



UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO

POSGRADO EN CIENCIA E INGENIERÍA DE LA COMPUTACIÓN

**DESARROLLO DE UN MÉTODO DE FUSIÓN DE IMÁGENES
BASADO EN ENERGÍA UTILIZANDO LA TRANSFORMADA
WAVELET DISCRETA**

T E S I S

QUE PARA OBTENER EL GRADO DE:

**MAESTRO EN INGENIERÍA
(COMPUTACIÓN)**

P R E S E N T A

URIEL RODRIGO NAVA VELAZCO

DIRECTOR: DR. BORIS ESCALANTE RAMÍREZ

México D.F.

2007



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 y abuela que también se desvelaron durante esta tesis.

A través de estas líneas deseo presentar mi más profundo agradecimiento al Dr. Boris Escalante que dedicó parte de su tiempo a orientarme y permitirme expresar incredulidad hacia la transformada Hermitiana, otorgándome total libertad en la elaboración de este trabajo. Al mismo tiempo, le hago llegar mi gratitud al Dr. Gabriel Cristóbal en el CSIC quien no sólo me ayudó en la tesis sino en hacer muy grata la estancia en Madrid.

A Rafa Redondo por explicarme la invariancia de las wavelets y por un excelente concierto de Jazz. También dentro del CSIC, agradezco a Noemí Carranza por su apoyo junto con Sylvain Fischer, Filip Šroubek, Jan Flusser y Javier Portilla por sus comentarios y sugerencias.

A mis compañeros en la maestría, les agradezco su amistad y los días de estudio, en especial a Gibrán con quien pase grandes momentos en la etapa de investigación.

A Iván, Javier, Alberto y Felipe que cuando necesité artículos de IEEE siempre los buscaron felizmente.

A la distancia envió mi cariño a Ishtar que libra su propia batalla.

Un agradecimiento a Diana, Lulú, Cecilia y Amalia por su apoyo en el aspecto administrativo.

A los sinodales que se tomaron la molestia en revisar el trabajo.

Finalmente, a CONACYT y DGEP por el apoyo económico durante estos dos años.

Gracias totales.

*Once upon a midnight dreary, while I pondered, weak and weary,
Over many a quaint and curious volume of forgotten lore,
while I nodded, nearly napping, suddenly there came a tapping,
As of someone gently rapping, rapping at my chamber door.
"This some visitor," I muttered, "tapping at my chamber door
Only this, and nothing more."*

[...]

*And the Raven, never flitting, still is sitting, still is sitting
On the pallid bust of Pallas just above my chamber door;
And his eyes have all the seeming of a demon's that is dreaming,
And the lamplight o'er him streaming throws his shadow on the floor;
And my soul from out that shadow that lies floating on the floor
Shall be lifted - nevermore!*

**The Raven
Edgar Allan Poe**

I

Índice.

*“What is the use of a book, without pictures or conversations?”
Luis Carroll in Alice in Wonderland*

Índice General.

Resumen.	v
Antecedentes.	1
1.1 <i>Introducción.</i>	1
1.2 <i>Esquemas de fusión.</i>	4
1.3 <i>Ventajas de la fusión de datos.</i>	5
1.4 <i>Estado del arte.</i>	5
1.5 <i>Modelos de calidad perceptiva.</i>	6
1.6 <i>Prerrequisitos y desarrollo de software.</i>	7
1.7 <i>Descripción de la tesis.</i>	7
1.8 <i>Referencias.</i>	8
Análisis Multirresolución.	9
2.1 <i>La necesidad del análisis multirresolución.</i>	9
2.2 <i>Representaciones multirresolución.</i>	11
2.3 <i>Espacios vectoriales.</i>	14
2.4 <i>Proceso de ortogonalización de Gram – Schmidt.</i>	16
2.5 <i>Espacios de Hilbert.</i>	19

2.6	<i>Perspectiva histórica de la teoría wavelet.</i>	21
2.7	<i>La transformada Wavelet Continua.</i>	23
2.7.1	<i>Ejemplos de wavelets continuas.</i>	25
2.8	<i>Wavelets y análisis multirresolución de Mallat.</i>	27
2.8.1	<i>Wavelets ortogonales.</i>	29
2.8.2	<i>Descomposición de la transformada Wavelet Discreta.</i>	31
2.9	<i>Wavelets y bancos de filtros.</i>	34
2.9.1	<i>Submuestreo y sobremuestreo.</i>	34
2.9.2	<i>Implementación de la transformada Wavelet usando filtros espejo en cuadratura.</i>	36
2.9.3	<i>Funciones escala y funciones wavelet.</i>	40
2.9.4	<i>Wavelets para imágenes.</i>	44
2.10	<i>Inconveniencias de las wavelets ortogonales.</i>	45
2.10.1	<i>Problemas con los bordes de las imágenes.</i>	47
2.11	<i>Referencias.</i>	48
	Fusión de Datos.	51
3.1	<i>Fundamentos de la fusión de datos.</i>	51
3.1.1	<i>Objetivos y retos de la fusión de imágenes.</i>	52
3.2	<i>Esquemas de fusión.</i>	54
3.3	<i>Técnicas de fusión de imágenes.</i>	56
3.3.1	<i>Intensity Hue Saturation.</i>	56
3.3.2	<i>Transformación de Brovey.</i>	57
3.3.3	<i>Análisis de Componentes Principales.</i>	58
3.3.4	<i>Filtrado paso-altas.</i>	59
3.3.5	<i>Problemas de los métodos de fusión.</i>	60
3.4	<i>Fusión de imágenes basada en la transformada Wavelet Discreta.</i>	62
3.4.1	<i>Reglas de fusión.</i>	64
3.4.2	<i>Ejemplos.</i>	66
3.5	<i>Referencias</i>	67
	Modelos de Calidad Perceptiva.	70
4.1	<i>Introducción.</i>	70
4.2	<i>Estado del arte en las métricas de calidad perceptiva.</i>	71
4.3	<i>Propiedades de la visión.</i>	73
4.3.1	<i>Adaptación visual. Sensibilidad a la intensidad luminosa.</i>	73
4.3.2	<i>Contraste.</i>	75

4.3.3	<i>Sensibilidad en frecuencia.</i>	77
4.3.4	<i>Enmascaramiento.</i>	78
4.4	<i>Clasificación de las métricas de calidad.</i>	79
4.5	<i>Métricas de calidad objetivas.</i>	80
4.5.1	<i>Métricas de calidad basadas en el error.</i>	80
4.5.2	<i>Métricas de calidad perceptiva.</i>	82
4.5.3	<i>Métricas de calidad híbridadas.</i>	84
4.6	<i>Implementación de los modelos visuales.</i>	87
4.6.1	<i>Adaptación de la luminancia.</i>	88
4.6.2	<i>Función de sensibilidad al contraste (CSF).</i>	88
4.6.3	<i>Descomposición en canales.</i>	92
4.6.4	<i>Enmascaramiento espacial.</i>	94
4.6.5	<i>Mapas e índices de error.</i>	94
4.7	<i>Referencias.</i>	95
	Algoritmo de Fusión.	97
5.1	<i>Algoritmo de Fusión.</i>	97
5.1.1	<i>Extracción de características.</i>	103
5.1.2	<i>Ejemplo del filtro de Canny.</i>	103
5.2	<i>Resultados experimentales.</i>	104
5.2.1	<i>Fusión de imágenes con referencia.</i>	105
5.2.2	<i>Fusión de imágenes sin referencia.</i>	117
5.3	<i>Referencias.</i>	121
	Conclusiones.	122
5.1	<i>Conclusiones.</i>	122
5.2	<i>Trabajo futuro.</i>	123

Índice de figuras.

Figura 1-1	Sistema de vigilancia multisensor.	2
Figura 1-2	Imágenes multifocales. (a) Imagen con el lado izquierdo desenfocado. (b) Imagen con el lado derecho desenfocado. (c) Imagen fusionada.	3
Figura 1-3	Fusión de imágenes de MR y CT: (a) Imagen de MR. (b) Imagen de CT. (c) Imagen compuesta obtenida utilizando la herramienta FUSIONLAB, la regla de selección empleada es MAX-MAX.	4

Figura 2-1 Ortogonalización de Gram–Schmidt para dos vectores, el vector y_2 se obtiene de restar a x_2 su proyección sobre x_1 .	19
Figura 2-2 Transformada de Fourier de $y = \sin(10\pi t) + \sin(40\pi t)$. (a) Señal sinusoidal en el dominio temporal. (b) Espectro de potencia de y .	21
Figura 2-3 Parte real de la wavelet compleja Morlet 1-5.	25
Figura 2-4 Wavelet <i>Mexican hat</i> con $\sigma = 1$.	26
Figura 2-5 (a) Función $x(t) = \text{chirp}$ con frecuencia 30 Hz en el tiempo $t = 0$ y desciende a 2Hz en el tiempo $t = 2$. (b) Vista 3D de la transformada Wavelet Continua de $x(t)$ utilizando <i>Mexican hat</i> .	27
Figura 2-6 Representación esquemática de la ecuación (2.66).	33
Figura 2-7 DWT de la señal $x(t)$ aleatoria ruidosa utilizando la wavelet Daubechies Db2.	33
Figura 2-8 Banco de filtros de M canales.	34
Figura 2-9 (a) Submuestreo de la señal $x(t)$ por un factor $N = 3$. (b) Sobremuestreo de $x(t)$ por un factor $N = 3$.	35
Figura 2-10 Filtros espejo en cuadratura QMF: (a) Filtro H_0 . (b) Filtro H_1 . (c) Respuesta en frecuencia de los QMF. (d) Potencia de los QMF.	37
Figura 2-11 Banco de filtros QMF y reconstrucción perfecta.	38
Figura 2-12 Wavelet Haar (a) Función Escala. (b) Wavelet Madre.	40
Figura 2-13 Wavelets Daubechies. (a) Daubechies de 4 coeficientes. (b) Daubechies de 6 coeficientes. (c) Daubechies de 12 coeficientes. (d) Daubechies de 20 coeficientes.	41
Figura 2-14 Wavelets Coiflets. (a) Coiflet de 6 coeficientes. (b) Coiflet de 12 coeficientes. (c) Coiflet de 18 coeficientes.	42
Figura 2-15 Wavelets Symlets (a) Symlet de 4 coeficientes. (b) Symlet de 6 coeficientes. (c) Symlet de 8 coeficientes. (d) Symlet de 16 coeficientes.	43
Figura 2-16 Esquema de descomposición de la transformada Wavelet Discreta en 2D.	44
Figura 2-17 Esquema de reconstrucción de la transformada Wavelet Discreta en 2D.	45
Figura 2-18 Ejemplo de DWT. (a) Imagen original de tamaño 512×512. (b) Descomposición en dos niveles, utilizando la wavelet Coif3. (c) Reconstrucción de la imagen.	46
Figura 2-19 Efecto de la traslación de la base biortogonal bior4.4 (Nomenclatura de Matlab). Cuando dos señales de entrada son exactamente las mismas, pero una está desplazada una muestra a la derecha. El primer renglón representa la señales de entrada, los siguientes renglones la primera y segunda descomposición wavelet respectivamente. El desplazamiento provoca una reorganización de los valores de los coeficientes. Ejemplo tomado con permiso de la tesis de doctorado de Rafael Redondo del CSIC, España (Redondo, 2007).	47
Figura 3-1 Imágenes sintéticas con diferente contraste. (a) Imagen con alto contraste. (b) Imagen con bajo contraste. (c) Imagen fusionada en FUSIONLAB, la regla de selección de coeficientes empleada es AVERAGE. Se aprecia que este tipo de métodos aritméticos como el promediado de coeficientes no son los más adecuados ya que se obtienen imágenes grises.	54
Figura 3-2 Esquemas de fusión de imágenes. (a) Fusión a nivel pixel. (b) Fusión a nivel características. (c) Fusión a nivel decisión o fusión simbólica.	55
Figura 3-3 Ejemplos de fusión de imagen pancromática y multiespectral. Las imágenes cubren un área de la ciudad de Sherbrooke, Canadá, tomadas el 20 de mayo de 2001. (a) Imagen pancromática de 1 m de resolución. (b) Imagen multiespectral de 4 m de resolución. (c) Imagen fusionada utilizando IHS. (d) Imagen fusionada utilizando BT. (e) Imagen fusionada utilizando PCA. (f) Imagen fusionada utilizando HPF. (g) Acercamiento 200% de la imagen fusionada con IHS. (h) Acercamiento 200% de la imagen fusionada con BT. (i) Acercamiento 200% de la imagen fusionada con PCA. (j) Acercamiento 200% de la imagen fusionada con HPF. Fuente (Zhijun, et al., 2005).	61

Figura 3-4 Esquema de fusión de imágenes basado en la transformada Wavelet Discreta (DWT). 64	64
Figura 3-5 Fusión de imágenes basada en comparación de energía pixel a pixel. (a) y (b) son las imágenes de entrada. (c) y (d) Son las imágenes fusionadas resultantes utilizando Db3 y los criterios de fusión son MAX-MAX y AVERAGE-AVERAGE, respectivamente.66	66
Figura 3-6 Fusión de imágenes basada en comparación de energía pixel a pixel. (a) y (b) Son las imágenes de entrada. (c) y (d) Son las imágenes fusionadas resultantes utilizando Db2, usando los criterios de fusión MIN-MIN y AVERAGE-AVERAGE, respectivamente.67	67
Figura 4-1 Esquema general para un modelo de calidad perceptible..... 72	72
Figura 4-2 Medida de la respuesta del SVH a cambios de intensidad. 74	74
Figura 4-3 Fenómenos relativos al procesamiento de contraste. (a) Contraste simultáneo. (b) Bandas de Mach. 75	75
Figura 4-4 CSF del sistema visual (Campbell & Green, 1965)..... 78	78
Figura 4-5 Curva de umbrales de elevación causados por enmascaramiento espacial. 79	79
Figura 4-6 Imágenes con PNSR = 26 dB. 82	82
Figura 4-7 Esquema del sistema de medida de SSIM. 84	84
Figura 4-8 Versiones distorsionadas de la imagen de un perico 256 × 256. (a) Imagen original. (b) Imagen distorsionada con ruido Gaussiano blanco aditivo. (c) Reconstrucción JPEG con un factor 88. (d) Imagen con interferencia de malla..... 87	87
Figura 4-9 Curvas de adaptación de luminancia. 88	88
Figura 4-10 CSF de Rust. (a) CSF para $\Theta = 45^\circ$ y 90° en 1D. (b) CSF 2D. 89	89
Figura 4-11 CSF de Daly. (a) CSF en 1D con $\Theta = 0^\circ$ y 2 valores de d . (b) CSF en 2D. 90	90
Figura 4-12 CSF de Barten. 91	91
Figura 4-13 Filtros de Gabor para 4 escalas y 4 orientaciones más el paso bajo. (a) En 1D en el dominio de frecuencia. (b) En 2D en el dominio espacial, parte real. El residuo paso bajo aparece representado en la primera fila de la presente figura..... 93	93
Figura 4-14 (a) Imagen original Lena. (b) Filtro Gabor de canal ($k = 3, l = 2$) en frecuencia. (c) Reconstrucción del canal ($k = 3, l = 2$)..... 93	93
Figura 5-1 Esquema del algoritmo de fusión propuesto. La fusión orientada a características extrae los bordes de ambas imágenes y los conserva durante todo el proceso de fusión..... 99	99
Figura 5-2 Primera derivada de Gaussiana en 2D..... 102	102
Figura 5-3 Supresión de <i>No Máximos</i> , en la dirección normal al borde. 103	103
Figura 5-4 Detección de bordes; (a) Imagen Original. (b) $\sigma = 2.5$. (c) $\sigma = 1.0$. (d) $\sigma = 0.6$ 104	104
Figura 5-5 Imágenes de entrada multifocales de 512 × 512. Servilletero y manzana. (a) Imagen de referencia o <i>ground truth</i> . (b) Imagen con el servilletero desenfocado y la manzana enfocada. (c) Imagen con el servilletero enfocado y la manzana desenfocada. 106	106
Figura 5-6 Resultado de los esquemas de fusión para el experimento Servilletero y manzana. (a) Imagen obtenida con la wavelet Haar a un nivel de descomposición, el criterio de selección de coeficientes fue MAX-MAX. (b) Imagen obtenida con Haar a un nivel de descomposición, los coeficientes se mezclaron usando el filtro de Canny y la regla de fusión MAX-MAX. (c) Imagen obtenida utilizando Haar con dos niveles de descomposición y la regla MAX-MAX. (d) Imagen obtenida usando Haar a dos niveles de descomposición y el filtro Canny, la regla de fusión fue MAX-MAX..... 107	107
Figura 5-7 Mapas de decisión para el experimento Servilletero y manzana. (a) y (b) Mapas de decisión del esquema 2. (c) y (d) Mapas de decisión del esquema 4. 108	108

Figura 5-8 Imágenes de entrada multifocales. Globos. (a) Imagen de referencia. (b) Imagen con los globos de la derecha desenfocados y el resto de la imagen en foco. (c) Imagen con el globo de la izquierda y los tres globos de la parte superior desenfocados.....	110
Figura 5-9 Resultado de los esquemas de fusión para el experimento Globos. (a) Imagen obtenida usando Cf6 con un nivel de descomposición y la regla de fusión AV-AV. (b) Imagen obtenida con Cf6 a un nivel de descomposición usando el filtro de Canny y la regla de fusión AV-AV. (c) Imagen obtenida usando Sym4 a un nivel de descomposición y la regla de fusión AV-AV. (d) Imagen obtenida con Sym4 a un nivel de descomposición usando Canny y la regla de fusión AV-AV.....	111
Figura 5-10 Mapas de decisión para el experimento Globos. (a) y (b) Mapas de decisión del esquema 2. (c) y (d) Mapas de decisión del esquema 4.	112
Figura 5-11 Imágenes de entrada multifocales. Pelota y fondo. (a) Imagen de referencia. (b) Imagen con el fondo desenfocado y el primer plano en foco. (c) Imagen con el fondo en foco y el primer plano desenfocado.	114
Figura 5-12 Resultado de los esquemas de fusión para el experimento Pelota y fondo. (a) Imagen obtenida aplicando Db3 a un nivel de descomposición y la regla de fusión MAX-MAX. (b) Imagen obtenida aplicando Db3 a un nivel de descomposición y el filtro de Canny, la regla de fusión es MAX-MAX. (c) Imagen obtenida aplicando Db3 a un nivel de descomposición y la regla de fusión AV-MAX. (d) Imagen obtenida usando Db3 a un nivel de descomposición y el filtro de Canny, la regla de fusión es AV-MAX.	115
Figura 5-13 Mapas de decisión para el experimento Pelota y fondo. (a) y (b) Mapas de decisión del esquema 2. (c) y (d) Mapas de decisión del esquema 4.	116
Figura 5-14 Experimento Relojes. (a) y (b) son las imágenes de entrada. (c) Imagen obtenida con Db2 a una escala y la regla de fusión AV-MAX. (d) Imagen obtenida con Db2 a un nivel de descomposición, aplicando Canny y la regla de fusión AV-MAX.	119
Figura 5-15 Mapas de decisión para el experimento Relojes. (a) y (b) Mapas de decisión del esquema 2.....	119
Figura 5-16 Experimento Refresco. (a) y (b) Imágenes de entrada. (c) Imagen obtenida con Sym4 y la regla de fusión MAX-MAX. (d).Imagen obtenida con Sym4 y el filtro de Canny, la regla de fusión es MAX-MAX.	121

R

Resumen.

"Turn away with fear and horror from this lamentable plague of continuous functions that do not have a derivative"
Charles Hermite

Resumen.

La búsqueda de métodos prácticos para combinar imágenes dio origen al concepto de fusión, el cual se utiliza para obtener una mejor interpretación de los datos integrados y reducir la ambigüedad, aumentando la precisión de la información. Los algoritmos de combinación están orientados hacia esquemas de fusión basados en píxeles, donde se ha obtenido cierto grado de madurez. Sin embargo, debido a que las características útiles en una imagen son generalmente más grandes que un píxel, las operaciones píxel por píxel no son las más apropiadas. En este trabajo de tesis se hace una revisión de estos métodos y se presenta la transformada Wavelet que descompone una imagen en una aproximación y varios detalles a diferentes escalas y en diversas direcciones. La aproximación contiene la información promedio y casi toda la energía de la imagen, mientras que los detalles contienen bordes o información de altas frecuencias.

Se propone una regla de fusión orientada a características. Tales características son los bordes que se extraen en el dominio de la transformada Wavelet utilizando el filtro de Canny. La evaluación se realiza mediante medidas tradicionales como el error cuadrático medio (MSE), la relación señal a ruido de pico (PNSR) y con esquemas de calidad perceptiva. Se implementa una modificación de la transformada Wavelet sobre el intervalo y se desarrollan dos *toolboxes*, uno en Java para fusión y otro en Matlab para la evaluación de la calidad en las imágenes.

Finalmente, los resultados demuestran que la regla de fusión orientada a características tiene un mejor rendimiento cuantitativo y cualitativo que las reglas tradicionales.

Palabras clave: *Fusión de Imágenes, Transformada Wavelet, Extracción de Características, Métricas de Calidad Perceptiva.*

Antecedentes.

*“Everything should be made as simple as possible, but not simpler”
Albert Einstein*

1.1 *Introducción.*

En la presente década, con los extraordinarios avances en tecnología de sensores, microelectrónica y comunicaciones, ha crecido la necesidad de diseñar técnicas de procesamiento que puedan combinar eficazmente la información proveniente de múltiples fuentes. En aplicaciones basadas en imágenes, el concepto de fusión ha surgido como un área de investigación novedosa que debe ser explotada.

Las técnicas de fusión proveen los medios para integrar datos de múltiples sensores y de información relacionada entre sí para lograr inferencias más específicas que las obtenidas a partir de un sólo sensor. El objetivo de la fusión de datos está orientado a obtener una mejor interpretación de los datos integrados y reducir la ambigüedad, aumentando la precisión de la información (Rogers, 1990). Por ejemplo, la fusión de imágenes médicas de resonancia magnética y tomografía computarizada pueden facilitar un diagnóstico médico.

El concepto de fusión de datos no es un tema nuevo, la mayoría de las técnicas para combinar o fusionar datos provienen de un conjunto de disciplinas tradicionales como el procesamiento digital de señales, la estimación estadística, la inteligencia artificial, la teoría de control entre otras.

En la naturaleza, la fusión de múltiples sensores es un proceso que los humanos y los animales han desarrollado para mejorar la capacidad de adaptación y sobrevivencia. Por ejemplo, cuando la visión de un animal es limitada, el sentido del olfato y el oído pueden proporcionar información para advertirle de depredadores cercanos. De manera similar, para juzgar la calidad de un vino no basta el sentido del gusto, es necesario utilizar los sentidos de la vista y el olfato. En ambos casos, el proceso de fusión nos permite juzgar de forma más precisa el ambiente que nos rodea.

Con la tecnología actual, se han incrementado las posibilidades de emular la capacidad natural de los seres humanos y animales para fusionar datos. La fusión de datos ha dejado de ser un conjunto de técnicas dispersas que se relacionan entre sí y se ha convertido en una disciplina con sistemas estandarizados y principios propios.

Las aplicaciones relacionadas con la fusión de datos son muy amplias. Como ocurre frecuentemente con la tecnología, las aplicaciones militares han impulsado la práctica de la fusión de datos. Actualmente las técnicas de fusión se utilizan en el rastreo e identificación de objetivos (armas inteligentes) (Hall, 1992), en los sistemas de vigilancia submarina como apoyo a las operaciones navales, en los sistemas de defensa tierra-aire y aire-aire para detección de aeronaves, en la conducción de vehículos autónomos, etc.

Cada una de estas aplicaciones involucra un foco u objetivo, un conjunto de sensores y un conjunto de inferencias, Figura 1-1. El reto es permitir la toma de decisión más precisa y rápida.

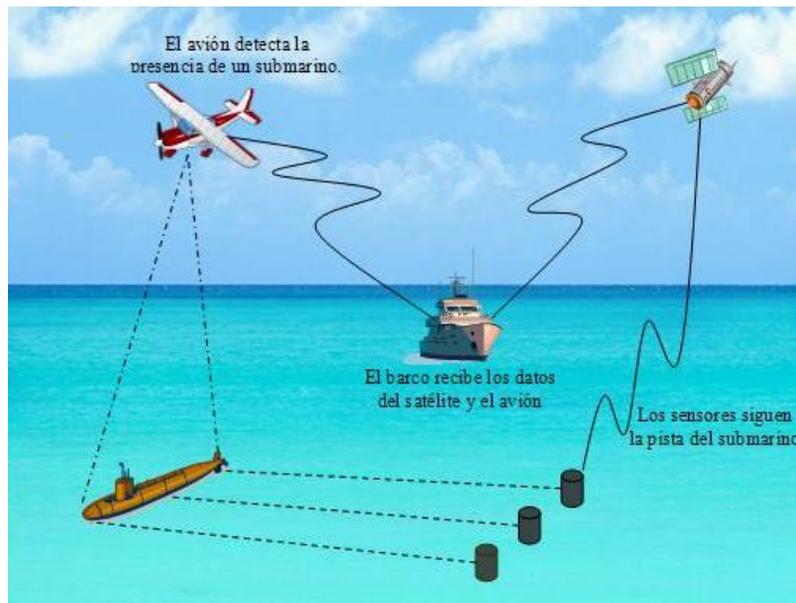


Figura 1-1 Sistema de vigilancia multisensor.

Por otro lado, las aplicaciones no militares se agrupan en problemas académicos, comerciales e industriales como son la implementación de robots, el control automático de sistemas de manufactura industrial, el desarrollo de edificios inteligentes, el monitoreo de equipo mecánico complejo (turbomaquinaria, turbinas de helicópteros), los sistemas de percepción remota y las aplicaciones médicas como la imagenología médica (rayos X, tomografía computarizada, resonancia magnética, ultrasonido, etc.) (Muñoz, 2005), (Klein, 1993).

Hablando particularmente de aplicaciones en el área de imágenes, la fusión de datos se utiliza para lograr realce espacial. Si fusionamos una imagen pancromática de alta resolución con una imagen multiespectral de baja resolución, podemos obtener una imagen que preserve la información

espectral y al mismo tiempo incorpore el contenido de alta resolución espacial, teniendo como resultado mayor detalle (Franklin, et al., 1993).

La fusión de datos también se utiliza para mejorar la precisión de la clasificación de imágenes, la redundancia en la información ayuda a discriminar las diferentes clases dentro de una imagen (Schistad-Soldberg, et al., 1994).

Otra aplicación es el caso de las imágenes multifocales, Figura 1-2. Debido a las limitaciones de los lentes ópticos para enfocar en los dispositivos CCD, no es posible obtener una imagen que contenga todos los objetos relevantes “en foco”. El proceso de fusión puede ayudar a seleccionar de entre varias imágenes los objetos enfocados para obtener la imagen deseada.

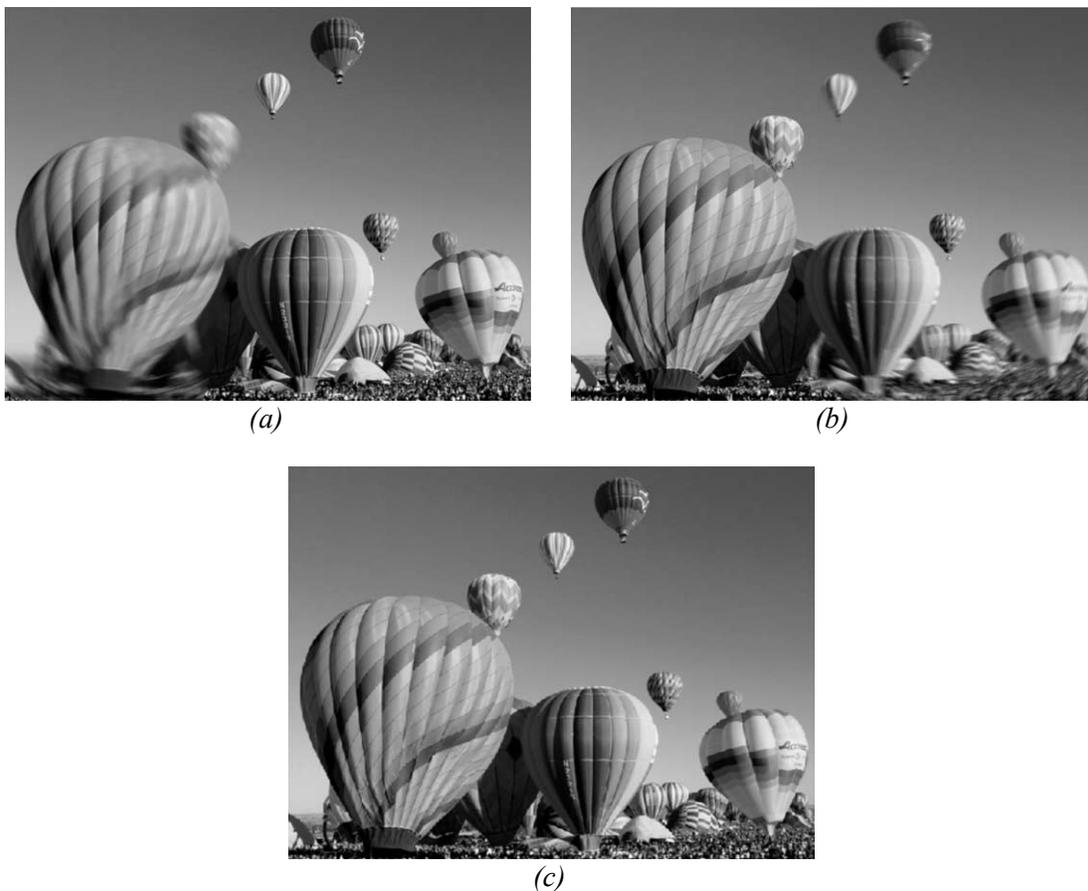


Figura 1-2 Imágenes multifocales. (a) Imagen con el lado izquierdo desenfocado. (b) Imagen con el lado derecho desenfocado. (c) Imagen fusionada.

En el caso de percepción remota, se fusionan los datos de sensores multispectrales para monitorear el clima, los recursos agrícolas y los desastres provocados por fenómenos naturales entre otras aplicaciones. Algunos de estos sensores, en los satélites, proporcionan las bandas necesarias para

distinguir características espectrales, pero no espaciales, mientras que otros sensores proporcionan resolución para distinguir características espaciales. Una aplicación interesante aparece en (Zhuo, et al., 1998), donde se combinan datos de Landsat TM y SPOT.

Otro ejemplo en el área de imágenes médicas es la fusión de resonancia magnética (MR) con tomografía computarizada (CT), Figura 1-3.

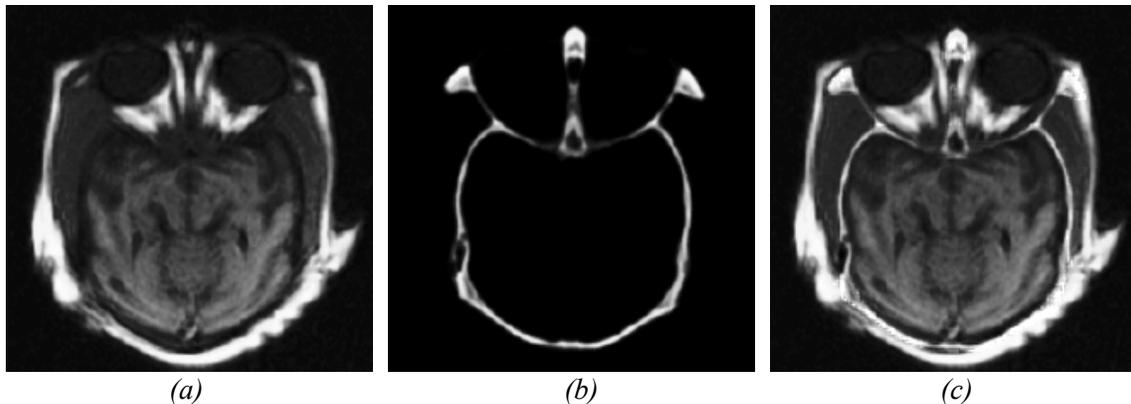


Figura 1-3 Fusión de imágenes de MR y CT: (a) Imagen de MR. (b) Imagen de CT. (c) Imagen compuesta obtenida utilizando la herramienta FUSIONLAB, la regla de selección empleada es MAX-MAX.

1.2 Esquemas de fusión.

Dependiendo de la etapa donde tenga lugar la fusión de datos, se pueden distinguir tres esquemas: (1) fusión directa de los datos de los sensores, (2) fusión basada en características ó (3) fusión basada en decisiones.

Si los datos de los sensores miden los mismos fenómenos físicos, entonces pueden ser fusionados directamente. La fusión de datos crudos ocurre en la primera etapa del proceso de fusión, siendo las técnicas utilizadas para este tipo de fusión de datos las basadas en métodos de estimación clásica como el filtrado de Kalman y operaciones aritméticas.

La fusión basada en características requiere la extracción de objetos o regiones representativas de los datos de entrada, que se conservan en el proceso de fusión. Este tipo de fusión involucra reconocedores de patrones, redes neuronales y algoritmos de cúmulos (*clustering*).

La fusión basada en decisiones combina información después de que cada sensor ha determinado preliminarmente los atributos de los datos, realizándose mediante la aplicación de la inferencia bayesiana o algún otro método como Dempster-Shafer (Huadong, et al., 2002).

1.3 Ventajas de la fusión de datos.

Si utilizamos varios sensores idénticos (p. ej. radares que siguen el movimiento de un objetivo), al combinar las observaciones obtenidas, tendremos una mejor estimación de la posición y velocidad de dicho objetivo. De hecho, obtendremos una ventaja estadística agregando N observaciones independientes. Una segunda ventaja involucra el uso de varios sensores para mejorar la observación de un proceso. Si dos sensores miden direcciones angulares, éstos pueden ser coordinados para determinar la posición de un objeto mediante triangulaciones. Una tercera ventaja es la redundancia en los datos que se obtiene al utilizar múltiples sensores (Waltz, et al., 1990).

Entre otras ventajas que proporciona la fusión de datos sobre la información obtenida a partir de un sólo sensor se pueden listar:

- Robustez y fiabilidad. Los sistemas basados en múltiples sensores pueden seguir operando incluso si una o varias fuentes de información no funcionan.
- Mayor cobertura en espacio y tiempo.
- Incremento de la dimensionalidad del espacio de datos. Esto aumenta la calidad para inferir información y reduce la vulnerabilidad de los sistemas que utilizan múltiples sensores.
- Reducción de ambigüedades. Información más completa proporciona mejor capacidad para discriminar entre todas las hipótesis disponibles.
- Proporciona una solución a la explosión de la información que se dispone en la actualidad.

1.4 Estado del arte.

Las tecnologías relacionadas con la fusión de datos de múltiples sensores han evolucionado rápidamente. Existe investigación para desarrollar nuevos algoritmos y reglas de fusión, así como para mejorar los algoritmos existentes. Además, se intenta unir todas estas técnicas en una sola arquitectura.

La fusión de imágenes multiespectrales, más que de ningún otro tipo de sensores, ha obtenido un crecimiento mayor. Los algoritmos de combinación están orientados hacia esquemas de fusión basados en píxeles, donde se ha obtenido cierto grado de madurez. Sin embargo, debido a que las características útiles en una imagen son generalmente más grandes que un píxel, las operaciones píxel por píxel no son las más apropiadas.

El esquema de fusión de imágenes más utilizado es la transformación por Análisis de Componentes Principales, que realza la resolución espacial pero produce distorsión espectral (Ehlers, 1991).

La transformada Hermitiana (Escalante-Ramírez, et al., 2003) es otra herramienta que se utiliza para la fusión de imágenes, la cual incorpora algunas propiedades importantes de percepción visual, como

el análisis a través de campos receptivos traslapados, permitiendo la construcción de esquemas piramidales multirresolución.

Existen también esquemas que proponen reglas de fusión basados en áreas o regiones, en las cuales las imágenes se dividen en pequeñas ventanas de 3×3 ó 5×5 y se calcula la actividad asociada con el píxel central.

La transformada Wavelet (Daubechies, 1992) es un método para el análisis multirresolución. Este método descompone una imagen en una aproximación y varios detalles a diferentes escalas y en diversas direcciones. La aproximación contiene la información promedio y casi toda la energía de la imagen, mientras que los detalles contienen bordes o información de frecuencias altas.

El esquema de fusión utilizando la transformada Wavelet ofrece varias ventajas sobre los esquemas de fusión tradicionales ya que proporciona información direccional, mientras que otras representaciones como la laplaciana no presenta orientación espacial en el proceso de descomposición (Li, et al., 1995). En este contexto, las wavelets de Gabor han demostrado que pueden representar a la información de manera óptima (Fischer, et al., 2006). La forma de esta wavelet es Gaussiana tanto en el dominio espacial como frecuencial, por lo que el filtro está estrechamente localizado tanto en el espacio como en frecuencia. Otra característica importante es que esta representación es adecuada en la descripción del procesado del sistema visual. Se ha demostrado que las funciones de Gabor pueden modelar el comportamiento de las células simples de la corteza visual.

Hasta ahora, la mayoría de las reglas de fusión son de naturaleza heurística y se basan en calcular la energía de un píxel o una vecindad. La evidencia confirma que los humanos utilizamos funciones cognitivas basadas en características para reconocer objetos, por lo que la siguiente frontera a alcanzar es diseñar una regla de fusión, también basada en características, que simule la forma en que los seres humanos procesamos la información de nuestros sentidos.

1.5 Modelos de calidad perceptiva.

Parte importante de este trabajo de tesis está orientado a evaluar el desempeño de los algoritmos de fusión midiendo la calidad perceptiva de las imágenes fusionadas. Sin embargo, debido a la ausencia de una métrica universal que proporcione una medida de calidad aceptable en un amplio rango de situaciones, se presentarán varias propuestas para evaluar la calidad de las imágenes.

Para obtener una estimación precisa de la calidad de una imagen se realizan pruebas subjetivas con observadores reales (seres humanos frente a un monitor), pero generalmente resultan ser procesos muy largos, costosos y difíciles de repetir, ya que requieren de condiciones de visualización especiales. Debido a estos inconvenientes, se utilizan métricas de calidad objetiva. Las más conocidas y que se emplean en casi todos los casos son las medidas matemáticas basadas en la relación señal a ruido de pico (PSNR), el error cuadrático medio (MSE), el error absoluto medio (MAE), etc. Sin embargo, estas medidas no guardan ninguna relación directa con la calidad de la imagen percibida, razón por la cual se busca desarrollar métricas más robustas que incorporen de alguna manera factores perceptivos.

El desarrollo de nuevas métricas objetivas ha sido posible gracias a los avances en el conocimiento del sistema de visión humano (SVH) y su funcionamiento. Las métricas de calidad perceptiva modelan en mayor o menor medida las primeras etapas del proceso de visión, desde que la luz es percibida por el ojo hasta la corteza visual primaria. Todos los modelos existentes comparten una estructura similar, en ellos se encuentra una etapa de calibración, filtrado de la imagen a diferentes frecuencias espaciales y orientaciones, ajustes de la sensibilidad al contraste (CSF), mecanismos para modelar el fenómeno del enmascaramiento y una etapa final para obtener el error como un número que describe la calidad de la imagen o como un mapa de distorsión.

La implementación de estas métricas es compleja y presenta un costo computacional elevado y aunque existen limitaciones como el hecho de no tomar en cuenta características del SVH de alto nivel (extracción de características, procesos cognitivos, reconocimiento de patrones, atención visual, etc.), presentan un avance con relación a las medidas tradicionales (Nava, et al., 2006).

1.6 Prerrequisitos y desarrollo de software.

Hasta ahora se ha presentado un breve panorama sobre la fusión de datos y su estado del arte. A partir de aquí, se hablará de la fusión de imágenes como un caso particular de la fusión de datos. Las reglas y algoritmos de fusión se diseñaron para el caso intermodal, aunque eventualmente podrían ser para imágenes multimodales. Los siguientes capítulos abordan la teoría del análisis multirresolución y la fusión de datos con el propósito de ser un trabajo auto contenido.

Las reglas de fusión de imágenes están basadas en métodos heurísticos y en comparaciones para discriminar que pixeles conservar, es por eso que son muy sensibles a errores de registro. En este trabajo de tesis se supondrá que todas las imágenes de entrada están alineadas con una precisión subpixel. Una guía interesante sobre registro de imágenes se puede encontrar en (Gottesfeld, 1992).

Parte del trabajo de esta tesis tiene que ver con desarrollo de software. Se programó una herramienta para fusión de imágenes llamada FUSIONLAB en lenguaje JAVA, versión 1.5, utilizando el IDE ECLIPSE y JAI (*Java Advanced Image*). Para la parte correspondiente a la evaluación de la calidad de imágenes se desarrollaron varios programas en MATLAB, versión R14. Finalmente, para la transformada de Wigner, que se utiliza para calcular la relación señal a ruido de Wigner (SNR_w), se usó una librería en C++.

1.7 Descripción de la tesis.

En las secciones previas se ha abordado brevemente el concepto de fusión y su estado del arte. En los siguientes capítulos los temas tratados aquí se profundizarán, para ello se ha dividido la tesis de la siguiente forma: en el capítulo dos se presenta una introducción al análisis multirresolución, su desarrollo histórico y la teoría wavelet. En el capítulo tres se revisan los métodos tradicionales de fusión y se muestran algunos ejemplos. En el capítulo cuatro se habla de las métricas de calidad perceptiva. En el capítulo cinco se presenta el método de fusión propuesto, se realizan experimentos

y comparaciones con los métodos tradicionales. Finalmente, en el capítulo seis, se muestran las conclusiones y trabajo futuro.

1.8 Referencias.

- **Daubechies I** Ten lectures on wavelets [Book]. - Filadelfia, E. U. : SIAM, 1992.
- **Ehlers M** Multisensor image fusion techniques in remote sensing [Journal] // ISPRS Journal of Photogrammetry and Remote Sensing. - 1991. - 1 : Vol. 46. - pp. 19-30.
- **Escalante-Ramírez B and López-Caloca A** Image fusion with the Hermite transform [Conference] // International Conference on Image Processing. - 2003. - Vol. 2. - pp. 145-152.
- **Fischer S, Cristobal G and Redondo R** Sparse overcomplete Gabor wavelet representation based on local competitions [Journal] // IEEE Transactions on Image Processing. - 2006. - 2 : Vol. 15. - pp. 265-272.
- **Franklin S and Blodgett C** An example of satellite multisensor data fusion [Journal] // Computer and Geoscience. - 1993. - 4 : Vol. 19. - pp. 577-583.
- **Gottesfeld L** A survey of image registration techniques [Journal] // ACM Computing Surveys. - 1992. - 4 : Vol. 24. - pp. 325-352.
- **Hall D** Mathematical Techniques in Multisensor Data Fusion [Book]. - Boston, E. U. : Artech House, 1992.
- **Huadong W [et al.]** Sensor fusion using Dempster-Shafter theory [Conference] // Instrumentation and Measurement Technology Conference. - 2002. - Vol. 1. - pp. 7-12.
- **Klein L** Sensor and data fusion concepts and applications [Book]. - Bellingham, E. U. : SPIE Opt. Engineering Press, 1993. - Vol. 14 : p. 252.
- **Li H, Manjunath B and Mitra S** Multisensor image fusion using the wavelet transform [Journal] // Graphical Models and Image Processing. - 1995. - 3 : Vol. 57. - pp. 235-245.
- **Muñoz J** Fusión de imágenes médicas a través de la transformada de Hermite. [Book]. - Tesis de Maestría. Mexico : Universidad Nacional Autónoma de México, 2005.
- **Nava R, Gallego A and Cristóbal G** Una nueva herramienta para la evaluación de la calidad perceptible en imágenes [Conference] // Congreso Anual de la Sociedad Española de Ingeniería Biomedica. - Pamplona, España : [s.n.], 2006. - pp. 77-80.
- **Rogers R** The history and status of merging multiple sensor data: an overview [Conference] // ACSM-ASPRS Annual Convention Image Processing and Remote Sensing. - 1990. - Vol. 4. - pp. 352-360.
- **Schistad-Soldberg A, Jain A and Taxt T** Multisensor classification of remotely sensed data fusion of landsat TM and SAR images [Journal] // IEEE Transactions on Geoscience and Remote Sensing. - 1994. - 4 : Vol. 32. - pp. 768-778.
- **Waltz E and Llinas J** Multisensor data fusion [Book]. - Massachusetts, E. U. : Artech House, 1990.
- **Zhuo J, Civco D and Silander J** A wavelet transform method to merge landsat TM and SPOT panchromatic data [Journal] // International Journal of Remote Sensing. - 1998. - 4 : Vol. 19. - pp. 743-757.

