4. Se utilizó, para el cálculo de los mapas, el esquema en ocho niveles de resolución y campos receptivos gaussianos con valor de $\mu = 2$ (relación radial entre campos receptivos centrales y periféricos). Primeramente se obtienen los distintos niveles resolutivos, se ensambla, y del mapa de ensamblaje se aplica el “ajuste local” detallado con antelación (cap. 4) para obtener finalmente el mapa de brillantez.

Se procura que la salida de este bloque sea de las mismas dimensiones espaciales que la imagen original, pero está representada en un dominio numérico distinto. Una vez obtenido el mapa de brillantez, trataremos de explotar al máximo sus cualidades estructurales y su inherente compresión: podremos comenzar entonces su transformación de energía. Primeramente, se somete a la matriz a la división en bloques. Esta “bloquificación” será llevada a cabo en un látice con un parámetro de muestreo $T=(N+1)/2$ y un defase de T entre columnas y filas nones y pares, como se muestra en la (fig. 6.3). Esta estructura fue adoptada por la similitud que presenta con la morfología de los campos receptivos visuales (ver cap. 2), sin introducir mucha redundancia y por su fácil traducción al mundo digital, con el empleo del muestreo mencionado. El producto de esta selección es un conjunto de campos visuales que a la vista asemejan hexágonos. La calidad en reproducción de imágenes en las que se acopla este modelo es visualmente muy superior a las que han sido muestreadas con bloques cuadráticos adyacentes (como JPEG), ya que en las fronteras entre bloques aparecen efectos no deseados y artefactos fácilmente visibles, sobre todo a altas tasas de compresión, que son sensiblemente disminuidos con el empleo de los *panales* hexagonales (fig. 6.4). Esto tiene que ver directamente con la estructura visual de ciertos insectos, como las abejas, que emplean unidades visuales básicas de formas poligonales.

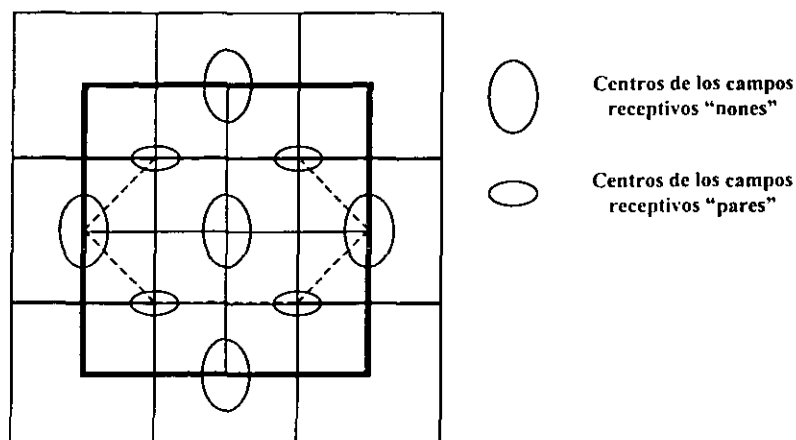


Figura 6.3 Muestreo del lattice, simplificado usando sólo dos "juegos" de rejillas: los nones y los pares, que generarán, en conjunto, estructuras de tipo hexagonal – véase la línea falsa (punteada).

Por motivos de coherencia con el esquema de cuantización vectorial, mismos que se explicarán más adelante, la dimensión del muestreo está fija con $N=7$. Esto determina tanto el tamaño de los vectores utilizados ($N+1 = 8$, $N+1 = 8$) como la separación de las muestras.

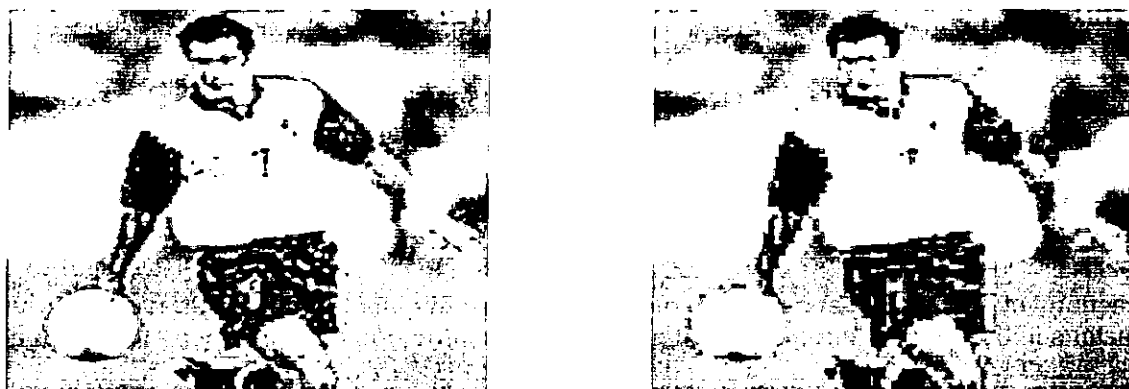


Figura 6.4 La imagen de la izquierda es producto de una decodificación con estructura hexagonal. En la de la derecha fue aplicada la tradicional estructura cuadrangular de bloques adyacentes. Note la diferencia en calidad visual. Ambas imágenes fueron severamente cuantizadas

La estructura arrojada por esta partición será transformada al dominio discreto de Hermite, que como ya se mencionó, es un caso particular de transformación polinomial

discreta en la que se utilizan ventanas binomiales. La transformación se lleva a cabo con un algoritmo rápido aplicado al látice explicado con anterioridad. El cálculo no se hará en un esquema multi-resolución, ya que las aplicaciones que este tipo de transformación tienen no son aplicables al esquema propuesto. La energía hermitiana resultante está “empaquetada” y, dependiendo de el tipo de matriz con la que se esté trabajando en ese momento, puede estar fuertemente orientada (revisar cap. 5), lo que la asignará a algún tipo de estructura (0D, 1D o 2D) en el bloque de “Rotación y Clasificación”. Se sugiere fuertemente al lector revisar la teoría asociada a este bloque de decisión, desarrollada en la sección (5.6) de esta tesis. Como se observa, dicho bloque tiene 3 salidas: una de ellas, los coeficientes rotados, otra, el ángulo que fue aplicado para rotar los coeficientes, y por último, el tipo de bloque del que se trata. Hemos descubierto que la determinación de tipos de estructura en el dominio de la brillantez puede llegar a ser más exacta y más confiable que en el dominio tradicional de la luminancia. Como se mostrará explícitamente en el último capítulo, los mapas de brillantez, al menos en cuestión estructural, tienden a “esconder” aquellos sectores donde la información perceptual no es muy relevante, aunque haya energía presente, mientras que suelen hacer más evidentes ciertas zonas en las imágenes con información perceptual presente, que podrían ser descartadas, o consideradas no-dimensionales, en el dominio de la luminancia.

El paso central del sistema de codificación es la cuantización vectorial, en la que se expondrán detalles más adelante. Éste bloque recibe las matrices recién transformadas, rotadas y clasificadas, y de acuerdo a la estructura que recibe en cada momento, utilizará el libro de códigos unidimensional (con vectores fuertemente tendientes a bordes, esquinas y fronteras de intensidad) o el libro de códigos bidimensional (con rica información textural). Las estructuras no-dimensionales (0D) no requieren un tratamiento vectorial, como se verá después. El cuantizador vectorial arrojará las matrices de etiquetas, a las que se podrá aplicar una sencilla adecuación de canal (codificación de canal) para poder escribir un archivo binario eficientemente, y ahorrar algo de bits en el proceso.

6.2.1 El proceso de transformación, rotación y clasificación hermitiano

El algoritmo de transformación hermitiana discreta recibe como entrada la imagen digital, muestreada con el régimen hexagonal, y arroja como salida un par de matrices, correspondientes a los coeficientes transformados de los campos receptivos nones y pares. A partir de este momento, y por esta razón, el tratamiento de la información transformada vendrá por “duplicado”. Otro algoritmo independiente, cuyo fundamento matemático se explicó con detalles en el capítulo pasado, se encarga de hacer dos tareas a ese par de matrices: primeramente, rotar cada campo receptivo de forma que su energía esté mayormente concentrada en el eje vertical² (ya que siempre habrá una rotación angular que dé mejores resultados que las demás), y después identificar el tipo de bloque que se está trabajando (clasificación de patrones) para etiquetarlo con esa condición. El proceso de identificación de clase sigue las políticas de energía direccional citadas en el capítulo 5.

La imagen cuya estructura espacial servirá de base para la codificación es la que posee sus valores dentro del dominio de la brillantez. Uno de los objetivos de este trabajo, como ya se dijo, es precisamente demostrar que la elección estructural con fines de compresión y representación de imágenes es más adecuada en el dominio de la brillantez que en el de la luminancia. También se mostrará que la construcción de libros de códigos para cuantizadores vectoriales en imágenes es más efectiva trabajando en este dominio. Sin embargo, si se desea mantener la cantidad de energía estable en todo el proceso de codificación, se tendrán que someter al cuantizador vectorial los coeficientes transformados en el dominio de la luminancia, y no los correspondientes a la brillantez³. De cualquier forma, se mostrarán en el siguiente capítulo resultados comparativos de la aplicación de ambos métodos.

Al final del proceso de rotación y clasificación, tendremos las matrices con los coeficientes rotados, el ángulo de rotación θ que se aplicó a cada matriz, y el tipo de estructura con la que se está lidiando. Ésta tarea se esquematiza a en la figura siguiente (fig. 6.5):

² Se tomó como referencia para alinear la energía el eje vertical, pero no existe razón alguna para preferirlo al horizontal. Simplemente se cambiaría el tratamiento matricial para tal efecto.

³ Si se utilizan los coeficientes en brillantez, la imagen reconstruida tendrá que ser adecuada para su observación, ajustando sus valores extremos a la escala de representación que se desee usar. Recordemos que el dominio de la brillantez es, por definición, un dominio comprimido.

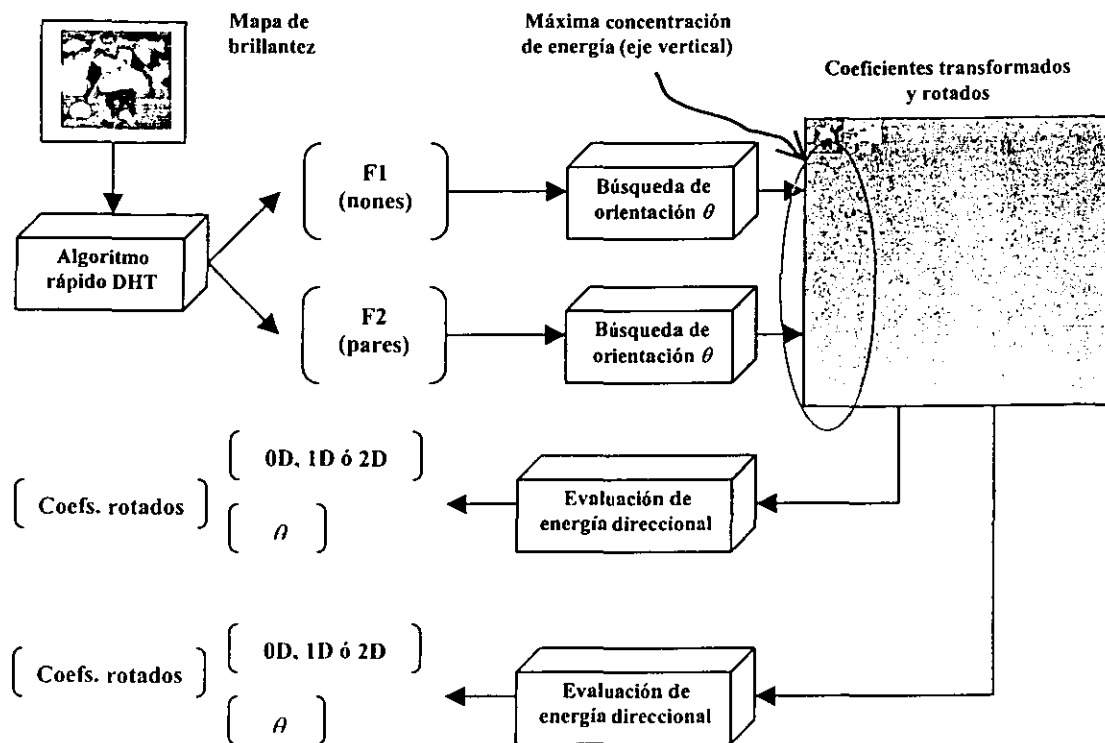


Figura 6.5 Sub-proceso de transformación, rotación y determinación de estructura local. Al final de dichos cálculos, tendremos un par de conjuntos de coeficientes rotados, con sus correspondientes ángulos de rotación y tipo de estructura asociada

Si no nos preocupara la determinación estructural, no tendría sentido la utilización de diversos libros de código. Esto se hace para optimizar, de acuerdo a cada tipo de estructura, la elección del mejor vector posible dentro del grupo. Entre mejor entrenado esté el libro de códigos correspondiente, más próxima a la realidad será la representación del bloque, como veremos en el siguiente sub-tema.

En el dominio de brillantez, transformado y rotado, la energía se halla sumamente concentrada y orientada, esta es una de las razones por las que la generación de libros de código con secuencias de entrenamiento de este tipo son muy eficientes. La entropía es menor y no se tienen que construir libros de código enormes para establecer un potente esquema de cuantización vectorial: simplemente hay que entrenar bien, con imágenes ricas en texturas, en estructuras orientadas, con varios tipos de regiones uniformes y con gran variedad de escenas. Observe la interesante gráfica que viene a continuación (fig. 6.6):

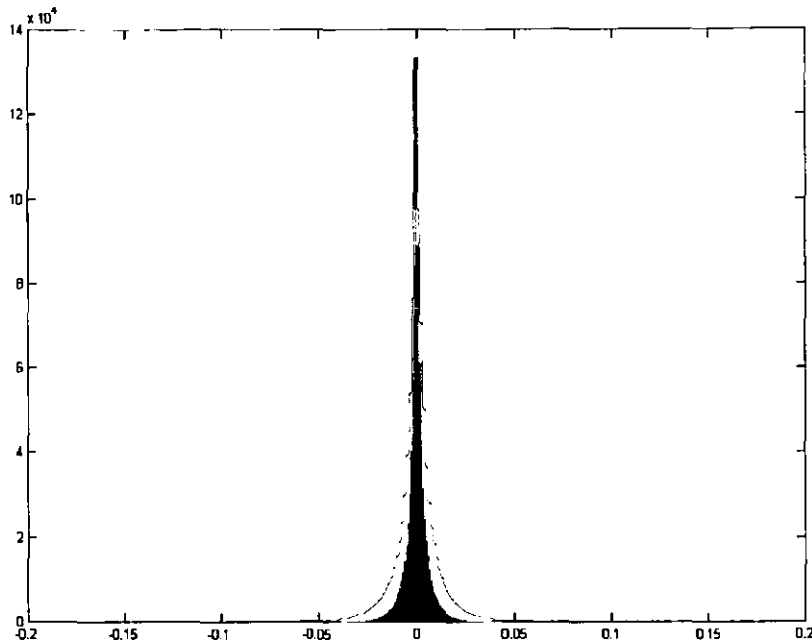


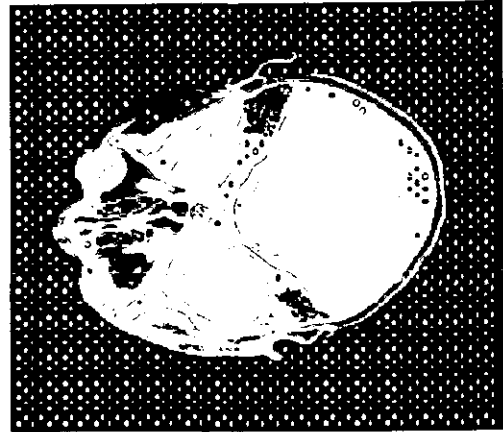
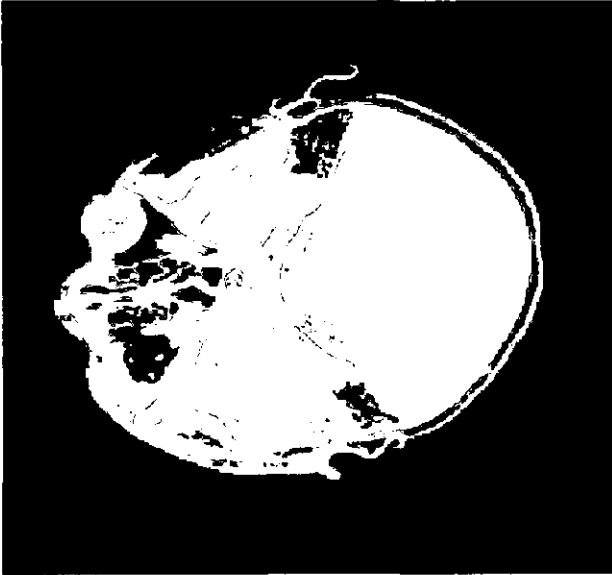
Figura 6.6. Histograma comparativo de la energía en los coeficientes de dos secuencias de entrenamiento: en rojo, la secuencia basada en luminancias. En azul, la secuencia basada en brillantez. Nótese la mayor dispersión en la gráfica roja.

