



ESTUDIO DE MÉTODOS DE INDEXACIÓN Y RECUPERACIÓN EN BASES DE DATOS DE IMÁGENES

MEMORIA PRESENTADA POR
José Orlando Maldonado Bautista

PARA OPTAR AL GRADO DE DOCTOR EN
INFORMATICA EN LA
UNIVERSIDAD DEL PAIS VASCO

DIRECTOR
Manuel Graña Romay

DEPARTAMENTO DE CIENCIAS DE LA COMPUTACIÓN E
INTELIGENCIA ARTIFICIAL

FACULTAD DE INFORMATICA

UPV/EHU

San Sebastián, abril 2008

Resumen

La tesis aborda el problema del acceso a bases de datos de imágenes basado en índices calculados a partir del contenido de la propia imagen, conocido por sus siglas en inglés Content Based Image Retrieval (CBIR). Realiza una revisión del estado del arte en este tema. Se realiza también una revisión de los principales resultados y conceptos relativos a Bancos de Filtros de Gabor y Transformada Discreta en Wavelets, dos técnicas de análisis de la imagen muy extendidas y que son instrumentales en una de las aplicaciones desarrolladas. La tesis contiene dos casos especiales de sistemas CBIR. El primero es un sistema de indexación de imágenes de papel reciclado. Además de servir para el acceso basado en contenidos, esta indexación puede ser utilizada para sistemas de control de calidad en la fabricación de papel reciclado. El segundo caso de estos sistemas trata con imágenes hiperespectrales de reconocimiento remoto. Se propone una medida de similitud espectral basada en los endmembers obtenidos mediante la aplicación de Memorias Autoasociativas Morfológicas para la detección de la independencia morfológica. Se proporcionan resultados experimentales de rendimiento de la recuperación calculados sobre bases de datos de imágenes sintéticas.

Agradecimientos

Deseo agradecer en primer lugar a la Universidad del País Vasco y la Universidad de Pamplona (Colombia), las cuales me han permitido iniciar y llevar a buen término mi formación doctoral. A mi tutor y director de tesis, DR. DON MANUEL GRAÑA ROMAY por su invaluable apoyo y colaboración desde el comienzo de mi doctorado, así como por su paciencia e influencia en todos los capítulos de este proceso. Estoy seguro que su contagioso entusiasmo por los procesos académicos y científicos me han marcado una ruta hacia nuevos y fructíferos desafíos. A los profesores de la Facultad de Informática y en especial del Departamento de Ciencias de la Computación e Inteligencia Artificial que de uno u otro modo han aportado a mi proceso formativo. A los miembros del Grupo de Inteligencia Computacional, que conforman un caudal creciente de conocimiento presto a aportar al desarrollo de todos sus miembros. A mis compañeros de Laboratorio: Elsa Fernández, Abdel Moujahid, Maite García, Miguel Veganzones, Ramón Moreno, Flavio Banterla, Alexandre Savio e Iván Villaverde, con quienes he compartido gratamente durante estos más de cinco años. También un agradecimiento especial a la secretaria Administrativa del Dpto. CCIA, Doña Elena Bidondo, por su ayuda y colaboración desde el momento mismo de mi llegada.

A mi familia: mis padres Rosa Bautista y José Maldonado, hermanos, sobrinos y demás, que han estado conmigo en todo momento. Sin su respaldo, nada de esto habría sido posible. A mis amigos, que aquí me han dado apoyo, compañía y buenos momentos, y que en Colombia siempre me han animado para continuar adelante, gracias por estar ahí, a pesar de la distancia.

A cada una de las personas que de algún modo han contribuido para el desarrollo y culminación de esta tesis, mi más sincero agradecimiento.

San Sebastián, Abril 21 de 2008.

Índice general

| | |
|--|-----------|
| 1. Introducción | 19 |
| 1.1. Motivación | 19 |
| 1.1.1. Imágenes de papel reciclado | 20 |
| 1.1.2. Imágenes de reconocimiento remoto | 21 |
| 1.2. Contribuciones más relevantes | 22 |
| 1.3. Objetivos de la tesis doctoral | 23 |
| 1.4. Publicaciones realizadas | 24 |
| 1.5. Publicaciones submitidas | 25 |
| 1.6. Estructura de la memoria de la tesis | 25 |
| | |
| 2. Recuperación de imágenes basada en contenidos | 29 |
| 2.1. Introducción | 29 |
| 2.2. Aspectos generales | 32 |
| 2.3. Ejemplos de Sistemas CBIR | 35 |
| 2.4. Arquitectura de los sistemas CBIR | 36 |
| 2.5. Extracción de características | 39 |
| 2.5.1. Características de Textura | 39 |
| 2.5.2. Características de Color | 41 |
| 2.5.3. Características de formas | 42 |
| 2.5.4. Relaciones espaciales de regiones y puntos de interés | 46 |
| 2.6. Métricas y funciones de similitud | 46 |
| 2.7. Métodos de acceso y búsqueda en base de datos | 49 |
| 2.8. Aprendizaje y realimentación por relevancia | 52 |
| 2.9. Evaluación en los sistemas CBIR | 53 |
| 2.10. Conclusiones | 56 |

| | |
|--|-----------|
| 3. Filtros de Gabor | 59 |
| 3.1. La transformada de Gabor | 60 |
| 3.2. Funciones de Gabor 2D | 62 |
| 3.3. Bancos de Filtros de Gabor | 65 |
| 4. Análisis Wavelet | 69 |
| 4.1. Wavelets Continuas | 69 |
| 4.1.1. Definición de wavelet | 69 |
| 4.1.2. Ejemplos de wavelets | 70 |
| 4.1.2.1. Wavelet de Haar | 70 |
| 4.1.2.2. Wavelet de Shannon | 71 |
| 4.1.2.3. Wavelet de Morlet | 72 |
| 4.1.3. La Transformada Wavelet Continua | 72 |
| 4.1.3.1. Definición | 73 |
| 4.1.3.2. Transformada Wavelet Inversa | 74 |
| 4.1.3.3. Propiedades | 75 |
| 4.2. La Transformada Wavelet Discreta | 76 |
| 4.2.1. Análisis Multiresolución | 78 |
| 4.2.2. Bases de wavelets ortonormales en el análisis multires- olución | 79 |
| 4.2.3. Algoritmo de descomposición piramidal | 80 |
| 4.3. Transformada Wavelet en dos dimensiones | 83 |
| 5. Control de calidad del papel reciclado | 87 |
| 5.1. Descripción del problema | 88 |
| 5.2. Consideraciones metodológicas | 91 |
| 5.3. Adquisición de las imágenes | 94 |
| 5.4. Etiquetado manual | 95 |
| 5.5. Definición de características | 99 |
| 5.5.1. Características basadas en Bancos del Filtros de Gabor | 100 |
| 5.5.2. Características basadas en coeficientes de la TWD . . . | 101 |
| 5.6. Clasificación automática | 102 |
| 5.6.1. Resultados con las características extraídas mediante BFG | 103 |
| 5.6.2. Resultados con las características extraídas mediante coeficientes wavelet | 104 |
| 5.6.3. Conclusiones de la clasificación | 108 |
| 5.7. Establecimiento de un índice de abollado | 108 |

| | |
|---|------------|
| 5.7.1. Validación por ordenación de pares | 111 |
| 5.8. Conclusiones | 117 |
| 6. Sistema CBIR para imágenes hiperespectrales | 119 |
| 6.1. imágenes hiperespectrales | 119 |
| 6.2. Descomposición espectral | 120 |
| 6.3. Algoritmos de extracción de endmembers | 123 |
| 6.4. Redes morfológicas e independencia morfológica | 125 |
| 6.4.1. Breve revisión de fundamentos | 126 |
| 6.4.2. Algoritmo heurístico de inducción de endmembers | 128 |
| 6.5. Resultados de segmentación | 130 |
| 6.6. Distancia entre imágenes hiperespectrales | 137 |
| 6.7. Resultados sobre bases de datos de imágenes sintéticas | 138 |
| 6.7.1. Experimento 1 | 138 |
| 6.7.2. Experimento 2 | 141 |
| 6.8. Conclusiones | 145 |
| 7. Conclusiones y líneas de trabajo futuro | 147 |
| 7.1. Caracterización visual de la calidad del papel | 147 |
| 7.2. Sistemas CBIR de imágenes hiperespectrales | 148 |
| 7.3. Aplicaciones de Lattice Computing a CBIR | 148 |
| A. Ejemplos Ilustrativos | 151 |
| A.1. Ejemplos ilustrativos sobre características de texturas basados en BFG | 151 |
| A.2. Ejemplos ilustrativos sobre características de texturas basados en la TWD | 159 |
| Bibliografía | 167 |

Lista de algoritmos

1. Algoritmo heurístico de inducción de los endmembers . . . 129

Índice de figuras

| | |
|---|----|
| 2.1. Esquema de una consulta mediante ejemplo, en un sistema CBIR. | 31 |
| 2.2. Agrupación de funcionalidades en módulos de un sistema CBIR | 38 |
| 3.1. Filtro de Gabor en el dominio espacial. a) Partes real e imaginaria de la sinusoidal compleja. b) Gaussiana rotada sobre el origen c) Máscaras formadas por las funciones sinusoidales real e imaginaria moduladas por la gaussiana. | 63 |
| 3.2. Elipse de puntos con respuesta igual a la mitad de la magnitud en el dominio de la frecuencia | 65 |
| 3.3. Recubrimiento del plano de Fourier por un Banco de Filtros de Gabor. a) Sin solapamiento en la respuesta de media-magnitud, b) Con solapamiento en la respuesta de magnitud media. | 67 |
| 4.1. Wavelet de Haar | 71 |
| 4.2. Función wavelet de Shanon | 72 |
| 4.3. Parte real de la Wavelet de Morlet | 73 |
| 4.4. Descomposición del plano mediante la discretización por muestro de la malla diádica | 77 |
| 4.5. Esquema para un nivel descomposición multiresolución de la imagen mediante el algoritmo piramidal | 84 |
| 4.6. Representación de una imagen con tres niveles de descomposición | 85 |
| 4.7. Esquema para la reconstrucción de una imagen multiresolución mediante el algoritmo piramidal | 86 |
| 5.1. Ejemplos de imágenes de papel reciclado con abollado. Las imágenes han sido preprocesadas para mejorar el contraste. . . | 92 |

| | | |
|-------|---|-----|
| 5.2. | Histogramas generados con los valores obtenidos a partir de la función discriminante de Fisher aplicada a los vectores generados con el BFG sin solapamiento. | 113 |
| 5.3. | Histogramas generados con los valores obtenidos a partir de la función discriminante de Fisher aplicada a los vectores generados con el BFG con solapamiento. | 114 |
| 5.4. | Histogramas generados con los valores obtenidos a partir de la función discriminante de Fisher aplicada a los vectores generados mediante los coeficientes de la TWD con wavelet madre Haar en los niveles 1-3. | 115 |
| 5.5. | Histogramas generados con los valores obtenidos a partir de la función discriminante de Fisher aplicada a los vectores generados mediante los coeficientes de la TWD con wavelet madre Haar en los niveles 4-6. | 116 |
| 6.1. | Ilustración de la captura de la imagen hiperespectral | 121 |
| 6.2. | Ilustración del cubo de datos que constituye una imagen hiperespectral | 121 |
| 6.3. | Ilustración de las causas de la mezcla lineal espectral | 122 |
| 6.4. | Endmembers encontrados por nuestro algoritmo heurístico basado en independencia morfológica sobre la imagen de Washington D.C. Mall | 131 |
| 6.5. | Imágenes de abundancia calculadas usando los endmembers de la figura 6.4 | 132 |
| 6.6. | Indian Pines 1992, verdad del terreno | 134 |
| 6.7. | Endmembers encontrados por el algoritmo heurístico de la sección 6.4.2 en la imagen Indian Pines | 134 |
| 6.8. | Abundancias calculadas usando los endmembers en la figura 6.7 | 135 |
| 6.9. | Resultado de la clasificación supervisada presentada en [150, 149] | 136 |
| 6.10. | Espectros de repositorio de la USGS usados como endmembers de la verdad del terreno en el primer experimento. | 139 |
| 6.11. | Un ejemplo de la interface de un sistema CBIR para imágenes hiperespectrales: una consulta y sus imágenes más cercanas de acuerdo al conjunto de endmembers inducido. | 141 |
| 6.12. | Endmembers verdad del terreno utilizados para el segundo experimento | 142 |

| | |
|--|-----|
| 6.13. Una instancia de las imágenes de abundancia generadas como campos basados en polinomios de Legendre para una imagen con cinco endmembers. | 143 |
| A.1. Imágenes construidas con orientación y frecuencia espacial específicas que han sido definidas para probar la respuesta de cada filtro, las frecuencias en las barras de cada imagen son de 1/4, 1/8, 1/16 y 1/32 ciclos/pixel, con orientaciones de 0, 45, 90 y 135 grados. | 152 |
| A.2. BFG sintonizado con los parámetros $F = 1/4, 1/8, 1/16, 1/32$ ciclos/pixel, y $\theta = 0^\circ, 45^\circ, 90^\circ, 135^\circ$ | 153 |
| A.3. Respuestas del FG con parámetros $\theta = 0^\circ, 45^\circ, 90^\circ, 135^\circ, F = 1/4$ ciclos/pixel | 154 |
| A.4. Respuestas del FG con parámetros $\theta = 0^\circ, 45^\circ, 90^\circ, 135^\circ, F = 1/8$ ciclos/pixel. | 155 |
| A.5. Respuestas del FG con parámetros $\theta = 0^\circ, 45^\circ, 90^\circ, 135^\circ, F = 1/16$ ciclos/pixel. | 156 |
| A.6. Respuestas del FG con parámetros $\theta = 0^\circ, 45^\circ, 90^\circ, 135^\circ, F = 1/32$ ciclos/pixel. | 157 |
| A.7. Imágenes de texturas con incrementos en los detalles y la variación en la frecuencia espacial de los niveles de gris. | 157 |
| A.8. Norma del vector de características obtenido a partir de BFG para cada una de las imágenes de la figura A.7. | 158 |
| A.9. Energía calculada para el primer nivel de descomposición wavelet. | 160 |
| A.10. Energía calculada para el segundo nivel de descomposición wavelet. | 161 |
| A.11. Energía calculada para el tercer nivel de descomposición wavelet. | 162 |
| A.12. Energía calculada para el cuarto nivel de descomposición wavelet. | 163 |
| A.13. Energía calculada para el quinto nivel de descomposición wavelet. | 164 |
| A.14. Norma del vector de características basado en los coeficientes de los detalles de la descomposición wavelet, calculado para las imágenes de la figura A.7 | 165 |

Índice de cuadros

| | |
|---|-----|
| 2.1. Métricas utilizadas para calcular la similitud de características en sistemas CBIR | 47 |
| 4.1. Filtros paso bajo y paso alto para las transformadas wavelet directa e inversa | 85 |
| 5.1. Comparación de la primera clasificación realizada por diferentes observadores, mediante múltiples matrices de confusión. | 96 |
| 5.2. Comparación de la segunda clasificación realizada por diferentes observadores, mediante múltiples matrices de confusión. | 97 |
| 5.3. Comparación de la primera y segunda clasificación realizada por los diferentes observadores, mediante múltiples matrices de confusión. | 97 |
| 5.4. La tabla muestra la concordancia entre cada observador en la primera evaluación. | 98 |
| 5.5. La tabla muestra la concordancia entre cada observador en la segunda evaluación. | 98 |
| 5.6. La tabla muestra la concordancia entre cada observador en la primera y la segunda evaluación | 99 |
| 5.7. Resultados iniciales de la clasificación con características basadas en BFG con y sin solapamiento en los campos receptivos. . . | 104 |
| 5.8. Resultados de la clasificación de los vectores de características basados en BFG utilizando Weka. | 104 |
| 5.9. Exito en la clasificación mediante el algoritmo k-NN con diferentes niveles de descomposición y wavelets madre db1 a db4. . | 105 |
| 5.10. Exito en la clasificación mediante el algoritmo k-NN con diferentes niveles de descomposición y wavelets madre db5 a db8. . | 106 |

| | |
|---|-----|
| 5.11. Exito en la clasificación mediante la red neuronal MLP con diferentes niveles de descomposición y diferentes wavelets madre | 107 |
| 5.12. Resultados de la clasificación de los vectores de características basados en los coeficientes de la TWD mediante Weka. | 108 |
| 5.13. Clasificación mediante la norma de los vectores de características | 109 |
| 5.14. Clasificación mediante los componentes principales calculados a cada | 110 |
| 5.15. Resultados de la clasificación realizada sobre los valores obtenidos mediante la aplicación de la función discriminante de Fisher a los vectores de características basados en BFG y coeficientes wavelets. | 111 |
| 5.16. Correspondencia entre el índice de abollado basado en la función discriminante de Fisher y la apreciación de los expertos. . | 112 |
| 6.1. Resultados de relevancia en las respuestas a las consultas sobre la base de datos de 400 imágenes sintéticas, usando la distancia definida en la ecuación 6.8 | 142 |
| 6.2. Resultados de relevancia basada en la distancia entre imágenes de abundancia para imágenes con abundancias generadas con polinomios de Legendre 2D. | 145 |

Capítulo 1

Introducción

Damos en este primer capítulo una breve motivación de los trabajos realizados. Indicamos las contribuciones que consideramos más relevantes al estado del arte. Tras enumerar las publicaciones realizadas y en preparación, presentamos la estructura de la tesis.

1.1. Motivación

El área temática cubierta por esta tesis, el acceso a bases de datos de imágenes basado en contenidos (CBIR en inglés) es un campo fértil de aplicaciones y desarrollo de herramientas. Los primeros sistemas se dirigían a colecciones muy heterogéneas, con pretensión de universalidad. La identificación del problema de la brecha semántica (*semantic gap*) dió mayor valor a los sistemas con dominio de imágenes restringido. Se desarrollaron sistemas para el acceso basado en contenidos a colecciones de imágenes de caras, de imágenes médicas, de imágenes astronómicas, y un largo etcétera. Estas aplicaciones se caracterizan porque el proceso de extracción de características es muy preciso y dirigido por la aplicación, también la distancia empleada como medida de similitud está definida muy precisamente. Así, los sistemas de CBIR de caras tempranos proponían la transformación en componentes principales (PCA en inglés) como proceso de extracción de características y la distancia Euclídea como medida de disimilitud. Otro ejemplo: los sistemas de CBIR sobre colecciones de imágenes médicas (i.e. MRI del cerebro) usan como proceso de extracción de características procesos de registro no lineal que miden las deformaciones entre las imágenes. Sobre éstas características las medidas de similitud pueden venir dadas, por ejemplo, por la transformada PCA del campo de deformación y la distancia Euclídea sobre los vectores así

obtenidos. Los trabajos de esta tesis se enmarcan en esta tendencia, puesto que los dos sistemas propuestos tienen un dominio de definición muy preciso y acotado, con procesos de extracción de características específicos y medidas de similitud o índices bien definidas. El *semantic gap* produjo también considerables esfuerzos en la definición de sistemas de realimentación de las consultas que permitirían acotar la semántica implícita en la consulta, interactuando con el usuario. En nuestro trabajo no hemos tocado estos temas de ninguna manera, puesto que al ser sistemas de semántica muy reducida no tiene interés el modelado semántico.

1.1.1. Imágenes de papel reciclado

El primer problema tratado en esta tesis, la caracterización e indexación de imágenes de papel reciclado con propósitos de control de calidad y realimentación a los procesos de producción expuesta en el capítulo 5, surge por la solicitud de una empresa del sector (Echezarreta SA que luego pasó a ser Paperalia SA) que trataba de caracterizar la calidad del papel en términos de un fenómeno que no está recogido en los estándares de calidad de papel: la formación de ondulaciones y abolladuras en el papel al cabo de un tiempo de almacenamiento o al pasar por condiciones especiales (ej. el recalentamiento en una fotocopidora). Esta relación se concretó en dos proyectos, uno con financiación del Gobierno Vasco en el programa Universidad-Empresa, y otro con financiación del Ministerio de Ciencia y Tecnología, en los cuales ha participado el doctorando. Parte de los resultados obtenidos es precisamente un sistema CBIR de gestión de las imágenes obtenidas por el departamento de calidad y que se instaló en la empresa. Los trabajos teóricos que se elaboraron a raíz de esta aplicación son los que se describen con detalle en el capítulo 5.

El problema del abollado es un excelente ejemplo de aplicación de las técnicas de caracterización de textura y nos ha dado la oportunidad de profundizar sobre ellas. Entre las distintas aproximaciones, como las matrices de coocurrencia, escogimos las aproximaciones basadas en Filtros de Gabor y Wavelets por que ofrecen la metodología más general y más sistemática. Las matrices de coocurrencia, por ejemplo, necesitan de bastantes intentos de prueba y error para fijar parámetros tales como la distancia entre píxeles considerada en cada tabla o las transformaciones adecuadas de la tabla para obtener características discriminantes.

El problema del abollado, además, tiene una característica específica, que se trata de caracterizar la textura global de la imagen. Esto se diferencia

mucho de las aplicaciones que realizan la segmentación de la imagen basada en texturas, puesto que la caracterización es local y homogénea, mientras que en una imagen pueden convivir varias texturas y no es evidente como definir un abollado global, ni siquiera cualitativamente. Parte de las dificultades encontradas se debe precisamente a la necesidad de inventar el método de etiquetado de las imágenes.

1.1.2. Imágenes de reconocimiento remoto

Hasta el momento, la literatura de sistemas CBIR para colecciones de reconocimiento remoto, que incluye las imágenes hiperespectrales, se basa en la extracción de características espaciales. Sin embargo, los sensores de alta resolución espectral dan información adicional que permitiría caracterizar las imágenes por los espectros de los elementos presentes en la escena, para distinguir imágenes con distribuciones espaciales similares de materiales distintos que tienen respuesta similar en sensores de baja resolución espectral (RGB o pancromático). Es por ello que parece deseable definir índices de las imágenes basados en la información espectral. También parece deseable que el proceso de indexación sea eficiente, aplicable en un tiempo razonable y con recursos razonables a un conjunto de imágenes relativamente grande.

Nuestro punto de partida ha sido el trabajo realizado en el grupo sobre la segmentación no supervisada de imágenes hiperespectrales usando redes neuronales morfológicas, más precisamente Memorias Autoasociativas Morfológicas. Esta segmentación se realiza mediante (1) la inducción de los endmembers a partir de la imagen hiperespectral y (2) el cálculo de las imágenes de abundancia, que nos dan la segmentación deseada. Las técnicas propuestas son no supervisadas y relativamente eficientes en términos computacionales. Su mayor inconveniente radica en su componente aleatorio, esto es, distintas ejecuciones del mismo algoritmo pueden dar distintas segmentaciones y distintas caracterizaciones espectrales si se parte de condiciones iniciales (i.e. endmember inicial aleatorio). Si fijamos el proceso tomando siempre el mismo endmember inicial (i.e. el primer pixel) el proceso es completamente determinista.

La aplicación al CBIR de imágenes hiperespectrales de estas técnicas supone una extensión a un dominio de aplicación nuevo y muy extenso. Además supone la prueba de los algoritmos sobre un conjunto de datos amplio (las imágenes sintéticas ocupan medio terabyte en formato binario).

1.2. Contribuciones más relevantes

Las aportaciones más relevantes de la presente tesis se encuentran en los capítulos 5 y 6. En ellos se realizan aplicaciones originales y se aportan métodos y técnicas novedosos. El capítulo 2 proporciona una revisión del estado del arte en sistemas CBIR que trata de situar las contribuciones de esta tesis en su marco justo. Los capítulos 3 y 4, y el apéndice dado por el capítulo 8, proporcionan una revisión de conceptos sobre Bancos de Filtros de Gabor y Transformada Discreta en Wavelets que son imprescindibles para la comprensión completa del trabajo expuesto en el capítulo 5, sin embargo pueden considerarse desde un punto de vista didáctico puesto que organizan los conceptos de forma sintética y muy accesible.

En el capítulo 5 describimos la aplicación de características de textura a la identificación de la calidad de imágenes de papel reciclado. El fenómeno de interés es la aparición de ondulaciones en la superficie del papel. A falta de un nombre estándar lo denominamos “abollado”. Esta aplicación es completamente nueva en la literatura conocida. Entre los precedentes, se encuentran aplicaciones de proceso de imágenes a nivel microscópico para tratar de determinar algunas propiedades del papel. Nuestras imágenes son macroscópicas y el tipo de estructuras que se buscan en las imágenes microscópicas no se encuentran en ellas. También existen algunas aplicaciones de análisis de imágenes para el estudio de fenómenos como el punto de ruptura del papel sujeto a tracción. En este caso el tipo de imágenes no tiene ninguna similitud con las nuestras y las técnicas de análisis son radicalmente distintas. Por último, algunos fenómenos de entintado del papel pueden ser similares en el aspecto general de las imágenes, sin embargo la observación detallada encuentra fuertes diferencias entre los tipos de imágenes que se producen en los dos casos.

La novedad del propio planteamiento del problema de cuantificación visual del abollado, hace que tenga especial interés el proceso de etiquetado que se realiza de forma rigurosa por observadores humanos independientes y se establecen las categorías mediante votación. Encontramos resultados relevantes sobre la concordancia de diversos observadores humanos de este fenómeno que son completamente novedosos. La selección de las características mediante la realización de procesos de clasificación automática supervisada da pie a la definición del índice de abollado que es la aportación final del capítulo. Este índice es validado mediante un proceso novedoso de comparación por pares

de los valores del índice y la observación humana.

En el capítulo 6 presentamos nuestras ideas sobre la construcción de sistemas CBIR para imágenes hiperespectrales. En concreto, la extracción de información espectral para caracterizar las imágenes es infrecuente en la literatura y cuando se hace se aplican soluciones triviales del tipo del espectro medio de la imagen, debido a la complejidad computacional y sensibilidad de los algoritmos de clustering que podrían ser aplicados para obtener información más detallada. Nuestra proposición es relativamente eficiente desde el punto de vista computacional puesto que implica sólo un paso sobre la imagen y las operaciones que se realizan son computacionalmente ligeras. Por ello pensamos que puede ser aplicado a sistemas reales con grandes colecciones de imágenes.

Nuestra aproximación toma un punto de vista novedoso en el sentido de que en lugar de tratar de encontrar valores promedio como caracterización del contenido espectral de la imagen, lo que es habitual cuando se considera un modelo de mezcla de gaussianas, buscamos valores extremos que definen un recubrimiento convexo de los datos (todos o gran parte de ellos). Este cambio de paradigma se debe a la adopción del modelo de mezcla espectral (*spectral mixing*) que trata de modelar los píxeles a resolución subpixel, buscando la composición fraccional de los píxeles.

Desde el punto de vista computacional, el algoritmo propuesto para la extracción de los endmembers es novedoso y forma parte de lo que podríamos llamar *Lattice Computing*: una colección de métodos computacionales basados en operadores de retículos o en Teoría de Retículos (*Lattice Theory*). El algoritmo está basado en un concepto novedoso: el de independencia morfológica (*lattice independence* en las nuevas tendencias de nomenclatura) y hace una utilización original de las Memorias Autoasociativas Morfológicas.

Las bases de datos de imágenes sintéticas pueden servir para la evaluación sistemática de algoritmos de proceso y segmentación de imágenes hiperespectrales. Existen pocos conjuntos de datos extensos accesibles para dichos procesos de evaluación y ninguno tiene la complejidad del que hemos construido para la realización de los experimentos.

1.3. Objetivos de la tesis doctoral

En esta sección vamos a enumerar los objetivos que persiguieron los trabajos de esta tesis doctoral. Algunos de ellos están explícitos en la estructura

de la propia memoria, otros han sido instrumentales para la realización de los trabajos y todos ellos marcan en alguna manera un hito o contribución. Algunos son puramente formativos, no nos olvidemos que la tesis doctoral es un periodo formativo del investigador:

- Revisión del estado del arte en sistemas CBIR.
- Investigación sobre procesos de extracción de características de texturas en imágenes digitales.
- Construcción de sistemas CBIR concretos: especializándose durante el trabajo en las aplicaciones al control de calidad del papel reciclado y a las imágenes hiperespectrales.
- Aplicación de algoritmos de *Lattice Computing* a algún problema de interés práctico no trivial.
- Construcción de una colección de imágenes hiperespectrales sintéticas no trivial para validación de algoritmos.
- Experimentación con algoritmos de clasificación automática en un dominio realista de datos (e.g. imágenes de papel reciclado).
- Prueba de la metodología experimental en un dominio no trivial y realista (formación).
- Revisión de métodos de segmentación de imagen (Filtros de Gabor y Wavelets) (formación).
- Revisión de técnicas y algoritmos de *Lattice Computing*.
- Transferencia de resultados a la empresa privada.

1.4. Publicaciones realizadas

- Orlando Maldonado, David Vicente, Manuel Graña, *CBIR Indexing Hyperspectral Images*. IGARSS 2005 IEEE Press, ISBN 0-7803-9051-2.

- Orlando Maldonado, David Vicente, Manuel Graña, Alicia d'Anjou *Content based retrieval of hyperspectral images using AMM induced endmembers* in *Knowledge-Based Intelligent Information and Engineering Systems* (KES 2005), Rajiv Khosla, Robert J. Howlett, and Lakhmi C. Jain (Eds.) LNAI 3681 : 827-832 Springer Verlag ISBN 3-540-28894-5.
- Manuel Graña, Orlando Maldonado, David Vicente *Morphological independence and hyperspectral image indexing* in *Mathematical Methods in Pattern and Image Analysis*, Jaakko T. Astola, Ioan Tabus, Junior Barrera, (eds) SPIE vol. 5916 pp: 213-222 ISBN 0-8194-5921-6.
- José Orlando Maldonado, David Vicente Herrera, Manuel Graña Rómay. *Visual texture characterization of recycled paper quality* in *Innovations in Hybrid Intelligent Systems*, Advances in Soft Computing 44 pp: 288- 295 Springer Verlag ISBN 978-3-540-74971-4
- M. Graña, I. Villaverde, J. O. Maldonado, C. Hernandez *Two Lattice Computing approaches for the unsupervised segmentation of Hyperspectral Images* Neurocomputing, Accepted.

1.5. Publicaciones submitidas

- J.O. Maldonado, M. Graña. *Recycled paper visual indexing for quality control*. Expert Systems with Applications. Under review.

1.6. Estructura de la memoria de la tesis

La presente memoria se estructura en los siguientes capítulos:

- En el capítulo 2 se proporciona una revisión general de los principios básicos de los sistemas de recuperación de imágenes basados en contenidos (CBIR). Tras una presentación de los aspectos generales de estos sistemas y algo de historia, se presentan los principales sistemas encontrados en la literatura. Presentamos la arquitectura típica de estos sistemas y desarrollamos en detalle cada uno de los elementos: las técnicas de extracción de características, las distintas medidas de similitud más frecuentes en la literatura, los sistemas de organización de las

búsquedas desarrollados a partir de los árboles de búsqueda, extendiéndolos a datos multivariante. Los procesos de retroalimentación de la búsqueda han cobrado protagonismo para solventar el problema de la brecha semántica y son presentados también. Por último, dedicamos un tiempo a discutir los procesos y variables de observación empleados en la validación de estos sistemas.

- En el capítulo 3 se introducen los Bancos de Filtros de Gabor (BFG), presentando su aplicación para la caracterización de la textura. La transformada de Gabor es la primera proposición de una transformada con localización espacial y frecuencial que permite diseñar de forma sistemática y elegante sistemas de filtros que explotan exhaustivamente la información en el espacio transformado de Fourier. En el apéndice damos algunos ejemplos didácticos de análisis de señales mediante filtros de Gabor. Los BFG han sido ampliamente utilizados para la caracterización de la textura presente en las imágenes.
- En el capítulo 4 presentamos la Transformada Discreta de Wavelets (TDW). Dicha transformada ha sido introducida como un medio de realizar análisis multiresolución de las imágenes, que trata de efectuar la detección y análisis de los objetos presentes en la imagen a distintas escalas. Presentamos el algoritmo de descomposición piramidal típico y la transformada en dos dimensiones que se utiliza en las imágenes. Una de las aplicaciones más extendidas de esta transformada es el análisis de texturas y es de interés especial para nuestra aplicación sobre las imágenes de papel reciclado.
- En el capítulo 5 presentamos la aplicación de técnicas de indexación de imágenes para la caracterización de la calidad del papel reciclado en términos de la aparición visual de un efecto que hemos denominado “abollado” a falta de una caracterización apropiada en los estándares actuales de calidad del papel. En este capítulo realizamos en primer lugar unas consideraciones metodológicas y describimos el etiquetado manual realizado sobre las imágenes proporcionadas por una empresa papelería que originalmente propuso trabajar en este problema. Definimos las características de textura que vamos a utilizar, basadas en Bancos de Filtros de Gabor y Transformada Discreta en Wavelets. El objetivo final es el establecimiento de un índice dado por un valor escalar que crezca monotonamente con el nivel de abollado percibido en

la imagen, para que pueda ser usado como una medida objetiva de la calidad del papel. Para establecer la calidad de las características generadas realizamos un experimento de clasificación supervisada probando una batería de sistemas de construcción de clasificadores. Finalmente, proponemos un índice de abollado con resultados de clasificación comparables a la observación humana y con las propiedades deseadas.

- En el capítulo 6 presentamos nuestras ideas para la construcción de sistemas CBIR para colecciones de imágenes hiperespectrales. Presentamos las imágenes hiperespectrales brevemente. La caracterización que buscamos se basa en la descomposición espectral de los píxeles. Para esta operación son críticos los sistemas de inducción de los endmembers en los que se basa dicha descomposición espectral. Presentamos nuestro algoritmo basado en la propiedad de independencia morfológica detectada mediante Redes Morfológicas. Para ello hacemos una breve revisión de sus fundamentos. Presentamos la distancia entre imágenes calculada entre los conjuntos de endmembers que caracterizan a las imágenes hiperespectrales en nuestra proposición. Para validar nuestra proposición realizamos experimentos de recuperación sobre bases de datos de imágenes sintéticas.
- En el capítulo 7 presentamos líneas de trabajo futuro y nuestras conclusiones sobre algunos aspectos de los temas tratados.
- El capítulo 8 contiene dos apéndices que presentan de forma didáctica el efecto de las transformadas de Gabor y de Wavelets.

Recuperación de imágenes basada en contenidos

2.1. Introducción

Desde que el hombre llegó a idear las representaciones gráficas, estas son una rica fuente de expresión y comunicación, sin embargo, nunca como hoy las imágenes han cobrado el protagonismo que ostentan en tantas áreas del quehacer humano como las artes, los medios de comunicación, la medicina, y la ciencia en general. El auge de los medios electrónicos y la informática han permitido el aumento en la producción y colección de imágenes digitales de todo tipo. Se puede apreciar en la vida diaria que el uso doméstico de las cámaras digitales y la fácil adquisición de medios de soporte y almacenamiento masivo de información permiten generar y mantener colecciones de información multimedial personal de gran tamaño.

En otras áreas más especializadas como la medicina, las imágenes son una herramienta diagnóstica cada vez más frecuente en muy diversas modalidades: resonancia magnética nuclear, PET, ultrasonidos, rayos X, etc.

En otras áreas de las ciencias de la vida, el uso de imágenes es cada vez más frecuente para tareas como la monitorización de especies animales y vegetales, incluyendo imágenes de reconocimiento remoto para monitorización de bosques, etc. Los microscopios electrónicos permiten capturar y contemplar imágenes que muestran las características y comportamientos presentes en el mundo molecular y celular. En otros campos de la ciencia, por ejemplo en Astronomía, se están generando constantemente nuevas imágenes procedentes de telescopios de todo tipo, desde los grandes telescopios en órbita, hasta los observatorios más locales.

Las técnicas de teledetección, como las basadas en satélites proporcionan imágenes que son una rica fuente de información con aplicaciones en la agricultura, las ciencias forestales, la geología o la seguridad entre otras.

La comunicación a través de Internet ha hecho posible que las imágenes generadas por muy diversas comunidades estén disponibles convirtiéndose en una especie de gigantesco repositorio de información. En la actualidad sitios web dedicados a compartir información multimedia están concentrando cantidades ingentes de esta información.

Son necesarios instrumentos de gestión y búsqueda en estas colecciones de imágenes. Una de las aproximaciones más elementales es proponer sistemas para el manejo de colecciones de imágenes desarrollados a partir de los paradigmas convencionales orientados a documentos de texto. Las imágenes son etiquetadas o documentadas mediante el uso de palabras clave, que describen el contenido de la imagen. La recuperación de imágenes en el sistema se realiza por medio de consultas textuales. El problema fundamental de esta aproximación, es que las consultas textuales requieren la documentación o etiquetado previo de cada una de las imágenes que conforman el repositorio, lo que conlleva dos problemas básicos [132]:

- Es un trabajo tedioso y costoso que implica gran cantidad de tiempo en el proceso de documentación o etiquetado de las imágenes
- El etiquetado es siempre un proceso subjetivo que depende de la opinión de la persona que lo hace. Por tanto, es fácil ver que, debido a la riqueza en información visual contenida en las imágenes, éstas pueden tener múltiples interpretaciones y no es fácil poner cada detalle en forma de texto.

Para evitar estos problemas, Google ofrece el servicio de recuperación de imágenes en web, basado en la información textual de la página en la cual está embebida la imagen. Obviamente la información contenida en la página web puede dar lugar a muchas ambigüedades debido a que la comunidad que crea y mantiene esta información es heterogénea y carece de control.

La alternativa a la búsqueda basada en anotaciones textuales, es realizar de forma automática la indexación de las imágenes mediante descriptores de su contenido calculados a partir de la propia imagen. Estos cálculos estarán basados en algoritmos de proceso de imagen digital y de visión por computador. Esta es la razón por la cual dichos sistemas se llaman sistemas de recuperación basados en contenidos (CBIR, por sus siglas en inglés).

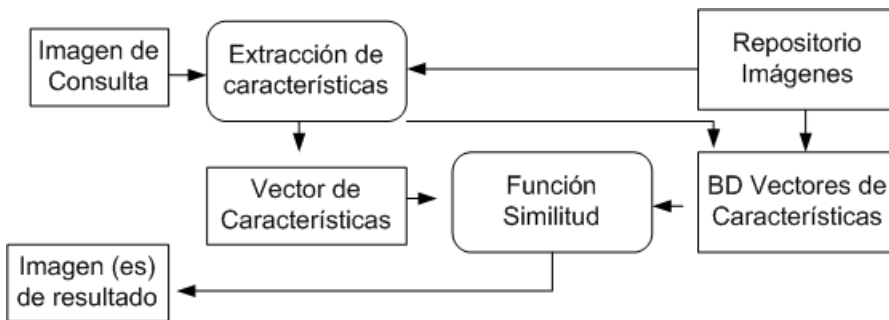


Figura 2.1: Esquema de una consulta mediante ejemplo, en un sistema CBIR.

Los sistemas más populares realizan las *consulta-mediante-ejemplo* (query-by-sample), en las que se presenta al sistema una imagen que contenga las características visuales que deseamos buscar. Sobre ésta, el sistema realiza el procesamiento necesario para extraer los descriptores que forman el índice de la imagen que llamaremos vector de características. Este vector es comparado con los vectores de características de cada una de las imágenes que conforman la base de datos. La comparación se realiza mediante una métrica o función de similitud que permite recuperar una lista con el(los) elemento(s) que se aproximen mejor a la consulta realizada. La figura 2.1 ilustra el diagrama de flujo del proceso de consulta, que es la estructura básica de un sistema de recuperación basado en contenidos, de cuyos componentes hablaremos con mayor detalle en la sección 2.3.

En este capítulo haremos una revisión sobre los sistemas de recuperación de imágenes basados en contenidos, abordando cada uno de los aspectos de su implementación, los cuales constituyen por si mismos amplias líneas de investigación. El capítulo está organizado de la siguiente manera. En la sección 2.2, comentaremos sobre algunos aspectos generales relacionados con los sistemas CBIR, como el dominio de conocimiento y la semántica. En la sección 2.3 enumeramos algunos de los sistemas más conocidos y se hace una descripción de la arquitectura típica de un sistema CBIR. La sección 2.5 está dedicada al proceso de extracción de características de la imagen, donde se hace referencia a los tipos habituales de características así como a las técnicas desarrolladas para su extracción. En la sección 2.6 hacemos una revisión de las diferentes métricas definidas sobre los espacios de características y su implementación dentro de los sistemas CBIR. La sección 2.7 está dedicada a revisar las técnicas más habituales para acceso rápido y eficiente a grandes

repositorios de datos. En la sección 2.8 se analizan las herramientas que pretenden la realimentación inteligente del proceso de consulta, las cuales tienen cada vez más importancia en los sistemas CBIR. Finalmente en la sección 2.10 se ofrecen algunas conclusiones alcanzadas tras el proceso de revisión bibliográfica que fue necesario para el desarrollo de este capítulo.

2.2. Aspectos generales

Para el diseño y desarrollo de sistemas de recuperación de imágenes basados en contenidos, es importante tener en cuenta los requerimientos de cada grupo de usuarios. Cabe preguntarse, qué buscan los usuarios, cómo lo buscan y cómo juzgan lo que encuentran. Las respuestas a dichos interrogantes demandan un conocimiento de las necesidades del usuario, que puede ser tan amplio como las áreas en que éstos desempeñan sus actividades.

Smeulders [142] habla de la importancia de tener en cuenta el contexto semántico de la imagen, a lo que llamaremos dominio de la imagen (*image domain*), para conseguir sistemas que lleguen a satisfacer las necesidades del usuario. En un dominio reducido, las imágenes presentan una variabilidad limitada y son predecibles en sus aspectos más relevantes, por lo cual es más fácil relacionar la interpretación semántica de la imagen con sus características visuales primitivas. Ejemplos de colecciones de imágenes que definen un dominio reducido son los catálogos litográficos y las colecciones de imágenes de rostros, con posición, iluminación y puntos de vista controlados. Por contra, en un dominio extenso, las imágenes se caracterizan por ser polisémicas y su semántica puede ser descrita solo parcialmente. Ejemplos de imágenes extraídas de un dominio extenso, son las colecciones fotográficas, o el conjunto de las imágenes disponibles en Internet. En este tipo de dominios, aspectos como la iluminación, la oclusión y recorte de objetos, y el registro desde diferentes puntos de vista, son dificultades importantes, que deben tenerse en cuenta en el momento de diseñar sistemas de recuperación. Estos aspectos tienen que ver con la *brecha sensorial*, “que se refiere a la diferencia existente entre el objeto del mundo real y la información digital (computacional) capturada o registrada de la escena”[142].

Para precisar los conceptos definimos *semántica* como la categorización de los objetos en función de algún criterio de similitud. De esta forma la semántica de un usuario cuando realiza una búsqueda es la categoría de objetos que tiene en mente, la búsqueda que realiza debería estar guiada por

la similitud que él tiene en mente. Por otro lado los algoritmos de extracción de características y la métrica definida sobre este espacio de características inducen una cierta categorización que puede venir dada por el *clustering* de las imágenes. La falta de coincidencia entre estas dos categorizaciones es conocida como la *brecha semántica*. Usualmente el usuario no especifica de ninguna manera su semántica por lo que la brecha semántica no es cuantificable ni formalizable. En dominios especializados (reducidos), la brecha semántica es usualmente pequeña contrario a los dominios extensos, donde la brecha semántica es considerablemente mayor.

En [142] se identifican además, tres tipos de búsqueda relacionados con los intereses o propósitos del usuario al acceder a un sistema:

- Las búsquedas por asociación, que permiten a los usuarios realizar una exploración sobre la colección de imágenes, refinando de manera iterativa la búsqueda. Estas son propias de colecciones generales de imágenes del mundo real.
- Las búsquedas específicas, en las cuales los usuarios buscan un elemento en particular, o una imagen que contenga un objeto como el que se ha suministrado en la imagen de ejemplo. Si el usuario tiene un objetivo preciso en mente, puede dar algunas de sus características e ir refinando la búsqueda hasta encontrar el objeto preciso. Estos sistemas pueden ser adecuados para búsquedas en catálogos de arte, de componentes industriales, etc.
- Las búsquedas por categoría, que permiten recuperar una imagen representativa de una clase o categoría específica. Pueden resultar útiles si se quiere introducir una imagen nueva al sistema, y es necesario establecer a qué clase pertenece. Este tipo de sistemas puede encontrarse en entornos especializados, como catálogos de especies biológicas.