Análisis Multirresolución.

*“An expert is someone who knows some of the worst mistakes that can be made in his subject and how to avoid them”
Werner Heisenberg*

2.1 *La necesidad del análisis multirresolución.*

Existen diversas formas para transformar una representación de una señal dada en otra. El ejemplo clásico es la transformada de Fourier, que descompone a una señal en ondas sinusoidales, que representan las componentes en frecuencia, las cuales a menudo son de gran importancia. Sin embargo, debido a la extensión de las funciones sinusoidales, las características locales de la señal se extienden sobre la representación entera quedando ocultas.

La forma en la cual las señales se pueden representar es importante en cualquier tarea de procesamiento. Distintas representaciones pueden enfatizar diferentes aspectos de la señal y hacer accesible cierta información relevante. Cuando se trata de caracterizar a las señales en términos de sus propiedades espectrales surgen problemas relacionados con el hecho de que es imposible describir localmente dichas propiedades.

Una señal en el tiempo discreto puede ser considerada como una representación en términos de funciones delta, estas funciones están perfectamente localizadas en el tiempo pero no en la frecuencia. Utilizando el análisis de Fourier, tendremos el valor exacto en el tiempo pero no tendremos información del momento en que ocurren esas frecuencias. Por ejemplo, dada una función de energía finita f con transformada de Fourier F , podemos reducir su tiempo de duración escalando por un factor $0 < a < 1$, tal que $f(t) = f(t/a)$. La transformada de Fourier de F es una función dilatada $F = |a|F(a\omega)$, es entonces que perdemos localización en frecuencia y la ganamos en el tiempo. De otra manera, si reducimos la duración en frecuencia perdemos localización en tiempo. Esto sugiere que hay un intercambio entre la localización en tiempo y frecuencia, el cual puede ser visto como una expresión del principio de incertidumbre de Heisenberg.

Este es un problema importante ya que las singularidades e irregularidades de las señales en ocasiones poseen la información de mayor interés. Otra limitación del análisis de Fourier es que diverge en funciones que no pertenecen al espacio $L^2(\mathbb{R})$ y en señales que caracterizan fenómenos no estacionarios.

En imágenes, las discontinuidades en la intensidad pueden representar la localización de contornos o bordes, los cuales son particularmente interesantes si queremos realizar extracción de características. Para otros tipos de señales como los electrocardiogramas o señales de radar, la información de interés es proporcionada por los máximos locales que contienen respuesta específica a instantes de tiempo específico.

Si deseamos describir completamente una señal debemos analizar un amplio rango de resoluciones. En algunas ocasiones las señales presentan un mismo comportamiento a diferentes escalas, esta propiedad se conoce como invariancia a la escala. Para señales que no son invariantes a la escala, no se puede conocer en general una escala en la cual la señal contenga la información más relevante. En principio, todas las escalas tienen la misma importancia, debido a esto, si deseamos tener una interpretación adecuada de la información de la señal, esta debe ser analizada a diferentes escalas de forma uniforme.

Una representación a múltiples escalas proporciona un marco jerárquico para interpretar la información de una imagen o una señal. A diferentes resoluciones o escalas los detalles en una imagen caracterizan diferentes estructuras físicas de una escena; a una escala muy grande los detalles corresponderán a grandes estructuras que proporcionan el contexto de la imagen, mientras que a una baja escala apreciaremos los detalles más finos de la escena. Es natural entonces, analizar primero los detalles de la imagen a una alta escala y después gradualmente disminuirla.

Una forma de entender el concepto de escala o resolución es pensando en un bosque. Si observamos a una distancia muy alejada, el bosque parece una cubierta sólida verde; esto es porque observamos la escena a una escala muy alta y los detalles se pierden. A medida que nos acercamos reduciendo la escala podemos percibir mayores detalles como los árboles, las ramas y las hojas.

En el caso de imágenes, es difícil analizar la información contenida en los píxeles de una imagen tomando directamente los valores de grises, ya que dependen de las condiciones de luz. Como se mencionó líneas arriba, las variaciones de intensidad en las imágenes son de mucha importancia. Si se desea analizar los objetos en las imágenes se debe adaptar el tamaño de la vecindad donde se calcula el contraste al tamaño de los objetos de estudio. El tamaño de la vecindad define la resolución para medir las variaciones locales (Rosenfeld, et al., 1971). Generalmente, las estructuras que queremos reconocer tienen diferentes tamaños, de aquí que no sea posible definir *a priori* una resolución óptima para analizar imágenes.

Existen diversos trabajos sobre algoritmos que procesan las imágenes a diferentes escalas (Hall, et al., 1976), (Marr, et al., 1979) y permiten reorganizar la información en un conjunto de detalles con distinta resolución. Sin embargo, todos poseen una idea en común. Dicha idea es calcular las aproximaciones de las señales en distintas escalas o resoluciones mediante proyecciones en diferentes espacios V_k que satisfacen (2.1), (2.2) y (2.3).

$$V_{k+1} \subset V_k \quad \dots (2.1)$$

$$\bigcup_{j \in \mathbb{Z}} V_j = L^2(\mathbb{R}) \quad \dots (2.2)$$

$$\bigcap_{j \in \mathbb{Z}} V_j = 0 \quad \dots (2.3)$$

2.2 Representaciones multirresolución.

Aunque este trabajo de tesis se enfoca a la representación wavelet, existen algunas otras representaciones multirresolución bastante populares que se mencionarán brevemente en los siguientes párrafos.

Una de las primeras aproximaciones multirresolución es la representación *piramidal* propuesta por (Burt, et al., 1983) y (Crowley, 1987). Un esquema piramidal consiste en una secuencia de versiones de una imagen original donde la resolución es gradualmente reducida mediante filtrado y submuestreo. El nivel más bajo o nivel cero x^0 de la pirámide es igual a la imagen original x , esta imagen es filtrada con un filtro paso-bajas y submuestreada para obtener el siguiente nivel x^1 el cual es filtrado y submuestreado de la misma forma para obtener el nivel x^2 , este procedimiento genera los subsecuentes niveles de la pirámide.

Para ilustrar este método, considere una imagen discreta x de tamaño $m \times n$, donde $x \in \mathbb{Z}^2$ y $x^0 = x$. Cada imagen sucesiva x^{k+1} , $k \geq 0$, se construye a partir del nivel anterior de la siguiente forma:

$$x^{k+1}(m, n) = \sum_{i \in \mathbb{Z}} \sum_{j \in \mathbb{Z}} w(i, j) x^k(2m - i, 2n - j) \quad \dots (2.4)$$

donde $w \in \mathbb{Z}^2$ son los coeficientes del filtro paso-bajas.

Generalmente se prefiere un filtro separable $w(m, n) = h(m)h(n)$ donde h es de longitud $2N + 1$. Si se elige adecuadamente el filtro h , la representación en cada nivel sucesivo corresponderá a estructuras más y más burdas o de más baja frecuencia. Burt y Adelson (Burt, et al., 1983) propusieron el siguiente criterio de selección de los coeficientes.

- Normalización: $\sum_{i=-N}^N h(i) = 1$
- Simetría: $h(i) = h(-i)$ para $\forall i$
- Contribución equitativa: $\sum_{i=-N}^N h(2i) = \sum_{i=-N}^N h(2i+1)$ esto es, que todas las muestras en un nivel dado contribuyen de la misma forma para obtener el siguiente nivel.

Una longitud mayor del filtro h incrementa el grado de libertad en el diseño, pero el costo computacional también aumenta. Burt y Adelson (Burt, et al., 1983) proponen una longitud igual a cinco. El filtro w más utilizado es el filtro Gaussiano y a la pirámide construida se le llama *pirámide Gaussiana*.

Mediante interpolación de cada imagen x^{k+1} de la pirámide Gaussiana y restándola de su predecesor, podremos obtener la diferencia de Gaussianas (*DOG*). La operación de interpolación está definida en (2.5) como:

$$\hat{x}^k(m, n) = 4 \sum_{i=-N}^N \sum_{j=-N}^N w(i, j) x^{k+1} \left(\frac{m-i}{2}, \frac{n-j}{2} \right) \quad \dots (2.5)$$

donde la imagen \hat{x}^k puede ser interpretada como la predicción de x^k . De este modo, los detalles de la imagen y^{k+1} de la pirámide laplaciana corresponden al error de aproximación x^k por \hat{x}^k .

$$y^{k+1}(m, n) = x^k(m, n) - \hat{x}^k(m, n) \quad 0 \leq k < K \quad \dots (2.6)$$

donde k corresponde al último nivel de la pirámide gaussiana.

Existe otro método multirresolución llamado *quadrees* (Hunter, et al., 1979) que divide a una imagen en pequeños cuadrantes de manera sucesiva. Si el cuadrante no satisface algún criterio de homogeneidad (p. ej. todos los píxeles deben tener un valor similar), entonces es subdividido y así sucesivamente hasta que todos los cuadrantes son homogéneos.

Este proceso puede ser realizado utilizando una estructura de datos tipo árbol, la imagen es representada por un nodo raíz, mientras que los subsecuentes cuadrantes son los nodos hijos y las hojas de los nodos corresponden a aquellos cuadrantes que no necesitaron seguir dividiéndose. Los *quadrees* son un método bastante rápido y se utiliza en algoritmos de reconocimiento y segmentación.

Las representaciones *espacio-escala*, por otra parte, forman un tipo especial de representación multirresolución que comprende un parámetro de escala continuo y preserva el muestreo espacial en todas las escalas. La introducción de esta teoría fue hecha por Witkin (Witkin, 1983) y Koenderink (Koenderink, 1984), quienes fueron los primeros en crear la noción de escala.

En la literatura se pueden encontrar algunas otras técnicas multirresolución basadas en *morfología matemática* (Soille, 1999). La morfología es una herramienta que basada en la teoría de conjuntos y se utiliza para extraer contornos y para dividir regiones basándose en el contenido de texturas. La morfología posee dos operaciones básicas: dilatación y erosión.

Definición 2.1

Sea $x(m, n) \in \mathbb{Z}^2$ una imagen binaria de entrada y $b(m, n)$ un elemento estructurante. La dilatación de x por b , representada por $x \oplus b$ se define como:

$$(x \oplus b)(s, t) = \max \{x(s - m, t - n) + b(m, n)\} \quad \dots(2.7)$$

la condición es que los dos conjuntos se traslapen al menos un elemento.

Definición 2.2

Sea $x(m, n) \in \mathbb{Z}^2$ una imagen binaria de entrada y $b(m, n)$ un elemento estructurante. La erosión representada por $x \ominus b$ se define como:

$$(x \ominus b)(s, t) = \min\{x(s - m, t - n) - b(m, n)\} \quad \dots (2.8)$$

De la misma forma que en la definición anterior, ambos conjuntos deben traslaparse al menos en un elemento. Adicionalmente, mediante la composición de estas dos operaciones básicas se pueden definir las operaciones de apertura y cierre (2.9) y (2.10) respectivamente.

$$x \circ b = (x \ominus b) \oplus b \quad \dots (2.9)$$

$$x \cdot b = (x \oplus b) \ominus b \quad \dots (2.10)$$

La teoría wavelet también habrá que mencionarla como otro método multirresolución muy popular. Las wavelets son funciones fuertemente localizadas en tiempo y frecuencia y pueden ser utilizadas para descomponer una señal en diferentes bandas de frecuencia con diferentes resoluciones de tiempo. Para el caso del análisis de imágenes nos interesa particularmente la transformada Wavelet Discreta (DWT).

La representación wavelet consiste de una banda paso-bajas y una banda paso-altas que se obtiene de forma iterativa. Esta transformación es invertible y no redundante y difiere de la descomposición piramidal, la cual es redundante. Ambas características antes mencionadas hacen de la transformada Wavelet una herramienta altamente eficiente, aplicable a un amplio rango de señales y a tareas de procesamiento de imágenes como compresión y eliminación de ruido (*denoising*).

Antes de abordar a detalle el análisis multirresolución estructurado y completado por Stéphane Mallat e Yves Meyer (Mallat, 1989), se presentan algunas definiciones del álgebra lineal relacionadas con el análisis multirresolución en el contexto de la teoría Wavelet y se revisará el concepto de espacios de Hilbert.

2.3 Espacios vectoriales.

En muchas ramas de las matemáticas se presentan conjuntos con elementos de diversa naturaleza como las matrices, los números reales, los polinomios, etc. Tales conjuntos junto con las leyes de adición y multiplicación por un escalar tienen en común un gran número de propiedades algebraicas y poseen una estructura común; tales estructuras tienen el nombre de espacios vectoriales.

De forma genérica, a los elementos de un espacio vectorial se les llama vectores, por lo que en este contexto, la palabra vector adquiere un significado más amplio que el de cantidad con magnitud, dirección y sentido.

Definición 2.3

Un conjunto no vacío V de números reales \mathbb{R} o números complejos \mathbb{C} es un espacio vectorial si tiene definidas dos leyes de composición llamadas adición y multiplicación por un escalar tal que:

- i. La adición asigna a cada pareja $x, y \in V$ un único elemento $x + y \in V$, llamado la suma de x y y .
- ii. La multiplicación por un escalar asigna a cada pareja $\alpha \in K, x \in V$ un único elemento $\alpha x \in V$ llamado el producto de α por x . K es el conjunto de escalares.
- iii. Conmutatividad en la suma. $x + y = y + x$
- iv. Asociatividad en la suma. $(x + y) + z = x + (y + z), (\alpha\beta)x = \alpha(\beta x)$
- v. Distributividad en la suma. $z(x + y) = zx + zy, (\alpha + \beta)x = \alpha x + \beta x$
- vi. Idéntico aditivo. $\exists 0 \in V$, tal que $x + 0 = x, \forall x \in V$
- vii. Inverso aditivo. $\forall x \in V, \exists -x \in V$, tal que $x + (-x) = 0$
- viii. Idéntico multiplicativo. $1 \cdot x = x, \forall x \in V$

En algunas ocasiones $x, y \in V$ pueden ser n -tuplas o secuencias, entonces definimos:

$$x + y = (x_1, \dots, x_k) + (y_1, \dots, y_k) = (x_1 + y_1, \dots, x_k + y_k) \quad \dots (2.11)$$

$$\alpha x = \alpha(x_1, \dots, x_k) = (\alpha x_1, \dots, \alpha x_k) \quad \dots (2.12)$$

Habiendo definido los conceptos de adición y multiplicación por un escalar en un espacio vectorial, surgen los términos de combinación e independencia lineal.

Definición 2.4

Sea V un espacio vectorial y $x_1, \dots, x_k \in V$. Un vector $x \in V$ es llamado una combinación lineal de vectores si existen escalares $\alpha_1, \dots, \alpha_k$ tal que:

$$x = \alpha_1 x_1 + \dots + \alpha_k x_k \quad \dots (2.13)$$

Definición 2.5

Un conjunto de vectores $\{x_1, \dots, x_k\}$ son linealmente independientes si:

$$\alpha_1 x_1 + \dots + \alpha_k x_k = 0 \Leftrightarrow \alpha_1 = \alpha_2 = \dots = \alpha_k = 0 \quad \dots (2.14)$$

Un conjunto de vectores arbitrarios es linealmente independiente si cada uno de sus subconjuntos es linealmente independiente. Un conjunto de vectores que no son linealmente independientes son dependientes. También podemos decir que V es linealmente independiente si ningún $x \in V$ es una combinación lineal de un número finito de vectores de V diferentes de x .

Hasta ahora, la definición de espacio vectorial se refiere a propiedades algebraicas, sin embargo no han sido consideradas las propiedades del espacio vectorial que pueden ser medidas. Una forma de introducir el concepto de métrica en un espacio vectorial es con la operación conocida como producto interno.

Definición 2.6

Sea V un espacio vectorial sobre \mathbb{C} . Un producto interno en V es una función $V \times V$ en \mathbb{C} que asigna a cada pareja ordenada $\langle x, y \rangle \in V$ un escalar que satisface las siguientes propiedades:

- i. $\langle x + y, z \rangle = \langle x, z \rangle + \langle y, z \rangle$
- ii. $\langle x, \alpha y \rangle = \alpha \langle x, y \rangle$
- iii. $\langle x, y \rangle^* = \langle y, x \rangle$
- iv. $\langle x, x \rangle \geq 0$ y $\langle x, x \rangle = 0$ iff $x \equiv 0$

La idea de magnitud de un vector se traduce en un espacio vectorial en la idea de una norma, la cual se define a partir del producto interno como sigue:

Definición 2.7

Sea V un espacio vectorial sobre \mathbb{C} con producto interno. Se llama norma de $x \in V$ y se representa como $\|x\|$, al número real no negativo definido por:

$$\|x\| = \sqrt{\langle x, x \rangle} \quad \dots (2.15)$$

El producto interno puede ser usado para definir la ortogonalidad de dos vectores $x, y \in V$.

Definición 2.8

Dos vectores son ortogonales sí y sólo sí $\langle x, y \rangle = 0$

Si dos vectores son ortogonales, se denota $x \perp y$, y satisfacen el teorema de Pitágoras (Speziale de Guzmán, 2003). El concepto de ortogonalidad se emplea también para un conjunto de vectores. Un vector x es ortogonal a un conjunto $S = \{y_i\}$ si

$$\langle x, y_i \rangle = 0, \forall i. \quad \dots (2.16)$$

De forma más general dos subespacios S_1, S_2 son ortogonales si todos los vectores de S_1 son ortogonales a todos los vectores de S_2 .

Un conjunto de vectores $\{x_1, \dots, x_k\}$ es llamado ortogonal si $x_i \perp x_j$ cuando $i \neq j$. Si los vectores tienen norma uno, entonces es un sistema ortonormal y satisface:

$$\langle x_i, x_j \rangle = \delta_{ij} N^2 \quad \dots (2.17)$$

Un sistema ortonormal en un espacio vectorial V es una base ortonormal si genera a V .

2.4 *Proceso de ortogonalización de Gram – Schmidt.*

A partir de un conjunto $G = \{x_1, x_2, \dots, x_k\}$ que genera a un espacio V , el procedimiento construye un conjunto $G' = \{y_1, y_2, \dots, y_k\}$ que es ortogonal y también genera el espacio V . La idea consiste en formar uno a uno los vectores y_i de tal manera que cada nuevo vector sea ortogonal a todos los vectores obtenidos anteriormente. A continuación se realizará el procedimiento de ortogonalización para llegar a una definición axiomática.

El primer vector de G' puede ser cualquier vector de G , puesto que no requiere ser ortogonal a ningún otro, entonces:

$$y_1 = x_1 \quad \dots (2.18)$$

A continuación proponemos un vector y_2 de la forma,

$$y_2 = x_2 - \alpha_1 y_1 \quad \dots (2.19)$$

Buscando que y_1 y y_2 sean ortogonales, es decir:

$$\langle y_2, y_1 \rangle = 0 \quad \dots (2.20)$$

Sustituyendo (2.19) en (2.20)

$$\langle x_2 - \alpha_1 y_1, y_1 \rangle = 0 \quad \dots (2.21)$$

$$\langle x_2, y_1 \rangle - \alpha_1 \langle y_1, y_1 \rangle = 0 \quad \dots (2.22)$$

$$\alpha_1 = \langle x_2, y_1 \rangle / \langle y_1, y_1 \rangle \quad \dots (2.23)$$

En consecuencia:

$$y_2 = x_2 - y_1 \langle x_2, y_1 \rangle / \langle y_1, y_1 \rangle \quad \dots (2.24)$$

satisface las condiciones requeridas. De forma gráfica se puede apreciar este procedimiento en la Figura 2-1.

El vector $y_1 \langle x_2, y_1 \rangle / \langle y_1, y_1 \rangle$ recibe el nombre de proyección de x_2 sobre y_1 , de esta manera el vector y_2 se obtiene restando a x_2 su proyección sobre y_1 .

Para obtener un tercer vector y_3 proponemos:

$$y_3 = x_3 - \alpha_1 y_1 - \alpha_2 y_2 \quad \dots (2.25)$$

Buscando que y_3 sea ortogonal a y_1 y a y_2 , el procedimiento es el mismo al que utilizamos para obtener y_2 . Por lo que se debe cumplir (18) y (19).

$$\langle y_3, y_1 \rangle = 0 \quad \dots (2.26)$$

$$\langle y_3, y_2 \rangle = 0 \quad \dots (2.27)$$

Sustituyendo (2.25) en (2.26)

$$\langle x_3 - \alpha_1 y_1 - \alpha_2 y_2, y_1 \rangle = 0 \quad \dots (2.28)$$

$$\langle x_3, y_1 \rangle - \alpha_1 \langle y_1, y_1 \rangle - \alpha_2 \langle y_2, y_1 \rangle = 0 \quad \dots (2.29)$$

Pero como $\langle y_2, y_1 \rangle = 0$ entonces:

$$\alpha_1 = \langle x_3, y_1 \rangle / \langle y_1, y_1 \rangle \quad \dots (2.30)$$

De la misma forma

$$\alpha_2 = \langle x_3, y_2 \rangle / \langle y_2, y_2 \rangle \quad \dots (2.31)$$

En consecuencia:

$$y_3 = x_3 - y_1 \langle x_3, y_1 \rangle / \langle y_1, y_1 \rangle - y_2 \langle x_3, y_2 \rangle / \langle y_2, y_2 \rangle \quad \dots (2.32)$$

Es decir, y_3 se obtiene restando a x_3 sus proyecciones sobre y_1 y y_2 . Generalizando los resultados llegamos al siguiente teorema.

Teorema 2.1

Sea V un espacio con producto interno y sea $G = \{x_1, \dots, x_k\}$ una base de V . El conjunto $G' = \{y_1, \dots, y_k\}$, donde:

$$y_l = x_l \quad \dots (2.33)$$

$$y_i = x_i - \sum y_j \langle x_i, y_j \rangle / \langle y_j, y_j \rangle, \text{ para } i = 1, \dots, k \quad \dots (2.34)$$

es una base ortogonal de V .

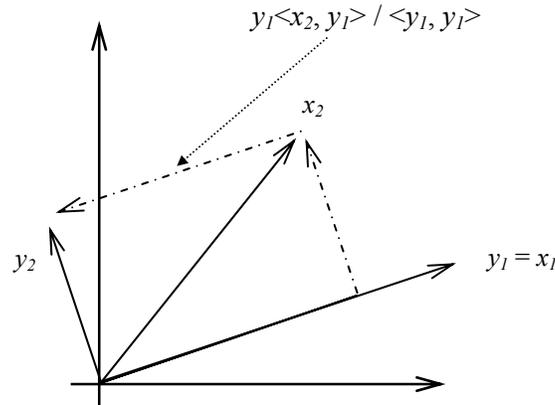


Figura 2-1 Ortogonalización de Gram–Schmidt para dos vectores, el vector y_2 se obtiene de restar a x_2 su proyección sobre x_1 .

2.5 Espacios de Hilbert.

La teoría de los espacios de Hilbert fue iniciada por David Hilbert (1862-1943) en su trabajo sobre formas cuadráticas que aplicó a la teoría de las ecuaciones diferenciales. Durante el período de 1904 a 1910 Hilbert publicó una serie de seis artículos que agrupó en su libro *Grundzüge einer allgemeinen theorie der linearen integralgleichungen* (Fundamentos de una teoría de ecuaciones integrales lineales). Dicho libro publicado en 1912 contiene las ideas generales de los espacios de Hilbert. Aunque él jamás utilizó el concepto de “espacio de Hilbert”, fueron otros matemáticos, particularmente Erhardt Schmidt y Frigyes Riesz, quienes acuñaron el término.

Algunos años después John von Neumann (1903-1957) fue el primero en formular una aproximación axiomática de los espacios de Hilbert y desarrolló la teoría moderna de los operadores sobre dicho espacio. La contribución más relevante fue proporcionar los fundamentos matemáticos en los que se basa la mecánica cuántica. Antes de dar una definición formal de un espacio de Hilbert, se definirá un concepto más: el espacio vectorial completo.

Definición 2.9

Una secuencia de vectores $\{x_n\}$, en un espacio V con norma, es llamada secuencia de Cauchy si para todo $\varepsilon > 0$ existe un número N , tal que para todos $n, m > N$ se cumple que $\|x_m - x_n\| < \varepsilon$.

Definición 2.10

Un espacio vectorial V es completo si cualquier secuencia de Cauchy de elementos pertenecientes al espacio V converge a un elemento en el espacio (Debnath, 1990).

Definición 2.11

Un espacio de Hilbert es un espacio vectorial V completo con producto interno.

Definición 2.12

Un espacio de Hilbert es llamado separable si contiene secuencias ortonormales completas. Los espacios de Hilbert de dimensión finita son considerados separables.

Definición 2.13

Una función f pertenece a $L^2(\mathbb{R})$ o espacio de funciones de cuadrado integrable si se cumple respecto a una medida μ de Lebesgue que:

$$\int_{\mathbb{R}} \|f\|^2 d\mu < \infty \quad \dots (2.35)$$

El cuadrado de la norma se puede interpretar físicamente como la energía asociada a la señal y por tanto, podemos decir que (2.35) significa que la señal tiene energía finita.

Definición 2.14

El espacio de funciones $L^2(\mathbb{R})$ es un espacio de Hilbert cuando se le asocia un producto interno de la forma:

$$\langle f, g \rangle = \int_{-\infty}^{\infty} f(x)g^*(x)dx \quad \dots (2.36)$$

donde $*$ es el complejo conjugado.

La teoría de los espacios de Hilbert juega un papel muy importante en el desarrollo del análisis de la transformada Wavelet; una de las características de las wavelets ortonormales es que forman un espacio $L^2(\mathbb{R})$. Por lo tanto, las wavelets ortonormales pueden ser definidas en términos del álgebra lineal.

2.6 *Perspectiva histórica de la teoría wavelet.*

A través de la historia, las matemáticas y la ingeniería han estado íntimamente relacionadas. En infinidad de casos ha sucedido que un matemático tiene una buena idea, desarrolla una teoría concreta, enfrenta gran oposición de prominentes figuras en el área, pero no obstante continúa su trabajo. Tiempo después aparece un ingeniero y un físico, reformulan y modifican dicha teoría para hacerla más accesible y eventualmente la idea se convierte en una herramienta estándar para investigación en muchos campos. El párrafo anterior describe la historia del desarrollo de la teoría wavelet.

En 1807, Joseph Fourier pronunció que todas las funciones periódicas pueden ser expresadas como una suma pesada de funciones trigonométricas (Bracewell, 1986). La idea fue criticada por Laplace debido a la falta de rigor matemático y Fourier tuvo que esperar más de quince años para convencerlo. Sin embargo, las ideas de Fourier jugaron un papel muy importante en la evolución de las matemáticas, incluso cuando dichas ideas fueron enunciadas, los conceptos de “función” e “integral” aun no recibían una definición precisa.

Una señal puede ser representada en términos de sus componentes de frecuencia mediante la transformada de Fourier; en este caso, las funciones utilizadas en la representación son armónicas puras localizadas en la frecuencia pero no en el tiempo, debido a que se extienden infinitamente en el tiempo, Figura 2-2.

La representación en frecuencia revela componentes frecuenciales de una señal pero oculta su comportamiento temporal. Ambas representaciones son matemáticamente equivalentes en el sentido de que toda la información está presente en cada representación.

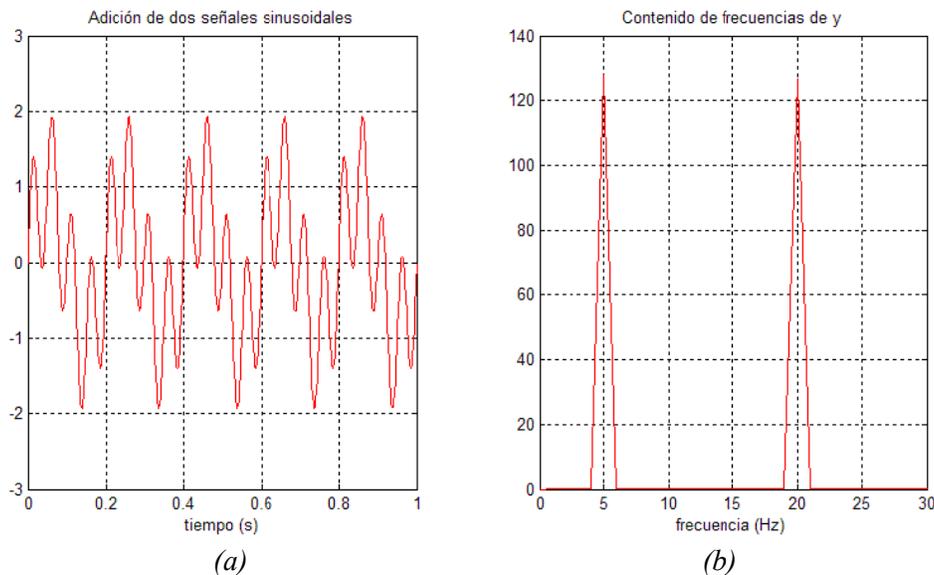


Figura 2-2 Transformada de Fourier de $y = \sin(10\pi t) + \sin(40\pi t)$. (a) Señal sinusoidal en el dominio temporal. (b) Espectro de potencia de y .

La representación de una señal en el dominio temporal no proporciona ninguna información cuantitativa sobre el contenido espectral de la señal. En otras palabras, la representación de Fourier proporciona un contenido espectral sin ninguna información sobre la localización en el tiempo de dichos componentes espectrales.

A principios de siglo XX, en el año de 1909, el húngaro Alfred Haar presentó la tesis doctoral *Zur Theorie der orthogonalen Funktionensysteme Erste Mitteilung* (Primer reporte sobre la teoría de sistemas de funciones ortogonales), bajo la supervisión de David Hilbert. El propósito del trabajo fue enfrentar problemas que no eran resueltos por la transformada de Fourier. La idea fundamental era representar una señal f como un límite de aproximaciones sucesivas correspondientes a diferentes niveles de resolución.

La contribución de Haar a la teoría wavelet es evidente. Creó una familia de wavelets que lleva su nombre. La wavelet Haar es la más simple y el concepto de familia es simple de entender; el padre es la función escala, el punto de partida y mediante escalamientos y translaciones obtenemos a la madre, hijos, hijas, etc. El trabajo de Haar fue continuado por Paul Levy, quien dedicó su investigación al movimiento Browniano.

En 1946, Denis Gabor realizó modificaciones a la transformada de Fourier. Segmentó a la señal utilizando una ventana y realizó el análisis para cada segmento, obteniendo localización de la señal en el dominio espectral. Con estas modificaciones, la transformada de Fourier en Tiempo Corto (STFT) se utilizó para representaciones tiempo-frecuencia (2.37).

$$STFT\{x(t)\} = \int_{-\infty}^{\infty} x(t)w(t-\tau)e^{-j\omega\tau} dt \quad \dots (2.37)$$

donde $w(t)$ es la función ventana, comúnmente una Gaussiana con media igual a cero.

En 1947 Jean Ville retomó el trabajo de Eugene Wigner sobre la ecuación de Schrödinger y propuso una forma alternativa de representación de señales. El método de Ville toma la energía de una señal y la presenta en un plano tiempo-frecuencia. El trabajo de Ville es conocido como la Distribución Wigner-Ville (Ville, 1948) y es ampliamente aplicado en química cuántica, óptica e ingeniería eléctrica. En (2.38) se muestra la distribución de Wigner-Ville.

$$WV\{x(t)\} = \int_{-\infty}^{\infty} x\left(t + \frac{\tau}{2}\right)x^*\left(t - \frac{\tau}{2}\right)e^{-i\omega\tau} d\tau \quad \dots (2.38)$$

Jean Morlet, un ingeniero geofísico graduado de la Escuela Politécnica de Francia, también se enfrentó con el problema de analizar señales cuando estudiaba el eco de los impulsos enviados bajo tierra para hallar petróleo. Cuando trabajó con la STFT descubrió que mantener fijo el tamaño de la ventana era una aproximación errónea. Trató las altas frecuencias de las señales con ventanas de tamaño angosto y las bajas frecuencias con ventanas de mayor tiempo de duración, además todas estas ventanas fueron generadas mediante dilataciones y compresiones de un prototipo de ventana

Gaussiana. El resultado fueron ventanas con soporte compacto en el tiempo y frecuencia, ya que la transformada de Fourier de una función Gaussiana es otra función Gaussiana. Morlet las llamó “wavelets de pendiente constante”.

En 1980, Morlet se unió a Alex Grossmann y orientaron sus esfuerzos a darle un rigor matemático a las wavelets y concibieron la transformación inversa (Grossman, et al., 1984).

En 1985, Yves Meyer, un matemático francés, continuó el trabajo de Morlet y desarrolló wavelets con mejores propiedades de localización en el tiempo y la frecuencia (Meyer, 1990). Al mismo tiempo, Ingrid Daubechies, bajo la tutela de Grossmann presentó su trabajo doctoral *Representation of Quantum Mechanical Operators by Kernels on Hilbert Spaces of Analytic Functions*. En dicho trabajo realiza una discretización de los parámetros en tiempo y frecuencia de la transformada Wavelet, permitiendo mayor libertad al construir bases ortogonales pero sacrificando la redundancia. Daubechies creó su propia familia de wavelets con soporte compacto, cumpliendo con la propiedad de ortogonalidad (Daubechies, 1992).

En 1986, Mallat desarrolló el análisis multirresolución (MRA) para la transformada Wavelet Discreta, del que líneas arriba se ha hablado brevemente. Propuso un algoritmo rápido para el cálculo de los coeficientes de las wavelets, un esquema muy similar que ya había sido desarrollado en los años 70's para reducir el ruido de cuantización usando filtros espejo en cuadratura (QMF).

En los últimos 10 años ha existido una búsqueda por encontrar otras bases con diferentes propiedades y modificaciones a los algoritmos de MRA. En 1992 Albert Cohen y Daubechies construyeron una base biortogonal de wavelets con soporte compacto (Cohen, et al., 1992) y Coifman desarrolló *wavelet packets*, una extensión de MRA (Coifman, et al., 1990).

2.7 *La transformada Wavelet Continua.*

La transformada Wavelet Continua es una operación lineal que descompone una señal compleja en bloques elementales que aparecen en diferentes escalas o resoluciones. Es una representación multirresolución basada en la diferencia de aproximaciones de información disponibles en dos resoluciones consecutivas 2^j y 2^{j+1} . Esta diferencia de información es llamada *detalles de la señal* en la resolución j y es igual a la proyección ortogonal de la señal sobre V_j y V_{j+1}

La selección de la wavelet depende de las características propias de la señal de estudio. La función base $\psi \in L^2(\mathbb{R})$ debe cumplir la condición de admisibilidad, es decir, que su transformada de Fourier Ψ satisfice:

$$\int_{-\infty}^{\infty} \frac{|\Psi(\omega)|^2}{|\omega|} d\omega < \infty \quad \dots (2.39)$$

(2.39) implica que la función ψ tiene media cero.

$$\int_{-\infty}^{\infty} \psi(t) dt = 0 \quad \dots (2.40)$$

Por simplicidad se asume que ψ está normalizada $\|\psi\| = 1$ y centrada en $t = 0$, ψ se conoce como *wavelet madre*.

Mediante traslaciones y dilataciones de ψ se construye la familia de wavelets:

$$\psi_{a,b}(t) = |a|^{-\frac{1}{2}} \psi\left(\frac{t-b}{a}\right) \quad a, b \in \mathbb{R}, a \neq 0 \quad \dots (2.41)$$

el factor $|a|^{-\frac{1}{2}}$ es de normalización y asegura que todas las funciones escaladas tengan la misma energía.

Podemos definir entonces la transformada Wavelet Continua (CWT) de x en la escala a y en el desplazamiento b como el producto interno de x con $\psi_{a,b}$, utilizando la Definición 2.14 se tiene:

$$CWT_{a,b}\{x(t)\} = \langle x, \psi_{a,b} \rangle = |a|^{-\frac{1}{2}} \int x(t) \psi^*\left(\frac{t-b}{a}\right) dt \quad \dots (2.42)$$

el parámetro b permite analizar partes de la señal x de forma separada. El parámetro de dilatación a determina el ancho de la función ventana ψ : para valores pequeños de a la transformada wavelet aumenta los pequeños detalles de la señal x . El tamaño de los detalles está relacionado con el rango específico de frecuencias, por esta razón a es llamado parámetro escala. La transformada Wavelet es una función continua, en el dominio del tiempo, la wavelet $\psi_{a,b}$ está centrada en b con una extensión proporcional a a . Mientras la escala aumenta la wavelet es más ancha en el tiempo y más angosta en el dominio frecuencial (la resolución temporal disminuye mientras que la resolución frecuencial aumenta).

Algunas de las aplicaciones donde es útil la transformada Wavelet son:

- Variando el parámetro de escala se puede hacer análisis de la señal en diferentes bandas de frecuencia, esto es particularmente útil para propósitos de compresión de datos.
- Mediante una adecuada wavelet, se pueden aproximar derivadas de la señal en direcciones específicas, esto permite detectar bordes y estimar la orientación de las imágenes.
- Adaptando la wavelet a un patrón deseado, la transformada Wavelet puede ser un método de reconocimiento de patrones realizando filtrado o correlación cruzada simultáneamente en todas las bandas de frecuencia.

2.7.1 Ejemplos de wavelets continuas.

- *Wavelet Morlet.*

La función gaussiana modulada es una de las más utilizadas en el análisis de señales, desarrollada por R. Kronland-Martinet y J. Morlet (Holschneider, et al., 1989).

$$\psi(t) = \pi^{-\frac{1}{4}} \left(e^{i\varepsilon_0 t} - e^{-\frac{\varepsilon_0^2}{2}} e^{-\frac{\beta^2 t^2}{2}} \right) \quad \dots (2.43)$$

Sin embargo, la wavelet Morlet no satisface la condición de admisibilidad (2.39), así que se tienen que variar los parámetros ε_0 y β para hacer a ψ aproximadamente admisible, donde ε_0 se calcula como la proporción de los dos valores máximos de la wavelet madre ψ . Además se debe cumplir que $\varepsilon_0 \geq 5\beta$. En la práctica se toma el valor de $\varepsilon_0 = 5$. Para este valor el segundo término dentro del paréntesis se desprecia por lo que podemos reescribir la wavelet Morlet de la forma:

$$\psi(t) = \pi^{-\frac{1}{4}} e^{i5t} e^{-\frac{t^2}{2}} \quad \dots (2.44)$$

ε_0 es el parámetro del ancho de banda y β es la frecuencia central. El nombre completo de la wavelet es entonces Morlet Compleja β - ε_0 . Figura 2-3.

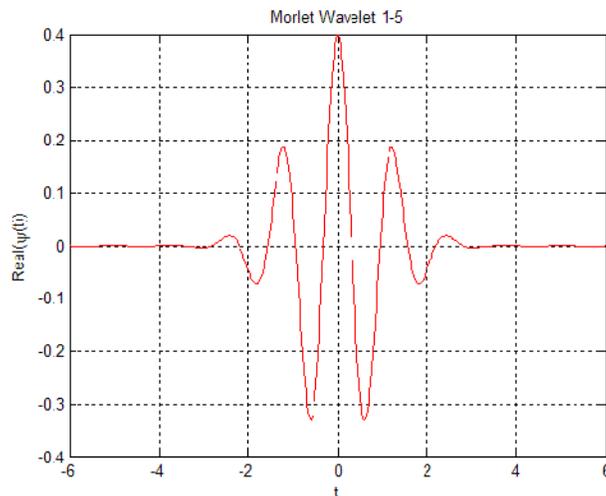


Figura 2-3 Parte real de la wavelet compleja Morlet 1-5.

- *Wavelet Mexican Hat.*

La segunda derivada de la gaussiana $e^{-\frac{x^2}{2}}$, es una función muy utilizada en visión computacional para detectar bordes multiescala (Witkin, 1983) y se conoce popularmente con el nombre de *Mexican hat*. Esta función está íntimamente relacionada con los procesos que ocurren en las primeras etapas de visión del ser humano.

$$\psi(t) = \frac{2}{\sqrt{3}\sigma} \pi^{-\frac{1}{4}} \left(1 - \frac{t^2}{\sigma^2}\right) e^{-\frac{t^2}{2\sigma^2}} \quad \dots (2.45)$$

La wavelet *Mexican hat* también se conoce en geofísica como wavelet *Ricker* y se emplea para modelar datos sísmicos. La generalización a mayores dimensiones se conocen como laplacianas de gaussiana. En la practica, esta wavelet es aproximada por la diferencia de funciones gaussianas. En la Figura 2-4 se muestra la forma de la wavelet *Mexican hat* y en la Figura 2-5 se realiza la transformación de una señal utilizando *Mexican hat*.

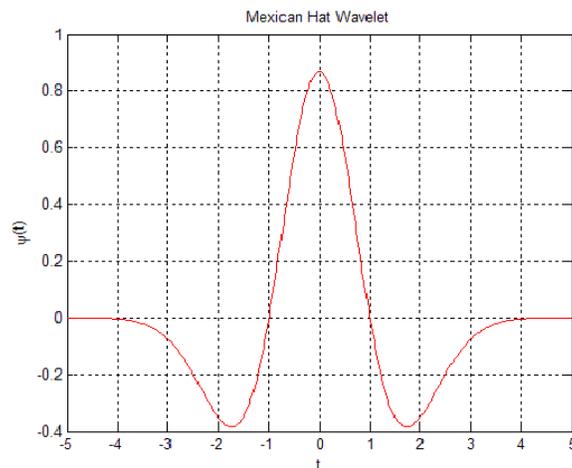


Figura 2-4 Wavelet *Mexican hat* con $\sigma = 1$.

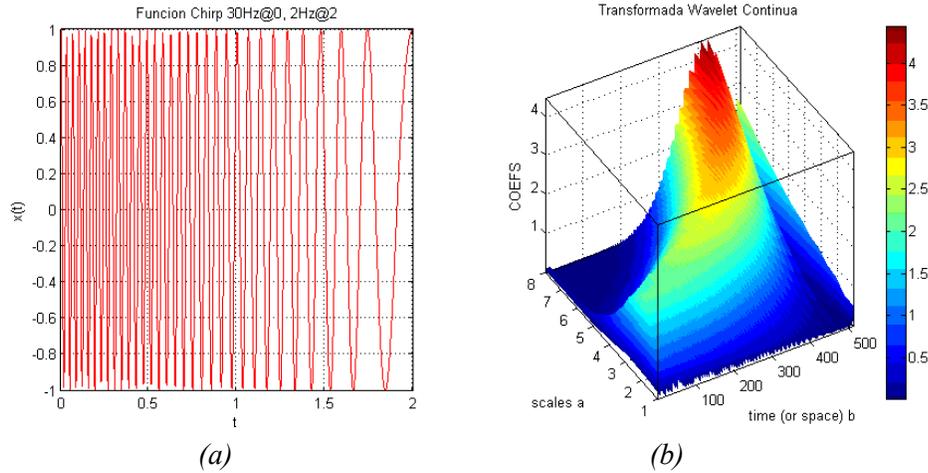


Figura 2-5 (a) Función $x(t) = chirp$ con frecuencia 30 Hz en el tiempo $t = 0$ y descendiendo a 2Hz en el tiempo $t = 2$. (b) Vista 3D de la transformada Wavelet Continua de $x(t)$ utilizando *Mexican hat*.