La distribución de energía en los coeficientes de los libros de código construidos con estas secuencias, tienen correspondencia topológica con la gráfica mostrada.

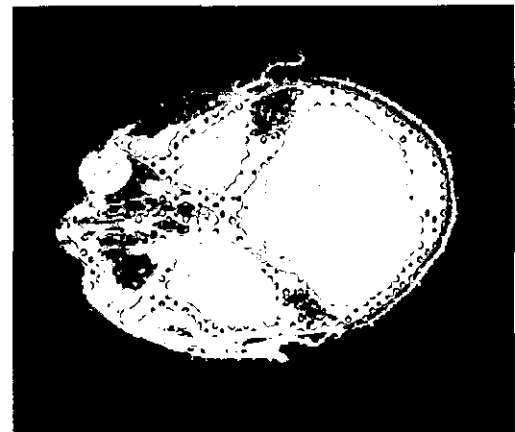
La correcta clasificación de las regiones es vital para el éxito de la codificación, ya sea en luminancia o en brillantez. Un ejemplo de clasificación de patrones para una imagen real se muestra en el esquema siguiente (fig 6.7). En ella puede observarse cómo las zonas homogéneas son asignadas a sectores no-dimensionales, las zonas que presentan orientaciones fuertes como bordes, esquinas o líneas son clasificadas como unidimensionales, mientras que algunas regiones que tienen fuerte información textural, no tan “orientable” como las anteriores, son clasificadas como bidimensionales

La importancia de una buena clasificación radica en que ése será el factor determinante para el uso de uno u otro libro de códigos, la aparición de zonas no-dimensionales donde la compresión perceptiva sea alta y el realce de zonas perceptivas importantes, que deberán tener mayor atención por parte del cuantizador vectorial, hallándoles un vector que esté en alta correspondencia visual con ellas. Creo que éste es el camino para que las imágenes finales sean agradables al ojo.

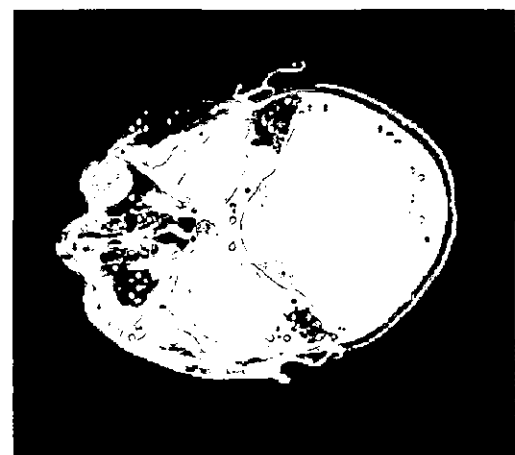
Original



0D



1D



2D

Figura 6.7 La imagen original (izquierda) es transformada, rotada y clasificada. En las tres imágenes de la derecha se sobrepone pequeños círculos de color en las zonas estructurales correspondientes (0D figura de arriba, 1D en medio, 2D figura de abajo)

6.2.2 Cuantización vectorial

El corazón de la etapa de codificación es la cuantización vectorial. Será el proceso encargado de determinar, dentro de un conjunto limitado de vectores bien seleccionados (libros de código), la mejor sustitución –en términos de distorsión- del vector que se desea codificar. En el proceso de cuantización citado aquí haremos uso tanto de cuantizadores vectoriales como cuantizadores escalares. Paso a detallar.

Como se había mencionado anteriormente, el sistema contará para la codificación con la ayuda de dos secuencias de entrenamiento. Una de ellas que servirá para la sustitución de los vectores unidimensionales (patrones con estructura 1D) y el otro para la sustitución de vectores bidimensionales (texturas, patrones con estructuras 2D). Se notó, por otro lado, que la distorsión resulta mínima si hacemos para los patrones no-dimensionales una simple cuantización escalar de su coeficiente más significativo, el coeficiente de *DC*. La siguiente tabla compara la energía, en porcentajes, de los coeficientes de AC para distintos tipos de estructuras típicas (tabla 6.1):

A) Típico bloque no-dimensional. Nótese la altísima concentración de energía en el coeficiente de DC, para una transformación hermitiana. No hay un patrón claro de disposición de energía en los coeficientes AC:

(Porcentajes de energía):

| | | | | | | | |
|--------|--------|--------|--------|--------|--------|--------|--------|
| 0.9258 | 0.0004 | 0.0025 | 0.0025 | 0.0025 | 0.0034 | 0.0007 | 0.0033 |
| 0.0007 | 0.0010 | 0.0007 | 0.0012 | 0.0014 | 0.0014 | 0.0016 | 0.0004 |
| 0.0003 | 0.0012 | 0.0008 | 0.0008 | 0.0000 | 0.0013 | 0.0010 | 0.0018 |
| 0.0015 | 0.0024 | 0.0006 | 0.0022 | 0.0007 | 0.0021 | 0.0007 | 0.0009 |
| 0.0000 | 0.0006 | 0.0017 | 0.0007 | 0.0007 | 0.0002 | 0.0006 | 0.0001 |
| 0.0020 | 0.0017 | 0.0007 | 0.0018 | 0.0002 | 0.0019 | 0.0003 | 0.0016 |
| 0.0000 | 0.0009 | 0.0013 | 0.0002 | 0.0007 | 0.0003 | 0.0006 | 0.0005 |
| 0.0022 | 0.0018 | 0.0005 | 0.0020 | 0.0006 | 0.0025 | 0.0003 | 0.0030 |

B) Bloque unidimensional con patrones verticales en la imagen. Nótese que hay menor concentración de energía en el coeficiente de DC, además de que hay energía asociada a la dirección vertical.



(Porcentajes de energía):

| | | | | | | | |
|--------|--------|--------|--------|--------|--------|--------|--------|
| 0.6580 | 0.0640 | 0.0106 | 0.0363 | 0.0297 | 0.0036 | 0.0345 | 0.0004 |
| 0.0155 | 0.0036 | 0.0049 | 0.0059 | 0.0027 | 0.0042 | 0.0070 | 0.0043 |
| 0.0015 | 0.0001 | 0.0000 | 0.0010 | 0.0046 | 0.0006 | 0.0054 | 0.0043 |
| 0.0079 | 0.0068 | 0.0010 | 0.0069 | 0.0014 | 0.0067 | 0.0018 | 0.0026 |
| 0.0030 | 0.0037 | 0.0016 | 0.0029 | 0.0013 | 0.0006 | 0.0030 | 0.0012 |

| | | | | | | | |
|--------|--------|--------|--------|--------|--------|--------|--------|
| 0.0028 | 0.0025 | 0.0010 | 0.0022 | 0.0021 | 0.0036 | 0.0007 | 0.0017 |
| 0.0046 | 0.0032 | 0.0016 | 0.0015 | 0.0003 | 0.0011 | 0.0017 | 0.0002 |
| 0.0045 | 0.0015 | 0.0021 | 0.0005 | 0.0023 | 0.0005 | 0.0016 | 0.0006 |

C) Bloque textural, o bidimensional. No hay orientaciones unidimensionales claras, pero hay cierta distribución complicada de energía. El coeficiente de DC tiene un poco menos de concentración de energía que en el caso anterior, y la energía restante no tiende a concentrarse en ninguna dirección clara.



(Porcentajes de energía):

| | | | | | | | |
|--------|--------|--------|--------|--------|--------|--------|--------|
| 0.5054 | 0.0199 | 0.0042 | 0.0044 | 0.0046 | 0.0009 | 0.0073 | 0.0058 |
| 0.0206 | 0.0126 | 0.0245 | 0.0051 | 0.0058 | 0.0084 | 0.0042 | 0.0030 |
| 0.0823 | 0.0155 | 0.0165 | 0.0049 | 0.0026 | 0.0042 | 0.0029 | 0.0033 |
| 0.0289 | 0.0068 | 0.0206 | 0.0011 | 0.0033 | 0.0059 | 0.0023 | 0.0059 |
| 0.0320 | 0.0000 | 0.0039 | 0.0056 | 0.0031 | 0.0049 | 0.0016 | 0.0011 |
| 0.0115 | 0.0102 | 0.0127 | 0.0013 | 0.0042 | 0.0032 | 0.0017 | 0.0010 |
| 0.0068 | 0.0067 | 0.0000 | 0.0079 | 0.0026 | 0.0058 | 0.0031 | 0.0027 |
| 0.0030 | 0.0027 | 0.0009 | 0.0041 | 0.0025 | 0.0002 | 0.0027 | 0.0067 |

Tabla 6.1 Ejemplos de concentración de energía en bloques transformados 0D, 1D y 2D

La cuantización escalar, además de hacerse para los coeficientes de DC en las estructuras no-dimensionales, se lleva a cabo también para las familias de ángulos de rotación en las estructuras unidimensionales y bidimensionales. Los algoritmos encargados de llevar a cabo las cuantizaciones fueron optimizados para que su rendimiento con matrices de tamaño fijo sea alto. No se incorporó ningún procedimiento heurístico a la búsqueda de vectores en el libro de códigos, sino que se escribió un programa de búsqueda exhaustiva con algunas rutinas inmersas que mejoran la eficiencia en la misma. Esta elección se tomó debido a que no se puede correr el riesgo, con un algoritmo heurístico de búsqueda, de no llegar a la mejor opción, sino a un mínimo local. Eso bajaría la calidad de la imagen reconstruida y va en contra de las políticas del proyecto, que apuntan a la evaluación e implementación de un método, y no tanto a la eficiencia en términos de tiempo de ejecución.

En la mayor parte de los casos, la cuantización escalar de los coeficientes de DC se hace a 8 bits, aunque con 7 la distorsión percibida es prácticamente imperceptible (perdiendo sólo el bit menos significativo) y para los ángulos se ha fijado la cuantización a 8 niveles angulares (3 bits), como lo muestra la figura siguiente (fig. 6.8).

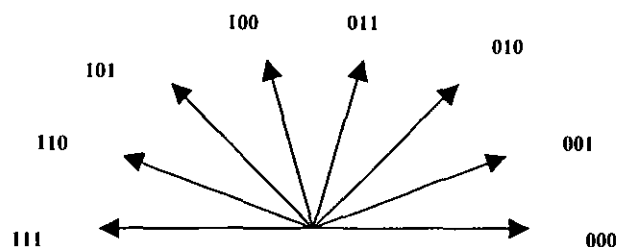


Fig. 6.8 Cuantizador escalar uniforme de 8 niveles para ángulos entre 0 y π [rad]

Resulta poco importante la ganancia cualitativa si se asignan más bits a la cuantización escalar angular, pero un bit menos sí restaría sensiblemente calidad a la reproducción, por lo que se decidió dejar fijo este valor.

Para llevar a cabo una buena cuantización vectorial, es esencial contar con un libro de códigos adecuado y generado de forma confiable. Hemos adoptado (y adaptado) el algoritmo LGB (Linde, 1980) como constructor de los libros de códigos debido a su confiabilidad, a la cantidad de proyectos que lo involucran y a su practicidad. Es relativamente sencillo de programar, y un listado bien diseñado dará lugar a un algoritmo aceptablemente rápido. El algoritmo tuvo que ser adaptado, ya que se tendrán al final no uno, sino dos libros de códigos correspondientes a las estructuras unidimensionales y bidimensionales, respectivamente. La secuencia de entrenamiento debe ser lo suficientemente grande y rica como para alimentar diversos tipos de patrones al libro, y el libro de códigos debe ser lo suficientemente grande para alcanzar un buen nivel de calidad SNR, aunque esto, como es de esperarse, disminuye la tasa de compresión del sistema. La dimensión básica de los vectores de trabajo es 8×8 . Esto por varias causas: el cálculo rápido y adecuado de coeficientes hermitianos, la estandarización de dimensiones vectoriales en la cuantización, por representar bloques de imágenes lo suficientemente grandes como para determinar estructuras espaciales y frecuenciales relevantes, y al mismo tiempo ser lo suficientemente chicos para no entorpecer y recargar los procesos de cómputo. Asimismo, las tasas de compresión y la relación compresión/calidad resultantes son favorables. Consideramos que libros de códigos de 2,048 elementos (direccionables con 11 bits), o libros de 4,096 elementos (direccionables con 12 bits) son una buena solución a esta relación calidad-compresión. En realidad, un libro de códigos bien obtenido no es una carga

pesada, al menos en cuestión de almacenamiento, al codificador y al decodificador: un libro de códigos de 2,048 elementos cuyos vectores sean unidimensionales (de dimensiones 7×1), y que cada elemento de los vectores tenga una gran resolución (punto flotante de 64 bits), nos dará un archivo final de:

$$(2,048 \times 7 \times 64) / 8 = 115 \text{ Kb}$$

Como ya se supondrá, la clasificación de los vectores de las secuencias de entrenamiento se lleva a cabo a priori, con objeto de determinar a cuál de los módulos LGB se dirigen.

Los vectores que componen el libro de códigos unidimensional final son de dimensiones 7×1 (sólo se están considerando los coeficientes rotados de AC en la dirección de máxima energía), por lo que se tienen arreglos de 7 datos; mientras que los vectores de estructuras de dos dimensiones medirán 8×8 , descartando el coeficiente de DC en ambos casos, lo que da un vector de 63 datos. El esquema que se muestra a continuación (fig. 6.9) explica los procesos de entrenamiento, generación de libros de códigos y morfología de los mismos.

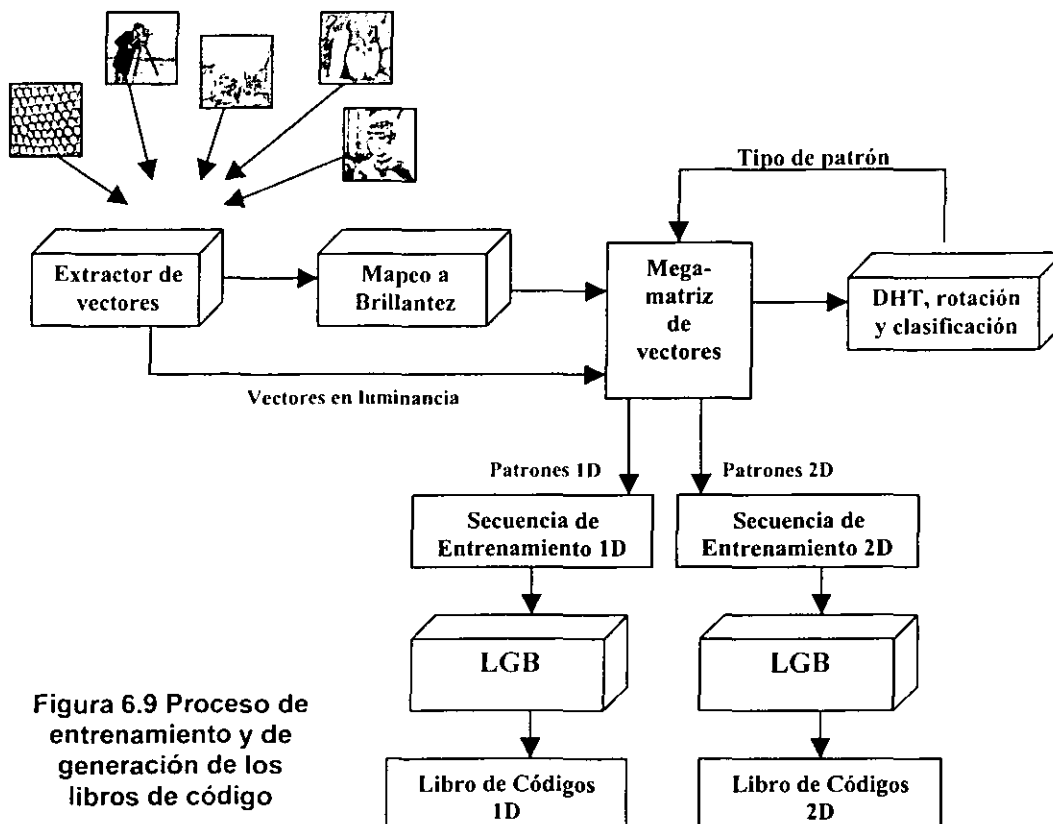


Figura 6.9 Proceso de entrenamiento y de generación de los libros de código

Hay que mencionar en este momento un dato importante. Se experimentó también con un esquema de generación de libros de código obtenidos exclusivamente del dominio de la brillantez, es decir, anulando el flujo de información etiquetada como “Vectores en luminancia” del esquema. Esto significa que tanto la clasificación de patrones como los patrones en sí están tomados de esta fuente. En el esquema inicial, la clasificación parte de los vectores mapeados a brillantez, pero la distribución energética corre a cargo de los valores en luminancia. Los libros de códigos, como es natural, deberán estar presente tanto en el dispositivo codificador como en el decodificador.