Smeulders además discute la necesidad de conocimiento a priori sobre el dominio de las imágenes para salvar las diferencias semánticas y sensoriales. Este conocimiento se puede especificar mediante reglas de similitud sintáctica, reglas de similitud perceptual, condiciones físicas y reglas topológicas y geométricas.

Eakins [43] da ejemplos de posibles atributos que los usuarios pueden utilizar para recuperar imágenes, tales como:

- la presencia de color, textura o forma particular, por ejemplo, un cuadrilátero rojo;
- la presencia de un arreglo o un tipo específico de objetos, por ejemplo, una bandada de pájaros; la descripción de un evento particular, por ejemplo, la entrega de premios a un deportista;
- la presencia de individuos, lugares o eventos conocidos, como la torre Eifel; emociones asociadas a una imagen, por ejemplo alegría;
- metadatos, tal como la fecha de creación de un fichero.

Exceptuando el primer tipo, cada posible consulta representa una abstracción de mayor nivel a la anterior, que requiere de alguna entidad de conocimiento externo para su validación. Así, Eakins hace una clasificación de las consultas en tres niveles:

- Nivel 1: Comprende recuperación por características primitivas, como color, textura o forma.
- Nivel 2: Comprende recuperación por características derivadas (o lógicas), que implican algún tipo de inferencia lógica sobre la entidad, a su vez las divide en dos: recuperación por objetos de un tipo dado (“recuperar imágenes con un coche”), o de un objeto o persona en particular (“recuperar imágenes de la torre Eifel”).
- Nivel 3: Comprende la recuperación por atributos abstractos, que requiere razonamiento de alto nivel sobre el significado y propósito de las escenas descritas. Las divide en dos a su vez: recuperación por eventos o actividades conocidas (“Encontrar imágenes con danzas folclóricas árabes”), y recuperación por imágenes con contenido emocional o religioso (“recuperar imágenes que describan sufrimiento”).

Podemos resaltar que la diferencia entre los niveles 1 y 2 está directamente relacionada con la brecha semántica, con lo cual, los resultados en las consultas de los niveles 2 y 3 pueden satisfacer a los usuarios en sistemas controlados, con un estrecho dominio de la imagen.

2.3. Ejemplos de Sistemas CBIR

La recuperación de imágenes basada en contenidos o CBIR (por sus siglas en inglés *Content based image retrieval*) es una de las áreas de investigación más prolíficas en los últimos años, muestra de ello es la gran cantidad de publicaciones que han surgido desde la década de 1980. Sin embargo, los avances más significativos se registran a partir de la década siguiente. Dentro de los sistemas CBIR más populares podemos citar el QBIC¹ de IBM (*Query By Image Content*) [48] y el *Virage*² [64] que han evolucionado de tal manera que han permitido su aplicación en la gestión de video e información multimedia. La mayoría de los sistemas desarrollados provienen de la academia y no han sido explotados comercialmente. Podemos mencionar algunos de los nombres más conocidos de sistemas o prototipos desarrollados en este ámbito tal como *Chabot* [115], *Photobook*³[119] o *Netra* [99], que utilizan características de color y textura para describir el contenido de la imagen. Desde el inicio de la presente década, la producción científica relacionada con los sistemas CBIR se ha llegado a incrementar de manera casi exponencial, como han comprobado Datta et al., en [36]. Entre los sistemas desarrollados en esta última década destaca el *Blobworld* [20], que no solo se fija en la extracción de características por cada pixel, sino que efectúa una segmentación en regiones teniendo en cuenta su ubicación espacial y su tamaño. *QuickLook* [27] es otro sistema que permite la recuperación de información visual en bases de datos extensas, teniendo en cuenta características del color y su distribución espacial en la imagen, así como la forma por detección de bordes. En [89] se presenta un sistema CBIR, que emplea información de la forma de los objetos en la imagen para recuperarla, mediante la extracción de bordes. Un sistema disponible y de libre distribución con licencia GNU es el GIFT⁴[143] (GNU image Finding Tool). En [85], [111], [44], [23], es posible encontrar ejemplos de otras propuestas de sistemas que implementan CBIR.

¹<http://www.qbic.almaden.ibm.com/>

²<http://www.virage.com>

³<http://vismod.media.mit.edu/vismod/demos/photobook/>

⁴<http://www.gnu.org/software/gift/>

2.4. Arquitectura de los sistemas CBIR

El abundante material escrito entorno a los sistemas CBIR en los últimos años ha sido recogido en varias revisiones exhaustivas. En [142], por ejemplo, se hace una revisión de las publicaciones realizadas hasta el año 2000, y se repasan las arquitecturas propuestas. El análisis revela que todos los modelos se ajustan a un marco para la implementación que tiene los siguientes componentes específicos:

- cálculo de características y análisis sensorial,
- un módulo de interpretación y dominio del conocimiento,
- un módulo de interacción e interfaz de usuario,
- y un módulo de indexación y almacenamiento.

Smeulders y sus colegas han concluido que en la mayoría de las aproximaciones encontradas en la literatura se han limitado a proponer innovaciones de uno o dos de esos componentes. Sugieren, además, que es necesario un marco (*framework*) para sistemas CBIR que “suministre una visión más balanceada de los cuatro componentes constituyentes”. El marco podría estar basado en “protocolos explícitos de comunicación”, que faciliten el diálogo entre cada uno de los módulos.

Dentro de los sistemas propuestos en la presente década, en [91] encontramos un modelo de arquitectura típico en el que existe un módulo de consulta que realiza la extracción de características (basadas en textura y color), un módulo de búsqueda en la base de datos, y un módulo que realiza un proceso de realimentación que afina la función de similitud con ayuda de la interacción del usuario, al igual que en [44], donde centran su atención en un módulo de entrenamiento de la función de similitud, para que identifique las imágenes relevantes en el dominio de imágenes médicas. En [45] se adopta un enfoque similar, aunque la caracterización de las imágenes se basa en la descomposición mediante quadrees, y no tiene en cuenta ningún proceso de realimentación. En [26] se desarrolla un modelo que incluye un motor de razonamiento lógico, que mediante cierta descomposición jerárquica de la imagen permite la comparación mediante analogías. En [72] se propone una “arquitectura de dos estados”, refiriéndose con ello en una técnica de

búsqueda basada en dos vectores de características por imagen, que finalmente podríamos encasillar en el marco clásico. Otras propuestas recientes pueden ser encontradas en [77], [144], [112], todas ellas son de algún modo un caso particular del modelo general propuesto por [142].

Otra revisión de la bibliografía ha sido detallada en [107], la cual centra su interés en las imágenes médicas, pero también cita los sistemas clásicos. En ella se hace un análisis de las necesidades y campos de aplicación presentes y futuros, llegando a proponer un sistema de recuperación modular en el que se puedan cambiar fácilmente los vectores de características y se puedan integrar con facilidad módulos que implementen nuevas técnicas de recuperación, o métodos eficientes de almacenamiento. Para la implementación “sería necesaria la definición de mecanismos sencillos de plug-in para los diferentes componentes”. Una revisión del estado del arte reciente y más completa puede ser consultada en [36].

Podemos recoger las ideas expuestas hasta ahora aportando una estructura general para un sistema CBIR. Si partimos de la figura 2.1 en la cual se aprecia el diagrama de flujo seguido en un proceso típico de consulta mediante ejemplos, podemos agrupar el sistema en módulos, como se muestra en la figura 2.2, donde hemos añadido un módulo de realimentación. La gran mayoría de los sistemas encontrados tienen una arquitectura similar a ésta, cuyos componentes pueden ser descritos por su funcionalidad de la siguiente manera:

- La interfaz de usuario: debe permitir realizar las consultas y visualizar los resultados de las mismas. En algunos casos incluso alimentar el sistema con nuevas imágenes o datos. Las consultas pueden ser realizadas mediante imágenes de ejemplo dadas al sistema, el diseño de gráficos, esquemas o dibujos sobre un editor de imágenes que ejemplifiquen las características gráficas de la(s) imagen que se está(n) buscando, los datos numéricos que conforman el vector de características de la consulta o usando como imagen de ejemplo una extraída de la propia base de datos del sistema CBIR.
- El análisis de imágenes: se encarga de realizar el procesamiento de las imágenes, su análisis y extracción de las características mediante técnicas de visión. Aquí se calcula y cuantifica la información relativa al color, a las texturas, a los objetos o formas presentes, y a puntos de interés en la imagen a procesar. Estas operaciones se han de realizar

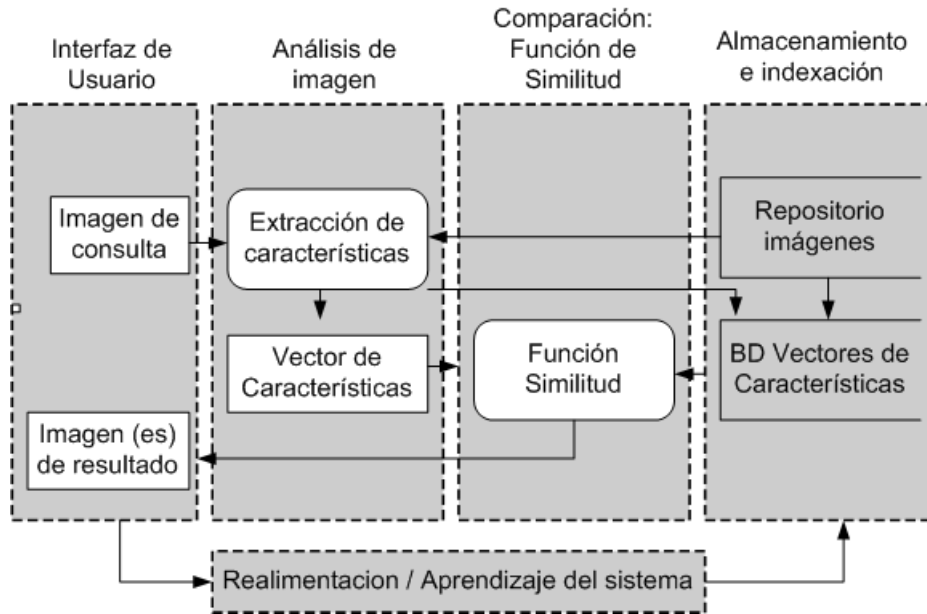


Figura 2.2: Agrupación de funcionalidades en módulos de un sistema CBIR

tanto en las imágenes que conforman la la base de datos del sistema como sobre las imágenes de consulta.

- La indexación y almacenamiento: se encarga de generar, mantener y acceder a la estructura de almacenamiento de la información relativa a las imágenes contenidas en el sistema CBIR, esencialmente los vectores de características. Un sistema CBIR contendrá una cantidad enorme de datos multidimensionales, por lo cual el tiempo de acceso requerido resulta crítico. Para afrontar de manera eficiente y robusta este problema, se utilizan técnicas de gestión de datos multidimensionales, sobre todo las basadas en estructuras de tipo árbol como los R-tree.
- Comparación y funciones de similitud: Este es un aspecto fundamental en los sistemas de recuperación y está estrechamente relacionado con la definición de los vectores de características. La métricas o funciones de similitud permiten comparar el vector de características de la consulta con los vectores de características de las imágenes almacenados en la base de datos del sistema, y establecer cuál(es) es(son) más cercana(s) o similar(es) a la consulta dada. Esta función puede ser la distancia

Euclídea, o una función más sofisticada según el tipo de imagen sobre el que se esté trabajando y la definición de su vector de características.

- La realimentación o aprendizaje del sistema: Este es un aspecto que ha venido tomando mayor relevancia a partir del artículo de revisión de Smeulders et al.[142]. Los expertos han observado que el sistema puede ser más eficiente si aprende de la interacción con los usuarios. Con cada consulta realizada, el sistema establece las preferencias en los resultados de las consultas así como la satisfacción ante esos mismos resultados. Con esta información el sistema puede modificar la función de similitud y/o la estructura de indexación de los datos, para mejorar los resultados en consultas posteriores. Con ésto se pretende aproximar la semántica existente en la mente del usuario.

2.5. Extracción de características

En este apartado describiremos los métodos utilizados en el proceso de extracción de características primitivas de la imagen, que es un aspecto clave en los sistemas CBIR y se apoya en las técnicas de visión por ordenador. Los autores suelen distinguir entre dos tipos de características visuales: las globales, que pueden referirse al color o texturas presentes en toda la imagen y las locales, que están definidas en áreas específicas, y requieren una segmentación en regiones [43],[98].

2.5.1. Características de Textura

No hay un consenso general para una definición formal o adecuada de textura, aunque podemos encontrar algunas definiciones como las siguientes:

- En [117], se describe la textura como la repetición de un patrón espacial básico, cuya estructura puede ser determinista o estocástica;
- Russ [133] la define como un descriptor de la variación en la iluminación local entre píxeles de una pequeña vecindad.

El análisis de texturas ha tomado un papel importante en áreas como el análisis de imágenes médicas, el reconocimiento remoto o la inspección industrial. También ha sido tenido en cuenta a la hora de caracterizar las imágenes en

sistemas CBIR [142], [107], [98]. Los métodos para análisis de textura pueden ser clasificados en cuatro categorías básicas: métodos estadísticos, métodos estructurales, métodos basados en modelos y métodos basados en transformadas.

- Los métodos estadísticos. Una de las más sencillas formas de comparar texturas es comparando sus estadísticos de primer orden, entendiéndose estos como los que involucran píxeles simples. Se puede utilizar en este caso el histograma normalizado de los niveles de gris de la imagen, que proporciona una estimación de la función de densidad de los niveles de gris, y comparar sus estadísticos como la media, mediana o varianza. Un método de más alto orden es la matriz de coocurrencias [65]: dada una imagen $f(x)$, su matriz de coocurrencias M_d para un vector de desplazamiento d se define como:

$$M_d(i, j) = \text{Card} \{ (s, r) : r - s = d, f(s) = i, f(r) = j \},$$

donde r, s son posiciones de píxeles en la imagen. Es decir, cada posición $M_d(i, j)$ de la matriz es el número de ocurrencias del par de niveles de gris i y j que están separados por el vector d . Sobre la matriz de coocurrencia se han definido un conjunto de descriptores como la energía, probabilidad máxima, entropía, correlación, etc, utilizados para procesos de comparación y clasificación.

- Métodos estructurales. Los métodos estructurales definen las texturas como una composición de elementos primitivos bien definidos, por ejemplo, líneas paralelas regularmente espaciadas [12], concibiendo las texturas reales o naturales, como una distorsión de estas texturas ideales. Otros enfoques pueden considerar a las texturas como una colección de objetos primitivos similares (pero no idénticos) distribuidos con algún patrón de repetición. La caracterización de las texturas bajo este enfoque puede realizarse calculando propiedades estadísticas sobre esta distribución de elementos, o las reglas de distorsión y/o ubicación .
- Métodos basados en modelos. Estas técnicas estiman los parámetros de un modelo sobre los píxeles de la imagen. Los parámetros del modelo describen las cualidades de las texturas. Ejemplos de tales técnicas son los modelos autoregresivos [137], los campos aleatorios de Markov [22] y los fractales [120].

- Métodos basados en transformadas y procesado de señal. Se han aplicado técnicas habituales en el procesamiento de señales, que permiten analizar la imagen, aplicando filtros para obtener características relacionadas con la orientación o la magnitud de los componentes frecuenciales presentes en las texturas. Dentro de estos métodos podemos encontrar los que se aplican en el dominio espacial, mediante aplicación de operadores de bordes como las máscaras Laplacianas o de Roberts [88], [116], y las basadas en momentos invariantes [70], [93]. También se han utilizado técnicas aplicadas en el dominio de Fourier, que proporcionan información sobre la potencia del espectro [106], o realizan una segmentación del plano transformado, discriminando los diferentes componentes frecuenciales [162]. Debido a que la transformada de Fourier proporciona información sólo en el dominio de la frecuencia, la transformada enventanada de Fourier, también llamada transformada de Gabor, ha sido una alternativa que permite operar tanto en el dominio espacial como en el de Fourier [37], [156], [153]. El diseño de Bancos de Filtros de Gabor ha sido ampliamente aplicado a la clasificación, segmentación y recuperación de imágenes texturadas [103], [166], [62]. En el mismo sentido, la Transformada Discreta Wavelet [101] también ha sido aplicada al análisis de texturas, [157], [71], y especialmente implementada en prototipos de sistemas CBIR [91], [89], [138].

2.5.2. Características de Color

El color ha sido una de las características más utilizadas en la recuperación de imágenes. Un espacio de color es una representación numérica mediante la que se puede especificar cualquier color. Por ejemplo, el espacio RGB, se basa en la representación de un color como la suma de tres señales en las bandas cromáticas básicas: El rojo, el verde y el azul (Red, Green, Blue). Por tanto, en una imagen con representación del color en el espacio RGB, cada píxel corresponde a un punto en un espacio tridimensional. Cada banda o componente, toma valores dentro de un rango determinado por el máximo valor en cada banda cromática. Así pues, es evidente que el poder de discriminación en un espacio de color, es superior al que tenemos en la escala de grises.

La extracción del histograma del color propuesta por Swain y Ballard [148], es una de las técnicas más utilizadas en los sistemas CBIR, en ella se determina la proporción de píxeles de cada color en la imagen. El histograma puede ser almacenado en la base de datos y en tiempo de búsqueda, el

usuario determina la proporción de color deseado en las imágenes recuperadas o envía una imagen de ejemplo para que su histograma sea extraído y comparado con los almacenados en la base de datos. Mejoras en la técnica del histograma de color se han introducido en [146] que incluyen histogramas de color acumulativos. Otras aproximaciones buscan reducir la cantidad de espacio de almacenamiento requerida por los histogramas, como la propuesta por [118], que utiliza características que denomina momentos de cromaticidad (*chromaticity moments*) que permiten capturar el contenido espectral de la imagen en una representación compacta. La indexación mediante Hashing de características de color (*Color-card*) invariantes a iluminación y puntos de observación son propuestas en [54]. La agrupación por regiones de color es sugerida en [55], como alternativa al uso de histogramas.

Hay dos asuntos críticos que son abordados con frecuencia en la bibliografía. Por un lado, la variabilidad que se puede presentar durante el registro de una imagen, en términos del punto de vista de la cámara, la iluminación o la reflectancia, para lo cual Gevers y Smeulder [54], por ejemplo, proponen un conjunto de características invariantes a tales aspectos. Por otro lado se ha propuesto el uso de otros espacios de color, los cuales parecen corresponder mejor a la percepción humana de la similaridad entre colores. Así, los espacios de color HSI y HSV (Matiz, saturación, intensidad), (*hue, saturation, intensity/value*) son habitualmente utilizados [53], [170], [168], [35], [24], debido primero a que la intensidad puede ser separada de la información del color en la imagen, segundo a que las componentes intensidad y saturación están muy relacionadas con la percepción humana del color, y tercero, por sus propiedades de invarianza frente a iluminación y orientación de la cámara, lo que lo hace adecuado para CBIR [142]. Otros espacios de color utilizados en recuperación de imágenes son el espacio *CIE XYZ* [118], *CIELUV* [55], el *CIELAB* [96], entre otros.

2.5.3. Características de formas

Aunque hay evidencia psicológica de que los objetos son reconocidos primeramente por su forma [13], la segmentación automática de los objetos en las imágenes es un problema no solucionado. Incluso en dominios muy específicos la segmentación totalmente automatizada causa muchos problemas y no es fácil de realizar. El problema es considerablemente complejo cuando hablamos de imágenes del mundo real que pueden no tener un fondo homogéneo, o en las cuales existen varios objetos que se solapan. La mayoría

de los sistemas CBIR que incluyen recuperación mediante la caracterización de formas, muestran un buen rendimiento con imágenes que presentan objetos fácilmente identificables, que pueden ser aislados del resto de la imagen, condición que no es habitual en la mayoría de las imágenes reales.

Para caracterizar las formas presentes en una imagen, podemos dividir las técnicas en tres categorías:

- Las basadas en el contorno del objeto. En este caso es habitual calcular una firma (una función 1D) del contorno de la imagen, que puede ser usada directamente como característica o sobre ella se pueden calcular otras características de dimensión inferior. [165], [167], [171] [154], [114], [4], [7].
- Las basadas en mapas de bordes, que se aplican a imágenes en las que es muy difícil realizar una identificación precisa de los objetos. Estos mapas de bordes no permiten obtener un contorno cerrado que defina el objeto, pero son utilizados definiendo distancias apropiadas. [79], [169], [54], [112], [89].
- Las basadas en regiones, donde las características se calculan como función de la región en la imagen ocupada por el objeto y no directamente sobre la curva 2D que define el contorno. [20], [5], [173].

En el caso de procesamiento de imágenes basado en el contorno del objeto, en [165] y [167] se propone calcular los descriptores de Fourier sobre la firma del contorno para caracterizar la imagen. Como firma se utiliza la distancia al centroide, que viene dada por la expresión:

$$f(t) = \sqrt{(x(t) - x_c)^2 + (y(t) - y_c)^2}, \quad (2.1)$$

donde (x_c, y_c) es el centroide de la imagen, $t = 1, \dots, N$, es el ángulo del radio sobre el que estamos midiendo la distancia del contorno al centroide y viene dado por un muestreo uniforme entre 0 y 360 grados; así, el contorno del objeto es digitalizado en N puntos. La distancia al centroide no está bien definida para imágenes con objetos que tengan contornos no convexos. La transformada discreta de Fourier de $f(t)$, viene dada por la expresión:

$$\begin{aligned}
F(u) &= \left(\frac{1}{N}\right) \sum_{t=0}^{N-1} f(t) e^{-j2\pi ut/N}, \\
&= \left(\frac{1}{N}\right) \sum_{t=0}^{N-1} f(t) (\cos(2\pi ut/N) - j \operatorname{sen}(2\pi ut/N)). \quad (2.2)
\end{aligned}$$

Dónde $u = 0, \dots, \frac{N}{2}$, y cada coeficiente $F(u)$ es un descriptor de Fourier. Para indexar la forma, Wong y sus colegas [165] toman la magnitud de la transformada, y la normalizan por el valor $F(0)$:

$$\begin{aligned}
|F(u)| &= \left(\frac{1}{N}\right) \sqrt{\left(\sum_{t=0}^{N-1} f(t) \cos(2\pi ut/N)\right)^2 + \left(\sum_{t=0}^{N-1} f(t) \operatorname{sen}(2\pi ut/N)\right)^2}, \\
DF &= \frac{|F(u)|}{|F(0)|}, u = 1, 2, \dots, \frac{N}{2}. \quad (2.3)
\end{aligned}$$

Así obtienen un conjunto de descriptores invariantes a traslaciones, escalado y rotación. Un interesante estudio sobre la aplicación de descriptores de Fourier a diferentes firmas del contorno es realizado en [171].

Como ejemplos de otras aproximaciones podemos ver la propuesta de Trazegnies y sus colegas [154], quienes utilizan los modelos ocultos de Markov para comparar las secuencias de esquinas obtenidas a partir del contorno. Los autores afirman que el método es resistente al desplazamiento o pérdida de las esquinas. En [114], se propone un método para recuperar imágenes en consultas donde el contorno es sólo parcialmente visible. En [4], se propone la función de diferencia de giro; el contorno es submuestreado en N puntos que determinan una resolución, y la información del ángulo entre cada par de segmentos que une dichos puntos es utilizada para caracterizar la imagen; para la comparación se tiene en cuenta la correspondencia entre regiones del polígono que son consideradas similares y el número de vértices contenidos en dichas regiones. La función de giro y los descriptores de Fourier sobre el contorno son utilizados por Antani y sus colegas en [7], para caracterizar el contorno sobre imágenes de vértebras tomadas mediante rayos X.

Respecto a las aproximaciones basadas en mapas de bordes, En [79] y [169] se propone el uso del histograma de dirección de bordes para representar información general de la forma en la imagen. En ambas propuestas,

los bordes son extraídos previamente con un operador de Canny; ya que esta aproximación puede ser invariante a escalado y traslación, pero no a la rotación, en [79] los autores proponen un suavizado del histograma para disminuir el efecto de la rotación. En [54], se proponen histogramas de bordes invariantes al color, con el fin de detectar objetos similares con independencia del punto de vista registrado. En [112] se utiliza un histograma de cambios de dirección del gradiente para representar la información global de la forma contenida en la imagen. En [89], se propone aplicar la Transformada Discreta Wavelet al mapa de bordes de la imagen, y utilizar los coeficientes normalizados como representación de la forma. En [38], se proponen técnicas basadas en contornos activos o plantillas deformables, para calcular la similitud de la silueta proporcionada como consulta y el objeto presente en la imagen, la cual ha sido preprocesada para extraer sus bordes. El grado de concordancia entre la plantilla deformada y el objeto, así como la energía requerida para deformarla, son utilizadas para derivar la función de similitud. En [6] se propone una variación de la transformada generalizada de Hough para comparar la silueta dada como consulta con el mapa de bordes de la imagen, aprovechando la robustez de la transformada para la identificación de objetos en imágenes no segmentadas.

En caso de las consultas basadas en regiones, en [20] se propone un método que utiliza información del color y la textura para agrupar píxeles en regiones similares y detectar objetos aislados, los cuales son indexados por cada región. En [122] se utiliza una representación de la forma basada en regiones, en la que tras una segmentación previa, y la ubicación del objeto de interés, se utiliza una rejilla de celdas cuadradas de tamaño fijo que es ubicada sobre el objeto para cubrirlo en su totalidad, se asigna uno a cada celda con al menos 25% de píxeles pertenecientes al objeto de interés, y cero a las demás; sobre esta rejilla se calculan algunas propiedades geométricas (mayor y menor eje, excentricidad, centro de gravedad, etc) que luego son almacenadas para su recuperación; los autores proponen algunas transformaciones para invarianza a escalado, traslación y rotación. En [173], se propone un sistema en el que se realiza una consulta introduciendo el dibujo a mano alzada de una silueta, y seleccionando el color y la textura del objeto deseado. El sistema busca regiones candidatas mediante la información de color y textura previas, y luego trata de ajustar la silueta de entrada, mediante técnicas de comparación de plantillas deformables. Esta aproximación, al igual que en [38], requieren de un alto coste de procesamiento.

2.5.4. Relaciones espaciales de regiones y puntos de interés

Cuando se tienen características calculadas sobre diferentes entidades en una imagen, las relaciones entre ellas también pueden ser usadas para procesos de recuperación. La información espacial es un aspecto utilizado, por ejemplo, en sistemas de información geográfica. Si se tiene en cuenta la estructura de la imagen, junto a las características visuales de las partes u objetos dentro de ésta, se pueden representar las relaciones espaciales, como por ejemplo, algún tipo de orden jerárquico u otra relación entre los objetos.

Entre las técnicas utilizadas para realizar las búsquedas en colecciones de imágenes usando información sobre las relaciones espaciales entre objetos, está la indexación icónica formulada por Chang en [21], quien propuso una estructura para datos pictóricos llamada 2D-String. Según esta propuesta, la información espacial contenida en una imagen del mundo real, puede ser representada mediante una matriz de caracteres, donde cada celda corresponde a un objeto en la imagen y la organización de la matriz viene dada por la distribución espacial de los objetos de la imagen. Modificaciones a los algoritmos basados en dicha estructura son propuestos en [73], [32] y [92]. En [159] Wang propone una técnica similar, pero basada en los rectángulos envolventes mínimos (MBR) para representar los objetos en cada imagen y modelar la información espacial como las relaciones entre estos rectángulos envolventes. En [35] se implementa un sistema CBIR que se basa en grafos de proximidad espacial, construidos sobre los objetos de la consulta y las regiones de color detectadas en las imágenes. En [47], se propone un método de recuperación de imágenes basada en regiones, que comprende dos pasos principales: una segmentación gruesa, basada en la cuantización del color en el espacio RGB y una descripción fina de las regiones, considerando la distribución del matiz. La información espacial es almacenada mediante un grafo de regiones adyacentes, en el que cada nodo contiene información relativa a la regiones (área, distribución de color, posición, contornos).

2.6. Métricas y funciones de similitud

Una vez definidas las características empleadas para describir cada imagen, éstas se reúnen en un vector o un conjunto de vectores de características que representarán a la imagen. Datta [36] ha llamado a esta representación

| NOMBRE | EXPRESIÓN |
|-----------------------|--|
| Distancia Manhattan | $d(v_a, v_b) = \sum_{i=1}^n v_a(i) - v_b(i) $ |
| Distancia Ecuclídea | $d(v_a, v_b) = \sqrt{(v_a - v_b)^T (v_a - v_b)}$ |
| Distancia Minkowsky | $d(v_a, v_b) = (\sum_{i=1}^n (v_a(i) - v_b(i))^p)^{1/p}$ |
| Distancia Mahalanobis | $d(v_a, v_b) = \sqrt{(v_a - v_b)^T \Sigma^{-1} (v_a - v_b)}$ |
| Distancia Canberra | $d(v_a, v_b) = \sum_{i=1}^n \frac{ v_a(i) - v_b(i) }{ v_a(i) + v_b(i) }$ |
| Distancia Chebyshev | $d(v_a, v_b) = \max_{i,j=1}^n v_a(i) - v_b(j) $ |

Cuadro 2.1: Métricas utilizadas para calcular la similitud de características en sistemas CBIR

la firma de la imagen. Para realizar consultas, es necesario especificar una medida de similitud, o métrica que permita comparar imágenes y presentar las respuestas en una lista ordenada. En el caso más simple, si la imagen está representada por un vector, se puede adoptar una métrica conocida, e.g. la distancia euclídea, como medida de similitud en el espacio de características, tal como se hace en [79], aunque algunos autores sostienen que no es una buena aproximación a la percepción de similitud en los humanos [135]. En los casos más complicados, el conjunto de características constituye un conjunto de datos heterogéneos para los que no está definida una única función de similitud.

Otras métricas alternativas a la distancia euclídea, son la distancia de Manhattan o city-block [27], al ser considerada estadísticamente más robusta, así como la distancia de Mahalanobis [47]; la distancia de Canberra es utilizada para comparar características de textura en [83]; la distancia de Chebyshev es utilizada en [94] para realizar la comparación de vectores de características una vez que han sido proyectados en el espacio topológico tangente al espacio de características de la imagen. La tabla 2.1 muestra un resumen de las métricas o medidas de distancia mencionadas junto a sus expresiones.

En los sistemas desarrollados para colecciones de imágenes con dominio extenso, e. g. colecciones fotográficas; las imágenes son representadas por

coleccionas de características heterogéneas, algunas de ellas expresadas en forma de vectores, grafos, etc. Esto es, se tienen distancias individuales entre vectores de características del mismo tipo, y una medida de similitud global que combina las individuales. Esta última se obtiene dando diferentes pesos a las distancias calculadas sobre las características comunes. Así, una definición recurrente de la medida de similitud (o disimilitud) general para dos imágenes I_1 e I_2 , está dada por la expresión siguiente [27] :

$$D(I_1, I_2) = \sum_{i=1}^n w_i d_i(v_{1,i}, v_{2,i}), \quad (2.4)$$

donde d_i es la métrica o distancia sobre la característica i ; $v_{1,i}, v_{2,i}$ son los datos extraídos correspondientes a dicha característica en cada imagen, y w_i , es el peso asignado a esta característica en la medida de similitud global. Esta definición de la similitud global permite su refinamiento de acuerdo a la importancia que los usuarios pueden darle a cada característica, mediante el proceso de realimentación por relevancia que será discutido más adelante, y que es una de las técnicas llamadas a reducir la brecha semántica [142].

Un trabajo interesante sobre las medidas de similitud, con referencias a estudios psicológicos puede ser encontrado en [135]. Santini y Jain parten de la asunción generalizada en la literatura de que la similitud entre imágenes (o disimilitud) es una distancia en algún espacio de características, que se asume como un espacio métrico. Los autores hacen una distinción entre la distancia percibida (calculada como una métrica en el espacio formal) y la distancia juzgada (la que es accesible experimentalmente) y posteriormente realizan una comparación entre distancia geométrica con sus axiomas, y las funciones de similitud propuestas en la literatura.

Una de las conclusiones más interesantes que presentan los autores se refiere a la verificación de los cuatro axiomas de las métricas sobre la distancia juzgada:

- Autosimilitud: A es tan parecido a A, como B es tan parecido a B.
- Minimalidad: La distancia de A a A, es más pequeña que la distancia de A, a cualquier otra cosa.
- Simetría: Distancia de A a B es igual a la distancia de B a A.
- Desigualdad triangular: La distancia de A a B es menor o igual que la suma de las distancias de A a C, y B a C.

En los experimentos con sujetos humanos, los dos primeros son cuestionables, el segundo no se cumple y el cuarto no se puede comprobar.

De otro lado, para características específicas, se han definido métricas puntuales de comparación. Cuando se extrae el histograma de color, por ejemplo, la intersección de los histogramas propuesta en por Swain y Ballard [148] es utilizada como función de similitud en gran cantidad de propuestas, aunque surgen otras alternativas como la distancia de Minkowsky [112] o los histogramas acumulativos.

Otras medidas de similitud más complejas pueden surgir, dependiendo de la representación que se haga de la imagen, como en [2], donde Ahmad y Grosky hacen una representación jerárquica de la imagen mediante quadrees, según la distribución de los puntos de interés encontrados en la imagen, -particularmente se usan las esquinas en esta propuesta-. Por tanto la comparación entre imágenes equivale a una comparación entre árboles, con lo cual la función de similitud se asocia al peso obtenido de los nodos de cada árbol.

2.7. Métodos de acceso y búsqueda en base de datos

En las bases de datos de los sistema CBIR, normalmente cada imagen mantiene una relación biunívoca con su vector de características que es el mismo que se utiliza para las operaciones de comparación mediante la función de similitud y la posterior recuperación. Si la base de datos es pequeña, y a pesar de que el vector de características tenga una alta dimensionalidad, una búsqueda exhaustiva secuencial, da resultados aceptables respecto a la rapidez de la consulta. Al crecer la base de datos, se hace necesario utilizar estructuras definidas sobre las bases de datos que permitan un acceso rápido y no se deterioren con el aumento del tamaño de la base de datos. El uso de estructuras típicas de acceso a datos de clave única resulta ineficiente por la alta dimensionalidad de los vectores de características, de manera que los árboles binarios, los métodos basados en B-trees, o tablas *hashing*, resultan poco adecuados para los sistemas CBIR. Ya que estas estructuras se basan en la existencia de un orden total, que en general no se garantiza en los espacios de alta dimensión.

Dentro de los tipos de consulta más habituales que se pueden realizar en

un sistema CBIR o en general en un sistema de datos multidimensionales están los siguientes: a) consultas exactas, que buscan un elemento específico en la base de datos; b) Consultas por rango, que buscan elementos dentro de un rango determinado; c) Consultas del vecino más cercano, en las que se busca el elemento más cercano a un elemento particular dado dentro del espacio de características. Este último tipo es el más habitual en sistemas CBIR.

Cuando hablamos de bases de datos de alta dimensionalidad, las estructuras de árbol han demostrado ser la más adecuadas para la gestión y la realización de consultas. Una de las propuestas pioneras es el R-tree [63], que es una estructura arborea balanceada, especial para datos bidimensionales, habituales por ejemplo, en los sistemas de información geográfica. Su eficiencia se basa en la óptima distribución de una jerarquía de rectángulos envolventes mínimos (*Minimal boundary rectangle* - MBR). En dicha estructura los nodos hoja contienen un identificador del objeto al que apuntan y el MBR que lo contiene. Los nodos internos representan una sucesión de regiones rectangulares minimales que cubren los nodos en el nivel inferior. Las regiones del mismo nivel pueden solaparse y su unión no necesariamente cubre todo el espacio.

Sin embargo, esta estructura inicialmente fue desarrollada para manejar datos en dos dimensiones, de manera que al aumentar la dimensión se pierde eficiencia. Variantes del R-tree, eficientes para tres dimensiones, han sido propuestas en la literatura, tal como el R^+ -tree [140], que maneja regiones disjuntas y el R^* -tree que optimiza los algoritmos cuando son usados en datos de más de dos dimensiones y disminuye el solapamiento de los nodos, aunque con un aumento en el coste de memoria y procesamiento en las operaciones de inserción de nodos. El R^* -tree ha sido utilizado en el QBIC [48], donde la textura se representa por vectores en tres dimensiones, mientras que la forma se representa mediante un vector de 20 valores. Para la textura se ha utilizado directamente el R^* -tree, y para la forma se realiza una reducción de la dimensionalidad mediante el cálculo de los componentes principales, obteniendo vectores de dos o tres componentes, que son adecuados para ser gestionados mediante dicha estructura.

Para poder gestionar datos en espacios de alta dimensión han surgido algunas propuestas basadas también en el R-tree tales como el TV-tree [97], que trata de utilizar en forma “dinámica” los vectores de características empleando solo aquellas características necesarias para discriminar los objetos; el X-tree [11] que añade un algoritmo de división de regiones que busca mini-

mizar el solapamiento, además de incluir “supernodos” con el fin de mantener una estructura jerárquica tanto como sea posible; el Pyramid-tree [10] que gestiona pirámides en lugar de rectángulos, el SPY-TEC [90] que divide el espacio vectorial en hiperesferas, etc. Una completa revisión sobre métodos de acceso y gestión de datos espaciales puede ser consultada en [52].

Algunos autores han tratado de optimizar el proceso de consulta, combinando las estructuras de árbol con otro tipo de técnicas como en [9], donde se propone un método que precalcula los resultados de una búsqueda del vecino más cercano, mediante el cálculo de las celdas de Voronoi de cada punto. La información obtenida es almacenada en una estructura de indexación que posibilita un acceso eficiente posterior. La estructura tiene la ventaja de ser dinámica y permitir su actualización. En [168], se propone un método de indexación, basado en la agrupación (*clustering*) de los vectores de características de la base de datos, soportados por una estructura de árbol; para optimizar el acceso a los nodos hoja, que pueden corresponder a *cluster* con varios vectores, se usa una técnica basada en la desigualdad triangular que reduce el número de comparaciones. En [2], se propone un método de indexación jerárquico, basado en el concepto de firmas de archivos (signatures files) y la comparación mediante quadrees. Cada nivel de la jerarquía reduce el espacio de búsqueda, permitiendo un nivel de búsqueda más refinado sólo para las imágenes potencialmente relacionadas en la base de datos.

Otros, como Ciocca [27], han evitado emplear estructuras complejas, implementando un método de filtrado bastante sencillo, que permite eludir la comparación secuencial de ítems en la base de datos; su técnica se basa en una variante de la desigualdad triangular. El método propone filtrar el número de imágenes que se van a comparar con una consulta Q , a través de la comparación previa de todas las imágenes I de la base de datos con una(s) imagen(es) de referencia K dentro de la base, llamada clave, con lo cual se determina y guarda la distancia entre ellas. Basado en que la medida de similitud cumple con la desigualdad triangular $d(I, Q) \geq |d(I, K) - d(Q, K)|$ se puede reducir de manera significativa el número de imágenes a comparar. En [49], se propone una técnica que realiza un mapeo de los vectores de alta dimensión al espacio $1D$, para luego explotar la eficiencia en la búsqueda ofrecida por el B^+ -tree.

2.8. Aprendizaje y realimentación por relevancia

Para mejorar el rendimiento de los sistemas CBIR se han propuesto algunas técnicas provenientes de la inteligencia artificial, dentro de las cuales, una de las más prominentes es la realimentación por relevancia (*relevance feedback*), la cual ya es bastante conocida en el mundo de la recuperación textual. Cuando un usuario realiza una consulta, el sistema despliega una serie de imágenes con el mayor grado de similitud a la imagen de consulta dada. Los sistemas que permiten la realimentación, dan opción al usuario para seleccionar aquellas imágenes respuesta que se relacionan mejor con el objetivo de su consulta y aquellas que no son importantes (realiza un juicio de relevancia); así el sistema puede aprender sobre las preferencias del usuario para refinar las búsquedas futuras. Esto es, el sistema “aprende” la semántica que el usuario tiene en mente.

Una de las primeras propuestas encontradas en la literatura es la explorada por Ciocca y Shettini [28]. En ella los autores proponen un análisis estadístico de la distribución de características de las imágenes sobre las que los usuarios han emitido algún juicio de relevancia, para determinar qué características han sido tomadas en cuenta y en qué medida. De esta manera se puede acentuar la influencia de esas características en la evaluación general de la similitud de las imágenes, así como en la formulación de una nueva consulta. El método utiliza una “métrica global”, que evalúa la similitud entre dos imágenes, y que consiste en una combinación lineal de las distancias entre las características individuales. Para esto, las distancias individuales son normalizadas, y los pesos de las distancias individuales son establecidos a partir del análisis estadístico de los juicios de relevancia.

En [91], se propone la incorporación de una red neuronal de Función de Base Radial (RBF) en un sistema CBIR para aprovechar la interacción entre el sistema y el usuario. El sistema permite que el usuario evalúe cada una de las imágenes que son desplegadas como respuesta de una consulta, etiquetándolas como “similar”, “parecida” o “diferente”. Dada una imagen de consulta, se calcula un vector de similitud con cada una de las imágenes en la base de datos, compuesto por las métricas relacionadas con cada característica, y este vector es suministrado como entrada a la red neuronal, que ha sido previamente entrenada de manera *off-line*. La red combina en forma no lineal dicho vector de métricas, arrojando un valor entre 0 y 1, que

corresponde al grado de similitud entre cada imagen y la consulta (cercana a 0 muy diferente, cercana a 1 muy similar). La red sigue su entrenamiento online, mediante la realimentación obtenida del usuario, quien puede refinar la búsqueda, valorando su impresión sobre cada imagen obtenida.

En [110], se explora una técnica de aprendizaje no supervisado para incorporar capacidades de autoaprendizaje a los sistemas CBIR. La propuesta busca minimizar la participación del usuario en el proceso de realimentación. Para ello, incorpora una red tipo *Self-organization tree map* (SOTM), que analiza los resultados de una consulta, y en un nuevo espacio de características, realiza un proceso de *clustering* en busca de las imágenes más y menos relevantes. Los resultados obtenidos en este proceso son aplicados para optimizar la realimentación por relevancia basado en la función de Base Radial.

El concepto de aprendizaje en sistemas CBIR es tratado también por Dong y Bhanu [39], quienes proponen modelar la distribución de las imágenes de la base de datos en el espacio de características como una mezcla de funciones de densidad gaussianas y estimar este modelo para conseguir el aprendizaje semántico, utilizando para ello la realimentación obtenida desde el historial de consultas de los usuarios del sistema. En [172], los autores asumen la realimentación por relevancia como un problema típico de clasificación de dos clases (relevantes y no relevantes) y desarrollan una técnica basada en aprendizaje bayesiano con la cual se puede determinar la probabilidad de que una imagen en la base de datos sea relevante o irrelevante a una consulta dada.

Otras propuestas que podemos encontrar en la literatura utilizan técnicas como los vectores de soporte finito [68], [66], el Análisis discriminante [74] o recientemente en [24] los autores proponen un método aplicado a imágenes médicas que permite al usuario definir un *ranking* de relevancia entre imágenes, que realimenta el sistema, para ajustar con dicha información los pesos del vector de características.

2.9. Evaluación en los sistemas CBIR

La evaluación es un aspecto fundamental para el desarrollo de sistemas que sean útiles en aplicaciones prácticas, además de una herramienta que permite comparar el rendimiento entre diferentes sistemas. Para evaluar sistemas CBIR, es importante tener en cuenta tres aspectos esenciales [108]:

- Una colección de datos apropiada para la evaluación.

- El establecimiento de relevancia verdadera (*judged relevance*) contra la que se hacen las validaciones.
- Una métrica de evaluación.

Respecto al primer punto, es deseable que el conjunto de datos sea lo suficientemente grande como para obtener una evaluación estadísticamente significativa. Es habitual dentro de la comunidad de investigadores, probar los sistemas con bases de datos generales comúnmente utilizadas, como estrategia para comparar su rendimiento con el de otros sistemas propuestos. Así, son muy conocidas las colecciones de imágenes de texturas de Brodatz⁵, Vistex⁶ o las colecciones de imágenes de color de Corel y Caltech 101⁷, entre otras. Sin embargo no todas son de libre uso y distribución, lo que limita las posibilidades de una estandarización en la comparación de resultados. En dominios más especializados, como la medicina o la teledetección, son escasas las librerías de imágenes de dominio público disponibles. Este ha sido un problema que preocupa a la comunidad de investigadores en CBIR, y han surgido algunas propuestas, para reunir repositorios de uso común [108], [109].