2.8 Wavelets y análisis multirresolución de Mallat.

Para llevar a cabo la fusión de imágenes, se necesita una versión discreta de la transformada Wavelet y un algoritmo que permita un cálculo rápido y estable. Además, la versión discreta de la transformada Wavelet debe heredar las propiedades de su versión continua. La transformada Wavelet Continua es redundante en el sentido que mapea una señal unidimensional en una función de dos dimensiones (Piella, et al., 2001). Esta redundancia puede ser reducida o incluso eliminada mediante submuestreo de los parámetros de escala y traslación. Por conveniencia restringimos los valores de $a = a_0^m$, donde $m \in \mathbb{Z}$ con $a_0 \neq 1$, $b = nb_0 a_0^m$ con $n, m \in \mathbb{Z}$ y $a_0 > 1$, $b_0 > 0$. Con esta selección de parámetros discretos aseguramos que la wavelet en el nivel m se comporte como su versión continua. Sustituyendo los parámetros discretos en (2.41) se tiene:

$$\psi_{m,n}(t) = a_0^{-\frac{m}{2}} \psi(a_0^{-m} t - nb_0) \quad \dots (2.46)$$

es la familia discretizada de wavelets.

Daubechies (Daubechies, 1992) demostró que la reconstrucción de $x \in L^2(\mathbb{R})$ a partir de los coeficientes wavelet $\langle x, \psi_{m,n} \rangle$, $m, n \in \mathbb{Z}$ es posible si y sólo si:

$$A \|x\|^2 \leq \sum_{m=-\infty}^{\infty} \sum_{n=-\infty}^{\infty} \|\langle x, \psi_{m,n} \rangle\|^2 \leq B \|x\|^2 \quad \dots (2.47)$$

Para ciertas constantes $0 < A \leq B < \infty$. Si la condición (2.47) se satisface, entonces la familia discreta de wavelets $\psi_{m,n}$ constituye un *frame* (Duffin, et al., 1952).

Definición 2.15

Si ψ_j con $j \in \mathbb{Z}$ es un *frame* en el espacio de Hilbert, entonces el operador *frame* F es un operador lineal definido como:

$$(Fx)_j = \langle x, \psi_j \rangle \quad \dots (2.48)$$

Aplicando el operador $(F^*F)^{-1}$ a $\psi_{m,n}$ donde F^* es el transpuesto de F , podemos obtener un *frame dual* $\tilde{\psi}_{m,n}$ dado por (2.49)

$$\tilde{\psi}_{m,n} = (F^*F)^{-1}\psi_{m,n} \quad \dots (2.49)$$

También podemos obtener $\tilde{\psi}_{m,n}$ calculando cada $\tilde{\psi}_{m,0}$ y trasladándola de acuerdo a:

$$\tilde{\psi}_{m,n}(t) = a_0^{-\frac{m}{2}} \tilde{\psi}(a_0^{-m}t - nb_0) \quad \dots (2.50)$$

La ecuación (2.46) y (2.50) satisfacen la condición de biortogonalidad.

$$\langle \psi_{m,n}, \tilde{\psi}_{l,k} \rangle = \delta_{m,l} \delta_{n,k} \quad m, n, l, k \in \mathbb{Z} \quad \dots (2.51)$$

Las wavelets que satisfacen (2.51) se llaman wavelets biortogonales y la reconstrucción de la señal está dada por:

$$x(t) = \sum_{m,n \in \mathbb{Z}} \langle x, \psi_{m,n} \rangle \tilde{\psi}_{m,n} = \sum_{m,n \in \mathbb{Z}} \langle x, \tilde{\psi}_{m,n} \rangle \psi_{m,n} \quad \dots (2.52)$$

Un caso especial son las wavelets ortogonales las cuales se obtienen cuando $\psi_{m,n} = \tilde{\psi}_{m,n}$.

2.8.1 Wavelets ortogonales.

En general, la expansión en (2.52) sigue siendo una representación sobrecompleta de x . De hecho, la serie de wavelets $\{\langle x, \psi_{m,n} \rangle\}_{m,n \in \mathbb{Z}}$ es simplemente una versión muestreada de la transformada Wavelet Continua. Eliminar la redundancia es equivalente a construir bases de $L^2(\mathbb{R})$, siendo el marco para la construcción de bases el análisis multirresolución (Mallat, 1989).

La aproximación de una función x en la resolución 2^j está definida por una malla discreta de muestras que proporcionan promedios locales de x sobre una vecindad de tamaño proporcional a 2^j . De una manera más formal, la aproximación multirresolución de Mallat de una función x en la resolución 2^j es definida como una proyección ortogonal en un espacio $V_j \subset L^2(\mathbb{R})$. El espacio V_j reagrupa todas las posibles aproximaciones en la resolución 2^j . La proyección ortogonal de x es una función $f_j \in V_j$ que minimiza $\|f - f_j\|$.

Existen muchos espacios que cumplen con las condiciones (2.1), (2.2) y (2.3), (páginas 10 y 11). Sin embargo no todos están relacionados con el análisis multirresolución, en consecuencia debemos agregar tres requerimientos más:

- Debe existir una función ϕ tal que $\{\phi(t - n)\}_{n \in \mathbb{Z}}$ es una base ortonormal de V_0 donde para todo $j, n \in \mathbb{Z}$ se tiene:

$$\phi_{j,n}(t) = 2^{-\frac{j}{2}} \phi(2^{-j}t - n) \quad \dots (2.53)$$

- Invariancia a la traslación: $f(t) \in V_j \Leftrightarrow f(t - 2^j n) \in V_j$ donde $j, n \in \mathbb{Z}$.
- $f(t) \in V_j \Leftrightarrow f\left(\frac{t}{2}\right) \in V_{j+1}$ donde $j \in \mathbb{Z}$.

La última propiedad asegura que todos los espacios son versiones escaladas del espacio central V_0 , esto es que la aproximación en la resolución 2^j contiene toda la información necesaria para calcular la aproximación en la siguiente resolución 2^{j-1} .

Ya que la función $\phi \in V_0 \subset V_{-1}$ y $\phi_{-1,n}$ son bases ortonormales en V_{-1} se tiene que:

$$\phi(t) = \sum_n h_n \phi_{-1,n} \quad \dots (2.54)$$

La secuencia $\{h(n)\}_{n \in \mathbb{Z}} \in L^2(\mathbb{R})$ determina la función escala ϕ y el correspondiente análisis multirresolución. Además, ϕ tendrá soporte compacto solo si el soporte de h es finito.

$$h_n = \langle \phi, \phi_{-1,n} \rangle \quad \text{y} \quad \sum_{n \in \mathbb{Z}} |h_n|^2 = 1 \quad \dots (2.55)$$

Sustituyendo (2.53) en (2.54) obtenemos la *función escala*:

$$\phi(t) = \sqrt{2} \sum_n h_n \phi(2t - n) \quad \dots (2.56)$$

El principio básico del análisis multirresolución es que dada una función escala ϕ existe una función wavelet ortonormal ψ tal que: para toda $x \in L^2(\mathbb{R})$ se cumple que:

$$P_{j-1}x = P_jx + \sum_{n \in \mathbb{Z}} \langle x, \psi_{j,n} \rangle \psi_{j,n} \quad \dots (2.57)$$

donde P_j es la proyección ortogonal sobre V_j .

Para poder explicar cómo se relacionan la función escala y la función wavelet debemos definir el complemento ortogonal.

Definición 2.16

Para todo $j \in \mathbb{Z}$, definimos al espacio W_j o *espacio de detalles*, donde $W_j \perp W_k$ con $j \neq k$, como el complemento ortogonal de V_j y contiene toda la información necesaria para ir de una aproximación V_j a una aproximación V_{j-1} , esto es:

$$V_{j-1} = V_j \oplus W_j \quad \dots (2.58)$$

De la definición anterior podemos decir que ψ puede ser escrita en términos de ϕ ya que forma una base de V_{j-1} (Mallat, 1989).

$$\psi = \sum_n (-1)^{n-1} h_{-n-1} \phi_{-1,n} \quad \dots (2.59)$$

$$\psi(t) = \sqrt{2} \sum_n (-1)^{n-1} h_{-n-1} \phi(2t - n) \quad \dots (2.60)$$

Reescribiendo la ecuación (2.60) tendremos la *función wavelet* definida en términos de la secuencia $\{g(n)\}_{n \in \mathbb{Z}} \in L^2(\mathbb{R})$.

$$\psi = \sqrt{2} \sum_n g_n \phi(2t - n) \quad \text{con} \quad g_n = (-1)^n h_{-n+1} \quad \dots (2.61)$$

Las ecuaciones (2.56) y (2.60) son referidas en la literatura como *ecuaciones de refinamiento*. Como veremos más adelante, h corresponde a la respuesta impulso de un filtro paso-bajas, mientras que g corresponde a la respuesta impulso de un filtro paso-altas. Las ecuaciones de refinamiento también existen para las funciones duales $\tilde{\phi}$ y $\tilde{\psi}$ produciendo los filtros discretos duales \tilde{h} y \tilde{g} respectivamente.

2.8.2 Descomposición de la transformada Wavelet Discreta.

El análisis multirresolución conduce de forma natural a un esquema jerárquico para la obtención de los coeficientes wavelet. El proceso puede verse como el cálculo de aproximaciones sucesivas de la señal x , junto con la diferencia de la información entre cada dos niveles sucesivos. Supongamos que tenemos la señal original x , entonces $x^0 = P_0 x$, recordando de (2.57) P_j es la proyección ortogonal sobre V_j ; además denotamos a Q_j como la proyección ortogonal sobre W_j .

De la ecuación (2.58) podemos descomponer a $x^0 \in V_0 = V_1 \oplus W_1$ en:

$$x^0 = x^1 + d^1 \quad \dots (2.62)$$

donde $x^1 = P_1 x^0$ es la siguiente aproximación de x y $d^1 = Q_1 x^0$ son los detalles que se pierden en la transición de $x^0 \rightarrow x^1$. Los detalles también se pueden expresar como: $d^1 = P_0 x - P_1 x^0$. En cada uno de estos espacios V_j, W_j se tienen bases ortogonales ϕ, ψ tal que:

$$x^0 = \sum_n a_n^0 \phi_{0,n}, \quad x^1 = \sum_n a_n^1 \phi_{1,n}, \quad d^1 = \sum_n b_n^1 \psi_{1,n} \quad \dots (2.63)$$

La siguiente aproximación $x^1 \in V_1 = V_2 \oplus W_2$, de nueva cuenta podemos descomponer de la misma forma que en (2.62):

$$x^1 = x^2 + d^2 \quad \dots (2.64)$$

$$x^2 = \sum_n a_n^2 \phi_{2,n}, \quad d^2 = \sum_n b_n^2 \psi_{2,n} \quad \dots (2.65)$$

De las ecuaciones (2.62) y (2.64) podemos presentar una expresión general conocida como *descomposición o análisis* de la transformada Wavelet Discreta:

$$x^{j-1} = x^j + d^j = \sum_n a_n^j \phi_{j,n} + \sum_n b_n^j \psi_{j,n} \quad \dots (2.66)$$

con los coeficientes de escala y wavelet definidos respectivamente como:

$$a_k^j = \sum_n \bar{h}_{n-2k} a_n^{j-1}, \quad b_k^j = \sum_n \bar{g}_{n-2k} a_n^{j-1} \quad \dots (2.67)$$

donde

$$a_n^0 = \langle x, \phi_{0,n} \rangle = \int x(t) \phi(t-n) dt \quad \dots (2.68)$$

La ecuación (2.68) representa un problema cuando no existe una representación analítica de $x(t)$ o es una función con valores discretos. En este caso, se acostumbra utilizar los valores de $x(t)$ como coeficientes de a_n^0 y continuar el análisis utilizando (2.67). Esta es una operación segura ya que el esquema piramidal, Figura 2-6, produce una reconstrucción perfecta y los coeficientes a_n^0 constituyen un promedio local de la señal proporcionada por la función escala.

En general, podemos escoger:

$$a_k^0 = \sum_n c_n x(k-n) \quad \dots (2.69)$$

Para que se cumpla la suposición anterior se debe considerar a $c_n = \delta_{0,n}$.

El esquema de descomposición wavelet se esquematiza en la Figura 2-6. En la Figura 2-7 se muestra un ejemplo de descomposición en cuatro niveles para la señal $x(t)$.

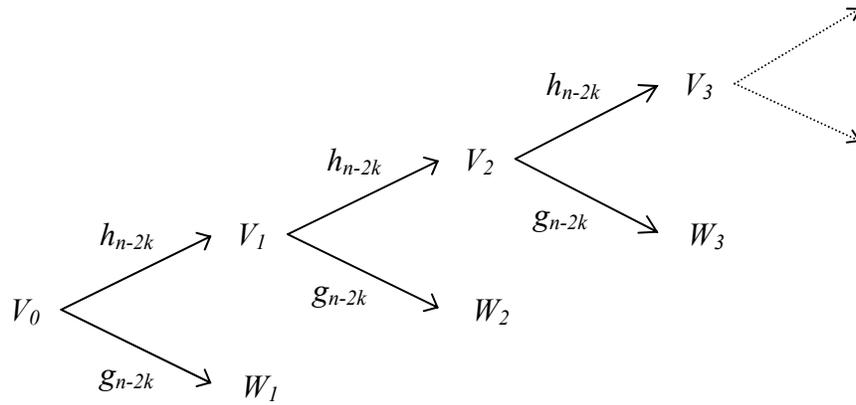


Figura 2-6 Representación esquemática de la ecuación (2.66).

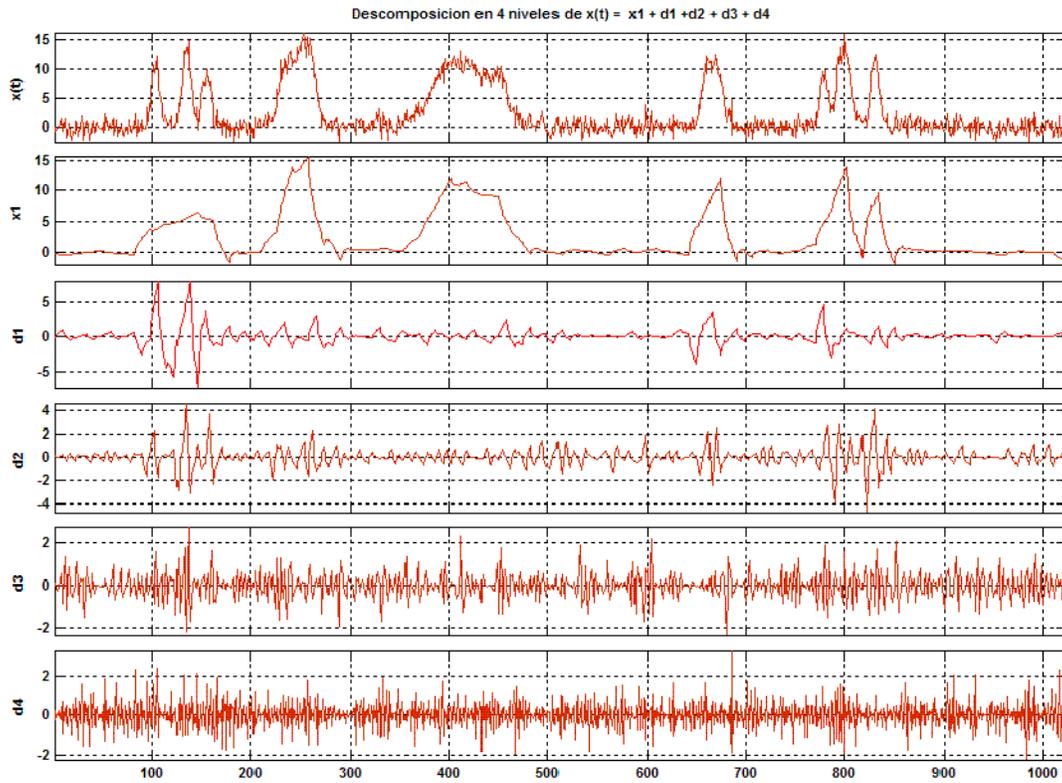


Figura 2-7 DWT de la señal $x(t)$ aleatoria ruidosa utilizando la wavelet Daubechies Db2.

2.9 Wavelets y bancos de filtros.

Los bancos de filtros juegan un papel muy importante en el desarrollo de la teoría wavelet (Vaidyanathan, 1987). Estos bancos son arreglos de filtros paso-bajas, paso-altas y paso-banda usados para el análisis y síntesis de las señales. La razón de su popularidad es el hecho que permiten extraer fácilmente las componentes espectrales de la señal, separándola en M canales llamados subbandas. En particular, los filtros de espejo en cuadratura (QMF) son una representación óptima para la codificación de subbandas. (Kirac, et al., 1998).

En la Figura 2-8 se muestra un banco de filtros de M canales. La señal de entrada es descompuesta en M subbandas aplicando M filtros de análisis. De esta forma, cada subbanda lleva información de la señal de entrada en una banda de frecuencia específica. A la salida de cada filtro de análisis existen bloques con flechas apuntando hacia abajo que indican submuestreo por un factor N , esta operación sirve para eliminar redundancias en las M subbandas de la señal. Los bloques con las flechas apuntando hacia arriba indican sobremuestreo por un factor N , esto permite recuperar la tasa de muestreo original. A la salida de los bloques de sobremuestreo se encuentran los filtros de síntesis que reemplazan los ceros por valores significativos. En este caso, el número de canales es igual a la tasa de muestreo $N = M$, hablamos de un *submuestreo crítico* ya que este es el máximo factor que permite la reconstrucción perfecta. La reconstrucción perfecta significa que la señal de salida es una copia de la señal de entrada sin más distorsiones que un desplazamiento en el tiempo y la variación de la amplitud.

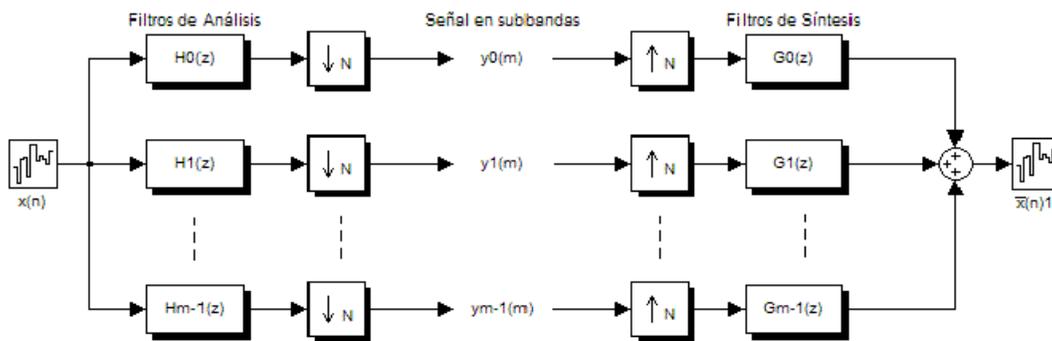


Figura 2-8 Banco de filtros de M canales.

2.9.1 Submuestreo y sobremuestreo.

Se pueden distinguir dos operaciones básicas en los bancos de filtros, la primera de ellas es el submuestreo también llamada compresor. El submuestreo es el proceso de reducir la tasa de muestreo de una señal por un factor entero N , tomando sólo las N -ésimas muestras de la señal,

Figura 2-9a. Un compresor toma una secuencia de entrada $x(n)$ y produce una secuencia de salida $y(n)$ de la forma:

$$y(n) = x(Nn) \quad \dots (2.70)$$

La constante N es el factor de submuestreo. Un aspecto importante con el submuestreo es que puede producir *aliasing*. Esto es, que hay un traslape de las frecuencias de entrada, si esto ocurre, no es posible recuperar la señal original. El *aliasing* puede ser evitado si $x(n)$ es una señal de banda base, limitada a $|\omega| < \frac{\pi}{M}$. Esto es equivalente a decir que se cumple con el criterio de Nyquist.

El sobremuestreo, a veces llamado interpolación, es el proceso de incrementar la tasa de muestreo de la señal por un factor entero N , Figura 2-9b. El sobremuestreo toma una secuencia de entrada $x(n)$ y produce una secuencia de salida $y(n)$

$$y(n) = \begin{cases} x\left(\frac{n}{N}\right) & \text{si } n \text{ es múltiplo de } N \\ 0 & \text{otro caso} \end{cases} \quad \dots (2.71)$$

Donde N es un entero. La constante N es el factor de sobremuestreo. En términos simples, el resultado es la inserción de $N-1$ ceros entre las muestras de la señal original. Ya que la forma del espectro no se altera con el sobremuestreo, no hay pérdida de información y la señal original puede ser recuperada.

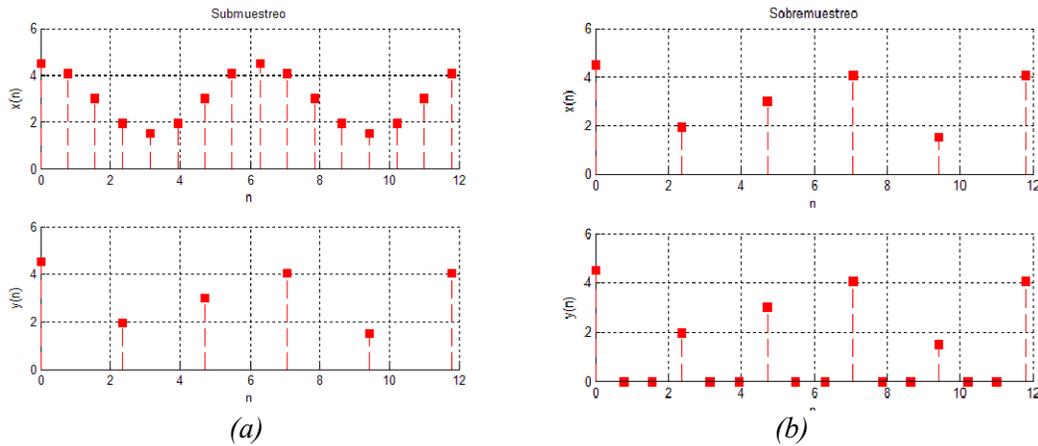


Figura 2-9 (a) Submuestreo de la señal $x(n)$ por un factor $N = 3$. (b) Sobremuestreo de $x(n)$ por un factor $N = 3$.

2.9.2 Implementación de la transformada Wavelet usando filtros espejo en cuadratura.

Existen muchas configuraciones de bancos de filtros, pero la que más nos interesa es la propuesta por Smith y Barnwell (Smith, et al., 1984). Ambos trabajaron en el desarrollo de bancos de filtros ortogonales de dos canales utilizando filtros espejo en cuadratura para obtener una reconstrucción perfecta.

Definición 2.17

Sea $h_0(n)$ un filtro FIR paso-bajas con coeficientes reales. El filtro espejo se define como:

$$h_1(n) = (-1)^n h_0(n) \quad \dots (2.72)$$

O su equivalente en el dominio de la transformada Z

$$H_1(z) = H_0(-z) \quad \dots (2.73)$$

Y su filtro en cuadratura se define como:

$$H_2(z) = z^{-(N-1)} H_0(-z^{-1}) \quad \dots (2.74)$$

Además se debe cumplir que la magnitud sea una función tal que:

$$\left| H_1 \left(e^{j \left(\frac{\pi}{2} - \omega \right)} \right) \right| = \left| H_0 \left(e^{j \left(\frac{\pi}{2} + \omega \right)} \right) \right| \quad \dots (2.75)$$

La ecuación (2.75) nos lleva a que H_0 y H_1 deben ser de potencia complementaria

$$\left| H_0(e^{j\omega}) \right|^2 + \left| H_1(e^{j\omega}) \right|^2 = 1 \quad \dots (2.76)$$

Los diagrama de polos y ceros de los filtros espejo en cuadratura se reflejan sobre el eje imaginario del plano Z de ahí el nombre de espejo. Los filtros espejo en cuadratura son utilizados para eliminar el *aliasing* en los codificadores subbanda de dos canales (Esteban, et al., 1977). Las propiedades arriba mencionadas se ilustran en la Figura 2.10.

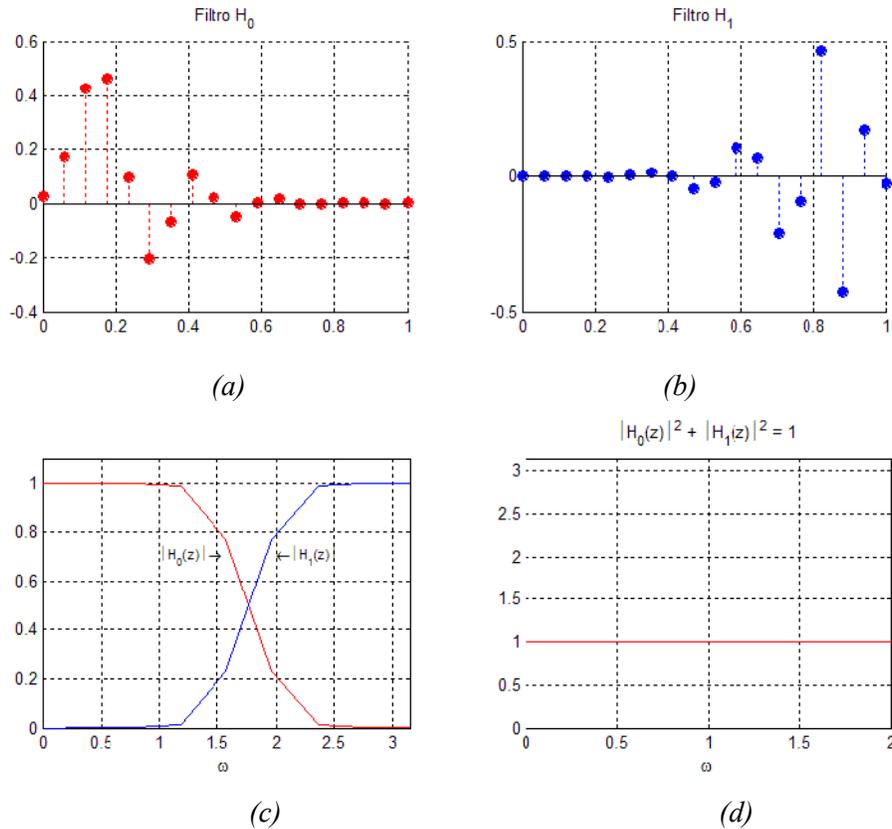


Figura 2-10 Filtros espejo en cuadratura QMF: (a) Filtro H_0 . (b) Filtro H_1 . (c) Respuesta en frecuencia de los QMF. (d) Potencia de los QMF.

El banco de filtros de dos canales que se muestra en la Figura 2-11 divide al espectro de la señal de entrada x en dos bandas iguales en el rango de frecuencias de $0 \leq \omega \leq \pi$ mediante los filtros de análisis $H_0(z)$ y $H_1(z)$. El resultado de la convolución de la señal de entrada con los filtros paso-bajas y paso-altas respectivamente es submuestreado por dos para proporcionar las bandas de las señales $V_1(z)$ y $W_1(z)$ como salida de la etapa de análisis. Esta etapa es de hecho la transformada Wavelet Discreta en una escala, repitiendo el proceso para la señal obtenida $V_1(z)$ se pueden obtener las siguientes escalas.

La etapa de síntesis recibe las bandas de la señal y se sobremuestran por dos para obtener $V_2(z)$ y $W_2(z)$, esto es intercalar ceros entre las muestras de las bandas de frecuencia. Estas señales se

convolucionan con los filtros $G_0(z)$ y $G_1(z)$ para poder sumarlas y producir la señal reconstruida, esta etapa de síntesis es de hecho la inversa de la transformada Wavelet Discreta.

Esta es la única combinación de filtros paso-bajas y paso-altas que permite pasar todas las frecuencias de la señal.

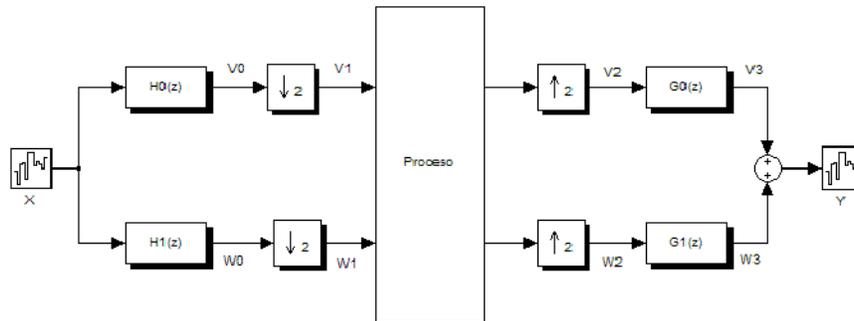


Figura 2-11 Banco de filtros QMF y reconstrucción perfecta.

Consideremos ahora la rama superior del banco de filtros de la Figura 2.11.

$$\begin{aligned} V_0(z) &= H_0(z)X(z) \\ V_3(z) &= G_0(z)V_2(z) \end{aligned} \quad \dots (2.77)$$

Del trabajo de Vaidyanathan (Vaidyanathan, 1987), podemos decir que la operación de submuestrear y sobremuestrear una señal está dada por:

$$(\uparrow 2)(\downarrow 2)V_2(z) = \frac{1}{2}[V_2(z) + V_2(-z)] \quad \dots (2.78)$$

De las ecuaciones (2.78) y (2.77) podemos reescribir para obtener el valor de $V_3(z)$:

$$V_3(z) = \frac{1}{2}G_0(z)[H_0(z)X(z) + H_0(-z)X(-z)] \quad \dots (2.79)$$

De la misma forma podemos obtener $W_3(z)$ de la rama inferior:

$$W_3(z) = \frac{1}{2}G_1(z)[H_1(z)X(z) + H_1(-z)X(-z)] \quad \dots (2.80)$$

Sumando (2.79) y (2.80) para obtener la señal reconstruida tenemos:

$$\begin{aligned} Y(z) &= \frac{1}{2}G_0(z)[H_0(z)X(z) + H_0(-z)X(-z)] + \frac{1}{2}G_1(z)[H_1(z)X(z) + H_1(-z)X(-z)] \\ &= \frac{1}{2}[G_0(z)H_0(z) + G_1(z)H_1(z)]X(z) + \frac{1}{2}[G_0(z)H_0(-z) + G_1(z)H_1(-z)]X(-z) \quad \dots (2.81) \\ &= \frac{1}{2}T(z)X(z) + \frac{1}{2}S(z)X(-z) \end{aligned}$$

Para cumplir con la reconstrucción perfecta se debe cumplir que:

$$T(z) = G_0(z)H_0(z) + G_1(z)H_1(z) = 2z^{-l} \quad \dots (2.82)$$

$$S(z) = G_0(z)H_0(-z) + G_1(z)H_1(-z) = 0 \quad \dots (2.83)$$

(2.82) se conoce como el término de distorsión y (2.83) es el término de *aliasing*.

Esto proporciona varias relaciones entre los cuatro filtros, dejándonos la libertad de escoger a partir de varias consideraciones:

$$\begin{aligned} H_1(z) &= H_0(-z) \\ G_0(z) &= H_1(-z) \\ G_1(z) &= -H_0(-z) \end{aligned} \quad \dots (2.84)$$

(2.84) satisface la condición en (2.83), la salida de la señal no contiene *aliasing*, pero la amplitud puede presentar distorsiones. Ya que los filtros son de potencia complementaria podemos tomar las siguientes ecuaciones que satisfacen (2.82)

$$\begin{aligned} H_1(z) &= z^{-(N-1)}H_0(-z) \\ G_0(z) &= z^{-(N-1)}H_0(z) \\ G_1(z) &= z^{-(N-1)}H_1(z) \end{aligned} \quad \dots (2.85)$$

donde N es el número de coeficientes del filtro.

Con esta relación podemos construir un banco de filtros con reconstrucción perfecta a partir de un filtro paso-bajas. De las ecuaciones en (2.85) se muestra ejemplo para un filtro de cuatro coeficientes.

$$\begin{aligned}
 h_0 &= [a_0 \quad a_1 \quad a_2 \quad a_3] \\
 h_1 &= [a_3 \quad -a_2 \quad a_1 \quad -a_0] \\
 g_0 &= [a_3 \quad a_2 \quad a_1 \quad a_0] \\
 g_1 &= [-a_0 \quad a_1 \quad -a_2 \quad a_3]
 \end{aligned}
 \quad \dots (2.86)$$

2.9.3 Funciones escala y funciones wavelet.

Con el propósito de ilustrar las diferentes wavelet que se implementaron en el *toolbox* para java, FUSIONLAB, se presentan las funciones de las wavelets ortogonales más populares: Haar, Daubechies, Coiflets y Symlets. El desarrollo teórico ampliamente explicado y el método de cálculo se pueden encontrar en (Daubechies, 1992), (Mertins, 1999), (Akansu, et al., 2001) y (Bogges, et al., 1999).

La Figura 2-12(a) muestra la función escala ϕ de la wavelet Haar y la Figura 2-12(b) muestra la función madre ψ de Haar. Las funciones se expresan en segundos [s].

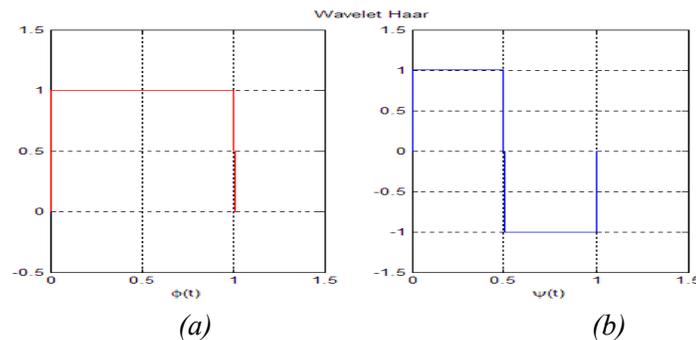
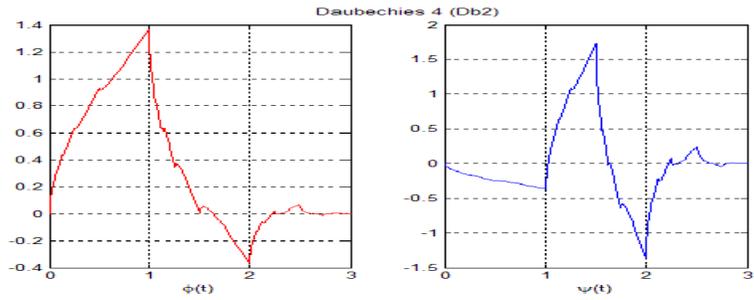
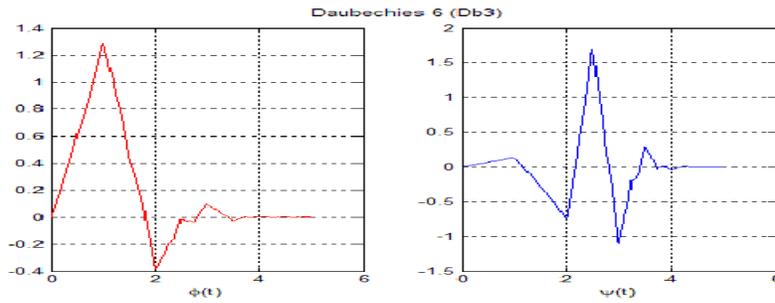


Figura 2-12 Wavelet Haar (a) Función Escala. (b) Wavelet Madre.

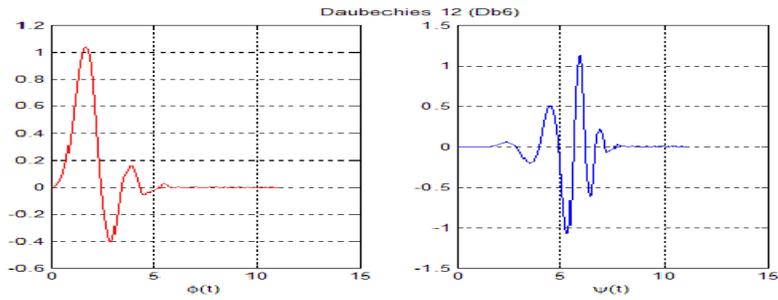
La Figura 2-13 muestra la familia de wavelets Daubechies (Db). La Figura 2-13(a) es Db2, la Figura 2-13(b) es Db3, la Figura 2-13(c) es Db6 y la Figura 2-13(d) muestra la wavelet Db10. Las funciones escala ϕ se encuentran del lado izquierdo y en rojo, mientras que las funciones madre ψ aparecen en el lado derecho y en azul. Las unidades son en segundos [s].



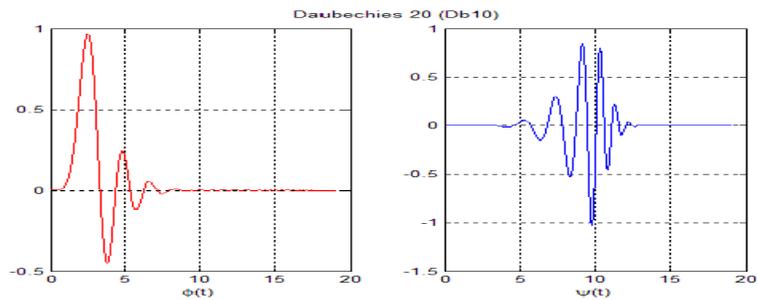
(a)



(b)



(c)



(d)

Figura 2-13 Wavelets Daubechies. (a) Daubechies de 4 coeficientes. (b) Daubechies de 6 coeficientes. (c) Daubechies de 12 coeficientes. (d) Daubechies de 20 coeficientes.

La Figura 2-14 muestra la familia de wavelets Coiflets (Coif). La Figura 2-14(a) es Coif1, la Figura 2-14(b) es Coif2 y la Figura 2-14(c) muestra la wavelet Coif3. Las funciones escala ϕ se encuentran del lado izquierdo y en rojo, mientras que las funciones madre ψ aparecen en el lado derecho y en azul. Las unidades son en segundos [s].

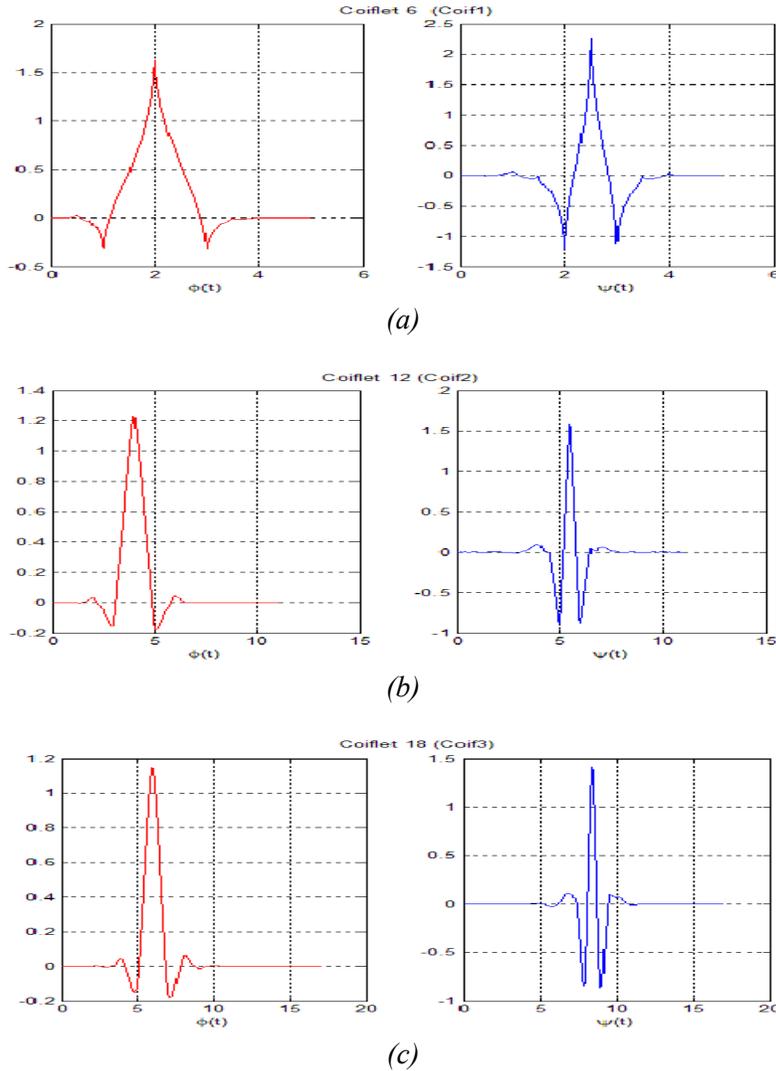
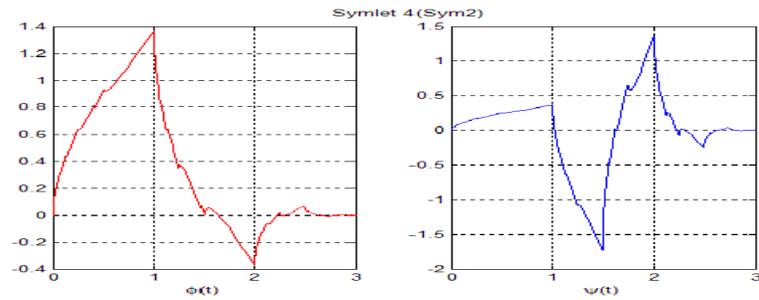
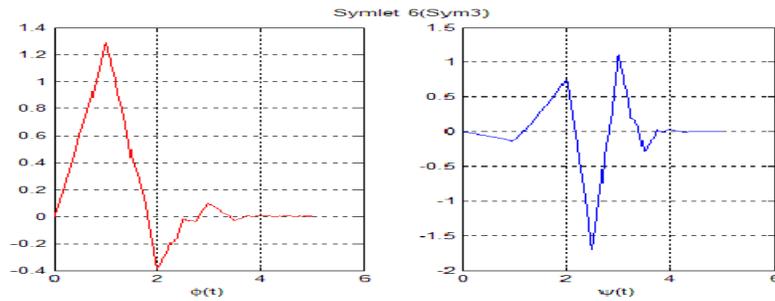


Figura 2-14 Wavelets Coiflets. (a) Coiflet de 6 coeficientes. (b) Coiflet de 12 coeficientes. (c) Coiflet de 18 coeficientes.

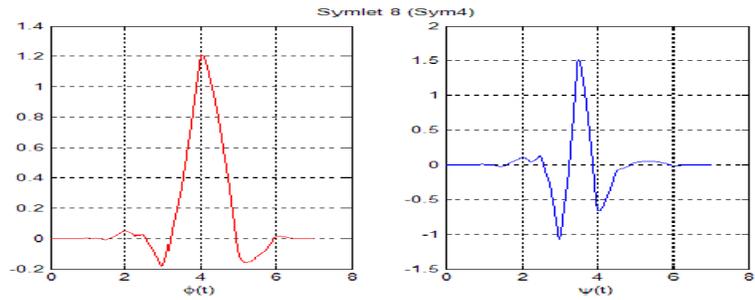
La Figura 2-15 muestra la familia de wavelets Symlet (Sym). La Figura 2-15(a) es la Sym2, la Figura 2-15(b) es la Sym3, la Figura 2-15(c) es la Sym4 y la Figura 2-15(d) es la wavelet Sym8. Las funciones escala ϕ se encuentran del lado izquierdo y en rojo, mientras que las funciones wavelet madre ψ aparecen en el lado derecho y en azul. Las unidades son en segundos [s].



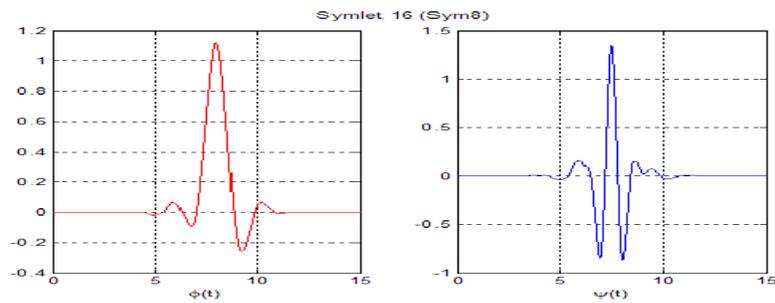
(a)



(b)



(c)



(d)

Figura 2-15 Wavelets Symlets (a) Symlet de 4 coeficientes. (b) Symlet de 6 coeficientes. (c) Symlet de 8 coeficientes. (d) Symlet de 16 coeficientes.