Similar a la generación de los libros de código, el proceso de cuantización vectorial podrá trabajar en dos modalidades, de acuerdo al carácter de sus libros. Si éstos fueron obtenidos en el dominio de la luminancia, los vectores de entrada al CV deberán estar definidos en este dominio, y en caso de que hayan sido mapeados a brillantez, los vectores para computar la cuantización también deberán de haber sido mapeados. Los resultados obtenidos con la utilización de ambos estilos serán reportadas en el siguiente capítulo de esta tesis.

6.2.3 La escritura del archivo

A la salida del proceso de cuantización, tanto escalar como vectorial, tendremos un conjunto de matrices disponibles para escribir el archivo comprimido: la matriz con la información de todos los ángulos de rotación, la matriz con los identificadores de tipo de patrón, y la matriz con las etiquetas asignadas a los coeficientes transformados y rotados. Estas etiquetas son las que representan a algún vector del libro de código. De este amplio conjunto de datos escribiremos un archivo ordenado que sea eficiente en su estructura y que sea decodificable sin ambigüedades, simplemente teniendo a la mano los mismos libros de códigos. Como consecuencia del muestreo hexagonal, tendremos las matrices mencionadas arriba, pero por duplicado (por el efecto de entramado explicado en la figura 6.3).

Como se puede constatar, todos nuestros datos consisten de números enteros. Esto facilita la escritura del archivo y nos obliga a fijar la resolución de los datos en enteros, sin tener que lidiar con punto flotante.

El archivo consistirá de palabras secuenciales, cada una de las cuales tendrá la siguiente estructura (fig. 6.10):

| Clase 1 | Componente de DC 1 | Ángulo 1 | Etiqueta de LC 1 | Clase 2 | Componente de DC 2 | Ángulo 2 | Etiqueta de LC 2 |
|---------|--------------------|----------|--------------------------|---------|--------------------|----------|--------------------------|
| 2 bits | 7 bits | 3 bits | $\log(LC1)/\log(2)$ bits | 2 bits | 7 bits | 3 bits | $\log(LC2)/\log(2)$ bits |

Figura 6.10 Estructura de la palabra en el archivo codificado. En la parte de abajo se hallan los tamaños, en bits, de cada sección. Nótese que el tamaño de las secciones "etiquetas" dependen del tamaño de los libros de código utilizados

A la estructura propuesta pueden hacerse varias modificaciones que derivarán en un uso más eficiente del espacio. Una de ellas es la de especificar en el decodificador, que al aparecer, en el campo de clase, el valor cero, no lea más que el componente de DC y que continúe al siguiente grupo de datos.

La ecuación que dictaría el tamaño final del archivo, en bits, previa a cualquier tipo de compresión entrópica de canal (que podría verse como una modificación futura al sistema), es la siguiente:

$$T = (2) \times (9) \times (E0) + (12 + LC1) \times (E1) + (12 + LC2) \times (E2) \quad (6.1)$$

donde,

E0: Cantidad de estructuras no-dimensionales en la imagen

E1: Cantidad de estructuras unidimensionales en la imagen

E2: Cantidad de estructuras bidimensionales en la imagen

LC1: Cantidad de bits asignados a cada etiqueta del libro de códigos unidimensional

LC2: Cantidad de bits asignados a cada etiqueta del libro de códigos bidimensional

De esta forma, un proceso de codificación que haya arrojado, para una imagen de 512 x 512 [pix] a 8 [bpp], 2000 bloques 0-D, 1000 bloques 1-D y 1096 bloques 2-D, con libros de código bidimensional de 2048 elementos y unidimensional de 1024 elementos, cuantizando a 7 bits el coeficiente de DC y a 3 bits los ángulos, tendrá como tamaño final (sin ningún tipo de compresión entrópica):

$$2 \times (2000(9) + 1000(12+10) + 1096(12+11)) = 130,416 \text{ bits} = 16.302 \text{ bytes}$$

lo que produce una tasa de compresión de

$$(512[\text{pix}] \times 512[\text{pix}] \times 1[\text{bytepp}]) / (16,302[\text{bytes}]) = 16.1$$

6.3 El Decodificador

Una vez que la imagen ha sido procesada y almacenada propiamente en un archivo binario, el decodificador tiene la importante tarea de “volver a la vida” a la imagen, reconstruyendo sus componentes a partir del archivo binario y llevando a cabo las operaciones pertinentes de referencia en libros de código, antirotación, antitransformación y en caso de requerirlo, ajustes previos (de los que hablaremos más adelante) a la visualización final de la imagen reconstruida. A grandes rasgos, la tarea de la codificación incluye los procesos utilizados por el codificador pero ordenados de forma inversa. Siempre habrá algunos detalles que complicarán un poco su estructuración, pero el resultado deberá ser una imagen digital cuya calidad dependerá de la rigidez de la compresión a la que se haya sometido su original. El esquema global del decodificador luciría de la siguiente forma (fig. 6.11):

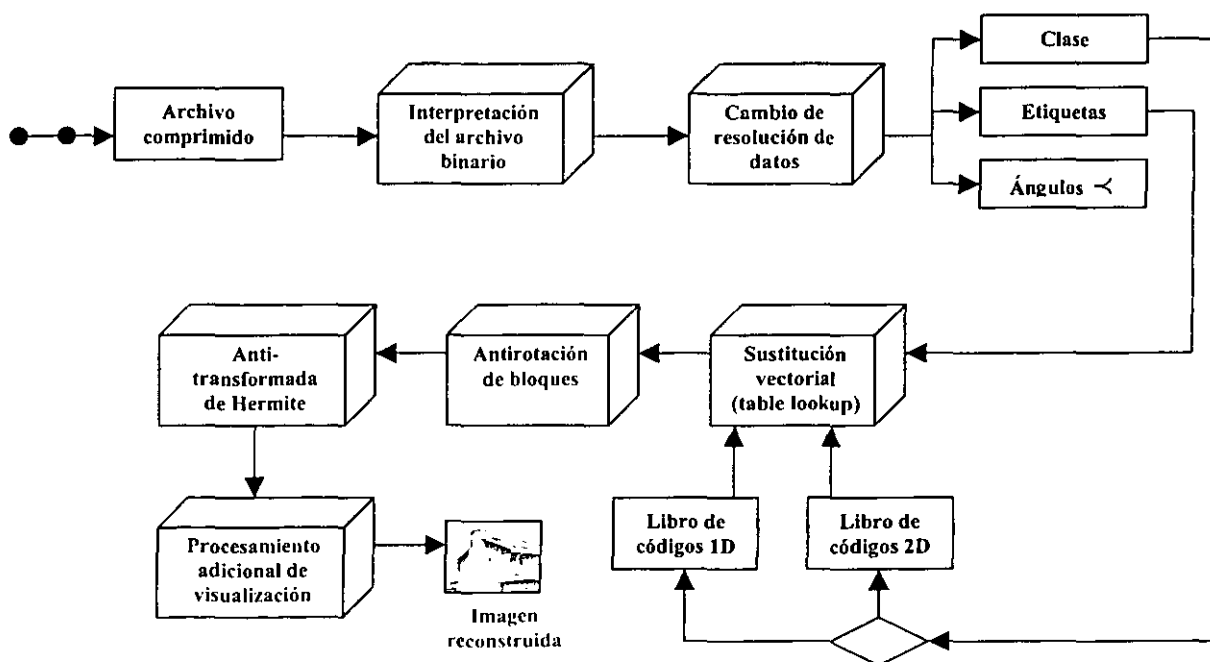


Figura 6.11 Esquema general del decodificador

El archivo tiene que ser re-interpretado. Para esto, se recurre a la estructura de las palabras propuestas en la sección anterior, y se regeneran varios grupos de matrices que contendrán los datos con la nueva estructura. Hay que mencionar que se cambia la resolución de los mismos, que eran enteros en el archivo y MatLab los adopta como entidades *dobles*. Así, las matrices que se tendrán a la mano en el espacio de trabajo serán las de etiquetas, las de clase de patrón y las de ángulos. Con las matrices de etiqueta se lleva a cabo una simple sustitución *look-up*⁴ en los vectores de los libros de códigos correspondientes. El libro de código a utilizar en cada caso estará dictado por el tipo de estructura mencionado en la matriz de clasificación. Ya con los coeficientes originales, y con las matrices de ángulo, se rota la distribución de energía un ángulo exactamente inverso al de codificación para tener las matrices hermitianas sin rotar, que pueden ser antitransformadas rápida y eficientemente para devolver los datos al dominio espacial, que como se explicó en el proceso de codificación, puede ser en luminancias o en la brillantez. En este punto es en el que la imagen decodificada finalmente puede ser vista por el usuario.

Las modificaciones adicionales que pueden ser hechas a la imagen antes de ser vista por el usuario están dirigidas sobre todo a aquellas que la reconstrucción refiera a valores en brillantez. Como ya se ha mencionado, el dominio de la brillantez es, por definición, un dominio comprimido, comparado al de la luminancia. Para lograr una versión agradable visualmente de la imagen, habrá que adecuar su distribución de intensidades a una escala convencional, por ejemplo, la de grises de 256 tonos (8 bits por píxel). Otro tipo de procesamientos agregados se requieren cuando la imagen, por efectos de filtrado en el mapeo de brillantez, presenta algunos píxeles (generalmente en los límites de la imagen) con valores afuera del rango estable normalizado (0 a 1). En este caso se evalúa, de acuerdo a la vecindad del píxel, que éste adopte un valor distinto permitido. La cantidad de píxeles que se hallan en estos casos es ínfima (menor del 0.01% en casi todos los casos), pero su conversión es necesaria para cuestiones de visualización de la escala de tonalidad, principalmente.

⁴ La operación look-up se hace sustituyendo el índice representativo (etiqueta) por su correspondiente vector, o matriz, del libro de códigos.

6.4 La Interfaz de usuario

Para mayor facilidad de manipulación del codificador, se diseñaron tres interfaces gráfica dentro del ambiente de programación MatLab llamadas, respectivamente, *gui_bvq*, *gui_open_bvq* y *gui_lc_bvq*. La primera interfaz agrupa con algunos “botones de comando” todas las tareas necesarias para codificar una imagen: manipulación de la imagen original, visualización de su configuración estructural, elección de los libros de código y elección del tipo de cuantización vectorial a realizar. Asimismo, hay ventanas en las que se despliega la información correspondiente al proceso (dimensiones y nombre de la imagen original, tasa de bits por píxel, número de vectores en los libros de código, tasa de compresión alcanzada en la codificación, tamaño del archivo comprimido y relación señal a ruido final).

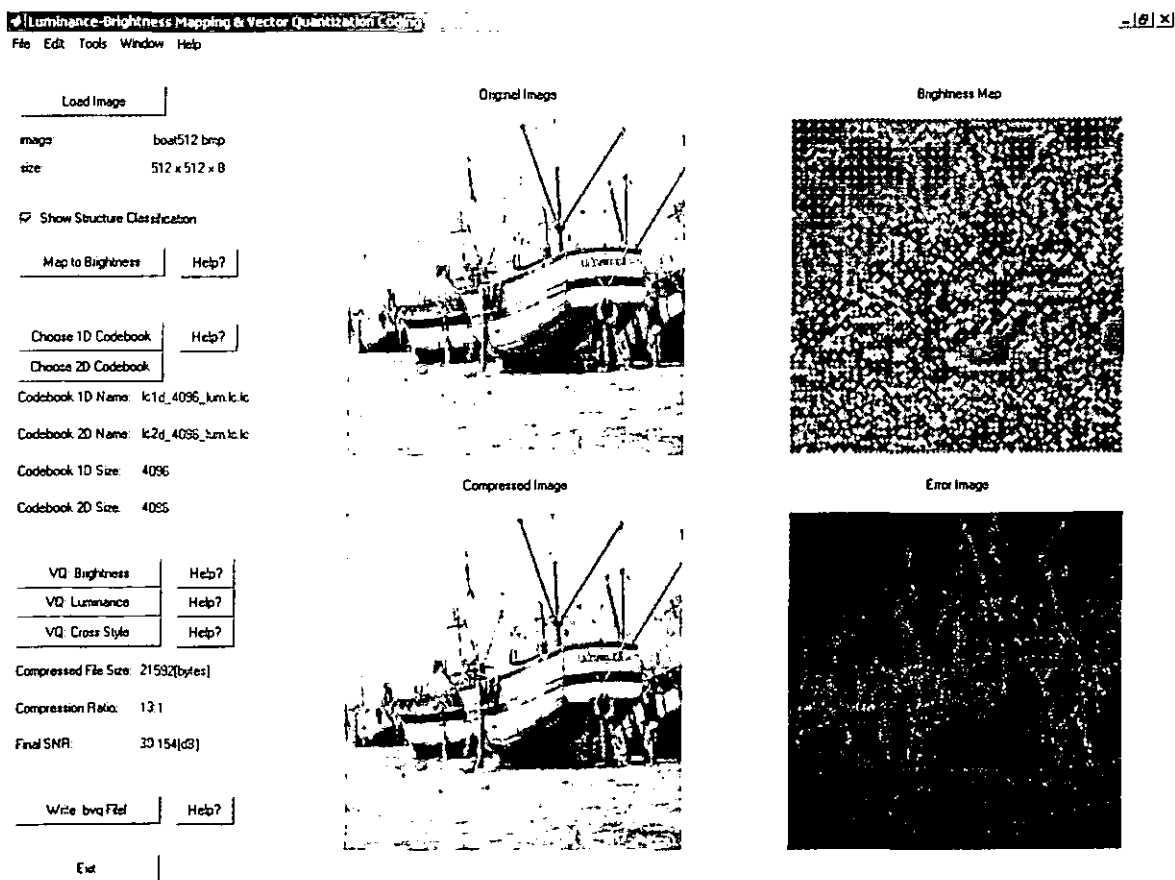


Figura 6.12 Vista general de la interfaz principal de usuario en funcionamiento

Se ofrecen en la interfaz los tres tipos de cuantización vectorial tratados en este trabajo de tesis: compresión en el dominio de la brillantez, en el dominio de la luminancia y en modo de “cruz”.

Se incluye, además, un grupo de botones de ayuda para orientar al usuario al momento de elegir un estilo de cuantización vectorial. En la página anterior se muestra una vista general de la interfaz de usuario (fig. 6.12).

La interfaz permite, a manera de conclusión, la escritura del archivo binario que contendrá la imagen comprimida, referenciada con la extensión *.bvq*. Este archivo binario de imagen podrá ser leído, para mayor facilidad, con la interfaz de decodificación que se muestra a continuación (fig. 6.13). En dicha interfaz, llamada *gui_open_bvq*, sólo se permiten las tareas de decodificación y despliegue en función de los libros de código dados.

