



Universidad
de Alcalá

ESCUELA POLITÉCNICA SUPERIOR
Tesis Doctoral

**Caracterización automática
de especies de madera mediante
técnicas de clasificación de imágenes**

Autor: Jesús Alpuente Herмосilla

Directores: Dr. D. Pablo Luis López Espí

Dr. D. Juan Antonio Martínez Rojas

Año 2014

Pablo Luis López Espí y Juan Antonio Martínez Rojas, Profesores Titulares de Universidad del Departamento de Teoría de la Señal y Comunicaciones de la Universidad de Alcalá,

HACEN CONSTAR

Que la tesis “**Caracterización automática de especies de madera mediante técnicas de clasificación de imágenes**”, presentada por D. Jesús Alpuente Hermosilla, realizada en el Departamento de Teoría de la Señal y Comunicaciones bajo nuestra dirección, reúne los méritos suficientes para optar al grado de Doctor, por lo que puede procederse a su depósito y lectura.

Alcalá de Henares, 7 de noviembre de 2014.

Fdo.: Dr. D. Pablo Luis López Espí

Fdo.: Dr. D. Juan Antonio Martínez Rojas

D. Jesús Alpuente Herмосilla ha realizado en el Departamento de Teoría de la Señal y Comunicaciones, bajo la dirección de los Doctores D. Pablo Luis López Espí y D. Juan Antonio Martínez Rojas, la tesis doctoral titulada “**Caracterización automática de especies de madera mediante técnicas de clasificación de imágenes**”, cumpliéndose todos los requisitos para la tramitación que conduce a su posterior lectura.

Alcalá de Henares, 7 de noviembre de 2014.

EL DIRECTOR DEL DEPARTAMENTO

Fdo.: Dr. D. Saturnino Maldonado Bascón

Caracterización automática de especies de madera
mediante técnicas de clasificación de imágenes

Jesús Alpuente Hermosilla

2014

*A mis padres, que en su cuarta edad
me han prestado su apoyo permanente.*

Agradecimientos

Quiero, con estas breves líneas, expresar mi agradecimiento a todas aquellas personas que han hecho posible la realización de este trabajo.

En primer lugar a los compañeros del Departamento por sus ánimos, comprensión y apoyo permanente desde que iniciara esta aventura que hoy culmina y que no es más que otra etapa en mi vida personal y profesional. A quienes me animaron a emprender este camino y, en particular, a los miembros de la Unidad Docente de Electromagnetismo, por ser los que más me han sufrido en el quehacer docente diario.

Mención especial merecen mis directores de tesis, Pablo Luis López Espí y Juan Antonio Martínez Rojas, por su dedicación, seguimiento y ayuda en todas las fases de realización de este trabajo, por su esfuerzo en la revisión de esta memoria y por aguantar mis nervios, silencios y dudas en muchos momentos.

Sin duda, este trabajo no habría sido posible sin la colaboración del Profesor D. Santiago Vignote Peña, Catedrático de la E.T.S.I. de Montes de la Universidad Politécnica de Madrid e impulsor indirecto de la idea que ha llevado a desarrollar esta tesis, que junto con Isaac Martínez, en su momento gerente de Maderas de El Espinar S.A., me han facilitado todas las muestras de madera con las que he realizado las diferentes pruebas reflejadas en este documento, despejándome cuantas dudas sobre el mundo de la madera me hayan podido surgir.

Sin todos ellos, esta tesis no habría visto la luz.

Resumen

Caracterizar un material consiste en determinar los atributos peculiares del mismo de modo que se distinga claramente de los demás. El material en esta tesis va a ser la madera, en forma de chapas fáciles de escanear para así obtener sus propiedades fotométricas y texturales y su dimensión fractal, como peculiaridades de cada una de ellas.

El problema consiste en que las diferencias entre especies pueden llegar a ser menores que las existentes dentro de una misma especie o individuo. Para resolverlo y poder identificar la especie a la que pertenece cada una de las muestras se va a recurrir al uso de técnicas de reconocimiento de patrones basadas en la teoría de la decisión.

Se propone en este trabajo de investigación el poner las bases para la automatización del proceso de clasificación de maderas, de bajo coste por el uso de un escáner y un ordenador personal junto con paquetes informáticos de dominio público (*ImageJ* y *Weka*). La digitalización de las chapas de madera y el posterior procesamiento de las mismas para obtener las características predichas a partir de las cuales poder clasificar cada especie forman la base fundamental de esta tesis.

Sin embargo, se plantea en este trabajo establecer la influencia que tienen estos mismos parámetros medidos sobre las componentes de color de la imagen, junto con las propiedades multiescala de la madera. Para éstas se han desarrollado los procedimientos de creación de las correspondientes imágenes, basados en la microscopía de contraste por interferencia diferencial y en los patrones periódicos subyacentes a toda superficie.

La combinación de todas estas imágenes y su procesamiento hace que se formen patrones de gran dimensión, que requieren de una reducción de la matriz de datos. Diferentes algoritmos llevan a obtener distintos patrones, con los que se prueban clasificadores lineales,

basados en árboles de decisión, con entrenamiento basado en casos, lineales o combinados aleatorios de varios del mismo tipo.

Los resultados obtenidos están a la altura de los conseguidos por otros investigadores, si bien éstos utilizan equipos de elevado coste o, en su defecto, complicados procesos. Estos resultados, además, se presentan con un alto grado de fiabilidad ya que en la mayoría de los trabajos revisados el número de especies utilizado es más limitado que en el caso presente.

Abstract

Materials can be characterized determining their physical parameters. Those parameters describe the differences among them. Wood is the material studied in this thesis in its veneer form. Wood veneers are ideal for optical analysis of parameters like photometric properties, texture and fractal dimensions, which will be used to distinguish them.

The main problem in wood identification is the large difference among samples from the same tree species or even from the same tree. Those differences can be greater than the differences among wood from several tree species. Pattern matching techniques based on Decision Theory will be used to solve this problem in order to identify the wood species.

This work proposes firm foundations for the automatic classification of wood samples, using a low cost approach based on optical scanning with a desktop scanner combined with image processing and data mining operations, performed on a personal computer, by public domain software, Image J and Weka. The data of this thesis come from digital images of wood veneers, forming a set of properties enough for wood species identification.

The influence of these parameters over the colour components of the wood images is studied. Also, the multiscaling properties of the wood images, derived from their structure, are considered. Using a novel approach inspired by Differential Contrast Interference Microscopy, pseudo-3D images and periodic patterns of the wood surface are obtained.

The combination of a large number of samples and the processing of an even greater number of parameters associated with them, make necessary the reduction of the dimensions of the data matrices. Several data mining and pattern matching algorithms are used to accomplish this reduction strategy, among them are linear classifiers based on decision trees with different training methods.

The results of this thesis are as good as or better than the obtained by state-of-art techniques described in the scientific literature. A high reliability has been achieved thanks to the study of a large number of species in comparison with the works of other researchers.

Índice general

Agradecimientos	I
Resumen	III
Abstract	V
Índice general	VII
Lista de figuras	XI
Lista de cuadros	XIII
1. Introducción	1
1.1. El sector de la madera en cifras.	1
1.2. La madera como material industrial.	3
1.3. Contexto del desarrollo de la tesis.	5
1.4. Objetivos del trabajo.	8
1.5. Organización de la memoria.	8

2. Revisión de la literatura científica	11
2.1. Técnicas de procesado de imágenes ópticas.	12
2.1.1. Clasificación de materiales por su textura.	12
2.1.2. Clasificación de maderas.	14
2.1.3. Clasificación de defectos.	17
2.2. Otras técnicas no destructivas.	19
3. Materiales y métodos	23
3.1. Materiales y equipos.	24
3.1.1. Material de trabajo.	24
3.1.2. Equipos informáticos.	26
3.1.3. Programas informáticos.	28
ImageJ	28
Weka	30
3.2. Metodología.	31
3.2.1. Digitalización de la madera.	32
3.2.1.1. Proceso de calibración.	32
3.2.1.2. Obtención de las imágenes digitales.	34
3.2.2. Preparación de las muestras.	35
3.2.3. Extracción de características.	36
3.2.3.1. Imágenes componentes de color.	37
3.2.3.2. Imágenes “microscópicas”.	38
3.2.3.3. Imágenes de patrones periódicos subyacentes.	41
3.2.3.4. Obtención de características.	43
Características fotométricas.	44

Características de textura	46
Dimensión fractal	48
3.2.4. Selección de características.	49
3.2.4.1. Reducción del vector de características.	50
Selector basado en la correlación	51
Análisis de componentes principales	53
3.2.5. Clasificación.	56
3.2.5.1. Árbol C4.5.	57
3.2.5.2. k-vecinos más próximos.	59
3.2.5.3. Bosque aleatorio.	60
3.2.5.4. Bosque de rotación.	61
3.2.5.5. Máquina de vectores soporte.	62
3.2.5.6. El proceso de clasificación.	64
3.3. Análisis de resultados.	66
4. Resultados y discusión.	67
4.1. Obtención de imágenes.	67
4.2. Construcción del vector de características.	68
4.3. Reducción del número de características.	70
4.3.1. Selección basada en la correlación.	71
4.3.2. Selección basada en el análisis de componentes principales.	74
4.3.3. Comparativa de los algoritmos de selección.	75
4.4. Clasificación de patrones.	76
4.4.1. Clasificador J48.	76
4.4.2. Clasificador IBk.	79

4.4.3. Clasificador RandomForest.	83
4.4.4. Clasificador RotationForest.	86
4.4.5. Clasificador LibLINEAR.	86
4.5. Comparativa de clasificadores.	90
5. Conclusiones	93
5.1. Aportaciones originales.	94
5.2. Líneas futuras de investigación.	95
5.3. Difusión de la investigación.	96
Bibliografía	100

Lista de figuras

1.1. Corte transversal de un tronco.	4
3.1. Maderas utilizadas en la tesis.	27
3.2. Barra de herramientas de ImageJ.	29
3.3. Interfaz de usuario de Weka.	30
3.4. Carta IT8.7/2 de EGM Laboratoris Color.	32
3.5. Carta base del perfil de color del escáner Epson Perfection V750 Pro.	33
3.6. Chapa de madera con defecto de origen en el lado izquierdo.	35
3.7. Ejemplo de imagen original y el orden en su división.	36
3.8. Ejemplo de imagen digitalizada y sus componentes primarias de color.	38
3.9. Ejemplo de fractal de 64 x 64 píxeles generados con XOR y su histograma.	38
3.10. Fractal de 5x5 píxeles.	39
3.11. Fractal 5x5 (a), y sus réplicas giradas 45° (b), 90° (c), 135° (d), 180° (e), 225° (f), 270° (g) y 315° (h).	40
3.12. Secuencia de reconstrucción 3D con incrementos angulares de 45°.	41

3.13. Ejemplo de imagen “microscópica” obtenida a partir de una imagen digitalizada.	41
3.14. Proceso de obtención de los patrones periódicos subyacentes a una imagen digitalizada.	43
3.15. Visualización de la periodicidad de la del patrón subyacente.	43
3.16. Conjunto de imágenes que forman el patrón correspondiente a una muestra digitalizada.	44
3.17. Interpretación gráfica de la dimensión fractal.	49
3.18. Gráfico de sedimentación.	55
4.1. Porcentaje de especies por tipo de madera.	68
4.2. Gráfico de sedimentación	75

Lista de cuadros

1.1. Consumo aparente de madera en rollo (millones de m ³): España 2012. . . .	2
1.2. Consumo aparente de madera. Industria de primera transformación (millones de m ³): España 2012.	2
3.1. Especies que componen el estudio.	26
3.2. Características del escáner utilizado.	28
4.1. Reducción de características inicial.	71
4.2. Parámetros seleccionados: CfsSubsetEval con estrategia LinnearForward-Selection.	72
4.3. Parámetros seleccionados: CfsSubsetEval con estrategia BestFirst.	74
4.4. Comparativa de algoritmos de selección.	76
4.5. Resultados del clasificador J48 con selección de características basada en correlación y estrategia hacia adelante.	77
4.6. Resultados del clasificador J48 con selección de características basada en correlación y estrategia el mejor primero.	78

4.7. Resultados del clasificador J48 con selección por análisis de componentes principales.	79
4.8. Resultados del clasificador J48 sin reducción de atributos.	80
4.9. Resultados del clasificador IB3 con selección de características basada en correlación y estrategia hacia adelante.	81
4.10. Resultados del clasificador IB1 con selección de características basada en correlación y estrategia el mejor primero.	81
4.11. Resultados del clasificador IB4 con selección por análisis de componentes principales.	82
4.12. Resultados del clasificador IB4 sin reducción de características.	83
4.13. Resultados del clasificador RandomForest con selección de características basada en correlación y estrategia hacia adelante.	84
4.14. Resultados del clasificador RandomForest con selección de características basada en correlación y estrategia el mejor primero.	84
4.15. Resultados del clasificador RandomForest con selección por análisis de componentes principales.	85
4.16. Resultados del clasificador RandomForest sin reducción de características.	86
4.17. Resultados del clasificador RotationForest con selección por análisis de componentes principales y árbol J48.	87
4.18. Resultados del clasificador LibLINEAR con selección de características basada en correlación y estrategia hacia adelante.	87
4.19. Resultados del clasificador LibLINEAR con selección de características basada en correlación y estrategia el mejor primero.	88
4.20. Resultados del clasificador LibLINEAR con selección por análisis de componentes principales.	89
4.21. Resultados del clasificador LibLINEAR sin reducción de características.	90

Capítulo 1

Introducción

La madera es uno de los materiales naturales de mayor utilización, debido a sus múltiples aplicaciones, dando lugar a la existencia de una industria cuya actividad abarca su transformación en productos de consumo.

El conocimiento del sector industrial de la madera desde el punto de vista coyuntural, de la estructura, propiedades y defectos de la madera como material industrial y el planteamiento de los temas en los que se ha de investigar en lo relativo a la caracterización y clasificación permiten establecer los objetivos de esta tesis, así como la estructura a dar a la presente memoria.

1.1. El sector de la madera en cifras.

Para entender el entorno industrial de la madera se presentan a continuación una serie de indicadores socio-económicos globales relacionados con la industria forestal y de la madera en España.

En 2012 la superficie forestal española ocupaba 277.150 kilómetros cuadrados (el 54,77% de la superficie total), siendo España el tercer país de la Unión Europea en superficie arbolada con 186.360 kilómetros cuadrados. De la superficie total de bosques, según la Comisión Económica de las Naciones Unidas para Europa (UNECE) [1] en 2012 se extrajeron 15,72 millones de metros cúbicos de madera en bruto, de los cuales el 75,19%

se dedicó a usos industriales y el resto a leña. Considerando únicamente la madera en rollo, el consumo aparente fue de 15,513 millones de metros cúbicos en global y de 11,716 millones de metros cúbicos para la madera en rollo de uso industrial, tal y como se recoge en el cuadro 1.1.

MADERA EN ROLLO	GLOBAL	USO INDUSTRIAL
Producción	15,527	11,627
Importaciones	1,735	1,727
Exportaciones	1,749	1,638
Consumo aparente	15,513	11,716

Cuadro 1.1: Consumo aparente de madera en rollo (millones de m³): España 2012.

Dentro del sector industrial relacionado con la madera se pueden distinguir entre la industria de primera transformación (empresas de tableros y de aserrado y preparación industrial de la madera) y la industria de segunda transformación (empresas de envases y embalajes, de muebles, carpinterías,...). En la industria de primera transformación, los consumos aparentes de madera aserrada, chapas y tableros se muestran en el cuadro 1.2.

	MADERA ASERRADA	CHAPAS	TABLEROS
Producción	1,971	0,960	3,587
Importaciones	0,865	0,072	1,191
Exportaciones	0,140	0,044	2,819
Consumo aparente	2,696	0,124	1,959

Cuadro 1.2: Consumo aparente de madera. Industria de primera transformación (millones de m³): España 2012.

Considerando el conjunto de la industria de primera y segunda transformación en 2012 [2], se puede deducir la importancia del sector de la madera en España, tanto social como económicamente, a partir de las 15.065 empresas -el 12,38 % del total de empresas industriales- que ocupan a un total de 109.319 personas -el 5,69 % del empleo industrial- y producen anualmente 10.079,844 millones de euros -el 2,54 % de la producción industrial total-. De estas cifras se desprende que el sector de la madera es un sector mayoritariamente minifundista, con un nivel de empleo medio por empresa de 7,26 empleados, hecho

éste que, sin restarle importancia al sector, dificulta a las empresas la realización de tareas de I+D de forma autónoma, por lo que se ven abocadas a recurrir a asociaciones de investigación como el Instituto Tecnológico del Mueble, Madera, Embalaje y Afines (AIDIMA) o la Asociación de Investigación de las Industrias de la Madera (AITIM).

1.2. La madera como material industrial.

Desde el punto de vista industrial, la madera es un material ortotrópico obtenido del tronco y de las ramas de los vegetales leñosos, que presenta una estructura multiescala compleja. En una primera aproximación a la estructura de la madera, la composición química de la misma contiene celulosa (40-50 %), hemicelulosa (20-25 %), lignina (24-28 %) y sustancias de impregnación [3], si bien las proporciones varían tanto entre especies como dentro de la misma especie, e incluso dentro del mismo individuo.

Las macromoléculas de celulosa se agrupan formando cristales de escala de Armstrong, dando lugar a microfibrillas de escala nanométrica. Considerando la madera como un conjunto de células, en su mayoría alargadas en forma de tubos, dispuestas en la dirección del eje del árbol, la variabilidad de su longitud, forma, espesor de sus paredes y dimensiones de sus lúmenes se tienen escalas micro o milimétricas. Si se observa un trozo de madera se ven los diversos elementos característicos que la forman a escala centimétrica, y además, se puede apreciar que no se trata de un material homogéneo.