2.9.4 Wavelets para imágenes.

En dos dimensiones, la representación wavelet puede ser obtenida con un algoritmo similar al presentado en la sección 2.9.2. El algoritmo es un esquema piramidal y puede utilizar filtros separables y no separables. El desarrollo aquí presentado está basado en filtros separables, encontrándose algunos trabajos sobre filtros no separables en (Vetterli, et al., 1995) y (Akansu, et al., 2001)

La transformada Wavelet Discreta 2D es una extensión del algoritmo unidimensional. En cada paso de descomposición obtendremos un conjunto de coeficientes *LL*, *LH*, *HL* y *HH*. El algoritmo se ilustra en la Figura 2-16. Primero se convoluciona cada renglón de la imagen con los filtros *L* (h_0) y *H* (h_1) de una dimensión, reteniendo los renglones noes. El proceso se repite convolucionando las columnas de la matriz resultante con los filtros unidimensionales explicados en la sección 2.9.2 y conservamos solo las columnas noes. En otras palabras, primero se aplica la DWT a los renglones y después se aplica la DWT a las columnas de la matriz resultante. Los detalles se concentran en el coeficiente *HH* y la imagen promediada se concentra en *LL*. El siguiente nivel de descomposición se obtiene repitiendo el algoritmo tomando como imagen de entrada el coeficiente *LL*.

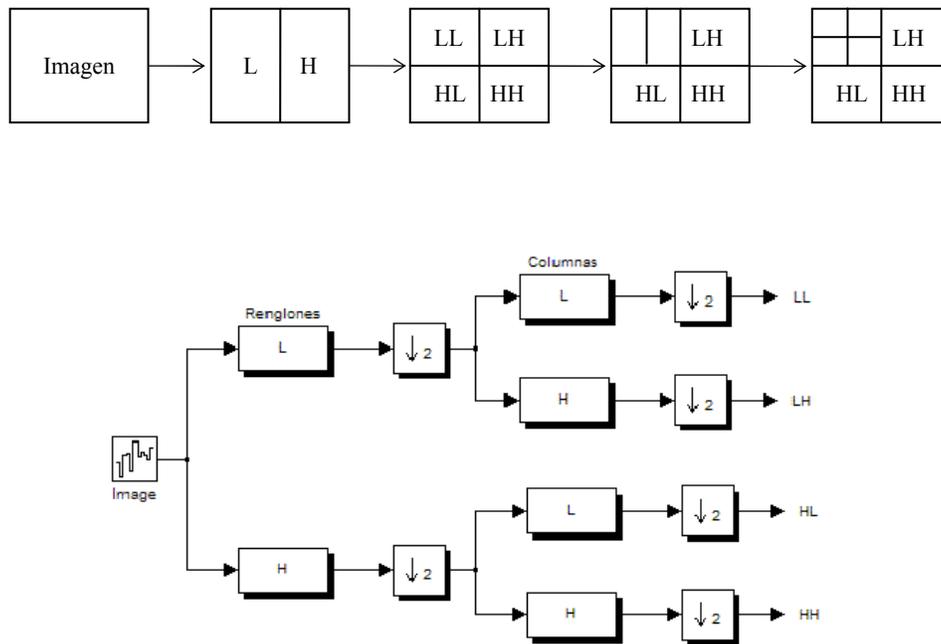


Figura 2-16 Esquema de descomposición de la transformada Wavelet Discreta en 2D.

El algoritmo de reconstrucción también es una extensión del caso unidimensional explicado en la sección 2.9.2. En cada paso de reconstrucción o síntesis, la imagen es obtenida a partir de los coeficientes *LL*, *LH*, *HL* y *HH* del nivel anterior. Agregamos entre cada columna una columna de ceros y convolucionamos los renglones con los filtros unidimensionales, al resultado le agregamos un renglón de ceros entre cada renglón y convolucionamos las columnas con los filtros

unidimensionales. El algoritmo se muestra en la Figura 2.17. La imagen reconstruida es de muy buena calidad y los errores que pueden presentarse son debidos a la precisión del tipo de datos utilizados.

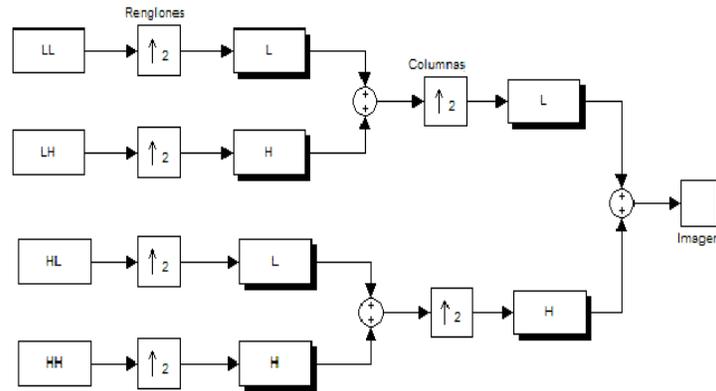


Figura 2-17 Esquema de reconstrucción de la transformada Wavelet Discreta en 2D.

La Figura 2-18 muestra un ejemplo de DWT aplicada a una imagen. En este caso, se utilizó el wavelet Coiflet3 (Coif3). La Figura 2-18(a) es la imagen original, la Figura 2-18(b) es la descomposición wavelet en dos niveles y la Figura 2-18(c) es la imagen reconstruida.

2.10 *Inconveniencias de las wavelets ortogonales.*

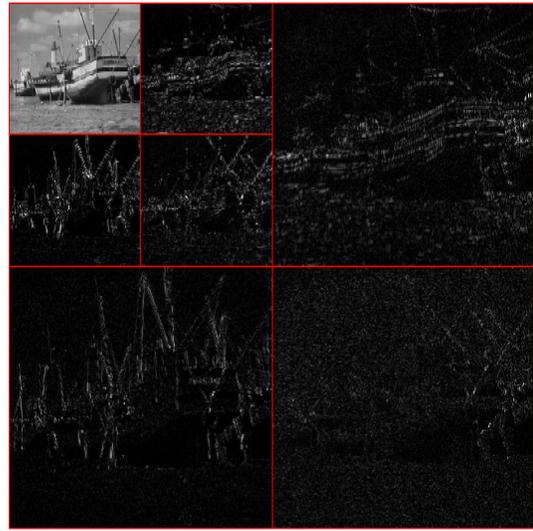
La propiedad de ortogonalidad de las wavelets involucra ciertas inconveniencias en su representación. La mayor dificultad se presenta con la falta de invariancia a la traslación: el contenido de las subbandas de la wavelet es inestable bajo traslaciones de la señal de entrada. La transformada Wavelet también es inestable con respecto a dilataciones de la señal de entrada, y en dos dimensiones, presenta problemas con rotaciones de la imagen de entrada. Todos estos problemas se engloban en un tipo de invariancia a traslación llamada *shiftability* (Simoncelli, et al., 1992).

Una de las principales razones de que las wavelets no sean invariantes a la traslación es el proceso de decimación. Esto significa, que si desplazamos una señal antes de obtener su transformada Wavelet, obtendremos resultados completamente diferentes, lo cual es indeseable en cualquier tarea de procesamiento, Figura 2-19.

Una solución al problema de invariancia es con el algoritmo “à trous”, una versión de la transformada Wavelet sin decimación (Shensa, 1992).



(a)



(b)



(c)

Figura 2-18 Ejemplo de DWT. (a) Imagen original de tamaño 512×512. (b) Descomposición en dos niveles, utilizando la wavelet Coif3. (c) Reconstrucción de la imagen.

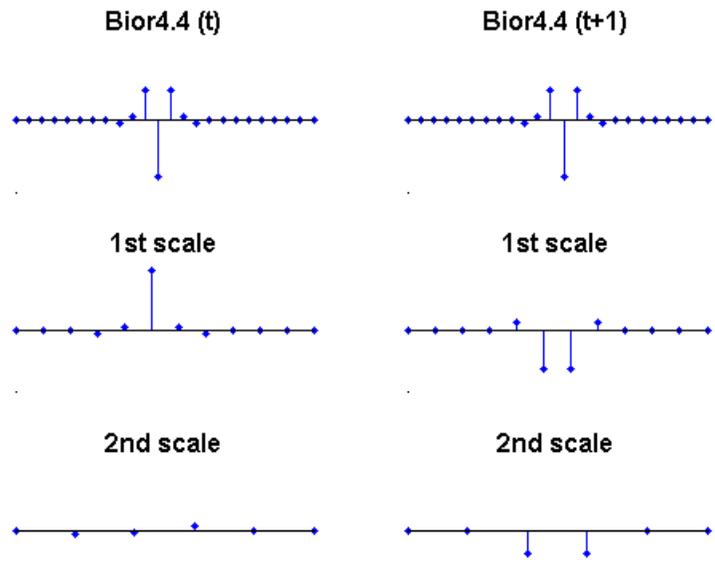


Figura 2-19 Efecto de la traslación de la base biortogonal bior4.4 (Nomenclatura de Matlab). Cuando dos señales de entrada son exactamente las mismas, pero una está desplazada una muestra a la derecha. El primer renglón representa la señales de entrada, los siguientes renglones la primera y segunda descomposición wavelet respectivamente. El desplazamiento provoca una reorganización de los valores de los coeficientes. Ejemplo tomado con permiso de la tesis de doctorado de Rafael Redondo del CSIC, España (Redondo, 2007).

2.10.1 Problemas con los bordes de las imágenes.

En el desarrollo de los temas se ha asumido que se aplican para señales infinitas. Sin embargo esto trae problemas al momento de procesar las imágenes con los bancos de filtros, ya que las muestras cercanas a los bordes de la imagen no están bien definidas.

Existen varios métodos que tratan sobre estos problemas comúnmente referidos como wavelets sobre un intervalo (Cohen, et al., 1992). (Lee, et al., 2007) y (Williams, et al., 1995)

De manera general encontramos dos enfoques para tratar con los bordes. El primero es extender de alguna forma la imagen o señal de tal forma que todas las muestras estén definidas, tal enfoque es conocido como extensión de la señal. El segundo enfoque es cambiar el filtro en los límites de la señal o imagen para evitar los valores no definidos, esta aproximación se conoce como filtrado de fronteras.

El enfoque más utilizado es el de extensión de la señal y los métodos que se utilizan son: agregar ceros (*zero-padding*), extensión periódica y suavizado.

- *Zero-padding.*

Este método asume que los valores fuera de la imagen son ceros. La desventaja es que se crean discontinuidades artificiales en los bordes de la imagen.

- *Extensión periódica.*

Un método para disminuir los efectos de los bordes es colocar valores similares fuera de la imagen a través de reflejos periódicos o simetrías. En el caso del reflejo periódico, el método asume que la señal o imagen se repite fuera de los bordes. La desventaja es que crea discontinuidades producto de las derivadas de primer orden en los bordes. Sin embargo, este método funciona bien para imágenes.

- *Suavizado*

Existen distintos métodos de extensión de las imágenes mediante suavizado. La idea es que las muestras fuera de la imagen o soporte de la señal sean obtenidas mediante la extrapolación de los valores de sus derivadas. Este método funciona muy bien para señales suaves.

2.11 Referencias.

- **Akansu A and Haddad R** Multiresolution signal decomposition. Transforms, subbands and wavelets [Book]. - New Jersey, E. U. : Academic Press, 2001.
- **Bogges A and Narcowich F** A First course in wavelets with Fourier analysis [Book]. - Texas, E. U. : Prentice Hall, 1999.
- **Bracewell R** The Fourier transform and its applications [Book] / ed. McGraw-Hill. - New York, E. U. : [s.n.], 1986.
- **Burt P and Adelson E** The laplacian pyramid as a compact image code [Article] // IEEE Transactions on Communications. - 1983. - 4 : Vols. C-31. - pp. 532-540.
- **Cohen A [et al.]** Multiresolution analysis, wavelets and fast algorithms on an interval [Journal] // Comptes Rendus de l'Académie des Sciences. Série I. Mathématique. - 1992. - 5 : Vol. 316. - pp. 417-421.
- **Cohen A, Daubechies I and Feauveau J** Biorthogonal bases of compactly supported wavelet [Journal] // Comm. Pure & Appl. Math. - 1992. - Vol. 45. - pp. 485-560.
- **Coifman R and Meyer Y** Orthonormal Wave Packet Bases [Report]. - New Heaven, E. U. : Dept. Math. Yale University, 1990.
- **Crowley J** A representation for visual information [Informe] : CMU-RI-TR-82-07 / Robotics Institute ; Carnegie Mellon University. - 1987.
- **Daubechies I** Ten lectures on wavelets [Book]. - Filadelfia, E. U. : SIAM, 1992.
- **Debnath L** Introduction to Hilbert spaces with applications [Book] / ed. Press Academic. - 1990.
- **Duffin R and Schaeffer A** A class of nonharmonic Fourier series [Journal] // Transactions of the American Mathematical Society. - 1952. - 2 : Vol. 72. - pp. 341-366.

- **Esteban D and Galan C** Application of quadrature mirror filters to split band voice coding schemes [Conference]. - 1977. - Vol. 2. - pp. 191-195.
- **Grossman A and Morlet J** Decomposition of hardy functions into square integrable wavelets of constant shape [Journal] // SIAM Journal on Mathematical Analysis. - 1984. - 4 : Vol. 15. - pp. 723-736.
- **Hall E, Rouge J and Wong R** Hierarchical search for image matching [Conference]. - 1976. - Vol. 15. - pp. 791-796.
- **Holschneider M, Kronland-Martinet R and Morlet J** A real-time algorithm for signal analysis with the help of the wavelet transform [Conference] // Proceedings of the International Conference Wavelets, Time-Frequency Methods and Phase Space. - 1989. - pp. 286-297.
- **Hunter G and Steiglitz K** Operations on images using quadtrees [Journal] // IEEE Transactions on Pattern Analysis and Machine Intelligence. - 1979. - 2 : Vols. PAMI-1. - pp. 145-154.
- **Kirac A and Vaidyanathan P** Optimality of Orthonormal Transforms for Subband Coding [Conference] // Proc. of 8th IEEE DSP Workshop. - Utah, E. U. : [s.n.], 1998.
- **Konderink J** The structure of images [Journal] // Biological Cybernetics. - 1984. - Vol. 50. - pp. 363-370.
- **Lee W and Kassim A** Signal and Image Approximation Using Interval Wavelet Transform [Article] // IEEE Transactions on Image Processing. - 2007. - 1 : Vol. 16. - pp. 46-56.
- **Mallat G** Multiresolution approximations and wavelet orthonormal bases of $L_2(\mathbb{R})$ [Journal] // Transactions of the American Mathematical Society. - 1989. - 1 : Vol. 315. - pp. 69-87.
- **Mallat S** A theory for multiresolution signal decomposition: The wavelet representation [Journal] // IEEE Transactions on Pattern Analysis and Machine Intelligence. - 1989. - 7 : Vol. 11. - pp. 674-683.
- **Marr D and Poggio T** A theory of human stereo vision [Article] // Proceedings of Royal Society. - 1979. - Vols. B-204. - pp. 301-328.
- **Mertins A** Signal Analysis. Wavelets, filter banks, time-frequency transforms and applications [Book]. - Sussex, Inglaterra : John Wiley and Sons, 1999.
- **Meyer Y** Ondelettes et opérateurs, volumen I: Ondelettes [Book]. - Paris : Hermann, 1990.
- **Piella G and Heijmans H** An adaptive update lifting scheme with perfect reconstruction [Conference] // Proceedings of the International Conference on Image Processing. - 2001. - Vol. 3. - pp. 190-193.
- **Redondo R** New Contributions on Image Fusion and Compression [Book]. - Madrid, España : Tesis Doctorado. Universidad Politécnica de Madrid, 2007.
- **Rosenfeld A and Thurston M** Edge and curve detection for visual scene analysis [Article] // IEEE Transaction on Computers. - 1971. - 5 : Vols. C-20. - pp. 562- 569.
- **Shensa M** The Discrete Wavelet Transform: Wedding the À Trous and Mallat Algorithms [Article] // IEEE Transactions on Signal Processing. - 1992. - 10 : Vol. 40. - pp. 2464-2482.
- **Simoncelli E [et al.]** Shiftable Multiscale Transform [Article] // IEEE Transactions on Information Theory. - 1992. - 2 : Vol. 38. - pp. 587-607.
- **Smith M and Barnwell T** A procedure for designing exact reconstruction filter banks for tree-structured subband coders [Conference]. - 1984. - pp. 434-441.
- **Soille P** Morphological Image Analysis [Book]. - Berlin : Springer-Verlag, 1999.
- **Speziale de Guzmán L** Apuntes de Algebra Lineal [Book] / ed. Limusa. - México : [s.n.], 2003.

- **Vaidyanathan P** Theory and design of m-channel maximally decimated quadrature mirror filters with arbitrary m, having the perfect-reconstruction property [Journal]// IEEE Transactions on Acoustic, Speech and Signal Processing. - 1987. - 4 : Vol. 3. - pp. 476-492.
- **Vetterli M and Kovacevic J** Wavelets and Subband Coding [Book]. - New Jersey : Prentice Hall, 1995.
- **Ville W** Theorie et applications de la notion de la signal analytique [Journal]// Cables et Transmission. - 1948. - Vol. 2a. - pp. 61-74.
- **Williams J and Amaratunga K** A Discrete Wavelet Transform Without Edge Effects Using Wavelet Extrapolation [Report]. - Massachusetts, E. U. : IESL Technical Report, 1995.
- **Witkin A** Scale-space filtering [Conference]. - Karlsruhe, Germany : [s.n.], 1983. - pp. 1019-1022.

Fusión de Datos.

"Black holes result from God dividing the universe by zero"
Unknown Author

3.1 Fundamentos de la fusión de datos.

Los avances en microelectrónica y comunicaciones han permitido que los sensores y satélites puedan recolectar una amplia cantidad de datos producto de la respuesta y detección de estímulos físicos, tales como el movimiento, la luz, el calor, etc. en una amplia variedad de longitudes de onda, resoluciones espaciales y diversas posiciones angulares. La combinación de todos estos datos provenientes de múltiples sensores resultan en beneficio de muchas aplicaciones y sistemas, ya que la fusión de datos extiende el rango de operación, la cobertura espacial y temporal, reduce incertidumbre, incrementa el tiempo de disponibilidad y compacta la información (Piella, 2003).

Aunque este trabajo de tesis está orientado a la fusión de información visual, la fusión de múltiples sensores abarca muchas otras fuentes de información que no están relacionadas con las imágenes.

Varias tareas de decisión pueden verse beneficiadas por la fusión de datos de múltiples sensores, incluyendo la estimación y las pruebas de hipótesis. Las tareas de decisión pueden ser vistas también como pruebas de detección de algún fenómeno (p. ej. una forma particular de un electrocardiograma o alguna ruptura en un modelo estadístico). Un campo donde la fusión ha sido ampliamente utilizada es en rastreo (*tracking*), el cual consiste en evaluar el movimiento de un objeto (Liu, et al., 2001). Esta tarea incluye decisiones parciales como la detección y posible reconocimiento de objetos de interés.

Otra área importante está relacionada con la clasificación y reconocimiento de patrones. En tales problemas, la decisión consiste en etiquetar objetos en una escena, atributos de un pixel o asignar una imagen a una determinada clase.

La fusión de datos abarca tantas áreas y su dominio es bastante amplio que es difícil proporcionar una definición precisa. Pohl y Genderen (Pohl, et al., 1998) proponen “La fusión de imágenes es la combinación de dos o más imágenes diferentes para formar una nueva usando cierto algoritmo”. Sin embargo, esta definición está restringida a imágenes. Mangoli (Mangoli, 1994) proporciona una definición más general y también se refiere a la calidad. Él define a la fusión de datos como un “Conjunto de métodos, herramientas y medios que utilizan datos provenientes de varios sensores de diferente naturaleza para incrementar la calidad (en el sentido amplio) de la información solicitada”. Una discusión más amplia sobre las definiciones de fusión de datos se puede encontrar en (Wald, 1999).

El propósito de la fusión de imágenes es crear nuevas imágenes que sean más adecuadas para los propósitos de la percepción visual humana, detección de objetos y reconocimiento de objetivos, mejorando el entendimiento de la escena, no sólo en términos de posición y geometría, sino en términos de interpretación.

Dependiendo del tipo de aplicación podemos reconocer las siguientes situaciones, donde los diferentes tipos de información están disponibles:

- Imágenes provenientes de un único sensor: este es el caso donde las imágenes provienen de varios canales de un satélite, de varias imágenes multi-eco en resonancia magnética (MR) o de una secuencia de imágenes de una escena. Los datos son homogéneos y corresponden a medidas físicas similares.
- Imágenes provenientes de diferentes sensores: este es el caso más utilizado, donde las medidas de diferentes fenómenos físicos llevan a diferentes y complementarios puntos de vista de una escena. En el campo de la percepción remota, la información de color es proporcionada por los tres sensores que cubren la longitud de onda del rojo, verde y azul. Estos sensores tienen un número pequeño de píxeles (baja resolución espacial) por lo que no pueden captar los objetos pequeños y los detalles. Estos objetos pequeños y los detalles pueden ser observados por un sensor diferente (pancromático), el cual tiene un número alto de píxeles (alta resolución espacial) pero sin información de color. Con un proceso de fusión podemos obtener una única imagen que contenga la alta resolución espacial y la información de color.
- Varias piezas de información extraídas de una imagen: esta situación se refiere a que varios sensores son aplicados a una imagen para extraer piezas de información. La información extraída puede pertenecer a un mismo objeto o diferentes objetos y entonces se busca la interpretación global de la escena.
- Imágenes y otras fuentes de información: la fuente adicional de información puede ser un modelo como un mapa, bases de conocimiento, reglas de inferencia, etc.

3.1.1 Objetivos y retos de la fusión de imágenes.

El objetivo de la fusión de imágenes es integrar la información redundante y complementaria, proveniente de las imágenes de entrada, para crear una imagen compuesta que contenga una mejor

descripción de la escena que cualquier imagen individual de entrada. Una de las primeras ventajas de la fusión es la reducción de los datos mejorando la eficiencia en el almacenamiento. Utilizando la información redundante, la fusión de imágenes puede mejorar la precisión, la confianza y, mediante el uso de la información complementaria, la fusión puede mejorar la capacidad de interpretación con respecto a tareas posteriores.

Considerando los objetivos y las ventajas de la fusión de imágenes, podemos imponer algunos requerimientos dentro del proceso de fusión (Rockinger, 1996.):

- ❑ El algoritmo de fusión no debe descartar cualquier información importante en las imágenes de entrada.
- ❑ El algoritmo de fusión no debe introducir artefactos o inconsistencias que puedan distraer al observador o alterar el siguiente paso de procesamiento.
- ❑ El algoritmo de fusión debe ser robusto y de confianza, es decir, capaz de ser tolerante a imperfecciones como ruido y errores de registro.

Los requerimientos arriba enlistados son muy difíciles de alcanzar y aun más difíciles de evaluar. El problema con la evaluación de los algoritmos recae en el hecho de que no existe una imagen perfecta o de referencia con la que podamos comparar. Este tema lo trataré con más detalle en el siguiente capítulo.

En el proceso de fusión destaca la importancia del conocimiento acerca de los datos de entrada y la aplicación de que se trata. Sin embargo, la fusión de imágenes debe ser independiente de las distintas modalidades de los datos.

Algunas dificultades que aparecen durante el proceso de fusión están relacionadas con:

- ❑ Manejo de información complementaria. Algunas características aparecen en una imagen pero no en la otra.
- ❑ Información común contradictoria. Existen objetos y regiones que aparecen en ambas imágenes pero con distinto contraste, en este caso, la fusión utilizando criterios aritméticos como la suma y el promedio no son una buena aproximación, Figura 3-1.
- ❑ Disparidad entre sensores. Las imágenes de entrada provienen de distintos tipos de sensores, los cuales tienen diferentes rangos y resoluciones.

Esta lista de problemas no es exhaustiva, ya que existen algunos otros problemas como la presencia de ruido, limitaciones en el hardware o la propia calibración de los sensores.

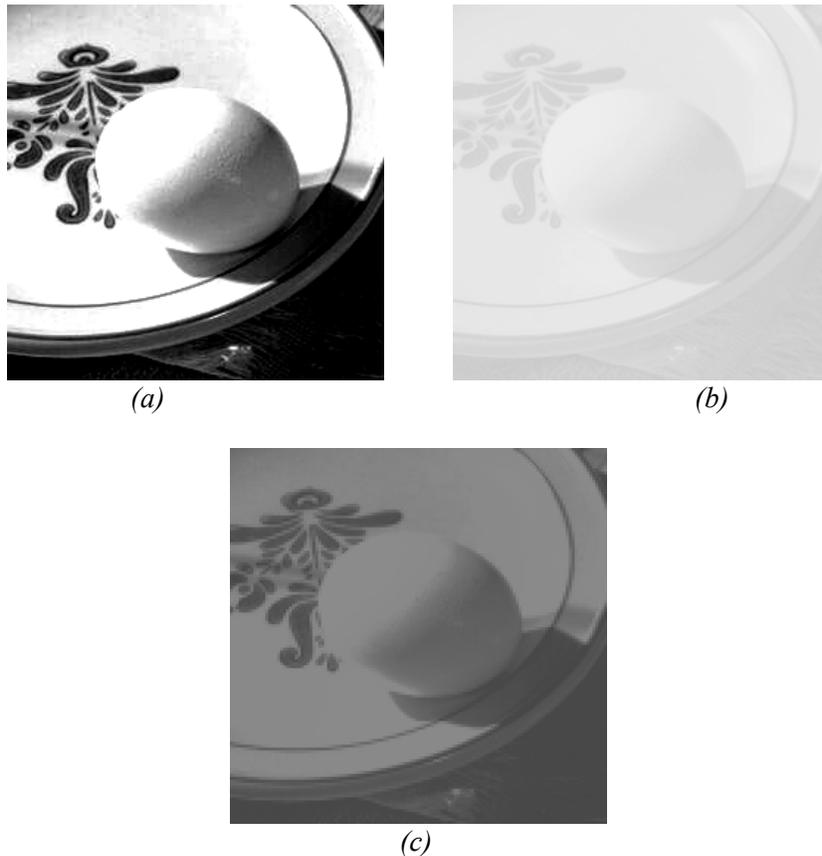


Figura 3-1 Imágenes sintéticas con diferente contraste. (a) Imagen con alto contraste. (b) Imagen con bajo contraste. (c) Imagen fusionada en FUSIONLAB, la regla de selección de coeficientes empleada es AVERAGE. Se aprecia que este tipo de métodos aritméticos como el promediado de coeficientes no son los más adecuados ya que se obtienen imágenes grises.

3.2 Esquemas de fusión.

El proceso de fusión de datos puede tomar lugar en diferentes niveles de representación de la información. Una clasificación común es distinguir tres niveles entre píxeles, características y simbólico (Pohl, et al., 1998), aunque estos niveles pueden combinarse entre sí (Lou, et al., 1995).

La fusión de imágenes a nivel pixel, Figura 3-2(a), es la fusión que se realiza en el nivel más bajo de procesamiento. Genera una imagen fusionada en la cual, los valores de cada uno de los píxeles provienen de un conjunto de imágenes de entrada. La fusión en este nivel puede ser realizada en el dominio espacial o frecuencial. Sin embargo, este tipo de fusión conduce a efectos no deseados como la reducción del contraste (Li, et al., 2001).

La fusión a nivel características, Figura 3-2(b), requiere primero la extracción de características relevantes de las imágenes de entrada. Las características pueden ser intensidades de píxeles, bordes o texturas dependiendo de la naturaleza de las imágenes. Los datos fusionados también se pueden utilizar para clasificación o detección.

La fusión simbólica o a nivel simbólico, Figura 3-2(c), permite combinar la información en el nivel más alto de abstracción. Las imágenes de entrada son procesadas de forma individual para extraer la información y clasificarla. Esto proporciona representaciones simbólicas que son fusionadas de acuerdo a reglas de decisión basadas en lógica difusa, métodos estadísticos, etc. Dependiendo de qué método se utilice, la fusión se clasifica en fusión dura o fusión suave (Kor, et al., 2004).

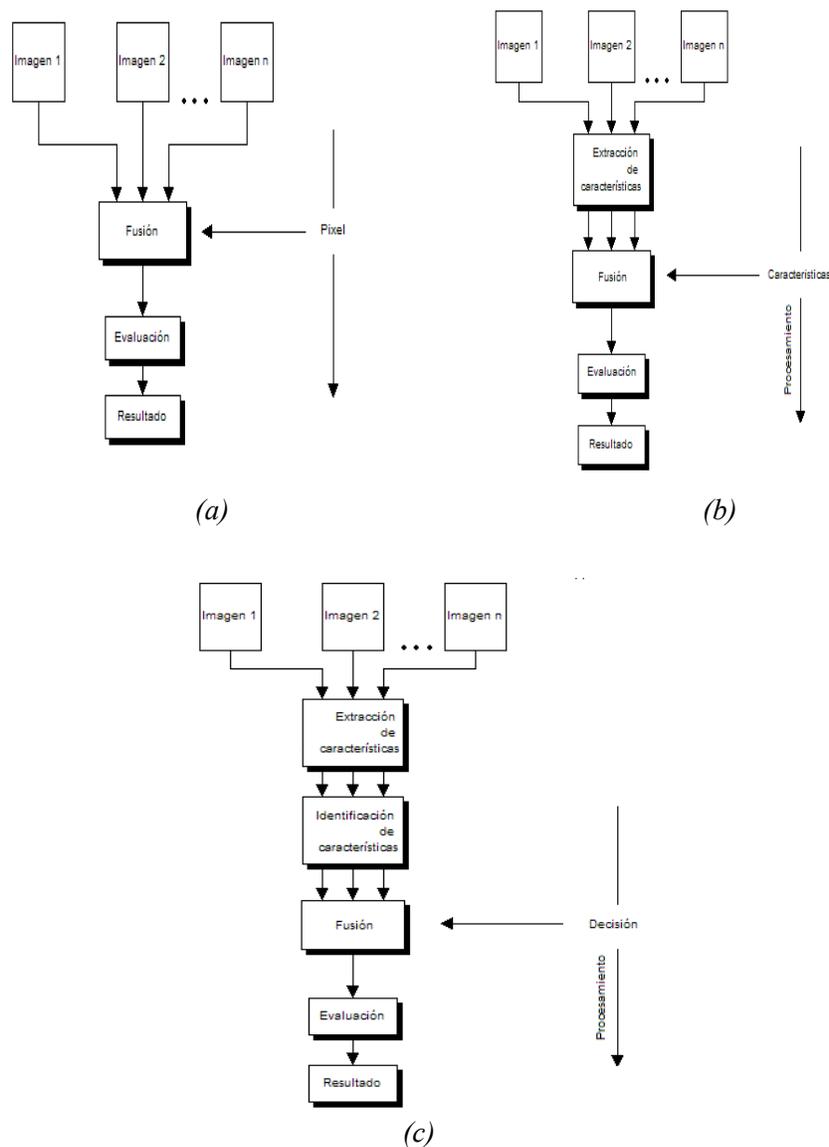


Figura 3-2 Esquemas de fusión de imágenes. (a) Fusión a nivel píxel. (b) Fusión a nivel características. (c) Fusión a nivel decisión o fusión simbólica.

3.3 Técnicas de fusión de imágenes.

En la sección anterior se mencionó que los algoritmos de fusión de imágenes deben ser independientes de las aplicaciones o modalidades de los datos de entrada. Sin embargo, las distintas técnicas de fusión que existen en la literatura han sido desarrolladas para aplicaciones específicas y poseen una naturaleza heurística. Existen algunos métodos de fusión que pueden producir imágenes multiespectrales de alta resolución a partir de imágenes pancromáticas de alta resolución e imágenes multiespectrales de baja resolución (Pohl, et al., 1998). Otros métodos de fusión están orientados a fusionar imágenes médicas, como tomografía por emisión de positrones (PET), tomografía computarizada (CT) y resonancia magnética (MR) (Muñoz, 2005). Inclusive, existen algunos otros métodos específicos para imágenes multifocales (Li, et al., 2001).

En general, las técnicas pueden ser agrupadas en dos clases: (1) técnicas relacionadas con el color y (2) métodos estadísticos y numéricos. Las primeras utilizan los tres canales de imágenes en el espacio de color RGB (p. ej. IHS). Mientras que las segundas realizan operaciones aritméticas como suma, promedio, entre otros. Algunos métodos más sofisticados calculan el nivel de actividad (energía) en una región y, en base a ella, elaboran una regla de fusión. Los métodos más poderosos utilizan un enfoque multirresolución como las wavelets o la transformada Hermitiana. La siguiente sección describe algunas de estas técnicas.

3.3.1 Intensity Hue Saturation.

El modelo de color IHS puede utilizarse para fusionar tres bandas espectrales de una imagen. Por ejemplo, las bandas espectrales de Landsat con su banda pancromática (Harrison, et al., 1990). Esta transformación separa la información espacial (I) y la espectral (H , S) de una imagen RGB estándar y está relacionada con la percepción humana del color. La ecuación de transformación se expresa en (3.1), (3.2) y (3.3) (Pratt, 1991).

$$\begin{bmatrix} I \\ v_1 \\ v_2 \end{bmatrix} = \begin{bmatrix} \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ -\frac{1}{\sqrt{6}} & -\frac{1}{\sqrt{6}} & \frac{2}{\sqrt{6}} \\ \frac{1}{\sqrt{6}} & -\frac{1}{\sqrt{6}} & 0 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix} \quad \dots (3.1)$$

$$H = \tan^{-1}\left(\frac{v_2}{v_1}\right) \quad \dots (3.2)$$

$$S = \sqrt{v_1^2 + v_2^2} \quad \dots (3.3)$$

Existen dos formas de aplicar la transformada IHS en fusión de imágenes: directa y por sustitución. La primera se refiere a la transformación de los tres canales asignados a I , H y S (Rast, et al., 1991). La segunda transforma los tres canales de datos RGB en un espacio de color IHS y uno de los canales es reemplazado por un cuarto canal. La transformación inversa de IHS a RGB se presenta en (3.4), (3.5) y (3.6).

$$v_1 = S \cos\{H\} \quad \dots (3.4)$$

$$v_2 = S \sin\{H\} \quad \dots (3.5)$$

$$\begin{bmatrix} R \\ G \\ B \end{bmatrix} = \begin{bmatrix} 1 & -\frac{\sqrt{6}}{6} & \frac{\sqrt{6}}{2} \\ 1 & \frac{\sqrt{6}}{6} & -\frac{\sqrt{6}}{2} \\ 1 & \frac{\sqrt{6}}{3} & 0 \end{bmatrix} \begin{bmatrix} I \\ v_1 \\ v_2 \end{bmatrix} \quad \dots (3.6)$$

La técnica IHS se ha convertido en un procedimiento estándar en el análisis de imágenes. Se utiliza para realzar el color de datos altamente correlacionados, para el realce de características y para el mejoramiento de la resolución espacial.

3.3.2 Transformación de Brovey.

La transformada de Brovey (Tu, et al., 2005) es una combinación de operaciones aritméticas que normaliza las bandas de las imágenes multiespectrales para poder desplegarlas en el espacio de color RGB y multiplica el resultado por otra imagen pancromática para agregar la componente de intensidad o brillantez. Cuando se utilizan bandas espectrales o información de color de un satélite Landsat TM y una imagen pancromática de un satélite SPOT el algoritmo está definido por:

$$\begin{aligned} R &= SPOT * banda5 / (banda2 + banda4 + banda5) \\ G &= SPOT * banda4 / (banda2 + banda4 + banda5) \\ B &= SPOT * banda2 / (banda2 + banda4 + banda5) \end{aligned} \quad \dots(3.7)$$

3.3.3 Análisis de Componentes Principales.

El Análisis de Componentes Principales (PCA) es útil en codificación de imágenes, compresión, realce de la imagen y fusión. Es una técnica estadística que transforma un conjunto de datos de variables correlacionadas a un nuevo conjunto de datos no correlacionados, formando un nuevo conjunto de ejes ortogonales.

Fundamentalmente es una forma de identificar patrones en los datos y expresarlos en forma distinta para resaltar sus similitudes y diferencias. La principal ventaja de PCA es que una vez encontrados los patrones se puede reducir la dimensión sin pérdida significativa de información

Técnicamente hablando, PCA es una transformación lineal ortogonal que transforma el conjunto de datos a un nuevo sistema de coordenadas de tal forma que la banda de mayor varianza es la primera coordenada, también llamada primer componente principal, la segunda banda de mayor varianza es la segunda coordenada y así sucesivamente.

Partiendo del hecho de que una imagen multispectral puede ser considerada como una variable aleatoria, entonces cada banda de la imagen puede ser representada como un vector y puede ser caracterizada por su media y su varianza. Dando una relación entre cada banda mediante los coeficientes de correlación.

Para que PCA trabaje adecuadamente, se debe restar la media de cada una de las bandas al conjunto de datos (Smith, 2003). Este procedimiento produce un conjunto de datos con media cero. Entonces podemos encontrar una transformación tal que:

$$Y = A^T X \quad \dots (3.8)$$

donde se cumple que la $cov(Y)$ es una matriz diagonal y $A^{-1} = A^T$.

Mediante sustitución se puede obtener:

$$\begin{aligned} cov(Y) &= E\{YY^T\} \\ &= E\{(A^T X)(A^T X)^T\} \\ &= E\{(A^T X)(X^T A)\} \quad \dots (3.9) \\ &= A^T E\{XX^T\} A \\ &= A^T cov(X) A \end{aligned}$$

$$\begin{aligned} Acov(Y) &= AA^T cov(X) A \\ &= cov(X) A \quad \dots (3.10) \end{aligned}$$

reescribiendo A como un vector columna $A = [A_1 \ A_2 \ \dots \ A_n]$ y

$$\text{cov}(y) = \begin{bmatrix} \lambda_1 & 0 & \dots & 0 \\ 0 & \lambda_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \lambda_n \end{bmatrix} \quad \dots (3.11)$$

Substituyendo (3.11) en (3.10) se consigue la siguiente relación.

$$[\lambda_1 A_1 \ \lambda_2 A_2 \ \dots \ \lambda_n A_n] = [\text{cov}(x)A_1 \ \text{cov}(x)A_2 \ \dots \ \text{cov}(x)A_n] \quad \dots (3.12)$$

se puede apreciar que $\lambda_i A_i = \text{cov}(x)A_i$, A_i son los vectores característicos de la matriz de covarianza de X . El resultado de la transformación de componentes principales es un conjunto de imágenes no correlacionadas ordenadas de acuerdo a la energía de sus varianzas.

Existen dos tipos de PCA que pueden ser aplicados a la fusión: selectiva o estándar. El primero utiliza todas las bandas disponibles de una imagen de entrada, (p. ej. Landsat TM 1-7). El segundo utiliza sólo una selección de bandas, las cuales son escogidas en base a un conocimiento *a priori* o dependiendo de la aplicación. En el caso de Landsat TM los tres primeros componentes principales (PC) contienen el 98 ó 99 por ciento de la variancia y por lo tanto son suficientes para representar la información (Yésou, et al., 1993).

El algoritmo realiza la transformación de Componentes Principales a una imagen multiespectral y se reemplaza la primera componente por una imagen diferente (substitución de componente principal), después se realiza la transformación inversa (Chavez, et al., 1991). Este método tiene como objetivo incrementar la resolución espacial de una imagen multiespectral introduciendo una imagen con mayor resolución.

3.3.4 Filtrado paso-altas.

El principio del filtrado paso-altas (HPF) es sumar la información de alta frecuencia de la imagen pancromática a las bandas de baja resolución de la imagen multiespectral. La información de alta frecuencia se calcula filtrando la imagen pancromática con un filtro paso-altas o tomando la imagen pancromática original y restándole sus bandas de baja resolución. Comparado con los métodos PCA e IHS, el método de HPF tiene mejor rendimiento (Dongmei, et al., 2001). No obstante, el método HPF tiene limitaciones cuando pasa información de texturas de la banda de alta resolución espacial a los datos de baja resolución (Tauch, et al., 1988).

3.3.5 Problemas de los métodos de fusión.

Los problemas y las limitaciones de los métodos de fusión tradicionales han sido analizados en muchos trabajos, entre ellos se encuentran: (Tu, et al., 2001) y (Varma, et al., 2003).

Uno de los problemas que surgió con los primeros métodos de fusión tiene que ver con la distorsión espectral que se incrementa con el cambio de saturación (S) durante el proceso de fusión. Los métodos de fusión antes mencionados utilizan todas las frecuencias de las imágenes pancromáticas durante el proceso de fusión de tal forma que mantienen casi la misma resolución, pero ocasionan distorsión espectral debido a la poca correlación que existe entre las imágenes pancromáticas y las imágenes multiespectrales. El problema de la distorsión espectral puede producir errores en tareas de clasificación y en reconocimiento de objetos.

La fusión de imágenes utilizando IHS y PCA incrementa la resolución de imágenes de percepción remota, pero sufren el problema que la radiometría en las bandas multiespectrales se modifica después del proceso de fusión, ya que las imágenes pancromáticas usualmente tienen rangos espectrales diferentes a los rangos de las imágenes multiespectrales (Li, et al., 2002).

Los métodos de fusión actuales pueden realzar la resolución espacial de una imagen, pero alteran la información espectral o de forma inversa, pueden preservar las características espectrales pero agregan artefactos a la información espacial. De nueva cuenta aparece el principio de incertidumbre de Heisenberg¹. Finalmente, los métodos de fusión piramidales son muy sensibles a errores de precisión de registro.

Los métodos de fusión basados en la transformada Wavelet utilizan dos variantes en el algoritmo: el algoritmo de Mallat (Mallat, 1989) y el algoritmo *à trous* (Shensa, 1992). El algoritmo de Mallat está basado en la transformación de wavelets diádicas, las cuales utilizan decimación lo que provoca que no sean invariantes a desplazamiento y presentan artefactos debido al *aliasing* en la imagen fusionada. El algoritmo *à trous*, no utiliza decimación, por lo que es invariante a desplazamiento, una característica que lo hace más adecuado para fusión de imágenes.

En la Figura 3-3 se presenta una comparación entre los métodos de fusión mencionados en esta sección. Las imágenes de entrada cubren un área de la ciudad de Sherbrooke en Canadá y fueron tomadas el 20 de mayo de 2001. La imagen de la Figura 3-3(a) es una imagen pancromática de 1 [m] de resolución y la Figura 3-3(b) es una imagen multiespectral de 4 [m] de resolución. El objetivo sólo es comparar visualmente los resultados. Las imágenes fueron tomadas con permiso del artículo (Zhijun, et al., 2005)

¹ En mecánica cuántica, el principio de incertidumbre de Heisenberg afirma que no se puede determinar, simultáneamente y con precisión arbitraria, ciertos pares de variables físicas. En otras palabras, cuanto mayor certeza se busca en determinar la posición de una partícula, menos se conoce su cantidad de movimiento lineal. Este principio fue pronunciado por Werner Heisenberg en 1927.