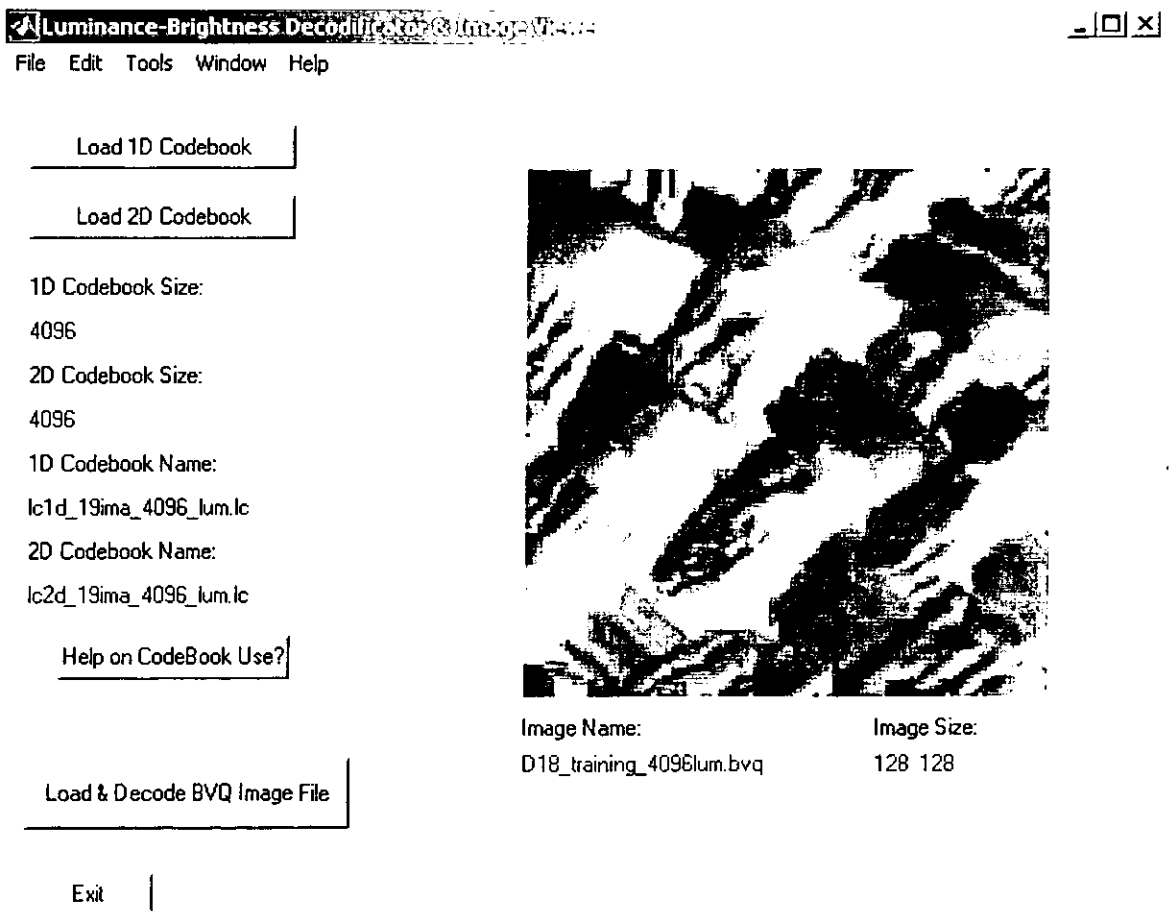


Figura 6.13 Interfaz para decodificación y despliegue de archivos de imagen BVQ

La tercera interfaz que se creó (*gui_lc_bvq*) tiene como objetivo primordial apoyar el entrenamiento de libros de códigos. Permite seleccionar hasta 20 imágenes como secuencia de entrenamiento y activa las funciones necesarias para obtener las secuencias de entrenamiento adecuadas para generar dichos libros de códigos, con las dimensiones y características deseadas. Éstos libros de códigos podrán ser posteriormente utilizados en los procesos de codificación y decodificación de imágenes .bvq. A continuación se muestra la vista general de la interfaz de entrenamiento: (fig. 6.14)

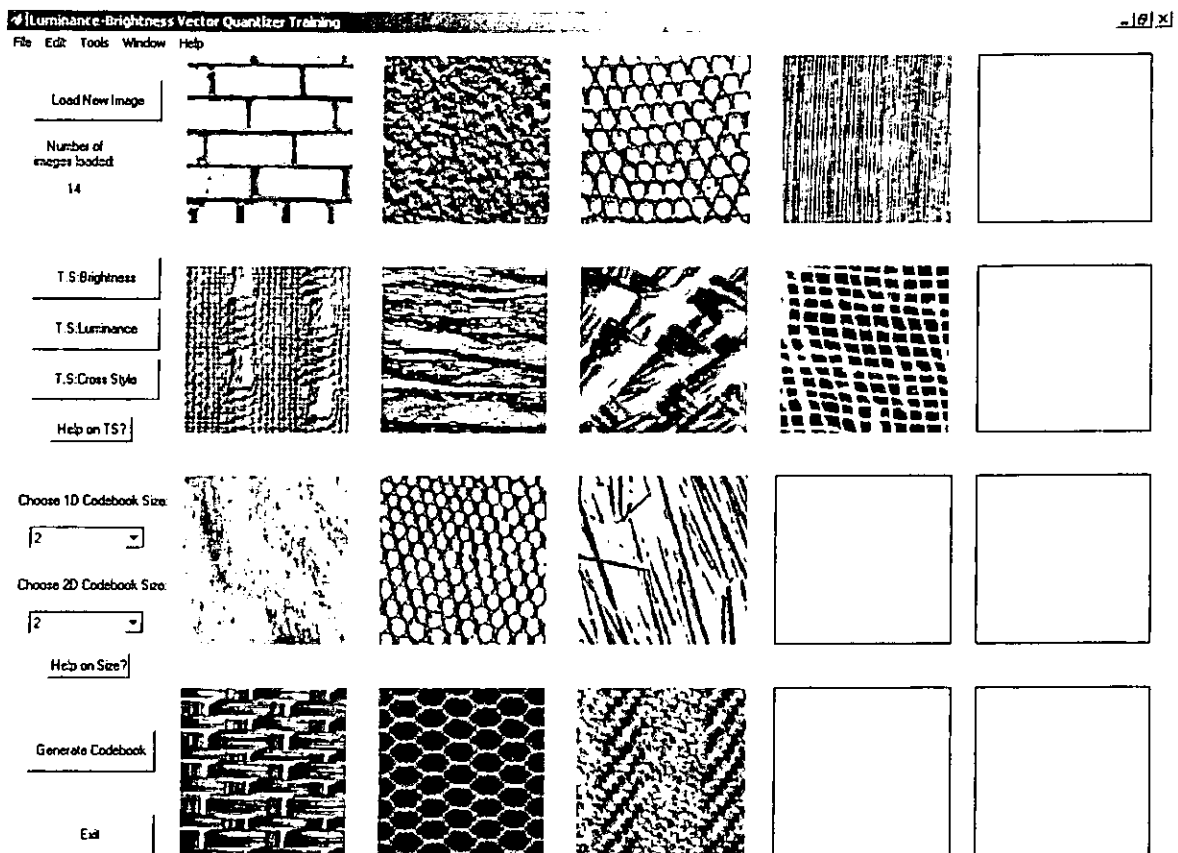


Figura 6.14 Vista general de la interfaz de entrenamiento para libros de código

Capítulo VII

Resultados y Conclusiones

El séptimo y último capítulo de esta tesis está dedicado a la exposición de resultados, a la presentación de los logros obtenidos (y las tareas inconclusas) así como a la explicación de las conclusiones derivadas del trabajo. Gran parte del tiempo dedicado a la realización de la tesis significó experimentación en el laboratorio; y es importante mencionar esto ya que en muchas ocasiones determinados resultados prácticos no correspondían exactamente con la expectativa que la teoría asociada había generado, por lo que había que revisar los métodos, desarrollos y procedimientos que me habían llevado a obtenerlos. Trataré de exponer algunas de las vías de investigación que se indagaron y enfatizaré en aquellas que nos dieron los mejores resultados, tanto visuales como analíticos.

Se mostrarán tablas y esquemas comparativos del desempeño de los métodos entre sí, y se enfrentarán con algoritmos conocidos y estandarizados (JPEG, por ejemplo), para ubicar en su reales dimensiones los resultados obtenidos.

Será importante también mencionar las posibles adiciones y/o modificaciones que se pueden llevar a cabo al presente trabajo, que como toda tesis, es perfectible. Resultaría harto interesante hacer un seguimiento permanente de esta poco explorada línea de investigación, los mapas de brillantez, (cosa que pudimos constatar en las revisiones bibliográficas) para continuar el desarrollo de proyectos que abran nuevas rutas de trabajo y que puedan generar resultados útiles.

7.1 Diferencias estructurales en campos de brillantez

Uno de los resultados que saltó a primera vista fue la gran diferencia, en cuanto a clasificación de patrones, que dieron los mapas de brillantez transformados, comparados con las imágenes transformadas en el dominio de la luminancia. Creo que las nuevas estructuras encontradas son de mucha ayuda para lograr una codificación psico-visual más correcta y más acertada. Los campos de brillantez, como se ve en las siguientes figuras

(figs. 7.1 y 7.2), obligan a poner más atención en algunas zonas que antes eran no-dimensionales, y las ubica en cualquiera de las dos clases, dependiendo de su importancia perceptiva.



LUMINANCIA

- Estructuras 0D
- Estructuras 1D
- Estructuras 2D



BRILLANTEZ

Figura 7.1 Diferencias en la clasificación de patrones. Arriba, clasificación de la imagen original "Venice" en luminancias. Abajo, clasificación del mapa de brillantez de "Venice". Los puntos rojos corresponden a estructuras 0-D, los azules a 1-D y los verdes a 2-D.



Figura 7.2 Obsérvese el mismo fenómeno que en la figura anterior, en esta ocasión para un detalle de la imagen "Goldhill", tanto en luminancia (arriba) como en mapa brillantez (abajo).

De la misma forma, resta importancia a regiones que eran consideradas texturales u orientadas, clasificándolas simplemente como no-dimensionales. Dependiendo de la imagen, la tasa de compresión obtenida puede variar de una modalidad a otro, en función de la cantidad de cambios estructurales que se presenten.

Más adelante se mostrarán imágenes codificadas de ambas formas, atendiendo a la clasificación estructural del dominio de la luminancia y en mapas de brillantez, con objeto de poder determinar cómo afecta esta serie variaciones en el producto final.

7.2 Decodificación de imágenes y calidad de reproducción

Sin duda uno de los parámetros de evaluación que tiene mayor peso es el de la calidad de reproducción de la imagen decodificada. Todo el proceso, largo y complejo, no tiene mayor importancia si el resultado final, la imagen que el usuario verá, no es suficientemente satisfactoria, al menos para la tasa de compresión que se desea, que es el otro gran aspecto evaluativo de esta sección del capítulo.

Las imágenes que se presentarán a continuación son las que el decodificador ha obtenido, por cualquiera de dos procesos detallados en la tesis. A saber: codificación en luminancia, con libros de códigos obtenidos en el dominio de la luminancia y codificación *cruzada*, con valores transformados desde el dominio de la luminancia pero con datos estructurales obtenidos desde el dominio de la brillantez.

Los dos grupos de imágenes que se muestran (figs. 7.3 y 7.4), contienen tanto la imagen original como sus dos decodificaciones, así como una imagen JPEG. Más adelante tendremos la evaluación numérica (objetiva) de su calidad. Las imágenes codificadas con los métodos propuestos en esta tesis fueron comprimidas utilizando los libros de códigos (el unidimensional y el bidimensional) de 2,048 elementos cada uno. Se hará una comparación con una imagen comprimida a la misma tasa, con el método JPEG. Las imágenes (*a*, *b*, *c* y *d*) son, respectivamente, la original, la comprimida en el dominio de la luminancia, la comprimida en cruz y la JPEG. Las tasas de compresión son 0.6 bpp tanto para nuestras imágenes como para la JPEG, sin aplicar ningún tipo de compresión entrópica en nuestros dos productos.

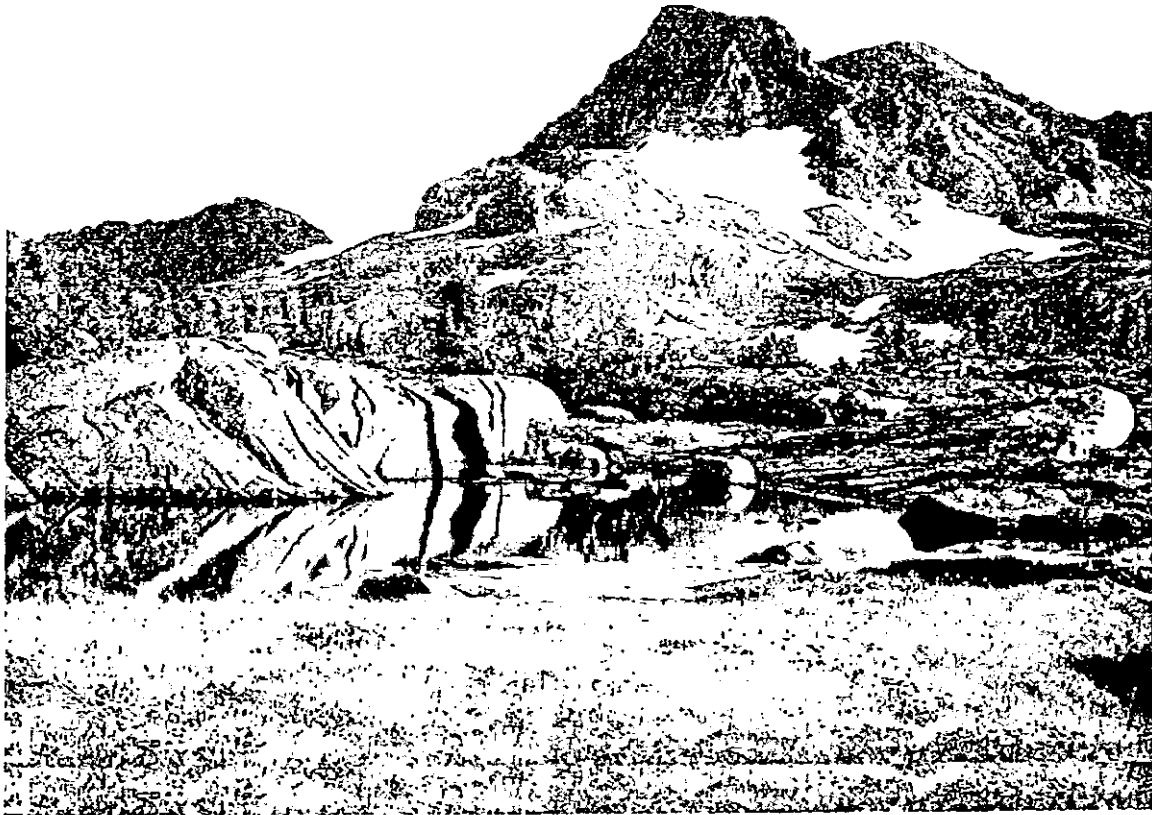
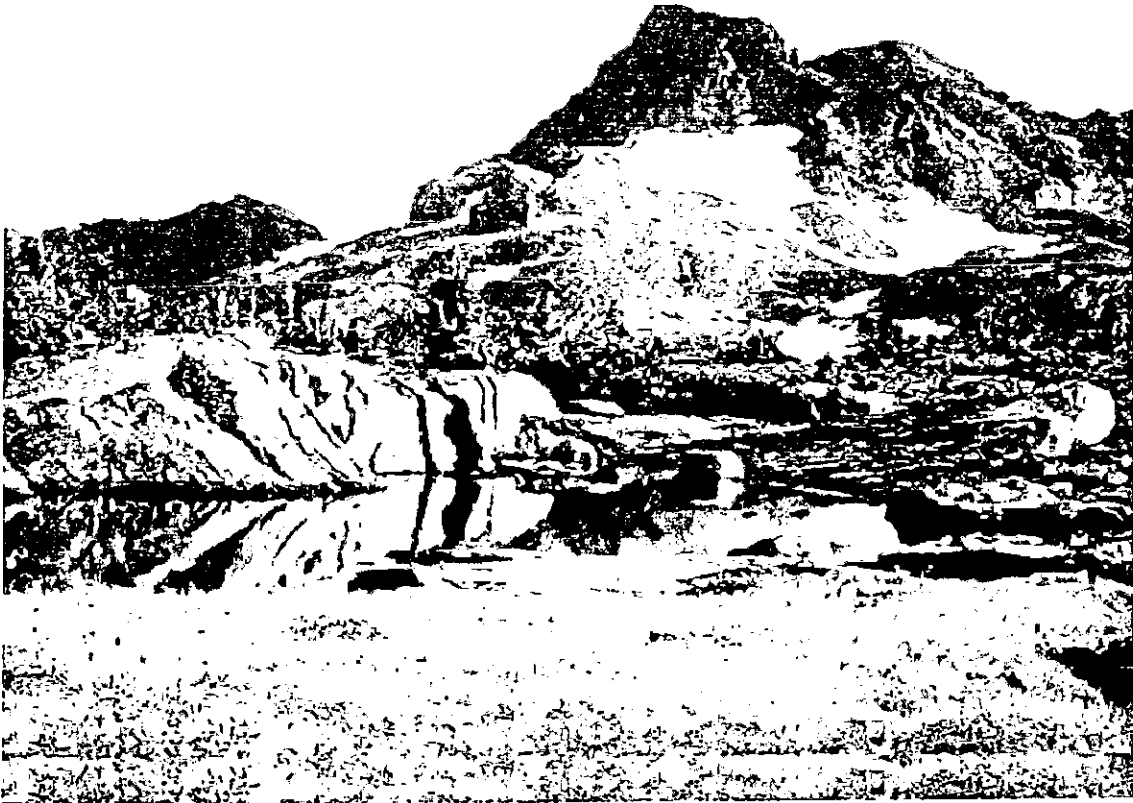


Figura 7.3 a) Imagen original (arriba), b) imagen comprimida con método en luminancias (abajo). Tasa de bits de 0.6 bpp



c) Imagen comprimida con método en cruz (arriba), d) imagen JPEG de misma compresión (abajo). Tasa de bits de 0.6 bpp

Los libros de códigos utilizados en el proceso de compresión de las presentes imágenes fueron entrenados con secuencias de entrenamiento obtenidas a partir de 20 imágenes de muy distintas clases, dentro de las cuales no se hallaba la que se codificó.