En cuanto a la relevancia verdadera, ésta es obtenida de diferente forma según las características de cada colección. Las bases de datos estándar, usualmente están catalogadas por tópicos (imágenes aéreas, cebras, etc.), e. g. la colección de Corel. Sin embargo, ítems pertenecientes a un mismo grupo pueden presentar características visuales diferentes, por lo cual es deseable una inspección visual previa. En áreas más especializadas, puede realizarse algún tipo de agrupamiento o categorización con la intervención de expertos, estableciendo la verdad del terreno en imágenes en las cuales las características de asociación no necesariamente son de fácil percepción visual. Para colecciones más generales, algunos autores han propuesto simular los juicios de similitud del usuario [158], añadiendo ruido a la métrica utilizada en el sistema CBIR. Sin embargo, como se discute en [108], la mejor manera de saber qué es lo que espera el usuario como respuesta a una consulta, es con su propia intervención, y debería hacerse a la totalidad de la base de datos, o a una parte representativa de ésta.

⁵www.ux.uis.no/~tranden/brodatz.html

⁶<http://vismod.media.mit.edu/vismod/imagery/VisionTexture/vistex.html>

⁷http://www.vision.caltech.edu/Image_Datasets/Caltech101/

Con respecto a las métricas de evaluación, los investigadores en sistemas CBIR han adoptado algunas provenientes del área de recuperación de información. Específicamente, son utilizadas con mayor frecuencia la medida de **precisión** (*precision*) y la medida de **recuerdo** (*recall*).

Si tenemos una colección de imágenes C , y una consulta q , supongamos que el conjunto puede ser dividido en dos subconjuntos de la siguiente forma:

- el subconjunto $R(q)$ de imágenes respuesta revelantes a la consulta,
- el subconjunto $\bar{R}(q)$, de imágenes respuesta irrelevantes a la consulta.

Además, denominamos como $A(q)$, al conjunto de imágenes devueltas por el sistema como respuesta a la consulta q . La precisión (*precision*), se refiere al porcentaje de imágenes devueltas por el sistema que son relevantes a la consulta:

$$p = \frac{A(q) \cap R(q)}{A(q)}. \quad (2.5)$$

El recuerdo (*recall*), se refiere al porcentaje de imágenes relevantes a la consulta q , que son devueltas por el sistema, esto es:

$$r = \frac{A(q) \cap R(q)}{R(q)}. \quad (2.6)$$

Los resultados obtenidos con estas dos medidas son a menudo presentados en un gráfico precisión/recuerdo. Aunque estas dos medidas han dado buenos resultados en el área de recuperación de información, que está centrada en información textual, en los sistemas CBIR se tienen algunos inconvenientes [142]. Por un lado, la selección del conjunto de imágenes relevantes de la base de datos por parte de sujetos humanos, conlleva el problema de la interpretación subjetiva que éste puede hacer de cada imagen, que puede ser ambiguo si no se contextualiza el objetivo de una consulta dada. De otro lado, la gran mayoría de los sistemas CBIR, retornan un número k de imágenes más cercanas o similares a la imagen de consulta, siendo k fijo, con lo cual, si el número de imágenes relevantes es mayor que k , el recuerdo (*recall*), pierde significado.

A pesar de dichos inconvenientes, estas medidas pueden ser útiles bajo circunstancias especiales, por ejemplo, en colecciones de imágenes de dominio estrecho, sobre las cuales se tenga información semántica consistente, suministrada mediante etiquetas o algún tipo de descripción textual.

Otras medidas propuestas son, la **evaluación del objetivo** [33], que cuenta el número de imágenes que el usuario tuvo que examinar, antes de encontrar su objetivo. La **tasa de error** [76], que determina la porción de imágenes no relevantes, del total de imágenes recuperadas. En [108] se proponen las siguientes medidas:

- Rango (R) y Rango normalizado (\tilde{R}): entendiéndose el rango como el número de imágenes de respuesta que anteceden la primera imagen relevante. El rango normalizado de imágenes relevantes está determinado en la siguiente expresión:

$$\tilde{R} = \frac{1}{NN_R} \left(\sum_{i=1}^{N_R} R_i - \frac{N_R(N_R - 1)}{2} \right), \quad (2.7)$$

donde R_i , es el rango en el cual, la primera imagen relevante aparece, y N y N_R son el número de imágenes y el número de imágenes relevantes respectivamente. Esta medida es 0 para un rendimiento perfecto, y se acerca a 1 cuando el rendimiento empeora.

- $P(20)$, $P(50)$ y $P(N_R)$: Medida de precisión después de 20, 50 y N_R imágenes recuperadas.
- $R_P(0,5)$ y $R(1000)$: Recuerdo, cuando la precisión es 0.5 y después de que 100 imágenes son recuperadas.

2.10. Conclusiones

En este capítulo hemos realizado una revisión de los aspectos generales tenidos en cuenta en el desarrollo de sistemas CBIR. Hemos observado que existen semejanzas estructurales en los sistemas propuestos, con lo cual es posible definir una arquitectura común. Sobre ésta, algunos investigadores señalan la importancia de la modularidad para permitir sistemas escalables. Con respecto a la extracción de características, hemos visto que en los años tempranos la atención estuvo centrada en el desarrollo e implementación de técnicas de procesamiento de la imagen, y aunque la dedicación a este aspecto ha disminuido frente a los demás, dicha área aun sigue desarrollándose. El acceso a grandes bases de datos de alta dimensionalidad es un área creciente, donde las estructuras jerárquicas de tipo árbol siguen marcado el rumbo.

Respecto al estudio de las medidas de similitud, este ha ido de la mano con la aplicación de métodos de aprendizaje y realimentación. Dos cuestiones que al parecer están marcando el rumbo hacia un posible estrechamiento de la brecha semántica, la cual está determinada por la limitada noción de similitud de los sistemas, en comparación con la gran riqueza semántica del usuario. Finalmente, el proceso de evaluación de los sistemas CBIR es un área abierta, influenciada por las técnicas heredadas desde la recuperación de información, y en la cual se requieren herramientas para la comparación de métodos en igualdad de circunstancias como bases de datos estándar, con información sobre la verdad del terreno si es el caso, pero sin perder de vista el hecho que dichos sistemas deben ser orientados al usuario final, con lo cual se debe tener en cuenta la subjetividad en cada contexto.

Capítulo3

Filtros de Gabor

Estudios pioneros en el campo de la neurofisiología realizados por Hubel y Weisel [75, 161] proporcionan evidencia sobre la forma en que las células del córtex visual en algunos mamíferos analizan los datos obtenidos por los campos receptivos localizados en la retina, sugiriendo que las células simples actúan como detectores de líneas, franjas o bordes rectilíneos. Además, cada uno de estos detectores es sensible a un número específico de orientaciones y frecuencias. En el mismo ámbito, investigadores [19] han concluido que el sistema visual humano (SVH) está constituido por un sistema de canales psicofísicos, cada uno sintonizado a una cierta banda de frecuencia espacial, los cuales operan linealmente y en paralelo e independientemente uno del otro. En [14] se discute el hecho de que estos canales son selectivos a un estrecho rango de frecuencias espaciales y a una dirección específica. Estos y otros estudios sobre la percepción humana han sugerido que el SVH hace un análisis multiresolución y multiescala, susceptible de ser modelado mediante un esquema piramidal que pueda descomponer una imagen en múltiples escalas y direcciones.

Bajo este paradigma algunos estudios han sugerido que los bancos de filtros permiten la extracción de la información en la imagen para una mejor segmentación o clasificación. En particular, en esta sección se tratará una aproximación basada en las funciones elementales de Gabor [51] que tienen la propiedad de ser altamente selectivas en el dominio del espacio, así como en el dominio de la frecuencia. Gabor realizó sus trabajos en señales con dominio unidimensional y sus principios fueron posteriormente extendidos por Daugman [37] a señales con dominio bidimensional (imágenes).

3.1. La transformada de Gabor

De la teoría del tratamiento de señales, sabemos que una señal puede ser representada tanto en el dominio temporal como en el de la frecuencia, mediante el uso de la transformada de Fourier. Para una función $f(x)$ su transformada de Fourier viene dada mediante la expresión:

$$\mathcal{F}(w) = \int_{-\infty}^{\infty} f(x) e^{-iwx} dx,$$

donde w es una variable real continua. La función original, puede ser reconstruida a partir de sus componentes de Fourier mediante la transformada inversa de Fourier:

$$f(x) = \int_{-\infty}^{\infty} \mathcal{F}(w) e^{iwx} dw.$$

La transformada de Fourier es ampliamente utilizada en el procesamiento y análisis de señales con resultados satisfactorios para el caso de señales periódicas las cuales presentan cierta regularidad, pero no ocurre lo mismo para el caso de señales cuyo espectro varía con el tiempo, denominadas *no-estacionarias*, puesto que cada representación nos ofrece una información completa de la señal en el dominio del tiempo, o en el de la frecuencia, pero no podemos tener a la vez la información de ambos dominios. Es decir, si se tiene la representación temporal, sabemos como se comporta una señal en un instante de tiempo determinado, pero no sus componentes frecuenciales en ese instante. De otro lado, si se tiene la representación frecuencial, es posible saber la aportación a la señal de una determinada frecuencia, pero sin tener alguna información sobre la localización temporal.

Una solución al problema de la representación conjunta en los dominios temporal y frecuencial fue aportada por Dennis Gabor en 1946 en su muy conocido artículo “*Teoría de la Comunicación*”. [51], en donde introdujo la que él denominó *función ventana*, mediante la cual se puede delimitar en el tiempo la función a analizar, antes de desarrollar la descomposición frecuencial. De esta manera, Gabor introdujo la denominada *Transformada de Fourier en Tiempo Corto*, también conocida como *transformada de Fourier con ventana deslizante* o *transformada de Gabor*.

El proceso de la transformada consiste en la elección de una función ventana $g(x)$ la cual debe ir deslizándose a lo largo del tiempo mediante un factor

de traslación τ . Para ello se toma la función a analizar, se multiplica por la función ventana y seguidamente se lleva a cabo la transformación frecuencial, como lo muestra la siguiente expresión:

$$G_{\tau}f(w) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} f(x) g(x - \tau) e^{-iwx} dx.$$

Con esta transformación lo que se busca es obtener información local sobre la distribución de frecuencias en el intervalo de análisis alrededor del instante τ , con lo cual se obtiene cierta localización temporal de la transformada de Fourier. Para buscar el compromiso óptimo de localización tanto en el dominio de Fourier como en el dominio del tiempo, Gabor utilizó una ventana con soporte fijo, y de hecho demostró que una *función Gaussiana* es la que presenta mejor localización en ambos dominios, con lo cual la función ventana adecuada resulta ser:

$$g_{\sigma}(x) = \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{1}{4\sigma}x^2},$$

donde el parámetro σ se mantiene fijo y determina el ancho de la ventana. Por tanto, la transformada de Fourier de tiempo Corto o transformada de Gabor queda determinada por la expresión:

$$G_{\tau}^{\sigma}f(w) = \int_{-\infty}^{\infty} f(x) g_{\sigma}(x - \tau) e^{-iwx} dx.$$

Ahora bien, si se define la familia de funciones $g_{\sigma,\tau,w}(x)$ en la forma:

$$g_{\sigma,\tau,w}(x) = e^{jwx} g_{\sigma}(x - \tau),$$

la transformada de Gabor se puede considerar de una manera alternativa, como sigue:

$$G_{\tau,w}^{\sigma}f(w) = \int_{-\infty}^{\infty} f(x) \bar{g}_{\sigma,\tau,w}(x) dx,$$

con lo cual se tiene una descomposición de $f(x)$ con la familia de funciones $g_{\sigma,\tau,w}(x)$ en el mismo dominio de $f(x)$. Esta descomposición alternativa está directamente ligada con la transformada Wavelet.

3.2. Funciones de Gabor 2D

Gracias a Daugman [37], los principios de Gabor fueron extendidos al dominio bidimensional, para ser aplicados al análisis de señales y desde entonces se han propuesto diversas aplicaciones en una gran cantidad de artículos científicos. En forma general una función de Gabor bidimensional puede ser expresada como:

$$g(x, y) = \left(\frac{1}{2\pi\sigma_x\sigma_y} \right) \exp \left[-\frac{1}{2} \left(\frac{x'^2}{\sigma_x^2} + \frac{y'^2}{\sigma_y^2} \right) \right] \exp [2\pi i (Ux + Vy)]$$

donde (x', y') son las coordenadas rectangulares (x, y) rotadas en el dominio del espacio:

$$\begin{aligned} x' &= x \cos(\theta) + y \operatorname{sen}(\theta), \\ y' &= -x \operatorname{sen}(\theta) + y \cos(\theta). \end{aligned}$$

Así, una función de Gabor es una función gaussiana modulada con una sinusoidal compleja. En la figura 3.1 la parte (a) ilustra la parte real e imaginaria de una sinusoidal compleja la parte (b) muestra la forma de la gaussiana y la parte c) las máscaras formadas por la parte real e imaginaria de la sinusoidal modulada. Los parámetros σ_x y σ_y determinan la extensión espacial y el ancho de banda del filtro. El eje mayor de la gaussiana se rota por un ángulo θ alrededor del eje positivo z .

Si se denota por (u, v) las coordenadas rectangulares en el dominio de la frecuencia, el punto (U, V) representa una frecuencia particular en el plano. La exponencial compleja, corresponde a una sinusoidal compleja de frecuencia $F = \sqrt{U^2 + V^2}$ y orientación $\gamma = \tan^{-1}(V/U)$.

La transformada de Fourier de $g(x, y)$ viene dada por la expresión:

$$H(u, v) = \exp \left\{ -\frac{1}{2} \left[(\sigma_x [u - U]')^2 + (\sigma_y [v - V]')^2 \right] \right\},$$

donde

$$[(u - U)', (v - V)'] = [(u - U) \cos\theta + (v - V) \operatorname{sen}\theta, -(u - U) \operatorname{sen}\theta + (v - V) \cos\theta].$$

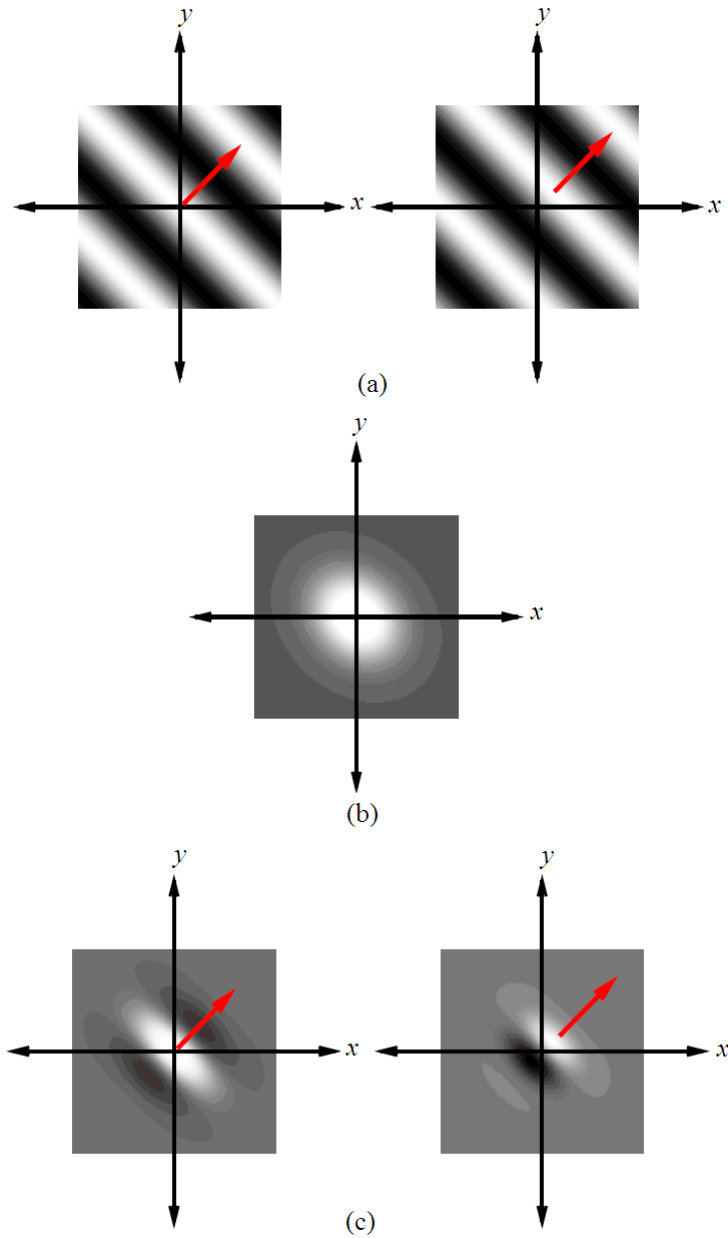


Figura 3.1: Filtro de Gabor en el dominio espacial. a) Partes real e imaginaria de la sinusoidal compleja. b) Gaussiana rotada sobre el origen c) Máscaras formadas por las funciones sinusoidales real e imaginaria moduladas por la gaussiana.

Así, la función de transferencia del filtro (en el dominio de la frecuencia) tiene la forma de una función gaussiana. La longitud del eje mayor de la gaussiana está determinado por (σ_x, σ_y) que está rotado por un ángulo θ respecto al eje positivo u , y la gaussiana está centrada en el punto (U, V) del plano de las frecuencias, de tal forma que la función de Gabor actúa como un filtro paso banda en el dominio transformado de Fourier.

Las funciones de Gabor han sido también implementadas para el análisis de texturas usando sólo su componente real. En tal circunstancia la función de respuesta al impulso del filtro está definido como:

$$g(x, y) = \left(\frac{1}{2\pi\sigma_x\sigma_y} \right) \exp \left[-\frac{1}{2} \left(\frac{x'^2}{\sigma_x^2} + \frac{y'^2}{\sigma_y^2} \right) \right] \cos [2\pi Fx].$$

En este caso la función de transferencia viene dada por dos gaussianas localizadas simétricamente, tal como sigue:

$$H(u, v) = \exp \left\{ -2\pi^2 [(\sigma_x [u - F])^2 + (\sigma_y v)^2] \right\} + \exp \left\{ -2\pi^2 [(\sigma_x [u + F])^2 + (\sigma_y v)^2] \right\}.$$

En cualquier caso, tomando tanto la versión compleja como la real de los filtros, hay que fijar un total de seis parámetros en el momento de implementarlos, a saber: $(F, \theta, \sigma_x, \sigma_y, B_F, B_\theta)$. El ancho de banda frecuencial (B_F) y el ancho de banda angular (B_θ) pueden ser establecidos como valores constantes, como frecuentemente se encuentra en la literatura, de acuerdo a los resultados obtenidos en los estudios psicovisuales. La frecuencia (F) y orientación (θ) definen la localización del centro del filtro. Para determinar las varianzas espaciales de la ventana gaussiana, se pueden usar las siguientes ecuaciones [29]:

$$\begin{aligned} \sigma_x &= \frac{\sqrt{\ln 2} (2^{B_F} + 1)}{\sqrt{2}\pi f (2^{B_F} + 1)} \\ \sigma_y &= \frac{\sqrt{\ln 2}}{\sqrt{2}\pi f \tan \left(\frac{B_\theta}{2} \right)} \end{aligned}$$

El ancho de banda frecuencial (B_F) y el ancho de banda angular (B_θ) se fijan habitualmente considerando las curvas (elipses) constituidas por el conjunto de puntos del dominio de Fourier con magnitud de respuesta igual a un medio de la magnitud máxima de la respuesta del filtro. En la figura 3.2 se representa la elipse formada por dichos puntos, y se indican las distancias que corresponden a B_F y B_θ .

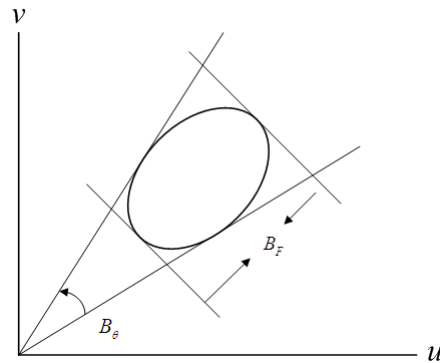


Figura 3.2: Elipse de puntos con respuesta igual a la mitad de la magnitud en el dominio de la frecuencia

3.3. Bancos de Filtros de Gabor

Los filtros de Gabor han demostrado ser adecuados para representar conjuntamente información en los dominios frecuencial y espacial, consiguiendo minimizar el principio de incertidumbre inherente al plano espacio/frecuencia [37], esto sumado a que permiten simular el comportamiento de los campos receptivos de células simples del cortex visual, ha conducido al desarrollo de un considerable número de propuestas disponibles en la literatura para el diseño de conjuntos de filtros, denominados comúnmente Bancos de Filtros de Gabor (BFG), donde cada filtro está sintonizado en una frecuencia y orientación particulares, para ser aplicados a una imagen y de este modo extraer sus características.

Durante los años 80's y 90's múltiples investigaciones han sido publicadas entorno a la eficacia de los Filtros de Gabor en el análisis de texturas. En este particular, podemos encontrar trabajos como el de Turner[156], quien describe una extensa experimentación que busca probar la efectividad de los filtros de Gabor en la discriminación de texturas. En su trabajo, explica cómo un conjunto de filtros sintonizados a diferentes frecuencias y orientaciones, fijando los parámetros de acuerdo a los modelos basados en la evidencia biológica para modelar la corteza visual, se aplican sobre imágenes con texturas sintéticas (generadas por ordenador). Estas texturas están diseñadas especialmente para probar su efectividad en la detección de patrones con marcada orientación, o elementos de textura con diferente distribución

espacial. Bovik y Clark [17], en la misma estela desarrollan ideas similares aplicando BFG a texturas reales y sintéticas. En cuanto a la optimización de los parámetros de los filtros, las propuestas de Dunn en [42][41], se enfocan en la selección de valores adecuados para un único filtro que permita distinguir entre dos texturas. Mas adelante Weldon [160] extiende estas ideas aumentando el número de filtros y la cantidad de texturas diferentes a distinguir. Otra propuesta de optimización de los BFG para la segmentación de imágenes texturadas sin un conocimiento *a priori* de las texturas se expone en [153]. Por la efectividad encontrada en los experimentos previos, los bancos de filtros se proponen para ser aplicados en problemas como la recuperación de imágenes [103][166], o la inspección industrial y detección de defectos [16].

Para aplicar los bancos de filtros se procede habitualmente de la siguiente manera: una imagen de entrada $I(x, y)$ se convoluciona con un filtro de Gabor $g(x, y)$ para obtener una imagen característica de Gabor $G(x, y)$ como se muestra en la siguiente expresión:

$$G(x, y) = \int \int I(m, n) g(x - m, y - n) dm dn.$$

De tal manera que un conjunto de filtros $\{g_{\Theta}, \Theta = (F, \theta, \sigma_x \sigma_y, B_F, B_{\theta})\}$ genera un conjunto de imágenes características. En aplicaciones de clasificación o recuperación de información, se realiza algún proceso de selección de características relevantes sobre estas imágenes respuesta o imágenes de salida, el cual depende del problema que se esté tratando. El proceso llevará a la configuración de un vector de características para cada imagen.

Dentro de las propuestas encontradas en la bibliografía sobre la configuración de los bancos de filtros podemos distinguir dos tipos. El primer tipo lo constituyen aquellos que proponen un solapamiento de las curvas de soporte de los filtros¹ en el dominio de Fourier con el objetivo de subsanar la no ortogonalidad de la familia de funciones de Gabor. Esta aproximación permite obtener una imagen aceptable en el proceso de reconstrucción de la señal. El segundo tipo lo constituyen aquellos que quieren evitar el solapamiento de las curvas de soporte de los filtros, para evitar la redundancia en la información de las imágenes características y obtener una mejor configuración de los vectores de características a nivel de pixel. La figura 3.3, ilustra estos dos tipos de particiones del espacio de Fourier.

¹Llamamos curva de soporte a la curva que tiene respuesta del filtro igual a la mitad de la magnitud máxima.

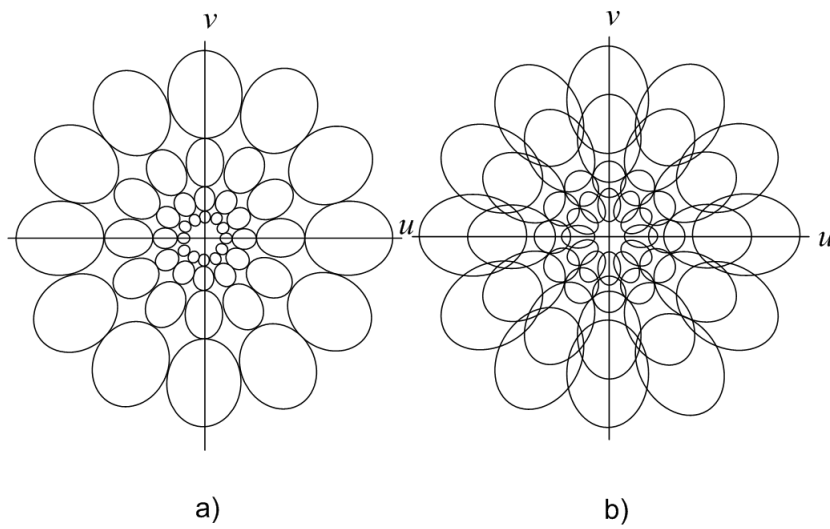


Figura 3.3: Recubrimiento del plano de Fourier por un Banco de Filtros de Gabor. a) Sin solapamiento en la respuesta de media-magnitud, b) Con solapamiento en la respuesta de magnitud media.

Capítulo 4

Análisis Wavelet

En este capítulo se hace una revisión general sobre el análisis de señales y en particular de imágenes mediante el uso de wavelets. Inicialmente se hace una definición de lo que en la literatura se conoce como wavelet¹, damos algunos ejemplos de dichas funciones y pasamos a definir formalmente la transformada continua wavelet, su inversa y sus propiedades. Posteriormente el capítulo se enfoca a la versión discreta de la transformada y la técnica conocida como análisis multiresolución que ha sido desarrollada para la implementación eficiente de la transformada discreta. Finalmente extendemos los conceptos al caso bidimensional, lo que nos permite la aplicación de la transformada wavelet discreta al análisis de imágenes.

Para la confección de esta revisión de los elementos teóricos sobre análisis wavelet nos hemos basado en los textos [125], [152], [145], [80], artículos clásicos [102], [101], [100], y otros artículos complementarios como [155], [95], [8].

4.1. Wavelets Continuas

4.1.1. Definición de wavelet

Una wavelet es una función que satisface ciertas condiciones especiales: muestra una concentración localizada de energía finita en su dominio y exhibe cierta oscilación. La primera condición atribuye a la función la propiedad de

¹En el desarrollo de este capítulo y de la tesis hemos decidido asumir la palabra wavelet como un neologismo. Conocemos algunas traducciones de esta palabra como “ondulillas”, “ondículas” u “onditas”, todas ellas nos parecen menos precisas para describir el concepto que la palabra inglesa.

estar bien localizada, mientras que a la segunda se debe que tenga forma de onda, por lo cual se da el nombre de wavelet, que puede traducirse como "ondita". Una wavelet simple puede generar una familia de funciones wavelets mediante traslaciones y dilataciones de la función original parametrizadas de forma discreta o continua.

Si $\psi(x)$ es una función wavelet básica, entonces el conjunto $\{\psi_{s,t}(x), s \in S, t \in T\}$ de todas las versiones de $\psi(x)$ dilatadas por s y trasladadas por t , es la familia wavelet generada por $\psi(x)$, donde S es el conjunto de los posibles escalados, T es el conjunto de las posibles traslaciones. Para esta familia de funciones se llama a $\psi(x)$ la *wavelet madre*, o *wavelet analizadora*. El principal interés en desarrollar familias de wavelets radica en su utilidad para representar y caracterizar señales. La transformada wavelet continua (TWC) de una función $f(x)$ se define como el producto interno de $f(x)$ con respecto a la familia wavelet generada por $\psi(x)$:

$$TWC(f) = \{\langle f(x), \psi_{s,t}(x) \rangle, s \in S, t \in T\}, \quad (4.1)$$

Para cada punto (s, t) , la magnitud de la transformada wavelet de una señal $f(x)$ describe cuánto se parece ésta a la versión dilatada y trasladada de $\psi(x)$.

La función wavelet madre $\psi(x)$ debe ser seleccionada de tal manera que permita reconstruir la función $f(x)$ a partir del conocimiento de su transformada wavelet, es decir la TWC debe ser invertible. Para asegurar la invertibilidad de la transformada se requiere que $\psi(x) \in L^1(\mathbb{R})$ y que tenga media cero; esto último es equivalente a decir que $\mathcal{F}(\psi)(0) = 0$, donde $\mathcal{F}(\cdot)$ denota la transformada de Fourier. Esta condición es suficiente para garantizar que se satisface el requerimiento de que la wavelet tenga alguna oscilación.

4.1.2. Ejemplos de wavelets

Existe una colección de funciones diversas que cumplen los requerimientos descritos anteriormente y por lo tanto pueden ser utilizadas como wavelet madre. En esta subsección se presentan algunos ejemplos de ellas.

4.1.2.1. Wavelet de Haar

Las Wavelets de Haar son ampliamente utilizadas debido a que, cuando los valores de dilatación y translación de la familia wavelet se restringen a potencias enteras de 2 (escala diádica), se obtiene una base ortonormal,

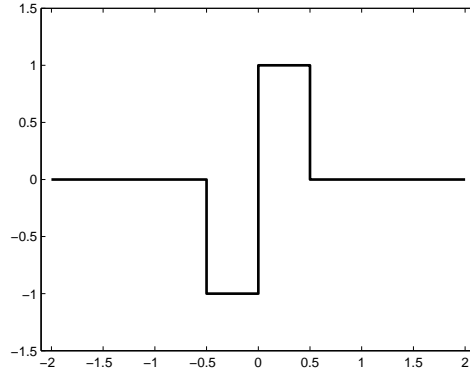


Figura 4.1: Wavelet de Haar

cualidad exhibida sólo por un pequeño grupo de funciones. La figura 4.1 muestra la wavelet de Haar.

La Wavelet Haar formalmente se define como:

$$\psi_{Haar}(x) = \begin{cases} 1 & 0 \leq x < 1/2, \\ -1 & 1/2 \leq x < 1 \\ 0 & \text{en otro caso} \end{cases} \quad (4.2)$$

4.1.2.2. Wavelet de Shannon

La característica más destacada de la wavelet de Shannon (o wavelet *sinc*) es que su transformada de Fourier es constante sobre algún intervalo de frecuencias excluyendo el origen, y cero en cualquier otro lugar del espacio de frecuencias. Es decir, es un filtro pasa banda ideal en el dominio de la frecuencia. La función wavelet madre está caracterizada en la ecuación siguiente y es representada en la figura 4.2.

$$\psi_{Shannon}(x) = 2sinc(2x) - sinc(2x),$$

donde

$$sinc(x) = \frac{\sin(\pi x)}{\pi x}.$$

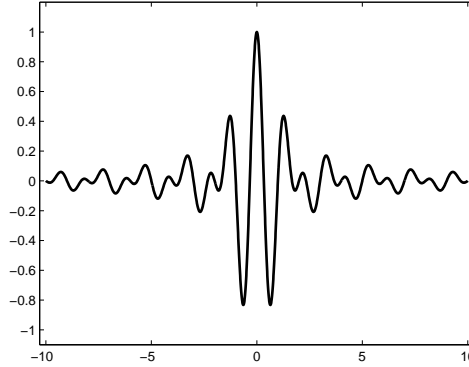


Figura 4.2: Función wavelet de Shannon

4.1.2.3. Wavelet de Morlet

Como las wavelets de Shannon, la wavelet de Morlet también se especifica en el dominio de la frecuencia. En este caso, es una función gaussiana localizada en torno a una frecuencia (media = w_0) y con un ancho de banda (varianza) específico. Sin embargo, ya que la wavelet de Morlet, en el dominio espacial, es una función gaussiana modulada por una función sinusoidal, el soporte de la wavelet en el dominio de la frecuencia no es un intervalo compacto, sino todo el eje de la frecuencia. Más precisamente la wavelet de Morlet y su transformada de Fourier están dadas por las expresiones siguientes:

$$\psi_{Morlet}(x) = \pi^{-1/4} \left(e^{jwx} - e^{-1/2w_0^2} \right) e^{-1/2x^2}.$$

$$\mathcal{F}(\psi_{Morlet})(w) = \pi^{-1/4} \left[e^{-\frac{1}{2}(w-w_0)^2} - e^{-\frac{1}{2}w^2} e^{-\frac{1}{2}w_0^2} \right].$$

La figura 4.3 muestra un ejemplo de la wavelet de Morlet. Esta wavelet está íntimamente relacionada con los bancos de filtros de Gabor que se describen en el capítulo anterior.

4.1.3. La Transformada Wavelet Continua

En esta sección se define en forma rigurosa la transformada wavelet continua (TWC) de una función y se discuten algunas de sus propiedades. La

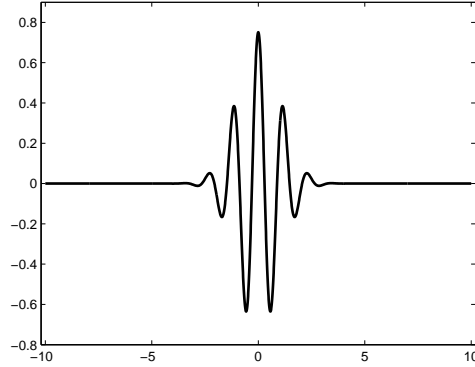


Figura 4.3: Parte real de la Wavelet de Morlet

TWC puede ser definida como una aplicación dependiente de la especificación de una función atómica $\psi(x)$ llamada wavelet madre y consiste en la representación de una señal $f(x)$ con respecto a la familia de wavelets generada por $\psi(x)$. Como se ha expuesto antes $\psi(x)$ debe satisfacer ciertas condiciones, pero existe una gran flexibilidad en cuanto a la posibilidad de elegir candidatas a funciones wavelet madre, las cuales se validan mediante el criterio de admisibilidad que se expone más adelante.

4.1.3.1. Definición

Para un espacio de Hilbert² general \mathcal{H} , la TWC puede ser definida como una aplicación $W_\psi : \mathcal{H} \rightarrow W_\psi(\mathcal{H})$ parametrizada por la función $\psi(x)$. Específicamente la TWC de una función unidimensional $f(x) \in L^2(\mathbb{R})$ está dada por:

$$W_\psi : L^2(\mathbb{R}) \longrightarrow W_\psi(L^2(\mathbb{R}))$$

$$f \longmapsto \left\{ \langle f, \psi_{s,t} \rangle_{L^2(\mathbb{R})} ; s \in S, t \in T \right\},$$

donde $\psi_{s,t}(x)$ es una versión de $\psi(x)$ dilatada por s y trasladada por t , es decir:

²Un espacio vectorial completo, dotado de una norma y producto interior, en nuestro caso tomamos como referencia a $L^2(\mathbb{R})$: el conjunto de funciones de cuadrado integrable.

$$\psi_{s,t}(x) = |s|^{-1/2} \psi\left(\frac{x-t}{s}\right). \quad (4.3)$$

Así, la TWC de una señal unidimensional $f(x)$ es una función bidimensional de variables reales $t, s > 0$ y puede ser escrita como sigue:

$$W_\psi(f)(s,t) = \int_{-\infty}^{\infty} \bar{\psi}_{s,t}(x) f(x) dx = \langle \psi_{s,t}, f \rangle = \Psi_{f,s,t}, \quad (4.4)$$

donde $\bar{\psi}_{s,t}$ es el conjugado complejo de $\psi(x)$.

La transformada wavelet asigna para cada valor particular de s y t un valor numérico (complejo) que describe el grado de similitud entre la función $f(x)$ y la versión dilatada y trasladada de $\psi(x)$.

4.1.3.2. Transformada Wavelet Inversa

La transformación inversa de la TWC, permite regresar al dominio original desde el dominio transformado. Para garantizar la reconstrucción de una señal a partir de los coeficientes de su transformada wavelet, la función wavelet madre debe cumplir con el criterio de admisibilidad (ecuación 4.6). Esta condición limita el número de funciones wavelet disponibles. La satisface cualquier función de media cero, que sea absolutamente integrable, esto es, cuya integral absoluta esté acotada.

Una señal $f(x)$ puede ser recuperada a partir de su transformada continua wavelet $W_\psi(f)(s,t)$ mediante la siguiente fórmula de reconstrucción:

$$f(x) = C_\psi^{-1} \int \int W_\psi(f)(s,t) \psi_{s,t}(x) dt ds \quad (4.5)$$

dónde C_ψ es un valor constante llamado *constante de admisibilidad* que depende de la función wavelet madre $\psi(x)$ elegida, y debe satisfacer la condición de admisibilidad:

$$C_\psi = \int_{-\infty}^{\infty} \frac{|\mathcal{F}(\psi)(\xi)|^2}{|\xi|} d\xi < \infty \quad (4.6)$$

siendo $\mathcal{F}(\psi)(\xi)$ la transformada de Fourier de $\psi(x)$. La condición anterior implica que $\mathcal{F}(\psi)(0) = 0$ con lo cual se debe cumplir que:

$$\int \psi(x) dx = 0 \quad (4.7)$$

de donde se puede deducir que cualquier función $\psi \in L^1(\mathbb{R})$ con media cero es una wavelet admisible.

4.1.3.3. Propiedades

Sean $a, b \in \mathbb{R}$, y $f, f_1, f_2 \in L^2(\mathbb{R})$. La TWC con respecto a una wavelet madre admisible $\psi(x)$, satisface las siguientes propiedades [152], [145].

1. Linealidad. Es una propiedad evidente a partir del producto escalar:
 $(W_\psi(af_1 + bf_2))(s, t) = a(W_\psi(f_1))(s, t) + b(W_\psi(f_2))(s, t)$,
2. Invarianza a traslación. Dada una función $f(x)$, con transformada $W_\psi(f)(s, t)$, al trasladarla $f(x+x_0) = f_1(x)$ se obtiene $W_\psi(f_1)(s, t) = W_\psi(f)(s, t+x_0)$. Es decir, la transformada wavelet de la función trasladada es la traslación de la transformada original.
3. Conservación de la energía.

$$\int_{-\infty}^{\infty} |f(x)|^2 dx = \frac{1}{C_\psi} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} |W_\psi(f)(s, t)|^2 \frac{dsdt}{s^2},$$

Es decir, la energía de la función puede medirse en cualquiera de los lados de la transformada. La transformada preserva la energía de la función original y la constante de admisibilidad es el factor de amplificación de la energía al realizar la transformada.

4. Escalado. Para una función dada $f(x)$ y su transformada wavelet $W_\psi(f)$ se cumple que al cambiar la escala de la función $f_1 = \frac{1}{\sqrt{s_0}} f\left(\frac{x}{s_0}\right)$, su transformada también se escala, $W_\psi(f_1)(s, t) = W_\psi(f)\left(\frac{s}{s_0}, \frac{t}{s_0}\right)$.
5. Localización. La TWC presenta una localización variable en el plano espacio-frecuencia. En particular, para altas frecuencias (correspondientes a escalas pequeñas), se tiene buena localización temporal y para las bajas frecuencias (que corresponde a escalas grandes) se tiene buena localización frecuencial.

6. Regularidad. La reconstrucción de funciones a partir de aproximaciones de muy baja resolución, mediante el análisis wavelet, solo es posible si las funciones wavelet madre y de escalado (que se introduce más adelante) son continuamente derivables. El orden de regularidad es el número de veces que es posible realizar la derivada sobre estas funciones. Este es un aspecto muy importante en el momento de realizar algún análisis específico como la descomposición piramidal.

4.2. La Transformada Wavelet Discreta

Como se ha visto en las secciones anteriores, la TWC es una función bidimensional definida en un espacio continuo, por lo que no se puede calcular en máquinas discretas de precisión finita. Sin embargo, es posible realizar una aproximación, con una precisión aceptable a la transformada continua, mediante un muestreo que permita la obtención de un conjunto enumerable de puntos del plano traslación-escala. Cada conjunto discreto de puntos en este plano permite especificar un conjunto contable de wavelets, que a su vez especifica una transformada wavelet discreta particular. Para que estas transformadas tengan algunas propiedades deseables, por ejemplo la invertibilidad, la discretización en el plano traslación-escala no puede obtenerse de manera arbitraria. Debido a que es la más generalizada, siempre que se aluda a la *transformada wavelet discreta* (TWD), se estará haciendo referencia a la transformada asociada con el muestreo de la malla diádica sobre el plano que se muestra en la figura 4.4, la cual viene dada por la expresión:

$$\Gamma_D \iff \{(2^{-n}m, 2^n)\}_{m,n \in \mathbb{Z}}$$

Este conjunto discreto de puntos permite una implementación eficiente del cálculo de la TWD, siempre y cuando se cumplan las siguientes condiciones:

1. El conjunto discreto muestreado debe ser la malla diádica Γ_D .
2. La familia wavelet $\psi_{t,s}(x)$, con $(s,t) \in \Gamma_D$, debe formar una base ortonormal.
3. La función wavelet madre debe ser de soporte compacto.

En el proceso de construcción de una versión discreta de la transformada wavelet, es deseable que los valores de dilatación se obtengan como las potencias enteras de una escala de referencia s_0 , normalmente mayor que 1. Así

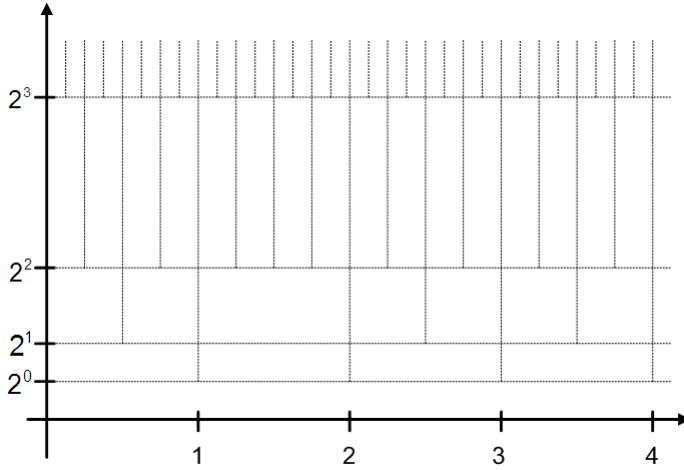


Figura 4.4: Descomposición del plano mediante la discretización por muestreo de la malla diádica

tendremos que $s = s_0^j$. Por otra parte, el parámetro de traslación, dependerá también del parámetro de escala, por lo cual $t = kt_0 s_0^j$, siendo k y t_0 valores enteros. De esta forma, para cada valor de la escala de referencia s_0 se define una discretización específica del espacio traslación-escala. Teniendo todo esto en cuenta, la familia wavelet discreta queda expresada como:

$$\psi_{j,k}(x) = s_0^{-j/2} \psi(s_0^{-j}(x - kt_0 s_0^j)) = s_0^{-j/2} \psi(s_0^{-j}x - kt_0). \quad (4.8)$$

Eligiendo adecuadamente los valores s_0 y t_0 , es posible cumplir las condiciones enumeradas antes. En particular una elección adecuada sería $s_0 = 2$ y $t_0 = 1$, con lo cual se obtiene un muestreo concordante con la malla diádica que se introdujo antes.

Así pues, dada una función $f(x) \in \mathcal{H}$ obtenemos una versión discretizada de su transformada wavelet continua que estará dada como:

$$W_\psi(f)(j, k) = \int f(x) \psi_{j,k}(x) dx, \quad (4.9)$$

donde las wavelets $\psi_{j,k}(x)$ son generadas a partir de la traslación y dilatación de la wavelet madre $\psi(x)$ en la siguiente forma:

$$\psi_{j,k}(x) = 2^{-j/2} \psi(2^{-j}x - k), \quad (4.10)$$

La función $f(x)$ puede ser reconstruida a partir de los coeficientes wavelets dicretos $W(j, k)$, mediante la fórmula

$$f(x) = \sum_j \sum_k \psi_{j,k}(x) W_\psi(f)(j, k). \quad (4.11)$$

4.2.1. Análisis Multiresolución

En esta sección se introduce un marco matemático para la construcción de bases de wavelets ortonormales en un espacio de Hilbert \mathcal{H} . En particular, en esta tesis nos restringimos al espacio $L^2(\mathbb{R})$. Para aplicaciones de la teoría wavelet a problemas del mundo real, es necesario dar con una adecuada representación discreta de la función que queremos analizar y de la wavelet madre. Sobre estas representaciones, a una escala particular, y en una región compacta de algún espacio, obtendremos una *representación wavelet aproximada* en dicha escala [125]. Usando el algoritmo de Mallat sobre esta aproximación, podemos obtener una representación multiresolución (y multiescala), de la función muestreada, sobre un número finito de escalas desde la más fina hasta la más gruesa.