Esta jerarquización de la estructura de la madera evidencia la anisotropía de la misma, relacionada con la disposición de las células en escala microscópica y con las tres direcciones de crecimiento en escala macroscópica, haciéndose necesario denominar de forma diferente los planos y direcciones que en cada caso se consideren, siendo las principales direcciones de la madera las designadas como axial, radial y tangencial [4].

Si se observa el tronco de un árbol, se ve que tiene forma troncocónica y que está formado por sucesivos anillos superpuestos: médula, duramen, albura, cámbium, líber, corteza y anillos anuales de primavera y otoño, algunos de los cuales se muestran en la figura 1.1 [5].

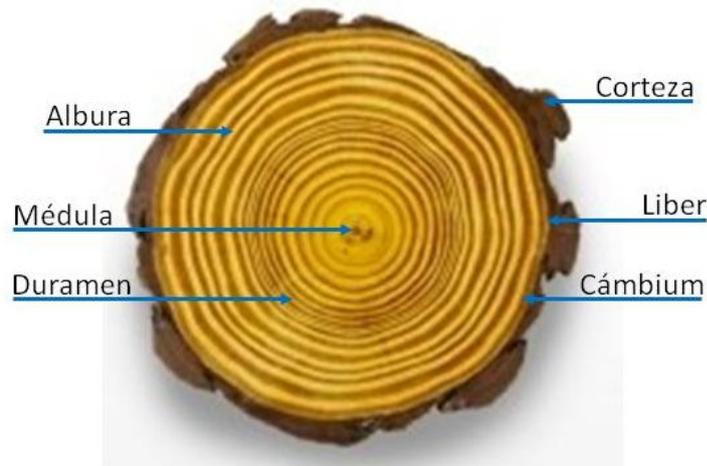


Figura 1.1: Corte transversal de un tronco.

En este entorno ha de tenerse en cuenta que la madera presenta una serie de propiedades y características muy convenientes para su uso como material estructural, y como tal se ha empleado desde los inicios de la civilización. Su buena resistencia, su ligereza y su carácter de material natural renovable constituyen las principales cualidades de la madera para su empleo, tanto en estructuras, como en cualquier otra aplicación industrial.

Las propiedades y características que presenta la madera se pueden clasificar en básicas -anisotropía, higroscopicidad y polaridad- [4], propiedades dependientes de la forma y distribución de los elementos anatómicos de la madera -estructura, textura, grano y densidad- [6], fotométricas -color y brillo- [7], mecánicas -elasticidad, flexibilidad, dureza y resistencia mecánica- [8], térmicas -calor específico, conductividad térmica y difusividad térmica- [9], eléctricas -constante dieléctrica, tangente de pérdidas y conductividad- [10] y acústicas, considerado el comportamiento desde los puntos de vista de aislamiento, reverberación y transmisión [11]. Los valores de los distintos parámetros son variables entre elementos de la misma o de distinta especie y medibles mediante técnicas que habrán de ser diferentes, dada la estructura multiescala de la madera, en función de la resolución a la que se pretenda trabajar [12].

Estas propiedades y características son susceptibles de sufrir variaciones debidas a distintos factores, así como a anomalías y defectos cuyo origen puede ser natural o deberse a la propia manipulación industrial de la madera. Entre los primeros, la humedad es uno de los que más influyen en los procesos industriales, siendo determinante en la variación

de las dimensiones de la madera, pudiendo dar lugar a la hinchazón o merma de la misma. Lógicamente, el porcentaje de humedad de la madera variará su densidad, dureza y resistencia mecánica, afectando a las propiedades térmicas y electromagnéticas, así como a la velocidad de transmisión acústica [4].

Entre las anomalías y defectos que producen variaciones en las propiedades de la madera se encuentran los nudos que disminuyen la resistencia mecánica y producen variaciones sobre la distribución del color, dirección de las fibras y comportamiento acústico de la madera [13, 14], las acebolladuras y fendas producidas por esfuerzos mecánicos, condiciones climáticas y tensiones de crecimiento que producen variaciones en los parámetros acústicos, densidad y características electromagnéticas de la madera [13, 15] y la impregnación natural de resina que disminuye la resistencia mecánica y produce variaciones en la transmisión acústica y en las propiedades electromagnéticas [13, 16]. Aparte de estas alteraciones, se pueden considerar otras que tienen orígenes diversos; entre ellas, las que tienen su base en agentes bióticos destructores de la madera -coloraciones, pudriciones, picaduras y perforaciones- que afectan a su color y brillo, influyendo en las propiedades mecánicas, electromagnéticas y acústicas, y las anomalías producidas por agentes abióticos de destrucción o degradación de la madera -oscurecimiento por degradación de la luz, meteorización y quemaduras- que afectan a la práctica totalidad de los parámetros característicos de la madera [4].

Todas estas alteraciones de la madera tienen un gran impacto económico, tanto para los fabricantes como para los usuarios de elementos de madera, dado que la calidad, la integridad estructural y el precio del producto final dependen del tamaño, localización y tipo de estos defectos, lo que ha llevado a desarrollar una vasta normativa para definir no sólo los parámetros y métodos de medida de los mismos, sino también para la clasificación de la madera por su calidad.

1.3. Contexto del desarrollo de la tesis.

Desde el inicio de la industria de la madera, el método más utilizado para su identificación y clasificación ha sido el basado en la inspección visual, que requiere una vasta

experiencia y una excelsa cualificación profesional, obteniéndose resultados transcurrido un tiempo prolongado.

El desarrollo de métodos automáticos no fue posible hasta el siglo XX en el que se dispuso de la instrumentación necesaria para la medida de las propiedades de la madera, reduciéndose los tiempos de respuesta y permitiendo la realización de estudios que, mediante la utilización de técnicas destructivas -microscopía electrónica, marcación molecular y análisis del ADN- o no destructivas -tomografía, termografía, densitometría, resonancia, reflectometría y espectrometría-, han facilitado la determinación de los valores de diferentes características y el análisis de los resultados mediante diversas técnicas de procesado, dando lugar al desarrollo de clasificadores que cubren diferentes objetivos, si bien la mayoría de los estudios van dirigidos a la identificación de defectos a partir de las medidas paramétricas realizadas con equipos que por lo general requieren un desembolso económico considerable o son equipos comerciales diseñados para unos fines concretos y considerando un número reducido de patrones de trabajo.

Teniendo en cuenta, por una parte, el carácter familiar de la mayoría de las empresas del sector y, por otra, la globalización de los mercados de la madera que implica un flujo continuo de importaciones y exportaciones tanto de madera en rollo como de productos de madera elaborados, es necesario que se disponga de las herramientas necesarias para que se pueda llevar a cabo una trazabilidad adecuada en la cadena de producción, utilizando para ello equipos económicos que, requiriendo cortos tiempos de procesado, permitan la obtención de parámetros de la madera que faciliten su clasificación en cuanto a especie, origen y calidad, así como en lo concerniente al procesado industrial a que ha sido sometida.

La investigación tendente al desarrollo de una tecnología que cumpla con los requisitos anteriormente mencionados ha de basarse en los siguientes pilares básicos:

1. *Medida de parámetros mediante tecnologías no destructivas de bajo coste.* La investigación debe orientarse hacia la utilización de las técnicas que mejor cumplan con el compromiso del coste. Las basadas en el uso de elementos ópticos de toma de datos (escáneres o cámaras digitales), al alcance de cualquier economía, se presentan así como las más adecuadas, sin que en ningún caso se descarten técnicas acústicas que

utilicen como tarjetas de adquisición las de sonido insertadas en ordenadores personales de sobremesa o portátiles. En cualquiera de los casos habrá de contemplarse el análisis de los parámetros obtenidos y el grado de validez y fiabilidad de sus valores.

2. *Creación de algoritmos eficientes de procesado de las muestras*, que integren el análisis de los valores característicos obtenidos junto con la extracción de defectos y la obtención de los parámetros identificativos de la muestra, así como la determinación de la calidad de la madera según la aplicación que se le vaya a dar. Es de interés que el procesado de los patrones permita también definir el grado de humedad y el azulado debido a la exposición a rayos ultravioleta de las muestras, a la vez que ha de permitir determinar el origen de las mismas.
3. *Generación de grandes bases de datos* a partir de las muestras existentes en distintas xilotecas, que permitan disponer de los datos necesarios para poder determinar la calidad de la muestra en estudio. Estas bases de datos aportarán las muestras de entrenamiento de los clasificadores que permitan estimar la calidad de las nuevas muestras.
4. *Uso de estructuras de clasificación* que posibiliten la rápida identificación de la madera, basada en las características que conformen las mencionadas bases de datos. La combinación adecuada de clasificadores encargados, cada uno de ellos, de verificar la pertenencia de la muestra a una u otra clase, tipo, orden, familia, especie, variante, origen, tratamiento,...- deberá concluir con una estimación de, al menos, la especie de la muestra.
5. *Diseño e implementación de equipos integrados* que contemplen los elementos de hardware y software necesarios para cubrir todas las fases del proceso: adquisición de datos, determinación de parámetros, clasificación de la muestra y volcado de resultados. Los equipos deberán cumplir los requisitos de ser portátiles, lo que facilitará su utilización en diferentes entornos o su compartición por parte de industrias próximas, y de bajo coste, requisito éste impuesto por el carácter estructural de la mayor parte de las industrias de la madera.

Necesariamente, aunque cada uno de estos pilares requiere de una investigación pormenorizada e individualizada, no ha de perderse de vista la interrelación entre ellos, de manera que las técnicas de medida de parámetros han de tener en cuenta cuáles de ellos van a ser utilizados para su procesamiento y adecuación a los de las bases de datos que den lugar a su posible clasificación y, finalmente, den lugar a un sistema integrado. Los parámetros característicos contemplados en la base de datos fijan las técnicas de adquisición de datos y su procesamiento. El diseño final de los sistemas automáticos ha de contemplar el control de los dispositivos de adquisición, dependientes de las características a considerar para su clasificación a partir de los valores de entrenamiento que forman la base de datos.

1.4. Objetivos del trabajo.

Teniendo en cuenta las bases de la investigación comentadas, el desarrollo de esta tesis doctoral fundamentará las bases para el diseño e implementación de un sistema automático de clasificación de especies de madera a partir de las imágenes en color obtenidas mediante el escaneado de muestras de las mismas.

El objetivo principal que se persigue es el de conseguir un sistema de bajo coste que presente bajos tiempos de respuesta y permita la clasificación de las distintas especies con un error de clasificación por debajo de los manejados por los sistemas utilizados actualmente. El sistema estará basado en la descripción de las características de la madera a partir de los parámetros derivados de la información medida mediante técnicas fotométricas que permita encontrar las propiedades distintivas de cada una de las especies para, mediante el uso de un clasificador a definir en función de una comparativa entre diferentes tipos, obtener un error de clasificación que, en media, esté por debajo del 10 %.

1.5. Organización de la memoria.

La presente memoria comienza con una introducción (capítulo 1) en la que se establece el marco socio-económico de la industria de la madera, junto con las características propias de este material y su variabilidad. Una breve descripción de las investigaciones que han

de tener como finalidad la obtención de un sistema de identificación de la madera que, contemplando la mayor parte de variables posibles, adapte su coste a la realidad de la industria maderera española, deriva en los objetivos perseguidos por esta tesis. Se cierra esta parte de la memoria con una breve descripción de la forma en que está organizada la memoria.

En el capítulo 2 se realiza una revisión del estado de la tecnología, a partir de las investigaciones publicadas sobre la identificación y clasificación de especies de madera o de sus defectos, haciendo especial énfasis en las fundamentadas en técnicas visuales basadas en la captación de imágenes e incluyendo la técnica utilizada en la adquisición de datos, las características medidas y, en su caso, el procesamiento dado a los valores obtenidos y el tipo de clasificador utilizado.

El capítulo 3 se dedica a la descripción del material y equipos y programas informáticos utilizados, así como al desarrollo de la metodología utilizada en la obtención de resultados, revisándose de forma sucinta los conceptos, técnicas y herramientas utilizados en su desarrollo, entre los que se encuentran las técnicas de procesado de las imágenes para obtener los valores de las características de la madera, las técnicas de reducción de datos y la descripción del funcionamiento de diferentes clasificadores.

En el capítulo 4 se recogen los resultados obtenidos en la investigación, siguiendo la metodología previamente definida, así como al análisis y discusión de los mismos, que dan lugar a la elección de los procedimientos y técnicas que, por comparación, permiten una mejor respuesta en la clasificación de especies de madera.

Las conclusiones generales a que da lugar el trabajo desarrollado se establecen en el capítulo 5, en el que también se hace un desglose de las principales aportaciones originales a la investigación, así como una referencia a los futuros trabajos de investigación basados en esta tesis. Se cierra este capítulo con una relación de las publicaciones que ha dado lugar el desarrollo de esta tesis, de forma directa o indirecta.

Configuran la última parte de esta memoria las referencias bibliográficas consultadas durante la elaboración del trabajo.

Capítulo 2

Revisión de la literatura científica

La utilización de métodos no destructivos en la caracterización de las propiedades de la madera viene siendo investigada desde hace varias décadas, cuando se publicaron los primeros trabajos en los que se establecían las relaciones fundamentales entre varios parámetros obtenidos mediante pruebas no destructivas y las correspondientes propiedades mecánicas. La celebración del simposio internacional celebrado en 1994 en Sopron (Hungría) sirvió como lanzamiento de la investigación europea en técnicas y evaluaciones no destructivas de la madera, que ha dado lugar a un gran esfuerzo en normalización, así como a la colaboración entre empresas y universidades e institutos de investigación para el desarrollo y fabricación de herramientas basadas en diferentes principios físicos, con diferentes usos finales [17].

Cada uno de los métodos utilizados, basados en técnicas acústicas, vibracionales, electromagnéticas, ultrasónicas, de infrarrojos y térmicas, radiográficas, láser y de procesado de imágenes ópticas, va a permitir la evaluación de diferentes parámetros relacionados con las propiedades físicas de la madera. Los datos obtenidos de las mediciones realizadas, convenientemente procesados, permitirán llevar a cabo una clasificación de las muestras sobre las que se hayan aplicado las técnicas de medida.

Aunque de todas estas técnicas solo unas cuantas están relacionadas con el tratamiento digital de imágenes obtenidas por métodos ópticos convencionales, a continuación se hace

una revisión de algunas de las investigaciones en que se ha utilizado cada una de ellas, considerando los casos cuya finalidad es la clasificación de maderas.

2.1. Técnicas de procesado de imágenes ópticas.

Dentro de estas técnicas se encuentra la inspección visual de la madera, que presenta un alto grado de subjetividad, dado que la clasificación realizada por dos operarios distintos observando las mismas tablas es coincidente en sólo un 60 % de los casos y el error medio de clasificación de un operario es de un 32 % [18].

El resto de técnicas que se incluyen en este apartado se basan en la utilización de dispositivos digitales de captación de imágenes –cámaras de vídeo o fotográficas y escáneres– que permiten evaluar la intensidad y el color de la luz reflejada, pudiendo detectarse características superficiales tales como defectos, color, textura o patrones del grano de las maderas [19]. En este punto han de tenerse en cuenta los trabajos tendentes a la clasificación por su textura de diferentes materiales, entre los que se encuentra la madera, los relacionados con la detección, identificación y clasificación de defectos, y aquellos que tratan de clasificar especies de madera, siendo estos últimos los que tienen una relación más directa con la temática de la tesis que se presenta, si bien los anteriores permiten analizar algunas de las técnicas de clasificación utilizadas a partir de la extracción de características de la madera conseguidas mediante distintos métodos de procesado de imagen.

2.1.1. Clasificación de materiales por su textura.

Utilizando imágenes procedentes de bases de datos de uso común, se han estudiado clasificadores de máxima verosimilitud, de mínima distancia y de k-vecino más próximo. El primer tipo de clasificadores ha sido utilizado por Ojala *et al.* a partir de las características resultantes del análisis multirresolución realizado a partir de la combinación de la medida del contraste local y el uso de patrones binarios locales [20], por Vichers y Modestino que los han aplicado a la probabilidad de coocurrencia de niveles de gris del histograma ecualizado para cada una de las partes resultantes de la segmentación de la imagen [21]

y por Thyagarajan *et al.* que han usado las matrices de coocurrencia de las imágenes en multirresolución obtenidas con la transformada ‘wavelet’ como entradas al clasificador [22].

Eligiendo como criterio de clasificación la mínima distancia, Haley y Manjunath se han basado en las características de escala, frecuencia de rotación e índice de correlación obtenidas al utilizar un filtro de Gabor para imágenes rotadas en pasos de 30° [23]; Akhloufi *et al.* en la distancia global de textura de las probabilidades de coocurrencia de los niveles de gris de las bandas de color de las imágenes [24]; Kashyap y Khotanzad en las características obtenidas por aplicación de un modelo autorregresivo de simetría circular a los grados de rugosidad y direccionalidad estimados por mínimos cuadrados de diferentes divisiones de las imágenes y distintas rotaciones de las mismas [25]; Chellappa y Chatterjee en las extraídas mediante modelos aleatorios de campos de Gauss–Markov [26], sirviendo el mismo criterio a Porter y Canagarajah para comparar los resultados obtenidos con éstas y las que resultan de aplicar filtros de Gabor con simetría circular o la transformada ‘wavelet’ a cada una de las particiones de la imagen original [27]. También Rivazhagan *et al.* han utilizado los clasificadores de mínima distancia para comparar las características obtenidas con bancos de filtros de Gabor con las estadísticas y las de coocurrencia de niveles de gris a partir de las imágenes a las que se les ha aplicado la transformada ‘wavelet’ [28].