En las imágenes anteriores se puede cuantificar, al menos subjetivamente, la calidad de las imágenes obtenidas. Las diferencias entre las dos imágenes producidas por los métodos en luminancia y en cruz, son muy sutiles y se expresan sobre todo en las estructuras finas, es decir, en la clasificación estructural. La compresión es relativamente alta (0.6 bpp, o lo que es lo mismo, un ratio de 1:13.5) para observar diferencias notables con la compresión JPEG.

El método JPEG, como la mayoría de los basados en transformadas, tiende a introducir artefactos desagradables desde tasas medias de compresión. Otro efecto no deseado es la "bloquificación", fenómeno que se presenta cuando son observables a simple vista los bloques regulares de pixeles que fueron la base de la codificación. En cambio, el método propuesto no introduce prácticamente artefactos y conserva una gran suavidad en zonas uniformes (no-dimensionales), además de que respeta mucho más las regiones estructurales u orientadas. La calidad (y el tamaño, por ende) de las dos imágenes propuestas podría elevarse aún más entrenando los libros de código con más imágenes o aumentando la cantidad de vectores dentro de cada libro.

La calidad subjetiva de las imágenes codificadas supera a la de el método JPEG y resulta una buena elección para lograr altas tasas de compresión. La siguiente tabla resume algunos datos interesantes (tabla 7.1):

| Tipo de codificación | PSNR | Tasa de compresión |
|-----------------------------------|------|--------------------|
| Codificación en luminancias | 23.1 | 0.6 bpp |
| Codificación en brillantez (cruz) | 23.7 | 0.6 bpp |
| JPEG | 21.8 | 0.6 bpp |

Tabla 7.1 Datos de la compresión de la imagen "Mount"

Aunque no sea notoria la superioridad, en PSNR, del método en brillantez contra el método en luminancias, la reproducción de pequeñas zonas finas y de ciertas texturas es más acertada. De igual forma, algunas secciones sin mayor importancia son consideradas no-dimensionales, a diferencia del primer método.



Figura 7.4 a) Imagen original (arriba), imagen comprimida con método en luminancias (abajo). Tasa de compresión de 0.51 bpp



b) imagen comprimida por método en cruz (arriba), imagen JPEG de misma compresión (abajo). Tasa de compresión 0.51 bpp

| Tipo de codificación | PSNR | Tasa de compresión |
|-----------------------------------|-------|--------------------|
| Codificación en luminancias | 28.2 | 0.51 bpp |
| Codificación en brillantez (cruz) | 28.77 | 0.51 bpp |
| JPEG | 27.9 | 0.51 bpp |

Tabla 7.2 Datos de la compresión de la imagen "Clown"

En este caso, tenemos una imagen ("Clown") un poco más comprimida en los tres casos, y aunque los valores PSNR son muy similares, la reproducción de los primeros dos métodos es muy limpia y suave, respetando las estructuras y haciendo agradable la figura. La imagen JPEG introduce gran cantidad de artefactos, muestra muchos bloques y abusa de la cuantización en zonas homogéneas que pueden tener algunas texturas asociadas. Recordemos una vez más que no se introduce ningún tipo de compresión entrópica a los métodos propuestos.

A continuación se muestra una tercera imagen (fig. 7.5) de características distintas, y codificada con un libro de códigos más grande. Se observa que, una vez disponible, un libro de códigos bien entrenado y de tamaño razonable, dará una buena calidad de reproducción con tasas de compresión altas. Una vez más se compara con su contraparte JPEG, obtenida con el mismo rango de compresión para establecer una comparación pareja.

En esta caso, la compresión JPEG para alta compresión, es menos defectuosa que en las imágenes anteriores, como los números de la tabla manifiestan, en general las estructuras son claras y aparecen menos artefactos; sin embargo, en las estructuras finas (mástiles, líneas), en la reproducción de bordes y en las texturas finas, se ve superada por las imágenes de los métodos vectoriales, donde otra vez el suavizado en ciertas zonas y el realce en otras la hace un fuerte competidor, -por lo menos en estos niveles- en calidad al conocido estándar. Veamos los números y las imágenes:

| Tipo de codificación | PSNR | Tasa de compresión |
|-----------------------------------|---------|--------------------|
| Codificación en luminancias | 30.1 dB | 0.65 bpp |
| Codificación en brillantez (cruz) | 30.0 dB | 0.65 bpp |
| JPEG | 31.0 dB | 0.65 bpp |

Tabla 7.3 Datos de la compresión de la imagen "Boat"

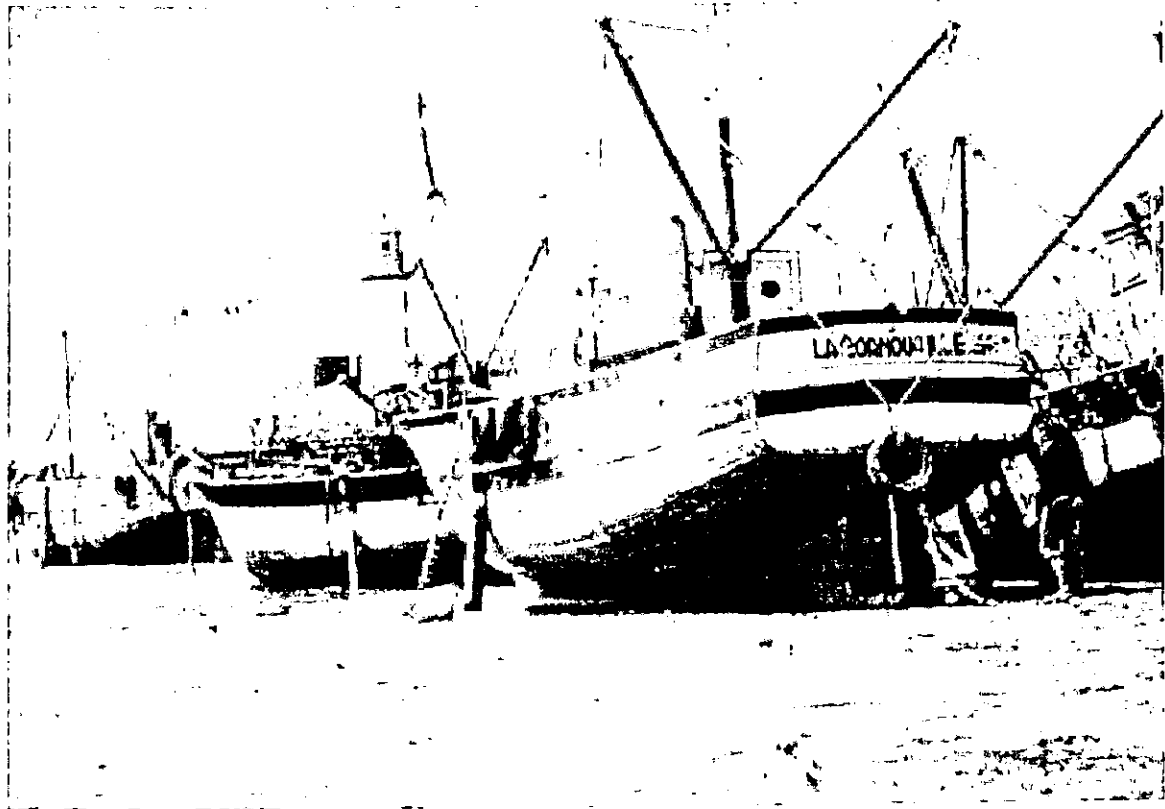


Figura 7.5 a) Imagen original (arriba), imagen comprimida con método en luminancias (abajo)



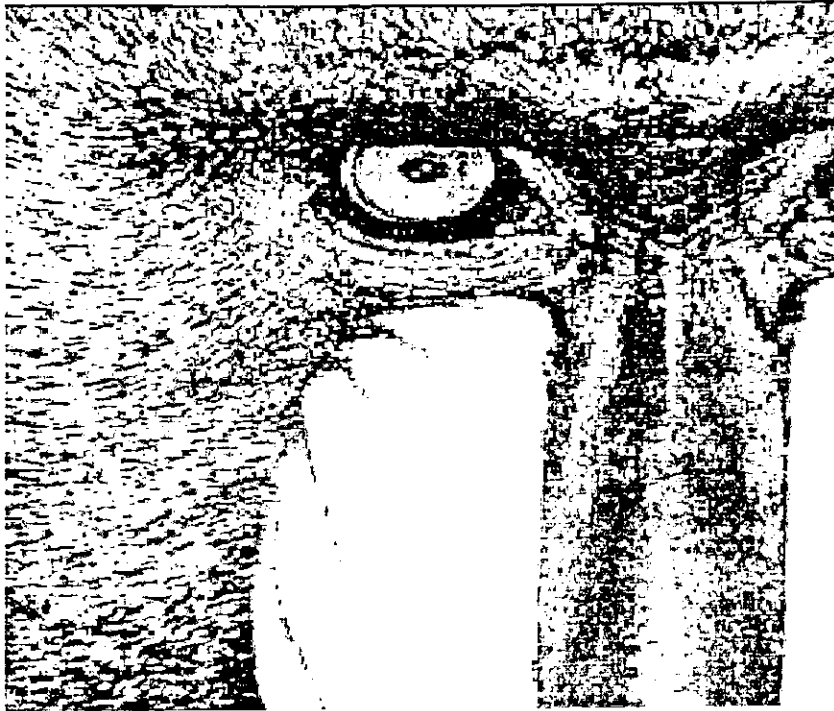
Fig. 7.5 b) imagen comprimida por método en cruz (arriba), imagen JPEG de misma compresión (abajo). La tasa de compresión lograda fue 0.65 bpp

7.3 La calidad de la imagen en función del tamaño y constitución del libro de códigos

En esta sección del capítulo de los resultados quisiera mostrar gráficamente como varía la calidad de una imagen sometida a cuantización vectorial en función del parámetro más poderoso para determinar su calidad subjetiva y objetiva: el tamaño del libro de códigos. Obviamente, la constitución del mismo, es decir, cómo se llevó a cabo el entrenamiento que derivó en su construcción, es un punto básico también; pero es más difícil de controlar en un sencillo proceso de comparación. El codificador se halla “atado” por otros lados. He fijado los parámetros de la transformada de Hermite para que el sistema trabaje únicamente con unidades básicas de 8x8 pixeles. En algún momento del desarrollo del proyecto se implementó un esquema *quadtree*¹ para utilizar libros de códigos de distinto tamaño, pero al momento de intentar asociarlo con las clasificaciones estructurales y la transformada hermitiana, no hubo forma de relacionarlos amigablemente. Había que fijar la unidad fundamental de trabajo, y así se hizo. Por otro lado, la distribución estructural depende de la riqueza contextual de cada imagen, por lo que no puede definirse a priori la cantidad de vectores de cada tipo (0D, 1D ó 2D) que serán cuantizados. Por esto se concluye que la “personalidad” de las imágenes obtenidas con este codificador será similar en todos los casos. Está enfocado a obtener altas tasas de compresión preservando características psico-visuales importantes. Una vía secundaria para aumentar el SNR de las imágenes codificadas sería asignando más bits a los cuantizadores escalares: el de DC y el angular, sin embargo no se obtendría mayor ganancia, y por el contrario, se disminuiría mucho la compresión. Es por esto que se mostrará a continuación (fig. 7.5) una imagen codificada en distintas tasas de compresión, obtenida con la ayuda de libros de código distinto.

La imagen “Baboon” es codificada con libros de código de 128, 256, 1024, 2048 y 4096 vectores. La tabla que se adjunta al lado de cada imagen incluye los radios de compresión y el SNR. Los libros de código fueron entrenados con 20 imágenes, dentro de las que no se hallaba “Baboon”:

¹ Ver capítulo 3



Detalle de la imagen original "Baboon", 512x512 pixeles

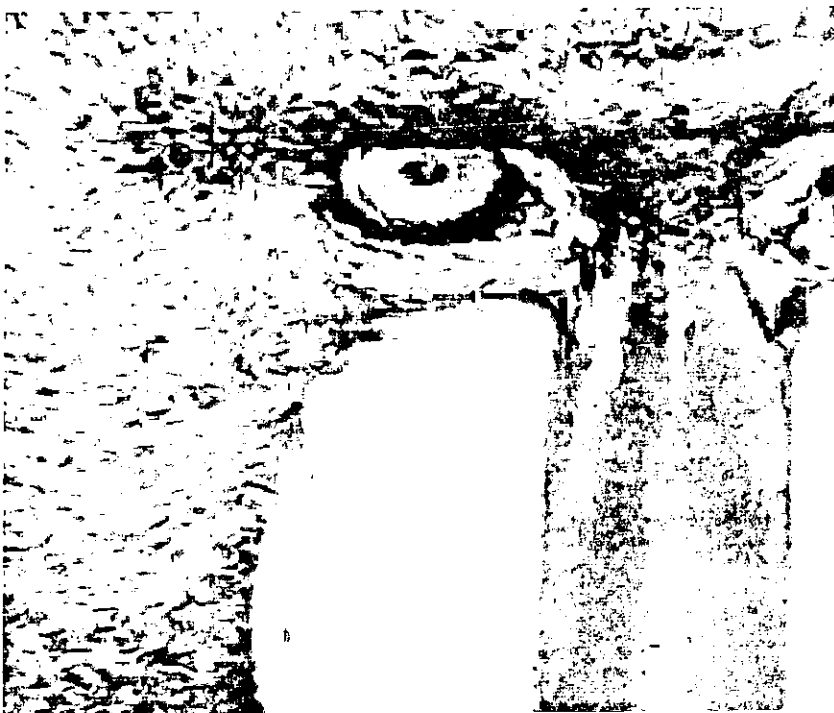


Imagen cuantizada con 2 libros de código de 128 vectores c/u
Tasa de compresión: 15:1
SNR: 21.77 dB



Imagen cuantizada con 2
libros de código de 256
vectores c/u

Tasa de compresión:
14:1

SNR: 22.01 dB



Imagen cuantizada con 2
libros de código de 1024
vectores c/u

Tasa de compresión:
13:1

SNR: 22.61 dB

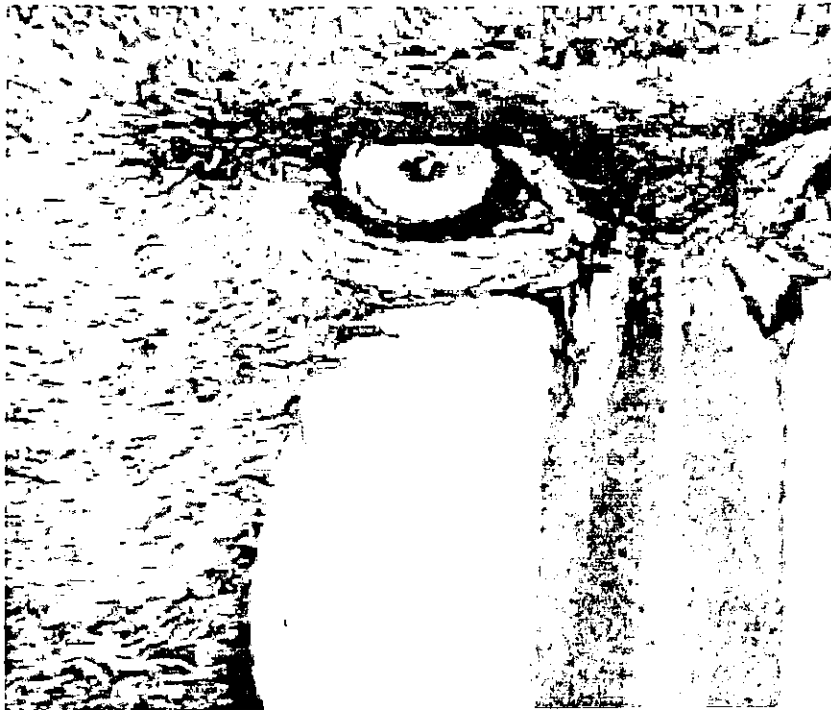


Imagen cuantizada con 2
libros de código de 2048
vectores c/u
Tasa de compresión:
13:1
SNR: 23.2 dB