El Análisis Multiresolución es un marco teórico propuesto por Mallat [?], mediante el cual se puede realizar el análisis de una señal discreta en múltiples bandas de frecuencia, y a distintos niveles de resolución espacial, proporcionando un algoritmo rápido de cálculo de la TWD derivado de la descomposición piramidal.

Definición: Sea $\{V_j : V_j \in \mathcal{H}, j \in \mathbb{Z}\}$ una secuencia creciente de subespacios y $\phi \in V_0$. La dupla $(\{V_j\}, \phi)$ es llamada un Análisis Multiresolución de \mathcal{H} si:

1. Existe una función de escalado $\phi \in V_0$, tal que $\{\phi(x - k)\}_{k \in \mathbb{Z}}$ es una base ortonormal para V_0 .
2. Si $f(x) \in V_j$, entonces $f(2x) \in V_{j-1}$, y
3. $\bigcup V_j = \mathcal{H}$ y $\bigcap V_j = \{0\}$.

La estructura básica del esquema multiresolución está formada por la secuencia de subespacios cerrados

$$\{0\} \subset \dots V_1 \subset V_0 \subset V_{-1} \dots \subset \mathcal{H}$$

así como por la *función de escalado* ϕ cuyas translaciones enteras $\{\phi(x - k)\}$ forman una base ortonormal en el subespacio V_0 . Cuando el índice $j \rightarrow -\infty$, la resolución se incrementa y el espacio V_j se aproxima cada vez más a \mathcal{H} . La escala sigue el orden inverso: conforme aumenta la resolución disminuye la escala, esto es, el soporte de la función de escalado correspondiente a ese nivel de resolución se hace más pequeño. Debido a que las funciones en V_0 sólo pueden ser combinaciones lineales de translaciones enteras de ϕ , éstas son menos detalladas que muchas otras funciones en \mathcal{H} , en el sentido de que su variación está restringida por la variación de ϕ . Es decir, podemos considerar a V_0 como una versión de baja resolución de \mathcal{H} . Los subespacios en un nivel superior tienen una resolución más alta que aquellos de niveles bajos. Para una función arbitraria $f(x) \in \mathcal{H}$, el subespacio V_j contendrá una función que aproxime $f(x)$ a la escala j , de tal manera que a menores valores de j se tienen espacios V_j correspondientes que contienen aproximaciones de $f(x)$ con más alta resolución.

De otro lado, es deseable que el aumento en el nivel de resolución de un espacio V_j al V_{j-1} sea uniforme, lo cual se asegura ya que cada espacio V_j tiene una base ortonormal relacionada mediante la dilatación por una potencia de 2 con la base generada por $\phi(x)$ en V_0 . Esto es consecuencia directa del punto 2 de la definición del análisis multiresolución, por el cual si $\{\phi(x - k)\}$ es una base ortonormal en V_0 , entonces $\{2^{-j/2}\phi(2^{-j}x - k)\}$ es una base ortonormal de V_j .

4.2.2. Bases de wavelets ortonormales en el análisis multiresolución

Para generar una base ortonormal sobre un espacio de Hilbert \mathcal{H} , es posible usar el marco de análisis multiresolución descrito previamente. El conjunto formado por las dilataciones y translaciones de la función de escalado, forman bases ortogonales para subespacios de resolución j . Sin embargo, para generar una base ortonormal sobre todo un espacio de Hilbert \mathcal{H} , se requiere la ortogonalidad entre las bases definidas en cada una de las resoluciones, lo cual no se cumple ya que cada subespacio está totalmente contenido en los de resolución superior, por lo cual la unión de las bases de cada subespacio V_j no es una base ortogonal general en el espacio \mathcal{H} .

Para conseguir una base ortogonal en un análisis multiresolución, se define una secuencia auxiliar de subespacios $\{W_j : W_j \subseteq \mathcal{H}\}$, donde cada W_j es complementario a V_j y ortogonal a V_{j-1} para cada j . Los W_j caracterizan las diferencias entre V_j y V_{j-1} . Estos son los *subespacios wavelet*, que se definen formalmente como el complemento ortogonal del subespacio V_j en el siguiente subespacio de mayor resolución V_{j-1} , es decir:

$$V_{j-1} = V_j \oplus W_j, \quad (4.12)$$

y

$$V_j \perp W_j.$$

Sin pérdida de generalidad, se puede considerar el subespacio W_0 . El objetivo es construir una función $\psi(x) \in W_0$ con la propiedad de que $\{\psi(x-k)\}$ forma una base ortonormal para W_0 . Si encontramos tal función, $\psi(x) \in W_0$, a la cual se denomina *wavelet madre*, el conjunto de sus traslaciones y dilataciones diádicas $\{2^{-j/2}\psi(2^{-j}x-k)\}_{j,k \in \mathbb{Z}}$ podría formar una base ortonormal para todo el espacio \mathcal{H} . Esto ocurre por dos razones: (i) cada subespacio W_j es el subespacio wavelet de j -ésima resolución de \mathcal{H} , por lo cual para un valor fijo j , el conjunto de funciones $\{2^{-j/2}\psi(2^{-j}x-k)\}_{k \in \mathbb{Z}}$ es una base ortonormal para W_j , y (ii) los subespacios wavelet son mutuamente ortogonales ya que:

$$V_{j-1} = V_j \oplus W_j = V_{j+1} \oplus W_{j+1} \oplus W_j = \dots = V_J \oplus \bigoplus_{i=j}^J W_i, \forall j < J. \quad (4.13)$$

donde \oplus denota la suma de espacios vectoriales. Como consecuencia de la propiedad de completitud $\bigcup V_j = \mathcal{H}$ de una descomposición multiresolución $\{V_j\}$ se tiene que $\lim_{j \rightarrow \infty} V_j = \{0\}$, de donde \mathcal{H} puede ser escrito como la suma directa de todos los subespacios wavelet, es decir:

$$\mathcal{H} = \dots W_1 \oplus W_0 \oplus W_{-1} \dots = \bigoplus_{j=-\infty}^{\infty} W_j. \quad (4.14)$$

4.2.3. Algoritmo de descomposición piramidal

La función de escalado y la función wavelet madre, junto con sus versiones trasladadas forman dos bases ortonormales distintas en el nivel de resolución básico. La dilatación de la función de escalado básica y de la wavelet

madre, respectivamente, producen sendas bases ortonormales a cada nivel de resolución. Si $\phi(x)$ es la función de escalado básica en el espacio V_0 , esta puede ser expresada como combinación lineal de los elementos de la base $\{\phi(2x - k)\}$ generado por $\phi(2x)$ (la función de escalado en el espacio V_{-1}), de manera que se satisface la siguiente relación:

$$\phi(x) = \sqrt{2} \sum_k g(k) \phi(2x - k), \quad (4.15)$$

esta expresión se puede interpretar como la proyección de la función de escalado $\phi(x)$ del espacio V_0 en un espacio de mayor resolución V_{-1} . Los términos de la secuencia $g(k)$ son denominados *coeficientes interescala* y corresponden a un filtro discreto paso bajo.

Si $\psi(x) \in W_0$ es la wavelet madre, ésta puede ser desarrollada en la base ortonormal $\{\phi(2x - k)\}$ del espacio V_{-1} según la siguiente expresión:

$$\psi(x) = \sqrt{2} \sum_k h(k) \phi(2x - k) \quad (4.16)$$

donde los términos de la secuencia $h(k)$ son denominados *coeficientes interescala* y corresponden a un filtro discreto paso alto.

Sea $f(x) \in V_0$, esto es, $f(x)$ es una función perteneciente al primer nivel de resolución. Esta puede ser escrita como una combinación lineal de las versiones trasladadas de la función de escalado, $\phi(x - k)$ como sigue:

$$f(x) = \sum_k c_0(k) \phi(x - k), \quad (4.17)$$

donde

$$c_0(k) = \langle f, \phi_{0,k} \rangle = \int f(x) \phi(x - k) dx. \quad (4.18)$$

En el siguiente nivel de menor resolución, donde $j = 1$, ya que $V_0 = V_1 + W_1$, podemos descomponer $f(x) \in V_0$ como:

$$f(x) = \sum_k c_1(k) \phi_{1,k} + \sum_k d_1(k) \psi_{1,k}, \quad (4.19)$$

donde

$$c_1(k) = 2^{-1/2} \sum_n g(n - 2k) c_0(n), \quad (4.20)$$

$$d_1(k) = 2^{-1/2} \sum_n h(n-2k) c_0(n), \quad (4.21)$$

siendo

$$\phi_{j,k} = \phi(2^{-j}x - k).$$

y

$$\psi_{j,k} = \psi(2^{-j}x - k).$$

En el segundo miembro de la ecuación 4.19 se aprecian dos términos, el primero de los cuales indica una aproximación suavizada de f en el espacio de resolución $j = 1$, mientras que el segundo término indica la información de detalle. La secuencia $d_1(k)$ representa las diferencias entre la función $f \in V_0$ y su aproximación $f_1 \in V_1$, y son denominados *coeficientes wavelets discretos*. Ya que los coeficientes c_j y d_j pueden ser calculados a partir de c_{j-1} y d_{j-1} respectivamente, la descomposición en aproximaciones suavizadas y detalles a menos resolución se puede generalizar como:

$$f_j(x) = \sum_k c_j(k) \phi_{j,k} + \sum_k d_j(k) \psi_{j,k}, \quad (4.22)$$

donde

$$c_j(k) = 2^{-j/2} \sum_n g(n-2k) c_{j-1}(n), \quad (4.23)$$

$$d_j(k) = 2^{-j/2} \sum_n h(n-2k) c_{j-1}(n). \quad (4.24)$$

De esta forma, es posible seguir la descomposición de manera iterativa hasta un nivel de descomposición J , representando a la función $f(x)$ mediante una aproximación menos fina, y una serie con los valores de detalle. Analíticamente esto es:

$$f(x) = \sum_k 2^{-J/2} c_J(k) \phi(2^{-J}x - k) + \sum_{j=1}^J \sum_k 2^{-j/2} d_j(k) \psi(2^{-j}x - k). \quad (4.25)$$

La secuencia original de valores $c_0(n)$ puede ser reconstruida a partir de las secuencias de valores $c_i(n)$ y $d_i(n)$, para $0 < j \leq J$, donde $j = J$ corresponde

a la menor resolución en la descomposición. El valor del coeficiente $c_{j-1}(n)$ puede ser obtenido mediante la expresión:

$$c_{j-1}(n) = 2^{-j/2} \sum_k c_j(k) g(n-2k) + 2^{-j/2} \sum_k d_j(k) h(n-2k), \quad (4.26)$$

donde g y h corresponden a los filtros paso alto y paso bajo que se convolucionan con los coeficientes de aproximación y con los coeficientes wavelet en el nivel j , respectivamente. Las dos expresiones anteriores son la base para el algoritmo de descomposición y reconstrucción piramidal propuesto por Mallat [101].

4.3. Transformada Wavelet en dos dimensiones

Para aplicaciones en imágenes digitales, la Transformada Wavelet se extiende al caso 2D. La Transformada Wavelet de una imagen digital $f(x, y)$ se define como:

$$W_f(u, v, s) = \frac{1}{s} \int \int f(x, y) \psi\left(\frac{x-u}{s}, \frac{y-v}{s}\right) d(x, y), \quad (4.27)$$

La descomposición wavelet multiresolución unidimensional, puede ser extendida al caso bidimensional mediante la introducción de funciones de escalado y wavelet 2D separables, derivadas de sus correspondientes unidimensionales. De esta manera se obtiene:

$$\phi_{LL}(x, y) = \phi(x) \cdot \phi(y), \quad (4.28)$$

$$\psi_{LH}(x, y) = \phi(x) \cdot \psi(y), \quad (4.29)$$

$$\psi_{HL}(x, y) = \psi(x) \cdot \phi(y), \quad (4.30)$$

$$\psi_{HH}(x, y) = \psi(x) \cdot \psi(y), \quad (4.31)$$

Típicamente la transformada wavelet 2D se implementa mediante la convolución de la imagen con filtros discretos 1D, los cuales deben ser de reconstrucción perfecta, es decir que la composición de la transformación directa

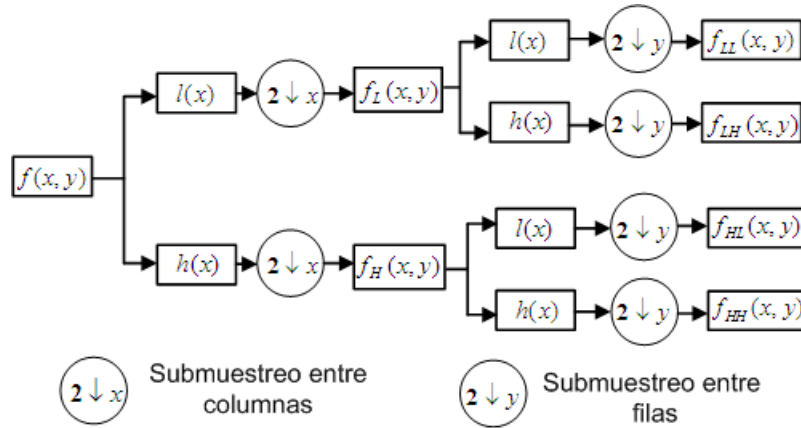


Figura 4.5: Esquema para un nivel descomposición multiresolución de la imagen mediante el algoritmo piramidal

y de la transformación inversa, sin cancelación o modificación de los coeficientes transformados, nos devuelve la señal original. La operación de análisis wavelet 2D es separable, se puede descomponer en dos transformadas 1D idénticas (dadas por los mismos filtros paso bajo L con funciones de respuesta al impulso $l(x)$ y paso alto H con funciones de respuesta al impulso $h(x)$) realizadas primero sobre las filas y luego sobre las columnas de la imagen. Primero, aplicando los filtros para cada fila en la imagen $f(x, y)$ y submuestreando por filas, se generan las matrices de coeficientes $f_L(x, 2^{-1}y)$ y $f_H(x, 2^{-1}y)$. Seguidamente se aplican los filtros para cada columna en $f_L(x, 2^{-1}y)$ y $f_H(x, 2^{-1}y)$ y se submuestrean las columnas, de donde se generan cuatro subimágenes: $f_{LL}^1(2^{-1}x, 2^{-1}y)$, $f_{LH}^1(2^{-1}x, 2^{-1}y)$, $f_{HL}^1(2^{-1}x, 2^{-1}y)$ y $f_{HH}^1(2^{-1}x, 2^{-1}y)$ para el primer nivel de descomposición. La subimágen f_{LL}^1 corresponde a una versión suavizada o menos detallada de la imagen original, mientras las restantes f_{LH}^1 , f_{HL}^1 y f_{HH}^1 son imágenes que representan los detalles en las direcciones horizontal, vertical y diagonal. El proceso de análisis se itera sobre la imagen $f_{LL}^j(x, y)$ para obtener sucesivas descomposiciones a resolución decreciente (escala creciente): $f_{LL}^{j+1}(2^{-j-1}x, 2^{-j-1}y)$, $f_{LH}^{j+1}(2^{-j-1}x, 2^{-j-1}y)$, $f_{HL}^{j+1}(2^{-j-1}x, 2^{-j-1}y)$ y $f_{HH}^{j+1}(2^{-j-1}x, 2^{-j-1}y)$.

La figura 4.5 ilustra un estado en el proceso de descomposición piramidal de una imagen. La figura 4.6 muestra los resultados de la descomposición en dos niveles.

| | | | |
|------------------------------------|------------|------------|--------------------------------------|
| f_{LL}^3 | f_{LH}^3 | f^2_{LH} | Horizontal detail $f^1_{LH}(x,y)$ |
| f_{HL}^3 | f_{HH}^3 | | |
| f^2_{HL} | | f^2_{HH} | Diagonal detail $f^1_{HH}(x,y)$ |
| Vertical detail $f^1_{HL}(x,y)$ | | | |

Figura 4.6: Representación de una imagen con tres niveles de descomposición

| | HAAR | DAUBICHIES |
|-----|------------------------------|--|
| PBD | $\frac{1}{\sqrt{2}} [1, 1]$ | $\frac{1}{4\sqrt{2}} [1 - \sqrt{3}, 3 - \sqrt{3}, 3 + \sqrt{3}, 1 + \sqrt{3}]$ |
| PAD | $\frac{1}{\sqrt{2}} [-1, 1]$ | $\frac{1}{4\sqrt{2}} [-1 - \sqrt{3}, 3 + \sqrt{3}, -3 + \sqrt{3}, 1 - \sqrt{3}]$ |
| PBI | $\frac{1}{\sqrt{2}} [1, 1]$ | $\frac{1}{4\sqrt{2}} [1 + \sqrt{3}, 3 + \sqrt{3}, 3 - \sqrt{3}, 1 - \sqrt{3}]$ |
| PAI | $\frac{1}{\sqrt{2}} [1, -1]$ | $\frac{1}{4\sqrt{2}} [1 - \sqrt{3}, -3 + \sqrt{3}, 3 + \sqrt{3}, -1 - \sqrt{3}]$ |

Cuadro 4.1: Filtros paso bajo y paso alto para las transformadas wavelet directa e inversa

La inversa de la transformada wavelet 2D puede ser implementada mediante dos transformadas 1D inversas idénticas (con filtros de reconstrucción paso bajo \tilde{L} con función de respuesta al impulso $\tilde{l}(x)$ y paso alto \tilde{H} con función de respuesta al impulso $\tilde{h}(x)$), aplicadas por filas y por columnas sucesivamente. Para la señal 1D (fila o columna) se realiza el sobremuestreo (intercalando valores cero), se aplica los filtros \tilde{L} y \tilde{H} , y se suman los resultados de estos filtros para obtener la imagen reconstruida por filas o columnas. La figura 4.7 muestra un estado en el proceso de reconstrucción.

Dentro de la amplia cantidad de filtros discretos propuestos para implementar la transformada wavelet discreta los más utilizados son los de Haar y Daubechies. En la tabla 4.1 se muestran los filtros directos e inversos.

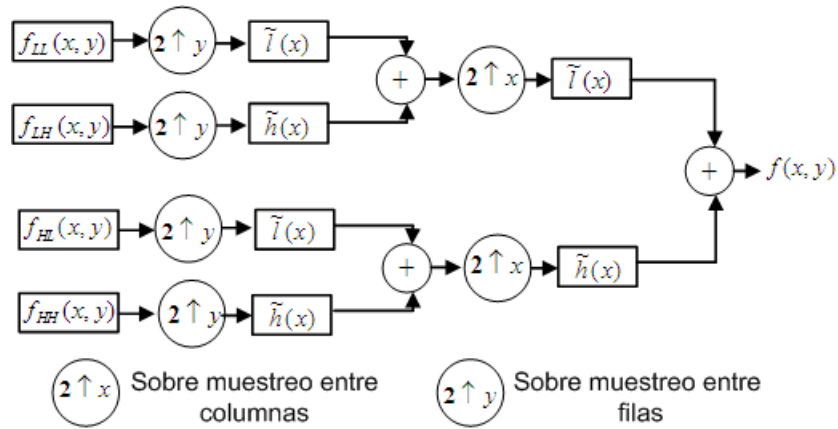


Figura 4.7: Esquema para la reconstrucción de una imagen multiresolución mediante el algoritmo piramidal

Capítulo 5

Control de calidad del papel reciclado

En el proceso de inspección de calidad del papel reciclado, un defecto que se aprecia en algunas ocasiones es la aparición de ondulaciones en las hojas de papel a escala macroscópica, que puede surgir algún tiempo después de su fabricación. Vamos a llamar a este fenómeno *abollado*. En este capítulo se explora la detección y medida de dicho fenómeno mediante técnicas de tratamiento de imágenes y reconocimiento de patrones, como instrumento de diagnóstico en el lugar de producción. Los trabajos aquí expuestos tienen origen en la solicitud de la empresa Papelaria S.A, para estudiar dicho fenómeno. El interés de la empresa era crear un control predictivo de la calidad del papel. Para ello el primer paso es definir un índice cuantitativo de calidad que permita formular la realimentación desde la observación de la calidad del papel a los factores de producción.

Este capítulo está dividido de la siguiente manera: En la sección 5.1 se realiza una introducción al problema, en la sección 5.2 se describen las consideraciones metodológicas tenidas en cuenta en el desarrollo del trabajo. El proceso de adquisición de las imágenes es descrito en la sección 5.3. La sección 5.4 está dedicada a explicar el procedimiento de etiquetado manual realizado para obtener la verdad del terreno que hemos tomado como base para la clasificación automática. El proceso de definición de características basado en técnicas de análisis de texturas es expuesto en la sección 5.5. En la sección 5.6 se explica como se realizó el proceso de clasificación automática sobre los vectores definidos en la sección anterior. La sección 5.7 está dedicada a describir la forma en que se obtuvo un índice para cuantificar el nivel de abollado y su proceso de validación. Finalmente, en la sección 5.8 se hace una discusión a los resultados del trabajo descrito en este capítulo.

5.1. Descripción del problema

En la producción del papel reciclado la gran variabilidad de las materias primas produce una gran variabilidad de la calidad del producto final. Las inhomogeneidades en la distribución de las fibras a nivel microscópico pueden ser la causa de la aparición de ondulaciones y protuberancias de diversos tamaños y formas a nivel macroscópico en la superficie de papel. Al hablar de nivel macroscópico nos estamos refiriendo a una escala espacial del orden de los centímetros. Llamamos a este defecto *abollado*. Es deseable contar con una herramienta de diagnóstico de la calidad del papel basada en las características a nivel macroscópico, para poder establecer un modelo predictivo que permita relacionar el índice de calidad del papel y la materia prima usada y otros factores fijados en el momento de la producción. Para poder realizar esta medida de forma no invasiva y no destructiva, las técnicas más apropiadas son las de visión por computador y reconocimiento de patrones.

La primera pregunta que nos planteamos al abordar este problema es si este defecto ha sido ya caracterizado por la industria de Pulpa y Papel, y si existen métodos de medida establecidos para este defecto.

Las industria de la Pulpa y Papel, así como la industria de Impresión, han establecido una serie de estándares de calidad, junto con sus procesos de medición. ISO y TAPPI son las organizaciones de estandarización más reconocidas por la industria. Han definido estándares para diversas características del papel. En una revisión exhaustiva¹ hemos encontrado la siguiente taxonomía de estándares de calidad más frecuentes en la industria:

- Propiedades físicas: Gramaje (ISO 536, Tappi T410), Contenido de humedad (ISO 286, Tappi T412), Grosor (ISO 534, Tappi T411), Absorción del agua (Tappi T441), Rugosidad (Tappi T555, Tappi T538).
- Propiedades de Resistencia: Resistencia a curvarse y Rigidez (ISO 2493, ISO 5628, Tappi T489, Tappi T543, Tappi T556, Tappi T566), resistencia a reventar (ISO 2758, ISO 2759, Tappi T403), *Ply Bond/ Scott Bond*

¹La metodología para esta revisión fue la siguiente: en primer lugar revisamos la literatura de fabricación de papel, en segundo lugar se exploró la información pública de los organismos de estandarización ISO y TAPPI, en tercer lugar se examinó la información comercial de las empresas que ofertan técnicas de evaluación de la calidad.

(Tappi T403), Rigidez (Tappi T489 and ISO 2491) resistencia al desgaste (ISO 1974, Tappi T414), Resistencia a la tensión (ISO 1924, Tappi T404, Tappi T494), Resistencia a los pliegues (ISO 5626, Tappi T511), Resistencia de cohesión Interna (Tappi T569), Tensión en dirección Z (Tappi T541).

- Propiedades ópticas: Color (Tappi T524, Tappi T527), Luminosidad (Tappi T452), Brillo (Tappi T480), Opacidad (ISO 2471, Tappi T425) Blancura (ISO 11475).
- Propiedades de Impresión: Moteado, Granosidad, Oscuridad, Contraste, etc. (ISO/IEC DIS 13660).

Ninguna de estas medidas guarda relación con el defecto en el cual estamos interesados. Por lo tanto nos hemos visto obligados a definir todo lo relacionado con el cálculo del índice de abollado.

Como antecedentes de nuestra aproximación, en la literatura hemos encontrado ejemplos en los que se han aplicado las técnicas de visión por ordenador para detección de defectos en el papel. En [18] se describe la arquitectura de un sistema para la detección de defectos sobre imágenes de pulpa de papel. Este sistema identifica algunos defectos previamente conocidos, como el *pitch* (mancha generada a partir de material resinoso, usualmente oscuro, proveniente de la madera y/o resinas internas) y el *shive* (un “bulto” generado por la separación incompleta de las fibras de madera, que puede aparecer como una imperfección sobre la hoja terminada). Además de ser muy específicos, estos defectos no están relacionados con la fabricación de papel reciclado, y son visualmente muy diferentes al abollado. En [31] se utilizan técnicas de visión por computador para analizar la distribución de la tensión local durante una prueba de resistencia, con el fin de determinar las concentraciones de tensión debidas a una distribución inhomogénea de las fibras. En [134] se hace una evaluación visual del *mottling* (moteado) que se define como el desvanecimiento indeseado de la densidad de impresión percibida, este fenómeno es claramente diferente al abollado, aunque algunas de las imágenes pueden guardar alguna similitud. Otros ejemplos de aplicación de técnicas de procesamiento de imagen en la industria del papel enfocados al análisis de estructuras y patrones microscópicos pueden ser consultados en [67] y [82], cuyo objetivo era establecer la relación entre la estructura microscópica y propiedades macroscópicas del papel como la permeabilidad transversal, el tamaño promedio de los poros en la mayor parte de la hoja, la anisotropía de

la resistencia, la luminosidad, la opacidad y la dispersión. Sin embargo, estos estudios se refieren a las fibras y microestructuras en el papel reciclado, por lo cual no hay una relación directa con el abollado. Aunque sería deseable la realización de estudios de este tipo para buscar las causas microscópicas del abollado, no disponemos de las herramientas ni la información para ello. Nuestros trabajos están en una fase previa que busca describir cuantitativamente el abollado.

En general podemos encontrar algunos ejemplos de aplicaciones de inspección de calidad basados en técnicas de visión y análisis de texturas en la industria textil en [136], [30], [25], [3], [86], [1], [113], [139], para la clasificación de losas de marmol [105], y para la detección de defectos en superficies de madera en [50], todas ellas bajo condiciones y con objetivos diferentes al problema en el cual estamos interesados.

Para plantear la construcción de un sistema de control de calidad mediante la clasificación visual de texturas es deseable que se cumplan las siguientes premisas:

- Las texturas a ser detectadas deben estar bien definidas mediante un conjunto de imágenes de ejemplo,
- La variabilidad dentro de cada clase de textura debe ser pequeña. Cada imagen de ejemplo debe estar razonablemente bien clasificada mediante una cuidadosa inspección realizada por expertos humanos. Esto significa que debe haber un alto grado de concordancia entre los distintos expertos en la clasificación de cada imagen

Sin embargo, al tratar de plantear el problema de caracterización del abollado de las hojas de papel reciclado mediante clasificación de imágenes en texturas, hemos encontrado las siguientes dificultades:

- La definición de clases no es trivial y no tenemos información de partida.
- La asignación de las muestras de imágenes dentro de las clases no está dada y no parece trivial.
- Hay una alta variabilidad en las características de las texturas en las imágenes que podrían pertenecer a una misma clase y hay una falta de concordancia entre los expertos humanos cuando realizan el etiquetado.

En consecuencia el mayor obstáculo que encontramos en el diseño de un sistema de inspección basado en la medida del abollado es la inexistencia de la propia definición cualitativa y/o cuantitativa del abollado. La definición cualitativa consistiría en la definición de un conjunto de clases de abollado sobre las que estuvieran de acuerdo los expertos humanos. La definición cuantitativa consistiría en un índice numérico que evalúe el grado de abollado sobre cuyos valores existiría una alta concordancia entre los expertos humanos. Por lo tanto el primer paso es concretar cualitativa y cuantitativamente dicha definición, mediante experimentos de etiquetado manual que implica la participación de expertos.

5.2. Consideraciones metodológicas

Como se ha comentado en la sección anterior no existen para este problema resultados formales anteriores, necesitamos establecer la metodología completa de trabajo y justificarla en relación a los resultados deseados. En esta sección describimos en detalle las consideraciones metodológicas que enmarcan nuestro trabajo.

- Exploración preliminar de las imágenes. En este punto hemos examinado visualmente las muestras, para familiarizarnos con el fenómeno. Este no ha sido un proceso formal por lo cual no se produce un resultado documental.
- Establecimiento de la verdad del terreno mediante etiquetado manual de las imágenes. Puesto que el concepto de abollado es un concepto difuso y no hay establecidas unas categorías a priori, el proceso de creación de la verdad del terreno no es un proceso trivial. Implica establecer primero un número de clases manejable por los expertos, es decir que sean capaces de discriminar adecuadamente las imágenes de cada clase. Además, los expertos tienen que obtener un alto grado de concordancia en sus observaciones para que la verdad del terreno quede bien establecida. Este proceso implica la repetición de los experimentos de clasificación, hasta obtener un adecuado grado de concordancia entre los observadores, estableciendo un nuevo número de clases o fusionando las clases definidas con anterioridad.

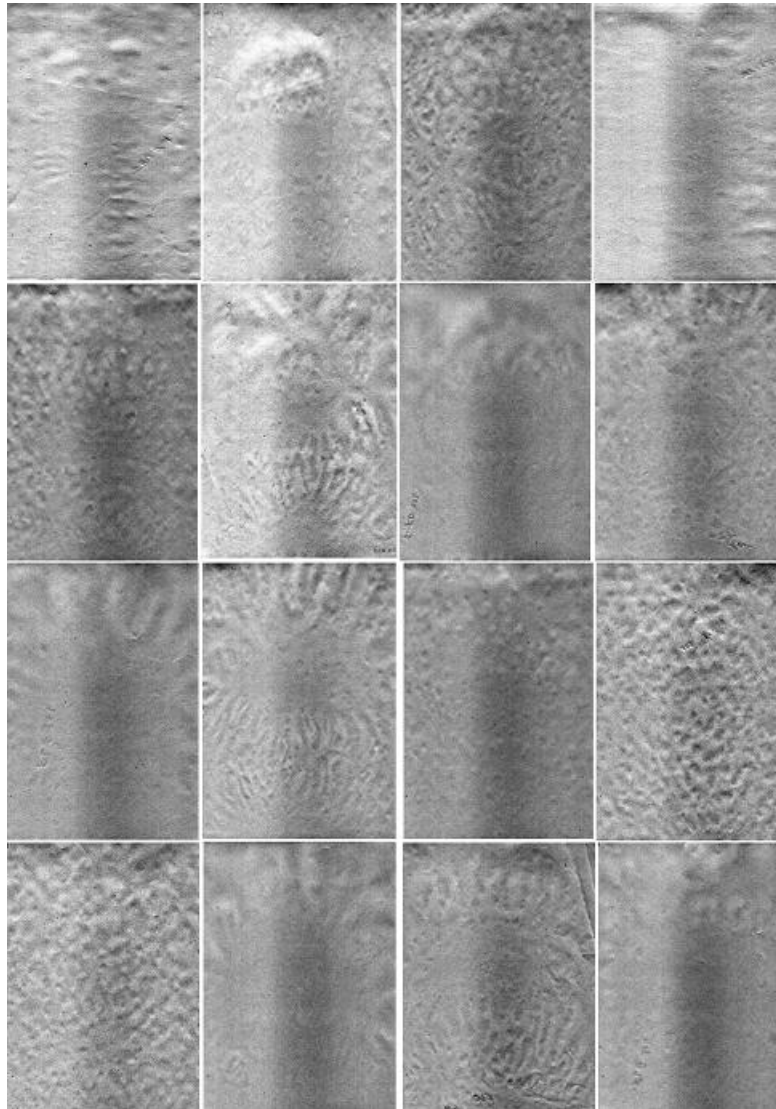


Figura 5.1: Ejemplos de imágenes de papel reciclado con abollado. Las imágenes han sido preprocesadas para mejorar el contraste.

- Extracción de características. El problema es considerado un problema de análisis de texturas, por lo cual se utilizan técnicas convencionales como los Bancos de Filtros de Gabor y Transformada Wavelet Discreta, para la construcción de los vectores de características. Los factores de diseño serán los distintos parámetros de configuración de los filtros y la transformada, tales como: la orientación de cada filtro, el tamaño de los campos receptivos, la frecuencia a la que son selectivos, la wavelet madre utilizada, el número de niveles de descomposición, etc.
- Clasificación automática de imágenes. Hemos probado dos técnicas convencionales, aplicadas sobre las características de textura, que son el K-NN y una red neuronal tipo Perceptrón Multicapa (MLP). Se han ejecutado una serie de experimentos de clasificación automática sobre la muestra de imágenes, con metodología de validación *two-fold cross validation*, donde la muestra se parte en dos conjuntos, uno de entrenamiento y otro para test. La principal utilidad de este proceso es la de evaluar el potencial discriminador de las características obtenidas mediante las técnicas de análisis de texturas, como paso previo al establecimiento de un índice continuo basado en dichas características. Además realizamos una segunda verificación del proceso de clasificación, mediante la aplicación de una batería de clasificadores implementados sobre una herramienta de software libre, y hemos comparado estos resultados con los obtenidos previamente.
- Evaluación de un índice continuo de abollado. Comprobada la utilidad de las técnicas de análisis de texturas aplicadas, hemos procedido a buscar un escalar que permita cuantificar el nivel de abollado en el papel, actuando como un índice continuo. Para ello hemos seleccionado una serie de valores, derivados de los vectores de características, y hemos realizado una clasificación similar a la del apartado anterior, para seleccionar el candidato idóneo. Hemos partido de la norma de los vectores y verificado su éxito en la clasificación, luego hemos hecho lo mismo utilizando los componentes principales derivados de cada vector, y finalmente hemos probado los escalares obtenidos mediante el Análisis Discriminante de Fisher. Con el valor que nos dá el mejor resultado en clasificación hemos realizado una nueva validación, esta vez no con un número específico de clases, para lo cual se hizo necesario una nueva verdad del terreno basada en la apreciación de los

expertos. Para obtener la nueva verdad del terreno hemos hecho una selección aleatoria de pares de imágenes para que sean ordenadas por los expertos de acuerdo a su nivel de abollado. Esto lo hacemos ya que una ordenación continua de todas las imágenes es imposible para el experto, pero resulta un procedimiento factible si seleccionamos pares de imágenes. Las ambigüedades que puedan surgir en los casos frontera, se resuelven de manera estadística. Finalmente hemos correlacionado el valor del escalar seleccionado en el proceso de clasificación, con la nueva verdad del terreno para ver la coincidencia del índice con la apreciación obtenida por los expertos.

- Conclusiones. En esta sección se dan las conclusiones generales a partir de los resultados obtenidos para el desarrollo de este capítulo y se plantean posibles trabajos futuros.

5.3. Adquisición de las imágenes

Las imágenes han sido capturadas mediante un escaner plano convencional con una resolución óptica de 600 dpi, con el cual se generaron imágenes de dimensiones 850 x 1170 píxeles. El tamaño original de las hojas escaneadas es el estándar A4: 24 cm x 29.7 cm. Las imágenes adquiridas son de color blanco con muy bajo contraste, lo que hace muy difícil la apreciación visual del abollado. Para mejorar el contraste se ha realizado el siguiente preprocesado: la selección de los niveles de gris superiores del histograma acumulativo de la imagen hasta cubrir un 90% y la renormalización del rango de intensidades. A estas imágenes se les aplica un filtro gaussiano para reducir el ruido impulsivo que es amplificado en el proceso de mejora de contraste. En las imágenes preprocesadas se observa la presencia de ciertas texturas que no guardan un patrón regular.

La figura 5.1 muestra una selección de imágenes de papel reciclado, en las que se ha aplicado el recorte y renormalización del histograma para realzar el contraste. En las imágenes se observa la presencia de ondulaciones y protuberancias con variación en la distribución espacial, la forma y el tamaño.

5.4. Etiquetado manual

Durante la inspección previa de las imágenes, se ha podido observar la gran variedad en la forma y tamaño de las ondulaciones y protuberancias que caracterizan al abollado, al igual que la variabilidad en su distribución espacial, tanto dentro de una misma imagen, como entre diferentes imágenes. Lo que diferencia este problema de otro problema de discriminación de texturas es que no existen clases de texturas diferenciadas a priori. Se trata de un defecto que produce una gradación de texturas cuyo orden no está establecido. Los primeros intentos de establecer una verdad del terreno los realizamos pidiendo a los expertos que realizaran el etiquetado en diez niveles de intensidad del abollado. Así, pudimos comprobar que resulta imposible para un observador humano realizar una clasificación estable en muchas clases graduadas de menor a mayor abollado. Por tanto hemos optado por establecer tres niveles de abollado: Alto, Medio y Bajo. Una vez definido el número de clases o niveles, se ha seleccionado un grupo de expertos para realizar el etiquetado manual, consistente en la asignación de cada imagen a alguna de las tres clases definidas, de acuerdo a su apreciación de la severidad del defecto. Se ha desarrollado una sencilla aplicación que presenta cada imagen al experto para que sea asignada a su clase correspondiente. Previamente se permite al usuario experto realizar una exploración de las imágenes para que se familiarice con ellas y pueda apreciar las variaciones y niveles antes del etiquetado. Cada experto realizó el etiquetado de las imágenes en tres ocasiones, con intervalos de descanso de un día, con el fin de evitar la fatiga. En el experimento participaron cinco expertos, con lo cual cada imagen ha sido etiquetada 15 veces. Con los resultados obtenidos se hizo una comparación para determinar el acuerdo de cada usuario consigo mismo en cada etiquetado, así como el acuerdo entre los etiquetados de los diferentes usuarios.

Las tablas 5.1, 5.2 y 5.3 muestran las matrices de concordancia/confusión sobre los etiquetados realizados por los observadores. Las tres etiquetas utilizadas $c1$, $c2$ y $c3$ corresponden a los tipos de abollado: Alto, Medio o Bajo, respectivamente. Cada entrada $C_{(O1,c1),(O2,c2)}$ muestra el porcentaje, sobre el número total de imágenes, asignado a la clase $c1$ por el observador $O1$ que han sido asignadas a la clase $c2$ por el observador $O2$, cuando las clases coinciden tenemos la concordancia entre los observadores, y cuando no coinciden tenemos la confusión. Las submatrices que están en la diagonal corresponden a la concordancia de un observador consigo mismo. En cada una de las subma-

| | | OBS1 | | | OBS2 | | | OBS3 | | | OBS4 | | | OBS5 | | |
|------|----|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|
| | | C1 | C2 | C3 | C1 | C2 | C3 | C1 | C2 | C3 | C1 | C2 | C3 | C1 | C2 | C3 |
| OBS1 | C1 | 0,64 | 0,00 | 0,00 | 0,23 | 0,36 | 0,05 | 0,27 | 0,36 | 0,01 | 0,50 | 0,13 | 0,00 | 0,44 | 0,19 | 0,01 |
| | C2 | 0,00 | 0,30 | 0,00 | 0,00 | 0,11 | 0,19 | 0,01 | 0,20 | 0,09 | 0,04 | 0,18 | 0,07 | 0,03 | 0,18 | 0,09 |
| | C3 | 0,00 | 0,00 | 0,06 | 0,00 | 0,00 | 0,06 | 0,00 | 0,01 | 0,05 | 0,00 | 0,01 | 0,05 | 0,00 | 0,00 | 0,06 |
| OBS2 | C1 | 0,23 | 0,00 | 0,00 | 0,23 | 0,00 | 0,00 | 0,17 | 0,06 | 0,00 | 0,23 | 0,00 | 0,00 | 0,23 | 0,00 | 0,00 |
| | C2 | 0,36 | 0,11 | 0,00 | 0,00 | 0,47 | 0,00 | 0,10 | 0,35 | 0,02 | 0,29 | 0,18 | 0,01 | 0,22 | 0,25 | 0,00 |
| | C3 | 0,05 | 0,19 | 0,06 | 0,00 | 0,00 | 0,30 | 0,01 | 0,15 | 0,13 | 0,03 | 0,14 | 0,12 | 0,02 | 0,12 | 0,16 |
| OBS3 | C1 | 0,27 | 0,01 | 0,00 | 0,17 | 0,10 | 0,01 | 0,28 | 0,00 | 0,00 | 0,25 | 0,03 | 0,00 | 0,23 | 0,05 | 0,00 |
| | C2 | 0,36 | 0,20 | 0,01 | 0,06 | 0,35 | 0,15 | 0,00 | 0,57 | 0,00 | 0,30 | 0,23 | 0,04 | 0,23 | 0,28 | 0,05 |
| | C3 | 0,01 | 0,09 | 0,05 | 0,00 | 0,02 | 0,13 | 0,00 | 0,00 | 0,15 | 0,00 | 0,06 | 0,09 | 0,00 | 0,05 | 0,10 |
| OBS4 | C1 | 0,50 | 0,04 | 0,00 | 0,23 | 0,29 | 0,03 | 0,25 | 0,30 | 0,00 | 0,55 | 0,00 | 0,00 | 0,45 | 0,10 | 0,00 |
| | C2 | 0,13 | 0,18 | 0,01 | 0,00 | 0,18 | 0,14 | 0,03 | 0,23 | 0,06 | 0,00 | 0,32 | 0,00 | 0,02 | 0,25 | 0,05 |
| | C3 | 0,00 | 0,07 | 0,05 | 0,00 | 0,01 | 0,12 | 0,00 | 0,04 | 0,09 | 0,00 | 0,00 | 0,13 | 0,00 | 0,02 | 0,11 |
| OBS5 | C1 | 0,44 | 0,03 | 0,00 | 0,23 | 0,22 | 0,02 | 0,23 | 0,23 | 0,00 | 0,45 | 0,02 | 0,00 | 0,47 | 0,00 | 0,00 |
| | C2 | 0,19 | 0,18 | 0,00 | 0,00 | 0,25 | 0,12 | 0,05 | 0,28 | 0,05 | 0,10 | 0,25 | 0,02 | 0,00 | 0,37 | 0,00 |
| | C3 | 0,01 | 0,09 | 0,06 | 0,00 | 0,00 | 0,16 | 0,00 | 0,05 | 0,10 | 0,00 | 0,05 | 0,11 | 0,00 | 0,00 | 0,16 |

Cuadro 5.1: Comparación de la primera clasificación realizada por diferentes observadores, mediante múltiples matrices de confusión.

trices 3x3 que relacionan a un observador con otro observador la suma de sus elementos es 1. La tabla 5.1 muestra la concordancia/confusión en el primer intento de etiquetado. La tabla 5.2 muestra los resultados para el segundo intento. La tabla 5.3 compara el primero con el segundo intento, donde las filas corresponden al primer intento y las columnas al segundo intento de etiquetado.