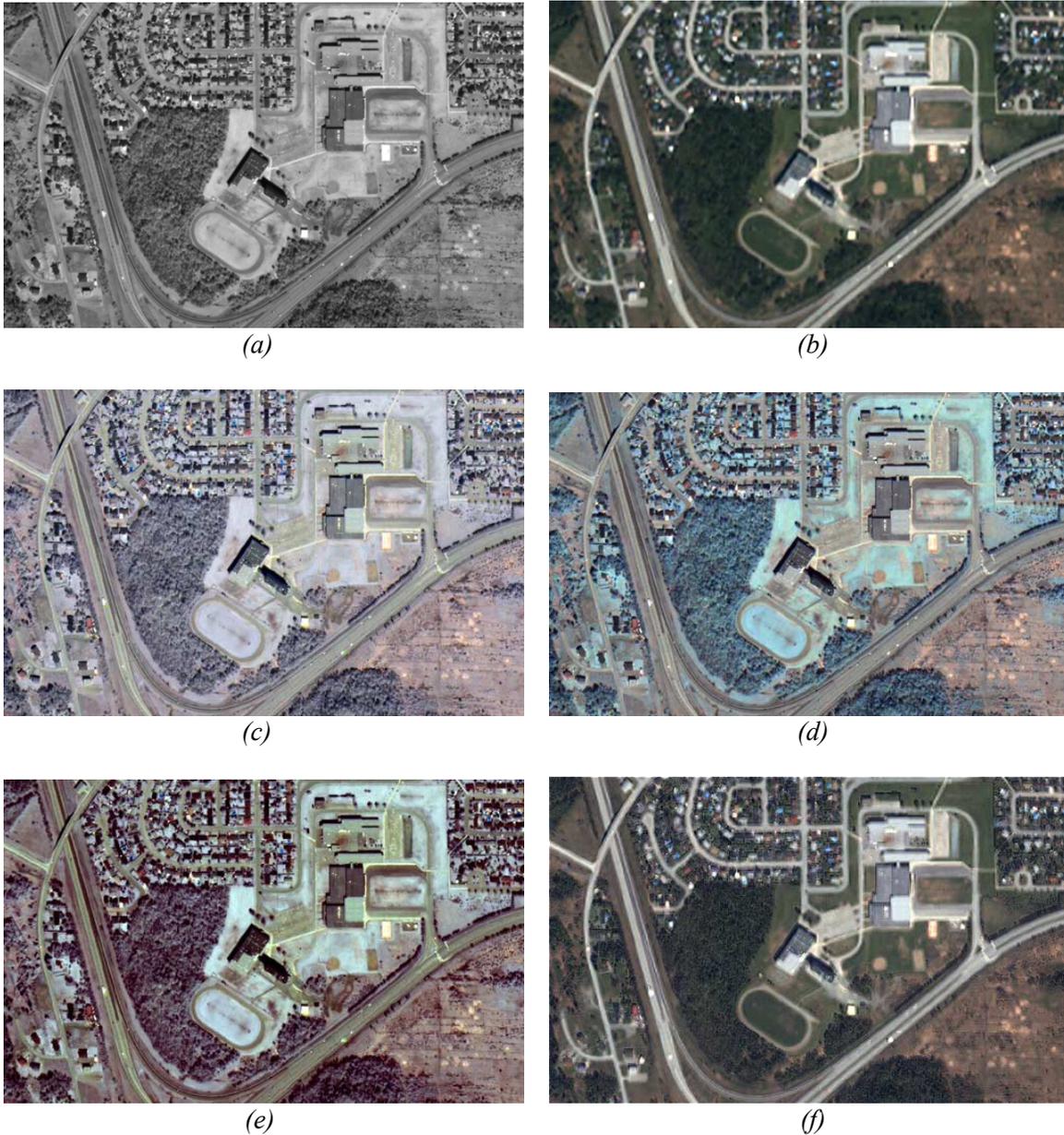


Figura 3-3 Ejemplos de fusión de imagen pancromática y multispectral. Las imágenes cubren un área de la ciudad de Sherbrooke, Canadá, tomadas el 20 de mayo de 2001. (a) Imagen pancromática de 1 m de resolución. (b) Imagen multispectral de 4 m de resolución. (c) Imagen fusionada utilizando **IHS**. (d) Imagen fusionada utilizando **BT**. (e) Imagen fusionada utilizando **PCA**. (f) Imagen fusionada utilizando **HPF**. (g) Acercamiento 200% de la imagen fusionada con **IHS**. (h) Acercamiento 200% de la imagen fusionada con **BT**. (i) Acercamiento 200% de la imagen fusionada con **PCA**. (j) Acercamiento 200% de la imagen fusionada con **HPF**. Fuente (Zhijun, et al., 2005).

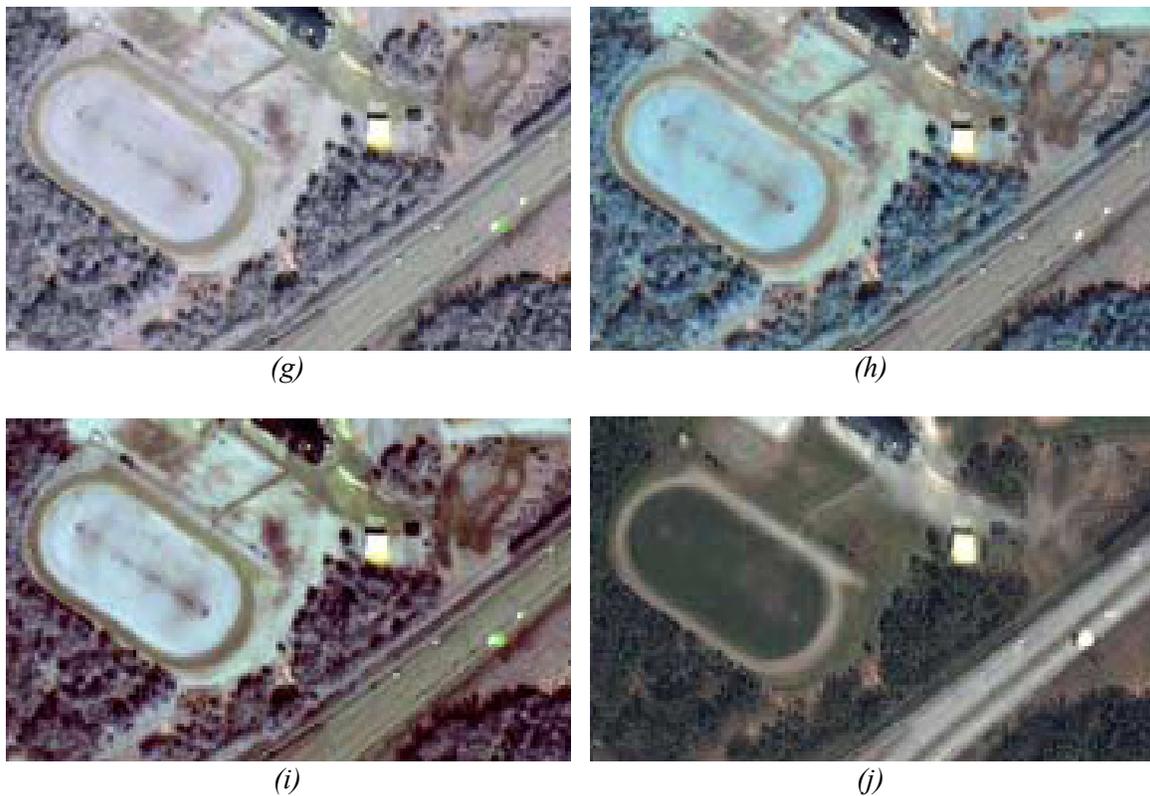


Figura 3-3 (cont.) Ejemplos de fusión de imagen pancromática y multispectral. Las imágenes cubren un área de la ciudad de Sherbrooke, Canadá, tomadas el 20 de mayo de 2001. (a) Imagen pancromática de 1 m de resolución. (b) Imagen multispectral de 4 m de resolución. (c) Imagen fusionada utilizando **IHS**. (d) Imagen fusionada utilizando **BT**. (e) Imagen fusionada utilizando **PCA**. (f) Imagen fusionada utilizando **HPF**. (g) Acercamiento 200% de la imagen fusionada con **IHS**. (h) Acercamiento 200% de la imagen fusionada con **BT**. (i) Acercamiento 200% de la imagen fusionada con **PCA**. (j) Acercamiento 200% de la imagen fusionada con **HPF**. Fuente (Zhijun, et al., 2005).

3.4 Fusión de imágenes basada en la transformada Wavelet Discreta.

En general, el problema de la fusión de imágenes está relacionado con realizar la combinación de información procedente de varias imágenes tomadas de la misma escena para obtener otra imagen que tendrá una mejor representación de la información que cualquiera de las imágenes de entrada.

La idea del esquema de fusión usando la transformada Wavelet Discreta (DWT) está basada en el conocimiento del sistema visual humano, el cual es principalmente sensible a los cambios locales de contraste (Rockinger, 1996.). Huble (Huble, 1987) menciona que los bordes son la característica más destacada en la percepción visual del ser humano.

Los métodos de fusión basados en la DWT han recibido considerable atención (Li, et al., 1994) y (Chipman, et al., 1995), y aunque originalmente los esquemas estaban orientados a fusionar imágenes

SPOT con otras imágenes multirresolución, han aparecido modificaciones para fusionar imágenes médicas (Guihong, et al., 2001) y multifocales (Zhi-guo, et al., 2004). La fusión de imágenes utilizando la DWT es apropiada para realizar las siguientes tareas:

- La DWT permite la descomposición en diferentes niveles preservando la información de la imagen ya que es un enfoque multiescala o multirresolución.
- Los coeficientes de diferentes imágenes obtenidos de la descomposición usando la DWT pueden ser combinados para obtener nuevos coeficientes.
- Una vez mezclados los coeficientes, la imagen fusionada final se logra con la transformada Wavelet Discreta Inversa (IDWT) que preserva la información fusionada en los coeficientes.

El paso clave en la fusión es la regla o método de combinación de los coeficientes en una forma apropiada para obtener la mejor calidad en la imagen fusionada. Esto puede lograrse con diferentes estrategias que se revisarán en la sección 3.4.1.

La Figura 3-4 ilustra el esquema de fusión basado en la DWT. El primer paso es calcular la transformada Wavelet Discreta de cada imagen de entrada en diferentes niveles de descomposición o escalas; es decir, cada una de las imágenes de entrada se descompone en subbandas, las cuales son tratadas como subimágenes o imágenes de menor tamaño. Los píxeles de las subimágenes corresponden a los coeficientes de la descomposición wavelet y tienen valores que fluctúan alrededor de cero excepto para el coeficiente *LL* que tiene valores positivos. En cada nivel existen cuatro coeficientes, pero es de especial importancia el coeficiente *LL* ya que es una aproximación de la imagen. Los coeficientes *LH*, *HL* y *HH* contienen los detalles horizontales verticales y diagonales de la imagen respectivamente. El siguiente nivel de descomposición se obtiene a partir del coeficiente *LL*.

Los valores absolutos más grandes en los coeficientes corresponden a cambios bruscos en la imagen como bordes, líneas y límites de regiones.

El siguiente paso es escoger una regla de fusión que seleccione los píxeles que se preservarán en cada uno de los coeficientes, de esta forma toma lugar la fusión en todos los niveles de resolución, conservando las características más importantes de las imágenes de entrada. Cabe mencionar que la fusión a nivel pixel es limitada, porque las características importantes en las imágenes son más grandes que un pixel.

Finalmente, la imagen fusionada se construye mediante la síntesis de los coeficientes fusionados (a través de la transformada Wavelet Discreta Inversa, IDWT). Algunos algoritmos agregan una etapa de filtrado para eliminar artefactos o mejorar la calidad de la imagen.

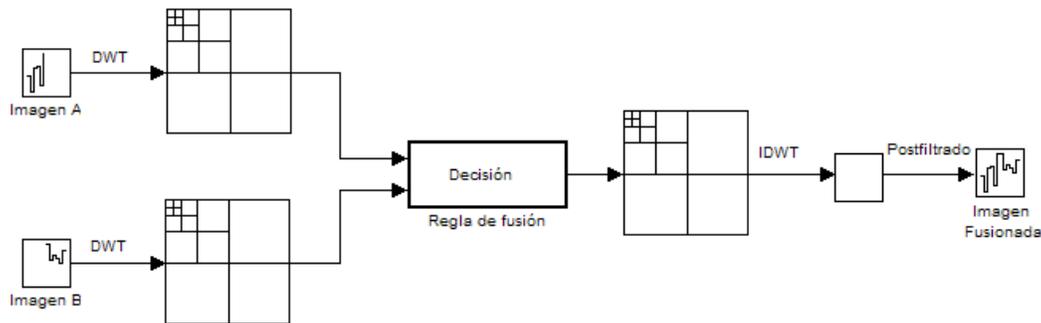


Figura 3-4 Esquema de fusión de imágenes basado en la transformada Wavelet Discreta (DWT).

3.4.1 Reglas de fusión.

Existe un gran número de publicaciones que se enfocan en las distintas reglas de fusión. Las posibilidades para combinar los datos son muy diversas y los detalles sobre ellas pueden encontrarse en (Bloch, 1996) y (Ehlers, 1991). En la mayoría de los sistemas de fusión de datos, la información extraída de las imágenes o sensores se representa de forma numérica o de forma simbólica. Por lo que es importante decidir qué tipo de operadores se usarán para realizar la fusión. Bloch (Bloch, 1996) proporciona una clasificación para los operadores, aquí sólo se mencionarán los operadores de comportamiento constante independientes del contexto, *Context Independent Constant Behavior*, (CICB) y los operadores de comportamiento variable independientes del contexto, *Context Independent Variable Behavior*, (CIVB).

- Operadores CICB.

Los operadores CICB son aquellos que tienen el mismo comportamiento independientemente del tipo de datos o los valores de la información a combinar. Estos operadores trabajan sin ayuda contextual o información externa. El primero de estos operadores se conoce como combinación selectiva o regla de selección del valor máximo, que para el caso de dos imágenes está dada en (3.13).

$$y_F^k = \begin{cases} a_A^k & \text{Si } a_A^k > a_B^k \\ a_B^k & \text{otro caso.} \end{cases} \quad \dots (3.13)$$

Donde a_A^k y a_B^k son los valores de los pixeles en el nivel de descomposición k y y_F^k es el valor del pixel en la imagen fusionada en el dominio wavelet.

Otras alternativas hacen uso de un valor mínimo (3.14), un valor promedio (3.15) o un valor aleatorio (3.16) o incluso un umbral (3.17).

$$y_F^k = \begin{cases} a_A^k & \text{Si } a_A^K \leq a_B^K \\ a_B^k & \text{otro caso.} \end{cases} \quad \dots (3.14)$$

$$y_F^k = \frac{(a_A^k + a_B^k)}{2} \quad \dots (3.15)$$

$$y_F^k = \text{random}\{a_A^k, a_B^k\} \quad \dots (3.16)$$

$$y_F^k = \begin{cases} a_A^k & \text{Si } a_A^K > T \\ a_B^k & \text{otro caso.} \end{cases} \quad \dots (3.17)$$

donde $T \in \mathbb{R}$.

Existen otros operadores que realzan el contraste. Un ejemplo es la multiplicación expresada en (3.18).

$$y_F^k = A(w_1 a_A^k + w_2 a_B^k) + B \quad \dots (3.18)$$

A y B son factores de escala y w_1 y w_2 son parámetros de peso. En particular, este método fue exitosamente aplicado en fusión de datos para Landsat TM y SPOT pancromático (Yésou, et al., 1993).

- *Operadores CIVB.*

Esta clase de operadores son independientes del contexto, como la clase anterior, pero dependen de los valores de a_A^k y a_B^k . Un ejemplo, es la fusión basada en regiones (Zhang, y otros, 1997). Las imágenes de entrada son descompuestas y se calcula la varianza de cada imagen utilizando ventanas de 3×3 ó 5×5 . La actividad es asociada al pixel central de cada ventana y si el valor de la varianza en ambas imágenes es cercano, se considera el promedio de los dos pixeles centrales; en otro caso se toma el valor del pixel más alto.

El cálculo de la variancia puede ser considerado como una operación de filtrado paso-altas no lineal. Sin embargo, este método no tiene ningún significado físico relacionado con la percepción humana.

Existen algunos otros métodos orientados a regiones, donde se utilizan ventanas que cambian de tamaño de acuerdo al valor de la varianza. Si la varianza es alta, indica que hay cambios bruscos dentro de esa ventana, por lo que se utilizan ventanas angostas; si el valor de la varianza es muy pequeño se utilizan ventanas mas grandes ya que no hay cambios bruscos en esa región.

Existen muchos otros métodos que aparecen en la literatura. Sin embargo, como en el caso del método de multiplicación (3.18), no tienen ningún significado físico asociado a la visión humana y son de naturaleza heurística.

3.4.2 Ejemplos.

En esta sección se presentan un par de ejemplos de fusión clásica utilizando la herramienta FUSIONLAB. La Figura 3-5 muestra un ejemplo de fusión aplicada a restauración. Las imágenes de entrada, Figura 3-5(a) y 3-5(b), fueron tomadas del *toolbox* para fusión de Matlab. Las Figuras 3-5(c) y 3-5(d) son las imágenes restauradas.

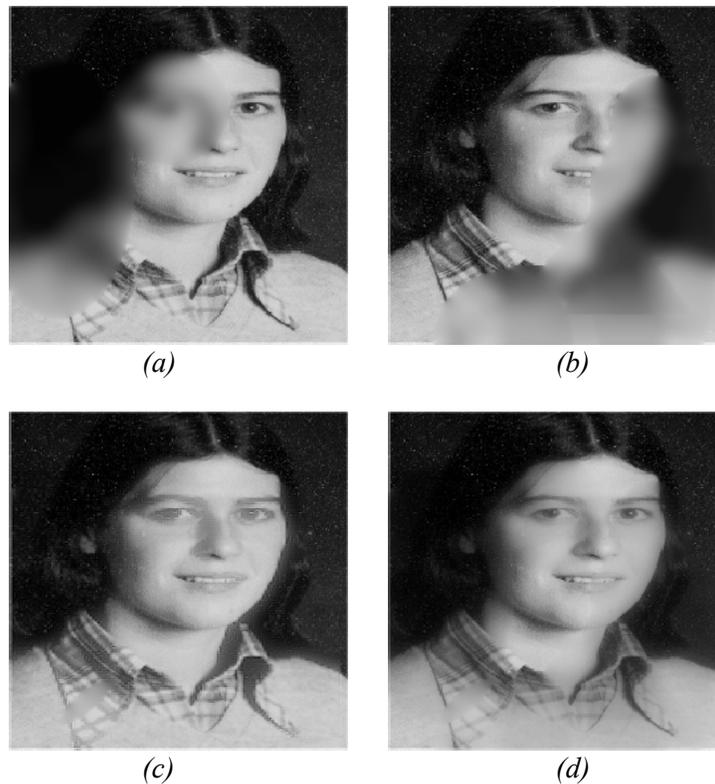


Figura 3-5 Fusión de imágenes basada en comparación de energía pixel a pixel. (a) y (b) son las imágenes de entrada. (c) y (d) Son las imágenes fusionadas resultantes utilizando Db3 y los criterios de fusión son MAX-MAX y AVERAGE-AVERAGE, respectivamente.

La Figura 3-6 muestra un ejemplo de fusión multifocal. La imagen de la Figura 3-6(a) tiene el fondo desenfocado pero la pelota se encuentra en foco, mientras que la Figura 3-6(b) presenta el fondo enfocado pero la pelota está borrosa. Las imágenes de las Figuras 3-6(c) y 3-6(d) son las imágenes resultantes.

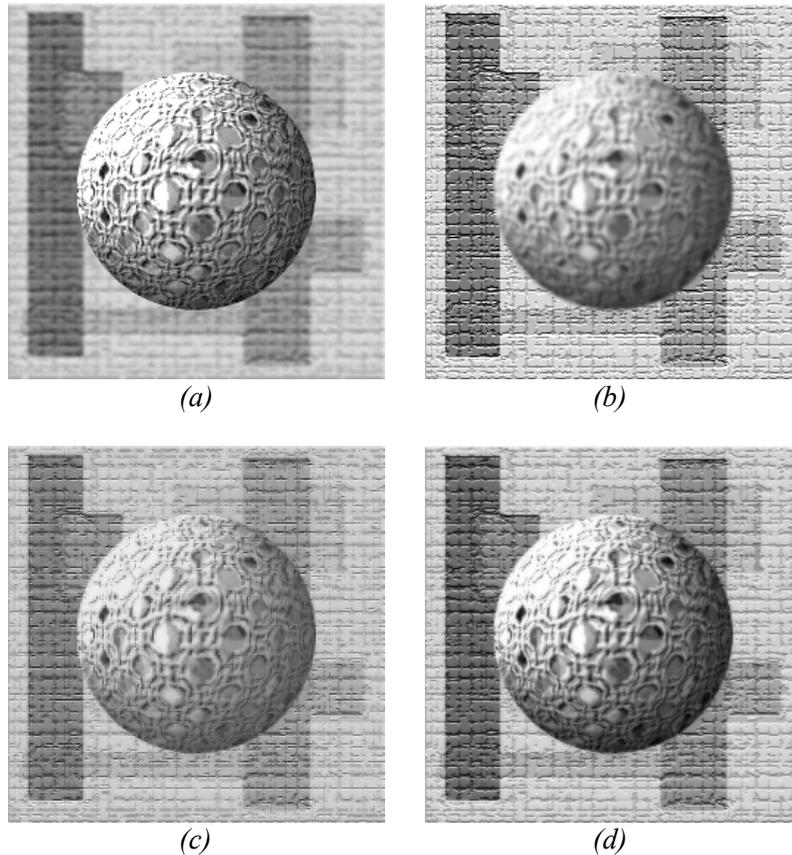


Figura 3-6 Fusión de imágenes basada en comparación de energía pixel a pixel. (a) y (b) Son las imágenes de entrada. (c) y (d) Son las imágenes fusionadas resultantes utilizando Db2, usando los criterios de fusión MIN-MIN y AVERAGE-AVERAGE, respectivamente.

3.5 Referencias.

- **Bloch Isabelle** Information Combination Operators for Data Fusion: A comparative Review with Classification [Journal]// IEEE Transactions on Systems, Man and Cybernetics. - 1996. - 1 : Vol. 26. - pp. 52-67.
- **Chavez P, Sides S and Anderson J** Comparison of three different methods to merge multi-resolution and multi-sectoral data: Landsat TM and SPOT Panchromatic [Journal]// Photogrammetric Engineering and Remote Sensing. - 1991. - 3 : Vol. 57. - pp. 295-303..

- **Chipman L, Orr T and Graham L** Wavelets and image fusion [Conference] // SPIE Wavelets Applications in Signal and Image Processing. - 1995. - Vol. 2569. - pp. 208-219.
- **Dongmei Y and Zhongming Z** Wavelet decomposition applied to image fusion [Conference] // International Conferences on Info-tech and Info-net.. - Beijing, China : [s.n.], 2001. - Vol. 1. - pp. 291-295.
- **Ehlers M** Multisensor image fusion techniques in remote sensing [Journal] // ISPRS Journal of Photogrammetry and Remote Sensing. - 1991. - 1 : Vol. 46. - pp. 19-30.
- **Guihong Q, Dali Z and Pingfan Y** Medical image fusion using two-dimensional discrete wavelet transform [Journal] // Proceedings of SPIE Data Mining and Applications. - 2001. - Vol. 4556. - pp. 86-95.
- **Harrison B and Jupp L** Introduction to image processing. MicroBrain Resource Manual [Report]. - Melbourne, Australia : CSIRO, 1990.
- **Huble D** Eye, brain and vision [Book]. - New York, E. U. : Freeman, 1987.
- **Kor S and Tiwary U** Feature level fusion of multimodal medical images in lifting wavelet transform domain [Conference] // International Conference of the Engineering in Medicine and Biology Society. - 2004. - Vol. 1. - pp. 1479-1482.
- **Li H, Manjunath B and Mitra S** Multisensor image fusion using the wavelet transform [Journal] // IEEE International Conference on Image Processing. - 1994. - 13 : Vol. 1. - pp. 51-55.
- **Li S, Kwok J and Wang Y** Combination of images with diverse focuses using the spatial frequency [Journal] // Information Fusion. - 2001. - 3 : Vol. 2. - pp. 169-176.
- **Li S, Kwok J and Wang Y** Using the discrete wavelet frame transform to merge Landsat TM and SPOT panchromatic images [Journal] // Information Fusion. - 2002. - 1 : Vol. 3. - pp. 17-23.
- **Liu J [et al.]** Wavelet based active contour model for object tracking [Conference] // International Conference on Image Processing. - 2001. - Vol. 3. - pp. 206-209.
- **Lou R and Kay M** Multisensor Integration and Fusion for Intelligent Machines and Systems [Book]. - Norwood, E. U. : Ablex Publishing Corporation, 1995.
- **Mallat G** Multiresolution approximations and wavelet orthonormal bases of $L_2(\mathbb{R})$ [Journal] // Transactions of the American Mathematical Society. - 1989. - 1 : Vol. 315. - pp. 69-87.
- **Mangoli M** Apport de la fusion d'image satellitaires multicapteurs au niveau en télédétection et photo-interprétation [Book]. - Antipolis, Francia : Tesis doctorado. Univ. Nice-Sophia, 1994.
- **Muñoz J** Fusión de imágenes médicas a través de la transformada de Hermite [Book]. - Mexico : Tesis de Maestría. Universidad Nacional Autónoma de México, 2005.
- **Piella G** A general framework for multiresolution image fusion: from pixel to regions [Journal] // Information Fusion. - 2003. - 4 : Vol. 4. - pp. 259-280.
- **Pohl C and Genderen J** Multisensor image fusion in remote sensing: Concepts, methods and applications [Journal] // International Journal in Remote Sensing. - 1998. - 5 : Vol. 19. - pp. 823-854.
- **Pratt W** Digital Image Processing [Book]. - New York : John Wiley, 1991.
- **Rast M, Jaskolla M and Aranson F** Comparative digital analysis of Seasat-SAR and Landsat-TM data for Iceland [Journal] // International Journal of Remote Sensing. - 1991. - 3 : Vol. 12. - pp. 527-544.

- **Rockinger O** Pixel-level fusion of image sequence using wavelets frames [Conference] // Proceedings in Image Fusion and Shape Variability Techniques. - Leeds, Reino Unido : [s.n.], 1996.. - pp. 149-154.
- **Shensa M** The Discrete Wavelet Transform: Wedding the À Trouns and Mallat Algorithms [Article] // IEEE Transactions on Signal Processing. - 1992. - 10 : Vol. 40. - pp. 2464-2482.
- **Smith L** A tutorial on PCA [Informe]. - Nueva York, E. U. : Cornel University, 2003.
- **Tauch R and Kähler M** Improving the quality of satellite image maps by various processing techniques [Conference]. - 1988. - Vol. 4. - pp. 238-247.
- **Tu T [et al.]** A new look at IHS-like image fusion methods [Journal] // Information Fusion. - 2001. - 3 : Vol. 2. - pp. 177-186.
- **Tu T [et al.]** Adjustable intensity-hue-saturation and Brovey transform fusion technique for IKONOS/QuickBird imagery [Journal] // Optical Engineering. - 2005. - 11 : Vol. 44. - pp. 116201-116211.
- **Varma H [et al.]** Confusion in data fusion [Journal] // International Journal of Remote Sensing. - 2003. - 4 : Vol. 24. - pp. 1969-1976.
- **Wald L** Some terms of reference in data fusion [Journal] // IEEE Transactions on Geoscience and Remote Sensing. - 1999. - 3 : Vol. 37. - pp. 1190-1193.
- **Yésou H [et al.]** Merging Seasat and SPOT imagery for study of geologic structures in a temperate agricultural region [Journal] // Remote Sensing of Environment. - 1993. - Vol. 43. - pp. 265-280.
- **Zhang Z y Blum R** Multisensor image fusion using a region-based wavelet transform approach [Publicación periódica] // Internatinal Proceedings of the DARPA IUW. - 1997. - págs. 1447-1451.
- **Zhi-guo J [et al.]** A wavelet based algorithm for multi-focus micro-image fusion [Conference] // Third International Conference on Image and Graphics. - 2004. - pp. 176-179.
- **Zhijun W [et al.]** A comparative analysis of image fusion methods [Journal] // IEEE Transactions on Geoscience and Remote Sensing. - 2005. - 6 : Vol. 43. - pp. 1391- 1402.

4

Modelos de Calidad Perceptiva.

"I have never let my schooling interfere with my education"
Mark Twain

4.1 *Introducción.*

En los capítulos anteriores se ha hablado de la fusión de imágenes. Ahora nos enfocaremos en evaluar el desempeño de los algoritmos de fusión utilizando modelos de calidad perceptiva. Sin embargo, es necesario dedicar algunas líneas a los antecedentes y conceptos básicos de las métricas de calidad.

Las imágenes digitales pueden verse afectadas por una amplia variedad de distorsiones que darían lugar a una degradación en la calidad visual. Es por ello que la evaluación de la calidad en imágenes es un aspecto esencial dentro de las tareas de procesamiento. En general, para obtener una estimación precisa de la calidad de las imágenes se utilizan pruebas subjetivas. Sin embargo, estas son generalmente costosas y lentas. El objetivo es definir métricas cuantitativas que incorporen características del sistema visual humano (SVH) de manera que se pueda evaluar de forma automática la calidad de las imágenes.

Durante su adquisición, procesamiento, almacenamiento, transmisión y reproducción, las imágenes sufren degradación de la calidad visual. El único método realmente correcto para evaluar la calidad de las imágenes en aplicaciones donde las imágenes serán observadas por personas, es mediante la valoración subjetiva que daría el observador. Sin embargo, dicha valoración resulta bastante complicada y en algunos casos, según la aplicación, puede llegar a ser inabordable.

Una métrica objetiva de calidad puede cumplir diferentes funciones según su aplicación. Por ejemplo, se puede utilizar para monitorear y ajustar dinámicamente la calidad de la imagen de un

servidor de video. También puede servir para optimizar algoritmos y parámetros en sistemas de comunicación (p. ej. filtros y algoritmos de codificación y decodificación).

Las métricas de calidad objetiva que resultan efectivas son aquellas que evalúan las degradaciones en las imágenes. Para ello, es necesario saber qué es lo que el observador percibe, siendo indispensable conocer el funcionamiento del SVH y modelarlo en la medida de lo posible.

Los modelos del sistema visual humano son el fundamento para la definición de métricas objetivas de calidad perceptiva. Lo ideal es que las principales características del SVH estén integradas directamente en el corazón de la herramienta diseñada. Este planteamiento adquiere importancia, ya que las capacidades limitadas de procesamiento del SVH permiten que una imagen pueda ser modificada sin que las distorsiones introducidas sean detectadas.

En el caso de compresión, el grupo de expertos JPEG2000 comenzó trabajos para incorporar características del SVH como herramienta de optimización dentro de su estándar. El método habitual para mejorar la compresión se basa en transformar las amplitudes de la imagen a un dominio perceptivo uniforme, ya que la respuesta global del sistema visual a la escala de grises puede aproximarse mediante una no linealidad cúbica. Esta técnica es denominada corrección gamma que está presente en la mayoría de los monitores de video.

Los otros dos elementos de optimización son el contenido de las frecuencias espaciales y el enmascaramiento. El primero involucra el cálculo de la función de sensibilidad al contraste (CSF) y el segundo está relacionado con la importancia del contexto. El fenómeno del enmascaramiento (*masking*) es un efecto perceptivo que se debe al hecho de que los estímulos de gran intensidad ocultan a estímulos de menor intensidad.

En el contexto de las imágenes médicas, la medida final de la calidad de una imagen reside en lo útil que es la imagen para conseguir un diagnóstico certero y preciso. La verdadera prueba para un radiólogo es la detección precisa de incluso pequeñas anomalías.

Finalmente, no se debe dejar de lado el hecho que los modelos del SVH no tienen sentido si no se puede evaluar *a posteriori* el valor de la calidad percibida de la imagen. De esta manera, necesitaremos métricas de calidad objetiva perceptiva para evaluar si se han conseguido los objetivos deseados después del procesamiento de la imagen.

4.2 Estado del arte en las métricas de calidad perceptiva.

Uno de los principales problemas de la evaluación de la calidad en imágenes ha sido la ausencia de una métrica universalmente aceptada que funcione en un amplio rango de situaciones (Wang, et al., 2002). Las pruebas subjetivas se utilizan para obtener una estimación más precisa de la calidad de las imágenes. Sin embargo, en la práctica resultan muy pesadas en cuanto a la carga temporal que conllevan, además de resultar costosas y difíciles de repetir, puesto que requieren condiciones de visualización especiales (ITU, 2000).

Debido a estos inconvenientes, es necesario el uso de métricas de calidad objetivas. De estas, las más conocidas y que se siguen empleando en casi todos los casos, son medidas simples matemáticas basadas en la relación señal a ruido de pico (PSNR), el error cuadrático medio (MSE) o el error absoluto medio (MAE); que comparan una imagen distorsionada con otra de referencia (libre de distorsión). El uso de este tipo de métricas presenta varias ventajas, entre ellas está el hecho de que son sencillas de calcular, operan en el dominio espacial y punto a punto de la imagen. El valor obtenido no depende de las características específicas del proceso de medida (características del monitor, distancia de visualización, etc.), ni del observador. Sin embargo, esto es un problema, ya que al no tomar en cuenta la sensibilidad del sistema visual todas las distorsiones tienen la misma importancia y por ello no predicen adecuadamente lo que realmente apreciará el observador (Nava, et al., 2006).

Se han definido algunas métricas que incorporan de alguna manera factores perceptivos. El desarrollo de estas métricas objetivas basadas en modelos del SVH ha sido posible gracias a los avances en el conocimiento del SVH y su funcionamiento. En general lo que se suele modelar es el procesado a bajo nivel desde el ojo hasta la corteza visual primaria. Las propiedades del SVH o características visuales que incorporan normalmente estas métricas son la adaptación de luminancia (Peli, 1990), la función de sensibilidad al contraste (CSF) (Rust, 1997), (Peli, 2001) y el enmascaramiento (Gallego, 2006), (Nava, et al., 2006). Los modelos visuales se obtienen gracias a los datos experimentales conseguidos mediante la realización de pruebas psicofísicas de detección de umbral con estímulos simples.

Las métricas de calidad perceptiva comparten una estructura básica similar y dependiendo de las aplicaciones para las que son propuestas tendrán diversas particularidades. En general, en todas ellas se puede encontrar una etapa de calibrado, filtrado de la imagen a diferentes frecuencias espaciales y orientaciones, ajustes según la sensibilidad al contraste, mecanismos para modelar el enmascaramiento y una etapa final que obtiene un error como número o como un mapa de visibilidad de la distorsión o ambos, Figura 4-1. En (Eckert, et al., 1998) se presenta una clasificación detallada.

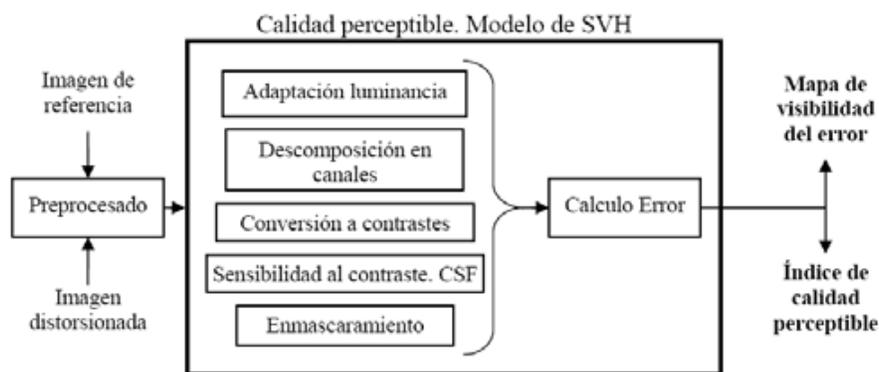


Figura 4-1 Esquema general para un modelo de calidad perceptible.

La implementación de estas métricas en la mayoría de casos es compleja y con un costo computacional elevado. Dos de los modelos de referencia para la mayor parte de las métricas que se han desarrollado son: “*Visible Difference Predictor*” VDP de Daly (Daly, 1993) y “*Visual Discrimination Model*” de Lubin (Lubin, 1993). Sin embargo, estos modelos presentan algunas limitaciones como son: no tener en cuenta características del SVH de alto nivel (extracción de características, procesos cognitivos, reconocimiento de patrones, atención visual, etc.).

En definitiva, es un área con distintos frentes de investigación abiertos, puesto que aún no hay una métrica estándar aceptada para la evaluación de la calidad percibida en las imágenes. Además, el estudio del comportamiento de las métricas ya existentes es complicado ya que la validación de las mismas se hace mediante pruebas subjetivas sobre un conjunto de imágenes dadas, con lo que en cualquier caso, su evaluación es parcial.

4.3 Propiedades de la visión.

La calidad de una imagen puede verse influenciada por diversos factores, entre los que destacan las técnicas de adquisición, de procesamiento y de visualización a las que se somete la imagen. En la mayoría de los casos el resultado obtenido será visto por un observador que formará parte activa en la evaluación de la imagen final. Por ello, para poder proporcionar una medida de calidad acorde con la percibida por ese observador, es necesario estudiar el SVH y los mecanismos de percepción (procesos de extracción de la información y la formación de representaciones) sobre los que pueda operar el sistema cognitivo.

En el estudio de la visión se pueden diferenciar dos etapas. Por un lado, se realiza un procesamiento a bajo nivel que no utiliza información de experiencias pasadas, ni razonamiento. Esto es lo que incluyen la mayor parte de las métricas de calidad basadas en las características del SVH y es en el que se centra este capítulo. Por otra parte, el procesamiento de alto nivel utiliza recursos como la atención y la memoria, que son mucho más complejos de modelar.

4.3.1 Adaptación visual. Sensibilidad a la intensidad luminosa.

El rango de niveles de intensidad luminosa que se presentan en la naturaleza y a los que el SVH se tiene que adaptar es muy amplio, alrededor de 10^{10} . Sin embargo, el SVH es capaz de discriminar sólo unos pocos niveles. La adaptación se realiza para un nivel de intensidad luminosa I_0 dentro de todos los niveles posibles del rango. Para este valor de I_0 , se pueden discriminar unos 50 valores distintos de intensidad luminosa. Por tanto, habrá un nivel por debajo del cual no se distingan otros niveles, que se verían como negro y de igual forma sucederá con la percepción del blanco.

A medida que el ojo recorre otros valores I_0 del rango total, define niveles de negro y blanco de forma análoga. La adaptación proporciona la capacidad de ver en condiciones distintas de iluminación. A pesar de ello, hay que tener en cuenta que la percepción visual varía mucho dependiendo de la iluminación con la que nos encontremos. En condiciones de baja iluminación, se detectan muy bien las diferencias de luminosidad, pero la distinción del color y del detalle es pobre.

Sucede lo contrario en condiciones de alta luminosidad. Además, el proceso de adaptación del SVH no es instantáneo, siendo más lenta la adaptación de luz a oscuridad.

En cuanto a la relación entre el estímulo percibido y el valor real del mismo se obtiene una respuesta no lineal. Para obtenerla, se considera una zona iluminada con una intensidad I y otra con $I + \Delta I$, Figura 4-2(a). Se mide el valor necesario de ΔI para que el observador vea dos zonas partiendo de un ΔI igual a cero. El valor obtenido de ΔI es se conoce como valor JND (*Just Noticeable Difference*). Este experimento se realiza sobre un amplio margen de intensidades (en visión fotópica, con valores sobre $10 \frac{cd}{m^2}$) dando como resultado la curva de la Figura 4-2(b). Se tiene una relación $\frac{\Delta I}{I}$ constante y de valor 0.02 que no depende del nivel de intensidad en el que se trabaja.

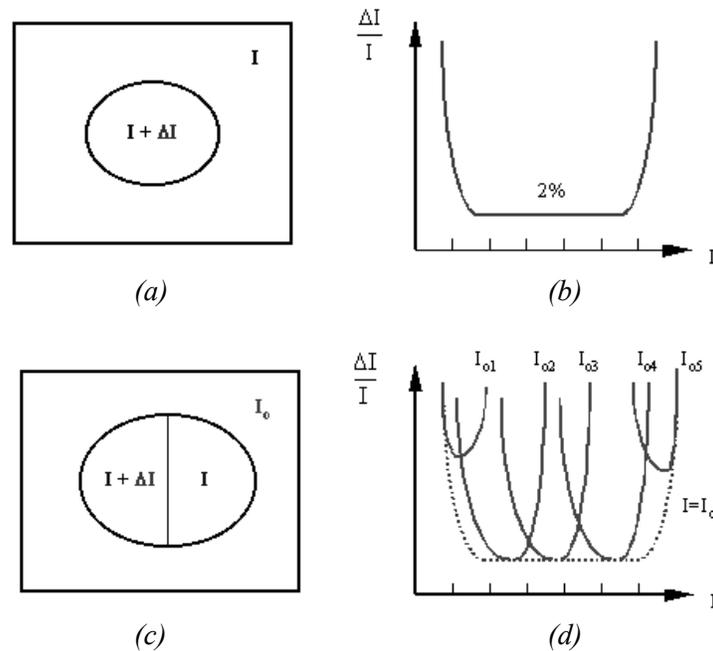


Figura 4-2 Medida de la respuesta del SVH a cambios de intensidad.

Si realizamos el mismo experimento para las condiciones indicadas en la Figura 4-2(c). Se obtienen curvas de respuesta que son constantes en un margen mucho menor, Figura 4-2(d).

A la relación $\frac{\Delta I}{I} = cte$ se le conoce como Ley de Weber y se verifica también para otros sentidos como el oído, el olfato, el gusto y el tacto. Este resultado nos interesa principalmente en el estudio de la percepción de ruido en imágenes. Basándonos en él, podemos explicar porque un nivel uniforme de ruido aleatorio es menos visible en las regiones más luminosas de la imagen.

Para que el ruido sea perceptivo con un fondo de imagen de nivel de intensidad luminosa I_1 será preciso un valor ΔI_1 mientras que para un $I_2 < I_1$ el valor de $\Delta I_2 < \Delta I_1$. Con un nivel inferior de ruido, este será más visible en la zona de luminosidad I_2 que en I_1 . En conclusión, para una imagen dada, se percibe más el ruido uniforme en las regiones oscuras. Para valores de luminancia menores a $10 \frac{cd}{m^2}$ la Ley de Weber deja de cumplirse y ΔI es proporcional a la raíz cuadrada de I que se conoce como relación de De Vries-Rose.

4.3.2 Contraste.

El avance en el conocimiento de la fisiología del SVH ha permitido saber que los mecanismos básicos de la visión trabajan con el contraste. En la Figura 4-3 se presentan experimentos psicofísicos que evidencian los fenómenos visuales relativos a la percepción de contraste. La Figura 4-3(a) es un ejemplo de contraste simultáneo, en el que se muestra un rectángulo del mismo nivel de intensidad, pero se percibe más claro u oscuro según el nivel de luminosidad del rectángulo exterior.

La Figura 4-3(b) son las Bandas de Mach que muestran el fenómeno de inhibición lateral producido en los campos receptores. A pesar de que cada barra representa un nivel constante, se aprecian diferencias subjetivas en los bordes de separación de las bandas.

Cómo el SVH es sensible al contraste y no a la luminancia en términos absolutos, es importante la definición que se haga del contraste. El objetivo es encontrar una medida de contraste que corresponda con la percepción visual que se tiene del mismo.



Figura 4-3 Fenómenos relativos al procesamiento de contraste. (a) Contraste simultáneo. (b) Bandas de Mach.

Una primera aproximación a la solución del cálculo de contraste es utilizar imágenes simples (patrones) tales como funciones sinusoidales o discos de luminancia. Michelson (Michelson, 1927) definió el contraste de este tipo de imágenes como:

$$C_{Michelson} = \frac{\Delta L}{L + \Delta L} \quad \dots (4.1)$$

donde $\Delta L = \frac{(L_{max} - L_{min})}{2}$, L_{max} y L_{min} son los valores máximos y mínimos de luminancia, y $L = L_{min}$.

Peli (Peli, 1990) propone el cálculo del contraste por bandas de frecuencia. Este es un contraste local y limitado en banda, que tiene en cuenta el nivel medio de luminancia local para la obtención de un valor en cada punto. Para calcular el contraste de Peli se necesita una versión filtrada de la imagen, ya que el contraste se calcula por bandas. El contraste en una banda frecuencial k está dado por:

$$C_{pelik}(i, j) = \frac{b_k(i, j)}{l(i, j)} \quad \dots (4.2)$$

donde $b_k(i, j)$ es la imagen filtrada para una determinada banda k y $l(i, j)$ es una versión paso-bajas de la imagen que contiene toda la energía por debajo de la banda k en la que se está realizando el cálculo del contraste. De esta forma tenemos un valor de contraste que no sólo depende de la energía en esa banda, sino también de la luminancia de fondo local.