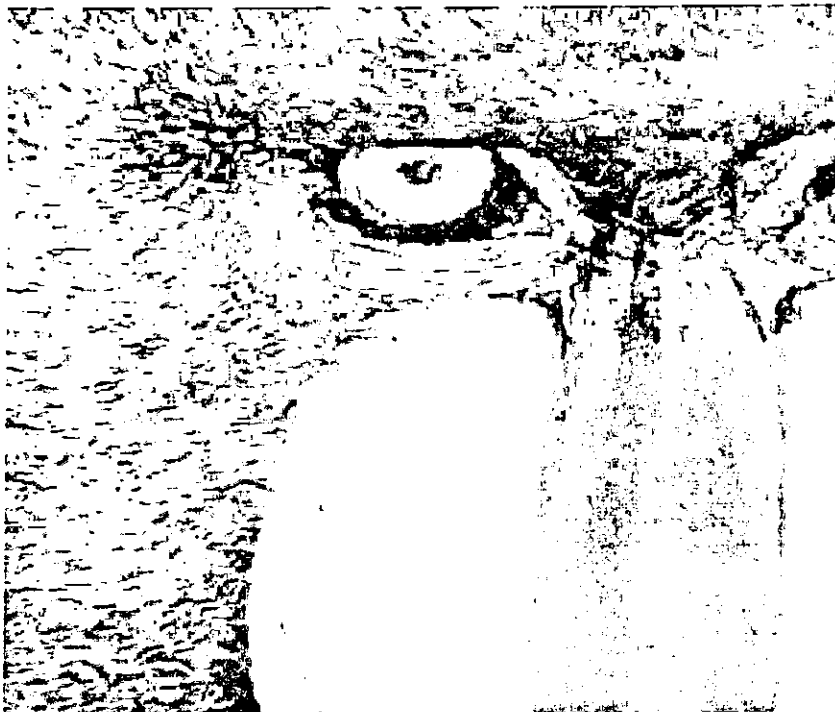


Imagen cuantizada con 2
libros de código de 4096
vectores c/u
Tasa de compresión:
12:1
SNR: 24.02 dB

Figura 7.6 La imagen "Baboon" fue comprimida utilizando libros de códigos de varias dimensiones, comenzando con uno de 128 elementos hasta uno de 4096. La primera imagen es la original. El arreglo de 5 imágenes comprimidas tiene como objetivo observar y comparar tasas de compresión, relación señal a ruido y, sobre todo, calidad subjetiva de reproducción en cada caso.

En la siguiente gráfica (fig. 7.6) se confrontan las tasas de compresión y las SNRs de las imágenes comprimidas "Baboon" mostradas en la figura anterior. Se observa, por lo menos para este caso, que la relación no guarda linealidad y que en teoría, al aumentar las dimensiones del libro de códigos obtendremos mejor rendimiento en la codificación, reflejado en un SNR más alto.

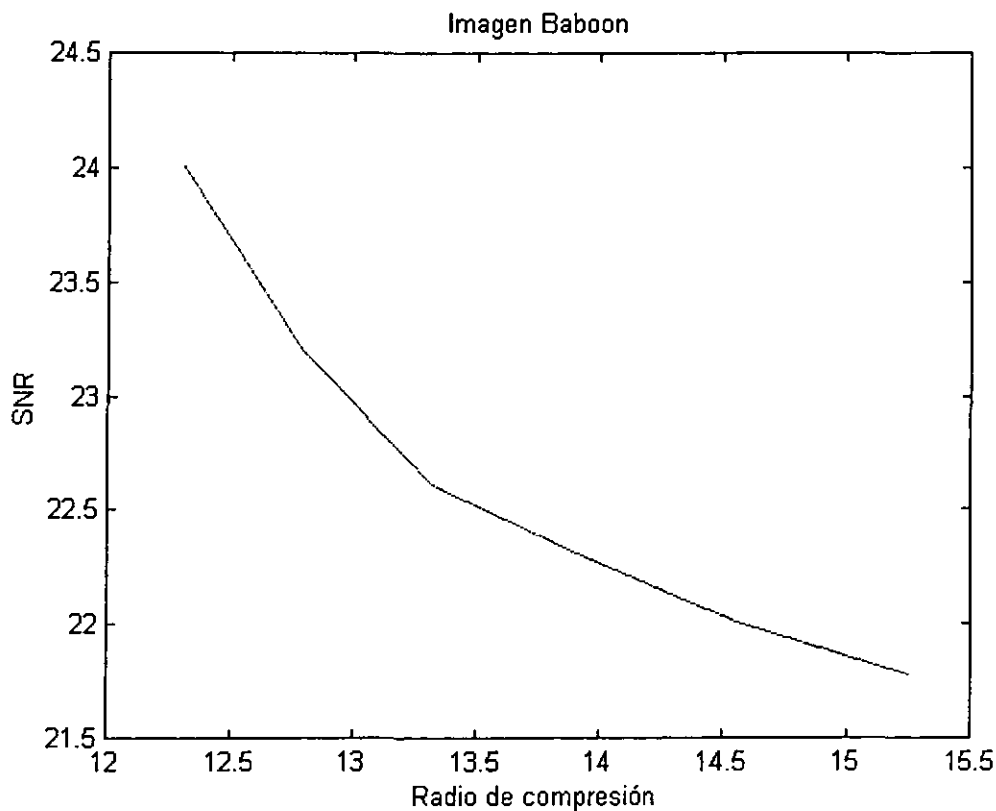


Figura 7.7 Confrontación de ratio de compresión y SNR para las distintas imágenes "Baboon" obtenidas en la figura anterior

Otro experimento interesante resultó de codificar la imagen "Lena", misma que se muestra en la figura siguiente (7.7), en el rango de libro de códigos de 64 a 4096 elementos, cuantificar su relación señal a ruido con respecto a la original y compararla con la misma imagen codificada con el estándar de compresión JPEG, sometido a las mismas condiciones de bits por píxel, sin tomar en cuenta la codificación entrópica que reduce aún más el tamaño de los archivos .JPG; esto con el objeto de establecer una "comparación pareja". Los resultados de dicha prueba se grafican a continuación (figura 7.7):



- Método en luminancias
- Método en cruz
- Estándar JPEG

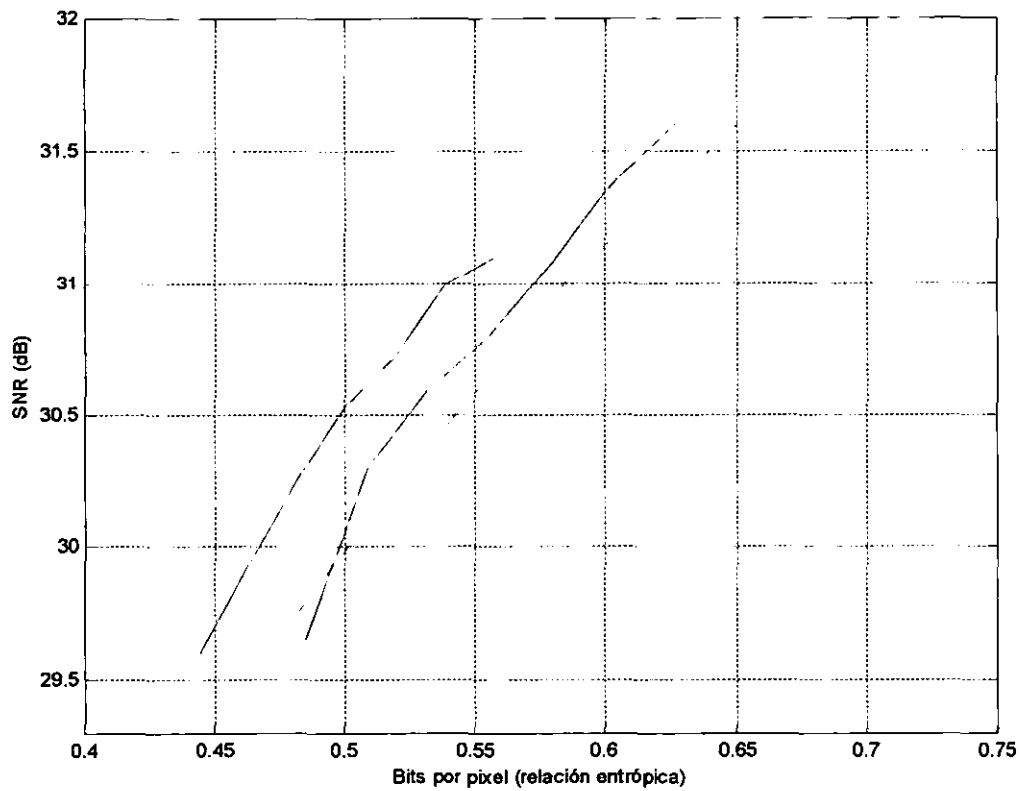


Figura 7.8 Comparación del rendimiento bits por pixel – SNR para los métodos de luminancia, cruz y el estándar JPEG en la imagen “Lena”, de 256x256 píxeles.

7.4 Baja distorsión entre imágenes codificadas en brillantez y su mapa correspondiente

Considero que una importante observación que se ha derivado del trabajo con campos de brillantez es su factibilidad para crear libros de código eficientes. Como se ha expuesto varias veces, el dominio de la brillantez es en sí una representación comprimida de la luminancia. El mapeo luminancia-brillantez tiene la virtud de preservar la información perceptiva más importante y descartar algunas zonas de contraste que no sean relevantes para el ojo humano sin destruir la morfología de la imagen. Para ilustrar esto de una forma sencilla, observe la siguiente figura (fig. 7.8). En ella se observa que la distribución en los histogramas cambia; invariablemente un mapa de brillantez necesitará menos información para ser representado.

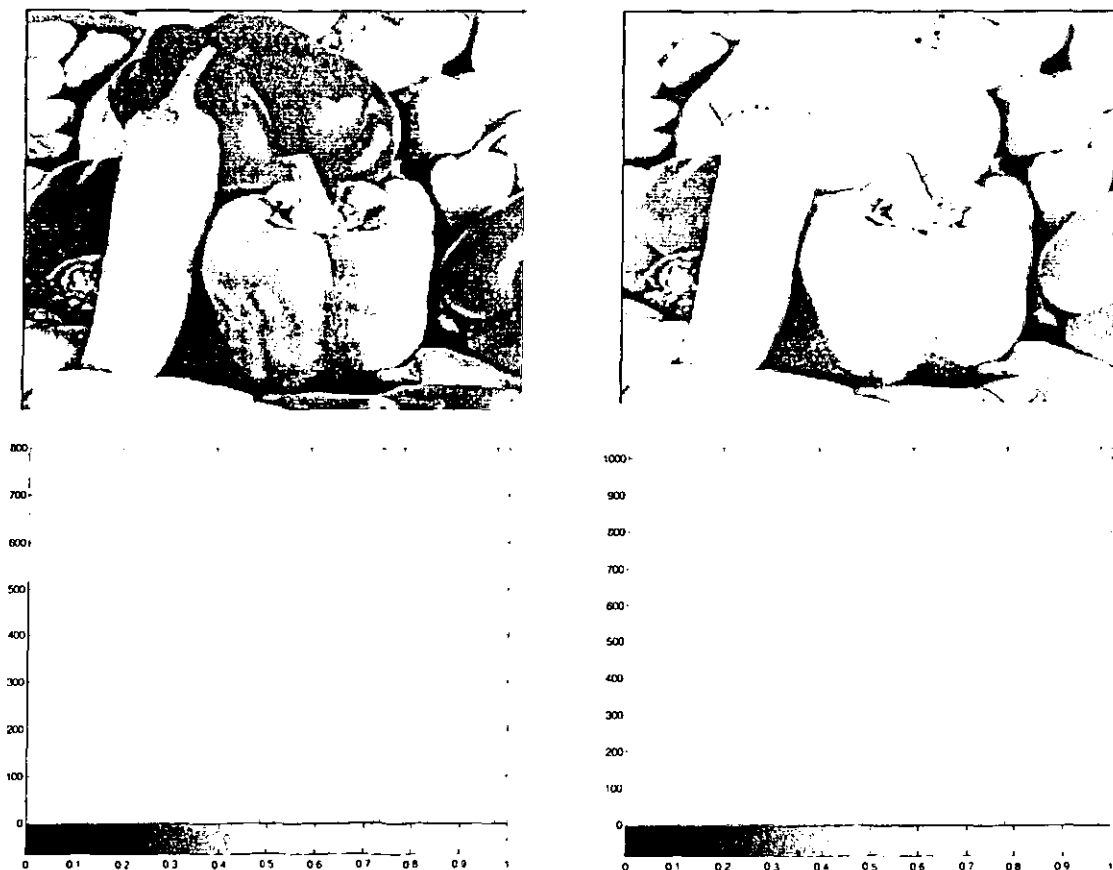


Figura 7.9 Imagen "Peppers" (arriba, izquierda) y su mapa de brillantez ecualizado para su representación visual (arriba, derecha). Sus histogramas se muestran debajo de ellos.

Desgraciadamente, los mapas de brillantez no pueden servir como el producto final, ya que no están diseñados para ser la representación visual definitiva de una imagen que se desea comprimir: si fuera así, nos limitaríamos a obtener mapas de brillantez y a almacenarlos en archivos para inventar un nuevo método de compresión. Se podría llevar a cabo, pero la utilidad más evidente (al menos como hemos podido constatar) de los mapas de brillantez es su incorporación, como guía, a métodos de compresión perceptiva.

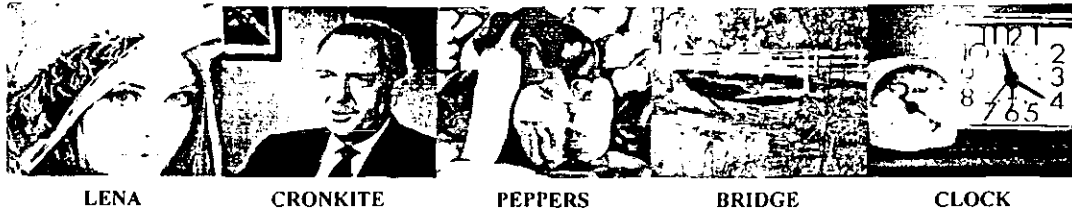
Uno de los métodos de codificación que propuso esta tesis consistía precisamente en codificar, con la ayuda de la transformada discreta de Hermite, vectores procedentes del dominio de la brillantez. Las imágenes decodificadas finales no superan en calidad a las de los otros dos métodos (aunque se hallan muy cerca al aplicar un ajuste al histograma, para visualizar correctamente la imagen), pero sirvieron para demostrar que la construcción de libros de códigos era muy eficiente. Este aspecto se tocó también en el pasado capítulo.

Para cuantificar esta afirmación, se muestra una tabla en la que se comparan las relaciones señal a ruido de imágenes cuantizadas con vectores de brillantez, calculadas contra su mapa de brillantez, no contra la imagen original. Los resultados son bastante superiores, al menos en cuanto a números se trata. La siguiente tabla (tab. 7.3) los compara con los valores PSNR de las imágenes codificadas con los métodos de luminancia y "cruz", calculados contra la imagen original.

Los resultados obtenidos muestran la factibilidad que presenta el dominio de la brillantez para elaborar libros de códigos eficientemente. Se observó, en todos los casos experimentados, que los datos en brillantez tienen menos entropía que en su dominio original. Este concepto puede ser explotado aún más en trabajos posteriores, incluyendo dichas ventajas en procesos de codificación avanzados.