Los clasificadores de k-vecino más próximo los han utilizado Palm *et al.* a partir de la descomposición de imágenes mediante la transformada de Gabor para distintas frecuencias centrales y orientaciones [29] y Liu *et al.* que se han basado en un método empírico de descomposición direccional alimentado por las características intrínsecas locales de banda estrecha y la dimensión fractal obtenida del cálculo de la distancia media entre una imagen y su rotada [30].

Por otra parte, Pyun *et al.* [31] han comparado diferentes estructuras fundamentadas en la mezcla de múltiples cuantificadores vectoriales gaussianos para cada textura, usando como datos de entrenamiento los de mínima distorsión en la discriminación de la información, habiendo servido las características establecidas por codificación mediante la transformada discreta de Fourier de la salida de un filtro en pirámide de Laplace orientada

para que Greenspan *et al.* hayan comparado el comportamiento de clasificadores de los tipos k-vecino más próximo, red neuronal con retropropagación y K-medios [32].

2.1.2. Clasificación de maderas.

En relación con la clasificación de la madera considerando el color y textura de sus imágenes, se han comparado diferentes clasificadores -bayesiano, k-vecino más próximo, k-vecino más próximo con decisión borrosa, red neuronal, clasificador de decisión borrosa, clasificador genético y máquina de vectores soporte-. Para el vector de características formado por la media de las componentes del espacio de color L^*a^*b y la homogeneidad de la luminancia y el matiz, trabajando con muestras de seis clases de color en tableros de roble rojo, Bombardier y Smichmitt han obtenido como resultado que el clasificador basado en la máquina de vectores soporte, con una función de base radial gaussiana como núcleo, tiende a una tasa de reconocimiento de clases del 85,5 % [33]. Trabajando con diez especies de madera de la India, Bremananth *et al.* han extraído las características de textura -entropía, energía, contraste, segundo momento angular y homogeneidad- a partir de las probabilidades de coocurrencia de los niveles de gris cada 45° , realizando la clasificación a partir del coeficiente de correlación de cada imagen de prueba con las de entrenamiento [34], si bien los resultados se limitan a comprobar el funcionamiento del sistema diseñado.

Liu y Furuno establecen la distribución superficial de la dimensión fractal correspondiente a tableros de quince especies distintas, partiendo de la división de las imágenes escaneadas en bloques de 128×128 píxeles y aplican a cada uno de los bloques el método de la superficie de un prisma triangular, en el que representar los valores de luminancia en las esquinas y su media en el centro [35].

Por otra parte, trabajando con las imágenes correspondientes a las secciones transversal y tangencial de sesenta especies chinas de madera, You y Cai han estudiado el grado de reconocimiento de las muestras comparando para cada sección métodos basados en el análisis de componentes principales y en el análisis discriminante lineal, obteniéndose las mejores tasas de aciertos, próximas al 87 %, cuando se utilizan técnicas bidimensionales y bidireccionales bidimensionales de análisis por componentes principales [36].

La medida de los parámetros de cromaticidad de la señal dada por una cámara de vídeo, reducidos mediante un análisis de componentes principales, han permitido a Lebow *et al.* comparar sendos clasificadores basados en el análisis discriminante, determinándose un mejor comportamiento en el caso cuadrático [37]. En la clasificación de la madera por su color, con el uso de un clasificador basado en un sistema de inferencia borrosa, autoconstruido mediante reglas lingüísticas difusas derivadas de la fusión de las componentes de color de las dos caras de un tablero, Schmitt *et al.* han conseguido una tasa de reconocimiento del orden del 84% para cada una de las seis clases de color perceptible preestablecidas [38].

Por otra parte, Johansson y Morén han experimentado sobre la resistencia a la flexión de madera de abedul tratada o no térmicamente, partiendo de la homogeneidad del color en el espacio $L^*a^*b^*$, determinando que la distribución de color en los tableros tratados no es homogénea y obteniéndose, mediante la regresión de mínimos cuadrados locales, una bondad de ajuste de 0,67 y una bondad de predicción de 0,64 [39].

Faria *et al.* han usado un sistema de inferencia mínimo-máximo borrosa de las funciones campana de pertenencia ajustadas a cada histograma de las bandas de color que ha permitido clasificar más del 95% de las treinta clases de color de tres especies de tableros de madera, tras segmentar las imágenes mediante ventanas solapadas de 400x400 píxeles y someterlas a un procedimiento de suavizado [40].

Wang *et al.* han conseguido un 91,7% de aciertos en la clasificación de veinticuatro especies de madera con una máquina de vectores soporte alimentada por las características obtenidas de la matriz de co-ocurrencia de niveles de gris [41], en tanto que utilizando un clasificador k-vecino más próximo Tou *et al.* han comparado diferentes métodos de extracción de características de textura, correspondientes a seis especies de madera, resultando un mejor porcentaje de aciertos del 85% con un filtro de Gabor basado en la matriz de covarianza [42]. Wang y Bai [43], trabajando con seis especies, han comparado un clasificador k-vecino más próximo con una red neuronal, logrado un 98,5% de aciertos de clasificación con el segundo, a partir de las características extraídas mediante la matriz de co-ocurrencia de niveles de gris, modelos de campos aleatorios de Gauss-Markov y la

dimensión fractal ‘wavelet’, reducidas mediante un proceso de normalización y fusión de datos por combinación.

Las redes neuronales se han utilizado como clasificador en el caso de treinta especies de madera de las que se han extraído las características de textura mediante un filtro de Gabor y la matriz de co-ocurrencia de niveles de gris, logrando Yusof *et al.* un 90,33 % de aciertos en clasificación [44], en la clasificación de maderas en función de las características de textura de su imagen Bai y Wang han utilizando el modelo aleatorio de Gibbs-Markov de segundo grado y una estimación de los parámetros por mínimos cuadrados [45]. Por su parte, Topalova y Tzokev mediante un controlador de lógica programable en una red neuronal de estructura modular, alimentada por la correlación entre cada dos clases de las características obtenidas en paralelo del histograma de cada una de las filas en que se segmenta la imagen, al considerar cinco rotaciones de la imagen han logrado un 95 % de muestras bien clasificadas [46].

A partir de muestras de seis tipos de chapas de madera, y una vez mejorado su contraste, Gonzaga *et al.* dividen cada imagen en treinta y seis regiones, extrayendo tres características del histograma de la diferencia entre píxeles y una cuarta de la suma de las regiones de una misma línea, de manera que con un clasificador formado por una red neuronal borrosa local, una red neuronal borrosa por línea y una lógica borrosa global consiguen algo menos del 86 % de chapas correctamente clasificadas [47].

Wang *et al.*, integrando redes neuronales con retropropagación han logrado un 90,25 % de aciertos en la clasificación de cinco especies de madera decorativa y de interior, partiendo del vector de características resultante de la determinación de la matriz de co-ocurrencia de niveles de grises para cada 45° [48]. Otros investigadores que han utilizado redes neuronales del tipo perceptrón multicapa han sido Khalid *et al.* para la clasificación por textura de 20 especies, a partir de la matriz de co-ocurrencia de niveles de gris, con una tasa de aciertos superior al 95 % [49], o Yi que, trabajando con cinco especies y las características extraídas mediante el método anterior a partir de la imagen tratada con un filtro de Gabor ha clasificado correctamente el 98,33 % de las muestras [50].

2.1.3. Clasificación de defectos.

En cuanto a los trabajos revisados que se han desarrollado relacionados con la detección, identificación y clasificación de defectos, se encuentran aquellos que parten de imágenes de maderas que incluyen defectos en su estructura, necesitándose de técnicas de segmentación para detectar los mismos, y los que se basan en el análisis de imágenes de defectos previamente seleccionados. Obviando, en su caso, los métodos de segmentación empleados, se han utilizado diferentes clasificadores de defectos, centrándose en ellos la inclusión de este apartado, por lo que pueda aportar al desarrollo de esta tesis.

Entre ellos, un clasificador bayesiano aplicado a las características de fondo, forma y textura de los defectos obtenidas tras la umbralización de la imagen [51] o a la varianza y la intensidad media de la imagen, junto con los resultados de aplicarle la transformada de Hough y la detección de bordes, una vez procesada la imagen con un filtro de media y mejorada en contraste [52].

Un clasificador basado en el método AdaBoost de agregación a partir de las características extraídas por combinación de métodos como la imagen integral, la transformada de Hough o el camino más eficiente entre dos puntos de los contornos [53], otro basado en el coeficiente de correlación de los estadísticos de color y las probabilidades de co-ocurrencia de niveles de gris [54] o en la reducción de bordes y la diferencia de las rugosidades debidas a diferentes iluminaciones [55], y una combinación de clasificadores k-vecino más próximo puro con otro del mismo tipo en árbol, que utiliza como características las correspondientes al color, geometría y estructura de los defectos [56], son otros ejemplos de clasificadores investigados. Se ha usado un clasificador k-vecino más próximo y una máquina de vectores soporte para clasificar los nudos de pinos en función de la resistencia a la ruptura de la madera a partir del parámetro de determinación de la correlación del tamaño, posición, longitud, volumen y tipo de defectos [57].

También se han utilizado máquinas de vectores soporte con estructura en árbol para la clasificación de nudos a partir de su tamaño y del pseudocolor medio [58], obteniéndose un mejor rendimiento con una máquina de vectores soporte que con una red neuronal del tipo perceptrón multicapa para un vector formado por características geométricas y de color, completadas con la probabilidad de co-ocurrencia de niveles de gris [59].

Las redes neuronales han sido ampliamente utilizadas, tanto a partir de las características extraídas mediante el criterio de rechazo de las variaciones inter e intra clase y de la correlación de características [60], como a partir de los estadísticos de la imagen y de las características de los bordes para diferentes umbrales, reducidos mediante un conjunto aproximado de precisión variable [61] o con el principio de consistencia [62]. En otro de los trabajos, a partir de las imágenes monocromas de chapas de madera se extraen los estadísticos de la imagen y se detectan los defectos por multiumbralización del histograma y detección de bordes laplaciano, alimentando los resultados obtenidos de la identificación de los defectos a un árbol de decisión en el que los distintos nodos son redes neuronales [63]. En [64] se utiliza una red neuronal con metaplasticidad artificial para caracterizar defectos una vez extraídas las características de densidad, forma y factores geométricos obtenidas tras pasar la imagen por un filtro de Gabor.

También se han utilizado redes neuronales del tipo perceptrón multicapa para clasificar defectos a partir del análisis por componentes principales al filtrado de Gabor de las imágenes [65], reduciendo el espacio de características geométricas, de color en los cuatro canales y estadísticas de los histogramas con un algoritmo de selección de características de información mutua y entrenando el clasificador con un algoritmo genético basado en el método cuasi Newton de segundo orden [66], o mediante la combinación de la agrupación autoorganizada de los coeficientes de Gabor, las características de los histogramas de color y de niveles de gris [67].

La combinación de redes tipo perceptrón multicapa aplicadas al vector formado por las características de brillo, tamaño, orientación y forma de los defectos [68], la suma de redes neuronales con entrenamiento por cuantización de vectores [69] o la comparativa de redes neuronales autoorganizada, de lógica borrosa y con propagación hacia adelante, utilizando los coeficientes de Gabor reducidos mediante los métodos de incorporación y análisis de componentes principales [70], han sido otros de los estudios llevados a cabo sobre la clasificación de defectos con redes neuronales.

Por último, en la clasificación de defectos también se ha recurrido al uso de mapas autoorganizados, considerando once percentiles [71] o centiles [72] del histograma de color, experimentándose diversas soluciones para la combinación de centiles del histograma con

probabilidades de co-ocurrencia y las características derivadas de la aplicación de patrones binarios locales [73].

2.2. Otras técnicas no destructivas.

Con el objetivo de identificar y clasificar especies de madera se ha investigado en el uso de diferentes técnicas no relacionadas con las imágenes ópticas, entre las que se encuentran las acústicas que, basadas en la medida de la velocidad de propagación o de los niveles de señal, permiten caracterizar mecánicamente las muestras bajo estudio. Recoger el sonido obtenido por impacto acústico sobre láminas de madera, se muestra como un método robusto de identificación de las mismas, considerando características como duración del impulso, número y amplitud de los picos de señal, tiempo de rizado e intensidad media de la onda acústica registrada [74], al igual que la monitorización de vigas a partir de las componentes cepstrales, reducidas mediante el análisis de componentes principales, aplicadas a clasificadores basados en máquinas de vectores soporte y a redes neuronales del tipo perceptrón multicapa [75], o la espectrometría por resonancia acústica, con la que se han logrado buenos resultados en la identificación de especies mediante técnicas de análisis multivariante [76], o el uso de técnicas basadas en la convolución del sonido producido por varios impactos acústicos sobre la misma muestra, dando lugar a un espectro que permite obtener características de frecuencia e intensidad diferenciadas para cada especie de madera [77].

En [78, 79] se lleva a cabo una revisión de técnicas de emisión acústica y acústico-ultrasónicas que permiten determinar el grado de humedad, la orientación de las fibras o la densidad y estructura de la madera.

Las técnicas ultrasónicas, con las que se obtiene información sobre la distribución y el flujo de la energía reflejada o transmitida al sonificar las muestras desde diferentes direcciones, hacen posible la evaluación de la elasticidad de la madera, así como la detección de defectos o su grado de descomposición. Utilizando ultrasonidos se ha comprobado que la propagación de la onda en la dirección longitudinal es la más útil en la determinación de las propiedades elásticas, siendo posible la identificación de especies a partir de la detec-

ción del grado de humedad mediante la medida del tiempo de propagación y la energía de la señal [80]. En cuanto a métodos de clasificación, se ha investigado la diferenciación entre cuatro especies, aplicando una red neuronal del tipo perceptrón multicapa a las características medidas de velocidad y amplitud de la onda que se propaga longitudinalmente [81].

En el caso de las técnicas vibracionales, la frecuencia natural de la vibración de la madera va a depender de la densidad y el módulo de elasticidad de la misma, proponiéndose un sistema de clasificación en cuatro clases para diecinueve especies de madera tropical, a partir del valor del módulo de elasticidad obtenido de los diferentes modos de vibración de las muestras [82].

Las técnicas de microondas tienen como objetivo fundamental la determinación de las propiedades dieléctricas de la madera que permitan la evaluación de la densidad, contenido de humedad o dirección de las fibras. En este contexto, se han dispuesto las muestras de madera entre dos antenas de bocina para medir la constante dieléctrica de la madera, pudiendo utilizarse los resultados para clasificar y estimar la densidad, el módulo de elasticidad y la resistencia de la madera [83]. Utilizando microondas también se ha propuesto un sistema de clasificación basado en máquinas de vectores soporte, a partir de la clasificación realizada para diversos materiales con una red neuronal del tipo perceptrón multicapa a la que se aplican las características de la imagen radar obtenida en el espacio libre [84].

Otra de las técnicas no destructivas utilizadas en la evaluación de la madera es la termografía, que permite medir la distribución de la temperatura superficial de la madera a partir de la radiación infrarroja que ésta emite, dependiendo los resultados logrados de las propiedades térmicas del material. La imagen térmica obtenida posibilita la distinción de diferentes tipos de madera por su densidad [85].

La espectrometría infrarroja, entendida como la medida del espectro de absorción de la energía de los rayos infrarrojos emitidos por una fuente cuando la frecuencia de vibración de las moléculas de la madera coincide con la frecuencia de la excitación, ha sido utilizada en diversos trabajos de investigación relacionados con la identificación del origen de doce especies tropicales mediante el análisis de agrupaciones tridimensionales

de la transformada de Fourier del interferograma obtenido con un equipo comercial [86], la clasificación de madera de apariencia visual análoga, aplicando el análisis discriminante al resultado de la regresión por mínimos cuadrados parciales [87], o la discriminación entre ocho especies utilizando la distancia de Mahalanobis entre clases, calculada para tres longitudes de onda, y aplicando el análisis discriminante con la segunda derivada del espectro [88].

Utilizando la espectrometría en las bandas del visible y del infrarrojo cercano, se ha utilizado la función de probabilidad de pertenencia de la muestra a una clase para clasificar material diverso, incluida la madera, a partir de la intensidad del infrarrojo y la luminancia del color [89], o bien el análisis discriminante cuadrático de las componentes principales de las curvas de reflectancia espectral y el color para la clasificación de defectos [90].

Por último, diferentes técnicas basadas en el uso de láseres han dado lugar a investigaciones relacionadas con la clasificación de especies de maderas. Así, al iluminar superficies de madera con un diodo láser desenfocado se obtiene una imagen de moteado de la muestra que, autocorrelada iterativamente, permite obtener las propiedades regulares de su estructura -tamaño, dirección de las fibras, lijado, tratamientos-, de manera que puede plantearse la utilización de este método en la clasificación de especies [91].

El resultado de aplicar la transformada de Fourier al espectro que la dispersión Raman produce al iluminar una superficie de madera con la luz de un láser en el rango del infrarrojo cercano permite la discriminación de diferentes tipos de madera mediante mapas autoorganizados de Kohonen de dos dimensiones teniendo en cuenta las características espectrales, basándose en las distancias euclídeas de la matriz de pesos [92]. También se ha utilizado la técnica de espectrometría de fluorescencia para, integrando la intensidad medida para diferentes bandas, comparar diferentes clasificadores basados en el análisis discriminante lineal y cuadrático, en una máquina de vectores soporte o en el k -vecino más próximo, con diferentes valores de k , cuando se quieren distinguir coníferas de frondosas de veintidós tipos diferentes [93], o procesando el espectro de fluorescencia emitido por la madera con diferentes bancos de filtros ópticos y utilizando una red neuronal [94].