Para cada observador, la submatriz 3x3 correspondiente, que está en la diagonal de las tablas 5.1 y 5.2, es una matriz diagonal, que nos da la estimación de la probabilidad a priori de cada clase según ese observador. Las submatrices fuera de la diagonal en estas tablas nos dan las concordancias y disensiones entre observadores. La traza (suma de la diagonal) de cada submatriz nos da el grado de concordancia de cada par de personas. La suma del resto de los elementos en la submatriz (1-traza) nos da el grado de disensión entre los observadores. La media de las concordancias nos da una medida del porcentaje de aciertos que podemos considerar como un resultado aceptable al construir un clasificador automático. En la tabla 5.3 las submatrices que están en la diagonal muestran la concordancia de un observador consigo mismo en dos intentos diferentes de etiquetado.

| | | OBS1 | | | OBS2 | | | OBS3 | | | OBS4 | | | OBS5 | | |
|------|----|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|
| | | C1 | C2 | C3 | C1 | C2 | C3 | C1 | C2 | C3 | C1 | C2 | C3 | C1 | C2 | C3 |
| OBS1 | C1 | 0,74 | 0,00 | 0,00 | 0,20 | 0,47 | 0,06 | 0,39 | 0,30 | 0,05 | 0,65 | 0,08 | 0,00 | 0,49 | 0,21 | 0,03 |
| | C2 | 0,00 | 0,23 | 0,00 | 0,00 | 0,10 | 0,12 | 0,01 | 0,13 | 0,08 | 0,08 | 0,11 | 0,04 | 0,05 | 0,09 | 0,09 |
| | C3 | 0,00 | 0,00 | 0,03 | 0,00 | 0,00 | 0,03 | 0,00 | 0,00 | 0,03 | 0,00 | 0,00 | 0,03 | 0,00 | 0,00 | 0,03 |
| OBS2 | C1 | 0,20 | 0,00 | 0,00 | 0,20 | 0,00 | 0,00 | 0,19 | 0,01 | 0,00 | 0,20 | 0,00 | 0,00 | 0,20 | 0,01 | 0,00 |
| | C2 | 0,47 | 0,10 | 0,00 | 0,00 | 0,58 | 0,00 | 0,20 | 0,34 | 0,04 | 0,49 | 0,08 | 0,01 | 0,33 | 0,21 | 0,04 |
| | C3 | 0,06 | 0,12 | 0,03 | 0,00 | 0,00 | 0,22 | 0,01 | 0,08 | 0,13 | 0,04 | 0,11 | 0,06 | 0,01 | 0,09 | 0,11 |
| OBS3 | C1 | 0,39 | 0,01 | 0,00 | 0,19 | 0,20 | 0,01 | 0,40 | 0,00 | 0,00 | 0,40 | 0,00 | 0,00 | 0,39 | 0,01 | 0,00 |
| | C2 | 0,30 | 0,13 | 0,00 | 0,01 | 0,34 | 0,08 | 0,00 | 0,44 | 0,00 | 0,32 | 0,10 | 0,02 | 0,15 | 0,23 | 0,05 |
| | C3 | 0,05 | 0,08 | 0,03 | 0,00 | 0,04 | 0,13 | 0,00 | 0,00 | 0,17 | 0,02 | 0,09 | 0,05 | 0,00 | 0,06 | 0,10 |
| OBS4 | C1 | 0,65 | 0,08 | 0,00 | 0,20 | 0,49 | 0,04 | 0,40 | 0,32 | 0,02 | 0,73 | 0,00 | 0,00 | 0,54 | 0,19 | 0,00 |
| | C2 | 0,08 | 0,11 | 0,00 | 0,00 | 0,08 | 0,11 | 0,00 | 0,10 | 0,09 | 0,00 | 0,19 | 0,00 | 0,00 | 0,11 | 0,08 |
| | C3 | 0,00 | 0,04 | 0,03 | 0,00 | 0,01 | 0,06 | 0,00 | 0,02 | 0,05 | 0,00 | 0,00 | 0,07 | 0,00 | 0,01 | 0,06 |
| OBS5 | C1 | 0,49 | 0,05 | 0,00 | 0,20 | 0,33 | 0,01 | 0,39 | 0,15 | 0,00 | 0,54 | 0,00 | 0,00 | 0,54 | 0,00 | 0,00 |
| | C2 | 0,21 | 0,09 | 0,00 | 0,01 | 0,21 | 0,09 | 0,01 | 0,23 | 0,06 | 0,19 | 0,11 | 0,01 | 0,00 | 0,31 | 0,00 |
| | C3 | 0,03 | 0,09 | 0,03 | 0,00 | 0,04 | 0,11 | 0,00 | 0,05 | 0,10 | 0,00 | 0,08 | 0,06 | 0,00 | 0,00 | 0,15 |

Cuadro 5.2: Comparación de la segunda clasificación realizada por diferentes observadores, mediante múltiples matrices de confusión.

| | | OBS1 | | | OBS2 | | | OBS3 | | | OBS4 | | | OBS5 | | |
|------|----|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|
| | | C1 | C2 | C3 | C1 | C2 | C3 | C1 | C2 | C3 | C1 | C2 | C3 | C1 | C2 | C3 |
| OBS1 | C1 | 0,61 | 0,03 | 0,00 | 0,20 | 0,42 | 0,02 | 0,39 | 0,25 | 0,01 | 0,61 | 0,03 | 0,00 | 0,49 | 0,15 | 0,00 |
| | C2 | 0,12 | 0,17 | 0,00 | 0,00 | 0,16 | 0,14 | 0,01 | 0,18 | 0,10 | 0,12 | 0,15 | 0,02 | 0,05 | 0,16 | 0,09 |
| | C3 | 0,00 | 0,03 | 0,03 | 0,00 | 0,00 | 0,06 | 0,00 | 0,01 | 0,06 | 0,00 | 0,02 | 0,05 | 0,00 | 0,00 | 0,06 |
| OBS2 | C1 | 0,23 | 0,00 | 0,00 | 0,16 | 0,07 | 0,00 | 0,22 | 0,01 | 0,00 | 0,23 | 0,00 | 0,00 | 0,23 | 0,00 | 0,00 |
| | C2 | 0,40 | 0,07 | 0,00 | 0,04 | 0,41 | 0,02 | 0,17 | 0,28 | 0,02 | 0,43 | 0,03 | 0,00 | 0,30 | 0,17 | 0,00 |
| | C3 | 0,11 | 0,16 | 0,03 | 0,00 | 0,10 | 0,20 | 0,01 | 0,14 | 0,15 | 0,07 | 0,16 | 0,07 | 0,01 | 0,14 | 0,15 |
| OBS3 | C1 | 0,28 | 0,01 | 0,00 | 0,16 | 0,12 | 0,00 | 0,23 | 0,05 | 0,00 | 0,28 | 0,00 | 0,00 | 0,24 | 0,04 | 0,00 |
| | C2 | 0,43 | 0,14 | 0,00 | 0,04 | 0,43 | 0,10 | 0,17 | 0,35 | 0,05 | 0,44 | 0,11 | 0,02 | 0,30 | 0,20 | 0,06 |
| | C3 | 0,04 | 0,08 | 0,03 | 0,00 | 0,03 | 0,12 | 0,00 | 0,04 | 0,11 | 0,02 | 0,08 | 0,05 | 0,00 | 0,06 | 0,09 |
| OBS4 | C1 | 0,51 | 0,03 | 0,00 | 0,20 | 0,33 | 0,02 | 0,38 | 0,16 | 0,00 | 0,54 | 0,01 | 0,00 | 0,51 | 0,04 | 0,00 |
| | C2 | 0,21 | 0,11 | 0,00 | 0,00 | 0,21 | 0,10 | 0,01 | 0,24 | 0,07 | 0,18 | 0,13 | 0,01 | 0,03 | 0,23 | 0,05 |
| | C3 | 0,02 | 0,08 | 0,03 | 0,00 | 0,03 | 0,10 | 0,00 | 0,04 | 0,09 | 0,02 | 0,06 | 0,06 | 0,00 | 0,03 | 0,10 |
| OBS5 | C1 | 0,43 | 0,03 | 0,00 | 0,20 | 0,26 | 0,01 | 0,37 | 0,10 | 0,00 | 0,46 | 0,01 | 0,00 | 0,45 | 0,01 | 0,00 |
| | C2 | 0,27 | 0,10 | 0,00 | 0,01 | 0,28 | 0,09 | 0,03 | 0,29 | 0,05 | 0,26 | 0,10 | 0,01 | 0,09 | 0,27 | 0,02 |
| | C3 | 0,03 | 0,09 | 0,03 | 0,00 | 0,04 | 0,12 | 0,00 | 0,05 | 0,11 | 0,02 | 0,08 | 0,06 | 0,00 | 0,03 | 0,13 |

Cuadro 5.3: Comparación de la primera y segunda clasificación realizada por los diferentes observadores, mediante múltiples matrices de confusión.

| | OBS1 | OBS2 | OBS3 | OBS4 | OBS5 |
|------|------|------|------|------|------|
| OBS1 | 1,00 | 0,40 | 0,52 | 0,73 | 0,68 |
| OBS2 | 0,40 | 1,00 | 0,65 | 0,53 | 0,63 |
| OBS3 | 0,52 | 0,65 | 1,00 | 0,56 | 0,61 |
| OBS4 | 0,73 | 0,53 | 0,56 | 1,00 | 0,80 |
| OBS5 | 0,68 | 0,63 | 0,61 | 0,80 | 1,00 |

Cuadro 5.4: La tabla muestra la concordancia entre cada observador en la primera evaluación.

| | OBS1 | OBS2 | OBS3 | OBS4 | OBS5 |
|------|------|------|------|------|------|
| OBS1 | 1,00 | 0,34 | 0,55 | 0,80 | 0,62 |
| OBS2 | 0,34 | 1,00 | 0,66 | 0,35 | 0,52 |
| OBS3 | 0,55 | 0,66 | 1,00 | 0,55 | 0,72 |
| OBS4 | 0,80 | 0,35 | 0,55 | 1,00 | 0,71 |
| OBS5 | 0,62 | 0,52 | 0,72 | 0,71 | 1,00 |

Cuadro 5.5: La tabla muestra la concordancia entre cada observador en la segunda evaluación.

Las tablas 5.4, 5.5 y 5.6 resumen las anteriores, mostrando la concordancia entre dos observadores, que se calcula como la suma de la diagonal de cada submatriz 3x3 de comparación entre personas (tablas 5.1, 5.2 y 5.3). Estas tablas dan una impresión más clara del grado de acuerdo entre los observadores.

Para las tablas 5.4, 5.5, 5.6, que comparan la primera y segunda observaciones de cada persona, así como la primera con la segunda, respectivamente, las concordancias promedio entre observadores son de: 0.69 , 0.67 y 0.63, respectivamente. Estos bajos porcentajes dejan ver la complejidad de la realización de una clasificación manual. El grado de concordancia entre los etiquetadores manuales puede ser asumido como una predicción de la

| | OBS1 | OBS2 | OBS3 | OBS4 | OBS5 |
|------|------|------|------|------|------|
| OBS1 | 0,81 | 0,42 | 0,62 | 0,81 | 0,71 |
| OBS2 | 0,34 | 0,77 | 0,65 | 0,33 | 0,55 |
| OBS3 | 0,45 | 0,70 | 0,69 | 0,43 | 0,53 |
| OBS4 | 0,66 | 0,51 | 0,71 | 0,73 | 0,84 |
| OBS5 | 0,57 | 0,59 | 0,77 | 0,62 | 0,85 |

Cuadro 5.6: La tabla muestra la concordancia entre cada observador en la primera y la segunda evaluación .

dificultad en la construcción de un clasificador automático.

Puesto que se ha observado una gran confusión de las clases de abollado Medio y Bajo, consideramos su fusión en una nueva clase. Dicha fusión consiste en asumir que las etiquetas Medio y Bajo puestas por los observadores pasan a ser de esta nueva clase. Esto es, no realizamos un nuevo etiquetado. Una vez realizada la fusión, los valores promedio de concordancias en cada tabla aumentan, obteniéndose 0.79, 0.75 y 0.73. Por tanto, asumimos 0.75 como valor de referencia para determinar el éxito en los algoritmos de clasificación automática propuestos, donde la verdad del terreno es el etiquetado realizado por los usuarios. Si obtenemos tal porcentaje de éxito en la clasificación, se puede decir que la clasificación automática tiene un rendimiento comparable a la realizada por los expertos humanos. Con estos datos como referencia, se ha procedido a implementar algunos métodos de clasificación automática. La verdad del terreno consiste en la asignación por mayoría de votos de cada imagen a una de las dos clases: abollado Alto, abollado Bajo.

5.5. Definición de características

En esta sección se detallan las características que se extraen de cada imagen y que van a ser usadas por los clasificadores automáticos. Hemos utilizado características de textura basadas en los Bancos del Filtros de Gabor (BFG) y en la Transformada Discreta Wavelet (TWD). Conviene precisar qué diferencia nuestro trabajo de otros encontrados en la literatura. Muchos

trabajos están dedicados a la segmentación de regiones en la imagen usando la textura para distinguir regiones. En estos casos las características de texturas se calculan para cada píxel y su vecindario, son características locales. En nuestro caso tratamos de distinguir imágenes completas. Por tanto el vector de características de textura es único para toda la imagen.

5.5.1. Características basadas en Bancos del Filtros de Gabor

Para calcular las características basadas en los BFG, hemos utilizados las versiones par (real) e impar (imaginaria) de cada filtro. Si $I(x, y)$ denota la imagen que vamos a analizar, y $G_r(x, y)$, $G_i(x, y)$ son las máscaras de convolución correspondientes a las versiones par e impar de cada filtro, previamente deducidos de la función de Gabor $g_\Theta(x, y)$, de parámetros: $\Theta = (F, \theta, \sigma_x \sigma_y, B_F, B_\theta)$ presentados en el Capítulo 3, calculamos la energía de Gabor de la respuesta del filtro mediante la expresión:

$$E_\Theta(x, y) = \sqrt{[G_r \bullet I(x, y)]^2 + [G_i \bullet I(x, y)]^2}, \quad (5.1)$$

donde \bullet denota la operación de convolución². Dado un Banco de Filtros de Gabor con m orientaciones y n frecuencias espaciales, hemos definido el siguiente vector de características formado por las medias y varianzas de las respuestas a cada uno de los filtros.:

$$V = [(A_{f,\theta}, S_{f,\theta}) : f = 1, \dots, n; \theta = 1, \dots, m], \quad (5.2)$$

donde

$$A_{f,\theta} = \frac{1}{MN} \sum_{x=1}^M \sum_{y=1}^N E_{f,\theta}(x, y), \quad (5.3)$$

$$S_{f,\theta} = \frac{1}{MN} \sum_{x=1}^M \sum_{y=1}^N (A_{f,\theta} - E_{f,\theta}(x, y))^2. \quad (5.4)$$

En las anteriores expresiones M y N denotan el tamaño de la imagen. Si consideramos que un elemento de textura es una función espacial periódica

²En la práctica, las convoluciones son calculadas mediante el producto de Hadamard de la función de transferencia del filtro por la transformada de la imagen en el dominio de Fourier

con determinada orientación y frecuencia, los filtros sintonizados con una orientación y frecuencia más cercanos a los propios de cada elemento de textura obtendrán una mayor respuesta de las imágenes en las que predomina ese elemento de textura. Con el vector de características definido en (5.2) se van a obtener representaciones discriminantes siempre y cuando la orientación y frecuencia de los diferentes elementos de textura caigan dentro de los campos receptivos de distintos filtros de Gabor. En el apéndice A.1 hemos incluido algunos ejemplos que ilustran el funcionamiento de los BFG como herramienta para la caracterización y discriminación de texturas.

5.5.2. Características basadas en coeficientes de la TWD

El análisis de texturas a través de características basadas en los coeficientes de la Transformada Wavelet Discreta ha sido propuesto en múltiples aplicaciones desde que Mallat [101] propuso su algoritmo de descomposición piramidal, y sigue vigente en la literatura especializada [46]. Como método de caracterización de texturas, nos hemos basado en los coeficientes de la TWD que contienen información del gradiente de la imagen. Hemos definido el vector global de características de textura de la imagen como las medias y las desviaciones estándar de cada una de las subimágenes de los detalles horizontal, diagonal y vertical, en un rango de niveles de descomposición convenientemente seleccionado, (i.e. $f_{HH}^i(x, y)$, $f_{LH}^i(x, y)$, $f_{HL}^i(x, y)$, con $i = n, ..m$). Formalmente, nuestro vector de características viene dado como:

$$V = [(A_{i,d}, S_{i,d}) : i = m, \dots, n; d \in \{HH, LH, HL\}], \quad (5.5)$$

donde m y n son los niveles de resolución inferior y superior considerados, respectivamente, y

$$A_{i,d} = \frac{1}{M_{i,d}N_{i,d}} \sum_{x=1}^{M_{i,d}} \sum_{y=1}^{N_{i,d}} f_d^i(x, y), \quad (5.6)$$

$$S_{i,d} = \frac{1}{M_{i,d}N_{i,d}} \sum_{x=1}^{M_{i,d}} \sum_{y=1}^{N_{i,d}} (A_{i,d} - f_d^i(x, y))^2, \quad (5.7)$$

donde $M_{i,d}$ y $N_{i,d}$ son las dimensiones de la matriz de coeficientes $f_d^i(x, y)$ de los detalles de orientación d , en el nivel de descomposición i , definida en la sección 4.3

Debido a la naturaleza recursiva de la TWD, y a la manera en que se realiza el submuestreo, cada nivel de descomposición corresponde a una octava en el espacio de frecuencias de la imagen. Los detalles en el nivel de descomposición i , muestran variaciones y regularidades que suceden en periodos del orden 2^i . El valor de la media en cada nivel de detalle nos da una descripción global de la existencia de variaciones en cada escala correspondiente, mientras la desviación estándar nos da información sobre la magnitud de la variabilidad de los cambios. En la mayoría de los casos los niveles de detalle más bajos no son de interés ya que éstos son afectados por cambios locales y ruido impulsivo. Hemos incluido algunos ejemplos en el apéndice A.2 que ilustran la caracterización de texturas usando los coeficientes de la TWD y la capacidad discriminatoria del vector de características definido antes.

5.6. Clasificación automática

Inicialmente hemos utilizado dos métodos para la contrucción automática de clasificadores [40] de las imágenes de papel reciclado. Primero, el clasificador del vecino más cercano k -NN (*K-nearest neighbor*), segundo, una red neuronal artificial de tipo Perceptrón Multicapa (MLP - *Multilayer Perceptron*) entrenada con el algoritmo de retropropagación (*backpropagation*). Posteriormente hemos extendido los resultados aplicando una batería de clasificadores implementados en la herramienta de software libre conocida como Weka³ [164]. Los algoritmos de clasificación seleccionados son los siguientes: Red Bayesiana (RB), Perceptrón Multicapa (MLP), Función de Base Radial (RBF), Regresión Logística (Log), Clasificador Multiclase para SVM (MCC), Optimizador Secuencias Minimales para SVM (SMO) , Clasificación Vía Regresión (CVR), LogitBost (LB), Árboles de decisión logistica (LMT).

En los experimentos de entrenamiento hemos utilizado una estrategia de validación cruzada en dos particiones (*two-fold cross validation*) sobre la muestra de los vectores de características calculados a partir de las imágenes de papel. Los experimentos realizados se han basado en los dos tipos de vectores globales de características de texturas discutidos previamente. El objetivo de estos experimentos de clasificación es determinar cual es el vector de características más adecuado para contruir un índice escalar de abollado.

³<http://www.cs.waikato.ac.nz/ml/weka/>

5.6.1. Resultados con las características extraídas mediante BFG

Se han usado dos Bancos de Filtros de Gabor distintos, cuya principal diferencia radica en que en el primero hay un solapamiento en las elipses formadas por los puntos del plano de Fourier correspondientes a la respuesta media del filtro, y en el segundo los parámetros han sido ajustados para evitar este solapamiento, y así, la redundancia de información. Los demás parámetros utilizados son los habitualmente citados en la literatura. El primer BFG está compuesto por 20 filtros, que han sido distribuidos sobre el plano con ángulos de orientación $\theta = 30, 60, 90, 120, 150$ grados y sintonizados en las frecuencias $F = 1/8, 1/16, 1/32$ y $1/64$ ciclos por pixel. Los valores para el ancho de banda angular y el ancho de banda radial seleccionados son de 35 grados y 1.4 octavas. En el segundo banco de filtros se han mantenido las orientaciones y anchos de banda angular y radial pero los centros de frecuencias de cada elipse han sido modificados para que las elipses no se solapen. La figura 3.3 ilustra los dos tipos de cubrimiento.

En relación a la estructura de los algoritmos de clasificación, para el algoritmo k-NN se realizaron pruebas con distintos valores de vecindario y se obtuvieron los mejores resultados con el valor $k=7$. La arquitectura del MLP consta de tres capas: la capa de entrada con el número de neuronas que corresponden al tamaño de cada vector de características, experimentalmente hemos encontrado que los mejores resultados se obtenían con 10 neuronas en la capa oculta, y la capa de salida con dos neuronas, una para cada clase de abollado. La función de activación es la tangente hiperbólica para todas las capas. La tabla 5.7 muestra los resultados de clasificación obtenidos sobre el conjunto de prueba de imágenes de papel reciclado, con los vectores de características construidos a partir de los BFG. Los mejores resultados han sido obtenidos con el MLP, aunque el clasificador 7-NN da mejores resultados sobre la clasificación de la primera clase, que es la más frecuente. La tasa de éxito es comparable con el nivel de concordancia entre los observadores humanos a la que se hace referencia en la sección previa.

Los resultados obtenidos con los clasificadores implementados en Weka se pueden ver en la tabla 5.8. Los mejores resultados se logran con el Algoritmo de Regresión Logística, y los más bajos mediante la Red Bayesiana. En la última columna podemos ver la media de los resultados obtenidos por el conjunto de clasificadores. Los valores concuerdan con los obtenidos por el MLP implementado inicialmente, lo que nos da indicación de la robustez de

| | Campos receptivos con solapamiento | | Campos receptivos sin solapamiento | |
|----|---------------------------------------|--------|---------------------------------------|--------|
| | 7-NN | MLP | 7-NN | MLP |
| C1 | 80 % | 78,5 % | 87,7 % | 80 % |
| C2 | 67,7 % | 76,9 % | 64,6 % | 75,4 % |
| | 73,8 % | 77,7 % | 76,2 % | 77,7 % |

Cuadro 5.7: Resultados iniciales de la clasificación con características basadas en BFG con y sin solapamiento en los campos receptivos.

| | RB | MLP | RBF | Log | SMO | MCC | CVR | LB | LMT | Prom |
|-------------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
| Sin solapam | 71.85 | 76.29 | 74.07 | 81.85 | 79.62 | 81.85 | 72.59 | 76.66 | 81.48 | 77,36 |
| Con Solap | 73.7 | 76.66 | 69.25 | 82.22 | 80.74 | 82.22 | 77.4 | 76.66 | 79.62 | 77,61 |

Cuadro 5.8: Resultados de la clasificación de los vectores de características basados en BFG utilizando Weka.

los vectores de características definidos frente a varios algoritmos de clasificación.

5.6.2. Resultados con las características extraídas mediante coeficientes wavelet

Hemos explorado las siguientes combinaciones de niveles de descomposición y tipos de función wavelet: rangos de niveles de descomposición [1,3], [2,4], [3,5], [4,6], [5,7] y [6,8], y funciones wavelet de Daubechies desde el orden 1 (conocida como wavelet de Haar), hasta el orden 8.

Al igual que en el caso de las características basadas en BFG, hemos utilizado el algoritmo k -NN y la red neuronal MPL. En las tabla 5.9 y 5.10 se puede ver el éxito alcanzado mediante el clasificador k -NN para cada fun-

| | K=1 | K=3 | K=5 | K=7 | K=1 | K=3 | K=5 | K=7 |
|-----|-------|-------|-------|--------------|-------|-------|-------|-------|
| | db1 | | | | db2 | | | |
| 123 | 57,69 | 60,00 | 62,31 | 66,15 | 52,31 | 60,00 | 65,38 | 60,77 |
| 234 | 67,69 | 68,46 | 70,00 | 69,23 | 72,31 | 76,15 | 76,92 | 72,31 |
| 345 | 76,15 | 76,92 | 76,15 | 77,69 | 74,62 | 74,62 | 75,38 | 76,92 |
| 456 | 73,85 | 77,69 | 83,08 | 83,85 | 73,08 | 73,85 | 73,08 | 70,00 |
| 567 | 76,92 | 77,69 | 78,46 | 80,00 | 61,54 | 54,62 | 60,77 | 56,15 |
| 789 | 66,92 | 70,77 | 69,23 | 70,00 | 55,38 | 53,08 | 56,15 | 55,38 |
| | db3 | | | | db4 | | | |
| 123 | 55,38 | 52,31 | 53,85 | 56,15 | 65,38 | 68,46 | 66,92 | 69,23 |
| 234 | 62,31 | 66,15 | 69,23 | 67,69 | 70,77 | 73,08 | 79,23 | 77,69 |
| 345 | 72,31 | 74,62 | 74,62 | 73,08 | 74,62 | 76,92 | 74,62 | 73,85 |
| 456 | 63,08 | 66,15 | 60,77 | 62,31 | 62,31 | 69,23 | 66,15 | 70,00 |
| 567 | 57,69 | 55,38 | 52,31 | 50,00 | 51,54 | 53,85 | 57,69 | 56,15 |
| 789 | 60,00 | 53,85 | 57,69 | 54,62 | 56,92 | 46,15 | 50,00 | 51,54 |

Cuadro 5.9: Exito en la clasificación mediante el algoritmo k-NN con diferentes niveles de descomposición y wavelets madre db1 a db4.

| | K=1 | K=3 | K=5 | K=7 | K=1 | K=3 | K=5 | K=7 |
|-----|-------|-------|-------|-------|-------|-------|-------|-------|
| | db5 | | | | db6 | | | |
| 123 | 63,08 | 64,62 | 63,85 | 62,31 | 61,54 | 60,77 | 58,46 | 64,62 |
| 234 | 73,08 | 73,85 | 76,92 | 70,77 | 73,85 | 75,38 | 76,15 | 74,62 |
| 345 | 72,31 | 77,69 | 73,85 | 76,92 | 67,69 | 69,23 | 70,77 | 71,54 |
| 456 | 71,54 | 70,77 | 69,23 | 63,85 | 67,69 | 68,46 | 71,54 | 70,77 |
| 567 | 56,15 | 60,00 | 58,46 | 57,69 | 60,77 | 56,92 | 60,00 | 65,38 |
| 789 | 54,62 | 49,23 | 56,92 | 52,31 | 56,15 | 54,62 | 53,85 | 50,00 |
| | db7 | | | | db8 | | | |
| 123 | 52,31 | 54,62 | 55,38 | 50,77 | 53,08 | 52,31 | 52,31 | 50,00 |
| 234 | 66,15 | 75,38 | 72,31 | 72,31 | 68,46 | 67,69 | 64,62 | 61,54 |
| 345 | 71,54 | 70,00 | 74,62 | 73,85 | 68,46 | 73,08 | 74,62 | 72,31 |
| 456 | 71,54 | 70,77 | 70,77 | 73,85 | 74,62 | 77,69 | 78,46 | 73,85 |
| 567 | 57,69 | 55,38 | 55,38 | 54,62 | 65,38 | 56,15 | 56,92 | 56,15 |
| 789 | 55,38 | 55,38 | 50,77 | 50,77 | 60,00 | 56,15 | 56,92 | 60,00 |

Cuadro 5.10: Exito en la clasificación mediante el algoritmo k-NN con diferentes niveles de descomposición y wavelets madre db5 a db8.

| | db1 | db2 | db3 | db4 | db5 | db6 | db7 | db8 |
|-----|-------|-------|-------|--------------|-------|-------|-------|-------|
| 1-3 | 78,46 | 69,23 | 53,85 | 42,31 | 66,15 | 63,85 | 60,77 | 62,31 |
| 2-4 | 76,15 | 76,15 | 73,85 | 71,54 | 70,77 | 70,77 | 64,62 | 65,38 |
| 3-5 | 79,23 | 78,46 | 78,46 | 71,54 | 82,31 | 76,15 | 73,85 | 74,62 |
| 4-6 | 80,00 | 46,92 | 80,77 | 85,38 | 82,31 | 78,46 | 76,15 | 83,08 |
| 5-7 | 77,69 | 52,31 | 51,54 | 44,62 | 43,85 | 76,15 | 56,92 | 76,15 |
| 6-8 | 50,00 | 50,00 | 50,00 | 47,69 | 50,00 | 46,92 | 44,62 | 50,00 |

Cuadro 5.11: Exito en la clasificación mediante la red neuronal MLP con diferentes niveles de descomposición y diferentes wavelets madre

ción wavelet madre y cada vector de características. Los mejores resultados fueron obtenidos con el vector construido desde los coeficientes del rango de niveles [4,6], con la madre wavelet de Haar (db1) y con el parámetro del vecino mas cercano $k=7$. Como era de esperar, los niveles bajos de descomposición no revelan las características espaciales del problema, siendo más representativos del ruido. Los niveles de descomposición más altos detectan características espaciales de mayor escala que las ocasionadas por el abollado en la hoja de papel. En la tabla 5.11 se muestran los resultados obtenidos mediante la clasificación con la red neuronal. En este caso los mejores resultados también corresponden a las características calculadas desde los niveles de descomposición [4,6], pero con la wavelet madre de Daubechies de orden 4 (db4). Estos resultados mejoran los obtenidos mediante el algoritmo k -NN, y también los obtenidos mediante las características basadas en los BFG. El éxito obtenido por ambos clasificadores supera notablemente el nivel de concordancia que habíamos establecido como deseable a partir de los resultados conseguidos entre los etiquetados manuales.

Los resultados obtenidos mediante los clasificadores implementados en Weka se muestran en la tabla 5.12. La última columna de la tabla tiene el promedio de los clasificadores y es muy superior a los resultados obtenidos con los vectores basados en BFG. Los mejores resultados por clasificador mejoran incluso la clasificación inicial descrita antes para los vectores basados en TWD.

| | RB | MLP | RBF | Log | SMO | MCC | CVR | LB | LMT | Prom |
|-------------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
| Wav:4-6.Db1 | 84.81 | 90.37 | 87.03 | 85.18 | 87.77 | 85.18 | 88.88 | 87.03 | 86.66 | 86,99 |
| Wav:4-6.Db4 | 85.18 | 82.59 | 83.33 | 82.59 | 85.92 | 82.59 | 88.14 | 88.51 | 85.55 | 84,93 |

Cuadro 5.12: Resultados de la clasificación de los vectores de características basados en los coeficientes de la TWD mediante Weka.

5.6.3. Conclusiones de la clasificación

1. Mediante la clasificación automática obtenemos resultados que son comparables, e incluso mejores que los obtenidos mediante la clasificación manual de los expertos.
2. Los vectores de características basados en los coeficientes de la TWD obtienen un mejor promedio de clasificación que los vectores basados en BFG, por tanto son los mejores candidatos para contruir un índice escalar para el abollado.

5.7. Establecimiento de un índice de abollado

Los experimentos de clasificación automática nos han permitido deducir cuales son los vectores de características más adecuados para el problema entre manos, que son los basados en la TWD, sin embargo mantenemos el paralelismo con los vectores basados en BFG. El objetivo final es definir un índice continuo de abollado de las imágenes de papel reciclado, de manera que puedan ser superadas las restricciones que conlleva la definición de un número limitado de clases.

Hemos probado diferentes funciones que transforman los vectores de características a un escalar. Para valorar estas transformaciones hemos efectuado dos procesos.

1. La repetición del experimento de clasificación realizado en la sección anterior sobre la nueva variable. Se mantiene el mismo etiquetado de las muestras utilizado previamente. Este proceso nos ha servido como filtro para seleccionar la mejor transformación a la que luego hemos validado con el siguiente proceso.

| | RB | MLP | RBF | Log | SMO | MCC | CVR | LB | LMT | Prom |
|-------------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
| Gabor sin. | 51,85 | 51,85 | 50 | 55,55 | 51,48 | 55,55 | 50 | 52,59 | 50 | 52,10 |
| Gabor con | 51,85 | 50,74 | 50,37 | 55,18 | 48,88 | 55,18 | 49,62 | 54,81 | 49,62 | 51,81 |
| Wav:4-6.Db1 | 62,22 | 60,37 | 63,33 | 60 | 59,62 | 60 | 61,85 | 65,18 | 60,74 | 61,48 |
| Wav:4-6.Db4 | 52,59 | 55,18 | 57,4 | 59,62 | 54,07 | 59,62 | 52,22 | 50,37 | 59,25 | 55,59 |

Cuadro 5.13: Clasificación mediante la norma de los vectores de características

- Validación mediante la concordancia de ordenación de pares de imágenes. Exige un nuevo etiquetado de las muestras, que es un etiquetado indirecto puesto que se ha comprobado la dificultad del etiquetado directo sobre un número de clases. El proceso se ha aplicado a la mejor transformación obtenida en la clasificación anterior.

Inicialmente probamos con la norma de cada vector como el escalar que buscamos. Los resultados no fueron satisfactorios como se puede apreciar en la tabla 5.13, donde el mejor valor promedio obtenido es de 61,48 para los vectores basados en los coeficientes wavelet.

La segunda aproximación probada ha sido utilizando el Análisis de Componentes Principales (ACP). Hemos calculado los componentes principales para los cuatro vectores de características (los dos basados en BFG y los dos mejores basados en TWD) y realizamos el experimento de clasificación para cada uno de los tres primeros componentes de cada vector por separado. Los resultados obtenidos se visualizan en la tabla 5.14. En este caso observamos una mejora en los resultados respecto a la clasificación por la norma, pero inferior a las clasificaciones obtenidas de los vectores completos en la sección anterior, salvo en un caso: la tercera componente del vector basado en los coeficientes wavelets de Daubichies de orden 4 ofrece un éxito promedio en la clasificación del 80,61 %. Este es un porcentaje que mejora nuestro valor de referencia y es comparable al éxito obtenido en la clasificación con el vector de características.

Finalmente hemos considerado el uso de otra técnica clásica de reducción de dimensionalidad: el Análisis Discriminante de Fisher. Este nos permite obtener una proyección lineal de los datos sobre un eje que busca la mayor distancia entre los centroides de cada clase, con la menor varianza intraclase.

| GABOR SIN | | | | | | | | | | |
|------------------|-------|-------|-------|-------|-------|-------|-------|-------|-------|--------------|
| CP1 | 49,62 | 52,22 | 49,25 | 56,29 | 49,62 | 56,29 | 50,37 | 56,29 | 50,37 | 52,26 |
| CP2 | 75,92 | 75,55 | 74,81 | 68,88 | 63,7 | 68,88 | 73,33 | 75,18 | 75,55 | 72,42 |
| CP3 | 68,51 | 68,14 | 70,37 | 72,59 | 70,74 | 72,59 | 71,48 | 65,92 | 72,96 | 70,37 |
| GABOR CON | | | | | | | | | | |
| CP1 | 49,62 | 51,11 | 48,88 | 55,18 | 50,74 | 55,18 | 50 | 54,81 | 50 | 51,72 |
| CP2 | 74,44 | 75,55 | 74,81 | 68,88 | 64,44 | 68,88 | 74,44 | 73,7 | 74,81 | 72,22 |
| CP3 | 69,25 | 69,25 | 66,29 | 72,22 | 71,11 | 72,22 | 71,11 | 69,25 | 72,22 | 70,32 |
| WAV-DB1 | | | | | | | | | | |
| CP1 | 49,62 | 45,55 | 50,37 | 48,51 | 46,29 | 48,51 | 50,37 | 42,59 | 48,51 | 47,81 |
| CP2 | 71,11 | 68,14 | 70,74 | 71,11 | 70,74 | 71,11 | 70,74 | 70,74 | 70,74 | 70,57 |
| CP3 | 64,81 | 69,25 | 67,77 | 67,77 | 67,03 | 67,77 | 67,77 | 66,29 | 67,77 | 67,36 |
| WAV-DB4 | | | | | | | | | | |
| CP1 | 49,25 | 55,18 | 53,7 | 58,14 | 51,85 | 58,14 | 51,11 | 48,51 | 51,48 | 53,04 |
| CP2 | 49,62 | 55,18 | 55,18 | 60,37 | 60,37 | 60,37 | 60,37 | 54,07 | 60,37 | 57,32 |
| CP3 | 78,88 | 81,48 | 81,11 | 81,11 | 81,11 | 81,11 | 81,85 | 76,66 | 82,22 | 80,61 |

Cuadro 5.14: Clasificación mediante los componentes principales calculados a cada

| | RB | MLP | RBF | Log | SMO | MCC | CVR | LB | LMT | Prom |
|-------------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
| Gabor sin. | 80,74 | 83,7 | 84,07 | 85,18 | 84,07 | 85,18 | 83,33 | 83,7 | 84,44 | 83,82 |
| Gabor con | 83,33 | 83,7 | 84,81 | 83,7 | 84,07 | 83,7 | 82,59 | 82,96 | 84,07 | 83,66 |
| Wav:4-6.Db1 | 90,37 | 90,37 | 90,37 | 90,37 | 90,37 | 90,37 | 90,37 | 90,74 | 90,74 | 90,45 |
| Wav:4-6.Db4 | 78,88 | 80,74 | 82,59 | 82,59 | 81,48 | 82,59 | 81,85 | 82,59 | 81,85 | 81,68 |

Cuadro 5.15: Resultados de la clasificación realizada sobre los valores obtenidos mediante la aplicación de la función discriminante de Fisher a los vectores de características basados en BFG y coeficientes wavelets.

La tabla 5.16 nos muestra los resultados de la clasificación sobre los valores obtenidos con la función discriminante de Fisher. Los resultados son muy superiores incluso a los obtenidos mediante la clasificación de los vectores de características de la sección anterior, alcanzando en el mejor de los casos hasta un 90% de éxito en la clasificación promedio. Por tanto, la función discriminante de Fisher es el mejor candidato para definir un índice que represente en forma única el nivel de abollado de cada imagen.

5.7.1. Validación por ordenación de pares

Dado que en el ejercicio de clasificación sólo comprobamos la capacidad de discriminación entre dos clases de abollado, se hace necesario el establecimiento de una nueva verdad del terreno. Sin embargo, ya hemos comprobado antes que para un experto humano es difícil clasificar en un número de clases mayor que dos, por lo tanto hemos diseñado otra técnica que permita validar el índice con una referencia basada en la observación de los expertos:

- Seleccionamos un conjunto aleatorio de pares de imágenes que fueron presentadas a cada experto humano,
- Para cada par:
 - Se presenta indicando al experto un orden, la imagen izquierda es menos abollada que la imagen derecha.
 - El experto indica si está de acuerdo o no con el orden mostrado. Las respuestas son seleccionadas dentro de las opciones de una es-

| CARACTERÍSTICAS | CORRESPONDENCIA (%) |
|---|---------------------|
| BFG sin solapamiento | 87.33 |
| BFG con solapamiento | 85 |
| Coefficientes wavelet nivel 4-6 con DB1 | 87.3 |
| Coefficientes wavelet nivel 4-6 con DB4 | 89.33 |

Cuadro 5.16: Correspondencia entre el índice de abollado basado en la función discriminante de Fisher y la apreciación de los expertos.

cala de Likert⁴ de tres puntos (de acuerdo, en desacuerdo, ninguno de los dos).

- Se calcula el orden inducido por los valores del discriminante de Fisher calculado sobre el vector de características. Menos valor del discriminante corresponde a mayor abollado. Esto se puede observar en las figuras 5.2 a la 5.5, correspondientes a los histogramas generados a partir de estos valores. En todos los casos observamos que la media en los histogramas correspondientes a las imágenes con abollado Alto, es menor que la media en los histogramas correspondientes a las imágenes con abollado Bajo.
 - La ambigüedad de la escala de Likert corresponde con una diferencia de los valores del discriminante de Fisher entre el par de imágenes menor o igual al 10% del rango entre todos los valores calculados para el conjunto total de muestras.
- Finalmente se calcula la correlación entre las respuestas obtenidas de los expertos y los valores deducidos del discriminante de Fisher.

La tabla 5.16 muestra los resultados obtenidos. Así se confirma que el discriminante de Fisher es adecuado para construir un índice que cuantifica el nivel de abollado en las imágenes de papel reciclado.

⁴La escala de Likert, es un tipo de escala psicométrica de respuestas usada con frecuencia en cuestionarios, siendo la más usada en encuestas de investigación. Al responder el ítem de un cuestionario, la respuesta especifica el nivel de concordancia del encuestado con el enunciado en cuestión.

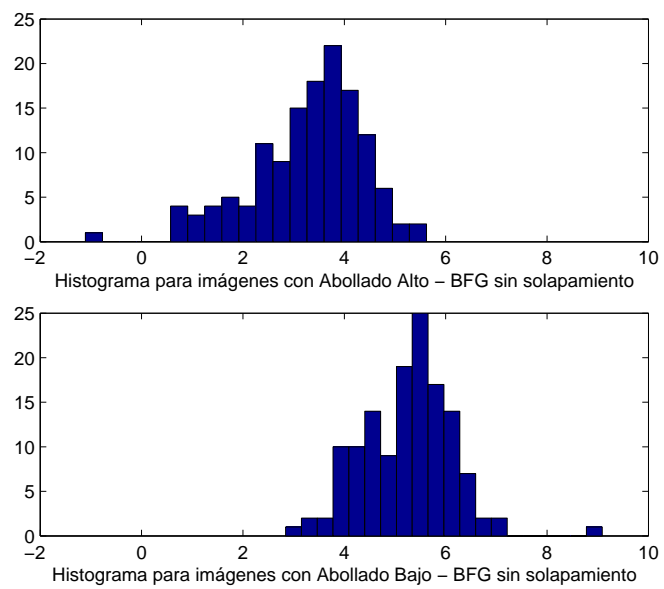


Figura 5.2: Histogramas generados con los valores obtenidos a partir de la función discriminante de Fisher aplicada a los vectores generados con el BFG sin solapamiento.

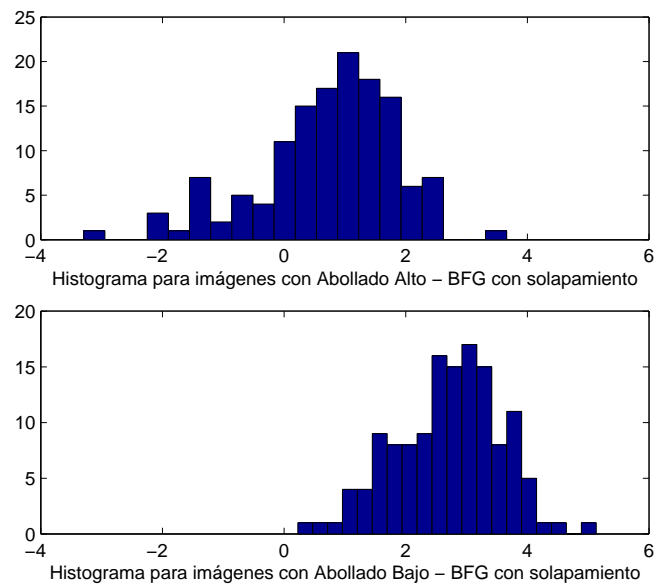


Figura 5.3: Histogramas generados con los valores obtenidos a partir de la función discriminante de Fisher aplicada a los vectores generados con el BFG con solapamiento.

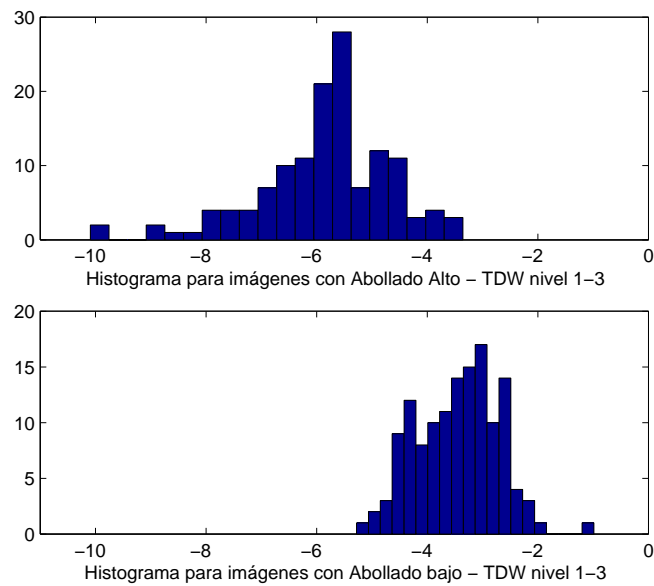


Figura 5.4: Histogramas generados con los valores obtenidos a partir de la función discriminante de Fisher aplicada a los vectores generados mediante los coeficientes de la TWD con wavelet madre Haar en los niveles 1-3.