Se han apuntado también en la tabla, además de las relaciones señal a ruido, los bits por píxel (relación entrópica) para cada caso. Como es de esperarse, dicha tasa arroja valores idénticos para las imágenes codificadas en cruz y en el dominio de la brillantez, ya que estructuralmente, partieron de la misma distribución de patrones. En algunos casos la relación señal a ruido es superior para la codificación en luminancias que en cruz, pero nunca la supera en entropía. Siempre es ligeramente más eficiente el uso de estructuras en brillantez (cruz). En cuanto a la calidad subjetiva de las imágenes codificadas en cruz, resulta agradable la decodificación de todas, resaltando "Cronkite", "Peppers" y

“Bridge” sobre “Lena” y “Clock”. En ninguno de los casos aparecen artefactos notorios ni se pierde información visual muy relevante.



| Imagen | 1: SNR (Original-imagen codif. en “cruz”) | 2: SNR (Original-imagen codif. en luminancias) | 3: SNR (Mapa de brillantez- codif. en brillantez) | % de ganancia del método 3 con respecto a 1 y 2 |
|-----------------------|---|--|---|---|
| LENA (256x256) | 31.1 dB (0.54 bpp) | 31.3 dB (0.6 bpp) | 36.8 dB (0.54 bpp) | 15 % |
| CRONKITE (256x256) | 35.0 dB (0.42 bpp) | 35.4 dB (0.475 bpp) | 40.3 dB (0.42 bpp) | 13% |
| PEPPERS (256x256) | 27.15 dB (0.544 bpp) | 27.3 dB (0.614 bpp) | 31.7 dB (0.544 bpp) | 14% |
| BRIDGE (512x512) | 27.0 dB (0.66 bpp) | 27.0 dB (0.677 bpp) | 32.11 dB (0.66 bpp) | 16% |
| CLOCK (512x512) | 37.3 dB (0.44 bpp) | 37.0 dB (0.46 bpp) | 42.51 dB (0.44 bpp) | 13% |

Tabla 7.4 Comparación de SNR para varios experimentos de codificación. Se utilizaron dos libros de códigos (1D y 2D) de 2048 elementos ambos. Las imágenes que se comprimiron (mismas que se muestran arriba) no estaban en las secuencias de entrenamiento. Se anotan también los datos de entropía de cada algoritmo de compresión, en bits por pixel.

Los números mostrados en la tabla no hacen más que confirmar que la compresión inherente que adquieren las imágenes al ser mapeadas a brillantez, disminuye la entropía de sus datos y hacen más factible su asociación. En otras palabras, se requieren menos vectores en un libro de códigos de brillantez para lograr un SNR determinado, que con un libro de códigos en luminancia, si de codificar imágenes en ambos dominios se trata. A continuación se muestra otro ejemplo (fig.7.9). Se trata de una imagen, su mapa de brillantez, y las decodificaciones de sus imágenes comprimidas.

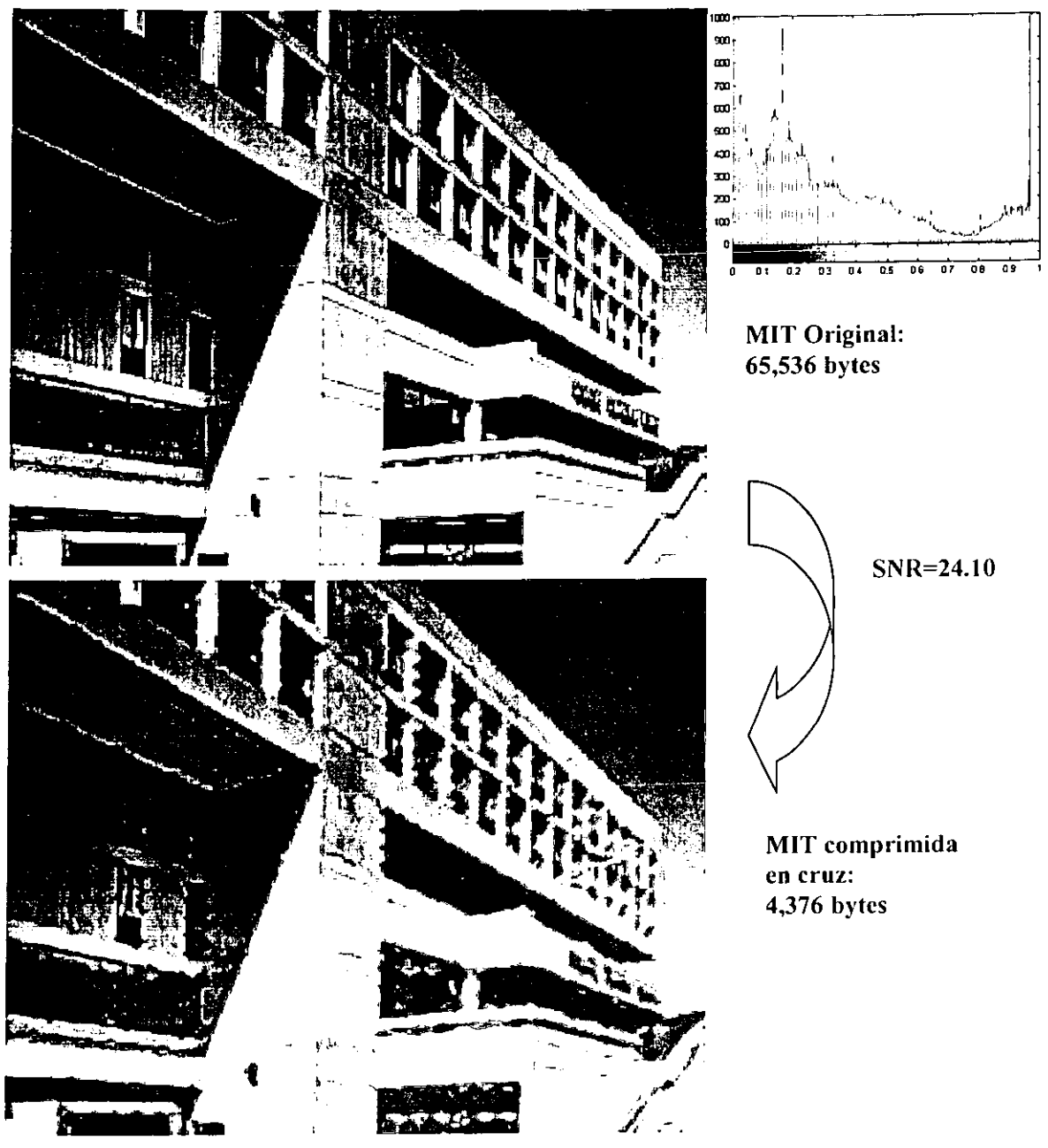


Figura 7.10 1) Imagen MIT con histograma, 2) reconstrucción del original (método en cruz), con tasa de 0.5 bpp,

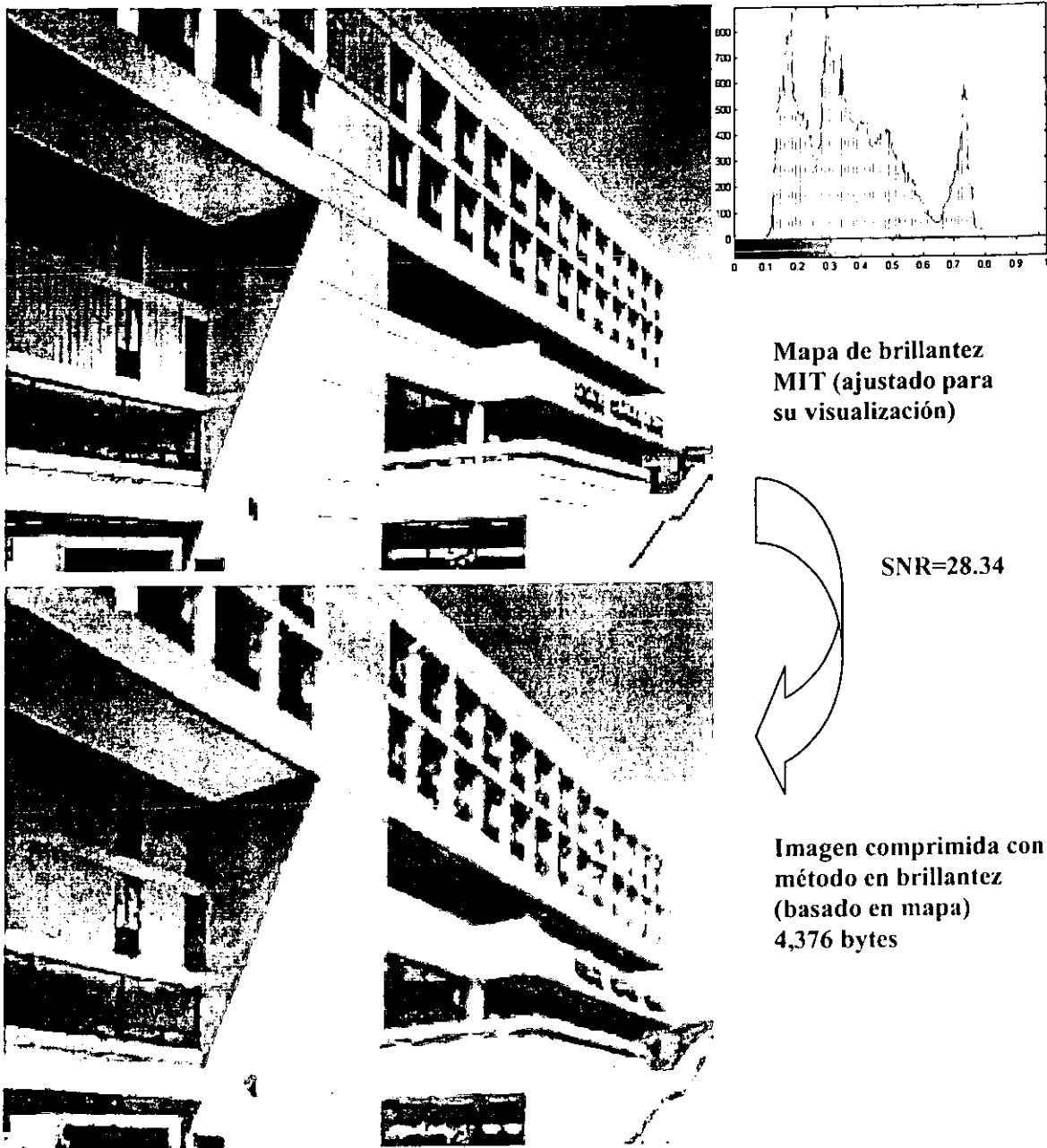


Figura 7.10 3) mapa de brillantez de MIT con histograma, 4) reconstrucción, en brillantez, del mapa

La imagen en cuestión (M.I.T.), es de 256x256 píxeles y 8 bpp. La codificación se hizo con la ayuda de un libro de códigos de sólo 1,024 elementos, logrando así una alta compresión. Lo que se quiere hacer notar en esta secuencia es la alta confiabilidad que dan los valores en brillantez al ser sometidos en un proceso de cuantización. La reconstrucción

del mapa, con un libro de códigos del mismo tamaño, da mucho mayor relación señal a ruido que en el caso tradicional. Es interesante, asimismo, observar los histogramas para darnos una idea del tipo de modificaciones que sufren las imágenes al ser mapeadas. Evidentemente, el mapa que se despliega aquí está modificado (con fines de despliegue), para observarlo con todo su contraste.

7.5 Posibles mejoras y/o modificaciones al trabajo

A lo largo de este trabajo se ha generado una interesante acumulación de información y de conclusiones. Como toda tesis, puede ser mejorada, ampliada o modificada aún más y es mi deber sugerir posibles vías para hacerlo. Creo que las líneas de investigación que convergieron en la elaboración de este proyecto son tópicos vastos en los que, a pesar de que se han estudiado exhaustivamente en la División de Estudios de Posgrado de la Facultad de Ingeniería, es necesario profundizar aún más. Algunas mejoras que tengan fin de darle continuidad al trabajo pueden ser las siguientes:

- Llevar a cabo experimentación exhaustiva con los campos de brillantez para asociarlos con distintos métodos de compresión de imágenes. Este tema puede ser aún trabajado ampliamente y puede redituarse en resultados muy interesantes. El refinamiento en la obtención de mapas de brillantez, o la especialización en el análisis matemático de los métodos existentes pueden abrir nuevas rutas de investigación para aprovechar de mejor forma el trabajo en este dominio. El mapeo ha demostrado ser una vía interesante para analizar estructuralmente las imágenes digitales, por lo que se antoja probable la concordancia de su teoría perceptiva básica con esquemas de procesamiento alternos a la cuantización vectorial y a la compresión de señales visuales.
- “Migrar” la codificación que se ha hecho del proyecto a lenguaje “C”, con objeto de hacer rápida su ejecución y más eficiente el manejo de la memoria. De esta forma podríamos aumentar la complejidad del filtrado en la obtención de los mapas de brillantez (más resoluciones) y podríamos manejar eficientemente aún más vectores (tanto de entrada como en los libros de código) de forma cómoda.

- Incorporar algún sencillo procedimiento de codificación entrópica (tablas de Huffman, por ejemplo) a la etapa de escritura del archivo o agregar códigos de seguridad (codificación de canal) para aumentar un poco más la compresión total y la confiabilidad del *bitstream* lograda en el codificador.
- Optimizar el código, migrarlo a lenguaje ensamblador y llevar a *hardware* el proceso de codificación para incorporarlo en codificación de imágenes, o de secuencia de imágenes, en tiempo real. Puede ser utilizado, para tal efecto, un dispositivo de procesamiento de señales como los DSP's, por ejemplo.
- Continuar la experimentación con secuencias de entrenamiento de diversas personalidades y configuraciones, con objeto de lograr libros de códigos más propicios para la compresión de determinado tipo de imágenes.
- Investigar la factibilidad de utilizar más libros de códigos en el proceso. Determinar cómo serían éstos y qué características intrínsecas tendrían que tener los vectores pertenecientes a ellos. Es probable que tanto los patrones unidimensionales como los bidimensionales puedan ser re-clasificados.
- Ampliar la capacidad de la interfases de usuario, sobre todo si se incorporan los módulos programados en MatLab a un proyecto de mayor envergadura.

7.6 Conclusiones

A pesar de que a lo largo de los últimos dos capítulos se ha profundizado en los resultados que el proyecto arrojó, siempre es conveniente hacer una última puntualización de las principales ideas y los datos más relevantes en que fructificó este estudio.

Primeramente, el hecho de indagar en cualquier cantidad de artículos y bibliografía acerca de percepción sensorial, me orientó para poder determinar en qué aspectos perceptivos se enfocaría el trabajo. Cualquier esquema de codificación perceptual de señales tiene una molesta disputa entre rapidez y potencialidad. A mayor complejidad en el análisis perceptivo de la señal, se obtiene una aproximación más fiel a la misma, pero se cae en un mayor uso de memoria y en un mayor consumo de tiempo. Se decidió, en este

caso, que el sistema se ubicara a la mitad en esa recta, elaborando un codificador que intentara ser preciso en su análisis sin “robar” demasiados recursos y demasiado tiempo.

En cuanto al análisis perceptivo en sí, se pudo determinar fehacientemente lo que es ampliamente comentado en la bibliografía: el hecho de que las estructuras finas, los patrones orientados y las texturas deben tratar de ser preservados en su mayoría para evitar la degradación de la imagen. Sin embargo, no deben tener la misma importancia subjetiva todos los elementos mencionados, ya que se observó como son fuertemente dependientes del contexto en que se hallan, como lo aportó el mapeo de brillantez. La cantidad de contraste en la vecindad, asimismo, es un factor de peso en la sensación visual que causa una escena el ojo humano. Esto pudo constatarse al comparar las diferencias en clasificación de estructuras entre las imágenes en luminancia y sus mapas de brillantez. Al cambiar al dominio transformado de la brillantez, algunas estructuras cobraban importancia mientras que otras tendían a ser absorbidas por el contexto. Esta última idea se llevó a la práctica en la política de codificación denominada “cruz”, en la que se consideraban los coeficientes hermitianos transformados a partir del dominio original, pero con la información estructural del mapa de brillantez de la imagen. Las diferencias entre la codificación de luminancias transformadas y la codificación en “cruz” son muy sutiles, pero reflejan el punto clave de la tesis, la explotación del mapa de brillantez. Desgraciadamente, éste no es un resultado apto para ser representado visualmente, ya que generalmente se trata (al menos a primera vista) de una versión “clarificada”, o con mayor energía, de la imagen original. Sin embargo, su aportación reside en su relación estructural con su contraparte en luminancias. Se hicieron esfuerzos para hacer del campo de brillantez una imagen codificada *per se*, pero probablemente no sea ése el camino correcto. Creemos firmemente que un mapa de brillantez bien obtenido es una guía para conocer más ampliamente las características perceptivas de una imagen, en un dominio distinto pero con una representación morfológicamente similar.