Capítulo 3

Materiales y métodos

La revisión realizada en el capítulo 2 sobre los procedimientos empleados para la clasificación de maderas arroja que son muchas las técnicas de medida sobre las que se ha investigado, llegándose incluso al desarrollo de equipos comerciales. En el caso de las no destructivas, en forma mayoritaria se sugiere el uso de equipamiento específico y costes relativamente altos. En términos económicos, las técnicas que mejor se adaptan al tipo de empresa mayoritario en el sector español de la madera son las basadas en la captación de imágenes que precisan de un ordenador y una cámara fotográfica o un escáner, cuyos precios son asumibles por industrias incluso de tipo familiar.

En cuanto a la finalidad de los estudios citados, pocos son los que utilizando técnicas basadas en la captación de imágenes para la identificación o clasificación de especies, dedicándose la mayoría de las investigaciones a identificar o clasificar defectos de la madera.

También del análisis de las diferentes investigaciones se desprende que las técnicas utilizadas en el procesado de las características suelen ser, a veces, bastante complejas, en tanto que los clasificadores basados en redes neuronales, k-vecinos más próximos, máquinas de vectores soporte y análisis discriminante, lineal o cuadrático, son los que más aparecen en la literatura científica relacionados con la clasificación de especies de madera.

En cuanto a los resultados obtenidos, en la mayor parte de los casos recogidos se han obviado por su poca relevancia, al estar trabajándose con una sola especie, o un solo

tronco de árbol, o tratar de clasificar defectos, algo que está fuera del objetivo de este trabajo.

Por tanto, este es el momento de definir con qué materiales, equipos y paquetes informáticos se va a trabajar y qué métodos y procedimientos van a ser utilizados en la consecución de los objetivos planteados en esta tesis, partiendo de la preparación de las muestras y pasando por la extracción de características, su reducción por métodos diferentes y la utilización de diferentes clasificadores. En aquellos casos que se considere adecuado, se aclararán o definirán los conceptos necesarios para dejar expedito el camino del entendimiento de la labor que se realiza en cada momento.

3.1. Materiales y equipos.

En este apartado se indican los materiales utilizados en el trabajo, junto con los equipos y los programas informáticos que se han empleado en el desarrollo de la tesis.

3.1.1. Material de trabajo.

En principio se ha dispuesto de chapas de madera de 10 x 10 cm, aproximadamente, correspondientes a 42 especies diferentes, proporcionadas por el Departamento de Economía y Gestión Forestal de la E.T.S.I. de Montes de la Universidad Politécnica de Madrid. Todas las chapas se han obtenido con un proceso de corte y secado análogo, y no han sido sometidas a ningún tipo de lijado o tratamiento posterior a su corte.

Las distintas especies, su denominación científica y las siglas que se han asignado en esta tesis, como etiqueta de identificación, a cada una de ellas se muestran en el cuadro 3.1. Ha de añadirse que se dispone de cuatro muestras iniciales de cada una de las especies, a excepción de la especie *Sipo* de la que sólo se tienen dos.

Siglas	Especie	Denominación científica
AFR	Afromosia	<i>Pericopsis elata</i> (Harms) van Meeuwen
ANT	Antiaris	<i>Antiaris africana</i> Engl.
ARCE	Arce europeo	<i>Acer pseudoplatanus</i> L.
BIL	Bilinga	<i>Nauclea diderrichii</i> (De Wild. & T.Durand) Merrill
BOSC	Bossé claro	<i>Guarea laurentii</i> De Wild.
BOSO	Bossé oscuro	<i>Guarea thompsonii</i> Sprague et Hutch
BUB	Bubinga	<i>Guibourtia arnoldiana</i> (De Wild. & T.Durand) J. Léonard
CAOA	Caoba africana	<i>Khaya ivorensis</i> A. Chev.
CAS	Castaña	<i>Castanea sativa</i> Miller.
CED	Cedro	<i>Cedrela odorata</i> L.
CER	Cerezo	<i>Prunus avium</i> L.
CERA	Cerezo americano	<i>Prunus serotina</i> Ehrh.
DOU	Doussié	<i>Afzelia africana</i> Smith
ETI	Etimoé	<i>Copaifera salikounda</i> Heckel
EUC	Eucalipto	<i>Eucalyptus globulus</i> Labill.
FRE	Fresno	<i>Fraxinus excelsior</i> L.
HAY	Haya	<i>Fagus sylvatica</i> L.
IRO	Iroko	<i>Milicia excelsa</i> (Welw.) CC Berg
KOT	Koto	<i>Pterygota macrocarpa</i> K. Schum
LIM	Limba	<i>Terminalia superba</i> Enger. & Diels.
MAN	Mansonia (Bete)	<i>Mansonia altissima</i> A. Chev.
MAP	Maple (Arce)	<i>Acer saccharum</i> Marsh
MOA	Moabi	<i>Baillonella toxisperma</i> Pierre
MON	Mongoy	<i>Guibourtia ehie</i> (A. Chev.) J. Léonard
MUK	Mukali	<i>Aningeria robusta</i> (A. Chev.) Aubr. & Pellegr.
MUKM	Mukali mallado	
NOGA	Nogal americano	<i>Juglans nigra</i> L.
NVE	Nvero	<i>Lovoa trichilioides</i> Harms.

Continúa en la página siguiente.

Siglas	Especie	Denominación científica
PALR	Palo rosa (Tipuana)	<i>Tipuana tipu</i> (Benth.) Kuntze
PER	Peral	<i>Pyrus communis</i> L.
PINI	Pino insignis	<i>Pinus radiata</i> D. Don
PINM	Pino melis	<i>Pinus palustris</i> Mill.
PINP	Pino pinaster	<i>Pinus pinaster</i> Ait.
PINS	Pino silvestre	<i>Pinus sylvestris</i> L.
ROBA	Roble americano	<i>Quercus alba</i> L.
ROBE	Roble europeo	<i>Quercus robur</i> L.
SAP	Sapelli	<i>Entandrophragma cylindricum</i> Harms
SAPM	Sapelli moaré	
SIP	Sipo	<i>Entandrophragma utile</i> (Dawe & Sprague) Sprague
TUL	Tulipífero	<i>Liriodendron tulipifera</i> L.
UKO	Ukola	<i>Tieghemella africana</i> Pierre
WEN	Wengue	<i>Millettia laurentii</i> De Will.

Cuadro 3.1: Especies que componen el estudio.

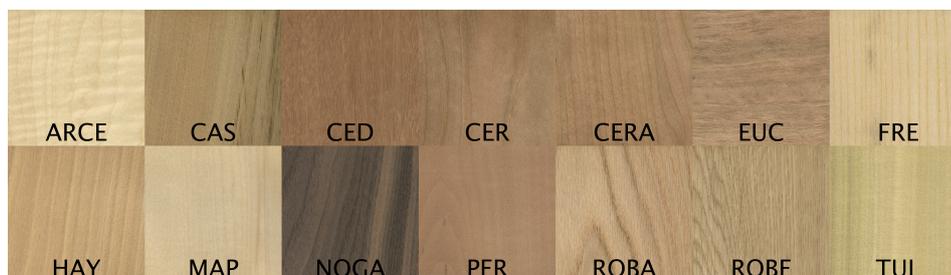
Estas especies se pueden clasificar en maderas coníferas (*Pino insignis*, *Pino melis*, *Pino pinaster* y *Pino silvestre*), frondosas (*Arce europeo*, *Castaño*, *Cedro*, *Cerezo*, *Cerezo americano*, *Eucalipto*, *Fresno*, *Haya*, *Maple*, *Nogal americano*, *Peral*, *Roble americano*, *Roble europeo* y *Tulipífero*) y tropicales (*Afromosia*, *Antiaris*, *Bilinga*, *Bossé claro*, *Bossé oscuro*, *Bubinga*, *Caoba africana*, *Doussié*, *Etimoé*, *Iroko*, *Koto*, *Limba*, *Mansonia*, *Moabi*, *Mongoy*, *Mukali*, *Mukali mallado*, *Nvero*, *Palo Rosa*, *Sapelli*, *Sapelli moaré*, *Sipo*, *Ukola* y *Wengue*), mostrándose en la figura 3.1 las especies de madera utilizadas en la tesis agrupadas según esta clasificación.

3.1.2. Equipos informáticos.

Para la realización de los diferentes trabajos vinculados con esta tesis se han utilizado indistintamente un ordenador portátil con procesador Intel Core i3 a 2,4 GHz y sistema



(a) Maderas coníferas.



(b) Maderas frondosas.



(c) Maderas tropicales.

Figura 3.1: Maderas utilizadas en la tesis.

operativo Windows 7 y un ordenador de sobremesa con procesador Intel Core i5 a 3,2 GHz y sistema operativo Linux, ambos con una capacidad de memoria RAM de 4.096 MB y disco duro de 500 MB, haciéndose necesario el uso de unidades externas de almacenamiento (discos duros o memorias “flash”) para el correspondiente trasvase de ficheros.

En el proceso de digitalización de la madera se ha utilizado un escáner plano *Epson Perfection V750 Pro*, cuyas principales características se muestran en el cuadro 3.2 [95].

Especificaciones	
Resolución óptica	6.400 ppp
Resolución de impresión	50-6.400 (paso de 1 ppp), 9.600, 12.800 ppp
Profundidad de color	48 bits (entrada y salida)
Sensor óptico	Matriz de CCD
Fuente luminosa	Lámpara fluorescente de cátodo frío
Método de escaneado	Documento fijo y carro móvil
Velocidad de escaneado	12 s/página

Cuadro 3.2: Características del escáner utilizado.

3.1.3. Programas informáticos.

En cuanto a los programas informáticos específicos utilizados, en las fases de preparación de las muestras y extracción de características se ha utilizado *ImageJ 1.47* [96], en tanto que en las fases de reducción de características y clasificación se ha trabajado con el programa *Weka 3.7.11* [97], ambos de dominio público, programados en Java, de los que a continuación se hace una breve descripción.

ImageJ.

ImageJ es un programa de análisis y procesado de imágenes desarrollado por Wayne Rasband para los Institutos Nacionales de Salud de Estados Unidos [98] y diseñado como una arquitectura abierta que permite la ampliación de sus funcionalidades mediante plugins Java y macroinstrucciones, aplicables a imágenes individuales o a conjuntos de imágenes incluidas en una misma carpeta [99]. Puede ejecutarse como una aplicación empujada en un navegador web o como aplicación de escritorio en cualquier ordenador que disponga de una máquina virtual *Java 5* o superior, pudiendo descargarse para cualquier sistema operativo y estando disponible el código fuente gratuitamente.

Con una corta curva de aprendizaje, *ImageJ* admite el análisis, procesado, visualización y edición de imágenes de 8, 16 y 32 bits en formatos TIFF, PNG, GIF, JPEG, DICOM, BMP, PGM y FIT, pudiendo visualizarse múltiples imágenes en una única ventana, siempre que sean del mismo tamaño y tengan la misma profundidad binaria, de manera que una operación de análisis o procesado se pueda aplicar a todas las imágenes del conjunto visualizado.

En la figura 3.2 se muestra el interfaz de usuario de *ImageJ*, donde se disponen, de arriba a abajo, los siguientes elementos:

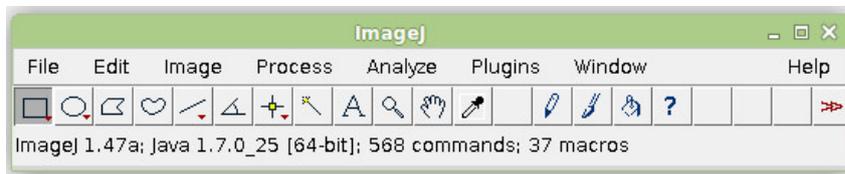


Figura 3.2: Barra de herramientas de ImageJ.

- *Menús de comandos.* Incluye todos los comandos disponibles, estando organizado en ocho menús desplegables que permiten realizar operaciones básicas con ficheros (*File*), la edición, operaciones de dibujo y configuración general (*Edit*), la conversión, transformación geométrica y modificación de imágenes (*Image*), el procesado de imagen, incluyendo operaciones aritméticas entre imágenes (*Process*), la medida de estadísticos y la representación de histogramas y perfiles (*Analyze*), la creación, edición y gestión de “plugins”, “scripts” y macroinstrucciones (*Macros*), la selección y gestión de las ventanas abiertas (*Window*) y la documentación e información sobre la versión (*Help*).
- *Barra de herramientas,* que contiene las que permiten hacer selecciones, dibujos, ampliaciones o desplazamientos sobre la imagen seleccionada.
- *Barra de estado.* Muestra las características del píxel sobre el que se encuentra el cursor o, cuando se ejecuta alguna acción el tiempo transcurrido y la velocidad de procesamiento. Haciendo clic en esta barra se muestran las versiones del programa y de Java, la memoria utilizada y la disponible.

- *Barra de progreso*, ausente en esta imagen, está situada en la parte derecha de la barra de estado y muestra el progreso de las operaciones que se están ejecutando.

Weka.

Por su parte, *Weka - Waikato Environment for Knowledge Analysis-*, programa desarrollado por la universidad neozelandesa de Waikato [100], se presenta como un paquete que contiene herramientas de visualización y algoritmos para el aprendizaje automático y la minería de datos [101]. Implementado en Java, Weka puede ser utilizado en cualquier plataforma y contiene una amplia colección de técnicas de minería de datos (preprocesado, agrupamiento, clasificación, regresión, visualización y selección) con una interfaz gráfica de usuario (figura 3.3) que hace que su uso sea fácil.



Figura 3.3: Interfaz de usuario de Weka.

En esta interfaz de usuario aparecen cuatro posibilidades de acceso a las funcionalidades del programa que son las siguientes:

- *Explorer* da acceso a las principales componentes de trabajo, como son el preprocesado (importación de datos y preprocesamiento de los mismos mediante filtros para transformarlos y eliminar registros o atributos) la clasificación (uso de clasificadores y análisis de regresión a los datos, estimando su exactitud y permitiendo visualizar parámetros de error o el propio modelo generado), asociación (acceso a reglas que intentan identificar las interrelaciones entre los atributos de los datos), agrupamiento (incorpora los algoritmos de “*clustering*”), selección de atributos (proporciona algo-

ritmos para la identificación de los atributos más predictivos del conjunto de datos) y visualización (muestra una matriz de puntos dispersos que pueden seleccionarse para su análisis en detalle).

- *Experimenter* permite comparar diferentes algoritmos aplicados al mismo conjunto de datos.
- *Knowledge Flow* implementa las mismas funciones que el explorador de forma gráfica, mediante bloques que se arrastran y sueltan, correspondientes a cada una de ellas.
- *Simple CLI* o consola que permite acceder a todas las funcionalidades de Weka.

3.2. Metodología.

En este apartado se desarrolla la metodología seguida en el desarrollo de la tesis, que, partiendo del material disponible, comprende las fases enumeradas a continuación:

1. Digitalización de las chapas de madera.
2. Preparación de las muestras.
3. Extracción de características.
4. Reducción de características.
5. Clasificación de las muestras.
6. Análisis de resultados.

A continuación se desarrolla cada uno de estos apartados, describiendo en cada caso las bases teóricas que sustentan cada una de las técnicas o procedimientos a utilizar.

3.2.1. Digitalización de la madera.

A partir de las chapas de madera disponibles (véase el apartado 3.1.1) se procede a la obtención de las imágenes digitales correspondientes a las mismas. Para ello se va a utilizar el escáner cuyas características técnicas se recogen en el cuadro 3.2. Ahora bien, previo al proceso de digitalización se hace necesaria la calibración del equipo de adquisición para que la respuesta sea la misma ante un mismo objeto con independencia de la posición o el momento en que se realice, lo que puede no ocurrir debido a la no uniformidad de la iluminación en toda la superficie del escáner, la degradación de los colores, la obtención de imágenes borrosas o la presencia de líneas verticales en la imagen escaneada.

3.2.1.1. Proceso de calibración.

Para evitar la variabilidad de resultados en el proceso de digitalización se utiliza la carta IT8.7/2 de EGM Laboratoris Color (ver figura 3.4), de 20 x 27 cm, utilizada en escaneados reflectantes, tal y como establece el estándar ANSIT8.7/2 [102].

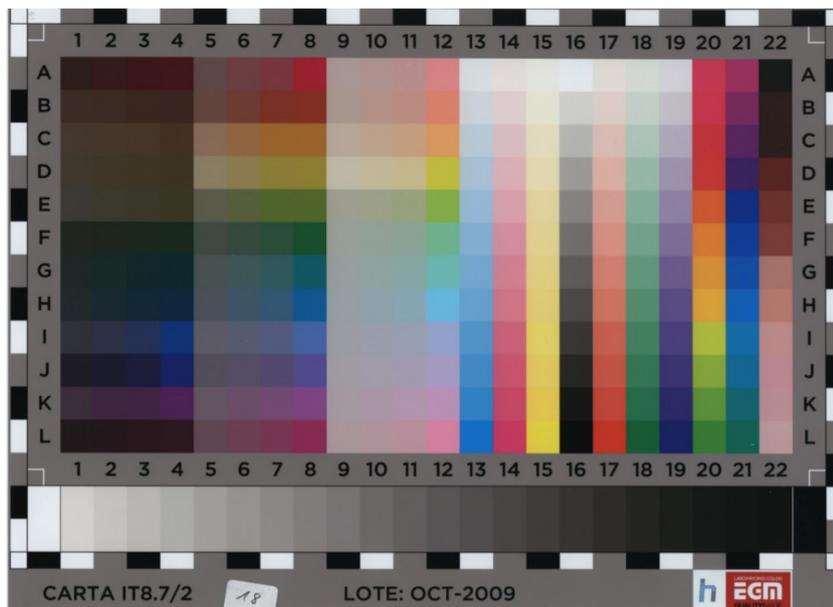


Figura 3.4: Carta IT8.7/2 de EGM Laboratoris Color.