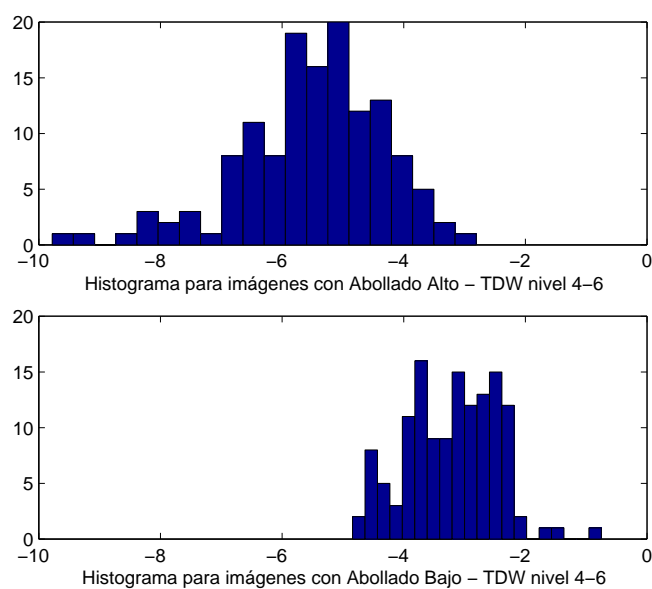


Figura 5.5: Histogramas generados con los valores obtenidos a partir de la función discriminante de Fisher aplicada a los vectores generados mediante los coeficientes de la TWD con wavelet madre Haar en los niveles 4-6.

5.8. Conclusiones

La detección y medida de las ondulaciones y protuberancias a nivel macroscópico sobre la superficie del papel reciclado, que hemos denominado abollado, puede realizarse por medio de técnicas de procesamiento de imágenes y reconocimiento de patrones. Hemos realizado una serie de experimentos con el objeto de encontrar un escalár que pueda ser utilizado como índice continuo. Así, el problema es planteado inicialmente como problema de clasificación: cada clase corresponde a un grado de abollado. Contrario a otros problemas de clasificación de texturas para inspección de calidad, la definición de las clases de textura y su relación con la medida del abollado no es trivial. Debido a las dificultades encontradas por los observadores humanos en el proceso de etiquetado manual, el número de clases ha sido reducido a tres: *Sin abollado*, *abollado Medio* y *abollado Alto*. La clase *Sin abollado* es trivial, ya que la imagen carece de alguna característica relacionada con el abollado. Hemos utilizado dos técnicas de análisis de texturas: los Bancos de Filtros de Gabor y la Transformada Discreta Wavelet en busca de una apropiada definición de las características texturales. Los algoritmos de clasificación utilizados inicialmente han sido los estándar k -NN y la red neuronal tipo MLP. Posteriormente hemos verificado los resultados obtenidos, mediante un conjunto de clasificadores implementados en una herramienta software. Específicamente el uso de Bancos de Filtros de Gabor (BFG) para calcular características de texturas de las imágenes nos ha permitido la construcción de un clasificador automático que ha alcanzado el éxito en la clasificación comparable con la obtenida por etiquetadores humanos. El uso de características de texturas basadas en los coeficientes de la Transformada Wavelet Discreta (TWD), mejora notablemente los resultados obtenidos por los BFG y la concordancia observada en los etiquetadores humanos, por lo tanto, las características basadas en TWD pueden considerarse óptimas para solucionar el problema en cuestión. Después de esto hemos querido derivar de los vectores de características utilizados, un escalár que cuantifique adecuadamente el abollado en el papel, como un índice continuo. Aunque los mejores resultados de clasificación fueron obtenidos por los vectores basados en la TWD, hemos usado también los vectores basados en BFG en este proceso. Primero hemos realizado una clasificación similar a la realizada con los vectores de características, pero con los valores escalares derivados con ellos. Específicamente hemos utilizado como escalares: la norma derivada de

los vectores, los componentes principales, y el valor obtenido para cada vector mediante el Análisis Discriminante de Fisher. Los mejores resultados han sido obtenidos con los valores derivados mediante el Análisis Discriminante de Fisher, alcanzando en el mejor de los casos un éxito superior al de las clasificaciones de los vectores de características completos. Para validar dicho escalar como un índice completo, hemos tenido que construir una nueva verdad del terreno. Para ello realizamos una selección aleatoria de pares de imágenes. Cada par ha sido mostrado a los usuarios expertos, con la hipótesis de que guardan cierto orden: la imagen más a la derecha es menos abollada que la imagen a la izquierda. El usuario decide si está o no de acuerdo con esa ordenación o si no puede decidir. Los resultados de esta comparación subjetiva han sido contrastados con el valor obtenido mediante la función discriminante de Fisher, en el que menor valor significa, mayor presencia de abollado. La correlación entre los valores obtenidos y las comparaciones realizadas por los usuarios coinciden hasta en un 89.9% en el mejor de los casos, el cual utiliza las características basadas en la TWD.

Podemos afirmar que el análisis de texturas basado en la TWD, combinado con el Análisis Discriminante de Fisher permiten derivar un índice continuo del nivel de abollado que podría ser incorporado en el proceso de control de calidad mediante un proceso de realimentación con los parámetros de producción.

Sistema CBIR para imágenes hiperespectrales

En este capítulo se describen los elementos para un sistema CBIR sobre imágenes hiperespectrales utilizando información espectral e información espacial. Para obtener características espectrales se aplica un tipo de redes neuronales, redes neuronales morfológicas, obteniendo el conjunto de endmembers de la imagen. Estos endmembers permiten el cálculo de la descomposición espectral de la imagen produciendo las imágenes de abundancias que se pueden utilizar para la búsqueda basada en la información espacial. Presentamos primero una revisión de conceptos relativos a memorias asociativas morfológicas. Después mostramos los resultados de segmentación de las imágenes comparando con algunas otras aproximaciones para el cálculo de los espectros endmember. Dichos resultados nos sirven de apoyo para esta aproximación como caracterización de las imágenes para su uso en la construcción de sistemas CBIR de imágenes hiperespectrales.

6.1. imágenes hiperespectrales

Las imágenes hiperespectrales corresponden a un muestreo fino del espectro de la luz. Habitualmente la región del espectro muestreada incluye el infrarrojo cercano (NIR) y el espectro visible. Los sensores para obtener estas imágenes utilizan alguna forma de descomponer la luz, mediante prismas o redes de difracción. Cada píxel de la imagen corresponde a un vector de alta dimensión que puede ser visualizado como una función de la frecuencia. La figura 6.1 [141] ilustra esta idea. Un sensor remoto (e.g. satélite) sobrevuela la zona visualizada. Cada píxel de la imagen obtenida corresponde a la señal

producida en el sensor por la luz reflejada en un área física de la superficie. En la figura se muestran los píxeles que se obtienen en distintas porciones de terreno. Se puede apreciar la variación típica entre espectros debida a la composición del suelo y de su cubierta. El espectro correspondiente a la vegetación tiene un aspecto distinto al suelo mineral y al agua. Los expertos humanos usualmente concentran su observación en las zonas del espectro de menor respuesta, las denominadas bandas de absorción, para distinguir entre materiales. Las imágenes hiperespectrales forman un cubo de información que puede ser tratado de diversas formas. En la figura 6.2 [141] se muestra una ilustración de esta idea. Las dimensiones horizontal y vertical corresponden a las dimensiones espaciales de la imagen, mientras que la profundidad corresponde a la frecuencia de la luz. Fijando una frecuencia obtenemos una imagen, que habitualmente se denomina *banda*. Una de las formas de proceso más realizada es la clasificación de los píxeles individuales para la obtención de mapas temáticos. La alta resolución espectral permite la distinción fina de los elementos presentes en la escena. Para esta clasificación se pueden utilizar las herramientas de reconocimiento estadístico de patrones [87]. El objetivo es obtener métodos automatizados de extracción de información de la imagen, los denominados *mapas temáticos*. La aportación fundamental de Landgrebe fue la formulación de estos mapas como un problema de clasificación automática, considerando los espectros como vectores en un espacio de alta dimensión, en lugar de tratar de imitar los procesos de clasificación manual seguidos por los expertos.

6.2. Descomposición espectral

Los espectros de los píxeles pueden considerarse como una unidad o como una combinación de espectros elementales, los denominados *endmembers*, en cuyo caso se trata de obtener la descomposición espectral del píxel, o la clasificación a nivel subpíxel. La composición de los píxeles puede ser lineal o no lineal. En la figura 6.3 se ilustra el fenómeno que da lugar a señales en el sensor (satélite) que pueden ser interpretadas como mezcla lineal de los espectros de materiales elementales. En este caso el área correspondiente a un píxel de la imagen cubre regiones homogéneas de materiales diversos, cada una de ellas devuelve un espectro definido pero lo que percibe el sensor es la radiación promedio del área, pesada por el porcentaje correspondiente a cada material. Los coeficientes de *abundancia* indican el porcentaje del píxel

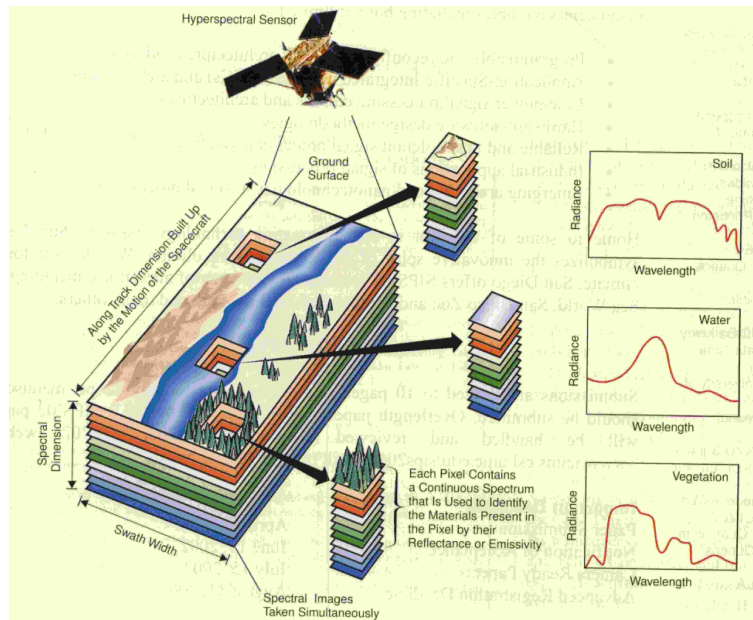


Figura 6.1: Ilustración de la captura de la imagen hiperespectral

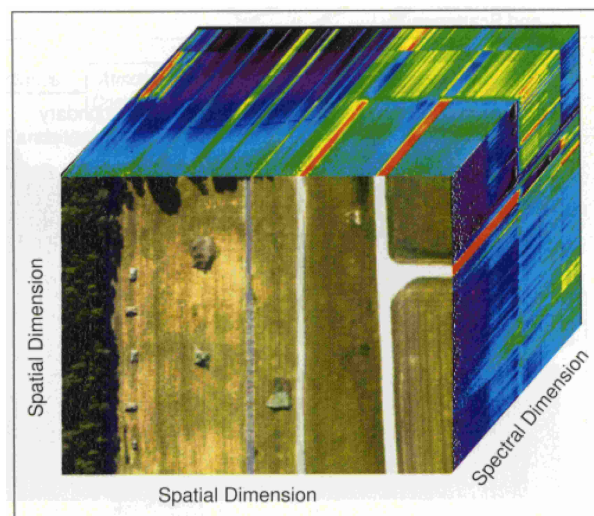


Figura 6.2: Ilustración del cubo de datos que constituye una imagen hiperespectral

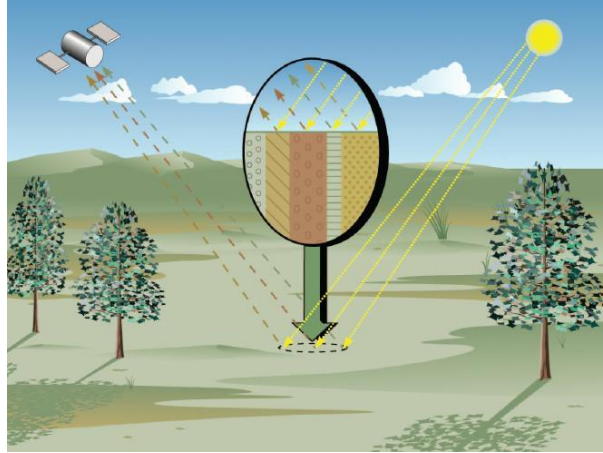


Figura 6.3: Ilustración de las causas de la mezcla lineal espectral

correspondiente a cada material. Al proceso de extraer estos coeficientes, conocidos los espectros de los materiales elementales presentes en el área de la superficie visualizada por el píxel, lo denominamos descomposición espectral [81].

El modelo de mezcla lineal [81] puede expresarse como sigue:

$$\mathbf{x} = \sum_{i=1}^M a_i \mathbf{s}_i + \mathbf{w} = \mathbf{S}\mathbf{a} + \mathbf{w}, \quad (6.1)$$

donde \mathbf{x} es el vector d -dimensional correspondiente al espectro recibido en el sensor, \mathbf{S} es la matriz $d \times M$ cuyas columnas son los vectores d -dimensionales correspondientes a los endmembers $\mathbf{s}_i, i = 1, \dots, M$, \mathbf{a} es el vector M -dimensional de abundancias fraccionales, y \mathbf{w} es un vector aleatorio d -dimensional con distribución gaussiana que modela el ruido aditivo de la observación. El modelo de mezcla lineal está sujeto a dos restricciones sobre los coeficientes de abundancia. Primeramente, para que sean significativos físicamente, todos los coeficientes de abundancia deben ser no negativos:

$$a_i \geq 0, i = 1, \dots, M. \quad (6.2)$$

En segundo lugar, para que la composición sea completa (no quede un “trozo” de píxel sin explicar o “sobren” componentes) los coeficientes de abundancia

deben ser completamente aditivos a la unidad:

$$\sum_{i=1}^M a_i = 1. \quad (6.3)$$

Una vez que los endmembers han sido determinados de alguna manera, la tarea de análisis de la imagen consiste en calcular la inversión que nos da la abundancia fraccional de cada endmember en cada espectro de cada pixel. La aproximación más sencilla es la de la estimación de mínimo error cuadrático sin restricciones, dada por:

$$\hat{\mathbf{a}} = (\mathbf{S}^T \mathbf{S})^{-1} \mathbf{S}^T \mathbf{x}. \quad (6.4)$$

Los coeficientes de abundancia que resultan de estos cálculos no necesariamente cumplen las condiciones 6.2 y 6.3. Es posible forzar cada condición por separado pero es bastante difícil hacerlo simultáneamente [81]. En nuestro caso se añade al problema el hecho de que los endmembers encontrados no definen un conjunto convexo que cubre todos los datos. Por esta razón y por simplicidad usaremos la estimación sin restricciones (6.4) para calcular las imágenes de abundancia. Las imágenes resultado se escalan en intensidad para facilitar la interpretación de los resultados.

6.3. Algoritmos de extracción de endmembers

El problema de la determinación de los endmembers es un problema clave para la realización de la descomposición espectral. Una aproximación inicial es la referencia a librerías estándar de espectros [81] conocidos a priori los materiales presentes en la escena. Esta aproximación necesita de un experto y tiene la limitación de que los espectros a utilizar tienen que ser obtenidos por el mismo sensor sobre el que se está trabajando. Las diferentes sensibilidades de las redes de difracción, y otros elementos del sensor, pueden hacer que datos provenientes de sensores distintos no sean comparables. Una de las primeras reflexiones [15] sobre la naturaleza de los endmembers desde un punto de vista geométrica es la constatación de que los endmembers son los vértices que definen un conjunto convexo que cubre completamente los datos de la imagen hiperespectral en un espacio de alta dimensión.

En el artículo de referencia clásico [34], Craig comienza con la observación de que los gráficos de dispersión (scatter plots) de los datos tienen forma

piramidal o de lágrima (*tear shaped*) cuando se consideran dos o tres bandas. El ápice cae en el denominado punto negro (*dark point*). La detección de los endmembers se convierte en la búsqueda de planos no ortogonales que envuelven a los datos de la imagen, formando un simplex de volumen mínimo. El método del Simplex de Volumen Mínimo es caro computacionalmente y requiere la especificación *a priori* del número de endmembers buscados.

Otro método de inducción automática de endmembers basado en criterios geométricos es el Analisis Cónico (*Conical Analysis*) (CCA) [78] con aplicaciones en detección de objetivos. En este método se realiza en primer lugar una transformación en componentes principales. El CCA selecciona tantos autovectores con los mayores autovalores como endmembers se desean. Estos autovectores definen la base del cono convexo que cubre los datos de la imagen. Los vértices del cono convexo corresponden a espectros con tantos coeficientes nulos como el número de autovalores seleccionados menos uno en su expresión lineal en términos de los autovectores seleccionados. El método es similar al de Craig en su fundamentación geométrica, pero no requiere la resolución de problemas de programación lineal. Sin embargo, requiere la exploración combinatoria de la selección de bandas y la solución de un sistema lineal para cada combinación. La complejidad de la búsqueda de estos vértices es $O(b^c)$ donde b es el número de bandas y c el número de autovectores seleccionados. En nuestros experimentos usamos un algoritmo de búsqueda aleatoria bruto. Obviamente, también requiere la especificación previa del número de endmembers deseado. El método N-FINDR [163] intenta expandir un simplex dentro de la nube de datos hasta que consigue cubrirla.

Otra aproximación distinta es el modelado mediante Campos Markovianos Aleatorios de la imagen para la detección de regiones espacialmente consistentes, cuyos espectros se asumen como endmembers [124]. Una aproximación basada en la morfología matemática se presenta en [121]. Los autores generalizan los operadores de erosión y dilatación morfológicos basándose en la distancia entre los píxeles dentro de un vecindario. Introducen una medida de excentricidad dada como una medida de la varianza en el vecindario que permite decidir sobre la validez del vecindario para la extracción de un endmember. El método usa para este fin vecindarios localizados en las regiones suaves y descarta los que se encuentran en regiones frontera. En general, los métodos de inducción de endmembers son intensivos computacionalmente y requieren varios pases sobre la imagen. Los métodos propuestos en [128] y [59, 60] sólo realizan un paso sobre la imagen y son ligeros computacionalmente.

6.4. Redes morfológicas e independencia morfológica

Lattice Computing, que podría traducirse como “Computación Reticular” puede ser identificada como una rama de la Inteligencia Computacional cuyos algoritmos se construyen utilizando el álgebra reticular $(\mathbb{R}, \vee, \wedge, +)$ o que aplican Teoría de Retículos para la generalización de algoritmos y paradigmas computacionales previos [56]. Las Memorias Asociativas Reticulares (*Lattice Associative Memories* y *Lattice Auto-associative Memories*) que nombraremos por su abreviatura en inglés (LAM) fueron propuestas bajo el nombre de Memorias Morfológicas (*Morphological Associative* y *Auto-associative Memories*) (MAM) [127, 126, 131] como la contraparte no lineal de las famosas Memorias Asociativas de Hopfield [69]. Admiten construcciones duales basadas en los operadores duales Min y Max. Las MAM se construyen como matrices de correlación calculadas usando el producto matricial Min o Max. La robustez y sensibilidad selectivas de las MAM a tipos específicos de ruido (erosivo y dilatativo) está en la raíz de su aplicación a la descomposición espectral de imágenes hiperespectrales. Fue establecido en [127, 126] que las memorias MAM son capaces de almacenar y recuperar conjuntos de patrones con la propiedad de Independencia Reticular Fuerte (*Strongly Lattice Independent*) (SLI). Específicamente la versión autoasociativa de las MAM tiene propiedades de recuperación universal bajo condiciones exentas de ruido. La robustez frente a ruido general fue investigada en [123, 130, 131]. Las nociones de Independencia Reticular e Independencia Reticular Fuerte fueron introducidas en [131] para estudiar la construcción de MAM con propiedades de recuperación robustas. Siguiendo esos trabajos, los estudios sobre los puntos fijos de las MAM han llevado [129] a una definición más precisa de la Independencia Reticular y de la SLI, incluyendo la relación entre SLI e independencia afín. Un resultado clave es el establecimiento de las condiciones para que un subconjunto de las columnas de la matriz de una MAM sean SLI y, por tanto, un conjunto de endmembers. Una transformación sencilla convierte la región convexa definida por estos endmembers en una aproximación del envolvente convexo de los datos en la imagen [128]. El punto de partida de nuestros trabajos es la búsqueda de métodos para la segmentación no supervisada de las imágenes hiperespectrales [59, 60] para detectar regiones que pueden necesitar un análisis más detallado, siguiendo las ideas avanzadas por [15]. La aproximación que más nos interesa es la

de detección de objetivos basada en filtrado lineal [104] y el modelo de descomposición espectral [81]. En [57, 58] presentamos algunos trabajos sobre el uso de las coordenadas convexas de un píxel como características para la construcción de sistemas de clasificación supervisada, que no han tenido continuidad.

6.4.1. Breve revisión de fundamentos

Los trabajos sobre Memorias Morfológicas surgen de la consideración de la estructura algebraica basada en los operadores Min y Max de los retículos $(\mathbb{R}, \vee, \wedge, +)$ como alternativa a la estructura algebraica $(\mathbb{R}, +, \cdot)$ convencionalmente usada para la definición de las Redes Neuronales Artificiales [127, 126]. Los operadores \vee y \wedge denotan, respectivamente, los operadores discretos máx y mín (resp. sup y ínf en el marco continuo). Dado un conjunto de pares de patrones input/output $(X, Y) = \{(\mathbf{x}^\xi, \mathbf{y}^\xi); \xi = 1, \dots, k\}$, una red neuronal heteroasociativa lineal basada en la correlación cruzada de los patrones [69] se construye como $W = \sum_{\xi} \mathbf{y}^\xi \cdot (\mathbf{x}^\xi)'$. Imitando este procedimiento constructivo [127, 126] proponen la siguiente construcción para las Memoras Asociativas Morfológicas (MAM):

$$W_{XY} = \bigwedge_{\xi=1}^k [\mathbf{y}^\xi \times (-\mathbf{x}^\xi)'] \quad \text{and} \quad M_{XY} = \bigvee_{\xi=1}^k [\mathbf{y}^\xi \times (-\mathbf{x}^\xi)'], \quad (6.5)$$

donde \times es cualquiera de los operadores \boxtimes ó \boxminus . Aquí \boxtimes and \boxminus denotan [127, 126] los productos matriciales máx y mín, respectivamente definidos como sigue:

$$C = A \boxtimes B = [c_{ij}] \Leftrightarrow c_{ij} = \bigvee_{k=1..n} \{a_{ik} + b_{kj}\}, \quad (6.6)$$

$$C = A \boxminus B = [c_{ij}] \Leftrightarrow c_{ij} = \bigwedge_{k=1..n} \{a_{ik} + b_{kj}\}. \quad (6.7)$$

Si $X = Y$ entonces las MAM son memorias autoasociativas. Las condiciones de recuperación perfecta de los patrones almacenados en la MAM autoasociativa demostradas en [127, 126] han empujado la investigación sobre ellas, porque en el caso continuo, la MAM es capaz de almacenar y recuperar cualquier conjunto de patrones. Formalmente:

$$W_{XX} \boxtimes X = X = M_{XX} \boxtimes X,$$

para cualquier conjunto de patrones X . Sin embargo, este resultado es aplicable sólo cuando tratamos con patrones libres de ruido. La búsqueda de técnicas de recuperación robustas a todo tipo de ruido [123, 126, 131, 147] basadas en los llamados patrones kernel, condujo a la noción de independencia morfológica, en el sentido erosivo y dilatativo, y, finalmente, a la de Independencia Morfológica Fuerte (*Strong Lattice Independence*) (SLI). A continuación enumeramos una serie de resultados extraídos de [129, 128] que proporcionan el respaldo teórico para la aplicación de las MAM a la tarea de inducción de endmembers a partir de los datos de la imagen.

Definición Dado un conjunto de vectores $\{\mathbf{x}^1, \dots, \mathbf{x}^k\} \subset \mathbb{R}^n$, una *combinación minimax lineal* de vectores de este conjunto es cualquier vector $\mathbf{x} \in \mathbb{R}_{\pm\infty}^n$ dado por una *suma minimax lineal* de estos vectores:

$$x = \mathcal{L}(\mathbf{x}^1, \dots, \mathbf{x}^k) = \bigvee_{j \in J} \bigwedge_{\xi=1}^k (a_{\xi j} + \mathbf{x}^\xi),$$

donde J es un conjunto finito de índices y $a_{\xi j} \in \mathbb{R}_{\pm\infty} \forall j \in J$ y $\forall \xi = 1, \dots, k$.

Definición El *espacio lineal minimax* de vectores $\{\mathbf{x}^1, \dots, \mathbf{x}^k\} = X \subset \mathbb{R}^n$ es el conjunto de todas las sumas minimax lineales de subconjuntos de X , denotado $LMS(\mathbf{x}^1, \dots, \mathbf{x}^k)$.

Definición Dado un conjunto de vectores $X = \{\mathbf{x}^1, \dots, \mathbf{x}^k\} \subset \mathbb{R}^n$, un vector $\mathbf{x} \in \mathbb{R}_{\pm\infty}^n$ es *morfológicamente dependiente* si y solo si $x = \mathcal{L}(\mathbf{x}^1, \dots, \mathbf{x}^k)$ para alguna suma minimax lineal de vectores de X . El vector \mathbf{x} es *morfológicamente independiente* si y solo si no es morfológicamente dependiente de X . El conjunto X se dice que es *morfológicamente independiente* si y solo si $\forall \lambda \in \{1, \dots, k\}$, \mathbf{x}^λ es morfológicamente independiente de $X \setminus \{\mathbf{x}^\lambda\} = \{\mathbf{x}^\xi \in X : \xi \neq \lambda\}$.

Definición Un conjunto de vectores $X = \{\mathbf{x}^1, \dots, \mathbf{x}^k\} \subset \mathbb{R}^n$ se dice que es *max dominante* si y solo si para cada $\lambda \in \{1, \dots, k\}$ existe un índice $j_\lambda \in \{1, \dots, n\}$ tal que

$$x_{j_\lambda}^\lambda - x_i^\lambda = \bigvee_{\xi=1}^k (x_{j_\lambda}^\xi - x_i^\xi) \forall i \in \{1, \dots, n\}.$$

Similarmente, X se dice que es *min dominante* si y solo si para cada $\lambda \in \{1, \dots, k\}$ existe un índice $j_\lambda \in \{1, \dots, n\}$ tal que

$$x_{j_\lambda}^\lambda - x_i^\lambda = \bigwedge_{\xi=1}^k (x_{j_\lambda}^\xi - x_i^\xi) \forall i \in \{1, \dots, n\}.$$

Definición Un conjunto de vectores independientes morfológicamente $\{\mathbf{x}^1, \dots, \mathbf{x}^k\} \subset \mathbb{R}^n$ se dice que es *fuertemente independiente morfológicamente* si y solo si X es max dominante o min dominante o ambos.

Definición El conjunto de puntos fijos de W_{XX} y M_{XX} es el mismo conjunto de todos los puntos fijos para ambas transformaciones

$$F(X) = \{\mathbf{x} \in \mathbb{R}^n : W_{XX} \boxtimes \mathbf{x} = \mathbf{M}_{XX} \boxtimes \mathbf{x}\}.$$

Teorema 6.4.1 Para cada $k \in \{1, \dots, k\}$ existe un conjunto de vectores $X \subset \mathbb{R}^n$ tal que $\dim F(X) = k$ y X es fuertemente independiente morfológicamente.

Corolario 6.4.2 Si $X = \{\mathbf{x}^1, \dots, \mathbf{x}^k\} \subset \mathbb{R}^n$ es fuertemente independiente morfológicamente entonces $k \leq n$.

Teorema 6.4.3 Si $X = \{\mathbf{x}^1, \dots, \mathbf{x}^k\} \subset \mathbb{R}^n$ es fuertemente independiente morfológicamente entonces X es independiente afín.

Teorema 6.4.4 Sea $X = \{\mathbf{x}^1, \dots, \mathbf{x}^k\} \subset \mathbb{R}^n$ y sea W (M) el conjunto de vectores que consiste de las columnas de la matriz W_{XX} (M_{XX}). Existen $V \subset W$ y $N \subset M$ tales que V y N son fuertemente independientes morfológicamente y $F(X) = F(W) = F(N)$ o, equivalentemente, $W_{XX} = W_{VV}$ y $M_{XX} = M_{NN}$.

6.4.2. Algoritmo heurístico de inducción de endmembers

En el algoritmo 1 se muestra el pseudo-código para el algoritmo de inducción de endmembers originalmente propuesto en [59, 60, 61]. Al desarrollar este algoritmo se asumía de forma intuitiva (heurística) la relación entre la independencia morfológica y la independencia afín, esto es, que los vectores independientes morfológicamente serían independientes afín y podrían

Algorithm 1 Algoritmo heurístico de inducción de los endmembers

1. Desplazar los espectros al origen del espacio
 $\{\mathbf{f}^c(i) = \mathbf{f}(i) - \boldsymbol{\mu}; i = 1, \dots, n\}$.
2. Inicializa el conjunto de endmembers con un espectro escogido al azar en la imagen $E = \{\mathbf{e}_1\}$. Inicializa el conjunto de firmas binarias morfológicamente independientes

$$X = \{\mathbf{x}_1\} = \{(e_k^1 > 0; k = 1, \dots, d)\}$$

3. Construye las MAM basadas en las firmas binarias de los endmembers : M_{XX} y W_{XX} .
 4. Para cada pixel $\mathbf{f}^c(i)$
 - a) calcula las firmas binarias del vector tras la dilatación y erosión por el término de ruido $\mathbf{f}^+(i) = (\mathbf{f}^c(i) + \alpha\boldsymbol{\sigma} > \mathbf{0})$ y $\mathbf{f}^-(i) = (\mathbf{f}^c(i) - \alpha\boldsymbol{\sigma} > \mathbf{0})$
 - b) calcula $y^+ = M_{XX} \boxtimes \mathbf{f}^+(i)$
 - c) calcula $y^- = W_{XX} \boxtimes \mathbf{f}^-(i)$
 - d) si $y^+ \notin X$ o $y^- \notin X$ entonces $\mathbf{f}^c(i)$ es un nuevo endmember a añadir a E , ejecuta una vez 3 con el nuevo conjunto de endmembers E y continúa la exploración de la imagen.
 - e) si $y^+ \in X$ y $\mathbf{f}^c(i) > \mathbf{e}_{y^+}$ el espectro del nuevo pixel considerado es una dilatación del endmember almacenado, substituye \mathbf{e}_{y^+} por $\mathbf{f}^c(i)$ en E .
 - f) si $y^- \in X$ y $\mathbf{f}^c(i) < \mathbf{e}_{y^-}$ el espectro del nuevo pixel considerado es una erosión del endmember almacenado, substituye \mathbf{e}_{y^-} por $\mathbf{f}^c(i)$ en E .
 5. El conjunto final de endmembers son los espectros originales en $\mathbf{f}(i)$ correspondientes a los vectores seleccionados en E .
-

considerarse vértices de un conjunto convexo. Para determinar si un vector es dependiente morfológicamente el algoritmo utiliza las MAM construidas sobre los vectores binarios de signo de los componentes del espectro desplazado por la media de los espectros en la imagen.

En el algoritmo $\{\mathbf{f}(i) \in \mathbb{R}^d; i = 1, \dots, n\}$ denota los datos de alta dimensión que corresponden a los píxeles de la imagen multi o hiperespectral. Los vectores $\boldsymbol{\mu}$ and $\boldsymbol{\sigma}$ representan, respectivamente, el vector media y el vector de las desviaciones estándar de las bandas, calculado sobre los datos de la imagen. Denotamos α el factor las desviaciones estándar por banda que sirve para descartar los vectores que pueden considerarse como meras perturbaciones ruidosas de alguno de los endmembers ya encontrados. Finalmente E es el conjunto de endmembers descubiertos hasta el momento. Por tanto, consideramos que la varianza del ruido en la ecuación 6.1 viene dada por $\boldsymbol{\sigma}^2$. En el algoritmo 1 la forma de tratar el ruido consiste en crear dos patrones que corresponden a la erosión y dilatación del patrón por $\alpha\boldsymbol{\sigma}$, antes de ser binarizados y presentados a las MAM construidas con los vectores signo de los endmembers. De esta manera, el parámetro α controla la flexibilidad en el descubrimiento de nuevos endmembers.

La expresión $\mathbf{x} > \mathbf{0}$ denota la construcción de un vector binario

$$(\{b_i = 1 \text{ if } x_i > 0; b_i = 0 \text{ if } x_i \leq 0\}; i = 1, \dots, n).$$

En realidad el algoritmo busca hipercuadrantes vacíos que tengan datos que no puedan ser explicados por el ruido a partir de los endmembers detectados.

6.5. Resultados de segmentación

En esta sección mostramos algunos resultados de segmentación no supervisada de imágenes hiperespectrales de referencia que demuestran el poder del algoritmo heurístico propuesto. En primer lugar consideramos la imagen del Whashington D.C. Mall proporcionada con el libro de Landgrebe [87]. La imagen tiene 191 bandas. Fue obtenida con el sensor AVIRIS (220 bandas) y las bandas nulas fueron eliminadas. Tiene la ventaja de ser una escena bien conocida y conteniendo estructuras que pueden identificarse con facilidad.

La aplicación del algoritmo heurístico descrito en la sección 6.4.2 sobre toda la imagen de Whashington D.C. Mall, con el factor de filtrado del ruido fijado a $\alpha = 2$ produce los seis endmembers graficados en la figura 6.4 y las imágenes de abundancia mostradas en la figura 6.5. Las tres primeras

imágenes de abundancia parecen tener la mayor parte de la información, mientras que las otras tres son más parecidas a imágenes residuales. La imagen de abundancia en la figura 6.5 parece detectar áreas de vegetación. Tiene valores altos para las zonas ajardinadas, árboles y prados, mientras que las zonas negras corresponden a edificios, agua y carreteras. La imagen de abundancia en la figura 6.5(2) detecta áreas de gran absorción. Las más altas respuestas se encuentran en las zonas de sombra (obelisco del monumento a Washington), agua como la piscina frente al monumento a Lincoln, y las carreteras. Por último, la imagen de abundancia 6.5(1) detecta tejados y estructuras construidas en general.

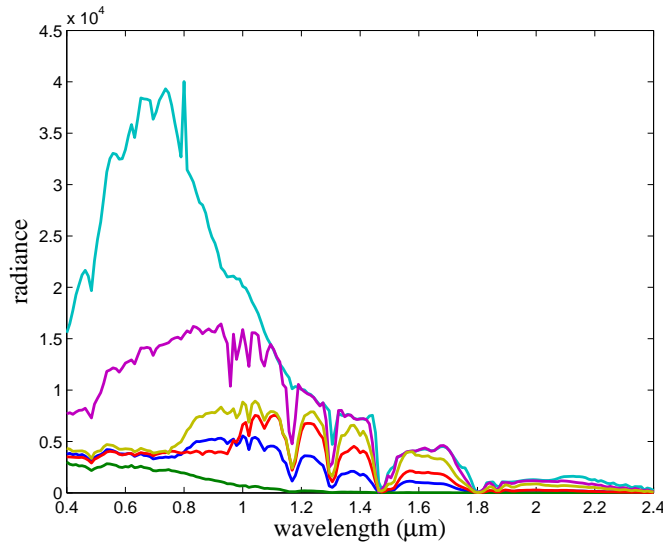


Figura 6.4: Endmembers encontrados por nuestro algoritmo heurístico basado en independencia morfológica sobre la imagen de Washington D.C. Mall

El segundo ejemplo clásico que mostramos es la imagen Indian Pines 1992 obtenida por el sensor AVIRIS, que tiene 224 bandas espectrales contiguas que cubren una región espectral desde 0.4 a 2.5 μm en intervalos de 10 nm. La imagen tiene 145 por 145 píxeles con 220 bandas espectrales. La escena contiene una distribución de dos tercios de terreno agrícola y un tercio de bosque y otros elementos (dos autopistas, un ferrocarril y algunas casas y carreteras menores). La verdad del terreno disponible para esta imagen se

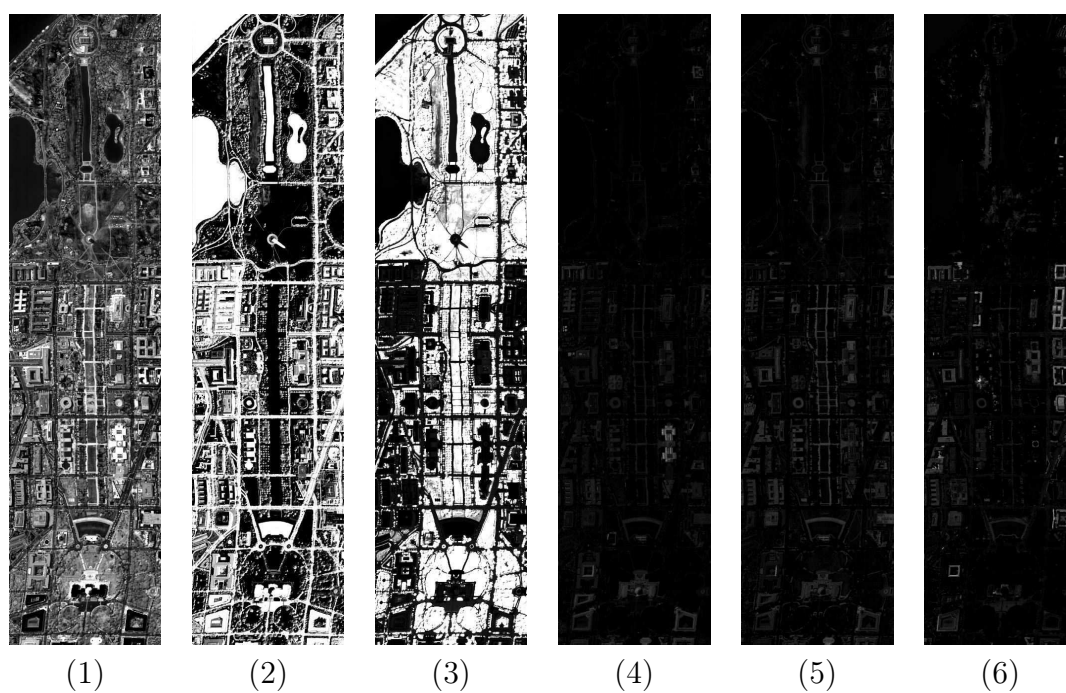


Figura 6.5: Imágenes de abundancia calculadas usando los endmembers de la figura 6.4

muestra en 6.6. Designa 16 clases de cubierta del terreno, que son clases mutuamente excluyentes. Las áreas etiquetadas como fondo (*background*) contienen píxeles de otras clases, sobre todo vegetación. La verdad del terreno tampoco tiene etiquetadas algunas estructuras como las carreteras que resultan evidentes cuando se realizan procesos de segmentación no supervisada o clasificación supervisada. Algunos píxeles asignados a clases de vegetación son difíciles de discriminar debido a que la imagen corresponde a una época temprana en el crecimiento de las plantas, por lo que gran parte del área corresponde a suelo mineral. También ocurre que algunas de las clases de cubierta están presentes marginalmente en áreas etiquetadas por otra clase. Esto es debido a que el etiquetado se realizó en base al conocimiento cualitativo del área y no a la observación directa. Nuestro algoritmo no supervisado descubre este tipo de características de la imagen.

Esta imagen ha servido de benchmark para validar algoritmos de clasificación supervisada, por ejemplo en [149, 150, 151]. La figura 6.7 muestra los endmembers encontrados por una ejecución de nuestro algoritmo descrito en la sección 6.4.2 sobre la imagen Indian Pines 1992, con un valor del factor de filtrado del ruido $\alpha = 2$. La figura 6.8 muestra las imágenes de abundancia correspondientes a los endmembers de la figura 6.7. Algunas estructuras, como las torres de acero en la parte superior de la figura, se aprecian con mucha claridad. La imagen de abundancia #5 resalta los campos cultivados, probablemente debido a que detecta las superficies minerales más bien que la cubierta vegetal. La figura 6.9 muestra el resultado de un experimento de clasificación supervisada. La comparación con las imágenes de abundancia en la figura 6.8 permite apreciar que estas últimas detectan muchas de las estructuras presentes en la imagen. Por ejemplo, en la imagen de abundancia #7 se detectan las zonas boscosas que en la verdad del terreno estaban no identificadas (asignadas a la clase fondo) de la misma manera que en la clasificación supervisada de la figura 6.9.

El propósito de esta sección era demostrar la efectividad del algoritmo de la sección 6.4.2 para la caracterización de las imágenes a través de las segmentaciones no supervisadas obtenidas como imágenes de abundancia a partir de los endmembers inducidos por el algoritmo.