El trabajo que se ha llevado a cabo en la DEPFI, ya desde hace varios años, en relación al estudio de las transformadas polinomiales en procesamiento digital de imágenes siempre ha dado buenos, e innovadores resultados. En esta tesis, la transformada discreta de Hermite es uno de los tres pilares básicos de su soporte, junto con los ya mencionados mapas de brillantez y con la cuantización vectorial, que es el agente de codificación en sí.

Era casi imperativo incorporarla al codificador porque sabíamos de antemano las bondades que presentaba tanto en compresión como en manipulación matemática. Afortunadamente, el esquema se prestaba perfectamente a que esto fuera así, y de esta forma lograr un compresor mucho más serio que pudiera ser comparado, cualitativa y cuantitativamente con estándares comerciales, como se hizo con el JPEG. La experiencia previa en el tema de transformación ayudó mucho, se contaba ya con la bibliografía adecuada y con varios módulos a los que se hicieron modificaciones pertinentes. Otros aspectos de este tema, sin embargo, fueron programados e investigados desde cero.

La transformada discreta de Hermite tiene una pequeña desventaja: induce necesariamente redundancia, dependiente sobre todo de la elección de las funciones ventana. Afortunadamente, con el empleo del "enrejado hexagonal", se alivió este fenómeno y no hubo que tratar con demasiada información, hecho que hubiera deteriorado mucho las tasas de compresión. Sólo se utilizaron dos bloques de matrices por imagen (coeficientes *nones* y coeficientes *pares*), además de la información angular y los coeficientes de DC. Es un hecho que la transformada hermitiana *acomoda* mejor la energía que lejanos parientes de ella, como la DCT. Esto, aunado a la rotación de coeficientes, genera un manejo potencial de estructuras mucho más amplio que con otras transformadas. Otros trabajos desarrollados en esta división han mostrado este hecho (Silván, 1998).

La cuantización vectorial es un arma interesante para llevar a cabo codificación de señales en general. Puede adecuarse fácilmente a múltiples esquemas tanto en señales unidimensionales (voz, audio) como en señales bidimensionales (imágenes, superficies de representación) y en señales tridimensionales (video digital, imágenes tridimensionales). Desde el arranque del proyecto tenía la inquietud de llevar a cabo el ensamble de los elementos con un cuantizador vectorial, que por supuesto tiene múltiples ventajas pero también cuenta con desventajas. En primer lugar, la dependencia de los libros de códigos hace que la tasa de compresión y la calidad de reproducción estén intrínsecamente ligados con la constitución de los mismos. Esta característica fue un reto en la edificación del sistema, y se optó por utilizar dos libros de códigos a la vez (uno para estructuras 1D y otro para estructuras 2D) que fueran entrenados con una cantidad considerable de estructuras provenientes de imágenes de todo tipo. Se logró que, con un número no muy bajo de vectores en los libros, la codificación de imágenes pudiera hacerse con libros de código en

los que la imagen de entrada no fuera considerada en el entrenamiento. Todos los resultados reportados en esta tesis cumplen con esa característica. Creo que es muy importante entrenar los libros con una gran variedad de vectores, ya que así enriquecerán su acervo y tendrán una capacidad potencial de representación mucho más grande.

El tiempo empleado en crear libros de códigos de prueba fue muy alto. El proceso de generación (obtención de las secuencias de entrenamiento en luminancia y brillantez y algoritmo LGB modificado) consume mucho tiempo y acapara gran cantidad de recursos de cómputo. Sin embargo, una vez obtenidos, los libros de códigos no son particularmente grandes y pueden ubicarse fácilmente en los dispositivos de codificación y de decodificación.

Por experiencia, puedo decir que las imágenes decodificadas que fueron alguna vez comprimidas con cuantizadores vectoriales, suelen tener muchos defectos (bloquificación, pobreza textural, enborronamiento², etc.), y creo que los resultados a los que se llegó salvan esas desventajas. Una de las causas por las que decidí que el sistema funcionara regido por cuantización vectorial fue por la capacidad de compresión que daría la rotación en la transformada de Hermite, facilitando el trabajo con un número limitado de vectores (transformados y rotados) de reproducción a la salida.

La comparación, aunque odiosa, se hace necesaria en este tipo de productos. Decidí que el estándar JPEG, por amplitud de uso, fama y características intrínsecas de funcionamiento (manejo en bloques de 8x8, transformada DCT, cuantización escalar de coeficientes y compresión entrópica de canal) podía ser un buen parámetro “a vencer”. Se compararon muchas imágenes comprimidas con el algoritmo presentado e imágenes JPEG con una tasa de compresión similar (tomando en cuenta que no hay compresión entrópica en el archivo construido) y los resultados fueron satisfactorios: la preservación de estructuras, la casi perfecta preservación de bordes y el manejo adecuado de las zonas uniformes hacen del codificador un buen *sparring* al universal JPEG, que en tasas altas de compresión, como las utilizadas, introduce gran cantidad de artefactos, “bloquifica” frecuentemente las regiones (notable sobre todo en las uniformes) y destruye las texturas finas. Sin embargo, JPEG tiene una gran ventaja: puede establecer a priori, y con mucha exactitud, el grado de compresión que se va a utilizar, aspecto que es mucho más difícil de

² Conocido en la bibliografía de imágenes digitales como “deblurring”

controlar en un sistema de cuantización vectorial perceptiva como el mío. Al haber fijado tanto la dimensión de los bloques básicos de la transformada de Hermite en 8x8 píxeles, como la dimensión de los libros de códigos, es difícil variar a discreción la tasa de compresión del sistema. Podemos decir que más bien trabaja en áreas de alta compresión, comparado con la mayoría de realizaciones JPEG, que puede cubrir casi todo el rango de compresión posible.

Un tópico notable fue la alta confiabilidad lograda con libros de códigos construidos a partir de vectores de mapas de brillantez. Las imágenes reconstruidas con estos vectores eran mucho más cercanas a su mapa de brillantez original que las imágenes cuantizadas sobre luminancias, con libros de código en ese dominio, con respecto a sus originales. Ya se presentaron con antelación números al respecto, y yo concluyo de esto la factibilidad de utilizar modelos de brillantez en todo tipo de codificaciones de imágenes. El principal problema es, como ya dije, que el dominio de la brillantez, al menos como se trabajó aquí, no está diseñado ni es apto para la visualización directa de la imagen.

Quisiera poner fin a la escritura de este trabajo de tesis manifestando mi satisfacción por haber realizado la presente tesis de maestría en un tema tan interesante, tan vasto y con tantas aplicaciones en el mundo actual, como el procesamiento digital de imágenes. He aprendido enormemente a lo largo de estos dos años de maestría y he enriquecido muchísimo mi espíritu académico de investigación y de estudio.

VIII

Referencias Bibliográficas

1. ANDERSEN, Richard; How we See; *Aerospace Conference 1998 IEEE*, vol. 1, 1998, pp. 3-8.
2. ANG, A.; DURRANI, T.; Adaptive Transform Coding using Vector Quantization; *Conf. Record, Second International Conference on Image Processing and its Applications*; 1986; pp. 11-16.
3. ANG, Y.; BI, M; et al.; Adaptive Vector Quantization of Wavelet Coefficient for Image Compression; *ICICS Proceedings*, 1997, vol. 1, pp. 500-504.
4. BANISTER, B.A.; FISCHER, T.R; Quadtree classification and TCQ image coding; *Data Compression Conference, 1999. Proceedings*; 1999; pp. 149 –157.
5. BARLOW, H.B.; *Handbook of Sensory Physiology*, ch. Dark and Light Adaptation: Psychophysics. L. Hurvich and D. Jameson, New York, 1972.
6. BLOMMAERT, Frans J.; MARTENS, Jean-Bernard; An object-oriented model for brightness perception, *Spatial Vision*, vol. 5, no. 1, 1990, pp. 15-41.
7. CASTLEMAN, K.; *Digital Image Processing*; Prentice Hall, New Jersey, USA, 1996.
8. CHING, Yang Wang; SHIH, Jao Liao; Wavelet Transform Coding Using Variable Blocksize Vector Quantization with Optimal Quadtree Segmentation; *TENCON '97. IEEE Region 10 Annual Conference, Proceedings*, 1997, pp. 715-718.
9. CIE (1987). *International lighting vocabulary*. CIE publication no. 17, Gêneve: Bureau Central de la Comisión Electrotechnique Internationale.
10. CONSTANTINESCU, J., STORER, J.A.; Improved Techinques for Single-Pass Adaptive Vector Quantization; *Proceedings of the IEEE*, vol. 82, 1994.
11. CORNSWEET, T.; *Visual Perception*; Academic Press, New York, USA, 1970.
12. COSMAN, P; GRAY, R; et. al.; Using Vector Quantization for Image Processing; *Proceedings of the IEEE*, vol. 81, no. 9, 1993. pp. 1326-1341.
13. CUPERMAN, V; GERSHO, A.; Adaptive Differential Vector Coding of Speech; *Conference Record, Globe-COM 82*, 1982, pp. 1092-1096.
14. DAUGMAN, J.; Six Formal Properties of Two-Dimensional Anisotropic Visual Filters: Structural Principles and Frequency/Orientation Selectivity; *IEEE Trans. Syst., Man, Cybernetics*, vol. 13, 1983, pp. 882-887.

15. DINET, E.; LAGET, B.; Morphological Sampling Based on Foveal Mechanisms; *IPA 97, Conference Publication*, no. 443, julio 1997, pp. 799-803.
16. ESCALANTE, B.; SILVÁN, J.L.; Multiresolution Directional-Oriented Image Transform Based on Gaussian Derivatives, *SPIE Proceedings*; vol. 4478; USA 2001.
17. FOSTER, J.; GRAY, R.; Finite-State Vector Quantization; *Abstracts of the 1982 International Symposium on Information Theory*, 1982.
18. FREEMAN, W.T.; ADELSON, E.H.; The Design and use of Steerable Filters; *IEEE Transactions Patt. Anal. Machine Intelligence*; No. 9, 1991, pp. 891-906.
19. GERSHO, Allan; GRAY, Robert; *Vector Quantization and Signal Compression*; Kluwer Academic Publishers, USA, 1992.
20. GONZÁLEZ, R.; WOODS, R.; *Digital Image Processing*. Addison-Wesley, Massachusetts, USA, 1992.
21. GRAY, Robert; Vector Quantization; *IEEE ASSP Magazine*, vol. 1, abril 1984; pp. 4-29
22. HALL, C.; HALL, E.; A Nonlinear Model for the Spatial Characteristics of the Human Visual System, *IEEE Trans. Syst. Man Cybern.*; USA, 1977, pp 161-170;
23. HEUCKE, Lars; MIRKO, Knaak; et al; A New Image Segmentation Method Based on Human Brightness Perception and Foveal Adaptation; *IEEE Signal Processing Letters*, vol. 7, no. 6, june 2000, pp.129-131.
24. HUFFMAN, D.; A Method for the Construction of Minimum Redundancy Codes, *Proc. IRE*, 40, 1952.
25. JAYANT, N.; JOHNSTON, J., et. al; Signal Compression Based on Models of Human Perception, *Proceedings of the IEEE*, vol. 81, no. 10, 1993, pp. 1385-1422.
26. LARSON, R.; HOSTETLER, R.; *Cálculo y Geometría Analítica*, McGraw Hill, 3ª edición, México, 1989.
27. LEGGE, G.E.; FOLEY, J. M.; Contrast Masking in Human Vision; *Journal of the Optical Society of America*; 70(12); 1980; pp. 1458-1471.
28. LI, Jia; CHADDHA, Navin; et. al; Asymptotic Performance of Vector Quantizers with a Perceptual Distorsion Measure; *Information Theory. 1997. Proceedings., 1997 IEEE International Symposium on* , 1997, 55 pág.
29. LINDE, Y.; BUZO, A.; et. al.; An Algorithm for Vector Quantizer Design, *IEEE Transactions on Communications COM-28*; 1980, pp. 84-95.
30. LLOYD, S.; Least Squares Quantization in PCM, *Bell Laboratories Special Issue on Quantization*, 1982.

31. MARTENS, J.B.; The Hermite Transform – Theory; *IEEE Transactions on Acoustics, Speech and Signal Processing*; vol. 38 no. 9, 1990; pp. 1595-1606.
32. MORENO, M.; et. al; *Implementación Orientada a Objetos de un Codificador MPEG-I, en MatLab, Aplicando Distintas Técnicas de Estimación de Movimiento*, Tesis de Licenciatura, F.I., UNAM, 1999.
33. OLSHEN, R.; COSMAN, P.; et al.; Evaluating Compressed Medical Images; *Proc. COMCON III*, 1991, pp. 2-11.
34. OPPENHEIM, A., SCHAFER, R.; *Discrete-Time Signal Processing*; Prentice-Hall, USA, 1989.
35. RABINER, L; SCHAFER, R.; *Digital Processing of Speech*; Prentice-Hall, USA, 1978.
36. RUIZ, I.; et al.; Image Compression via Multiresolution Feature-Based VQ of Hermite Transforma Coefficients Using Kohonen Neural Network; *Image Processing and its Applications*, 1995, pp. 549-553.
37. SCHARF, B; Critical Bands; *Foundations of Modern Auditory Theory*; Academic Press, 1970, N.Y., USA, pp. 159-202.
38. SCHIRILLO, J., et. al.; Perceived Lightness, but not brightness, of achromatic surfaces depends on perceived depth information . *Perception and Psychophysics*, vol. 48, 1990, pp. 82-90.
39. SCHOUTEN, Gerard; *Luminance-Brightness Mapping: the Missing Decades*. Technische Universiteit Eindhoven, Holanda, 1993.
40. SHAPIRO, L; ROSENFELD, A.; *Computer Vision and Image Processing*, Academic Press, San Diego, USA, 1992.
41. SHANNON, C.; A Mathematical Theory of Communication; Bell Systems Tech. J.; num. 27, 1948, pp. 379-423, 623-656.
42. SHIG-MO, Ch.; KWOK-TUNG, L.; Variance Oriented Dynamic Codebook Adaptive Vector Quantization System for Image Coding; *IEEE International Conference on Communications*, Vol. 1 , 1996; pp. 137 –141.
43. SILVÁN, J.L.; *Compresión de Imágenes Digitales utilizando la Transformada Polinomial*; Tesis de Licenciatura, F.I., UNAM, 1999.
44. SILVÁN-CÁRDENAS, J.L., ESCALANTE, B.; Image Coding with a Directional-Oriented Hermite Transform on an Hexagonal Sampling Lattice; SPIE Proceedings; Vol. 4472; USA, 2001.
45. SZENTAGOTHAJ, J.; ARBIB, M., Neurosciences Research Program Bulletin, 12, 319-353.

46. VAN DIJK, A.; MARTENS, J.B.; Feature-Based Image Compression with Steered Hermite Transforms; *Image Processing, 1996. Proceedings, International Conference on*, Vol. 1, 1996, pp. 205 -208.
47. WANG, L., et. al.; Unified Variable-Length Transform Coding and Image-Adaptive Vector Quantization; *IEEE J. Selected Areas Comm.*; vol. 10, 1982, pp. 892-901.
48. WANG, S.; SEKEY, A.; et. al; An Objective Measure for Predicting Subjective Quality of Speech Coders; *IEEE J. Select. Areas Commun.*; junio 1992; pp. 819-829.
49. WATT, R.J.; *Visual Processing: Computational psychophysical and cognitive research*; Hove and London: Lawrence Erlbaum Associates, 1988.
50. WESTEN, S.; LAGENDIJK, R., et. al.; Perceptual Image Quality Based on a Multiple Channel HVS Model; *Acoustics, Speech, and Signal Processing, 1995. ICASSP-95*, vol. 4, pp. 2351-2354.
51. YOUNG, R., et al; Recursive Gaussian derivative filters, *Pattern Recognition, Proceedings. Fourteenth International Conference on*, vol. 1, 1998, pp 509 –514.
52. YU-CHEN, H.; CHIN-CHEN, Ch.; Variable rate vector quantization scheme based on quadtree segmentation; *Consumer Electronics, IEEE Transactions on*, vol. 45, issue 2, 1999; pp. 310 –317.
53. ZEMANSKY, M, et al.; *University Physics*; Addison Wesley, 6° ed., 1988, USA.
54. ZETZSCHE C.; HAUSKE G.; Principal Features of Human Vision in the Context of Image Quality Models; *Image Processing and its Applications*, 1989, pp. 102 –106.
55. INT-1: <http://www.mgarrison.com/ksucourses/psy20090/lecturesetspsy200.html>