El proceso a seguir para calibrar el escáner, recogido en las instrucciones de uso dadas por el fabricante de la carta [103], consiste en escanear la carta en modo RGB con máxima

profundidad de color y una gamma de 2,2, desactivando todos los ajustes automáticos y el filtro de enfoque. A continuación se guarda la imagen obtenida en formato TIFF y se abre con Photoshop.

Las esquinas (cuadros blancos) se utilizan para corregir la posible falta de homogeneidad en la iluminación, en tanto que el ajuste del color y de la escala de grises se realiza haciendo coincidir los números de la parte superior dentro del cuadro observado al ampliar la imagen, consiguiéndose así una carta corregida como la mostrada en la figura 3.5, en la que se han separado cada uno de los colores y grises de la carta original. Este resultado se archiva en formato TIFF sin comprimir, como base para la creación del perfil de color del escáner.

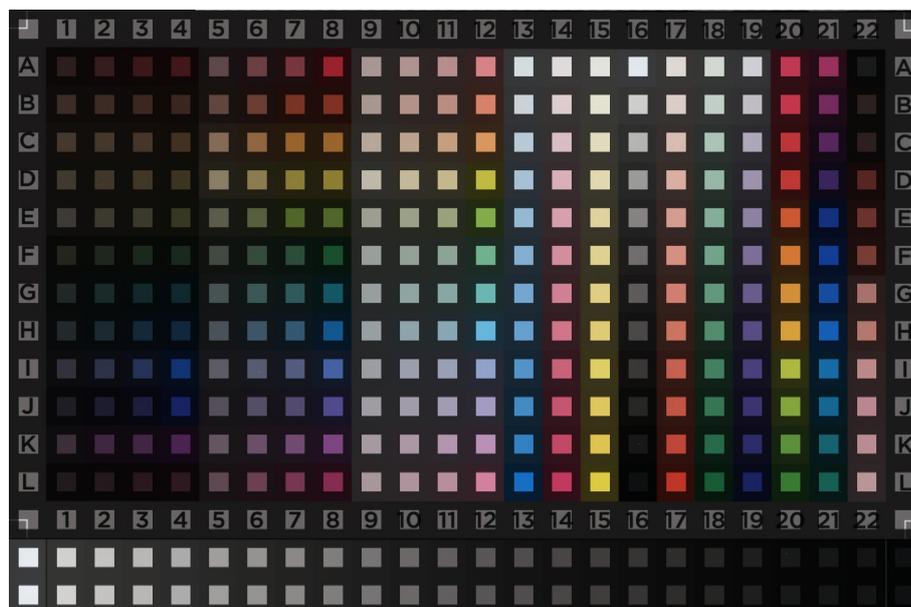


Figura 3.5: Carta base del perfil de color del escáner Epson Perfection V750 Pro.

Utilizando el software proporcionado con la carta, a partir de la imagen anterior se define el perfil de color o perfil ICC (*International Color Consortium*) correspondiente al escáner y que una vez guardado se podrá incrustar en cada toma de imágenes realizada con éste, corrigiéndose en cada operación de escaneado el color y la iluminación capturados.

3.2.1.2. Obtención de las imágenes digitales.

Una vez calibrado el escáner se procede a la digitalización de las chapas de madera, asegurando la limpieza de la cama de cristal sobre la que se depositan para evitar que existan residuos de escaneados anteriores que puedan alterar el resultado de la digitalización.

La adquisición de las imágenes se realiza con las siguientes especificaciones:

- Profundidad de color: 24 bits.
- Resolución: 1.200 ppp.
- Administrador de imágenes en color en origen (escáner): perfil ICC obtenido en la calibración del Epson Perfection V750 Pro.
- Administrador de imágenes en color en destino (visualización): perfil correspondiente a Adobe RGB.
- Imagen digital recortada a un tamaño de 4.096 x 4.096 píxeles.
- Formato de salida de las imágenes: TIFF.

La profundidad de color de 24 bits es la que da lugar al “color verdadero”, en el que en cada píxel la información codificada dedica un octeto a representar la intensidad luminosa de cada una de las componentes roja, verde y azul de la imagen, pudiendo así representarse 16,7 millones de colores diferentes.

La selección de una resolución de 1.200 ppp se justifica al ser lo suficientemente pequeña como para que las imágenes no sean muy pesadas computacionalmente y, por otra parte, para asegurar que aunque las imágenes con las que se va a trabajar son de pequeñas dimensiones contengan un número de píxeles suficientemente grande para los análisis que se van a realizar sobre ellas.

El recorte de las chapas de madera en el proceso de digitalización, reduciendo sus dimensiones a 4.096 x 4.096 píxeles, tiene como finalidad evitar posibles defectos producidos en los procesos de cortado o manipulación de las mismas, como se refleja en el ejemplo representado en la figura 3.6.



Figura 3.6: Chapa de madera con defecto de origen en el lado izquierdo.

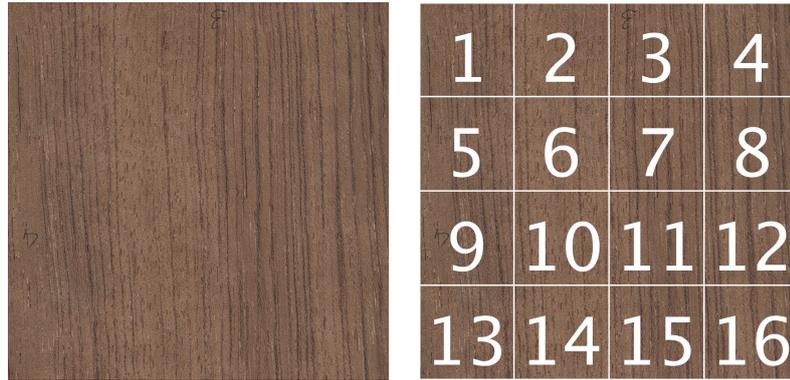
Por último, el formato elegido a la hora de guardar las imágenes digitales hace que a éstas no se les aplique ningún tipo de compresión, evitando así cualquier posible pérdida de información.

3.2.2. Preparación de las muestras.

El número de imágenes digitales obtenidas, una por cada chapa de madera, es demasiado bajo para los fines planteados en esta tesis, dado que en la clasificación se estaría trabajando con un máximo de cuatro patrones por especie, de los cuales al menos tres serían necesarios para el entrenamiento de los clasificadores y uno sólo para la prueba de los mismos, que daría lugar a errores de clasificación en esta fase de prueba del 0 ó del 100 %, según la muestra se clasifique correcta o incorrectamente.

Se podrían utilizar todos los patrones para entrenamiento y volver a aplicar el modelo de clasificador obtenido sobre el mismo conjunto de patrones, pero esta solución provoca una falta de generalización en el clasificador.

Se hace necesario un mayor número de patrones por especie, optándose por dividir cada una de las imágenes en 16 no solapadas de 1.024 x 1.024 píxeles (ver figura 3.7). Estas nuevas imágenes tienen unas dimensiones físicas aproximadas de 2,17 x 2,17 cm, suficientes para contener diferentes componentes estructurales del corte de madera, evitando una excesiva homogeneidad de las imágenes, lo que ocurriría si la división se realizase a tamaños menores.



(a) Imagen escaneada.

(b) División de la imagen.

Figura 3.7: Ejemplo de imagen original y el orden en su división.

Esta partición se realizará con *ImageJ* mediante un procesamiento por lotes a todas las imágenes, archivadas en una misma carpeta, utilizando la secuencia de comandos:

1. Abrir imagen: **File** → **Open**.
2. Dividir imagen en 16 y apilarlas: **Image** → **Stacks** → **Tools** → **Montage to Stack**.
3. Guardar secuencia de imágenes: **File** → **SaveAs** → **Image Sequence**.

Las imágenes se archivan en formato TIFF con el nombre de la original seguido del número correspondiente a la partición dado en la figura 3.7.

3.2.3. Extracción de características.

Partiendo de un planteamiento inicial basado en el estudio de las características estadísticas de las imágenes patrón relacionadas con su fotometría, textura y dimensión fractal, se presenta la incertidumbre de qué influencia puedan tener en la clasificación de especies las componentes de color de la imagen, o las propiedades que presenta la madera a diferentes escalas.

Para poder valorar esta influencia antes de la extracción de características para cada imagen digitalizada se van a obtener las imágenes correspondientes a sus componentes de color y las representativas de las distintas escalas, estas últimas utilizando técnicas

de mejora de las imágenes inspiradas en la microscopía de contraste por interferencia diferencial o en la búsqueda de las estructuras periódicas subyacentes de la madera.

3.2.3.1. Imágenes componentes de color.

En el espacio colorimétrico RGB una imagen de color está formada por la combinación píxel a píxel de tres imágenes en escala de grises, denominadas canales, que se corresponden con los colores primarios rojo, verde y azul de la imagen original.

Con *ImageJ* la descomposición de imágenes de color en sus componentes primarias de color se conseguirá utilizando los comandos que, para cada una de las imágenes, se indican a continuación:

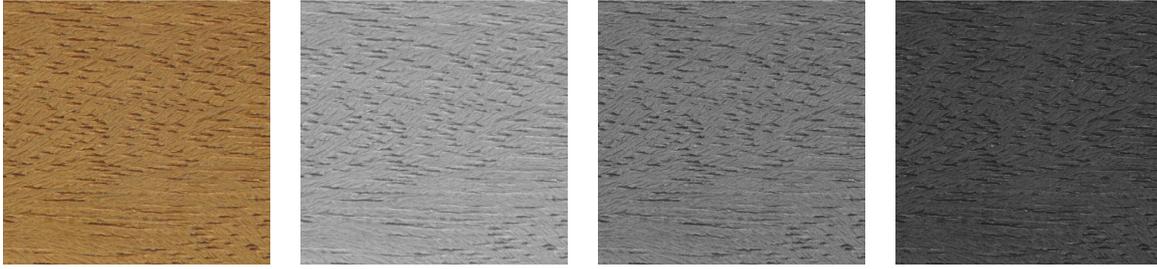
1. Abrir imagen: **File** → **Open**.
2. Convertir la imagen formando una ‘pila’ de sus componentes primarias de color:
Image → **Type** → **RGB Stack**.
3. Separar las componentes apiladas: **Image** → **Stacks** → **Stacks to Images**.
4. Archivar las componentes: **File** → **SaveAs**.

Las componentes se archivan en directorios diferentes, en el orden azul, verde y rojo, en formato TIFF.

En este proceso se ha tenido en cuenta la curva de sensibilidad relativa del ojo [105], que da lugar a que la luminancia de la imagen (Y) sea la suma ponderada de las luminancias de las imágenes roja (R), verde (G) y azul (B),

$$Y = 0,30 \cdot R + 0,59 \cdot G + 0,11 \cdot B \quad (3.1)$$

En la figura 3.8 se muestra un ejemplo de una de las imágenes originales y sus tres componentes de color.



(a) Imagen digitalizada. (b) Componente roja. (c) Componente verde. (d) Componente azul.

Figura 3.8: Ejemplo de imagen digitalizada y sus componentes primarias de color.

3.2.3.2. Imágenes “microscópicas”.

Para la obtención de estas imágenes se va a utilizar el método de mejora de imágenes inspirado en la microscopía de contraste por interferencia diferencial y basado en la utilización de fractales que ha sido planteado por Rojas *et al.* [106].

Los fractales son estructuras básicas que se repiten a diferentes escalas; los utilizados en este trabajo se generan aplicando la operación lógica XOR píxel a píxel a dos imágenes mutuamente ortogonales consistentes en un gradiente de niveles de gris.

En la figura 3.9 se representa un fractal generado de esta forma de 64 x 64 píxeles junto con su histograma, cuya característica principal es que todas sus componentes tienen la misma amplitud.

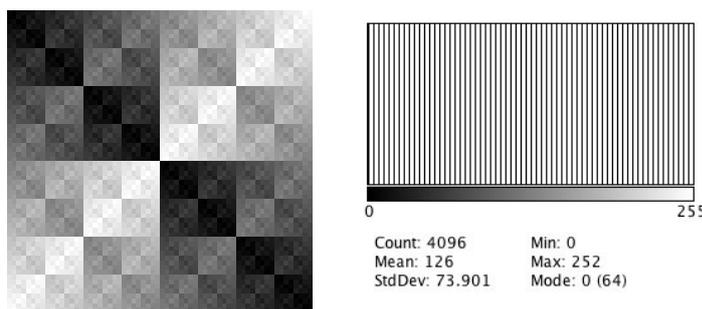


Figura 3.9: Ejemplo de fractal de 64 x 64 píxeles generados con XOR y su histograma.

Teniendo en cuenta que el fractal actúa como un filtro paso bajo utilizado para eliminar el ruido de las imágenes, cuya frecuencia de corte es mayor a medida que el orden del fractal es menor, se ha elegido un fractal de 5 x 5 píxeles, al no ser el ruido en nuestro caso un factor crítico.

Utilizando *ImageJ*, para obtener la imagen “microscópica” de cada imagen patrón se genera el fractal de 5 x 5 píxeles, resultando la imagen de la figura 3.10, que se archiva como una imagen en formato texto.

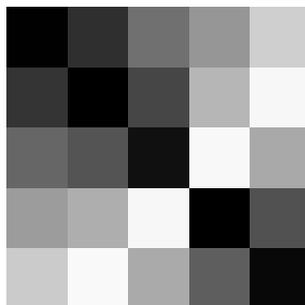


Figura 3.10: Fractal de 5x5 píxeles.

A continuación, se convoluciona la imagen patrón con el fractal y con las réplicas de éste giradas, sin interpolación, en pasos de 45° , con lo que se consigue la eliminación de componentes de alta frecuencia en cada una de las direcciones marcadas por el giro del fractal.

Para cada imagen, en *ImageJ* se sigue la secuencia de comandos siguiente:

1. Abrir la imagen: **File** → **Open**.
2. Convolucionar la imagen con cada uno de los fractales rotados: **Process** → **Filters** → **Convolve**. Los fractales se utilizan en formato texto y corresponden a los mostrados en la figura 3.11.
3. Convertir cada una de las imágenes resultantes a un formato de 32 bits: **Image** → **Type** → **32 bits** de *ImageJ*.

A continuación se procede al cálculo del cociente entre imágenes. Para ello, la imagen resultante de la convolución con el fractal girado 45° se divide entre la imagen de la convolución de la imagen con el fractal a 0° ; el resultado de la convolución con el fractal girado 90° se divide entre la de convolución con el fractal a 0° y entre la del fractal girado 45° ; y así sucesivamente, dividiendo cada una de las imágenes entre cada una de las obtenidas con menores giros del fractal. Para ello se utiliza el comando **Process** →

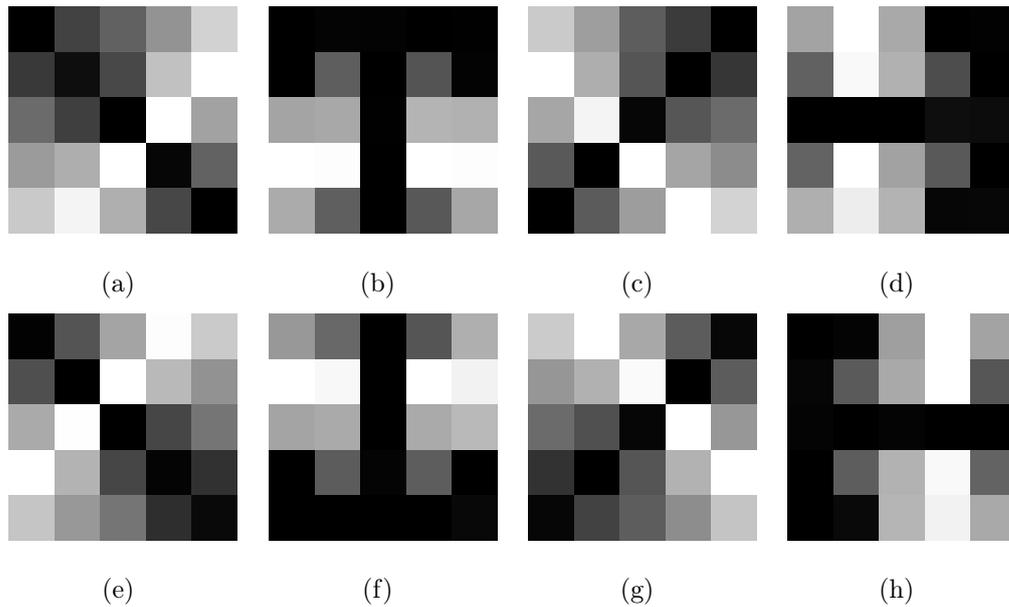


Figura 3.11: Fractal 5x5 (a), y sus réplicas giradas 45° (b), 90° (c), 135° (d), 180° (e), 225° (f), 270° (g) y 315° (h).

Image Calculator de *ImageJ*, obteniéndose una secuencia de reconstrucción pseudotridimensional de la imagen original como la que se representa en la figura 3.12, formada por las imágenes cociente obtenidas en formato de 32 bits.

Las imágenes de esta secuencia se convierten a imágenes de 8 bits con `Image → Type → 8 bits`, apilándose la secuencia con el comando `Image → Stacks → Images to Stack` en un “stack”.