Otras posibles métricas de contraste para imágenes complejas realizan una modificación del contraste de Michelson adaptándolo al filtrado previo de la imagen como se muestra en la ecuación (4.3)

$$C_k(i, j) = \frac{b_k(i, j) - \bar{b}_k}{\bar{b}_k} \quad \dots (4.3)$$

donde $b_k(i, j)$ es la imagen filtrada para una banda k y \bar{b}_k es la media de esa banda. El problema de esta definición del contraste es que la media es cero para todas las bandas de frecuencia excepto para la banda base. Por ello se deben hacer modificaciones. Daly (Daly, 1993) propone un cálculo global del contraste y otro local.

$$C_{Global_k}(i, j) = \frac{b_k(i, j)}{\bar{b}_k} \quad C_{Local_k}(i, j) = \frac{b_k(i, j)}{b_K(i, j)} \quad \dots (4.4)$$

donde $b_K(i, j)$ es la banda base y \bar{b}_k es la media de la misma banda.

El contraste en un píxel de una imagen también puede calcularse en función de la luminancia de los píxeles que hay en su vecindario. Esta técnica (Taylor, 1998) considera los valores máximos y mínimos de luminancia en los píxeles vecinos a la izquierda y derecha del píxel para el que se calcula el contraste. Tanto en este caso, como en todos los anteriores, se debe tener en cuenta que la

definición que se hace del contraste debe ser consecuente con la implementación del modelo visual que se esté llevando a cabo, así como con el objetivo que va a tener este modelado.

Una vez expuestos los métodos para el cálculo de contrastes, queda saber qué método tiene una mayor correspondencia con la percepción del contraste del SVH. Para esto, es necesario el uso de pruebas psicofísicas que evalúen esta relación. Un estudio comparativo se encuentra en (Peli, 1997), los patrones utilizados son Gabores y el resultado de las pruebas realizadas mantiene que la fórmula de Peli es la mejor.

4.3.3 Sensibilidad en frecuencia.

Una de las caracterizaciones más importantes que se hacen del SVH consiste en la máxima frecuencia capaz de detectar o diferenciar. Esta capacidad se denomina agudeza visual.

La agudeza visual proporciona únicamente un límite y es insuficiente si se quiere conocer la variación en la detección de un estímulo en función de su frecuencia espacial. Con este objetivo se empezó a medir la función de sensibilidad al contraste, *contrast sensitivity function* (CSF) que refleja la sensibilidad o capacidad de detección del SVH a estímulos de distinta frecuencia.

El procedimiento para medir la CSF debe consistir en mantener un estímulo de contraste constante y variarlo en frecuencia. Sin embargo, esto es imposible ya que no se puede medir el contraste de la imagen percibida, por lo que el procedimiento en la práctica es el contrario. El método se basa en reducir el contraste de un estímulo sinusoidal en la que la luminancia media se mantiene constante hasta que se alcanza el umbral (el estímulo deja de ser visible). El inverso de este valor umbral es el valor de la sensibilidad para esa frecuencia. Al realizar este proceso para las diferentes frecuencias obtenemos la CSF.

Algunas consideraciones sobre la CSF son: método válido sólo para estímulos simples. Aunque cualquier estímulo complejo pueda ser analizado como una serie de estímulos simples, el SVH es no lineal y por ello, la respuesta varía y no puede ser calculada como una combinación lineal. Además, la CSF no es válida para condiciones por encima del umbral.

En (Nadenau, 2000) se presenta un resumen de algunas CSF y los parámetros que se utilizaron la obtención de la función. La forma de típica de la CSF se muestra en la Figura 4-4. Es un filtro paso banda con un pico que se sitúa entre los 4 y 8 ciclos/grado para niveles de iluminación fotópicos.

Se produce una fuerte atenuación en altas frecuencias debida a factores ópticos como imperfecciones en la córnea. Además, el muestreo espacial que realizan los fotorreceptores de la retina impone un límite a partir del cual no se detectan frecuencias espaciales mayores (agudeza visual). El valor de la CSF en frecuencia cero es imposible de obtener para el método de medida que se ha explicado.

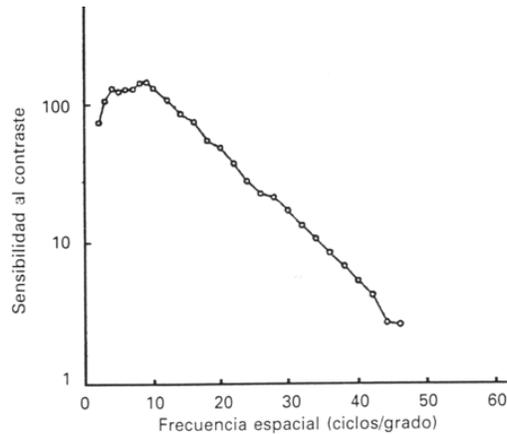


Figura 4-4 CSF del sistema visual (Campbell & Green, 1965).

4.3.4 Enmascaramiento.

El enmascaramiento es un fenómeno que hace que la visibilidad de una señal en presencia de otra disminuya o quede oculta. El enmascaramiento es un paso más de complejidad en el modelado del SVH. El objetivo es conocer como varían los umbrales de visibilidad del estímulo cuando se encuentran en el entorno cambios espaciales y también temporales de luminancia. Es entonces que aparecen dos fenómenos de enmascaramiento:

- *Enmascaramiento temporal.*

Está asociado a la reducción visual que se produce sobre el estímulo, cuando está rodeado de cambios temporales de luminancia. Su estudio es complicado ya que toma en cuenta las características de movimiento del ojo; hay una dependencia con la velocidad de movimiento de la retina. Por otro lado, está demostrado que cuando hay un cambio de escena, la resolución espacial puede ser reducida drásticamente sin que se llegue a percibir, siempre que se restablezca la resolución original en un periodo breve de tiempo (en torno a 100 ms).

- *Enmascaramiento espacial.*

Está asociado a los cambios espaciales de luminancia que se producen alrededor del estímulo. Recibe también el nombre de enmascaramiento de contraste, ya que el efecto del enmascaramiento es menor en regiones uniformes que en zonas de gran contraste. Si consideramos una imagen natural a la que se le añade ruido aleatorio uniforme, la visibilidad del ruido en los bordes y texturas de la imagen será menor que en zonas uniformes.

El comportamiento se muestra en la Figura 4-5. C_T representa el valor de contraste para que el estímulo sea visible y C_M el contraste de la señal que enmascara.

Para un valor pequeño de C_M , el umbral de visibilidad del estímulo C_T viene dado por la CSF y se representa como C_0 . Este valor se mantiene hasta que se alcanza el punto en que $C_M = C_0$. A partir de ahí un incremento en C_M hace que se eleve el umbral de detección, es decir, se necesitará aumentar el contraste del estímulo para que sea visible. La línea discontinua de la figura ha sido observada en algunos de los experimentos psicofísicos. Sin embargo, este fenómeno sólo se observa cuando la señal y el ruido están en fase.

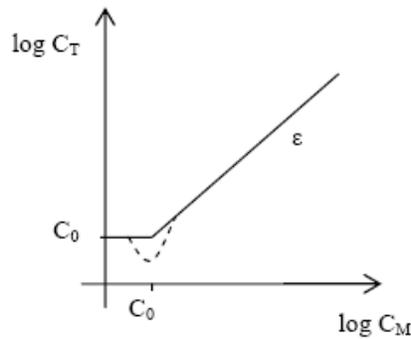


Figura 4-5 Curva de umbrales de elevación causados por enmascaramiento espacial.

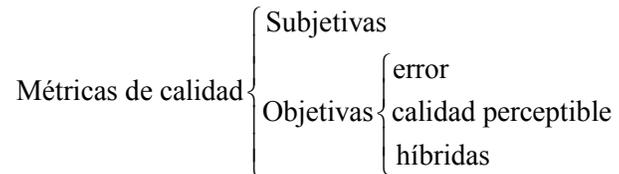
El valor de la pendiente ε depende de la complejidad que se tenga del estímulo y del tipo de señal utilizada para el enmascaramiento. El valor de ε varía entre 0.6 y 1 dependiendo del tipo de señal utilizada para el enmascaramiento. Además, aparece un efecto de aprendizaje. Si a un observador se le muestra el mismo patrón de ruido repetidas veces, el valor de ε decae, ya que el observador será capaz de reconocerlo con un valor menor de contraste.

4.4 Clasificación de las métricas de calidad.

Antes de proporcionar una clasificación de las métricas de calidad, habrá que hacer un comentario sobre la definición de calidad de la imagen. Se define la calidad de una imagen como: “la capacidad que tiene una imagen de representar el objeto original”. En las métricas de calidad, en vez de tener una imagen y un objeto, la definición se extiende a dos imágenes, donde una se considera la imagen de referencia u original y la otra es la imagen con alguna distorsión. En este caso, la medida de calidad es una relación de semejanza entre la imagen original y la distorsionada.

La definición de calidad de imagen arriba mencionada se conoce como fidelidad de la imagen y también tiene algunos inconvenientes. No es claro que la visibilidad del error esté relacionada con la pérdida de calidad. De hecho, algunas distorsiones que pueden ser claramente visibles, no son a pesar de ello, molestas para el observador.

En cuanto a las métricas de calidad, se diferencian dos grupos principales: las métricas de calidad subjetivas (utilizando observadores) y las métricas de calidad objetiva (medidas matemáticas), estas a su vez, se dividen dependiendo si incorporan o no el SVH.



Las *métricas de error* obtienen una medida de calidad sin tener en cuenta las características del SVH y están orientadas a medir la diferencia entre imágenes. Las métricas de calidad perceptiva son aquellas que incorporan las características del SVH. Finalmente, existe un tercer grupo de métricas híbridas que a pesar de no incluir un modelo de SVH, tratan de obtener valores que se adecuen a la calidad que el observador percibe explotando otras cualidades relativas a la percepción.

4.5 Métricas de calidad objetivas.

En las siguientes secciones proporcionaré una descripción detallada de las métricas de calidad objetivas y su implementación en el *toolbox* para MATLAB *Perceptual Models* que acompaña esta tesis. Sobre las métricas de calidad subjetiva se puede encontrar una explicación amplia en (Gallego, 2006).

Las métricas de calidad objetiva proporcionan un valor de la calidad de una imagen utilizando operaciones matemáticas en las que eventualmente pueden incluirse funcionalidades del SVH. Se pueden utilizar para la optimización de algoritmos en procesamiento de imágenes, siendo una de sus principales ventajas frente a los métodos subjetivos el que proporcionan una medida casi inmediata y por ello sirven para monitorizar de forma dinámica la calidad de la imagen.

4.5.1 Métricas de calidad basadas en el error.

Las métricas basadas en el error no incorporan características del SVH. Se basan en funciones matemáticas simples y generalmente trabajan en el dominio espacial realizando operaciones punto a punto en la imagen. Estos métodos obtienen un valor de calidad en términos de variaciones entre la imagen procesada y la original.

Los más utilizados son el error cuadrático medio MSE (*Mean Square Error*) (4.5) y la relación señal a ruido de pico PSNR (*Peak Signal to Noise Ratio*) (4.6).

$$MSE = \frac{1}{MN} \sum_{i=0}^{M-1} \sum_{j=0}^{N-1} (I(i, j) - \hat{I}(i, j))^2 \quad \dots (4.5)$$

$$PSNR = 10 \log \frac{(2^b - 1)^2}{MSE} \text{ dB} \quad \dots (4.6)$$

donde N y M son el número de columnas y filas de la imagen, respectivamente, $I(i, j)$ es el valor de intensidad del pixel en la posición (i, j) en la imagen original, $\hat{I}(i, j)$ es el valor del pixel en el mismo punto en la imagen procesada y b es el número de bits por pixel.

Una modificación a MSE para reducir la sensibilidad que presenta a cambios globales en la intensidad es el NMSE (*Normalized Mean Square Error*), que normaliza las intensidades haciendo que las imágenes tengan media cero y variancia uno. Después se aplica MSE.

Las ventajas de este tipo de métricas son su simplicidad y velocidad. Sin embargo, al no depender de las características del proceso que evalúa, trata todas las distorsiones de la misma forma dándoles la misma importancia, por lo que no predicen de forma correcta la calidad que realmente es apreciada por el observador.

Cuando se utilizan sobre diversas escenas o sobre una misma pero con distintos tipos de distorsión, se pone de manifiesto el hecho de que para un mismo valor de PSNR o MSE, se observan calidades diferentes, Figura 4-6.

Las cuatro imágenes que se muestran están comprimidas con JPEG y todas tienen un valor aproximado de PSNR de 26 dB, sin que ello implique que las calidades percibidas sean iguales. De hecho, se aprecian calidades distintas, siendo mejor la imagen de la Figura 4-6(d) y la peor la imagen de la Figura 4-6(b).

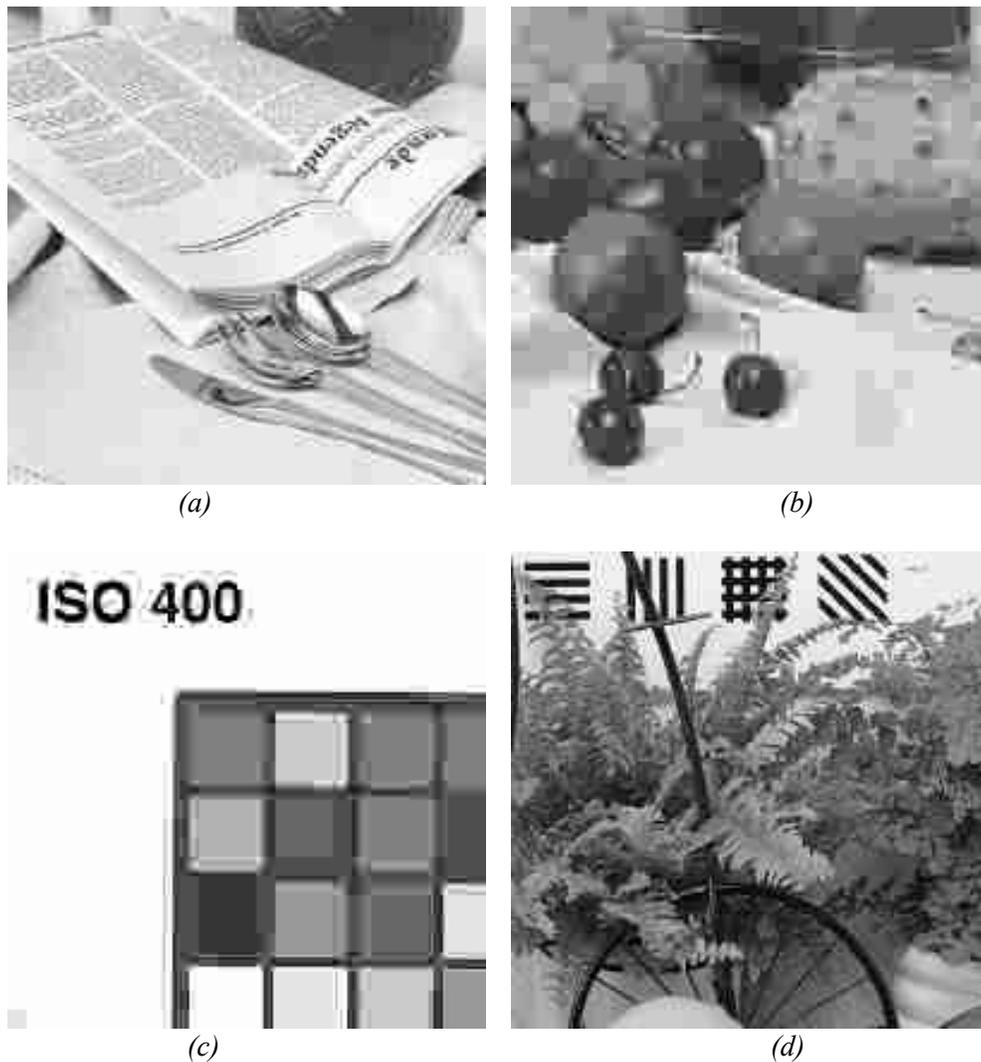


Figura 4-6 Imágenes con PNSR = 26 dB.

4.5.2 Métricas de calidad perceptiva.

Las métricas de calidad perceptiva pueden proporcionar valores numéricos que cuantifican la “satisfacción” del observador con la imagen procesada y también presentar mapas de visibilidad del error.

La principal ventaja de este tipo de métricas consiste en que al incorporar de alguna forma las características del SVH pueden proporcionar resultados reales de la calidad que se percibe. La incorporación de estas características hace que sean modelos más complejos. De la *Figura 4-1*, se

puede distinguir que estas métricas comparten un esquema común que varía dependiendo de las aplicaciones en las que se empleen.

- *Preprocesado.*

La etapa de preprocesado no forma parte en sentido estricto del modelo de calidad, pero es importante porque se encarga de escalar y alinear las imágenes de entrada.

Otra funcionalidad correspondiente a esta etapa es la conversión de valores de luminancia almacenados en memoria a valores digitales con los que se visualiza en pantalla. Esta conversión en monitores CRT se corresponde con una transformación no lineal que se conoce como función gamma. Esto implica una calibración del monitor y obtener la relación existente entre la entrada digital de niveles de gris y la salida de valores de luminancia en $\frac{cd}{m^2}$. En general, se requiere una caracterización precisa del monitor dentro de cualquier modelo de calidad, ya que el SVH es capaz de ver sólo aquello que sea reproducible en la pantalla

Por último, ya que las métricas están basadas en comparaciones, se necesita una correspondencia punto a punto para que sean consistentes (Gottesfeld, 1992).

- *Algunas limitaciones.*

Estas métricas son un avance importante en la evaluación de la calidad de imágenes, pero aun presentan algunos puntos débiles. La calidad perceptiva, en todos ellos, se estima mediante una cuantificación de la visibilidad del error, que se consigue introduciendo en la métrica las funcionalidades de las primeras etapas del SVH. Sin embargo, la complejidad del SVH y su no linealidad no permiten obviar que el modelado del SVH usando operadores lineales o casi lineales es bastante simplista. Se asume que la calidad percibida se determina en las primeras etapas del sistema visual.

Otra limitación en este tipo de modelos está relacionada con los experimentos psicofísicos de los que se extraen las curvas de la CSF y de los fenómenos de enmascaramiento. Estas curvas se obtienen con estímulos simples como puntos, barras o señales sinusoidales que no se corresponden con la variedad del contenido presente en las imágenes reales evaluadas.

La estructura básica que comparten estos modelos incluye, la suposición de que el valor global de la distorsión percibida aumenta de forma monótona con la suma del error para todos los canales y en todas las posiciones. Por tanto, a la hora de conseguir un mapa de visibilidad del error o un índice global se utiliza la métrica de Minkowski, (Pratt, 1991), que tiene la forma típica:

$$E = \left(\sum_k (e_k)^\beta \right)^{\frac{1}{\beta}} \quad \dots (4.7)$$

donde e_k es el valor del error en el canal k . El uso de este tipo de métrica para el cálculo final de los errores tiene inconvenientes, ya que se asume que los errores en diferentes localizaciones son independientes.

4.5.3 Métricas de calidad híbridas.

Este grupo de métricas proporcionan un valor que se adecua a la imagen percibida sin incorporar el modelo de visión humano. Primero, consideraré un modelo de calidad que se basa en la información estructural de la imagen para obtener una evaluación de la calidad. Después, revisaré algunas ventajas que presenta la distribución de Wigner y como se puede obtener una medida de calidad a partir de ella.

- *Medida de calidad basada en la información estructural.*

Es un modelo propuesto por Wang y Bovik (Wang, et al., 2002), (Wang, et al., 2004), proporciona un valor de calidad en el rango [0,1] al que denominan índice de similitud estructural SSIM (*Structural Similarity*). Se basa en la idea de que el SVH extrae información de la estructura de los objetos del campo visual e incorpora este aspecto a la métrica.

El valor del índice de similitud estructural dependerá de tres factores, dos de los cuales, luminancia y contraste, no representan ningún aspecto nuevo en métricas de calidad. Primero se realiza un cálculo de luminancia que consiste en una media matemática, seguido del cálculo de contraste usando la desviación estándar y finalmente los atributos que representan a la estructura de los objetos en escena se modelan como una correlación, Figura 4-7.

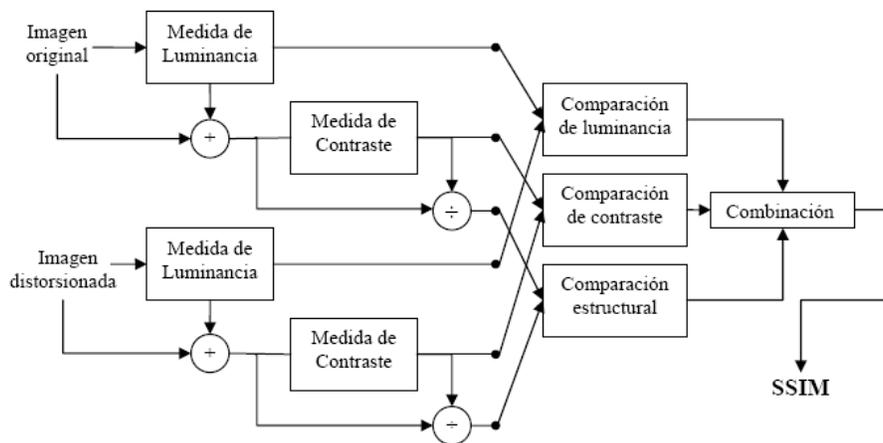


Figura 4-7 Esquema del sistema de medida de SSIM.

El valor de SSIM se modela con la ecuación (4.8)

$$SSIM(I, \hat{I}) = \frac{\sigma_{\hat{I}}}{\sigma_I \sigma_{\hat{I}}} \frac{2\mu_I \mu_{\hat{I}}}{(\mu_I)^2 + (\mu_{\hat{I}})^2} \frac{2\sigma_I \sigma_{\hat{I}}}{\sigma_I^2 \sigma_{\hat{I}}^2} \quad \dots (4.8)$$

donde μ_I es la media de la imagen original y $\mu_{\hat{I}}$ la media de la imagen distorsionada; σ es la varianza y $\sigma_{\hat{I}}$ la covarianza.

En la ecuación anterior, el primer término corresponde a la comparación estructural (correlación entre la imagen original I y la imagen distorsionada \hat{I}), el segundo término es la distorsión de la luminancia y el tercero evalúa la diferencia de contraste.

Una de las ventajas de este método es su sencillez, portabilidad y poco costo computacional. Además proporciona un índice y un mapa de la variación espacial de la calidad.

- *Medida de calidad basada en la distribución Wigner.*

La distribución de Wigner-Ville (WVD) se ha utilizado con éxito en aplicaciones con señales no estacionarias, así como para analizar las características espacio-frecuenciales. En (Beghdadi, et al., 2006) se propuso una medida de calidad basada en la WVD que permite modelar las primeras etapas del procesamiento visual

Para incluirla la distribución en una métrica de calidad se debe obtener la WVD discreta, para el caso de las imágenes, se convierte en una transformación en 4D que incluye una transformada de Fourier en cada punto de la imagen, por lo que esta métrica es costosa desde el punto de vista computacional. Ejemplos de aplicaciones de esta distribución se dan en (Hormigon, et al., 2003) y (Reed, et al., 1991) donde se emplea para segmentación de texturas.

La idea de utilizar la WVD en una métrica de calidad consiste en aprovechar la representación conjunta espacio-frecuencia teniendo presente que en las primeras etapas del SVH se crea el mismo tipo de representación. Una de las propiedades más importantes de la distribución de Wigner-Ville para su aplicación sobre imágenes es que es una función real y al mismo tiempo contiene la información de fase. La distribución continua 2D de una imagen $I(x, y)$ se define como:

$$W_f(x, y, f_x, f_y) = \int_{\mathbb{R}^2} f\left(x + \frac{\alpha}{2}, y + \frac{\beta}{2}\right) f^*\left(x - \frac{\alpha}{2}, y - \frac{\beta}{2}\right) e^{-j2\pi(\alpha f_x + \beta f_y)} d\alpha d\beta \quad \dots (4.9)$$

donde x e y son las coordenadas espaciales y f_x y f_y las frecuencias espaciales. Esta función localiza las estructuras en espacio y frecuencia. Sin embargo, la ecuación (4.9) no es útil, ya que es necesario calcular la WVD discreta. Esto presenta dos problemas, el *aliasing* y la dispersión espectral por el uso de ventanas espaciales.

La implementación discreta de la WVD se denomina *pseudo* WVD. No se entrará en detalles del análisis de la PWDV, sólo se mostrará su uso para el cálculo de una SNR de Wigner.

$$SNR_W = 10 \log_{10} \frac{\sum_i \sum_j \max_{u,v} (|PWVD_I(i, j, u, v)|)}{\sum_i \sum_j \max_{u,v} (|PWVD_I(i, j, u, v) - PWVD_{\hat{I}}(i, j, u, v)|)} \dots (4.10)$$

Se calcula la PWVD para la imagen original I y para la imagen distorsionada \hat{I} , donde (i, j) son las coordenadas espaciales y (u, v) son las frecuencias espaciales.

En cada posición (i, j) se guarda la componente de máxima energía de la PWVD. Se trata de simular una especie de enmascaramiento al suponer que esta componente máxima oculta las contribuciones de los demás componentes.

La medida de SNR_W presentada da resultados más precisos (mejores en cuanto a la relación entre este valor y lo que percibe el observador) que los obtenidos mediante medidas como la SNR. Esto se debe a que las distorsiones que presenta una estructura clara son mucho más molestas para el SVH que las que no la poseen, Figura 4-8.

Finalmente, el uso de la WVD presenta la ventaja principal que es una representación conjunta espacio y frecuencia; posee buena localización de las estructuras. Sin embargo, el precio que se paga por ello es elevado al incrementar la dimensionalidad hasta 4D.

El cálculo es costoso en tiempo y las aplicaciones que pueden darse a esta medida serían limitadas. Además, el valor SNR_W es un índice global y sería interesante poder disponer de mapas de visibilidad del error. Por todo ello, si se utilizan este tipo de distribuciones convendría poder explotar mejor sus propiedades e incluirlas en un modelo más complejo de sistema visual humano incluyendo el enmascaramiento, etc.

4.6.1 Adaptación de la luminancia.

Basados en la ley de Weber-Fechner que establece que las diferencias de luminancia en un estímulo son proporcionales a la luminancia media del mismo, las métricas deben reflejar que se puede tolerar un mayor nivel de distorsión si la luminancia local aumenta.

En la herramienta para MATLAB, se implementaron tres curvas de adaptación de luminancia en el dominio espacial y en las que el nivel de adaptación para un píxel de la imagen depende únicamente de ese píxel. En la Figura 4-9 se muestran las opciones propuestas para el modelo de luminancia: la función lineal, la función raíz cuadrada, la función logaritmo y una función propuesta por Daly (Daly, 1993), la curva se calcula con la ecuación (4.11).

$$I_l(x, y) = \frac{I(x, y)}{I(x, y) + 12.6I(x, y)^{0.63}} \quad \dots (4.11)$$

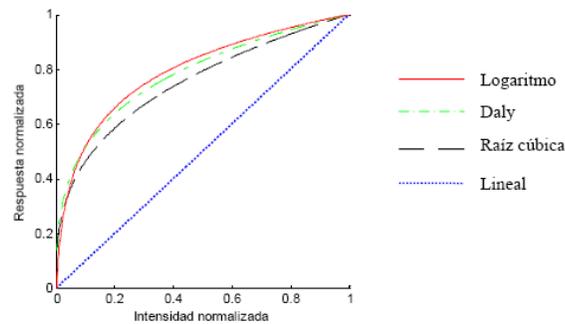


Figura 4-9 Curvas de adaptación de luminancia.

Algunos autores (Daly, 1993) afirman que la función logarítmica sobrestima la sensibilidad visual en regiones de baja intensidad por lo que es mucho mejor la función cúbica. En general, la adaptación de luminancia no tiene gran impacto en las métricas, a menos que las imágenes tengan mucho contraste.

4.6.2 Función de sensibilidad al contraste (CSF).

La CSF es una pieza clave para el funcionamiento de las métricas de calidad ya que describe las variaciones de la sensibilidad visual en función de la frecuencia espacial. En la práctica, la CSF depende de muchos parámetros y es bastante compleja de obtener, por lo que se implementaron modelos tres modelos ya propuestos:

- ❑ CSF de Rust (Rust, 1997).
- ❑ CSF de Daly (Daly, 1993).
- ❑ CSF de Barten (Eriksson, et al., 1998)

Para estos modelos se consideran dos esquemas, el primero utiliza a la CSF como un filtro, una normalización en frecuencia; el segundo utiliza a la CSF como una función de peso. Ambos esquemas se explicarán páginas adelante.

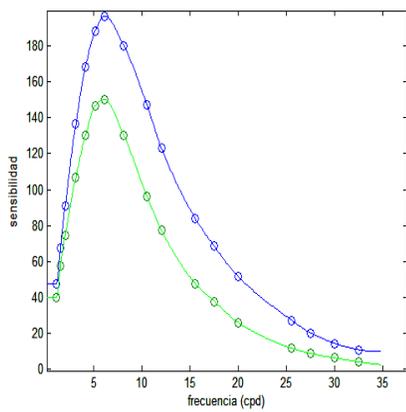
- *CSF de Rust*

Rust toma medidas de la sensibilidad al contraste para 17 valores de frecuencia espacial de 1.1 a 32.5 ciclos por grado (cpd) en dos orientaciones (45° y 90°). Estas medidas permiten modelar el carácter anisotrópico de la CSF que se observa en la Figura 4-10. Rust utiliza una *spline* cúbica para ajustar los valores medios (4.12), donde S_i son los valores medios para cada frecuencia ρ_i y poder expresar la CSF como la ecuación (4.13).

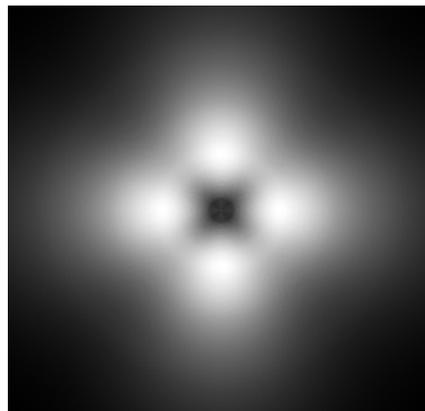
$$A(\rho_i) = \frac{1}{2} [S_{90^\circ}(\rho_i) - S_{45^\circ}(\rho_i)] \quad \dots (4.12)$$

$$B(\rho_i) = \frac{1}{2} [S_{90^\circ}(\rho_i) + S_{45^\circ}(\rho_i)]$$

$$CSF(\rho, \Theta) = B(\rho) + A(\rho) \cos(4\Theta) \quad \dots (4.13)$$



(a)



(b)

Figura 4-10 CSF de Rust. (a) CSF para $\Theta = 45^\circ$ y 90° en 1D. (b) CSF 2D.

Uno de los problemas de la CSF de Rust es la ambigüedad en el origen, ya que no se dispone de un valor para la frecuencia cero, por lo que se decidió que los valores medidos fijen la CSF.

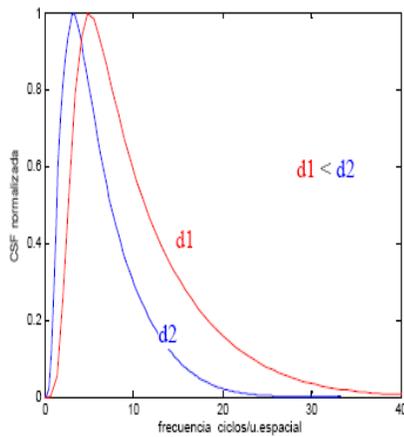
- *CSF de Daly.*

Daly proporciona un modelo analítico para la CSF, ecuación (4.14). La ventaja fundamental del modelo analítico es que permite una mayor flexibilidad y es posible estudiar la forma de las curvas. La CSF depende de la frecuencia ρ en (cpd), la orientación Θ en (grados), la distancia de visualización d en (metros), el tamaño de la imagen i^2 en (grados visuales), el nivel de adaptación de luminancia l en ($\frac{cd}{m^2}$) y la excentricidad e en (grados)

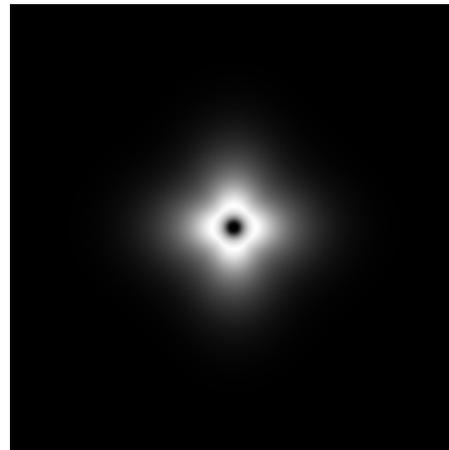
$$CSF(\rho, \Theta, l, i^2, d, e) = P * \min \left[S_1 \left(\frac{\rho}{r_a \cdot r_e \cdot r_\Theta}, l, i^2 \right), S_1(\rho, l, i^2) \right] \dots (4.14)$$

$$S_1(\rho, l, i^2) = \left(\left(3.23(\rho^2 \cdot i^2)^{-0.3} \right)^5 + 1 \right)^{-\frac{1}{5}} \cdot A_l 0.9 \rho \cdot e^{-(B_l 0.9 \rho)} \cdot \sqrt{1 + 0.06 e^{(B_l 0.9 \rho)}} \dots (4.15)$$

Donde r_a , r_e y r_Θ caracterizan la CSF en función de la distancia, la excentricidad y orientación respectivamente. La variación con l se determina por A_l y B_l . En la Figura 4-11(a) se muestra la curva para $\Theta = 0^\circ$ y dos distancias de visualización.



(a)



(b)

Figura 4-11 CSF de Daly. (a) CSF en 1D con $\Theta = 0^\circ$ y 2 valores de d . (b) CSF en 2D.

Una diferencia con esta función de CSF es que si se utiliza como un filtro se eliminan las bajas frecuencias, ya que $CSF(0, \Theta, l, i^2, d, e) = 0$. La función es anisotrópica. En general, esta función es menos suave que la de Rust, por lo que elimina mayor cantidad de frecuencias.

- *CSF de Barten.*

Se incluye un tercer modelo de CSF que aparece en (Daly, 1993) y (Westen, et al., 1995). Barten obtuvo la siguiente fórmula analítica basándose en datos experimentales:

$$CSF(\rho, l, i^2, d) = a \cdot \rho \cdot e^{-b\rho} \sqrt{1 + ce^{b\rho}} \quad \dots (4.16)$$

donde

$$a = \frac{540 \left(1 + \frac{0.7}{l}\right)^{-0.2}}{1 + \frac{12}{i \left(1 + \frac{\rho}{3}\right)^2}}, \quad b = 0.3 \left(1 + \frac{100}{l}\right)^{0.15}, \quad c = 0.06 \quad \dots (4.17)$$

Una diferencia de esta CSF con las anteriores es que no aparece la función Θ , ya que es isotrópica. Otra consideración es el valor pico de esta función, para los mismos parámetros que el modelo de Daly, el pico de sensibilidad es mayor.

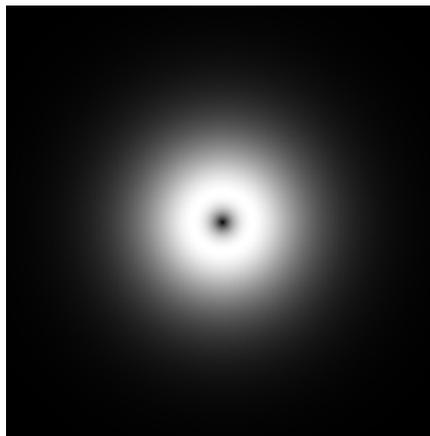


Figura 4-12 CSF de Barten.

4.6.3 Descomposición en canales.

El estudio del SVH ha mostrado que la representación de la señal visual se lleva a cabo mediante una serie de células que están sintonizadas en canales de frecuencia espaciales y con un rango de orientaciones determinadas. De este hecho se infiere que la forma más adecuada para el tratamiento de imágenes dentro de un modelo de SVH es el análisis con múltiples escalas y orientaciones.

Para la implementación presentada en esta tesis se consideran dos opciones de descomposición multicanal:

- Banco de filtros con canales diseñados para un determinado rango de frecuencias espaciales y orientaciones: filtros de Gabor (Aznar, 2000), (Fischer, 2000) y la transformada Cortex (Westen, et al., 1995).
- Banco de filtros con canales selectivos en frecuencia únicamente: filtros anillo (Westen, et al., 1995)

Todas las implementaciones realizan una descomposición en bandas en el dominio de frecuencia como una multiplicación de la imagen de entrada por cada uno de los canales. En el siguiente apartado se revisarán los filtros de Gabor; el desarrollo del resto de filtros pueden revisarse en (Gallego, 2006) ya que su estudio y desarrollo quedan fuera del tema de esta tesis.

- *Filtros de Gabor.*

Gabor demostró que puede representar una señal mediante paquetes elementales de información que minimizan la relación de incertidumbre de Heisenberg (Gabor, 1946). En este contexto, los filtros 2D dan la máxima resolución en los dos dominios. Otra característica de estos paquetes o *logons* es que son adecuados en la descripción del procesamiento del sistema visual. Se ha demostrado que las funciones de Gabor pueden modelar el comportamiento de las células simples de la corteza visual. Los filtros de Gabor 2D en el dominio espacial y de frecuencia tienen la siguiente forma:

$$\begin{aligned}
 G(X) &= e^{-\pi(\alpha(X-X_0))^2} e^{-2\pi i U_0(X-X_0)} \\
 G(U) &= e^{-\pi\left(\frac{1}{\alpha}(U-U_0)\right)^2} e^{-2\pi i X_0(U-U_0)}
 \end{aligned}
 \quad \dots (4.18)$$

donde $X = (x, y)$, $U = (u, v)$, X_0 , U_0 , α y β son vectores en \mathbb{R}^2 . X_0 representa las coordenadas en el dominio espacial del centro de la envolvente Gaussiana, U_0 las coordenadas cartesianas de las frecuencias espaciales de sintonía de los filtros, α es el vector determina la anchura de banda radial y la forma de la envoltura Gaussiana.

En la Figura 4.13 se muestran los filtros de Gabor para 1D y 2D. En la implementación se realizaron ajustes de las varianzas para que los filtros se traslapen adecuadamente. Además, se modificaron los Gabores para que tengan respuesta nula en frecuencia cero.

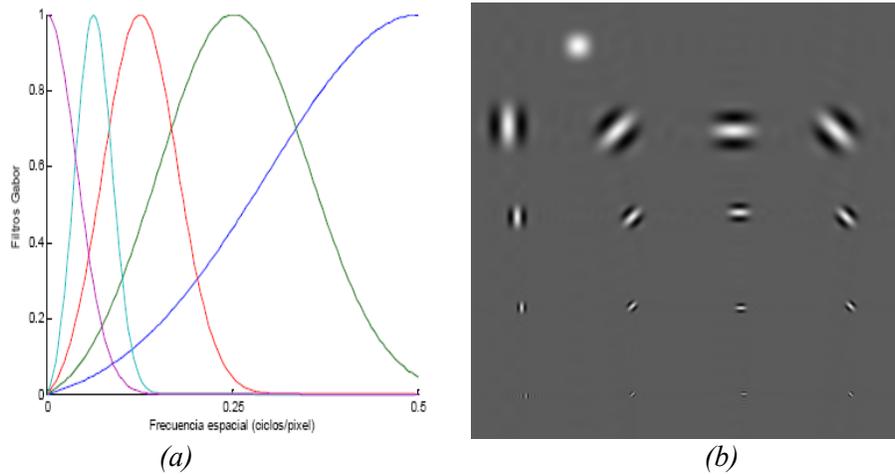


Figura 4-13 Filtros de Gabor para 4 escalas y 4 orientaciones más el paso bajo. (a) En 1D en el dominio de frecuencia. (b) En 2D en el dominio espacial, parte real. El residuo paso bajo aparece representado en la primera fila de la presente figura.

En la práctica, el análisis o descomposición de la imagen de entrada mediante el banco de filtros se hace en el dominio frecuencial como una multiplicación de los filtros de Gabor por la imagen transformada. En la Figura 4.14 se muestra el procedimiento de descomposición para un canal ($k = 3, l = 2$).

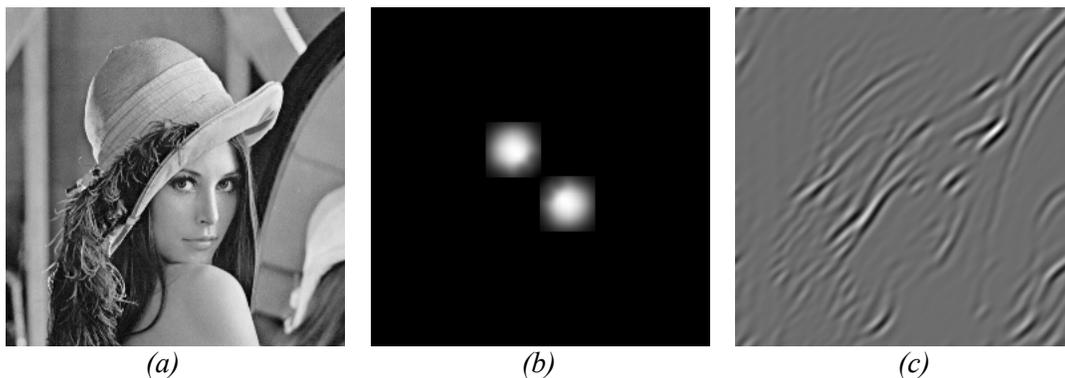


Figura 4-14 (a) Imagen original Lena. (b) Filtro Gabor de canal ($k = 3, l = 2$) en frecuencia. (c) Reconstrucción del canal ($k = 3, l = 2$).

4.6.4 Enmascaramiento espacial.

Como se vio en la sección 4.3.4, el enmascaramiento espacial es la reducción de la visibilidad de un componente en la imagen debido a la presencia de otro que lo oculta. Si no existiera el enmascaramiento, la imagen original (su contenido) no influiría en la visibilidad del error, que en este caso es la diferencia entre la imagen de referencia y la distorsionada. Sin embargo, lo que sucede en realidad es que la imagen original sí afecta al umbral de visibilidad del error. Los efectos del enmascaramiento se modelan mediante funciones de elevación del umbral que describen el incremento que debe producirse en el valor de un estímulo para que resulte visible dentro la imagen.

El propósito es determinar si una distorsión es visible dentro de la imagen. Por tanto, el contenido de la imagen es la fuente del enmascaramiento y a efectos prácticos se considerará ruido, mientras que la distorsión es la señal cuya detectabilidad se quiere predecir.

En (Eriksson, et al., 1998) se presentan tres modelos básicos de enmascaramiento de diferente complejidad, según incorporan o no los efectos entre canales. La implementación en la herramienta de calidad sólo se centra en el enmascaramiento de un mismo canal.

Así, al ver una imagen, estructuras simples como bordes y zonas uniformes enmascaran menos que las texturas. Esto se debe a que el observador sabe o puede predecir fácilmente que apariencia deben tener los bordes y zonas uniformes, lo que no sucede generalmente con texturas.

4.6.5 Mapas e índices de error.

El paso final en un sistema de medida de calidad perceptiva es la combinación de los mecanismos descritos anteriormente para obtener, o bien una imagen en la que cada píxel tiene un valor indicativo de la visibilidad del error en esa localización, o bien un valor numérico que describa la calidad global de la imagen, o ambos. Todo ello sin olvidar que tanto el mapa como el índice deben corresponderse con la valoración que se obtendría al mostrar la imagen de referencia y la distorsionada a un observador.

En el modelo descrito hasta el momento se ha asumido una descomposición de la imagen de referencia y la distorsionada en canales. El conjunto de imágenes resultado de este filtrado varía en número según la elección que se haga del banco de filtros.