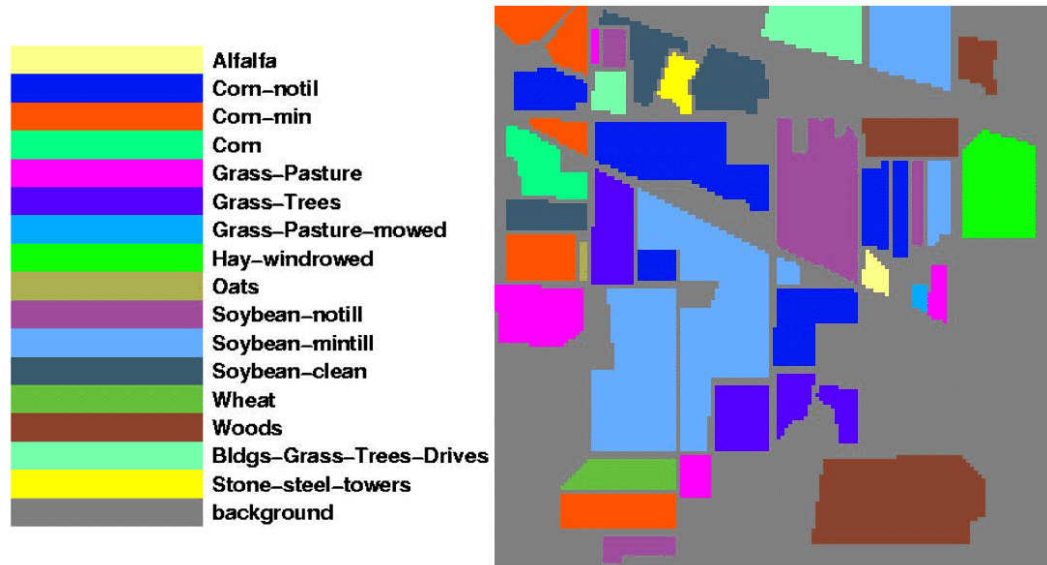


Figura 6.6: Indian Pines 1992, verdad del terreno

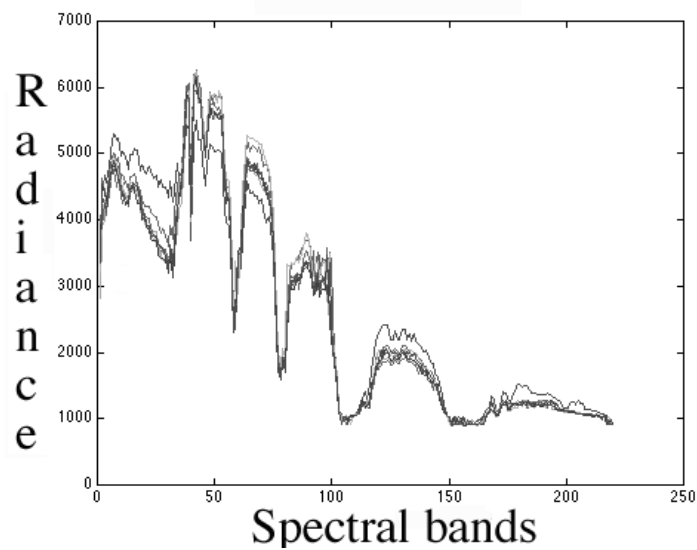


Figura 6.7: Endmembers encontrados por el algoritmo heurístico de la sección 6.4.2 en la imagen Indian Pines

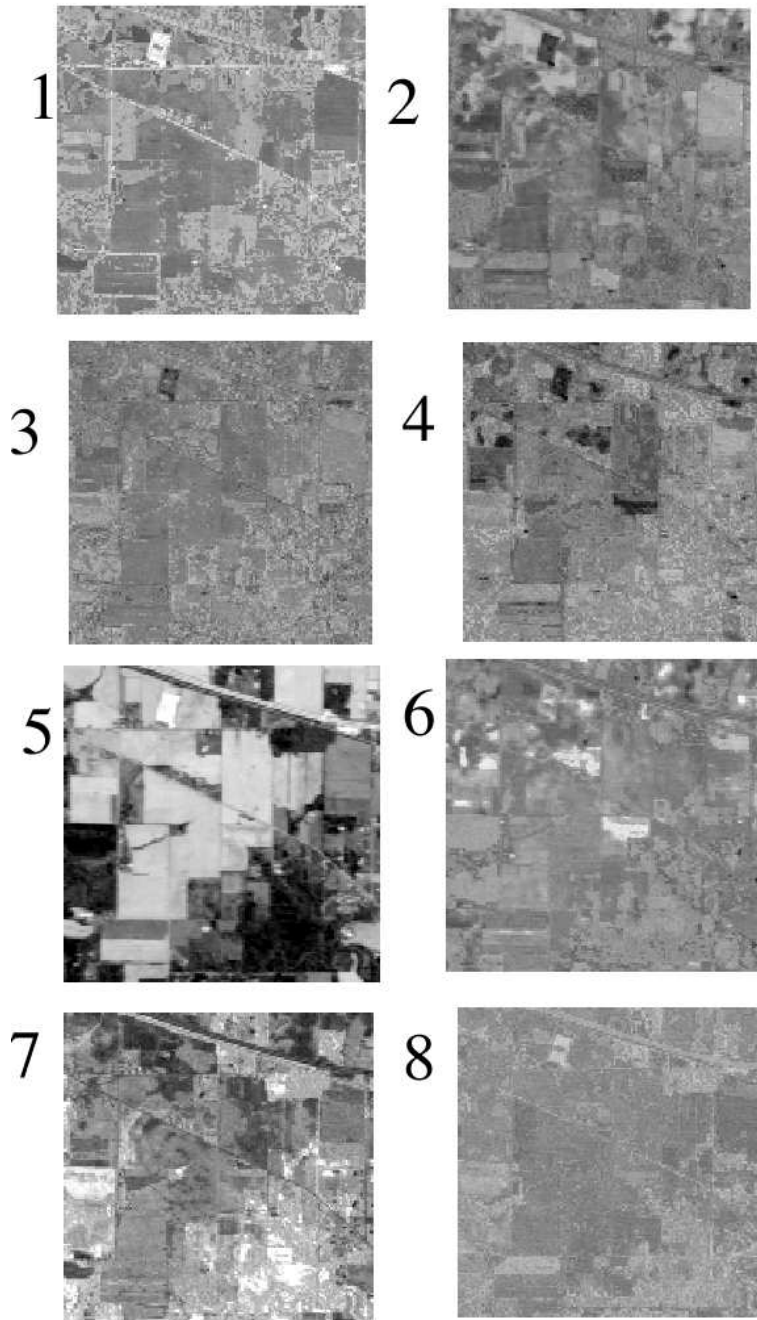


Figura 6.8: Abundancias calculadas usando los endmembers en la figura 6.7

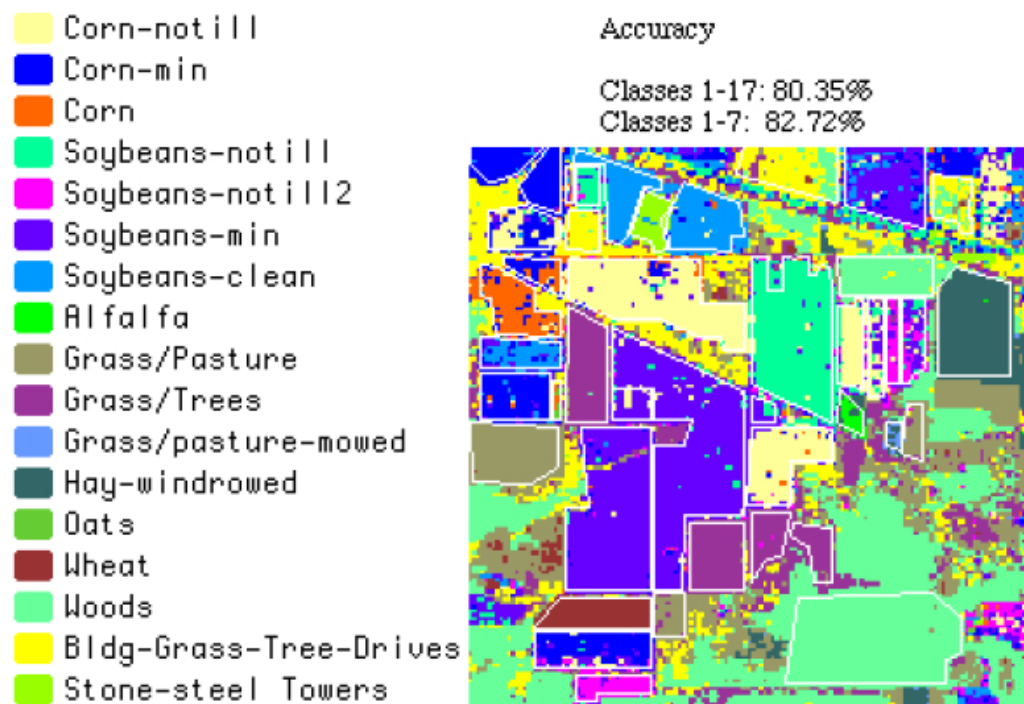


Figura 6.9: Resultado de la clasificación supervisada presentada en [150, 149]

6.6. Distancia entre imágenes hiperespectrales

En esta sección elaboramos la distancia entre imágenes hiperespectrales basada en la caracterización dada por los endmembers inducidos por el algoritmo de la sección 6.4.2. Sea $E^k = [e_1^k, \dots, e_{n_k}^k]$ el conjunto de endmembers de la k -ésima imagen en la base de datos $f_k(x, y)$. Dadas dos imágenes hiperespectrales, $f_k(x, y)$ y $f_l(x, y)$, sus conjuntos característicos de endmembers E^k y E^l pueden tener un número distinto de endmembers, i.e. $n_k \neq n_l$. Esto implica que los espacios de características son diferentes, sin relación funcional entre ellos. El cálculo de la similitud entre las imágenes comienza con el cálculo de la matriz de distancias euclídeas entre todos los posibles pares de endmembers correspondientes entre imágenes:

$$D_{k,l} = [d_{i,j}; i = 1, \dots, n_k; j = 1, \dots, n_l]$$

donde

$$d_{ij} = \|e_i^k - e_j^l\|.$$

Calculamos el vector de valores mínimos por filas

$$\mathbf{m}_k = [m_i^k = \min_j \{d_{ij}\}],$$

y el vector de valores mínimos por columnas

$$\mathbf{m}_l = [m_j^l = \min_i \{d_{ij}\}].$$

Calculamos la similitud entre imágenes hiperespectrales como sigue:

$$d(f_k, f_l) = (\|\mathbf{m}_k\| + \|\mathbf{m}_l\|) (|n_k - n_l| + 1). \quad (6.8)$$

La medida de similitud de la ecuación 6.8 es una composición de dos vistas asimétricas: cada vector de distancias mínimas mide como de cerca están los endmembers de una imagen respecto de los endmembers de otra imagen. El caso peor es cuando todos los endmembers E^k de la imagen f_k están muy cerca de un subconjunto E_*^l de los endmembers E^l de la imagen f_l , y lejos de los restantes $E^l - E_*^l$. En este caso, la magnitud del vector de distancias mínimas por filas será muy pequeña, ya que todas las filas tendrán un valor muy pequeño. Sin embargo, la magnitud del vector de distancias mínimas por columna será mucho mayor, ya que las columnas correspondientes al subconjunto de endmembers $E^l - E_*^l$ tendrán un valor mínimo relativamente

grande. Por tanto la medida de similitud de la ecuación 6.8 puede detectar la asimetría de la situación. Además, la formulación de la ecuación 6.8 evita el problema combinatorio de decidir cual es el emparejamiento óptimo de los endmembers. Cuando el número de endmembers es diferente de una imagen a otra, su diferencia se introduce como un factor de amplificación. La medida es independiente del tamaño de la imagen. También puede calcularse en un tiempo aceptable puesto que el algoritmo de la sección 6.4.2 es muy rápido. El conjunto de endmembers no plantea ningún problema de almacenamiento. Esta aproximación puede complementarse con el cálculo de la similitud espacial de las imágenes de abundancias.

6.7. Resultados sobre bases de datos de imágenes sintéticas

Las imágenes hiperespectrales usadas para los experimentos computacionales han sido generadas como mezclas lineales de un conjunto de espectros, que llamaremos los *endmembers verdad del terreno*, usando como coeficientes imágenes de abundancia sintéticas. Los endmembers han sido seleccionados aleatoriamente de un subconjunto de la librería espectral de la U.S. Geological Survey correspondiente a los vuelos AVIRIS. La figura 6.10 muestra los espectros que se utilizan en las imágenes con 5 endmembers.

6.7.1. Experimento 1

Para el primer experimento computacional, las imágenes de abundancia verdad del terreno fueron generadas en un proceso en dos pasos. En primer lugar, se simula cada una como un campo gaussiano aleatorio con una función de correlación de Mattern con parámetros variando entre 2 y 20. Aplicamos los procedimientos propuestos en [84] para la generación eficiente de campos markovianos de gran dominio. En segundo lugar, para asegurar que hay regiones cuyos espectros son endmembers casi puros, seleccionamos para cada píxel el coeficiente de abundancia con el mayor valor y normalizamos los restantes para asegurar que la suma de coeficientes de abundancia es uno. Se puede apreciar en las imágenes de abundancia que cada endmember tiene varias regiones de endmembers casi puros, que aparecen como regiones casi blancas al visualizar las imágenes de abundancia. El tamaño de las imágenes

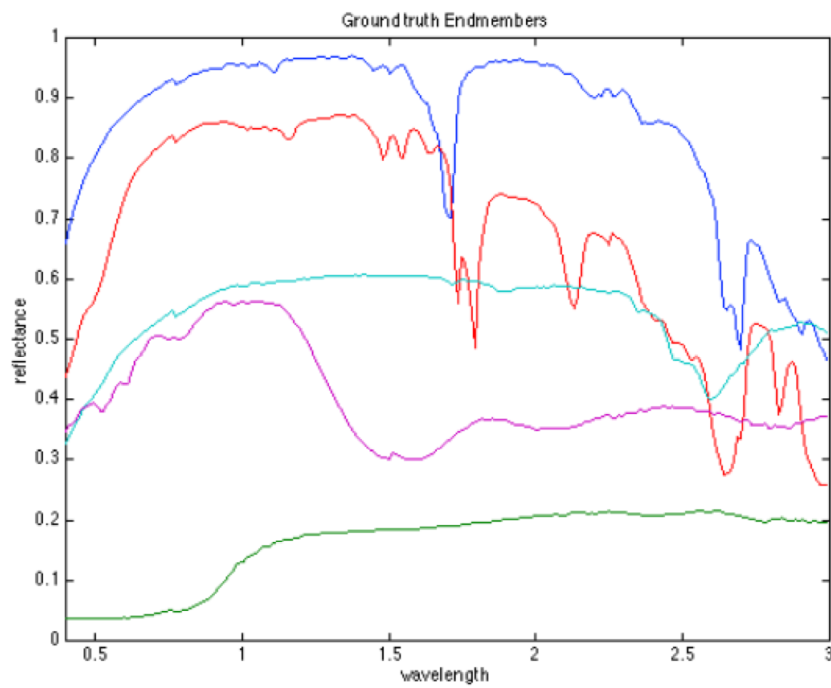


Figura 6.10: Espectros de repositorio de la USGS usados como endmembers de la verdad del terreno en el primer experimento.

es 256x256 píxeles con 224 bandas espectrales. Hemos generado colecciones de imágenes con 2 a 5 endmembers. En total, 400 imágenes, 100 para cada número de endmembers.

El experimento realizado consiste en los siguientes pasos:

1. Calcula las similitudes entre las imágenes en la base de datos, definida por la ecuación 6.8, usando los endmembers verdad del terreno. Las distancias se calculan entre imágenes con el mismo número de endmembers y con todas las demás.
2. Extraemos los endmembers de las imágenes usando el algoritmo descrito en la sección 6.4.2, con un valor de filtrado del ruido $\alpha = 0,5$
3. Calcula las similitudes entre las imágenes en la base de datos, definida por la ecuación 6.8, usando los endmembers inducidos por el algoritmo de la sección 6.4.2. Las distancias se calculan entre todas las imágenes, independientemente del número de endmembers inducidos.
4. Consideramos las R imágenes más cercanas a cada imagen en la base de datos, para las dos matrices de distancias calculadas en base a los endmembers verdad del terreno y a los inducidos como las respuestas en cada caso a una potencial consulta representada por dicha imagen.
5. Las imágenes que aparecen en ambas respuestas se consideran las imágenes relevantes o respuestas correctas.

En la tabla 6.1 presentamos los resultados del experimento con 400 imágenes, en términos del número promedio de respuestas correctas (imágenes relevantes). La primera fila presenta los resultados cuando la búsqueda se extiende a todas las imágenes, independientemente del número de endmembers. Las siguientes filas presentan los resultados cuando restringimos la búsqueda a la subcolección de imágenes con el mismo número de endmembers que la imagen consulta.

Se puede apreciar en la tabla 6.1 que la consideración de todas las imágenes como respuestas a la consulta introduce alguna confusión y reduce el número promedio de imágenes relevantes obtenidas en la consulta. Este efecto se debe a que el algoritmo de la sección 6.4.2 puede encontrar un número de endmembers distinto a la verdad del terreno, haciendo posible que se produzcan emparejamientos fuera de su colección natural de imágenes. Entonces,

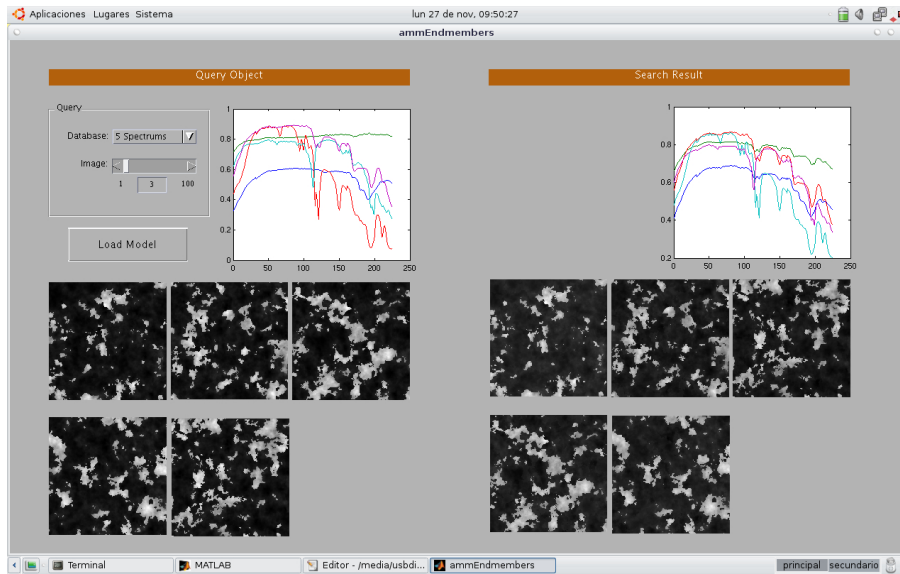


Figura 6.11: Un ejemplo de la interface de un sistema CBIR para imágenes hiperespectrales: una consulta y sus imágenes más cercanas de acuerdo al conjunto de endmembers inducido.

imágenes con diferentes números de endmembers en la verdad del terreno puede llegar a ser similares como para entrar en los respectivos conjuntos de respuestas.

Cuando se restringe la búsqueda a las colecciones con estrictamente el mismo número de endmembers verdad del terreno, los resultados mejoran excepto para $R = 1$. Obtenemos que cerca del 50% de las respuestas son significativas cuando $R > 1$. El caso $R = 1$ puede interpretarse como la probabilidad de obtener la imagen más cercana en la base de datos de acuerdo a la distancia definida en la ecuación 6.8 o la probabilidad de éxito en la clasificación. Puede verse que es muy cercana a 1 en todos los casos, excepto para las colecciones con 2 endmembers en la verdad del terreno.

6.7.2. Experimento 2

Para el segundo experimento hemos generado un conjunto de imágenes con un nuevo conjunto de posibles endmembers. En la figura 6.12 se muestra el conjunto de espectros que se han utilizado como endmembers. Para gener-

| | R=1 | R=3 | R=5 | R=10 |
|--------------------|------|------|------|------|
| Todas las imágenes | 0.94 | 1.21 | 1.61 | 2.96 |
| 2 endmembers | 0.81 | 1.55 | 2.27 | 4.67 |
| 3 endmembers | 0.98 | 1.44 | 2.21 | 4.96 |
| 4 endmembers | 0.99 | 1.53 | 2.36 | 4.81 |
| 5 endmembers | 1.00 | 1.57 | 2.37 | 4.74 |

Cuadro 6.1: Resultados de relevancia en las respuestas a las consultas sobre la base de datos de 400 imágenes sintéticas, usando la distancia definida en la ecuación 6.8

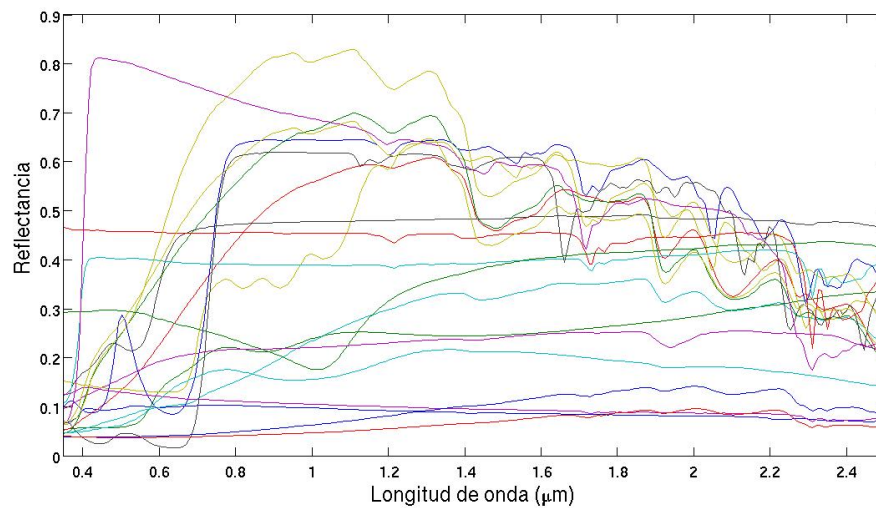


Figura 6.12: Endmembers verdad del terreno utilizados para el segundo experimento



Figura 6.13: Una instancia de las imágenes de abundancia generadas como campos basados en polinomios de Legendre para una imagen con cinco end-members.

ar las imágenes se han seleccionado aleatoriamente un número entre 2 y 5 de estos espectros para construir la imagen en base a las abundancias sintéticas de forma similar al experimento 1. Las imágenes de abundancia son campos 2D generados como productos de polinomios de Legendre. Se han generado 400 imágenes, con 2 a 5 endmembers. En la figura 6.13 se muestra una instancia de las imágenes de abundancia generadas para construir una imagen hiperespectral con cinco endmembers. En este segundo experimento se considera la distancia euclídea entre las imágenes de abundancias en lugar de los endmembers. De esta forma estamos probando la capacidad de nuestra aproximación para obtener caracterizaciones espaciales y espectrales, puesto que el cálculo de las abundancias lleva consigo la caracterización espectral adecuada.

El esquema de realización del experimento es similar al del primer experimento:

1. Se calculan las distancias entre las imágenes de abundancia de la verdad del terreno, sustituyendo los endmembers en la ecuación 6.8 por las propias imágenes de abundancia.
2. Se calculan las imágenes de abundancia a partir de los endmembers inducidos por el algoritmo de la sección 6.4.2.
3. Se calculan los conjuntos de respuestas para cada una de las imágenes en ambos casos: verdad del terreno e imágenes inducidas. La intersección de las respuestas se considera el número de respuestas correctas o imágenes relevantes obtenidas por la consulta.

En la tabla 6.2 se muestra los resultados de relevancia al realizar la consulta en base a la información espacial para las 400 imágenes de la base de datos. En general los resultados son inferiores a los presentados en la tabla 6.1 lo que es indicativo de que nuestra medida espacial por sí sola no añade nada a la medida de similitud basada en los espectros. Sin embargo, debemos notar que los endmembers presentados en la figura 6.12 presentan pocas diferencias en muchos casos y, en algunos casos, no son muy marcadamente independientes. Por otro lado, las imágenes de abundancia parecen tener pocas regiones de píxeles puros, por lo que el algoritmo de la sección 6.4.2 puede tener especiales dificultades para obtener endmembers similares a los de la verdad del terreno. La eficiencia de la recuperación es muy alta con 2 endmembers, pero decae mucho para un mayor número de endmembers.

| | R=1 | R=3 | R=5 | R=10 |
|--------------------|-----|------|------|------|
| Todas las imágenes | 1 | 1,38 | 1,79 | 2,89 |
| 2 endmembers | 1 | 2,48 | 4,03 | 8,02 |
| 3 endmembers | 1 | 1,04 | 1,14 | 1,58 |
| 4 endmembers | 1 | 1,03 | 1,12 | 1,51 |
| 5 endmembers | 1 | 1,01 | 1,09 | 1,31 |

Cuadro 6.2: Resultados de relevancia basada en la distancia entre imágenes de abundancia para imágenes con abundancias generadas con polinomios de Legendre 2D.

6.8. Conclusiones

En este capítulo se ha introducido una técnica para la indexación de imágenes hiperespectrales que permite la búsqueda en bases de datos con imágenes heterogéneas en el número de endmembers. La medida de similitud entre imágenes propuesta se basa en la distancia entre conjuntos de endmembers extraídos de las imágenes hiperespectrales por medio de un algoritmo de detección de patrones fuertemente independientes morfológicamente, que usa las Memorias Autoasociativas Morfológicas. Los experimentos de segmentación no supervisada muestran que el algoritmo obtiene endmembers que caracterizan el contenido espectral de la imagen. La medida de similitud propuesta es capaz de solventar situaciones asimétricas y es independiente de que las imágenes tengan distintos números de endmembers. Hemos realizado dos experimentos computacionales con imágenes sintéticas que demuestra la potencia de la aproximación. Considerando la imagen más cercana se obtiene una relevancia próxima al 100% y para tamaños mayores de respuestas se obtienen altas relevancias en las respuestas en el caso de la distancia basada en los espectros únicamente. En el segundo experimento, cuando se consideran las similitudes en base a la distancia espacial entre las imágenes de abundancia, el rendimiento desciende considerablemente. Parte de la responsabilidad recae en problemas de diseño de la base de datos experimental, que dificultan el trabajo del algoritmo de extracción de endmembers. De todas maneras, es posible que una nueva distancia que combine explícitamente ambas fuentes

de información (espacial y espectral) mejore los resultados.

Conclusiones y líneas de trabajo futuro

En este capítulo vamos resumir algunas de las conclusiones que se deducen de nuestros trabajos en el marco de la tesis doctoral y proyectar algunas líneas de trabajo futuro que el propio doctorando piensa seguir o que esperamos sirvan de inspiración para investigadores en algún tiempo futuro.

7.1. Caracterización visual de la calidad del papel

En su formulación inicial, el proyecto de trabajo con la empresa Echezarreta SA llegaba hasta la formulación de mecanismos de realimentación de la medida de la calidad del papel en la formulación de consignas de producción en tiempo real. De esta forma, el proceso de control de calidad integraba el CBIR que asociaría las medidas tomadas sobre la producción actual con medidas anteriores y sus condiciones de producción.

Los trabajos realizados han llegado hasta el punto de obtener un índice escalar que refleja bastante bien el nivel de abollado de la imagen del papel. Estos trabajos se han realizado siguiendo una metodología rigurosa con validación por parte de los usuarios. De entre las posibles técnicas de caracterización de la textura en la imagen, hemos escogido las que ofrecen la mayor flexibilidad y capacidad de caracterización. No parece probable que surgan nuevas técnicas que mejoren a las utilizadas. Tampoco parece probable que una experimentación más extensa pueda aportar nuevas luces.

El problema de la caracterización visual nos parece ya cerrado, puesto que se han conseguido los objetivos propuestos y no hay posibilidad de continuar

con la línea de integración del sistema CBIR en el proceso de producción.

7.2. Sistemas CBIR de imágenes hiperespectrales

Hemos realizado extensas (en relación a la dimensión de los conjuntos experimentales referidos en la literatura) pruebas experimentales que demuestran la potencia de los algoritmos de extracción de endmembers y la bondad de la medida de similitud definida sobre los conjuntos de endmembers. Para contrastar la aproximación propuesta será necesario continuar con la experimentación aplicando otros algoritmos de extracción de endmembers. El mayor inconveniente es el gran costo computacional de estos algoritmos que hacen muy lenta la experimentación con medios convencionales. Plantearse llevar la experimentación a una plataforma de sistemas multiprocesador supone abrir una nueva línea de trabajo (e investigación) relacionada con las implementaciones paralelas, que cae fuera del ámbito de la tesis y de las propias líneas de trabajo del doctorando. De todas maneras, se tratará en la medida de lo posible de obtener contrastes sobre estas bases de datos usando otros algoritmos de extracción de endmembers.

En nuestros trabajos los mejores resultados los hemos encontrado cuando aplicamos la distancia sobre el conjunto de endmembers. Cuando los aplicamos sobre las imágenes de abundancias obtenemos considerablemente peores respuestas. La combinación de ambas informaciones, espectral y espacial, en una sola función parece una vía prometedora para mejorar el rendimiento de esta aproximación. En principio, nos parece que el término más importante sería la información espectral, puesto que es la que distingue entre distintas composiciones materiales de la escena. El término espacial serviría para eliminar ambigüedades entre escenas con el mismo contenido espectral.

7.3. Aplicaciones de Lattice Computing a CBIR

Una de los efectos de la revisión del estado del arte ha sido la realización de la similitud entre los denominados R-tree y algunas aproximaciones *fuzzy*, concretamente Fuzzy-ARTMAP. Una de las líneas actuales de trabajo en *Lattice Computing* es el *Fuzzy Lattice Neurocomputing* (FLN) propuesto por

Kaburlasos¹. En esta aproximación se generaliza el Fuzzy-ARTMAP permitiendo tipos de datos heterogéneos. El concepto básico es la integración de toda la información bajo la Teoría de Retículos. Si es posible construir una función de evaluación, entonces es posible definir la distancia entre objetos y la medida de inclusión que permite realizar el proceso de “aprendizaje”.

Los objetos en colecciones multimedia tienen caracterizaciones que pueden ser datos muy heterogéneos (vectores de textura, histogramas, grafos de distribución espacial, etc.). El modelado de estos datos bajo la perspectiva del FLN puede abrir nuevas avenidas de investigación.

¹Kaburlasos, V.G., Towards a Unified Modeling and Knowledge Representation based on Lattice Theory. 2006: Springer Verlag.

Ejemplos Ilustrativos

A.1. Ejemplos ilustrativos sobre características de texturas basados en BFG

Para ilustrar lo expuesto en la sección 5.5.1, podemos ver el efecto de aplicar un BFG a un conjunto de imágenes de texturas. Cada imagen está formada por una señal sinusoidal 2D que produce el efecto visual de una serie de barras en tonos de gris, con frecuencias espaciales de $1/4$, $1/8$, $1/32$ y $1/64$ ciclos/pixel. La orientación de la señal está determinada por ángulos específicos de 0 , 45 , 90 y 135 grados respecto a la horizontal. La figura A.1 muestra dicho conjunto, conformado por 16 imágenes de dimensiones 256×256 píxeles.

Hemos construido un banco de filtros diseñado para que cada filtro dé una respuesta máxima para una imagen específica. Esto es, se utilizaron los parámetros $\theta = 1/4, 1/8, 1/16, 1/32$ ciclos/pixel y $F = 0, 45, 90, 135$ grados, para configurar cada filtro. En la figura A.2 podemos observar la parte real de las respuestas del conjunto de filtros en el dominio de Fourier.

Los resultados obtenidos de aplicar el BFG al conjunto de imágenes se ilustran en las figuras A.3 a A.6. Para cada filtro se tiene un diagrama 2D, con las imágenes ubicadas en el eje de las abscisas, y la respuesta de cada imagen al filtro aplicado en el eje de las ordenadas. En cada gráfico aparecen los parámetros F y θ a los que se ha sintonizado el filtro en cuestión. El orden de las imágenes en el eje de abscisas es el siguiente: $(1/4, 0^\circ)$, $(1/4, 45^\circ)$, $(1/4, 90^\circ)$, $(1/4, 135^\circ)$, $(1/8, 0^\circ)$, $(1/8, 45^\circ)$, $(1/8, 90^\circ)$, $(1/8, 135^\circ)$, $(1/16, 0^\circ)$, $(1/16, 45^\circ)$, $(1/16, 90^\circ)$, $(1/16, 135^\circ)$, $(1/32, 0^\circ)$, $(1/32, 45^\circ)$, $(1/32, 90^\circ)$ y $(1/32, 135^\circ)$, respectivamente.

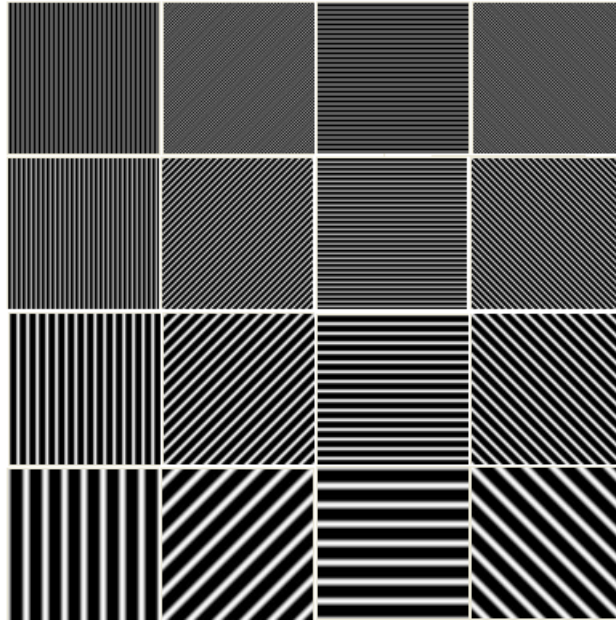


Figura A.1: Imágenes construidas con orientación y frecuencia espacial específicas que han sido definidas para probar la respuesta de cada filtro, las frecuencias en las barras de cada imagen son de $1/4$, $1/8$, $1/16$ y $1/32$ ciclos/pixel, con orientaciones de 0, 45, 90 y 135 grados.

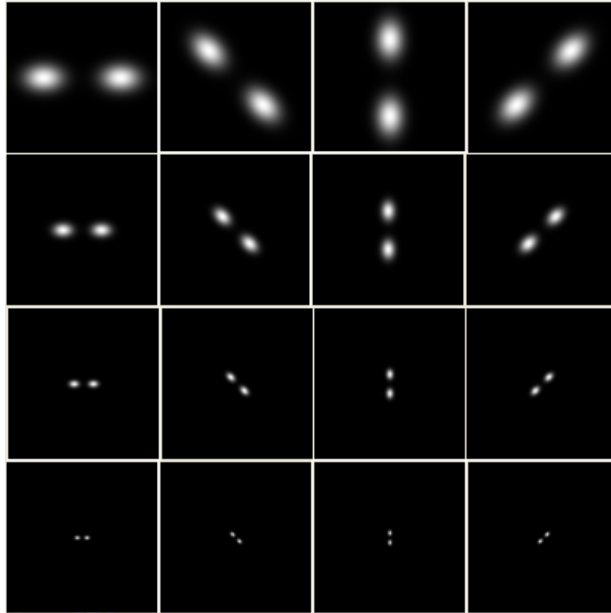


Figura A.2: BFG sintonizado con los parámetros $F = 1/4, 1/8, 1/16, 1/32$ ciclos/pixel, y $\theta = 0^\circ, 45^\circ, 90^\circ, 135^\circ$

Podemos constatar visualmente las siguientes propiedades de las respuestas de los filtros:

- Las imágenes que coinciden en orientación y frecuencia con cada filtro, responden con la más elevada magnitud de la energía.
- Se obtienen respuestas armónicas de menor magnitud para señales que tienen la misma orientación que el filtro, pero menor frecuencia, excepto en el último caso donde se aprecian los armónicos para frecuencias en una octava superior.
- Para señales que tienen distinta orientación que el filtro no hay respuesta.

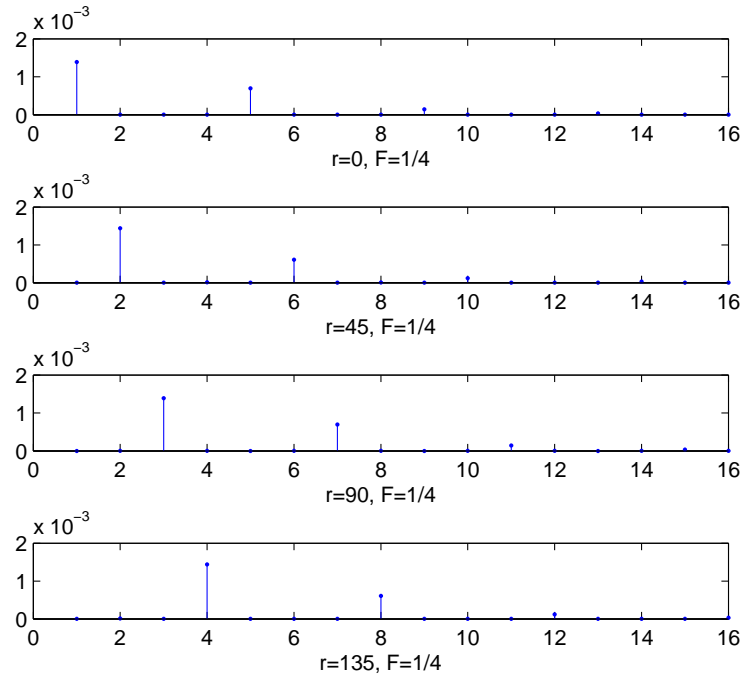


Figura A.3: Respuestas del FG con parámetros $\theta = 0^\circ, 45^\circ, 90^\circ, 135^\circ, F = 1/4$ ciclos/pixel

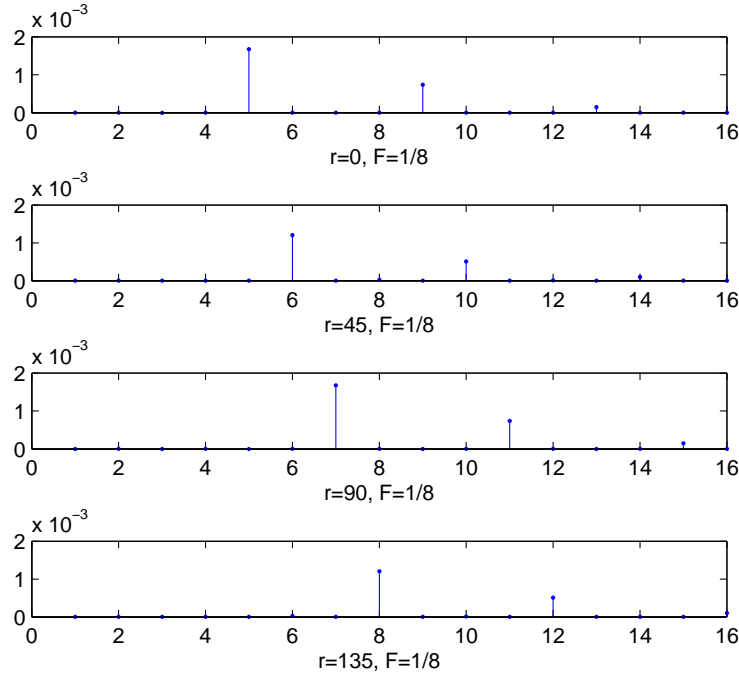


Figura A.4: Respuestas del FG con parámetros $\theta = 0^\circ, 45^\circ, 90^\circ, 135^\circ$, $F = 1/8$ ciclos/píxel.

Hemos diseñado otro experimento ilustrativo con un conjunto de imágenes de textura mallada en la que gradualmente se va aumentando la frecuencia de la malla. Las imágenes se pueden apreciar en la figura A.7.

Hemos aplicado el BFG diseñado para el experimento anterior a este nuevo conjunto de imágenes y hemos calculado para cada imagen el vector de características definido en la expresión 5.2. Calculamos la norma del vector de características para establecer el poder discriminante del BFG en este tipo de imágenes que gradualmente va aumentando los detalles de textura. Los resultados obtenidos se pueden apreciar en la figura A.8. Podemos ver que el valor de la norma se reduce gradualmente cuando aumentan los trazos horizontales y verticales de cada imagen, observándose cierta proporción inversa entre los detalles de textura y la magnitud de la norma del vector de características.

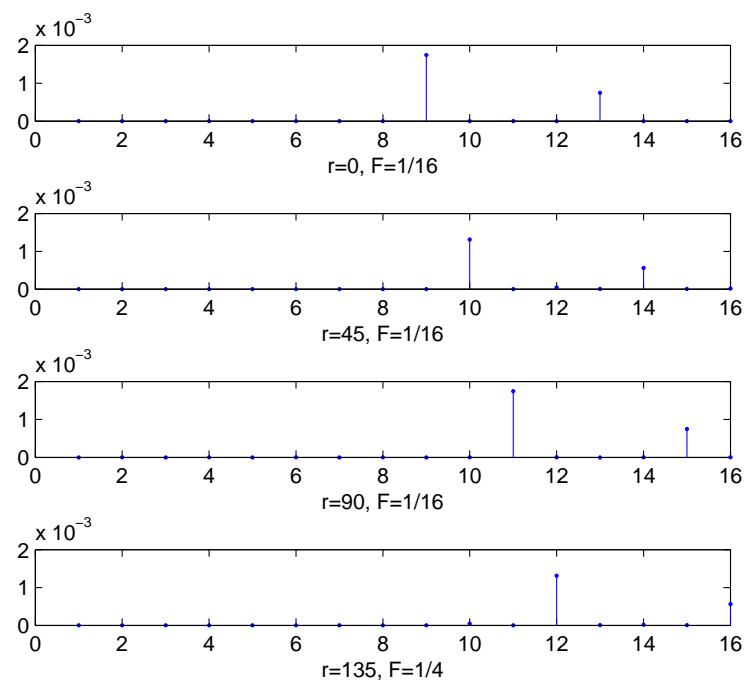


Figura A.5: Respuestas del FG con parámetros $\theta = 0^\circ, 45^\circ, 90^\circ, 135^\circ$, $F = 1/16$ ciclos/pixel.

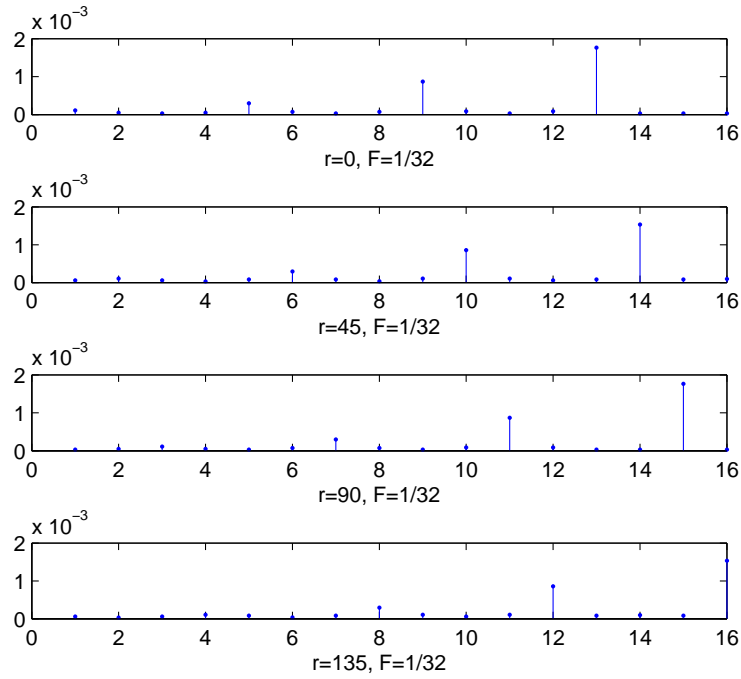


Figura A.6: Respuestas del FG con parámetros $\theta = 0^\circ, 45^\circ, 90^\circ, 135^\circ$, $F = 1/32$ ciclos/pixel.

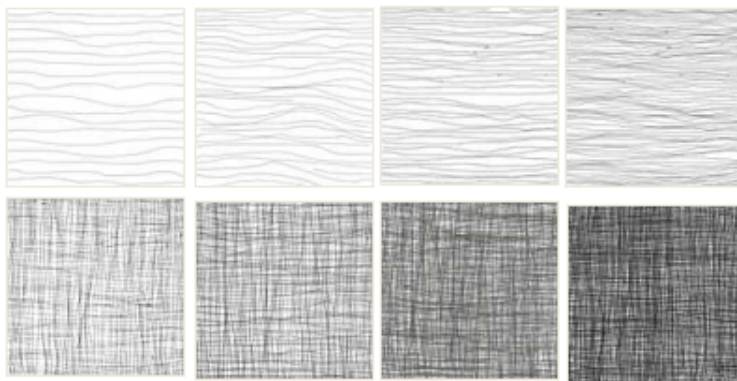


Figura A.7: Imágenes de texturas con incrementos en los detalles y la variación en la frecuencia espacial de los niveles de gris.

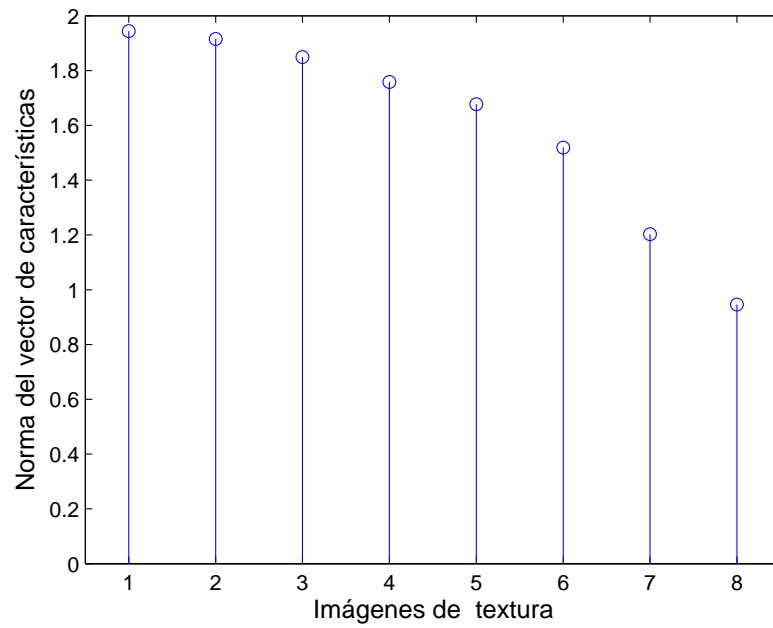


Figura A.8: Norma del vector de características obtenido a partir de BFG para cada una de las imágenes de la figura A.7.

A.2. Ejemplos ilustrativos sobre características de texturas basados en la TWD

Para poder apreciar la sensibilidad de las características de texturas calculadas sobre los coeficientes wavelet, hemos aplicado el algoritmo de descomposición piramidal y el cálculo de las características de texturas a las imágenes de la figura A.1. Los resultados para 5 niveles de descomposición se ilustran en la figuras A.9 a A.13. En cada figura se aprecian tres gráficas, una para cada tipo de detalle. En el eje de las abscisas se ubica cada imagen con el mismo orden usado en el apéndice A.1, el eje de las ordenadas muestra la energía aportada por el detalle correspondiente de cada imagen. En cada figura la gráfica superior muestra la energía de los detalles horizontales, la gráfica central la energía de los detalles diagonales y la gráfica inferior la energía de los detalles verticales. Se puede observar que las imágenes con barras verticales exhiben alta energía sólo en los detalles verticales. Las imágenes que muestran barras horizontales, sólo exhiben energía en los detalles horizontales. Las imágenes que están orientadas a 45 y 135 grados, muestran energía en los tres niveles, pero para cada imagen, la mayor magnitud de la energía se presenta en los detalles diagonales. Cuando aumenta el nivel de descomposición, se obtiene mayor respuesta en las imágenes con menos frecuencia en la señal sinusoidal con las que fueron generadas, mientras que las imágenes con altas frecuencias son menos susceptibles.

Como en el caso de los BFG, hemos diseñado otro experimento ilustrativo con el conjunto de imágenes de textura mallada de la figura A.7.