Determinando el valor medio, píxel a píxel, de la luminancia de las imágenes apiladas, con `Image → Stacks → Z Project`, se genera una imagen pseudotridimensional, como la de la figura 3.13, que recrea el efecto de visualización del gradiente del índice de refracción en microscopía de contraste por interferencia diferencial.

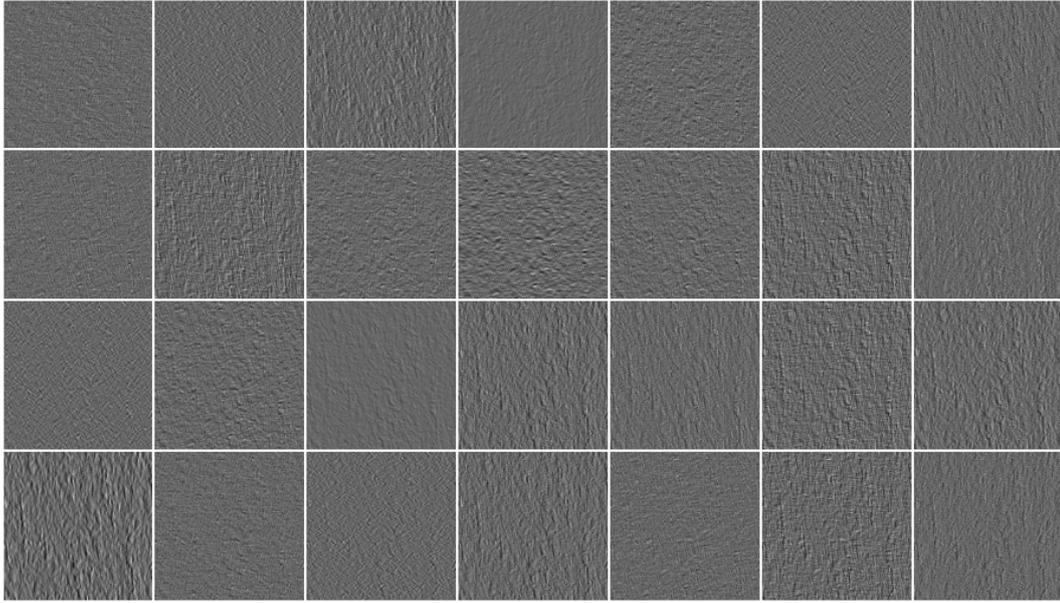


Figura 3.12: Secuencia de reconstrucción 3D con incrementos angulares de 45° .

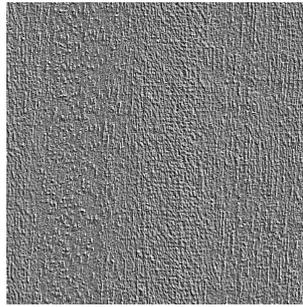


Figura 3.13: Ejemplo de imagen “microscópica” obtenida a partir de una imagen digitalizada.

3.2.3.3. Imágenes de patrones periódicos subyacentes.

Las superficies de la madera pueden ser muy complejas, dependiendo su rugosidad de muchos factores naturales o artificiales, pero las características anatómicas subyacentes de las mismas son más simples.

Para extraer los patrones de periodicidad ocultos en una imagen, Rojas *et al.* [91] han desarrollado una técnica, inspirada en la interferometría de moteado estelar, que utiliza la autocorrelación iterativa de imágenes de madera iluminadas con un láser, de manera que en cada iteración se filtran las características periódicas menos relevantes, siendo

suficientes tres o cuatro iteraciones para conseguir un patrón muy claro de la superficie iluminada.

La utilización de esta técnica ha tenido que modificarse porque la iluminación proporcionada por el escáner no crea en la imagen una señal de alta frecuencia como es el moteado generado por la iluminación láser.

La autoconvolución está relacionada con la densidad de potencia espectral mediante la transformada de Fourier y se utiliza habitualmente para detectar información periódica oculta por altos niveles de ruido. En nuestro caso, al no ser el ruido un parámetro crítico, se necesita generar un cierto grado de interferencia sobre la imagen para aplicar el método propuesto.

La forma de conseguir la interferencia buscada consiste en utilizar un fractal de 5 x 5 píxeles que se convoluciona con la imagen original, resultando una imagen que es convolucionada iterativamente con el mismo fractal girado sin interpolación en pasos de 90°. Se obtiene así una imagen borrosa, con cierto grado de interferencia.

A partir de aquí, los patrones periódicos subyacentes se consiguen dividiendo la imagen patrón original entre esta imagen con interferencia, realizando así las componentes de alta frecuencia.

Una vez mejorado el contraste de la imagen cociente (`Process` → `Enhance Contrast`) sin saturación, se somete el resultado a un proceso de autoconvolución iterativa con cuatro iteraciones, en cada una de las cuales se utiliza el comando `Process` → `FFT` → `FD Math` de *ImageJ*, eligiendo la operación de convolución.

En la figura 3.14 se recoge el proceso en forma de secuencia de imágenes. En la fila superior, y de izquierda a derecha se tienen la imagen original, la imagen obtenida tras la convolución iterativa con el fractal y sus réplicas giradas, la imagen cociente entre la original y la convolucionada iterativamente con el fractal y ésta última mejorada en contraste. En la fila inferior se muestra, también de izquierda a derecha el resultado de cada iteración de la autoconvolución de la imagen mejorada en contraste.

Analizando el patrón periódico resultante, a partir de la representación de niveles de una de sus líneas horizontales (figura 3.15), se aprecia que se corresponde con una señal

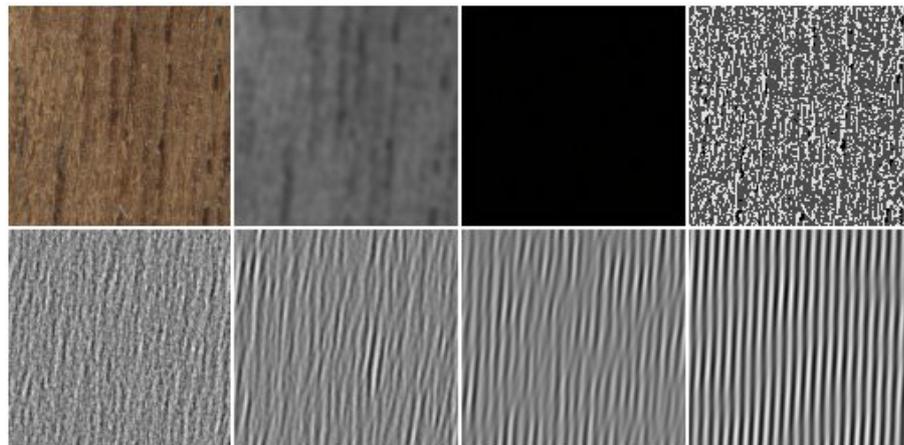


Figura 3.14: Proceso de obtención de los patrones periódicos subyacentes a una imagen digitalizada.

periódica de 12 píxeles de período dentro de una envolvente cuyo período es de 216 píxeles.

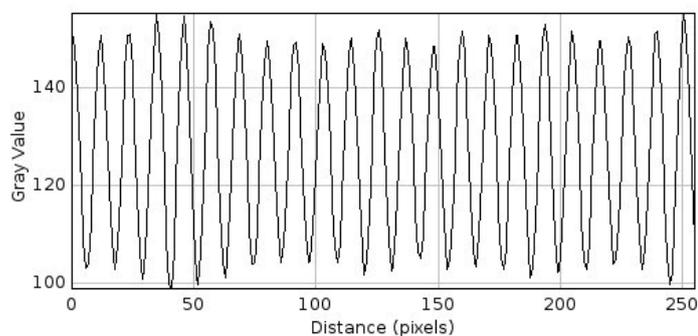


Figura 3.15: Visualización de la periodicidad de la del patrón subyacente.

3.2.3.4. Obtención de características.

En este punto, cada patrón se compone de un conjunto de seis imágenes: imagen obtenida al dividir la imagen digitalizada en 16 partes, componentes roja, verde y azul de la misma, imagen pseudotridimensional y patrón periódico subyacente.

En la figura 3.16 se representa el conjunto de imágenes de trabajo correspondientes a una de las particiones de 1.024 x 1.024 píxeles obtenidas tras la división de su co-

respondiente chapa de madera digitalizada, conjunto para el que se van a extraer las características correspondientes a la fotometría, textura y dimensión fractal.

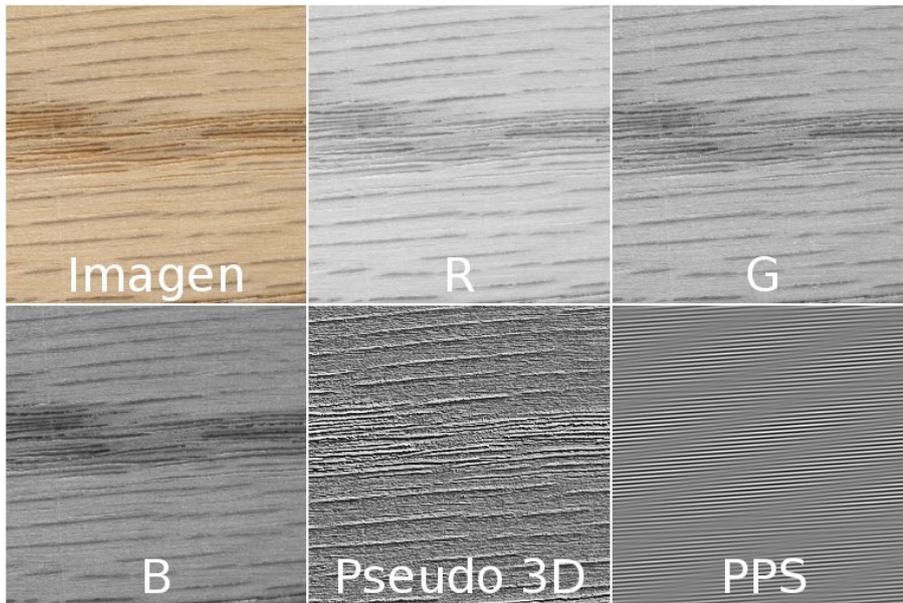


Figura 3.16: Conjunto de imágenes que forman el patrón correspondiente a una muestra digitalizada.

Características fotométricas.

Las características fotométricas se relacionan con los niveles de gris de los píxeles de la imagen. Para cada una de las imágenes se obtiene una serie de valores numéricos, correspondientes a parámetros estadísticos de medida de posición, dispersión o forma.

Los parámetros de medida de posición son valores de la variable caracterizados por la posición que ocupan en el rango de valores de aquélla. Dentro de estos se van a realizar las siguientes medidas de tendencia central:

- *Media.* Proporciona el gris medio de la imagen en una escala de 0 a 225, en la que el 0 corresponde al negro y el 255 al blanco. Siendo n el número de píxeles de la imagen y x_i el nivel de gris del píxel i , la media se determina como

$$\bar{x} = \frac{\sum_i x_i}{n} \quad (3.2)$$

- *Moda*. Es el valor de gris de mayor frecuencia, el más repetido en la imagen.
- *Mediana*. Determina el valor por encima y por debajo del cual se tiene el mismo número de niveles de gris en la imagen. Como se dispone de un número par de píxeles en la imagen, la mediana se determina como la media de los valores de la variable que ordenados ocupan las posiciones $\frac{n}{2}$ y $\frac{n}{2} + 1$.

Las medidas de dispersión dan idea de la heterogeneidad de los datos, midiéndose en este caso los siguientes parámetros absolutos:

- *Desviación típica*. Mide la media cuadrática de las distancias entre los datos y la media, es decir, el grado de dispersión de los niveles de gris con respecto a su valor medio. Para n valores característicos $\{x_1, x_2, \dots, x_n\}$ de media \bar{x} , se calcula mediante la expresión

$$\sigma = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n}} \quad (3.3)$$

- *Rango*. Dado por la diferencia entre los valores de los niveles de gris máximo y mínimo de la imagen.

Por último, las medidas de forma dan idea de cómo se distribuyen los valores de la variable comparados con una distribución normal. Dentro de éstas se van a medir los siguientes:

- *Asimetría*. Momento estadístico de tercer orden que mide la asimetría en la distribución de grises de la imagen. También denominado *sesgo*, este parámetro está dado por la expresión

$$\text{Asimetría} = \frac{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^3}{\left(\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2\right)^{3/2}} \quad (3.4)$$

- *Curtosis*. Mide el grado de aflación o los picos de la distribución de los grises de la imagen con respecto a la distribución normal. Este momento estadístico de cuarto orden, está dado por

$$\text{Curtosis} = \frac{\sum_{i=1}^n (x_i - \bar{x})^4}{n\sigma^4} - 3 \quad (3.5)$$

Para la obtención de los valores correspondientes a estas características de cada una de las imágenes se utiliza el comando **Analyze** → **Measure** de *ImageJ*.

Características de textura.

En las imágenes de objetos reales se observa que no existen regiones de intensidades uniformes, sino que se producen variaciones de intensidad que forman unos patrones repetidos a los que se denomina “textura visual”. Estos patrones repetidos, en el caso de la madera, son el resultado de las propiedades físicas de su superficie -aspereza u orientación de las fibras- o de la distribución del color de la misma. La textura en imágenes, por tanto, se define como una función de la variación espacial de las intensidades de los píxeles evaluadas por sus valores de gris [107].

Los parámetros de textura se van a obtener utilizando la matriz de co-ocurrencia de niveles de gris (GLCM-*Gray Level Co-occurrence Matrix*), en la que sus elementos representan las frecuencias relativas de las distintas combinaciones de los niveles de gris en una imagen. A partir de la matriz de co-ocurrencia se pueden obtener una serie de estadísticos relacionados con la textura de la imagen, entre los que se encuentran los estadísticos de primer orden, calculados a partir de la imagen original, no considerando las relaciones de vecindad entre píxeles, los estadísticos de segundo orden, en los que se consideran las relaciones entre dos píxeles vecinos, y los estadísticos de tercer orden o superiores, en los que se consideran las relaciones entre tres o más píxeles.

La matriz GLCM es una matriz cuadrada, estando definido el número de filas o de columnas por el nivel de cuantificación (niveles de grises) de la imagen, y simétrica con respecto a la diagonal, en la que el elemento $P_{ij} = P(i, j)$ representa la probabilidad de que los pares de píxeles para los que los valores de intensidad i y j estén separados una determinada distancia [108].

Los parámetros de textura se miden a 0, 45, 90 y 135° sobre imágenes monocromas con 256 niveles de gris [108, 109], y son los que se relacionan a continuación:

- *Contraste*. Medida de la cantidad de variaciones locales en los tonos de gris de la imagen, calculado como

$$\text{Contraste} = \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} P_{ij} (i - j)^2 \quad (3.6)$$

- *Homogeneidad local* o *momento diferencia inversa*. Es lo opuesto al contraste, estando dada esta medida por

$$\text{Homogeneidad local} = \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} \frac{P_{ij}}{1 + (i - j)^2} \quad (3.7)$$

Obsévese que en este caso intervienen los elementos de la diagonal de la matriz de co-ocurrencia, en tanto que en el contraste no.

- *Segundo momento angular*. Mide la regularidad en la ordenación de los píxeles y está dado por la expresión

$$\text{Segundo momento angular} = \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} P_{ij}^2 \quad (3.8)$$

- *Energía*. Parámetro que mide la uniformidad de la imagen, estando dado por la raíz cuadrada del segundo momento angular.
- *Entropía*. Parámetro cuyo valor indica cuan compleja es la imagen, a partir del grado de desorden de los píxeles o de utilización de los diferentes niveles de gris. Se calcula como

$$\text{Entropía} = - \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} P_{ij} \cdot \log_2(P_{ij}) \quad (3.9)$$

- *Prominencia*. Da idea de la relevancia de de los distintos píxeles con respecto a los que les rodean. Se determina como

$$\text{Prominencia} = \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} P_{ij} (i + j - \mu_i - \mu_j)^4 \quad (3.10)$$

- *Sombra*. Parámetro relacionado con la tonalidad de la imagen, medido como

$$\text{Sombra} = \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} P_{ij} \left(i + j - \frac{\mu_i + \mu_j}{2} \right)^3 \quad (3.11)$$

- *Varianza*. Medida de las desviaciones cuadráticas con respecto a la media, dada por

$$\text{Varianza} = \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} P_{ij} (i - \mu)^2 \quad (3.12)$$

- *Correlación.* Medida que, en el análisis de texturas, se puede considerar independiente de las anteriores, por lo que resulta de gran interés en combinación con ellas. Indica la dependencia lineal entre los niveles de grises de los píxeles vecinos, y está dada por

$$\text{Correlación} = \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} \frac{(i - \mu_i)(j - \mu_j)}{\sqrt{\sigma_i^2 \sigma_j^2}} \quad (3.13)$$

En todas estas expresiones los valores medios de las filas o las columnas de la matriz están dados por

$$\begin{aligned} \mu_i &= \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} i \cdot P_{ij} \\ \mu_j &= \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} j \cdot P_{ij} \end{aligned}$$

Las medidas de todos estos parámetros en cada una de las imágenes se realizan para la mínima distancia entre píxeles, y se han obtenido con el plugin `Image CorrelationJ` 1o de *ImageJ*.

Dimensión fractal.

La dimensión fractal es una cantidad estadística que indica, a partir de una zona de la imagen de un tamaño determinado, cuántas veces aparece esa misma estructura en la imagen completa, al ir reduciendo la escala de la imagen.