El procedimiento lógico para reducir la dimensión de esta información a 2D (mapa) o 1D (índice) consiste en sumar las respuestas de los distintos canales utilizando la métrica de Minkowski (4.7). Un paso previo a esta suma es el cálculo de la visibilidad del error en cada canal, que se denomina $V_n(x, y)$ y se expresa como:

$$V_n(x, y) = \frac{C_n(x, y) - \bar{C}_n(x, y)}{Tem_n(x, y)} \quad \dots (4.19)$$

donde $C_n(x, y)$ y $\bar{C}_n(x, y)$ son respectivamente el contraste de la imagen original y la distorsionada en el canal n . Tem_n es el umbral de visibilidad llamado enmascaramiento mutuo, definido como:

$$Tem_n(x, y) = \min [Te_n(x, y), \bar{Te}_n(x, y)] \quad \dots (4.20)$$

$$Te_n(x, y) = \left(1 + \left(k_1 \left(k_2 |b_n(x, y)| \right)^s \right)^4 \right)^{\frac{1}{4}} \quad \dots (4.21)$$

donde $b_n(x, y)$ es la banda de la imagen en el canal n (escala k y orientación l). Con esto se obtendrán finalmente los mapas e índices que muestran la calidad que el observador percibe al visualizar una imagen de referencia con su correspondiente imagen distorsionada.

4.7 Referencias.

- **Aznar J** Análisis multiescala y multiresolución de imágenes mediante un banco de filtros de gabor 2D [Book]. - Barcelona, España : Facultad de Psicología, Universidad de Barcelona, 2000.
- **Beghdadi A and Iordache R** Image quality assessment using the joint space/spatial-frequency representation [Article]// EURASIP Journal on Applied Signal Processing. - 2006. - Vol. 2006. - pp. Article ID 80537, 8 pages. - doi:10.1155/ASP/2006/80537.
- **Daly S** The visible difference predictor: An algorithm for assessment of image fidelity [Book Section]// Digital Images and Human Vision / book auth. Watson A. - Massachusetts, E. U. : MIT Press, 1993.
- **Eckert M and Bradley A** Perceptual quality metrics applied to still image compression [Journal]// Signal Processing. - 1998. - 3 : Vol. 70. - pp. 177-200.
- **Eriksson R, Andren B and Brunnström K** Modelling the perception of digital images: A performance study [Article]// Proceedings SPIE Conference of Human Vision and Electroning Imaging III. - 1998. - Vol. 3299. - pp. 88-97.
- **Fischer S** Modelos perceptuales en la cuantificación y codificación de imágenes [Book]. - Madrid, España. : Tesis Maestría. Universidad Politécnica de Madrid, 2000.
- **Gabor D** Theory of communication [Article]// Journal I.E.E.. - 1946. - 3 : Vol. 93. - pp. 429-457.
- **Gallego A** Modelos visuales en el análisis de la calidad de imagen [Book]. - Madrid, España : Tesis de Maestría. Dpto. de Ingeniería Electrónica, Instituto de Óptica "Daza de Valdés" (CSIC), 2006.
- **Gottesfeld L** A survey of image registration techniques [Journal]// ACM Computing Surveys. - 1992. - 4 : Vol. 24. - pp. 325-352.

- **Hormigon J and Cristóbal G** Image segmentation using the Wigner-Ville distribution [Book Section] // Advances in Electronics and Electron Physics / book auth. Hawkes P. - E. U. : Academic Press, 2003.
- **ITU** Methodology for the subjective assessment of the quality of television pictures. - 2000. - pp. 500-509.
- **Lubin J** The use of psychophysical data and models in the analysis of display performance [Book Section] // Digital Images and Human Vision / book auth. Watson A. - Massachusetts, E. U. : MIT Press, 1993.
- **Michelson A** Studies in optics [Book]. - Chicago, E. U. : University of Chicago Press, 1927.
- **Nadenau M** Integration of human color vision models into high quality image compression. [Book]. - Suiza : PhD. Thesis. Swiss Federal Institute of Technology, 2000.
- **Nava R, Gallego A and Cristóbal G** Una nueva herramienta para la evaluación de la calidad perceptible en imágenes [Conference] // Congreso Anual de la Sociedad Española de Ingeniería Biomedica. - Pamplona, España : [s.n.], 2006. - pp. 77-80.
- **Peli E** Contrast in complex images [Article] // Journal of the Optical Society of America A. - 1990. - 10 : Vol. 7. - pp. 2032-2040.
- **Peli E** Contrast sensitivity function and image discrimination [Article] // Journal of Optical Society of America A. - 2001. - 2 : Vol. 18. - pp. 283-293.
- **Peli E** In search of contrast metric: Matching the perceived contrast of gabor patches at different phases and bandwidths [Journal] // Vision Research. - 1997. - 2 : Vol. 37. - pp. 3217-3224.
- **Pratt W** Digital Image Processing [Book]. - New York, E. U. : John Wiley, 1991.
- **Reed T and Wechsler H** Spatial/spatial-frequency representations for image segmentation and grouping [Article] // Image and vision computing. - 1991. - 3 : Vol. 9.
- **Rust B** A new representation of the contrast sensitivity function for human vision [Conference] // International Conference on Image Science Systems and Technology. - Las Vegas, E. U. : [s.n.], 1997. - pp. 1-15.
- **Taylor C** Image Quality Assessment via a Human Visual System Model. PhD. Dissertation [Book]. - Indiana, E. U. : Purdue University, 1998.
- **Wang Z and Bovik A** A universal image quality index [Journal] // IEEE Signal Processing Letters. - 2002. - 3 : Vol. 9. - pp. 81-84.
- **Wang Z, Bovik A and Lu L** Image quality assessment from error measurement to structural similarity [Journal] // IEEE Transactions on Image Processing. - 2004. - 4 : Vol. 13. - pp. 600-612.
- **Westen S, Langendijk R and Biemond J** Perceptual image quality based on multiple channel HSV model [Conference]. - Detroit, E. U. : [s.n.], 1995. - Vol. 4. - pp. 2351-2354.

5

Algoritmo de Fusión.

*“An engineer is a device for turning coffee into theorems”
Paul Erdos*

5.1 *Algoritmo de Fusión.*

Para que un sistema visual, biológico o artificial, pueda interpretar de forma adecuada la información, debe simplificarla extrayendo las características más importantes de la imagen y almacenarla en forma eficiente. La mayoría de los investigadores están de acuerdo en que los bordes y líneas son en particular una rica fuente de información.

Existen detectores en el sistema visual humano especializados en responder a bordes y líneas, (Tolhurst, 1972) sugiere que los detectores visuales pueden aprovechar la simetría local de estas características. La evidencia psicofísica es consistente con la idea de los detectores visuales. Sin embargo, los mecanismos mediante los cuales los detectores localizan bordes y líneas no son obvios. El problema radica en que los detectores de bordes responden también a líneas ligeramente apartadas del centro de los campos receptivos. El mismo argumento aplica para los detectores de líneas. Así que aunque los detectores no dan una respuesta ambigua, se debe decodificar su respuesta con reglas adicionales.

Algunos modelos de detección de características emplean un operador. Por ejemplo, (Marr, et al., 1980) usan la función laplaciana de Gaussiana y buscan el cruce con cero en todas las escalas. Este modelo está diseñado para detectar bordes, no líneas. El procedimiento de coincidencia de cruce por cero elimina marcar líneas de forma inapropiada. Sin embargo, esta condición también elimina bordes importantes bajo ciertas condiciones. Un ejemplo ocurre cuando aparecen dos bordes muy juntos, el cruce por cero en escalas muy grandes aparece en medio de los bordes.

La dificultad más importante es que no existe una definición matemática satisfactoria de qué constituye una línea o borde para el sistema de visión humano. Sin embargo una buena aproximación es el filtro de Canny (Canny, 1986), que utiliza dos operadores, para simetría par e impar, buscando máximos locales.

En el desarrollo del algoritmo presentado en este trabajo de tesis, se tomó en cuenta la importancia que tiene para el sistema de percepción visual los bordes y líneas, por lo que se diseñó una regla de fusión orientada a la extracción de características.

En la sección 3.2 se habló de la fusión a nivel características. Este tipo de fusión requiere la identificación y extracción de información destacada de las imágenes de entrada y su transferencia a la imagen fusionada, removiendo la redundancia de información. Dependiendo de la naturaleza de las imágenes, las características destacadas pueden ser texturas, bordes o líneas.

En la Figura 5-1 se presenta el esquema del algoritmo propuesto en esta tesis. El primer paso es calcular la transformada Wavelet Discreta para cada una de las imágenes de entrada, en uno o varios niveles de resolución o escala. En (Dongmei, et al., 2001) se presenta una justificación para elegir el número adecuado de niveles de descomposición. Para el caso de fusión, se consideran uno o dos niveles de resolución, ya que utilizar más niveles en la etapa de análisis sólo eleva el costo computacional y no refleja una mejoría en los resultados de la fusión.

Segundo, se extraen las características más importantes que aparecen en los coeficientes wavelet (bordes y líneas), utilizando una implementación del filtro de Canny. Estas características se utilizan para crear un mapa de decisión que permite decidir que coeficientes se conservan. En (5.1) se muestra la regla de fusión.

$$y_F^k = \begin{cases} a_A^k & \text{Si } a_A^k \in b_A^k \ \& \ a_B^k \notin b_B^k \\ a_B^k & \text{Si } a_B^k \in b_B^k \ \& \ a_A^k \notin b_A^k \\ \max(a_A^k, a_B^k) & \text{Si } a_A^k \in b_A^k \ \& \ a_B^k \in b_B^k \\ CICB^* & \text{otro caso} \end{cases} \quad \dots (5.1)$$

Donde a_A^k y a_B^k son los valores de intensidad de los píxeles en el nivel de descomposición k , b_A^k y b_B^k son los bordes de las imágenes A y B , respectivamente, en el nivel de descomposición k y y_F^k es el valor del píxel en la imagen fusionada en el dominio wavelet.

En otras palabras, si un píxel perteneciente a los coeficientes wavelets de la imagen A es un borde y su correspondiente píxel en los coeficientes de la imagen B no lo es, entonces se asigna a y_F^k . De la misma forma si un píxel perteneciente a los coeficientes de la imagen B es un borde y su correspondiente en la imagen A no lo es, se asigna a y_F^k . Si ambos píxeles a_A^k y a_B^k son bordes, se conserva el borde de mayor energía.

En cualquier otro caso, el criterio de selección será mediante comparación de intensidades utilizando operadores de comportamiento constante independientes del contexto.

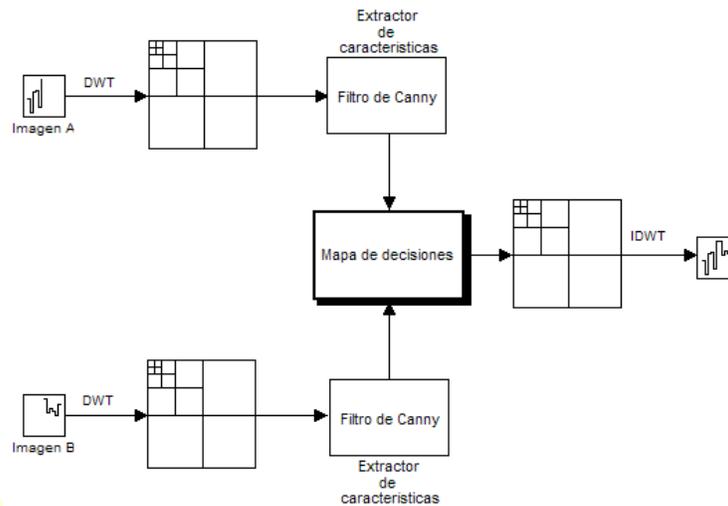


Figura 5-1 Esquema del algoritmo de fusión propuesto. La fusión orientada a características extrae los bordes de ambas imágenes y los conserva durante todo el proceso de fusión.

5.1.1 Extracción de características.

Las primeras técnicas para realce de bordes fueron desarrolladas a principios de 1970. Prewitt (Prewitt, 1970) y Sobel (Sobel, 1990) entre otros (una revisión completa la puede encontrar en (Pinho, et al., 1997)) propusieron convoluciones con pequeños kernels. En 1980 aparecieron modelos más complejos inspirados en el sistema visual humano basados en filtros laplacianos (p. ej. filtros paso-banda isotrópicos).

En 1986 Canny (Canny, 1986) propuso un método basado en el criterio de buena detección (baja probabilidad para no detectar bordes existentes), buena localización y detección única (evita detectar un borde dos veces) para construir filtros 2D antisimétricos para la extracción de bordes. El esquema de Canny implementa mecanismos adicionales para adaptarse a los niveles de ruido. El detector de Canny constituye un referente hasta nuestros días.

El detector de Canny primero suaviza la imagen para eliminar ruido aditivo y después encuentra el gradiente para realzar regiones donde el valor de las derivadas es máximo. El algoritmo sigue sobre estas regiones y suprime cualquier píxel que no sea un máximo (supresión de *No Máximos*). El gradiente se reduce mediante histéresis. La histéresis se utiliza para tomar decisiones sobre los píxeles que no han sido suprimidos utilizando dos umbrales. Si la magnitud de un píxel está por debajo del primer umbral, se elimina. Si la magnitud del píxel está por encima del segundo umbral, es marcado como borde. Los píxeles que se encuentran entre los dos intervalos se eliminan si no están conectados directamente con un borde.

- *Suavizado de la imagen.*

El primer paso para la implementación del filtro de Canny es eliminar el ruido aditivo en la imagen original, lo cual se logra aplicando un filtro Gaussiano. El resultado es una imagen suavizada (5.2).

$$g(x, y) = f(x, y) * \omega_G(x, y; \sigma) \quad \dots (5.2)$$

donde $f(x, y)$ es la imagen original y la función Gaussiana está dada por:

$$\omega_G(x, y; \sigma) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x^2+y^2)}{2\sigma^2}} \quad \dots (5.3)$$

El tamaño del filtro Gaussiano lo determina el parámetro σ , que es la varianza y el cual se relaciona de acuerdo con:

- El nivel de detalles deseados: bordes finos vs bordes globales.
- El nivel de ruido.
- El intercambio detección localización: principio de incertidumbre de Heisenberg.

- *Obteniendo el gradiente.*

Después de suavizar la imagen y eliminar el ruido, el siguiente paso es encontrar la intensidad de los bordes mediante el gradiente de la imagen. El operador Sobel realiza una medición del gradiente en 2D sobre la imagen. Después de esto, se puede encontrar el valor absoluto de la magnitud del gradiente en cada punto. El operador Sobel utiliza un par de máscaras de convolución de 3×3 , una estima el gradiente en la dirección x , bordes verticales G_x y la otra en la dirección y , bordes horizontales G_y (5.4).

$$G_x = \begin{bmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{bmatrix} \quad \dots (5.4a)$$

$$G_y = \begin{bmatrix} 1 & 2 & 1 \\ 0 & 0 & 0 \\ -1 & -2 & -1 \end{bmatrix} \quad \dots (5.4b)$$

Aplicando el operador gradiente, podemos obtener la magnitud y la orientación de los bordes (5.5).

$$\begin{aligned}\nabla g(x, y) &= \left[\frac{\partial}{\partial x} g(x, y), \frac{\partial}{\partial y} g(x, y) \right]^T \\ &= \left[g(x, y)_x, g(x, y)_y \right]^T\end{aligned}\quad \dots (5.5)$$

Una forma de implementación más eficiente se obtiene sustituyendo (5.2) en (5.5)

$$\begin{aligned}\nabla g(x, y) &= \nabla (f(x, y) * \omega_G(x, y; \sigma)) \\ &= \nabla (\omega_G(x, y; \sigma)) * f(x, y)\end{aligned}\quad \dots (5.6)$$

el filtro Gaussiano cumple con la propiedad de ser separable (5.7).

$$\omega_G(x, y; \sigma) = \omega_G(x) \cdot \omega_G(y) \quad \dots (5.7)$$

De esta forma, la implementación es una convolución con la primera derivada de Gaussiana, Figura 5-2. La aproximación al operador de la primera derivada de Gaussiana fue propuesto por Macleod (Macleod, 1970) para detectar la dirección de los bordes.

En dos dimensiones, marcar un borde como el valor máximo en la dirección del gradiente equivale a encontrar el cruce por cero con un filtro de segunda derivada.

Encontrar la intensidad de los bordes y la dirección del gradiente es trivial si ya se ha calculado (5.5). La magnitud se expresa en (5.8) y la dirección en (5.9).

$$M(x, y) = |\nabla g(x, y)| = \sqrt{g(x, y)_x^2 + g(x, y)_y^2} \quad \dots (5.8)$$

$$\Theta(x, y) = \tan^{-1} \frac{g(x, y)_y}{g(x, y)_x} \quad \dots (5.9)$$

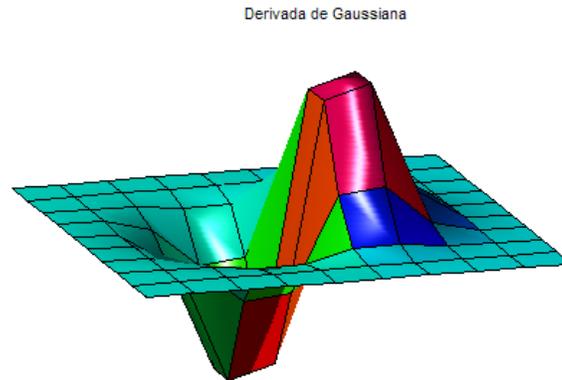


Figura 5-2 Primera derivada de Gaussiana en 2D.

- *Supresión de No máximos.*

Debido a las múltiples respuestas, la magnitud de los bordes, $M(x, y)$ puede contener líneas anchas sobre los máximos locales, por lo que se realiza una supresión de los valores *No máximos* para eliminar bordes fantasmas, preservando la conectividad de los contornos y ajustando el ancho a un píxel. La ecuación (5.10) describe el algoritmo de supresión de *No máximos*.

$$M(x, y) = \begin{cases} |\nabla g(x, y)| & \text{Si } |\nabla g(x, y)| > |\nabla g(x', y')| \ \& \ |\nabla g(x, y)| > |\nabla g(x'', y'')| \\ 0 & \text{Otro caso} \end{cases} \quad \dots (5.10)$$

donde (x', y') y (x'', y'') son los vecinos perpendiculares de (x, y) en $\nabla g(x, y)$, en la dirección normal al borde, Figura 5-3.

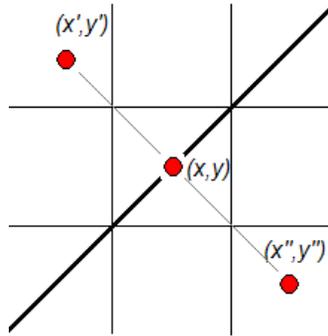


Figura 5-3 Supresión de *No Máximos*, en la dirección normal al borde.

- Histéresis.

Todos los esquemas de detección de bordes utilizan algún umbral. Si el umbral no se fija *a priori*, se determina de alguna manera mediante algún algoritmo. El presente algoritmo ajusta el umbral basado en estimaciones locales del ruido de la imagen. Una complejidad adicional es el uso de dos umbrales. El umbral inferior T_L elimina los falsos bordes y el umbral superior T_H fragmenta los bordes de la siguiente forma:

- Un píxel (x, y) es llamado **fuerte** si $M(x, y) > T_H$.
- Un píxel (x, y) es llamado **débil** si $M(x, y) \leq T_L$.
- El resto de píxeles son llamados **candidatos**.

En cada posición (x, y) , descartar el píxel (x, y) si es débil; si el píxel es fuerte, se marca como borde. Si el píxel es un candidato, verificar si está conectado a un píxel fuerte, en dado caso se marca como borde, en otro caso se descarta.

5.1.2 Ejemplo del filtro de Canny.

La Figura 5-4 muestra el resultado de la extracción de bordes con el filtro de Canny. A la imagen de la Figura 5-4(a) se le aplica el filtro de Canny, variando el valor de la varianz σ . Se puede apreciar que a una menor varianza, se detectan más líneas, esto es porque el filtro es más sensible con variancias pequeñas.

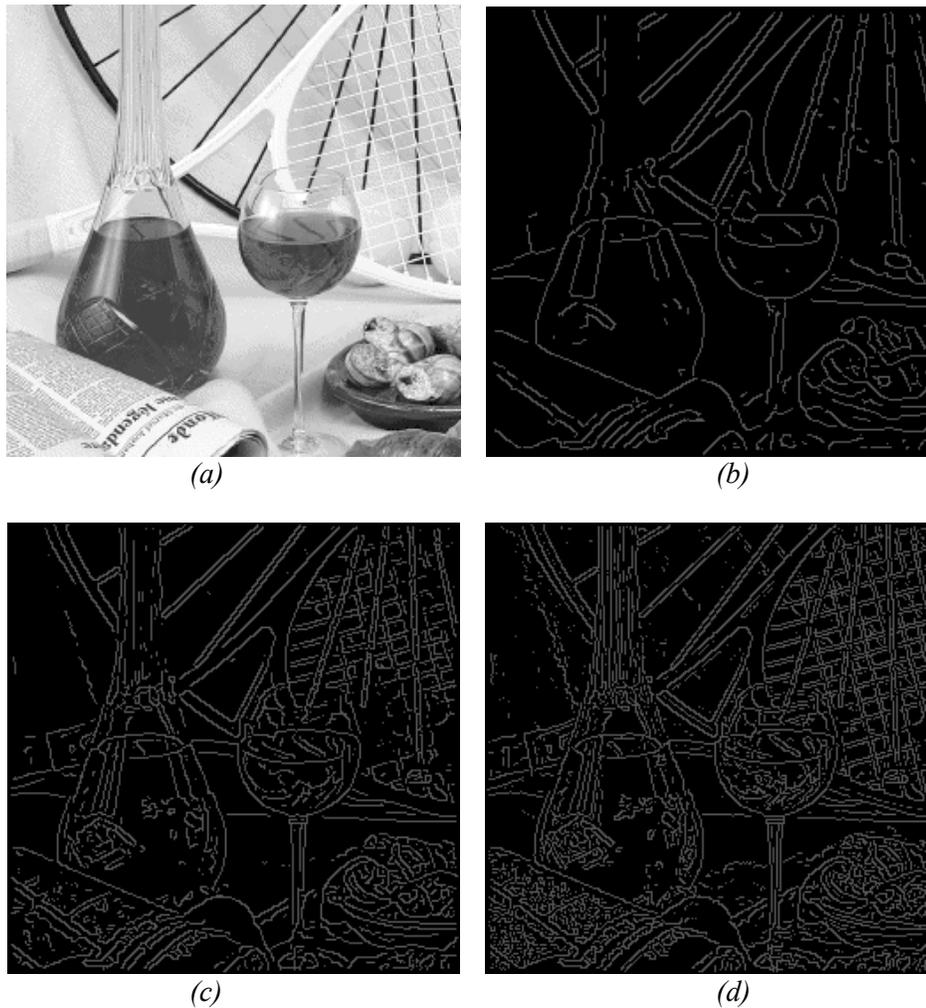


Figura 5-4 Detección de bordes; (a) Imagen Original. (b) $\sigma = 2.5$. (c) $\sigma = 1.0$. (d) $\sigma = 0.6$

5.2 Resultados experimentales.

En esta sección se presentan algunos experimentos de fusión de imágenes con el propósito de comparar el desempeño de los métodos tradicionales y el algoritmo propuesto en esta tesis. Las pruebas se realizan sobre conjuntos de imágenes multifocales.

Las imágenes multifocales son complementarias en diversas formas y una sola no es suficiente en términos de su respectivo contenido de información. Observar una serie de dichas imágenes de forma separada e individualmente no es útil ni conveniente. La ventaja de los datos multifocales puede ser explotada fusionando las regiones más nítidas y no desenfocadas de las diferentes

imágenes de entrada para obtener una sola imagen, aprovechando la sobrerrepresentación de los datos. Con el propósito de evaluar los resultados, los experimentos se han dividido en dos grupos:

- a) fusión de imágenes multifocales con referencia (*Ground Truth*¹).
- b) fusión de imágenes multifocales sin referencia.

Algunas imágenes utilizadas en estos experimentos son cortesía del Dr. Shutao Li del Departamento de Ciencias de la Computación de la Universidad de Ciencia y Tecnología de Hong Kong, Hong Kong, China y aparecen en el artículo (Li, et al., 2001). Se citará la referencia expresamente en su momento.

5.2.1 Fusión de imágenes con referencia.

Los criterios utilizados para evaluar la fusión de las imágenes con referencia son: la relación señal a ruido de pico (PSNR), el error cuadrático medio (MSE), la relación señal a ruido de Wigner (SNR_w), propuesta por (Beghdadi, et al., 2006), la medida de calidad basada en la información estructural (SSIM), creada por (Wang, et al., 2002). Todos estos criterios fueron explicados ampliamente en el Capítulo 4.

5.2.1.1 Servilletero y manzana.

Se consideraron dos imágenes multifocales en escala de grises, de tamaño 512 × 512 píxeles. La Figura 5-5(a) es la imagen de referencia, la Figura 5-5(b) tiene el servilletero desenfocado y la manzana enfocada, mientras que la Figura 5-5(c) enfoca el servilletero pero la manzana no. Se realizaron cuatro tipos de fusión utilizando la transformada Haar, el propósito en estos experimentos es sólo comparar la regla de fusión propuesta en esta tesis con las reglas tradicionales, por lo que el tipo de transformación deberá ser la misma para cada uno de los experimentos. A continuación se describe cada uno de los esquemas realizados.

- En el esquema 1 se fusionaron dos imágenes multifocales, Figura 5-5(b) y Figura 5-5(c), utilizando la wavelet Haar² con un nivel de descomposición. La regla de fusión conserva los valores más grandes para los coeficientes de promedio y detalles (MAX-MAX). El resultado aparece en la Figura 5-6(a).
- El esquema 2 utiliza el mismo tipo de descomposición que el esquema, pero ahora se aplica el método de extracción de características con los siguientes parámetros: *connectivity* = 8,

¹ *Ground Truth* es un término usado comúnmente en cartografía, meteorología y en general en percepción remota. Es de especial importancia, ya que permite relacionar los datos reales con los recolectados a distancia. De forma más específica, la colección de datos *ground truth* permite comparar la información de los píxeles de una imagen satelital con los datos reales tomados en sitio.

² Coeficientes Haar: $a_0 = \frac{1}{\sqrt{2}}, a_1 = \frac{1}{\sqrt{2}}$

$filter = \text{Prewitt}$; $\sigma = 0.6$, $low_threshold = -100$, y $high_threshold = -10$. El resto de coeficientes se seleccionaron con el criterio (MAX-MAX). El resultado aparece en la Figura 5-6(b).

- El esquema 3 utiliza la wavelet Haar con dos niveles de descomposición y la regla de selección de coeficientes empleada es la misma que en el esquema 1. El resultado se muestra en la Figura 5-6(c).
- El esquema 4 utiliza la misma descomposición que se aplicó en el esquema 3, pero ahora se aplica el extractor de características con los mismos parámetros que en el esquema 2. El resultado se muestra la Figura 5-6(d).

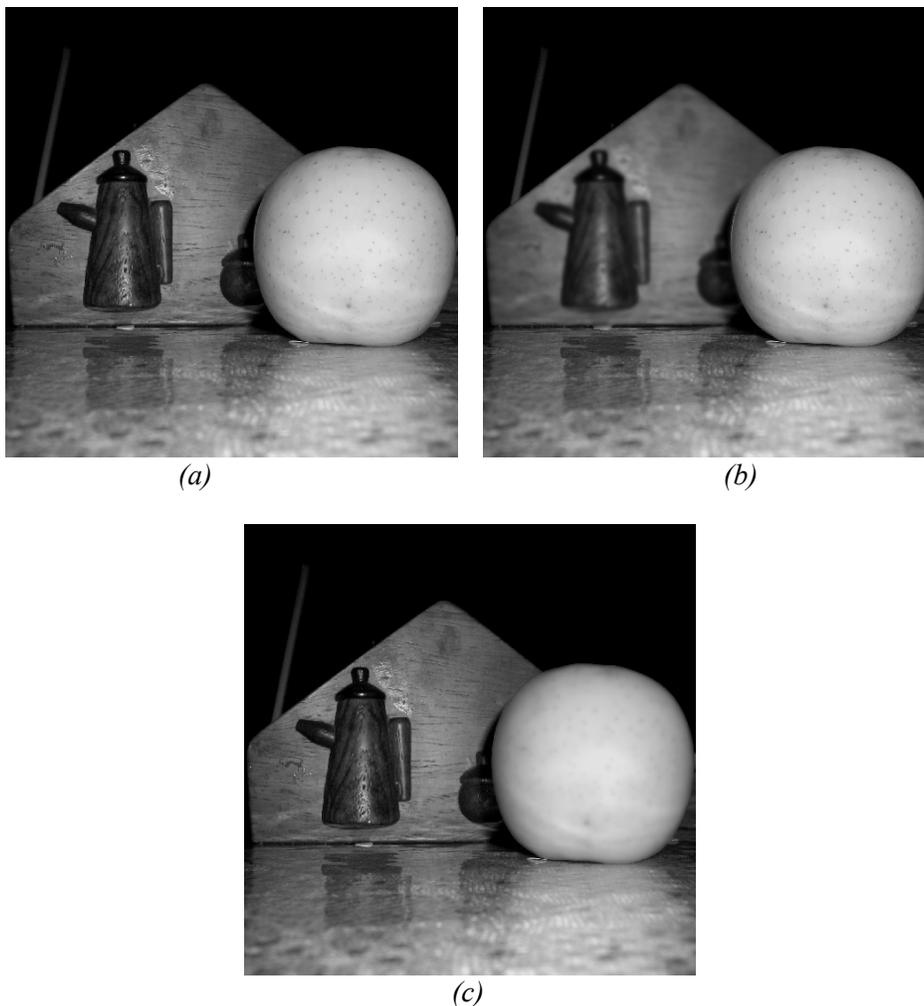


Figura 5-5 Imágenes de entrada multifocales de 512×512 . Servilletero y manzana. (a) Imagen de referencia o *ground truth*. (b) Imagen con el servilletero desenfocado y la manzana enfocada. (c) Imagen con el servilletero enfocado y la manzana desenfocada.

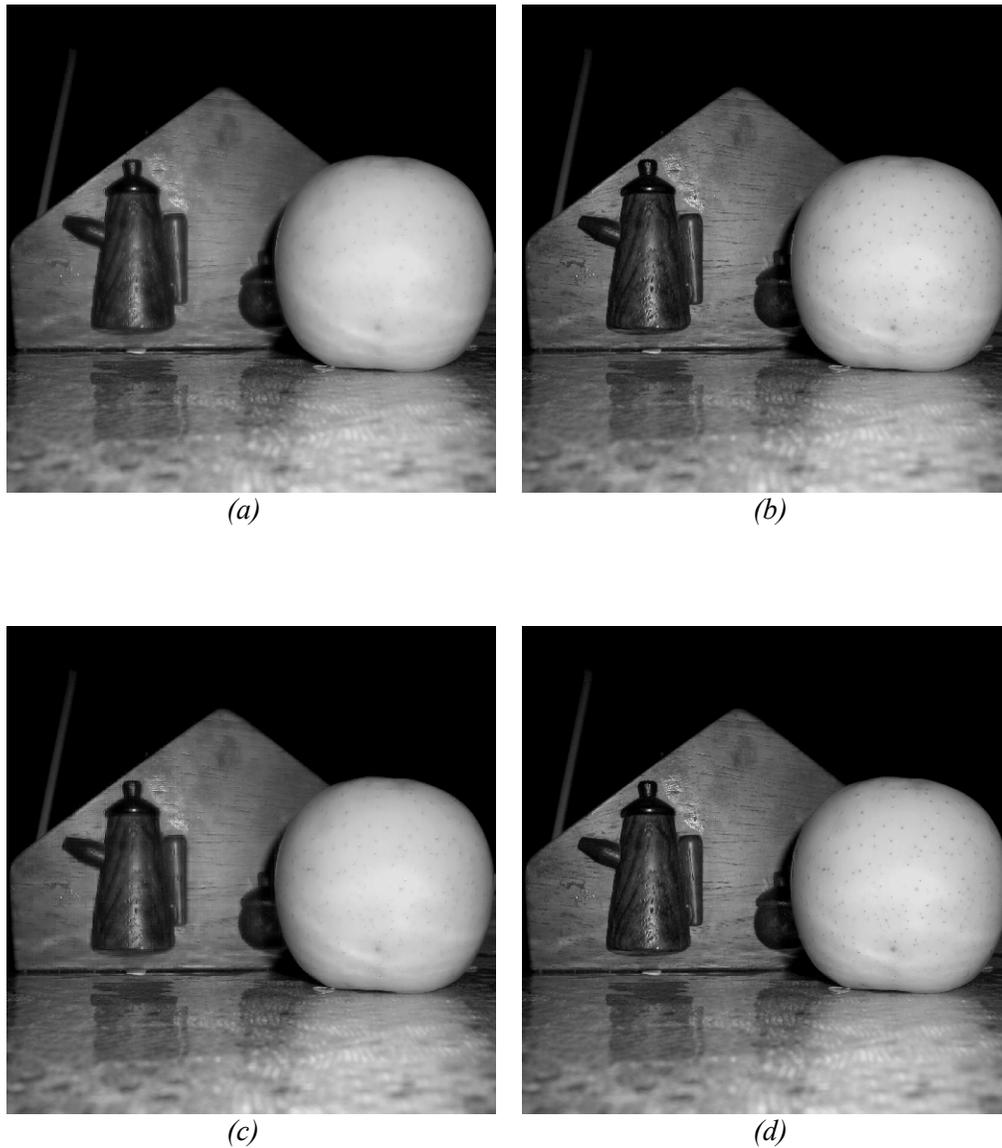


Figura 5-6 Resultado de los esquemas de fusión para el experimento Servilletero y manzana. (a) Imagen obtenida con la wavelet Haar a un nivel de descomposición, el criterio de selección de coeficientes fue MAX-MAX. (b) Imagen obtenida con Haar a un nivel de descomposición, los coeficientes se mezclaron usando el filtro de Canny y la regla de fusión MAX-MAX. (c) Imagen obtenida utilizando Haar con dos niveles de descomposición y la regla MAX-MAX. (d) Imagen obtenida usando Haar a dos niveles de descomposición y el filtro Canny, la regla de fusión fue MAX-MAX.

En la Figura 5-7 se muestran los mapas de decisión generados por el filtro de Canny. Los pixeles blancos corresponden a los valores que se conservan en la fusión. Las Figuras 5-7(a) y 5-7(b) corresponden al esquema 2. Las Figuras 5-7(c) y 5-7(d) corresponden al esquema 4.

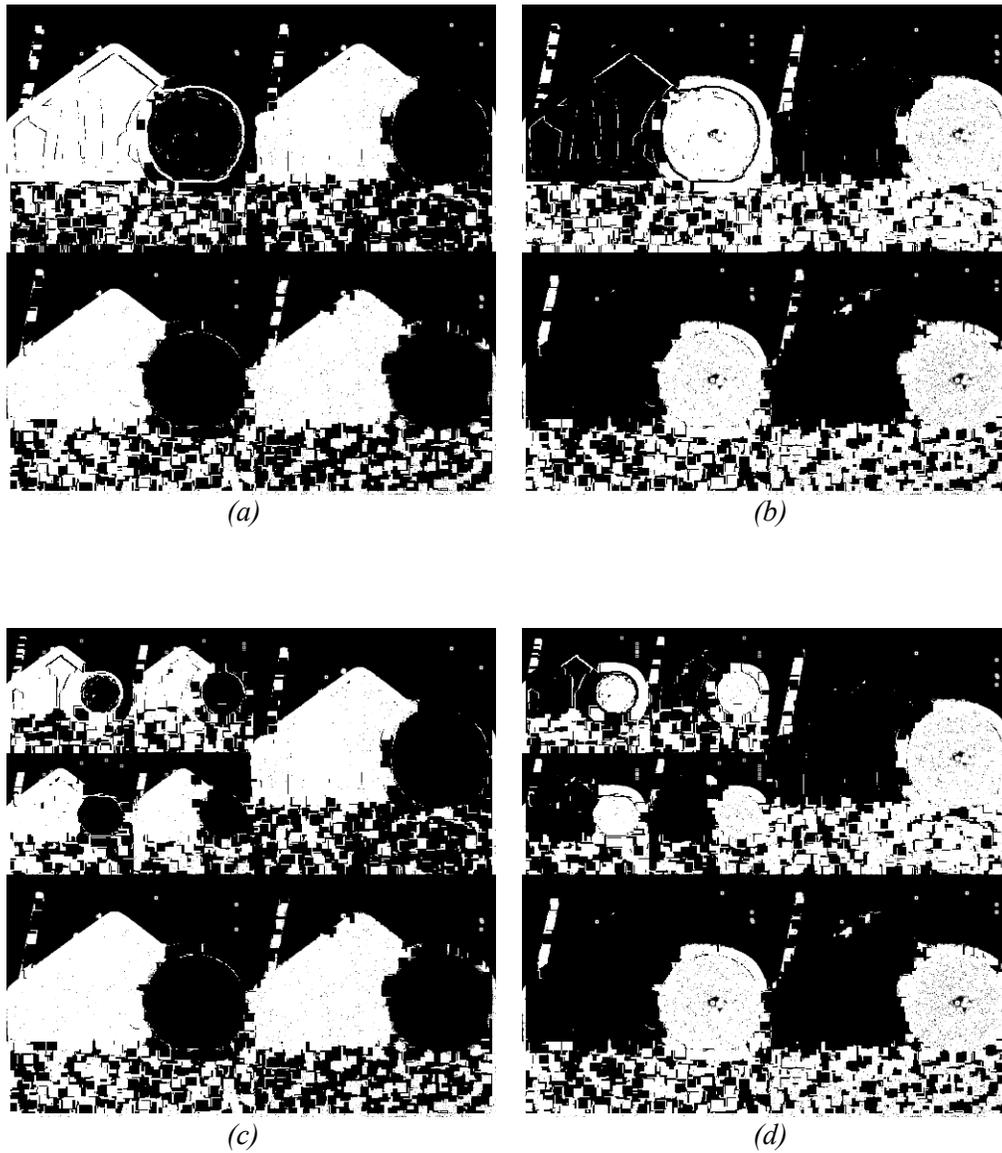


Figura 5-7 Mapas de decisión para el experimento Servilletero y manzana. (a) y (b) Mapas de decisión del esquema 2. (c) y (d) Mapas de decisión del esquema 4.

Los resultados obtenidos se presentan en la Tabla 5-1. Se realizan dos comparaciones, la primera entre el esquema 1 y el esquema 2; la segunda entre en esquema 3 y el esquema 4. Para ambas comparaciones, el algoritmo propuesto en esta tesis tiene mejor desempeño que la regla de fusión tradicional correspondiente.

Fusión	PSNR [dB]	MSE	SNR _w	SSIM
#1	30.4963	8.92e-4	9.9263	0.9594
#2	39.547	1.109e-4	14.0275	0.99712
#3	26.4666	2.256e-3	7.6	0.95271
#4	45.301	2.95e-5	22.5525	0.99626

Tabla 5-1 Resultados del desempeño de la fusión para el experimento **Servilletero y manzana**.

5.2.1.2 Globos.

Para este experimento se consideraron dos imágenes multifocales en escala de grises de tamaño 512×512 . Las imágenes de la Figura 5-8 son cortesía del Dr Shutao Li. La Figura 5-8(a) es la imagen de referencia. Las Figuras 5-4(b) y 5-4(c) son las imágenes con distintas regiones desenfocadas. Se realizaron cuatro tipos de fusión. Los dos primero esquemas utilizan Coiflet 6 y los esquemas 3 y 4 utilizan Symlet4. A continuación se describe cada uno de los esquemas realizados.

- En el esquema 1 se fusionaron las imágenes multifocales de las Figuras 5-8(b) y 5-8(c) utilizando la wavelet Coiflet 6 (Cf6)³ con un nivel de descomposición. Se seleccionaron los promedios de los coeficientes de detalle y de aproximación para reconstruir la señal (AV-AV). El resultado se muestra en la Figura 5-9(a).
- El esquema 2 utiliza la misma descomposición que el esquema 1, pero ahora se utiliza el extractor de características con los siguientes parámetros: *connectivity* = **8**, *filter* = **Prewitt**; σ = **0.6**, *low_threshold* = **-50**, y *high_threshold* = **-10**. El resto de coeficientes se seleccionaron con el mismo criterio que en el esquema 1. El resultado se muestra en la Figura 5-9(b).
- En el esquema 3 se utilizó la wavelet Symlet 4 (Sym4)⁴ con un nivel de descomposición. La regla de fusión es la misma que el esquema 1. El resultado se muestra en la Figura 5-9(c).
- El esquema 4 utiliza la misma descomposición que el esquema 3, pero se aplica el extractor de características con los siguientes parámetros: *connectivity* = **8**, *filter* = **Prewitt**; σ = **0.5**, *low_threshold* = **-30**, y *high_threshold* = **10**. El resto de coeficientes se seleccionaron con el criterio (AV-AV). El resultado se muestra en la Figura 5-9(d).

³ Coeficientes Coiflet (Cf6):

$a_0 = -0.0727, a_1 = 0.3379, a_2 = 0.8526, a_3 = 0.3849, a_4 = -0.0727, a_5 = -0.0157$

⁴ Coeficientes Symlet (Sym4):

$a_0 = 0.0322, a_1 = -0.0126, a_2 = -0.0992, a_3 = 0.2979, a_4 = 0.8037, a_5 = 0.4976, a_6 = -0.0296, a_7 = -0.0758$

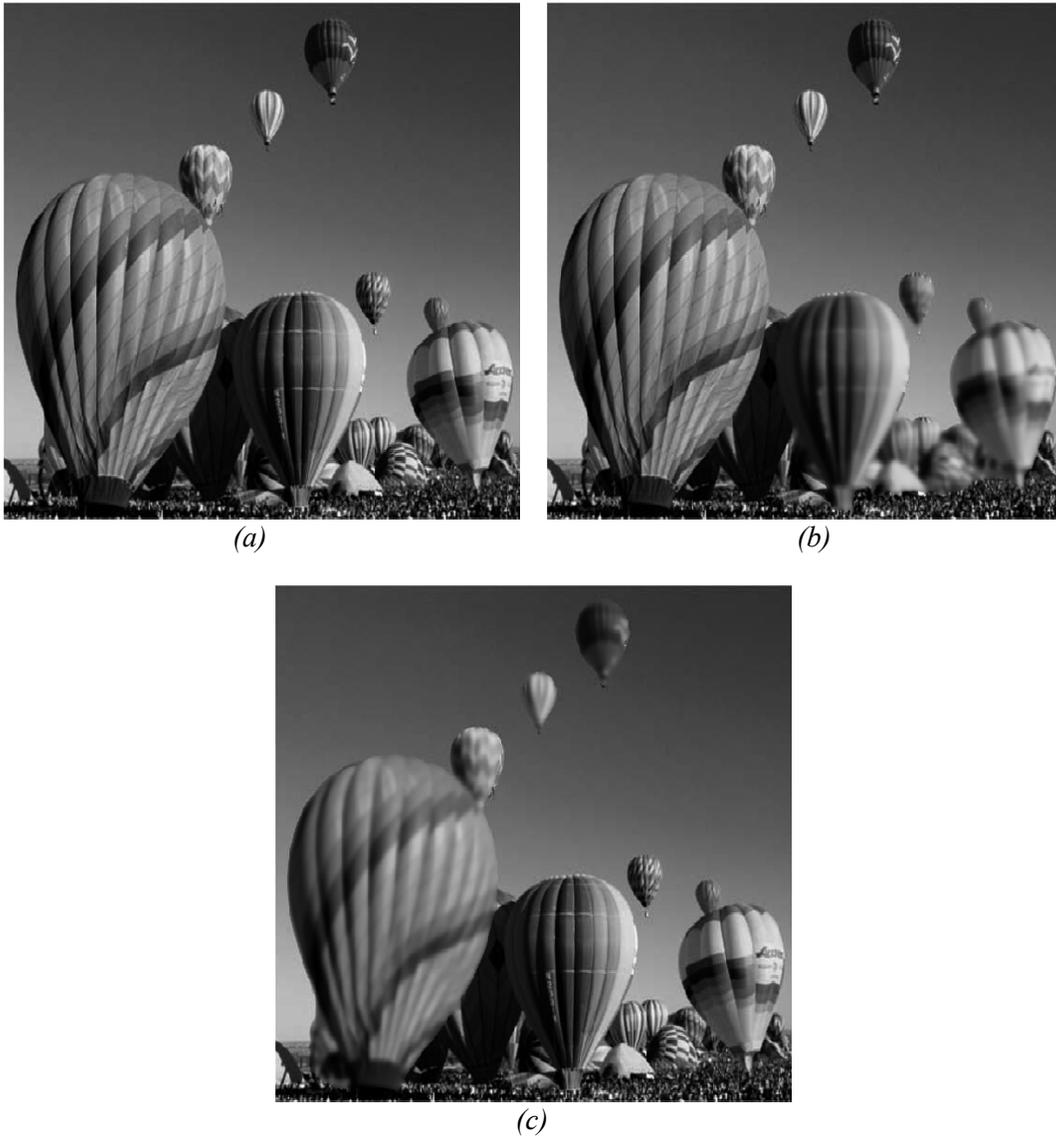


Figura 5-8 Imágenes de entrada multifocales. Globos. (a) Imagen de referencia. (b) Imagen con los globos de la derecha desenfocados y el resto de la imagen en foco. (c) Imagen con el globo de la izquierda y los tres globos de la parte superior desenfocados.