A estas imágenes se les ha aplicado el algoritmo de descomposición piramidal en un nivel y hemos utilizado los coeficientes wavelet de sus detalles horizontales, verticales y diagonales para calcular el vector de características de cada imagen, definido en la ecuación 5.5, y utilizando la wavelet madre de Haar. Hemos calculado la norma del vector para cada imagen y los resultados están plasmados en la figura A.14. Se observa que la magnitud de la norma del vector de características de textura se incrementa para cada imagen, en proporción a la cantidad de detalles de la misma, o dicho de otra manera cuando aumentan los componentes frecuenciales globales de cada imagen.

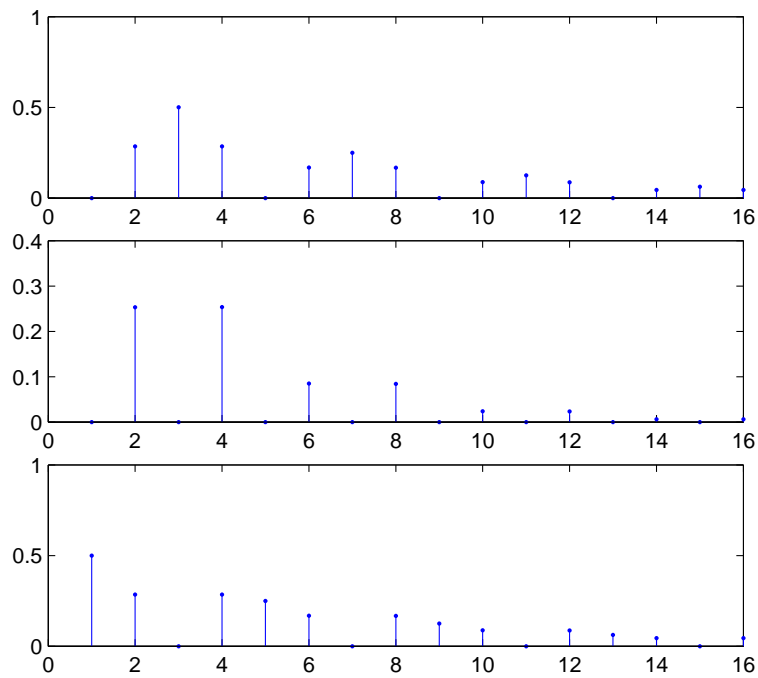


Figura A.9: Energía calculada para el primer nivel de descomposición wavelet.

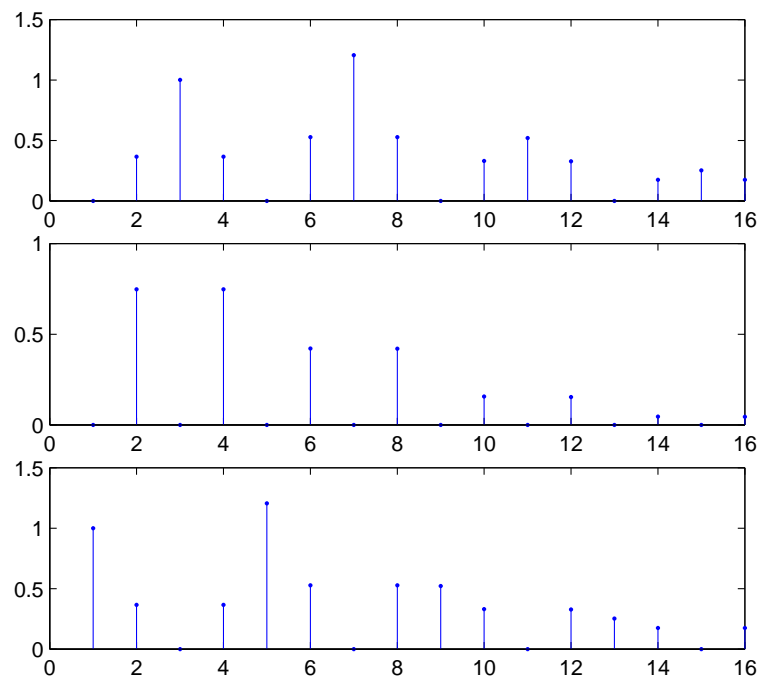


Figura A.10: Energía calculada para el segundo nivel de descomposición wavelet.

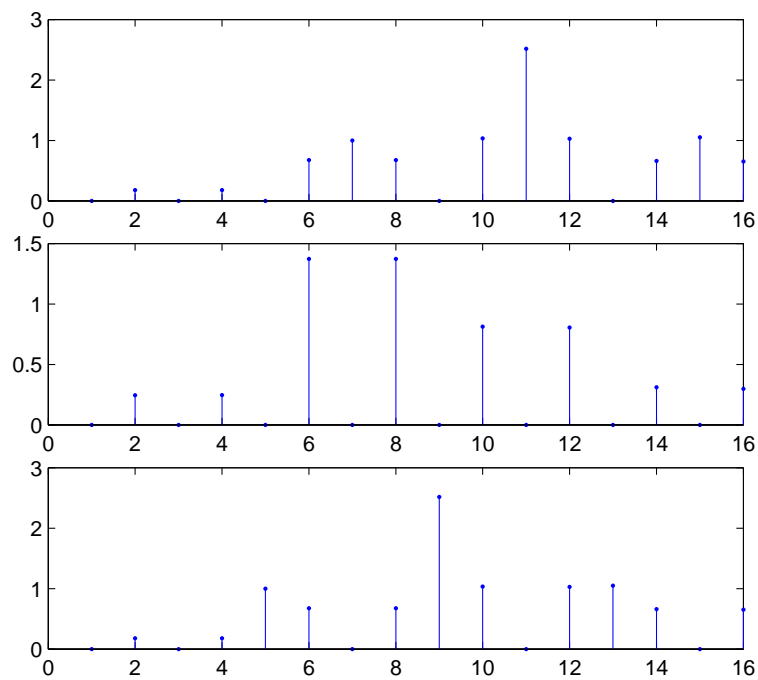


Figura A.11: Energía calculada para el tercer nivel de descomposición wavelet.

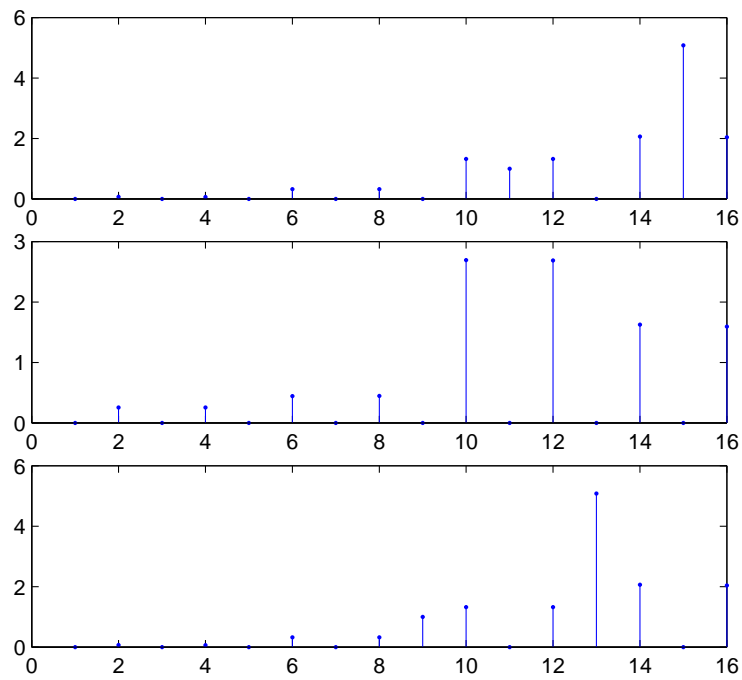


Figura A.12: Energía calculada para el cuarto nivel de descomposición wavelet.

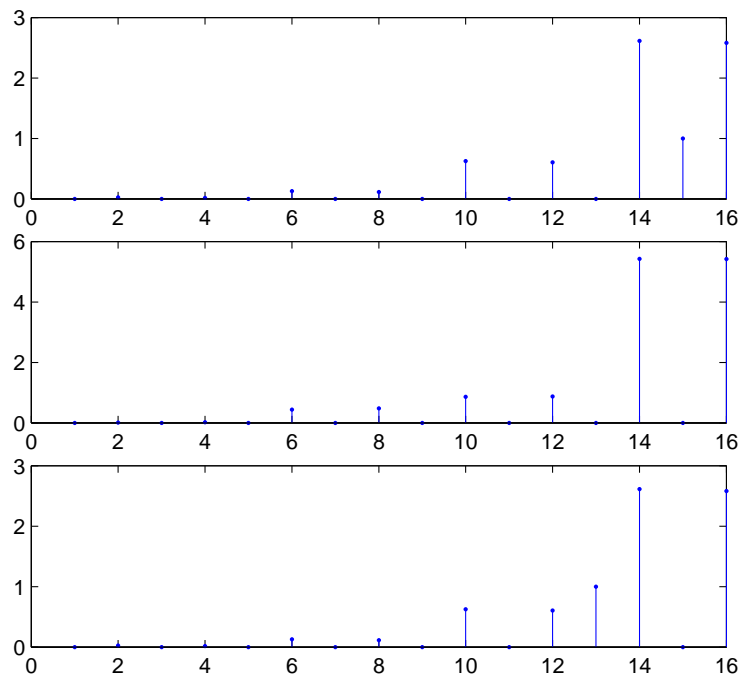


Figura A.13: Energía calculada para el quinto nivel de descomposición wavelet.

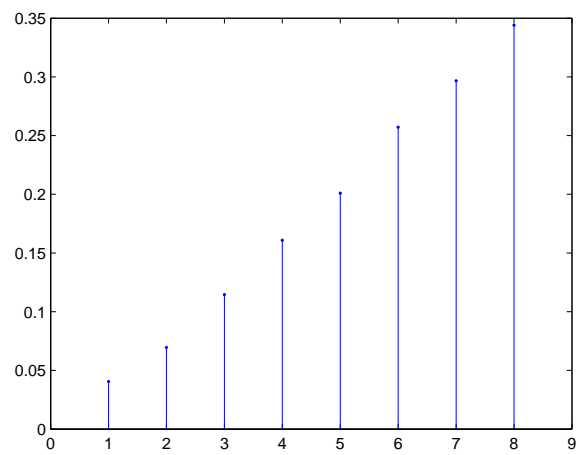


Figura A.14: Norma del vector de características basado en los coeficientes de los detalles de la descomposición wavelet, calculado para las imágenes de la figura A.7

Bibliografia

- [1] Ahmed Abouelelaa, Hazem M. Abbasb, Hesham Eldeeba, Abdelmonem A. Wahdanb, and Salwa M. Nassara. Automated vision system for localizing structural defects in textile fabrics. *Pattern Recognition Letters*, 26:1435–1443, 2004.
- [2] Imran Ahmad and William I. Grosky. Indexing and retrieval of images by spatial constraints. *J. Vis. Commun. Image R.*, 14:291–320, 2003.
- [3] C. Anagnostopoulos, Anagnostopoulos I., Vergados D., Kouzas G., Kayafas E., Loumos V., and Stassinopoulos G. High performance computing algorithms for textile quality control. *Mathematics and Computers in Simulation*, 60:389–400, 2002.
- [4] Ioannis Andreou and Nikitas M. Sgouros. Computing, explaining and visualizing shape similarity in content-based image retrieval. *Information Processing and Management*, 41:1121–1139, 2005.
- [5] D. Androutsos, K. N. Plataniotis, and A. N. Venetsanopoulos. A novel vector-based approach to color image retrieval using a vector angular-based distance measure. *Computer Vision and Image Understanding*, 75:46–58, 1999.
- [6] M. Anelli, L. Cinque, and Enver Sangineto. Deformation tolerant generalized hough transform for sketch-based image retrieval in complex scenes. *Image and Vision Computing*, 25:1802–1813, 2007.
- [7] Sameer Antania, D.J. Leeb, L. Rodney Longa, and George R. Thoma. Evaluation of shape similarity measurement methods for spine x-ray images. *J. Vis. Commun. Image R.*, 15:285–302, 2004.

-
- [8] S. Arivazhagan, L. Ganesan, and S. Padam Priyal. Texture classification using gabor wavelet based rotation invariant features. *Pattern Recogn. Lett.*, 27(9-10):1976–1982, 2006.
- [9] S. Berchtold, D.A. Keim, H.-P. Kriegel, and T. Seidl. Indexing the solution space: a new technique for nearest neighbor search in high-dimensional space. *Knowledge and Data Engineering, IEEE Transactions on*, 12(1):45–57, 2000.
- [10] Stefan Berchtold, Christian Bohm, and Hans-Peter Kriegel. The pyramid-technique: towards breaking the curse of dimensionality. *SIGMOD Rec.*, 27(2):142–153, 1998.
- [11] Stefan Berchtold, Daniel A. Keim, and Hans-Peter Kriegel. The x-tree: An index structure for high-dimensional data. In T. M. Vijayaraman, Alejandro P. Buchmann, C. Mohan, and Nandlal L. Sarda, editors, *Proceedings of the 22nd International Conference on Very Large Databases*, pages 28–39, San Francisco, U.S.A., 1996. Morgan Kaufmann Publishers.
- [12] Manish H. Bharati, J. Jay Liu, and John F. MacGregor. Image texture analysis: methods and comparisons. *Chemometrics and Intelligent Laboratory Systems*, 72:57–71, 2004.
- [13] Irving Biederman. Recognition-by-components: a theory of human image understanding. *Psychological Review*, 94:115–147, 1987.
- [14] C. Blakemore and F. W. Campbell. On the existence of neurons in the human visual system selectively sensitive to the orientation and size of the retinal image. *Journal of Psychology*, 203:237–260, 1969.
- [15] J.W. Boardman. Analysis, understanding and visualization of hyper-spectral data as convex sets in n-space. In *Imaging Spectrometry 1995*, volume Proc. SPIE 2480, pages 14–22. SPIE, 1995.
- [16] A. Bodnarova, M. Bennamoun, and S. Latham. Optimal gabor filters for textile flaw detection. *Pattern Recognition*, 35:2973–2991, 2002.
- [17] Alan Conrad Bovik, Marianna Clark, and Wilson Geisler. Multichannel texture analysis using localized spatial filters. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 12:55–73, 1990.

-
- [18] P. Calderon-Martinez, J.A.; Campoy-Cervera. A convolutional neural architecture: an application for defects detection in continuous manufacturing systems. *Circuits and Systems, 2003. ISCAS '03. Proceedings of the 2003 International Symposium on*, 5:V-749-V-752 vol.5, 25-28 May 2003.
- [19] F. W. Campbell and J. G. Robson. Application of fourier analysis to the visibility of gratings. *J. Physiol.*, 197:551-556, 1968.
- [20] C. Carson, S. Belongie, H. Greenspan, and J. Malik. Blobworld: image segmentation using expectation-maximization and its application to image querying. *Transactions on Pattern Analysis and Machine Intelligence*, 24(8):1026-1038, Aug 2002.
- [21] S.K. Chang, Q.Y. Shi, and C.W. Yan. Iconic indexing by 2d strings. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 9:413-428, 1987.
- [22] R. Chellappa and S. Chatterjee. Classification of textures using markov random field models. *Acoustics, Speech, and Signal Processing*, 33:959-963, 1985.
- [23] Yixin Chen, J.Z. Wang, and R. Krovetz. Clue: cluster-based retrieval of images by unsupervised learning. *Image Processing, IEEE Transactions on*, 14(8):1187-1201, Aug. 2005.
- [24] Pei-Cheng Chenga, Been-Chian Chienb, Hao-Ren Kec, and Wei-Pang Yangd. A two-level relevance feedback mechanism for image retrieval. *Expert Systems with Applications*, 34:2193-2200, 2008.
- [25] G.K.H. Chi-Ho Chan; Pang. Fabric defect detection by fourier analysis. *Industry Applications, IEEE Transactions on*, 36(5):1267-1276, Sep/Oct 2000.
- [26] Maurizio Cibelli, Michele Nappi, and Maurizio Tucci. Abi: analogy-based indexing for content image retrieval. *Image and Vision Computing*, 22:23-34, 2004.
- [27] G. Ciocca, J. Gagliardi, and R. Schettini. Quicklook: An integrated multimedia system. journal of visual languages and computing. *Journal of Visual Languages and Computing*, 12:81-103, 2001.

-
- [28] G. Ciocca and R. Schettini. A relevance feedback mechanism for content-based image retrieval. *Information Processing and Management*, 35:605–632, 1999.
- [29] David A. Clausi and M. Ed. Jernigan. Designing gabor filters for optimal texture separability. *Pattern Recognition*, 33:1835–1849, 2000.
- [30] Aura Concia and Claudia Belmiro Proença. A fractal image analysis system for fabric inspection based on a box-counting method. *Computer Networks and ISDN Systems*, 30:1887–1895, 1999.
- [31] J.M. Considine, C.T. Scott, R. Gleisner, and J.Y. Zhu. Use of digital image correlation to study the local deformation field of paper and paperboard. In *13th Fundamental Research Symposium Conference*, pages 613–630, 2005.
- [32] G. Costagliola, F. Ferrucci, G. Tortora, and M. Tucci. Non-redundant 2d strings. *Knowledge and Data Engineering, IEEE Transactions on*, 7(2):347–350, Apr 1995.
- [33] I.J. Cox, M.L. Miller, T.P. Minka, T.V. Papathomas, and P.N. Yianilos. The bayesian image retrieval system, pichunter: theory, implementation, and psychophysical experiments. *Image Processing, IEEE Transactions on*, 9(1):20–37, Jan 2000.
- [34] M. Craig. Minimum volume transformations for remotely sensed data. *IEEE Trans. Geos. Rem. Sensing*, 32(3):542–552.
- [35] M. Das and E.M. Riseman. Focus: A system for searching for multi-colored objects in a diverse image database. *Computer Vision and Image Understanding*, 94:168–192, 2004.
- [36] Ritendra Datta, Dhiraj Joshi, Jia Li, and James Z. Wang. Image retrieval: Ideas, influences, and trends of the new age. *ACM Computing Surveys*, 2008.
- [37] John G. Daugman. Uncertainty relation for resolution in space, spatial frequency, and orientation optimized by two-dimensional visual cortical filters. *J. Opt. Soc. Am. A.*, 2:1160–1169, 1985.

-
- [38] A. Del Bimbo and P. Pala. Visual image retrieval by elastic matching of user sketches. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 19(2):121–132, Feb 1997.
- [39] Anlei Dong and B. Bhanu. Active concept learning in image databases. *Systems, Man, and Cybernetics, Part B, IEEE Transactions on*, 35(3):450–466, June 2005.
- [40] R. O. Duda, P. E. Hart, and D. G. Stork. *Pattern Classification*. Wiley Interscience, 2001.
- [41] Dennis Dunn and William E. Higgins. Optimal gabor filters for texture segmentation. *IEEE Transactions on Image Processing*, 4:947–961, 1995.
- [42] Dennis Dunn, William E. Higgins, and Joseph Wakeley. Texture segmentation using 2-d gabor elementary functions. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 16:130–149, 1995.
- [43] John Eakins and Margaret Graham. Content-based image retrieval. Technical report, University of Northumbria at Newcastle, 1999.
- [44] I. El-Naqa, Yongyi Yang, N.P. Galatsanos, R.M. Nishikawa, and M.N. Wernick. A similarity learning approach to content-based image retrieval: application to digital mammography. *Medical Imaging, IEEE Transactions on*, 23(10):1233–1244, Oct. 2004.
- [45] Eyas El-Qawasmeh. A quadtree-based representation technique for indexing and retrieval of image databases. *Journal of Visual Communication and Image Representation*, 14:340–357, 2003.
- [46] J. Jr.; Panchanathan S. Fahmy, G.; Black. Texture characterization for joint compression and classification based on human perception in the wavelet domain. *Image Processing, IEEE Transactions on*, 15(6):1389–1396, June 2006.
- [47] Julien Fauqueur and Nozha Boujema. Region-based image retrieval: fast coarse segmentation and fine color description. *Journal of Visual Languages and Computing*, 15:69–95, 2004.

-
- [48] M. Flickner, W. Sawhney, H. ND Niblack, J. Ashley, Qian Huang, B. Dom, M. Gorkani, J. Hafner, D. Lee, D. Petkovic, D. Steele, and P. Yanker. Query by image and video content: the qbic system. *Computer*, 28(9):23–32, Sep 1995.
- [49] M.J. Fonseca and J.A. Jorge. Indexing high-dimensional data for content-based retrieval in large databases. *Database Systems for Advanced Applications, 2003. (DASFAA 2003). Proceedings. Eighth International Conference on*, pages 267–274, 2003.
- [50] J. W. Funck, Y. Zhong, D. A. Butler, C. C. Brunner, and J. B. Forrer. Image segmentation algorithms applied to wood defect detection. *Computers and Electronics in Agriculture*, 41:157–179, 2003.
- [51] D. Gabor. Theory of communication. *J. Inst. Electr. Eng.*, 93:429–457, 1946.
- [52] Volker Gaede and Oliver Gunther. Multidimensional access methods. *ACM Comput. Surv.*, 30(2):170–231, 1998.
- [53] C. Garcia and G. Tziritas. Face detection using quantized skin color regions merging and wavelet packet analysis. *Multimedia, IEEE Transactions on*, 1(3):264–277, Sep 1999.
- [54] Theo Gevers and Arnold W.M. Smeulders. Content-based image retrieval by viewpoint-invariant color indexing. *Image and Vision Computing*, 17:475–488, 1999.
- [55] Shen-Tat Goh and Kian-Lee Tan. Mosaic: A fast multi-feature image retrieval system. *Data & Knowledge Engineering*, 33:219–239, 2000.
- [56] M. Graña. State of the art in lattice computing for artificial intelligence applications. In R. Nadarajan, R. Anitha, and C. Porkodi, editors, *Mathematical and Computational Models*, pages 233–242. Narosa Publishing House, 2007.
- [57] M. Graña and A. d’Anjou. Feature extraction by linear spectral unmixing. In M. Negoita, R.J. Howlett, and L.C. Jain, editors, *Knowledge-Based Intelligent Information and Engineering Systems*, volume LNAI 3213 of *LNAI*, pages 692–697. Springer Verlag, 2004.

-
- [58] M. Graña, A. d'Anjou, and F.X. Albizuri. Morphological memories for feature extraction in hyperspectral images. In M. Verleysen, editor, *ESANN 2005*, pages 497–502, 2005.
- [59] M. Graña and J. Gallego. Associative morphological memories for endmember induction. In *Proc. IGARSS'2003*, volume 6, pages 3757–3759, 2003.
- [60] M. Graña, P. Sussner, and G.X. Ritter. Associative morphological memories for endmember determination in spectral unmixing. In *Proc. FUZZ-IEEE'03*, volume 2, pages 1285–1290, 2003.
- [61] M. Graña, I. Villaverde, R. Moreno, and F.X. Albizuri. Convex coordinates from lattice independent sets for visual pattern recognition. In V.G. Karbulasos and G.X. Ritter, editors, *Computational Intelligence Based on Lattice Theory*, pages 99–126. Springer Verlag, 2007.
- [62] Simona E. Grigorescu, Nicolai Petkov, and Peter Kruizinga. Comparison of texture features based on gabor filters. *Image Processing, IEEE Transactions on*, 11:1160–1167, 2002.
- [63] Antonin Guttman. R-trees: a dynamic index structure for spatial searching. In *SIGMOD '84: Proceedings of the 1984 ACM SIGMOD international conference on Management of data*, pages 47–57, New York, NY, USA, 1984. ACM.
- [64] Arun Hampapur, Amarnath Gupta, Bradley Horowitz, Chiao-Fe Shu, Charles Fuller, Jeffrey R. Bach, Monika Gorkani, and Ramesh C. Jain. Virage video engine. volume 3022, pages 188–198. SPIE, 1997.
- [65] R. M. Haralick, K. Shanmugam, and I. Dinstein. Texture features for image classification. *IEEE Trans. on Syst. Man Cybern*, 3:610–621, 1973.
- [66] Chu-Hong Hoi and M.R. Lyu. Group-based relevance feedback with support vector machine ensembles. *Pattern Recognition, 2004. ICPR 2004. Proceedings of the 17th International Conference on*, 3:23–26, 2004.

-
- [67] R. Holmstad, C. Antoine, J. Silvy, A. P. Costa, and J. Antoine. Modelling the paper sheet structure according to the equivalent pore concept. *In Action E11 Final workshop Conference. Espoo, Finland*, pages 1–14, 2001.
- [68] Pengyu Hong, Qi Tian i, and T.S. Huang. Incorporate support vector machines to content-based imageretrieval with relevance feedback. *Image Processing, 2000. Proceedings. 2000 International Conference on*, 3:750–753, 2000.
- [69] J.J. Hopfield. Neural networks and physical systems with emergent collective computational abilities. *Proc. Nat. Acad. Sciences*, 79:2554–2558, 1982.
- [70] Ming-Kuei Hu. Visual pattern recognition by moment invariants. *Information Theory, IEEE Transactions on*, 8:179–187, 1962.
- [71] Ke Huang and Selin Aviyente. Information-theoretic wavelet packet subband selection for texture classification. *Signal Processing*, 86:1410–1420, 2006.
- [72] P.W. Huang and S.K. Dai. Design of a two-stage content-based image retrieval system using texture similarity. *Information Processing and Management*, 40:81–96, 2004.
- [73] P.W. Huang and Y.R. Jean. Using 2d c+-string as spatial knowledge representation for image database systems. *Pattern Recognition*, 27:1249–1257, 1994.
- [74] Thomas S. Huang and Xiang Sean Zhou. Image retrieval with relevance feedback: From heuristic weight adjustment to optimal learning methods. *Image Processing, 2001. Proceedings. 2001 International Conference on*, 3:2–5, 2001.
- [75] D. G. Hubel and T. N. Wiesel. Receptieve fields, binocular interaction, and functional architecture in the cat’s visual cortex. *J. Physiol*, 160:106–154, 1962.
- [76] Wey-Shiuan Hwang, J.J. Weng, Ming Fang, and Jianzhong Qian. A fast image retrieval algorithm with automatically extracted discriminant

- features. In *Content-Based Access of Image and Video Libraries, 1999. (CBAIVL '99) Proceedings. IEEE Workshop on*, 1999.
- [77] Khalid Idrissi, Guillaume Lavoué, Julien Ricard, and Atilla Baskurt. Object of interest-based visual navigation, retrieval, and semantic content identification system. *Computer Vision and Image Understanding*, 94:271–294, 2004.
- [78] A. Ifarraguerri and C.I. Chang. Multispectral and hyperspectral image analysis with convex cones. *IEEE Trans. Geos. Rem. Sensing*, 37(2):756–770, 1999.
- [79] Anil K. Jain and Aditya Vailaya. Image retrieval using color and shape. *Pattern Recognition*, 29:1233–1244, 1996.
- [80] Gerald Kaiser. *A Friendly Guide to Wavelets*. Birkhäuser, 1994.
- [81] N Keshava and Mustard J.F. Spectral unmixing. *IEEE Signal Processing Magazine*, 19(1):44–57, 2002.
- [82] J.H. Klungness, A. Ahmed, N. Ross-Sutherland, and S. AbuBakr. Lightweight, high-opacity paper by fiber loading: filler comparison. *Nordic Pulp and Paper Research Journal*, 15:345–350, 2000.
- [83] M. Kokare, P.K. Biswas, and B.N. Chatterji. Texture image retrieval using new rotated complex wavelet filters. *Systems, Man, and Cybernetics, Part B, IEEE Transactions on*, 35(6):1168–1178, Dec. 2005.
- [84] B. Kozintsev. *Computations with gaussian random fields*, 1999.
- [85] R. Krishnapuram, S. Medasani, Sung-Hwan Jung, Young-Sik Choi, and R. Balasubramaniam. Content-based image retrieval based on a fuzzy approach. *Transactions on Knowledge and Data Engineering*, 16(10):1185–1199, Oct. 2004.
- [86] G.K.H. Kumar, A.; Pang. Defect detection in textured materials using gabor filters. *Industry Applications, IEEE Transactions on*, 38(2):425–440, Mar/Apr 2002.
- [87] D.A. Landgrebe. *Signal theory methods in multispectral remote sensing*. John Wiley & Sons, Hoboken, NJ, 2003.

-
- [88] K.I. Laws. Rapid texture identification. In *Proc. SPIE Conf. Image Processing for Missile Guidance*, 1980.
- [89] D. Lee and H.-J. Kim. A fast content-based indexing and retrieval technique by the shape information in large image database. *The Journal of Systems and Software*, 56:165–182, 2001.
- [90] Dong-Ho Lee and Hyoung-Joo Kim. Spy-tec: An efficient indexing method for similarity search in high-dimensional data spaces. *Data & Knowledge Engineering*, 34:77–97, 2000.
- [91] Hyoung Ku LEE and Suk In YOO. Intelligent image retrieval using neural network. *IEICE transactions on information and systems*, 84(12):1810–1819, 2001.
- [92] S.Y. Lee and F.J. Hsu. and 2d c-string: A new spatial knowledge representation for image database systems. *Pattern Recognition*, 23:1077–1087, 1999.
- [93] Bing-Cheng Li and Jun Shen. Fast computation of moment invariants. *Pattern Recogn.*, 24(8):807–813, 1991.
- [94] Hongyu Li, Rongjie Shi, Wenbin Chen, and I-Fan Shen. Image tangent space for image retrieval. *Pattern Recognition, 2006. ICPR 2006. 18th International Conference on*, 2:1126–1130, 2006.
- [95] Shutao Li and John Shawe-Taylor. Comparison and fusion of multiresolution features for texture classification. *Pattern Recognition Letters*, 26(5):633–638, 2005.
- [96] Stan Z. Li, Kap Luk Chan, and Changliang Wang. Performance evaluation of the nearest feature line method in image classification and retrieval. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 22:1335–1339, 2000.
- [97] King-Ip Lin, H. V. Jagadish, and Christos Faloutsos. The tv-tree: An index structure for high-dimensional data. *VLDB Journal: Very Large Data Bases*, 3(4):517–542, 1994.
- [98] Ying Liua, Dengsheng Zhanga, Guojun Lua, and Wei-Ying Ma. Asurvey of content-based image retrieval with high-level semantics. *Pattern Recognition*, 40:262–282, 2007.

-
- [99] Wei-Ying Ma and B. S. Manjunath. Netra: A toolbox for navigating large image databases. *Multimedia Systems*, 7(3):184–198, 1999.
- [100] Stephane G. Mallat. Multifrequency channel decompositions of images and wavelet models. *IEEE transactions on acoustics, speech and signal processing*, 37:2091–20110, 1989.
- [101] Stephane G. Mallat. A theory for multiresolution signal decomposition: The wavelet representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2:674–693, 1989.
- [102] Stephane G. Mallat. Wavelets for vision. In *Proceeding of the IEEE*, volume 4, pages 604–614, April 1996.
- [103] B. S. Manjunath and W. Y. Ma. Texture features for browsing and retrieval of image data. *IEEE Trans. Pattern Anal. Mach. Intell.*, 18(8):837–842, 1996.
- [104] D. Manolakis and G. Shaw. Detection algorithms for hyperspectral imaging applications. *IEEE Signal Proc. Mag.*, 19(1):29–43, 2002.
- [105] J. Martinez-Alajarin, J.D. Luis-Delgado, and L.M. Tomas-Balibrea. Automatic system for quality-based classification of marble textures. *Systems, Man and Cybernetics, Part C: Applications and Reviews, IEEE Transactions on*, 35(4):488–497, Nov. 2005.
- [106] Ruggero Milanese and Michel Cherbuliez. A rotation, translation, and scale-invariant approach to content-based image retrieval. *Journal of Visual Communication and Image Representation*, 10:186–196, 1999.
- [107] Henning Müller, Nicolas Michoux, David Bandon, and Antoine Geissbuhler. A review of content-based image retrieval systems in medical applications-clinical benefits and future directions. *International Journal of Medical Informatics*, 73:1–23, 2004.
- [108] Henning Müller, Wolfgang Müller, David McG. Squire, Stephane Marchand-Maillet, and Thierry Pun. Performance evaluation in content-based image retrieval: overview and proposals. *Pattern Recogn. Lett.*, 22(5):593–601, 2001.

-
- [109] Henning Müller, Antoine Rosset, Jean-Paul Vallee, Francois Terrier, and Antoine Geissbuhler. A reference data set for the evaluation of medical image retrieval systems. *Computerized Medical Imaging and Graphics*, 28:295–305, 2004.
- [110] P. Muneesawang and Ling Guan. Automatic machine interactions for content-based image retrieval using a self-organizing tree map architecture. *Neural Networks, IEEE Transactions on*, 13(4):821–834, Jul 2002.
- [111] P. Muneesawang and Ling Guan. An interactive approach for cbir using a network of radial basis functions. *Multimedia, IEEE Transactions on*, 6(5):703–716, Oct. 2004.
- [112] H. Nezamabadi-pour and E. Kabir. Image retrieval using histograms of uni-color and bi-color blocks and directional changes in intensity gradient. *Pattern Recognition Letters*, 14:1547–1557, 2004.
- [113] Henry Y.T. Ngana, Grantham K.H. Panga, S.P. Yungb, and Michael K. Ngb. Wavelet based methods on patterned fabric defect detection. *Pattern Recognition*, 38:559–576, 2005.
- [114] Hirobumi Nishida. Structural feature indexing for retrieval of partially visible shapes. *Pattern Recognition*, 35:55–67, 2002.
- [115] V.E. Ogle and M. Stonebraker. Chabot: retrieval from a relational database of images. *Computer*, 28(9):40–48, 1995.
- [116] Y. Ohta and T. Kanade. Stereo by intra- and inter-scanline search using dynamic programming. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 7:139–154, 1985.
- [117] Gonzalo Pajares and Jesus M. de la Cruz. *Vision por Computador: Imágenes digitales y aplicaciones*. Ra-Ma, 2001.
- [118] G. Paschos, I. Radev, and N. Prabakar. Image content-based retrieval using chromaticity moments. *Knowledge and Data Engineering, IEEE Transactions on*, 15(5):1069–1072, Sept.-Oct. 2003.
- [119] A. Pentland, R.W. Picard, and S. Sclaroff. Photobook: Content-based manipulation of image databases. *Int. J. Comput. Vis.*, 3:233–254, 1996.

-
- [120] A. P. Pentland. Fractal-based description of natural scenes. *IEEE Trans. Pattern Anal. Machine Intell*, 6:661–674, 1984.
- [121] A. Plaza, P. Martinez, R. Perez, and J. Plaza. Spatial spatial/spectral endmember extraction by multidimensional morphological operations. *IEEE Trans. Geoscience Remote Sensing*, 40(9):2025–2041, 2002.
- [122] B.G. Prasad, K.K. Biswas, , and S.K. Gupta. Region-based image retrieval using integrated color, shape, and location index. *Computer Vision and Image Understanding*, 94:193–233, 2004.
- [123] B. Raducanu, M. Graña, and X. Albizuri. Morphological scale spaces and associative morphological memories: results on robustness and practical applications. *J. Math. Imaging and Vision*, 19(2):113–122, 2003.
- [124] R.S. Rand and D.M.Keenan. A spectral mixture process conditioned by gibbs-based partitioning. *IEEE Trans. Geos. Rem. Sensing*, 39(7):1421–1434, 2001.
- [125] Howard L. Resnicoff and Raymond O. Wells. *Wavelet Analysis: The Scalable Structure of Information*. Springer, 1998.
- [126] G. X. Ritter, J. L. Diaz de Leon, and P. Sussner. Morphological bidirectional associative memories. *Neural Networks*, 12:851–867, 1999.
- [127] G. X. Ritter, P. Sussner, and J. L. Diaz de Leon. Morphological associative memories. *IEEE Trans. on Neural Networks*, 9(2):281–292, 1998.
- [128] G. X. Ritter, G. Urcid, and M.S. Schmalz. Autonomous single-pass endmember approximation using lattice auto-associative memories. *Neurocomputing*, page in press, 2008.
- [129] G.X. Ritter and P. Gader. Fixed points of lattice transforms and lattice associative memories. In P. Hawkes, editor, *Advances in Imaging and Electron Physics*, volume 144, pages 165–242. Elsevier, 2006.
- [130] G.X. Ritter and G. Urcid. Lattice algebra approach to single-neuron computation. *IEEE Trans Neural Networks*, 14(2):282–295, 2003.

-
- [131] G.X. Ritter, G. Urcid, and L. Iancu. Reconstruction of patterns from moisy inputs using morphological associative memories. *J. Math. Imaging and Vision*, 19(2):95–112, 2003.
- [132] Yong Rui and Thomas S. Huang. Image retrieval: Current techniques, promising directions, and open issues. *Journal of Visual Communication and Image Representation*, 10:39–62, 1999.
- [133] John C Russ. *The Image Processing Handbook, 5th Edition*. CRC Press, 2002.
- [134] A. Sadonikov, P. Salmela, L. Lensu, J.-K. Kamarainen, and H. Kälviäinen. Mottling assessment of solid printed areas and its correlation to perceived uniformity. In *In Proc. of the 14th Scandinavian Conf. of Image Processing (Joensuu, Finland)*, 1995.
- [135] S. Santini and R. Jain. Similarity measures. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 21(9):871–883, Sep 1999.
- [136] J.S. Jr Sari-Sarraf, H.; Goddard. Vision system for on-loom fabric inspection. *Industry Applications, IEEE Transactions on*, 35(6):1252–1259, Nov/Dec 1999.
- [137] A. Sarkar, K.M.S. Sharma, and R.V. Sonak. A new approach for subset 2-d ar model identification for describing textures. *Image Processing, IEEE Transactions on*, 6:407–413, 1997.
- [138] Ch.S. Sastry, Arun K. Pujari, B.L. Deekshatulu, and C. Bhagvati. A wavelet based multiresolution algorithm for rotation invariant feature extraction. *Pattern Recognition Letters*, 25:1845–1855, 2004.
- [139] J. Scharcanski. Stochastic texture analysis for measuring sheet formation variability in the industry. *Instrumentation and Measurement, IEEE Transactions on*, 55(5):1778–1785, Oct. 2006.
- [140] Timos K. Sellis, Nick Roussopoulos, and Christos Faloutsos. The r-tree: A dynamic index for multi-dimensional objects. In *The VLDB Journal*, pages 507–518, 1987.
- [141] G. Shaw and D. Manolakis. Signal processing for hyperespectral image exploitation. *IEEE Signal Processing Magazine*, 19(1):12–16, 2002.

-
- [142] A.W.M. Smeulders, M. Worring, S. Santini, A. Gupta, and R. Jain. Content-based image retrieval at the end of the early years. *Transactions on Pattern Analysis and Machine Intelligence*, 22(12):1349–1380, Dec 2000.
- [143] David McG. Squire, Wolfgang Müllera, Henning Müllera, and Thierry Puna. Content-based query of image databases: inspirations from text retrieval. *Pattern Recognition Letters*, 21:1193–1198, 2003.
- [144] Daniela Stan and Ishwar K. Sethi. eid: a system for exploration of image databases. *Information Processing & Management*, 39:335–361, 2003.
- [145] Gilbert Strang and Truong Nguyen. *Wavelets and Filters Banks*. Wellesley-Cambridge Press, 1997.
- [146] Markus A. Stricker and Markus Orengo. Similarity of color images. In *Storage and Retrieval for Image and Video Databases (SPIE)*, pages 381–392, 1995.
- [147] P. Sussner. Observations on morphological associative memories and the kernel method. In *Proc. IJCNN'2001*. IEEE, 2001.
- [148] M.J. Swain and D.H. Ballard. Color indexing. *International Journal of Computer Vision*, 7:11–32, 1991.
- [149] S. Tadjudin and D. Landgrebe. Classification of high dimensional data with limited training samples. Technical Report TR-ECE 98-8, School of Electrical & Computer Engineering Technical Report, Purdue University, 1998.
- [150] S. Tadjudin and D. Landgrebe. Robust parameter estimation for mixture model. *IEEE Trans. Geos. Rem. Sensing*, 38(1):439–445, 1999.
- [151] S. Tadjudin and D. Landgrebe. Robust parameter estimation for mixture model. *IEEE Trans. Geos. Rem. Sensing*, 38(1):439–445, 2000.
- [152] Anthony Teolis. *Computational Signal Processing with Wavelets*. Birkhäuser, 1998.

-
- [153] Andreas Teuner, Olaf Pichler, and Bedrich J. Hosticka. Unsupervised texture segmentation of images using tuned matched gabor filters. *IEEE Transactions on Image Processing*, 4(6):863–870, 1995.
- [154] C. de Trazegnies, C. Urdiales, A. Bandera, and F. Sandoval. Planar shape indexing and retrieval based on hidden markov models. *Pattern Recognition Letters*, 23:1143–1151, 2002.
- [155] Du-Ming Tsai and Bo Hsiao. Automatic surface inspection using wavelet reconstruction. *Pattern Recognition*, 34:1285–1305, 2001.
- [156] M. R. Turner. Texture discrimination by gabor functions. *Biol. Cybern.*, 55:71–82, 1986.
- [157] Michael Unser. Texture classification and segmentation using wavelet frames. *Image Processing, IEEE Transactions on*, 4:1549–1560, 1995.
- [158] Jeroen Vendrig, Marcel Worring, and Arnold W. M. Smeulders. Filter image browsing: Interactive image retrieval by using database overviews. *Multimedia Tools Applications*, 15:83–103, 2001.
- [159] Ying-Hong Wang. Image indexing and similarity retrieval based on spatial relationship model. *Information Sciences*, 154:39–58, 2003.
- [160] Thomas P. Weldon, William E. Higgins, and Dennis F. Dunn. Gabor filter design for multiple texture segmentation. *Optical Engineering*, 35(10):2852–2863, 1996.
- [161] D. G. Hubel and T. N. Wiesel. Sequence regularity and geometry of orientation columns in the monkey striate cortex. *J. Comp. Neurol.*, 158:267–293, 1974.
- [162] Roland Wilson and Michael Spamm. Finite prolate spheroidal sequences and their applications. ii. image feature description and segmentation. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 10:193–203, 1998.
- [163] M.E. Winter. An algorithm for fast autonomous spectral endmember determination in hyperspectral analysis,. In *Imaging Spectrometry*, volume 3753 of *Proc. SPIE*, pages 266–275. SPIE, 1999.

-
- [164] Ian H. Witten and Eibe Frank. *Data Mining: Practical machine learning tools and techniques", 2nd Edition*. 2005.
- [165] Wai-Tak Wong, Frank Y. Shih, and Jung Liu. Shape-based image retrieval using support vector machines, fourier descriptors and self-organizing maps. *Information Sciences*, 177:1878–1891, 2007.
- [166] P. Wu, B.S. Manjunath, S. Newsam, and H.D. Shin. A texture descriptor for browsing and similarity retrieval. *Signal Processing: Image Communication*, 16:33–46, 2000.
- [167] Raj Bahadur Yadav, Naveen K. Nishchal, Arun K. Gupta, and Vinod K. Rastogi. Retrieval and classification of shape-based objects using fourier, generic fourier, and wavelet-fourier descriptors technique: A comparative study. *Optics and Lasers in Engineering*, 45:695–708, 2007.
- [168] Hun-Woo Yoo, Dong-Sik Jang, Seh-Hwan Jung, Jin-Hyung Park, and Kwang-Seop Song. Visual information retrieval system via content-based approach. *Pattern Recognition*, 35:749–769, 2002.
- [169] Hun-Woo Yoo, Dong-Sik Janga, Seh-Hwan Junga, Jin-Hyung Parka, and Kwang-Seop Songb. Visual information retrieval system via content-based approach. *Pattern Recognition*, 35:749–769, 2002.
- [170] Hun-Woo Yoo, She-Hwan Jung, Dong-Sik Jang, and Yoon-Kyoon Na. Extraction of major object features using vq clustering for content-based image retrieval. *Pattern Recognition*, 35:1115–1126, 2002.
- [171] Dengsheng Zhang and Guojun Lu. Study and evaluation of different fourier methods for image retrieval. *Image and Vision Computing*, 23:33–49, 2005.
- [172] Ruofei Zhang and Zhongfei (Mark) Zhang. Balas: Empirical bayesian learning in the relevance feedback for image retrieval. *Image and Vision Computing*, 24:211–223, 2006.
- [173] Yu Zhong and Anil K. Jain. Object localization using color, texture and shape. *Pattern Recognition*, 33:671–684, 2000.