Si se dispone de una imagen cuyo tamaño es la unidad en un espacio euclídeo de dimensión D , y se reduce el tamaño de la imagen en un factor $1/\epsilon$ en cada dirección espacial, el número de imágenes autosimilares necesarias para cubrir la imagen original será $N = \epsilon^D$. Al ir reduciendo el tamaño de la imagen por el citado factor una y otra vez, se llega a la definición matemática de la dimensión fractal,

$$D = \lim_{\epsilon \rightarrow 0} \frac{\log N(\epsilon)}{\log \frac{1}{\epsilon}} \quad (3.14)$$

donde $N(\epsilon)$ es el número de imágenes de tamaño ϵ necesarias para cubrir la imagen completa [110], precisándose la conversión de la imagen a binaria.

Representando logarítmicamente el número de imágenes autosimilares en función del factor de reducción (figura 3.17), la pendiente de la recta de regresión resultante da el valor de la dimensión fractal.

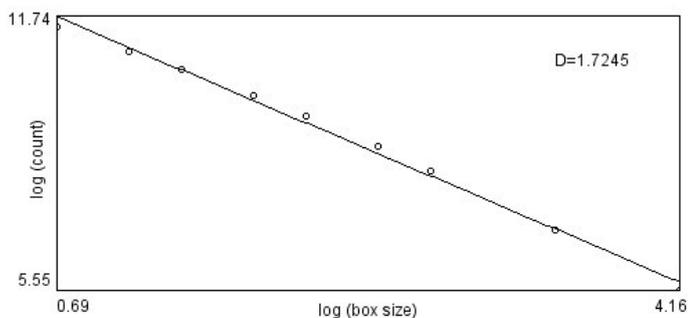


Figura 3.17: Interpretación gráfica de la dimensión fractal.

El comando `Analyze` → `Tools` → `Fractal Box Count` ha sido el utilizado en *ImageJ* para determinar los valores de la dimensión fractal.

3.2.4. Selección de características.

Antes de extraer las características planteadas en el apartado 3.2.3.4 para las seis imágenes de cada patrón, teniendo en cuenta la ecuación 3.1 que indica que la luminancia de la imagen digitalizada es una combinación lineal de las correspondientes a sus componentes de color, parece evidente que existe redundancia de información, ya que disponiendo del valor de la luminancia de tres de ellas se puede deducir fácilmente la correspondiente a la cuarta. Considerando siempre la presencia de la imagen digitalizada, bastará con mantener dos de las componentes de color. Realizando la correlación entre las series de las componentes de color se puede determinar de qué imagen se va a prescindir.

El resultado final da lugar a un vector de características formado por cinco imágenes - imagen digitalizada, imágenes de dos componentes de color, imagen pseudotridimensional e imagen de los patrones periódicos subyacentes-, obteniendo de cada una de ellas los valores de sus parámetros fotométricos y de textura y su dimensión fractal.

Se tiene así un patrón de una dimensión demasiado grande si se quiere construir un modelo estadístico de clasificación simple, al poder existir variables redundantes o grupos de variables correlacionados.

Antes de proceder a la reducción de la dimensión de los vectores de características se propone sustituir la desviación típica por el *coeficiente de variación de Pearson*, medida de variación relativa dada por el coeficiente entre la desviación típica y la media, por ser éste un parámetro que permite comparar el grado de dispersión de dos distribuciones con medias diferentes. Cuanto mayor es el coeficiente mayor es la dispersión y menor la representatividad de la media.

Para las características de textura se realizará la media direccional de cada una de ellas, quitando así la influencia debida al posible giro de la madera en el proceso de digitalización. Con ello, los valores obtenidos para un parámetro a partir de las matrices de co-ocurrencia a 0, 45, 90 y 135° se transforman en un único valor, media de los cuatro iniciales. En el caso de la correlación esto no sería necesario, dado que su valor es el mismo, independientemente del ángulo de cálculo de la matriz de co-ocurrencia.

Las características finales que forman la matriz de datos se normalizan para evitar que en la selección o en la propia clasificación los resultados estén sesgados por el propio peso de los valores de las variables. De igual forma se aleatoriza la posición de los vectores de características en la matriz de datos, lo que va a permitir la generalización de los procesos. La normalización se realiza dividiendo el valor por el máximo que puede adoptar su correspondiente característica, en tanto que para la aleatorización se utiliza la secuencia de comandos `Filter → Unsupervised → Instance → Randomize` de *Weka*.

3.2.4.1. Reducción del vector de características.

Tras la normalización y aleatorización se utiliza un algoritmo de selección de características de entre los descritos a continuación: selectores basados en la correlación y selector basado en el análisis de componentes principales.

Si se quiere realizar una selección de atributos utilizando el programa *Weka*, se necesita trabajar con datos de entrada en formato ARFF, acrónimo de *Attribute-Relation File Format*, cuya estructura es la siguiente:

- *Cabecera*. Define el nombre de la relación y su formato es el siguiente:

```
@relation <nombre de la relación>
```

- *Declaraciones de atributos*. Se declaran en esta sección los atributos que formarán el archivo junto con su tipo. La sintaxis, considerando que el nombre del atributo es del tipo cadena de texto (“string”) y el tipo puede ser “numeric” (números reales), “integer” (números enteros), “date” (fechas) o “string”, es la siguiente:

```
@attribute <nombre del atributo 1><tipo>
```

```
@attribute <nombre del atributo 2><tipo>
```

```
...
```

```
@attribute class {clase 1,clase 2,...,clase m}
```

- *Selección de datos*. Se declaran los datos que componen la relación separando con comas los atributos y con saltos de línea las relaciones. La sintaxis es la siguiente:

```
@data
```

```
dato 1, dato 2,..., dato n
```

Este tipo de ficheros se puede generar en un editor de texto o de notas, cuidando de guardarlo con la extensión **.arff**, o se puede convertir con un filtro desde un fichero en el que se disponga de los atributos separados por comas en un registro por línea, es decir, se pueden convertir a ARFF ficheros con formato CSV.

Selector basado en la correlación.

Existen diferentes algoritmos para la selección de características que se basan en la realización de una búsqueda en el espacio de subconjuntos de características que se ve afectada por el punto de comienzo de la búsqueda, seleccionado un punto del espacio de

subconjuntos de características desde el que comenzar la búsqueda. Ésta puede realizarse comenzando sin características y añadiendo éstas sucesivamente (búsqueda hacia adelante), o comenzando con todas las características y quitándolas una a una (búsqueda hacia atrás), o bien comenzando en un punto medio e ir desplazándose hacia el exterior del subconjunto (búsqueda bidireccional).

Las estrategias heurísticas de búsqueda, si bien no aseguran que se vaya a encontrar el subconjunto óptimo de éstas, suelen conseguir resultados aceptables, por lo que lo importante es la organización de la búsqueda, teniendo en cuenta lo prohibitivo de realizar una búsqueda exhaustiva.

La estrategia de evaluación de los subconjuntos de características es otra de las diferencias entre los diferentes algoritmos, existiendo diferentes propuestas, entre las que se encuentran las que evalúan el factor de mérito de los subconjuntos, el sesgo hacia determinados subconjuntos o la validación cruzada para medir la precisión del subconjunto final.

Por último, en estos algoritmos habrá de tenerse en cuenta el criterio de parada de la búsqueda, que depende de la estrategia de evaluación y puede consistir en dejar de añadir o quitar características cuando no se mejora la evaluación del subconjunto, continuar evaluando el subconjunto de características mientras no se degrade el parámetro de evaluación o continuar evaluando subconjuntos de características hasta el final y seleccionar el que mejores resultados obtenga.

El selector basado en la correlación [111], implementado en *Weka* con la denominación `CfsSubsetEval`, es un filtro simple que clasifica subconjuntos de características de acuerdo a una correlación basada en la siguiente función de evaluación heurística,

$$M_S = \frac{k\bar{r}_{fc}}{\sqrt{k + k(k-1)\bar{r}_{ff}}} \quad (3.15)$$

donde M_S es el heurístico del subconjunto S que contiene k características, \bar{r}_{fc} es el valor medio de la correlación característica-clase y \bar{r}_{ff} el valor medio de la correlación entre características. Esta función de evaluación está sesgada hacia subconjuntos que contienen características que guardan una relación directa con la clase y no están correlacionadas entre sí. Las características con una baja correlación con la clase se ignoran, y las muy

relacionadas con una o más del resto de características se consideran redundantes y se eliminan, e impone una clasificación de los subconjuntos de características en el espacio de búsqueda de todos los posibles subconjuntos de características.

En este tipo de selectores el primer paso consiste en crear una matriz con las correlaciones característica-clase y característica-característica a partir de los datos de entrenamiento. A continuación se busca el mejor subconjunto de características, habiéndose elegido dos estrategias de búsqueda para este selector: “selección hacia adelante” y “mejor primero”.

La selección hacia adelante, `LinearForwardSelection` en *Weka*, comienza sin características y las va añadiendo una una, desde el punto de arranque, hasta que no es posible obtener una evaluación más alta.

Por su parte, el mejor primero es una estrategia de búsqueda voraz que comienza con un conjunto vacío de características y genera todas las expansiones posibles de características individuales. Se elige el subconjunto con la evaluación más alta y se le van añadiendo características. Si esta expansión de características no presenta mejoras se pasa la búsqueda al siguiente mejor subconjunto no expandido y se repite el proceso. Cuando termina la búsqueda se devuelve el mejor subconjunto encontrado. La condición de parada de este algoritmo es que un determinado número de subconjuntos de características totalmente expandidos no introduzcan mejoras. Esta estrategia en *Weka* se denomina `BestFirst`.

Análisis de componentes principales.

El análisis de componentes principales, debido a Hotelling [112], tiene como objetivo la reducción de dimensionalidad del vector de características construyendo un nuevo vector en el que las nuevas características sean combinaciones lineales de las originales.

Con esta técnica se identifican las posibles variables latentes o no observadas que generan los datos y se transforman las variables originales, en general correladas, en nuevas variables incorreladas entre sí, generándose el llamado conjunto de componentes principales.

La obtención de las componentes principales se realiza buscando la combinación lineal de las variables que maximiza la variabilidad. Para ello, se parte de los datos disponibles de los n patrones con vectores característicos de dimensión p , dados en la matriz \mathbf{X} ,

$$\mathbf{X} = \begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1p} \\ x_{21} & x_{22} & \cdots & x_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{np} \end{pmatrix} \quad (3.16)$$

y a partir de ella se obtiene la matriz \mathbf{S} de varianzas y covarianzas, dada por

$$\mathbf{S} = \frac{1}{n-1} \mathbf{X}^T \mathbf{X} \quad (3.17)$$

Las combinaciones lineales de las variables observadas que se buscan han de ser incorreladas y sus varianzas progresivamente decrecientes, adoptando la forma

$$C_i = a_{i1}x_1 + a_{i2}x_2 + \cdots + a_{ip}x_p \quad (3.18)$$

donde a_{ij} representan los pesos de las distintas variables ($j = 1, \dots, p$) sobre la componente i , cumpliéndose que C_1 será la componente que explica la mayor variabilidad de los datos, C_2 , que será ortogonal a C_1 , explica la mayor parte de la variabilidad restante, y así sucesivamente, observándose que existen tantas componentes principales como variables originales.

En forma matricial, se puede escribir

$$\mathbf{C} = \mathbf{X}\mathbf{A} \quad (3.19)$$

donde $\mathbf{A}'\mathbf{A} = \mathbf{I}$. Por tanto, calcular las componentes principales equivale a aplicar una transformación ortogonal \mathbf{A} a las variables \mathbf{X} para obtener unas nuevas variables \mathbf{C} incorreladas entre sí. Esta operación puede interpretarse como elegir unos nuevos ejes coordenados que coincidan con los “ejes naturales” de los datos.

Para obtener estas componentes principales se calculan los valores propios (autovalores) o raíces características, $\lambda_1, \lambda_2, \dots, \lambda_n$ de la matriz \mathbf{S} mediante

$$|\mathbf{S} - \lambda\mathbf{I}| = 0 \quad (3.20)$$

siendo sus vectores asociados (autovectores) los \mathbf{a}_i que cumplen la condición

$$(\mathbf{S} - \lambda \mathbf{I}) \mathbf{a}_i = 0 \quad (3.21)$$

Los términos λ_i son reales, al ser la matriz \mathbf{S} simétrica, y positivos por definirse \mathbf{S} como positiva. Por ser \mathbf{S} simétrica, si λ_j y λ_k son dos raíces distintas, sus vectores asociados son ortogonales.

Las componentes principales son nuevas variables que conservan la variabilidad inicial, siendo la varianza acumulada la suma de la proporción de varianza de las componentes analizadas, y la proporción de varianza de la componente j el cociente entre su varianza (autovalor asociado al autovector que lo define) y la suma de las varianzas de las variables originales (suma de los autovalores), $\lambda_j / \sum_{i=1}^p \lambda_i$.

Se puede hacer una representación de los autovalores de las diferentes componentes calculadas, conformando el denominado gráfico de sedimentación (figura 3.18), de manera que se permite considerar aquellas componentes que se encuentran antes del punto de inflexión en el que se pasa de una pendiente pronunciada a una pendiente mínima.

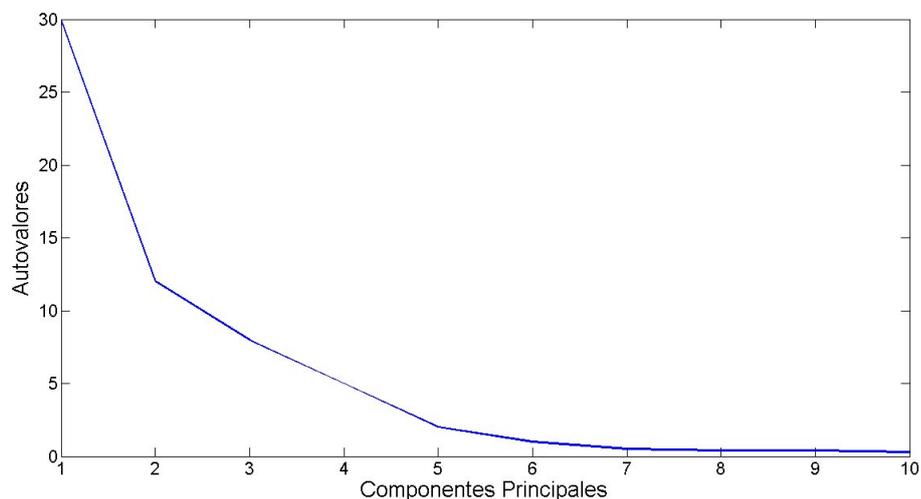


Figura 3.18: Gráfico de sedimentación.

Otro procedimiento habitual para elegir las componentes principales más significativas consiste en aplicar el criterio de Kaiser modificado, según el cual se eligen aquellas componentes cuyo autovalor asociado sea mayor que 0,7.

La función `PrincipalComponents` de *Weka* permite hacer una reducción de características aplicando el análisis de componentes principales.

Para aplicar este método de reducción de características es necesario comprobar a priori si es adecuado o no utilizarlo. Para ello existen diferentes pruebas, entre las que se encuentra el *índice KMO* o de Kaiser-Meyer-Olkin de adecuación de la muestra [115], calculado como

$$KMO = \frac{\sum_i \sum_{j \neq i} r_{ij}^2}{\sum_i \sum_{j \neq i} r_{ij}^2 + \sum_i \sum_{j \neq i} a_{ij}^2} \quad (3.22)$$

donde r_{ij} es el coeficiente de correlación entre las variables i y j y a_{ij} el coeficiente de correlación parcial entre las mismas variables. Estos coeficientes miden la correlación existente entre las variables una vez que se elimina la influencia del resto de variables sobre ellas o efectos de los factores comunes, que al ser eliminados hacen que a_{ij} , que representa la correlación entre los dos factores únicos de las dos variables, sea teóricamente nula.

Cuando hay correlación entre las dos variables, caso en el que resulta apropiado utilizar el análisis de componentes principales, estos coeficientes deben tener valores próximos a 0, por lo que el índice KMO se aproximaría a 1. En caso contrario, los valores del índice próximos a 0 desaconsejarían el uso de este análisis. Normalmente está aceptado que para valores del índice por debajo de 0,5 la adecuación muestral es inaceptable, entre 0,5 y 0,6 se considera mala, entre 0,6 y 0,7 regular, aceptable entre 0,7 y 0,8, buena entre 0,8 y 0,9 y excelente para valores del índice KMO por encima de 0,9.

3.2.5. Clasificación.

En la clasificación de patrones se pueden aplicar múltiples sistemas basados en la teoría de la detección. Ahora bien, cuando el problema consiste en clasificar materiales anisótropos como la madera, cuyas propiedades no siguen un modelo estadístico gaussiano, la utilización de clasificadores bayesianos no da buenos resultados. Revisando los modelos que se han venido utilizando en este tipo de aplicaciones, los sistemas basados en discriminantes lineales o cuadráticos tampoco se muestran eficaces. Utilizar redes neuronales, por la complejidad que presenta el dimensionamiento de las capas ocultas, sobre todo cuando se tiene gran número de clases a clasificar, junto con el tiempo que requiere la realización

de un gran número de iteraciones, tampoco es muy aconsejable. Por ello, para el proceso de clasificación se han elegido los siguientes clasificadores:

- *Clasificador basado en árboles de decisión*: árbol **C4.5**.
- *Clasificador con entrenamiento basado en casos*: k vecinos más próximos (**kNN**).
- *Clasificadores combinados con aleatorización*: bosque aleatorio (**Random Forest**) y bosque de rotación (**Rotation Forest**).
- *Clasificador lineal*: máquina de vectores soporte (**SVM**).

3.2.5.1. Árbol C4.5.

El árbol de decisión C4.5, que en *Weka* se implementa como el árbol J48, pertenece a los árboles de decisión por inducción descendente, que basan su estrategia en una secuencia de preguntas que se realizan en un recorrido desde la raíz a las hojas a través de nodos.