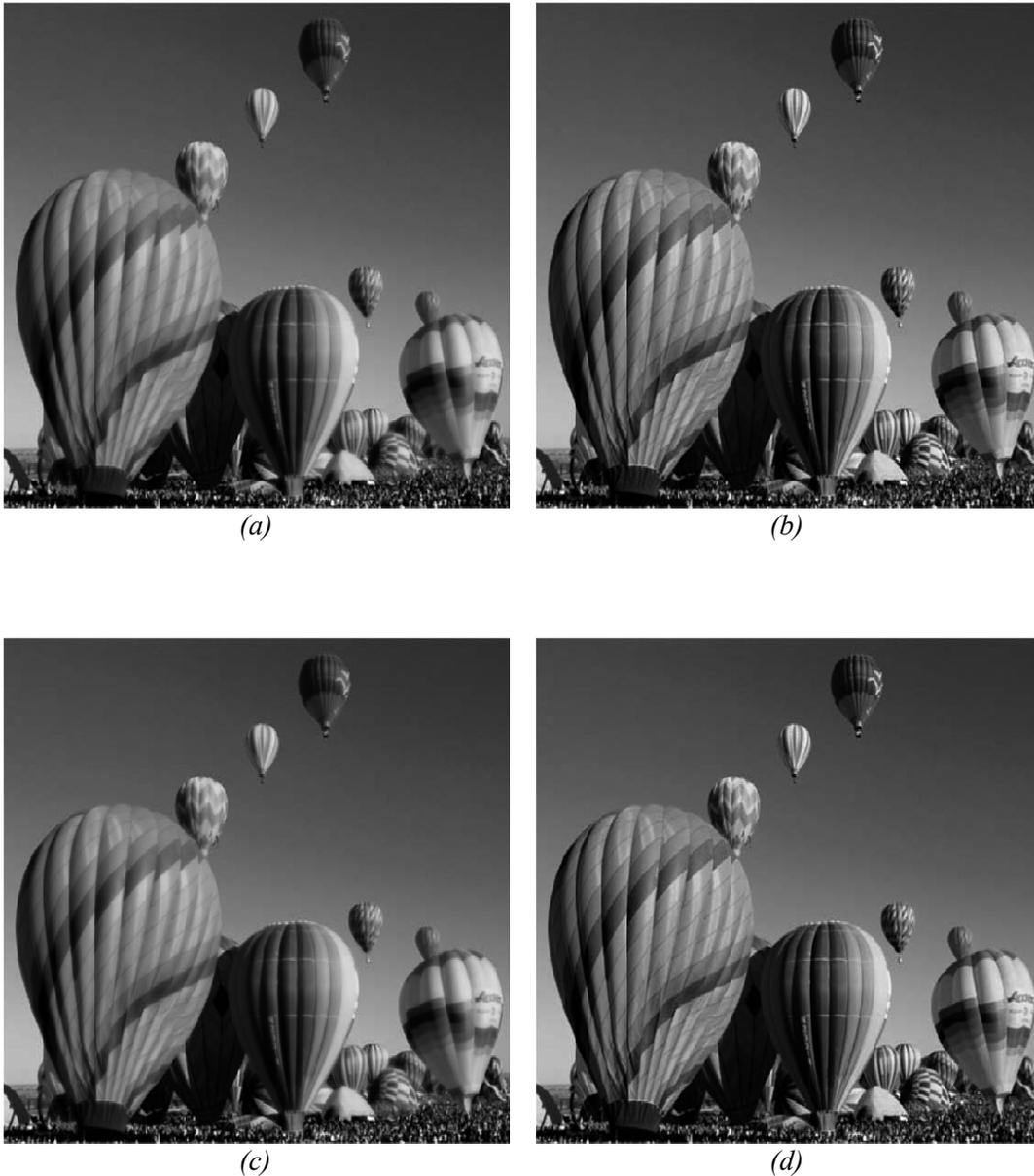


Figura 5-9 Resultado de los esquemas de fusión para el experimento Globos. (a) Imagen obtenida usando Cf6 con un nivel de descomposición y la regla de fusión AV-AV. (b) Imagen obtenida con Cf6 a un nivel de descomposición usando el filtro de Canny y la regla de fusión AV-AV. (c) Imagen obtenida usando Sym4 a un nivel de descomposición y la regla de fusión AV-AV. (d) Imagen obtenida con Sym4 a un nivel de descomposición usando Canny y la regla de fusión AV-AV.

En la Figura 5-10 se muestran los mapas de decisión generados por el filtro de Canny. Las Figuras 5-10(a) y 5-10(b) pertenecen al esquema 2 y las Figuras 5-10(c) y 5-10(d) pertenecen al esquema 4.

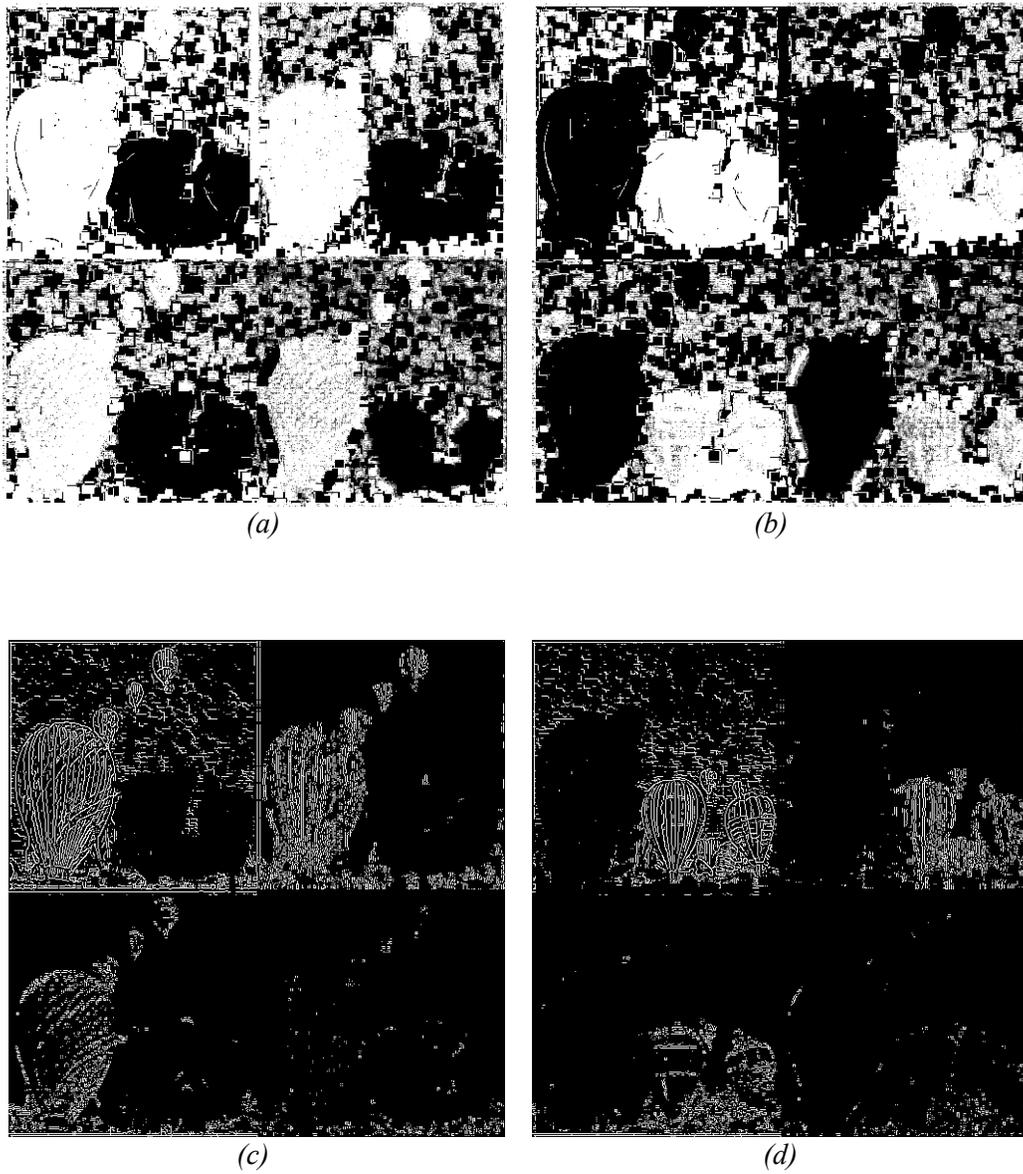


Figura 5-10 Mapas de decisión para el experimento Globos. (a) y (b) Mapas de decisión del esquema 2. (c) y (d) Mapas de decisión del esquema 4.

En la Tabla 5-2 se muestra el desempeño de los esquemas de fusión. El esquema 2 que utiliza el algoritmo propuesto en esta tesis presenta mejor desempeño cuantitativo sobre el esquema 1. De la misma forma, el esquema 4 tiene mejores resultados sobre el esquema 3.

Fusión	PSNR [dB]	MSE	SNR _w	SSIM
#1	35.2652	2.975e-4	17.8113	0.9777
#2	40.0984	1.2307e-5	28.7155	0.9987
#3	35.264	2.9758e-4	16.3956	0.97788
#4	46.3678	2.3079e-5	17.8852	0.99826

Tabla 5-2 Resultados del desempeño de la fusión para el experimento **Globos**.

5.2.1.3 Pelota y fondo.

En el siguiente experimento consideramos dos imágenes en escala de grises de tamaño 256×256 . La Figura 5-11(a) es la imagen *ground truth* o de referencia con el fondo y la pelota enfocados. La Figura 5-11(b) tiene el fondo desenfocado y la pelota en foco y la Figura 5-11(c) tiene el fondo en foco pero la pelota esta desenfocada. De nueva cuenta se realizaron cuatro experimentos que se describen a continuación.

- En el esquema 1 se utilizó la wavelet Db3⁵ con un nivel de descomposición. la regla de fusión conserva los valores más grandes para los coeficientes de promedios y detalles (MAX-MAX). El resultado se muestra en la Figura 5-12(a).
- El esquema 2 utiliza la misma descomposición que el esquema 1, pero ahora se aplica el extractor de características con los siguientes parámetros: *connectivity* = **8**, *filter* = **Prewitt**; $\sigma = 0.4$, *low_threshold* = **0**, y *high_threshold* = **20**. El resto de coeficientes se seleccionaron con el criterio (MAX-MAX). El resultado se muestra en la Figura 5-12(b).
- En el esquema 3 se utilizó la wavelet Db3 con un nivel de descomposición. Se seleccionó el promedio de los coeficientes de aproximación y los coeficientes más altos de detalles (AV-MAX). El resultado aparece en la Figura 5-12(c).
- El esquema 4 utilizó la misma descomposición que el esquema 3. Se aplica el filtro de Canny con los siguientes parámetros: *connectivity* = **8**, *filter* = **Prewitt**; $\sigma = 0.6$, *low_threshold* = **-100**, y *high_threshold* = **0**. El resto de coeficientes se seleccionaron con el criterio (AV-MAX). El resultado se muestra en la Figura 5-12(d).

⁵ Coeficientes Daubechies 3 (Db3):

$a_0 = 0.3326, a_1 = 0.8068, a_2 = 0.4598, a_3 = -0.135, a_4 = -0.0854, a_5 = 0.0352$

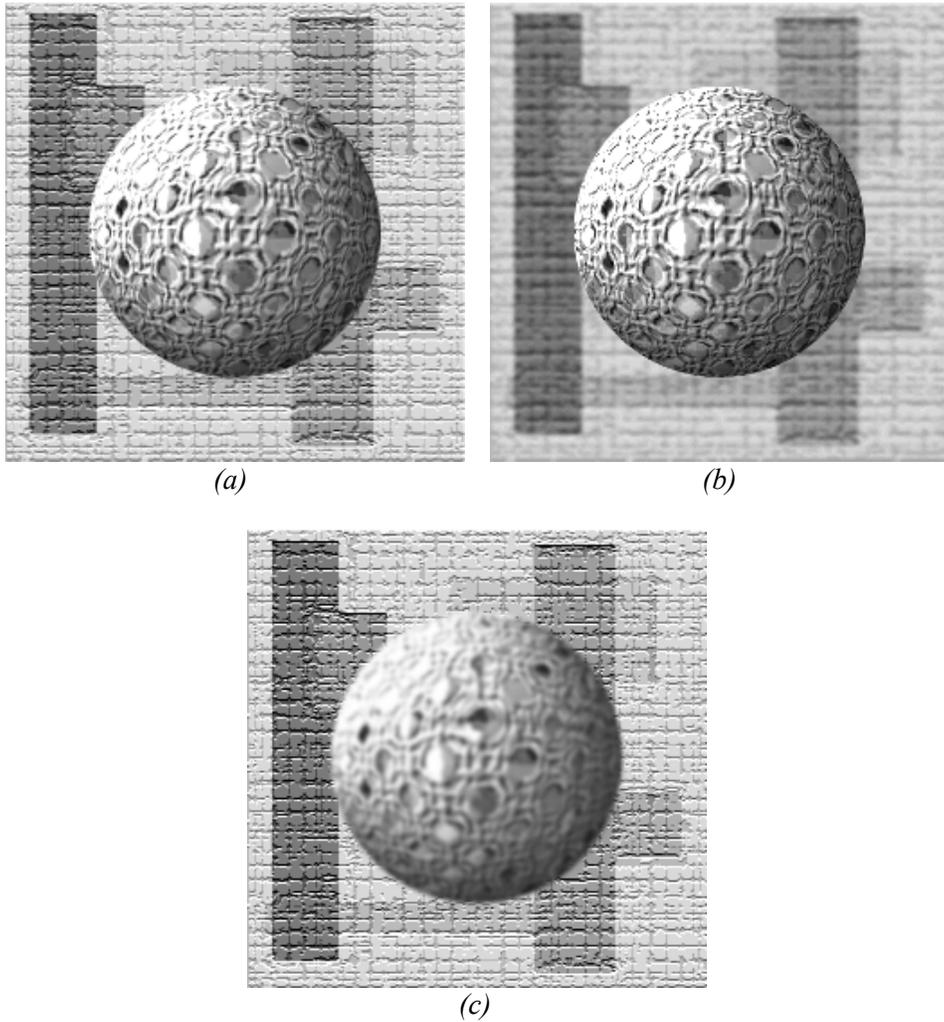


Figura 5-11 Imágenes de entrada multifocales. Pelota y fondo. (a) Imagen de referencia. (b) Imagen con el fondo desenfocado y el primer plano en foco. (c) Imagen con el fondo en foco y el primer plano desenfocado.

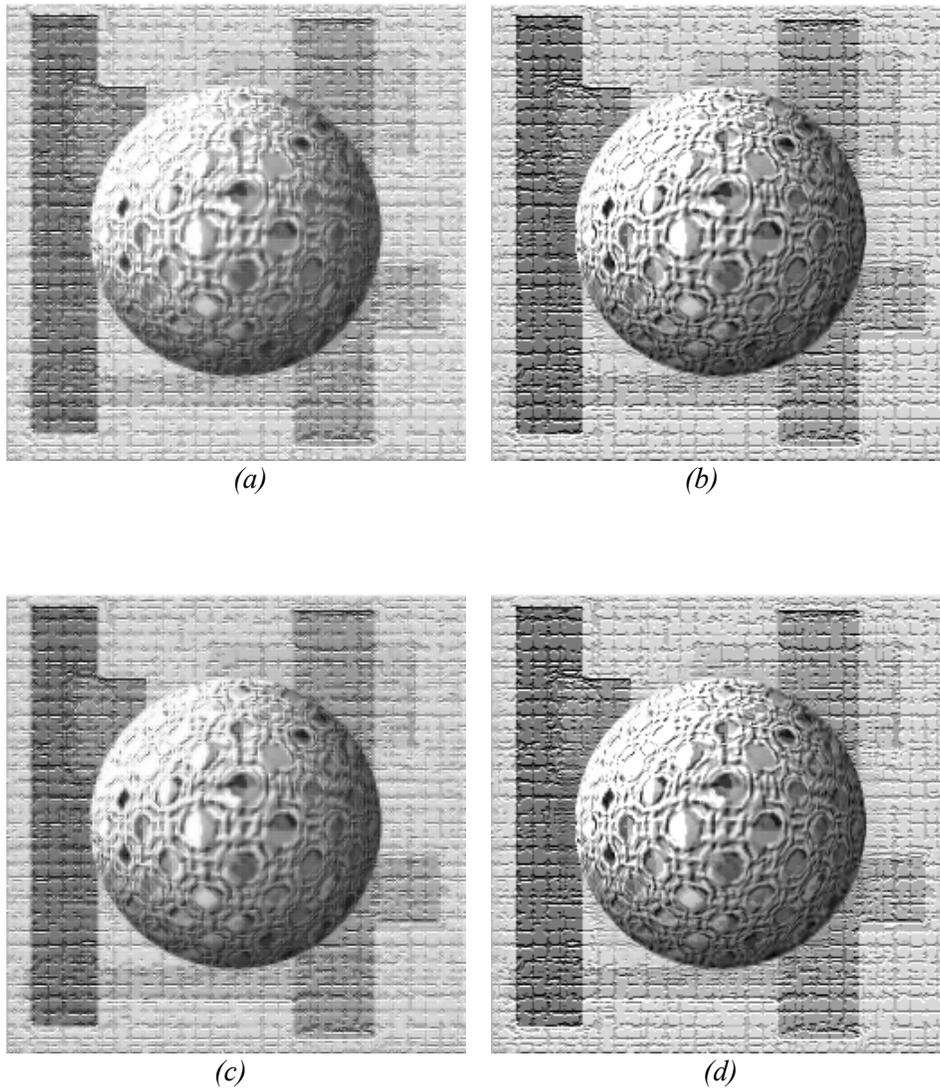


Figura 5-12 Resultado de los esquemas de fusión para el experimento Pelota y fondo. (a) Imagen obtenida aplicando Db3 a un nivel de descomposición y la regla de fusión MAX-MAX. (b) Imagen obtenida aplicando Db3 a un nivel de descomposición y el filtro de Canny, la regla de fusión es MAX-MAX. (c) Imagen obtenida aplicando Db3 a un nivel de descomposición y la regla de fusión AV-MAX. (d) Imagen obtenida usando Db3 a un nivel de descomposición y el filtro de Canny, la regla de fusión es AV-MAX.

La Figura 5-13 muestra los mapas de decisión de los experimentos de la imagen con la pelota y el fondo. Las Figuras 5-13(a) y 5-13(b) corresponden al esquema 2 y las Figuras 5-13(c) y 5-13(d) pertenecen al esquema 4.

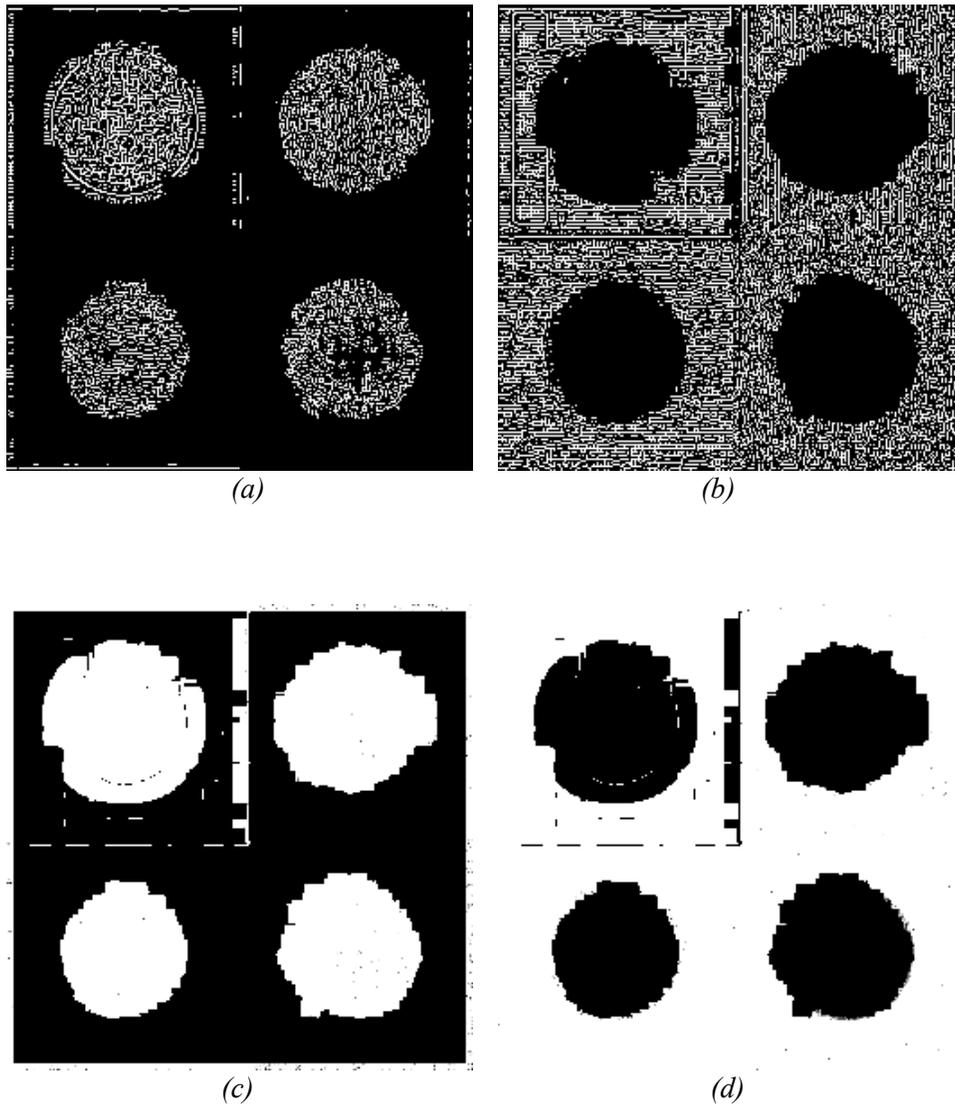


Figura 5-13 Mapas de decisión para el experimento Pelota y fondo. (a) y (b) Mapas de decisión del esquema 2. (c) y (d) Mapas de decisión del esquema 4.

Los resultados se muestran en la Tabla 5-3. El esquema 2 presenta mejores índices cuantitativos que su contraparte en el esquema 1, mientras que el esquema 4 tiene mejor desempeño que el esquema tradicional presentado en 3.

Fusión	PSNR [dB]	MSE	SNR _w	SSIM
#1	19.8449	1.0364e-2	10.7654	0.7217
#2	20.362	9.2003e-3	12.4006	0.8135
#3	20.2193	9.507e-3	11.2495	0.7347
#4	20.3561	9.2128e-3	12.4322	0.8096

Tabla 5-3 Resultados del desempeño de la fusión para el experimento **Pelota y fondo**.

5.2.2 Fusión de imágenes sin referencia.

Para el caso de la evaluación de la fusión de imágenes sin referencia, se utilizó una métrica basada en la entropía de Tsallis, presentada en (Nava, et al., 2007).

5.2.2.1 Relojes.

Para el siguiente experimento se utilizaron dos imágenes en escalada de grises de tamaño 512×512 . La Figura 5-14(a) tiene el reloj en primer plano desenfocado y el reloj de fondo enfocado y la Figura 5-14(b) de forma inversa. La descomposición se realizó con la wavelet Db2⁶ a una escala y se probaron dos reglas de fusión.

- En el esquema 1 se promediaron los coeficientes de aproximación y se mantuvieron los coeficientes de detalle más grandes para poder reconstruir la imagen (AV-MAX), Figura 5-14(c).
- En el esquema 2 se utilizó el filtro de Canny con los siguientes parámetros: *connectivity* = **8**, *filter* = **Prewitt**; $\sigma = 0.8$ *low_threshold* = **-20**, y *high_threshold* = **10**. El resto de coeficientes se seleccionaron con el criterio (AV-MAX), Figura 5-14(d).

En la Tabla 5-4 se muestran los resultados del experimento. De forma cualitativa se tienen mejores resultados con el algoritmo propuesto en esta tesis. La principal ventaja del método propuesto en esta tesis es que se conserva el contraste en la imagen. Aunque también se obtiene una mayor cantidad de información con el método propuesto, esta es muy pequeña, ya que la métrica depende del parámetro α de la entropía de Tsallis. En general, este parámetro se calcula de forma heurística y varía para cada imagen.

⁶ Coeficientes Daubechies 2 (Db2): $a_0 = \frac{1+\sqrt{3}}{4\sqrt{2}}$, $a_1 = \frac{3+\sqrt{3}}{4\sqrt{2}}$, $a_2 = \frac{3-\sqrt{3}}{4\sqrt{2}}$, $a_3 = \frac{1-\sqrt{3}}{4\sqrt{2}}$

Fusión	I(A, F)	I(B, F)	MI	NMI
#1	1.2388	1.3584	2.5472	0.4761
#2	1.3398	1.2702	2.61	0.4785

Tabla 5-4 Resultados del desempeño de la fusión para el experimento Relojes.

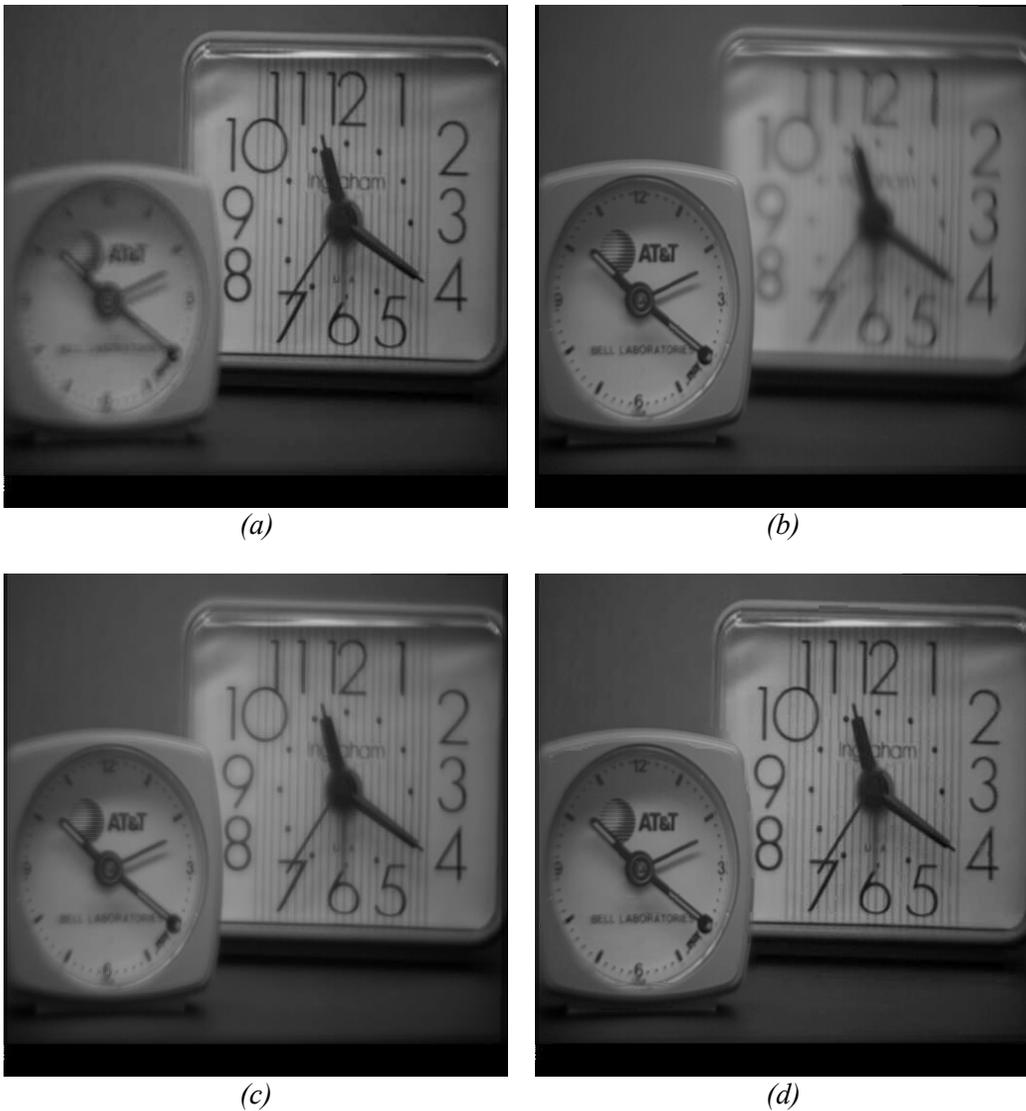


Figura 5-14 Experimento Relojes. (a) y (b) son las imágenes de entrada. (c) Imagen obtenida con Db2 a una escala y la regla de fusión AV-MAX. (d) Imagen obtenida con Db2 a un nivel de descomposición, aplicando Canny y la regla de fusión AV-MAX.

En las Figuras 5-15(a) y 5-15(b) se muestran los mapas de decisión del esquema 2.

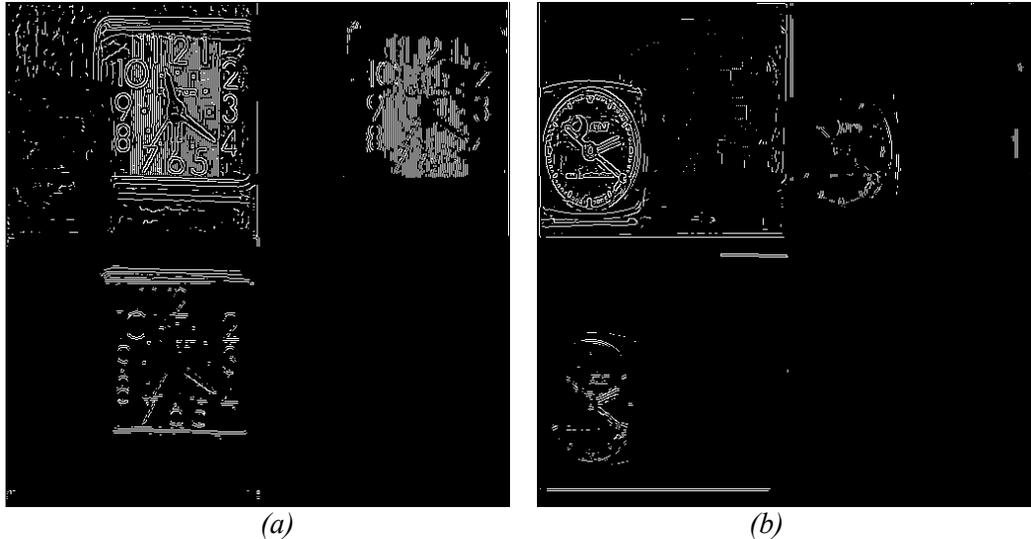


Figura 5-15 Mapas de decisión para el experimento Relojes. (a) y (b) Mapas de decisión del esquema 2.

5.2.2.2 Refresco.

Para este último experimento se utilizaron imágenes proporcionadas por el Dr. Shutao Li. Las imágenes son en escala de grises de tamaño 512×512 . La Figura 5-16(a) tiene la botella de refresco desenfocada y el fondo en foco, mientras que la Figura 5-16(b) tiene el fondo en foco y la botella de refresco borrosa. La descomposición se realizó con la wavelet Sym4 a una escala y se probaron dos reglas de fusión.

- En el esquema 1 se tomaron los valores más grandes para los coeficientes de detalles y aproximaciones. (MAX-MAX), Figura 5-16(c).
- En el esquema 2 se utilizó el filtro de Canny con los siguientes parámetros: *connectivity* = **8**, *filter* = **Prewitt**; $\sigma = 0.7$ *low_threshold* = **15**, y *high_threshold* = **25**. El resto de coeficientes se seleccionaron con el criterio (MAX-MAX), Figura 5-16(d).

Los resultados se presentan en la Tabla 5-5. Se puede apreciar de forma visual que el esquema 2 presenta una mejor calidad.

Fusión	I(A, F)	I(B, F)	MI	NMI
#1	2.2777	2.1813	4.459	0.3203
#2	2.3635	2.1557	4.519	0.3246

Tabla 5-5 Resultados del desempeño de la fusión para el experimento **Refresco**.

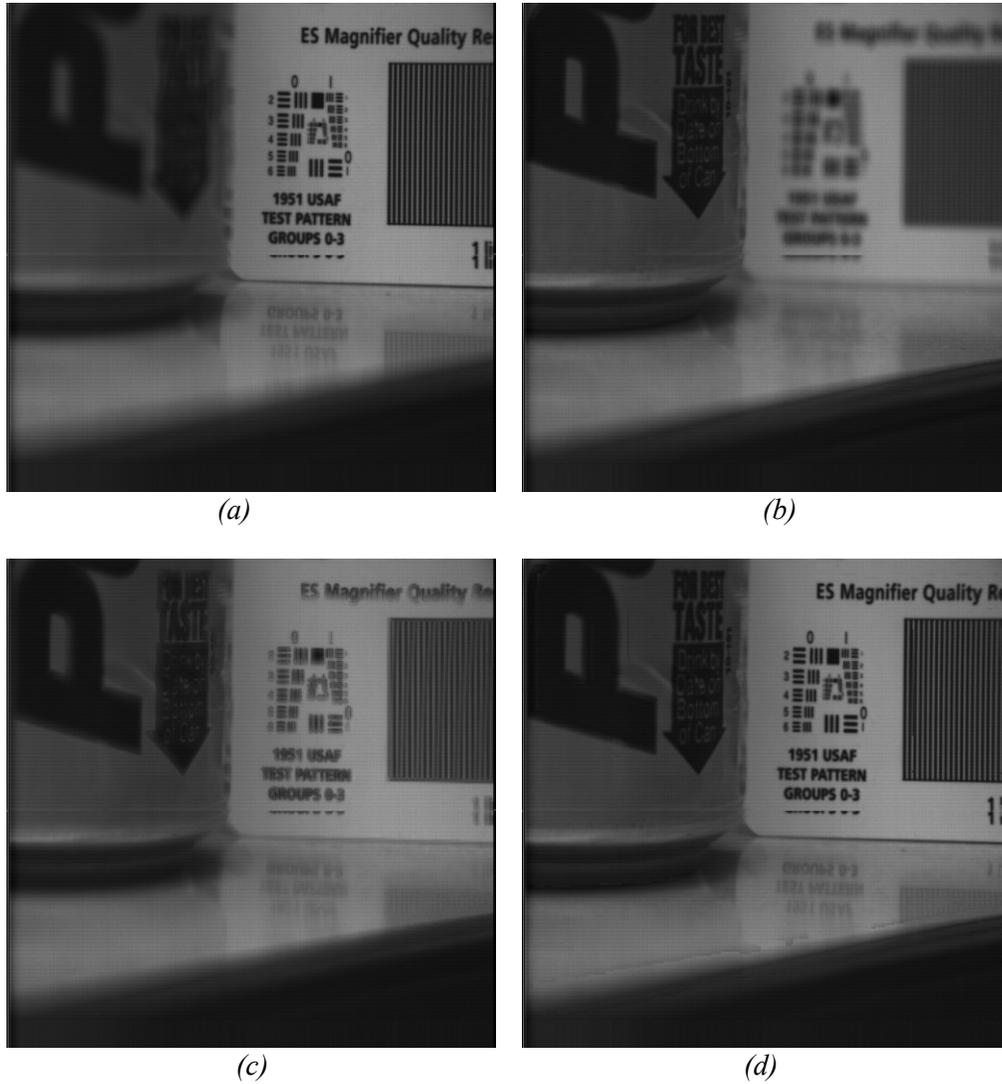


Figura 5-16 Experimento Refresco. (a) y (b) Imágenes de entrada. (c) Imagen obtenida con Sym4 y la regla de fusión MAX-MAX. (d) Imagen obtenida con Sym4 y el filtro de Canny, la regla de fusión es MAX-MAX.

5.3 *Referencias.*

- **Beghdadi A and Iordache R** Image quality assessment using the joint space/spatial-frequency representation [Article] // EURASIP Journal on Applied Signal Processing. - 2006. - Vol. 2006. - pp. Article ID 80537, 8 pages. - doi:10.1155/ASP/2006/80537.
- **Canny J** A Computational Approach to Edge Detection [Journal] // IEEE Transactions on Pattern Analysis and Machine Intelligence. - 1986. - 6 : Vols. PAMI-8. - pp. 679-698.
- **Dongmei Y and Zhongming Z** Wavelet decomposition applied to image fusion [Conference] // International Conferences on Info-tech and Info-net.. - Beijing, China : [s.n.], 2001. - Vol. 1. - pp. 291-295.
- **Li S, Kwok J and Wang Y** Combination of images with diverse focuses using the spatial frequency [Journal] // Information Fusion. - 2001. - 3 : Vol. 2. - pp. 169-176.
- **Macleod I** On finding structure in pictures [Book Section] // Picture Language Machines / book auth. Kaneff S. - [s.l.] : Academic Press, 1970.
- **Marr E and Hildreth E** Theory of Edge Detection [Journal] // Proceedings of the Royal Society of London. - 1980. - Vol. 207. - pp. 187-217.
- **Nava R, Cristóbal G and Escalante-Ramírez B** Non-reference image fusion evaluation procedure based on mutual information and a generalized entropy measure [Conference] // Bioengineered and Bioinspired Systems / ed. Lopez-Feliciano J, Sarmiento R and Kang S. - Maspalomas, España : SPIE, 2007. - Vol. 6592.
- **Pinho A and Almeida L** A review on edge detection based on filtering and differentiation [Journal] // Electrónica e Telecomunicacoes. - 1997. - 1 : Vol. 2. - pp. 113-126.
- **Prewitt J** Object enhancement and extraction [Book Section] // Picture Processing and Psychopictorics / ed. Lipkin B and Rosenfeld A. - New York, E. U. : Academic Press, 1970.
- **Sobel I** An isotropic 3x3 image gradient operator [Book Section] // Machine Vision for Three-Dimensional Scenes. - New York, E. U. : Academic Press, 1990.
- **Tolhurst D** On the possible existence of edge-detector neurones in the human visual cortex [Journal] // Vision Research. - 1972. - Vol. 12. - pp. 797-804.
- **Wang Z and Bovik A** A universal image quality index [Journal] // IEEE Signal Processing Letters. - 2002. - 3 : Vol. 9. - pp. 81-84.

Conclusiones.

"If you think dogs can't count, try putting three dog biscuits in your pocket and then giving Goofy only two of them"
Phil Pastoret

5.1 Conclusiones.

El objetivo fundamental de este trabajo de tesis ha sido proponer un algoritmo de fusión de imágenes que esté relacionado con la forma en que trabaja el sistema de visión humano (Capítulo 5). La evidencia confirma que el sistema de visión utiliza funciones cognitivas basadas en características para reconocer objetos, por lo que el algoritmo propuesto está dirigido a preservar las características visuales más importantes. En este sentido, se implementó el filtro de Canny como extractor de bordes que se conservan a través del proceso de fusión.

Con el propósito de comparar el algoritmo con las reglas de fusión tradicionales, fue desarrollada una herramienta para JAVA que permite evaluar diversas reglas de fusión utilizando la transformada Wavelet Discreta (Capítulo 3). Otra contribución importante en esta tesis se hizo en el área de la evaluación automática de la calidad de las imágenes (Capítulo 4). Fue implementada una herramienta en MATLAB para la evaluación de la calidad perceptiva de imágenes monocromáticas que proporcionar tres tipos de métricas: objetivas, híbridas y perceptivas e incorporar algoritmos de compresión, marcado de agua y de adición de ruido para su validación. Además se propuso una métrica para evaluar los algoritmos de fusión basada en la entropía, sin la necesidad de una imagen de referencia o *ground truth*. Los resultados obtenidos en este trabajo han sido discutidos a lo largo de los capítulos previos y basándonos en los resultados experimentales se puede afirmar que el algoritmo propuesto resulta en una mejor calidad de la imagen tanto de forma cualitativa como cuantitativa.

El trabajo desarrollado en esta tesis fue dividido en tres partes. En la primera parte (Capítulo 1 y 2) se presentó el análisis multirresolución. Se revisaron los algoritmos existentes en la literatura para descomposición a múltiples escalas de imágenes y se abordó ampliamente el desarrollo de la teoría

wavelet, primero para el caso unidimensional y después se extendió para imágenes, discutiendo el problema de la reconstrucción perfecta utilizando filtros espejo en cuadratura. En la última parte se revisó el problema del efecto de los bordes y la invariancia de las wavelets.

La segunda parte de esta tesis (Capítulo 3) presenta una justificación y las ventajas de la fusión de datos, se analizan y comparan los diferentes esquemas de fusión y se explica el caso del análisis por componentes principales.

La tercera parte de este trabajo (Capítulos 4) está dedicada a la evaluación de la calidad en las imágenes. Primero se analizó el estado del arte de las métricas de calidad y se estableció la importancia de proporcionar un método que evalúe de forma automática la calidad de las imágenes. Se revisaron las propiedades de la visión como la adaptación a la luminancia, el contraste, la sensibilidad y el efecto del enmascaramiento con el fin de introducirlas en las métricas de evaluación.

En el Capítulo 5 se presentó la regla de fusión y se mostró el extractor de características basado en el filtro de Canny. En la parte final del capítulo se realizaron experimentos para comparar el rendimiento del algoritmo de fusión propuesto contra los métodos tradicionales. Las pruebas sugieren una mejora en el contraste y una mejor calidad para el observador, además de que se obtienen mejores valores de la relación señal a ruido y menores niveles del error cuadrático medio. Finalmente, con las métricas de calidad perceptiva también se obtienen mejores resultados cuantitativos. Aunque en el caso de la métrica sin referencia, los resultados visuales obtenidos con el algoritmo propuesto son más agradables al observador, en términos de información, la mejora es muy pequeña. Esto se debe a que la entropía de Tsallis depende de un parámetro α que varía para cada tipo de imagen.

5.2 Trabajo futuro.

Este trabajo de tesis deja varias líneas de investigación abiertas. Por un lado, el uso de wavelets ortogonales para la descomposición multirresolución limita el esquema de fusión. Bajo este esquema, no se aprovecha completamente la redundancia en los datos, ya que siempre se cae en el caso de decidir sobre un conjunto de píxeles el más adecuado, desechando el resto. Trabajos futuros deberían aprovechar la sobrerepresentación de esquemas más poderosos como las wavelets “sobrecompletas de Gabor” (*overcomplete Gabor wavelets*) e incluir esquemas de multirresolución direccionales. Recordando que las wavelets solo descomponen en tres direcciones, vertical, horizontal y a 45 grados.

En este trabajo se propone el filtro de Canny como extractor de características, sin embargo es limitado y presenta el problema de doble respuesta a bordes y el problema de tratar a las texturas como ruido. Además de que depende de muchos parámetros que se modifican dependiendo del tipo de imágenes a fusionar. Esquemas futuros deberán extraer características más complejas como patrones, texturas y objetos, haciendo que no se dependa de un conocimiento previo de las imágenes.

Finalmente, una posible modificación a este algoritmo es tratar de manejar distintos valores de umbrales para las imágenes de entrada. Como se escribe líneas arriba, la extracción de características depende de varios parámetros, entre ellos varianza y los umbrales de histéresis, que se fijaron para ambas imágenes de entrada.