En cada nodo se prueba una característica y se baja por la rama asociada al valor que de ésta tiene el patrón, repitiéndose el proceso nodo a nodo hasta llegar a una hoja en la que se encuentra la predicción de clase del patrón. Cada rama del árbol es una restricción sobre los valores de la característica probada.

El algoritmo C4.5 se construye a partir del método de Hunt [116], que consiste en disponer de un conjunto de datos de entrenamiento pertenecientes a k clases previamente clasificados. Si este conjunto contiene uno o más casos, todos pertenecientes a una única clase el árbol de decisión para los datos de entrenamiento es una hoja que identifica a la clase; si el conjunto no contiene ningún caso el árbol es una hoja cuya clase se determina con información que no pertenece a los datos; por último, si el conjunto de entrenamiento tiene datos pertenecientes a varias clases se hace una división de los datos en subconjuntos de casos que tiendan a pertenecer a una sola clase.

En la construcción del árbol ha de tenerse en cuenta que el principio básico es la simplicidad [114], por lo que las decisiones han de dar lugar a un árbol simple y compacto, con pocos nodos. Por ello, se suelen utilizar árboles binarios, en los que el nodo raíz divide el conjunto completo de patrones de entrenamiento en subconjuntos que acceden a los

nodos sucesivos, dando lugar a distintas particiones que finalizan en las hojas del árbol, fijando el diseñador el número de decisiones a tomar en cada nodo.

Este mismo principio de simplicidad implica que en cada nodo se busca una característica de prueba que haga que los datos que alcanzan el siguiente nodo sean lo más puros posibles, estableciéndose así la impureza de los nodos, que básicamente se mide mediante la *ganancia de información*, métrica basada en la entropía,

$$i(N) = - \sum_j P(\omega_j) \cdot \log_2 P(\omega_j)$$

donde $P(\omega_j)$ es el porcentaje de patrones que en el nodo pertenecen a la clase ω_j , y que mide la heterogeneidad de los datos, determinando la cantidad mínima de bits necesarios para codificar los elementos del nodo.

La ganancia de información da idea de la reducción de la entropía del conjunto al clasificar los patrones en un determinado nodo. Si S es el conjunto de patrones clasificados en c clases, A es la característica de búsqueda del nodo y S_V son los patrones que tienen el valor v para la característica A , la ganancia de información se define como

$$\text{Ganancia}(S, A) = H(S) - \sum_{v \in \text{Valores}(A)} \frac{|S_V|}{|S|} H(S_V) \quad (3.23)$$

indicando el número de bits de ahorro cuando se conoce el valor de la característica.

Uno de los problemas a resolver en este tipo de clasificadores es encontrar una decisión óptima para un nodo, lo que depende de la forma de decidir, que normalmente se basa en las funciones de impureza. Así, como la optimización se realiza para cada nodo, no hay garantías de que decisiones óptimas locales en cada uno de los nodos lleven a una decisión óptima global, ni de que después del entrenamiento el árbol generado sea el menor posible. Con el fin de reducir la impureza en las hojas hasta valores razonables, se pueden seguir dividiendo los datos, teniendo en cuenta que la impureza no va a alcanzar el valor nulo.

Otro problema estriba en cuándo parar la ramificación del árbol durante el entrenamiento. Si se construye el árbol hasta que cada hoja tenga una impureza mínima aparece un problema de sobreajuste que no permitirá la generalización del clasificador, y que puede deberse a errores en los patrones de entrenamiento o a una insuficiente cantidad de éstos.

Para evitar el error que se produce por sobreajuste del árbol, se puede parar la construcción del árbol antes de que clasifique perfectamente todos los patrones o se puede construir el árbol e intentar una poda a posteriori.

La determinación del tamaño óptimo del árbol se puede realizar utilizando un conjunto de patrones de validación, aplicando técnicas estadísticas para evaluar si es conveniente agregar o quitar nodos o usando una medida de complejidad para codificar los datos y el tamaño del árbol, parando cuando esta medida se minimiza. Para evitar que la ramificación se pare demasiado pronto se utiliza la poda una vez que el árbol se ha desarrollado completamente, hasta que las hojas tienen una impureza mínima, momento en el que todas las hojas vinculadas a un nodo que al ser eliminadas produzcan una mejora en los aciertos sobre el conjunto de validación son eliminadas (podadas), sustituyendo cada uno de estos nodos por la valoración más común de los patrones que engloba.

3.2.5.2. k-vecinos más próximos.

Este tipo de clasificadores, conocidos por su acrónimo en inglés kNN (*k-Nearest Neighbours*) asignan al patrón a clasificar examinando la clase correspondiente a los k patrones más próximos en distancia y decidiendo por mayoría [117].

La implementación en *Weka* se denomina IBk y puede emplear entre los criterios de distancia los basados en la distancia euclídea, distancia de Chebyshev, distancia Manhattan o distancia de Minkowski, dadas por

$$D_{euclídea}(P, Q) = \sqrt{(p_1 - q_1)^2 + (p_2 - q_2)^2 + \dots + (p_n - q_n)^2} \quad (3.24)$$

$$D_{Chebyshev}(P, Q) = \max_i (|p_i - q_i|) \quad (3.25)$$

$$D_{Manhattan}(P, Q) = \sum_{i=1}^n |p_i - q_i| \quad (3.26)$$

$$D_{Minkowsky}(P, Q) = \left(\sum_{i=1}^n |p_i - q_i|^p \right)^{1/p}, \quad p \geq 1 \quad (3.27)$$

En la fase de entrenamiento se almacenan los vectores de características y las etiquetas de las clases correspondientes. En la fase de clasificación, cada patrón se representa por su vector de características y se calcula la distancia entre éste y los vectores almacena-

dos, seleccionando los k vectores más cercanos, asignándole al patrón a clasificar la clase correspondiente a la mayoritaria de los vectores cercanos.

El método supone que la mejor clasificación la proporcionan los vecinos más cercanos, lo que se determina utilizando el vector de características completo del patrón, existiendo el problema de que si existen muchas características irrelevantes éstas dominen en la clasificación. Este sesgo se puede corregir dándole mayor importancia a las características más relevantes mediante un peso que pondera las distancias de cada característica, si bien antes de asignar estos pesos se recomienda identificar y eliminar las características que se consideren irrelevantes.

La eficacia del clasificador recae en la adecuada elección de k , lo que depende de los datos disponibles, y se puede seleccionar utilizando una optimización de uso. Ha de tenerse en cuenta que no siempre un valor grande de k produce los mejores resultados, ya que, si bien reduce el efecto de ruido en la clasificación, puede crear límites entre clases parecidas.

3.2.5.3. Bosque aleatorio.

Este clasificador, `RandomForest` en la terminología de *Weka*, es una técnica de agregación desarrollada por Leo Breimman en 2001 [118] y utiliza varios árboles de clasificación aleatorios. Realizando un muestreo con reemplazo sobre los datos de entrenamiento se forma un conjunto de datos que puede contener patrones duplicados del conjunto de entrenamiento, pero que contiene el mismo número de patrones que éste. Usando esta técnica normalmente un tercio de los patrones de entrenamiento no se encuentran en este conjunto y con ellos se forma otro conjunto de patrones excluidos.

Para cada uno de los árboles se elige un número aleatorio de características y con ellas se construye el árbol, mediante algoritmos estándar, con la mayor profundidad posible y sin poda, repitiéndose el proceso para desarrollar múltiples árboles aleatorios de entrenamiento. La clasificación de cada patrón se hace por votación, con las salidas de los distintos árboles, a las que se asigna el mismo peso.

Una vez que se han construido los diferentes árboles se utilizan los patrones excluidos para probar tanto los árboles individuales como el bosque entero formado por todos

ellos. La media de los patrones mal clasificados por todos los árboles se conoce como error estimado de los patrones excluidos y predice el rendimiento del clasificador para los patrones de prueba, pudiendo utilizarse este dato para ponderar los pesos de los árboles de clasificación individuales en el bosque aleatorio.

Un parámetro importante en este tipo de clasificadores es la importancia de las variables. El algoritmo de entrenamiento analiza cada característica y revela la importancia que ésta tiene en la predicción de una correcta clasificación. Para ello se cuentan los patrones del conjunto de excluidos que se clasifican correctamente. A continuación se permutan aleatoriamente los valores de una de las características, formando un nuevo conjunto de patrones de prueba para el bosque, en busca de la clasificación correcta de los mismos. El promedio de los patrones correctamente clasificados en todos los árboles del bosque con este nuevo conjunto de prueba permite determinar un valor de importancia para la característica permutada. Esta información se puede utilizar para filtrar características innecesarias en el proceso de clasificación.

3.2.5.4. Bosque de rotación.

Este clasificador utiliza técnicas de empaquetado y su funcionamiento básico consiste en disponer una matriz de datos de entrenamiento y otra que contenga las clases correspondientes. Seleccionando los clasificadores a utilizar, para cada uno de ellos se divide aleatoriamente el conjunto de características en un número de subconjuntos previamente definido, que pueden ser disjuntos o solapados [119].

Del subconjunto de características seleccionado para un clasificador se elige un subconjunto de clases y se comienza el entrenamiento con un 75 % de los patrones, a los que se realiza un análisis de componentes principales para obtener los correspondientes coeficientes en forma matricial, formando la matriz de rotación en la que cada uno de los elementos de la diagonal se corresponde con el vector de coeficientes obtenido para cada subconjunto de características. Para obtener el conjunto de entrenamiento del clasificador se reorganizan las columnas de la matriz de rotación para que coincidan con las del conjunto inicial de características. Se construye así el nuevo conjunto de entrenamiento para

el clasificador multiplicando la matriz inicial de características por la matriz de rotación reorganizada.

Para un patrón dado se determina la probabilidad asignada por el clasificador a la hipótesis de pertenencia del mismo a una determinada clase, calculando el nivel de confianza para cada una de las clases como media de los obtenidos en los distintos clasificadores y se asigna al patrón la clase con mayor nivel de confianza.

Llamado *RotationForest* en *Weka*, que lo incluye en la categoría de metaclasificadores, aplica la selección de características mediante análisis de componentes principales.

3.2.5.5. Máquina de vectores soporte.

Conocidas como SVM (*Support Vector Machines*), las máquinas de vectores soporte tienen su origen en los trabajos sobre la teoría de aprendizaje estadístico y fueron introducidas en los años 90 [120], siendo un algoritmo de clasificación binario. Pertenecen a la categoría de clasificadores lineales por introducir separadores lineales o hiperplanos en el espacio original de las muestras o en un espacio transformado, según la separabilidad o no de las mismas.

El entrenamiento de estas máquinas tiene como objetivo encontrar un hiperplano de separación óptima entre clases, que se puede expresar como

$$\mathbf{w} \cdot \mathbf{x} + b = 0 \quad (3.28)$$

en la que el vector \mathbf{x} es el vector de características, \mathbf{w} es el vector que define al plano de separación y $b/\|\mathbf{w}\|$ es la separación entre el hiperplano y el origen de coordenadas en la dirección normal a \mathbf{w} .

Si los datos de entrenamiento son linealmente separables, se pueden encontrar dos fronteras que separen los datos, sin que existan puntos entre ellos, definidas como

$$y_i (\langle \mathbf{w} \cdot \mathbf{x}_i \rangle + b) \geq 1 \quad (3.29)$$

donde y_i adopta los valores -1 ó +1 según la clase a que pertenezca el patrón \mathbf{x}_i .

Para maximizar la separación entre las regiones de cada clase solo habrá que fijarse en los vectores de características que se encuentren en el límite de las mismas, vectores que se conocen como *vectores soporte*, diseñados para cumplir la relación 3.29.

Así, cada hiperplano encontrado para separar las clases es una combinación lineal de sus vectores soporte,

$$\mathbf{w} = \sum_{i=1}^n \alpha_i y_i \mathbf{x}_i \quad (3.30)$$

siendo α_i los multiplicadores de Lagrange que se definen positivos para los vectores soporte y nulos para el resto.

Si se quiere maximizar el margen sujeto a que los patrones estén bien clasificados habrá que minimizar $\|\mathbf{w}\|$. Estableciendo el lagrangiano

$$L(\omega, \alpha) = \frac{1}{2} \|\mathbf{w}\|^2 - \sum_{i=1}^n \alpha_i y_i (\langle \mathbf{w} \cdot \mathbf{x}_i \rangle + b - 1) \quad (3.31)$$

sujeto a $y_i (\langle \mathbf{w} \cdot \mathbf{x}_i \rangle + b - 1)$, surgen dos formulaciones diferentes:

1. *Formulación primaria*, en la que la minimización está sujeta a la condición para $\alpha_i \geq 1$ dada por $\alpha_i y_i (\langle \mathbf{w} \cdot \mathbf{x}_i \rangle + b) \geq 1$.
2. *Formulación dual*, que lleva a que el plano es la combinación lineal definida en la ecuación 3.30 de los patrones de entrada, pero sólo para los vectores soporte e imponiendo que la suma de los multiplicadores de Lagrange, contemplando el signo, es nula.

Cuando no se puede encontrar un hiperplano que separe limpiamente las clases, se pueden introducir variables de holgura ξ_i no negativas que midan el grado de clasificación errónea de los datos o pérdidas. Estas pérdidas se evalúan de forma distinta según los vectores soporte tipo L1-SVM que se encuentren en la frontera de su región,

$$L(\omega, \alpha) = \frac{1}{2} \|\mathbf{w}\|^2 + C \sum_{i=1}^l \xi_i \quad (3.32)$$

o los tipo L2-SVM que se encuentren fuera de la misma,

$$L(\omega, \alpha) = \frac{1}{2} \|\mathbf{w}\|^2 + C \sum_{i=1}^l \xi_i^2 \quad (3.33)$$

donde $0 < \alpha_i < C$ para los vectores soporte tipo L1-SVM, $\alpha_i = C$ para los tipo L2-SVM y $\alpha_i = 0$ para los que no son vectores soporte.

Si en el problema de clasificación los datos del conjunto de entrenamiento no son linealmente separables se recurre a transformarlos pasándolos a un espacio de dimensión superior en el que sí sean lineal o cuasi linealmente separables. En este nuevo espacio se podría aplicar la teoría del margen máximo y, por tanto, de las máquinas de vectores soporte, que permite encontrar un plano que separe con margen máximo.

Para no estar condicionados por el tipo de transformada a realizar, lo que redundaría en complicar las matemáticas a utilizar, la solución adoptada consiste en emplear una función de transformación o “kernel” al espacio de entrada, entre las que se encuentran las funciones de base radial,

$$K(\mathbf{x}, \mathbf{y}) = \langle \phi(\mathbf{x}), \phi(\mathbf{y}) \rangle = e^{-\gamma \|\mathbf{x} - \mathbf{y}\|^2} \quad (3.34)$$

o las polinomiales,

$$K(\mathbf{x}, \mathbf{y}) = \langle \phi(\mathbf{x}), \phi(\mathbf{y}) \rangle = (\gamma \langle \mathbf{x}, \mathbf{y} \rangle + c)^d \quad (3.35)$$

Antes de entrenar este clasificador se deberán conocer los valores de γ , d y C , que pueden variar entre 2^{25} y 2^{15} este último y entre 2^{-15} y 2^3 el primero, de manera que su ajuste lleva a procesos de construcción del clasificador de mucha duración.

Por ello se ha tomado la decisión de utilizar el paquete LibLINEAR [121], también disponible en *Weka*, que resuelve tanto la formulación primaria como la dual del problema de maximización del margen, mediante las funciones

$$L(\omega) = \frac{1}{2} \|\mathbf{w}\|^2 + C \sum_{i=1}^l (\text{máx}(0, 1 - \mathbf{w}^T \cdot \mathbf{x}_i)) \quad (3.36)$$

$$L(\omega) = \frac{1}{2} \|\mathbf{w}\|^2 + C \sum_{i=1}^l (\text{máx}(0, 1 - \mathbf{w}^T \cdot \mathbf{x}_i))^2 \quad (3.37)$$

no siendo estos clasificadores muy sensibles a la función de coste C . Si C es mayor que cierto valor, los modelos obtenidos tienen actuaciones similares.

3.2.5.6. El proceso de clasificación.

Tras la selección de los patrones, se procederá a la clasificación de las muestras. Para cada una de las matrices de datos resultantes del proceso de selección se va a realizar

su división en dos conjuntos diferentes. El primero de ellos, formado por el 75 % de los patrones, servirá como conjunto de entrenamiento de los clasificadores. El segundo, que contiene al resto de patrones, se utilizará como conjunto de prueba para el modelo generado durante el entrenamiento.

Cada uno de los clasificadores, a excepción del basado en un bosque rotatorio, será entrenado y probado con todos y cada uno de los conjuntos de datos seleccionados. Es decir, se combinarán todos los métodos de selección con todos los tipos de clasificadores elegidos. En el caso del bosque rotatorio los datos de entrada corresponderán a los patrones iniciales, antes de la selección de características, utilizando como filtro de proyección el basado en el análisis de componentes principales y como árbol el basado en el algoritmo C4.5.

En cada caso, antes de proceder a la clasificación de los patrones, utilizando los de entrenamiento se procederá a la optimización de los parámetros del clasificador, en aras a mejorar su eficiencia, lo que es una tarea tediosa, porque requiere múltiples entrenamientos del clasificador variando el parámetro a optimizar hasta encontrar cuándo el error de clasificación es mínimo.

Con los parámetros optimizados se procederá a realizar el entrenamiento del correspondiente clasificador. Para ello se utilizará la técnica de *validación cruzada*, con el fin de garantizar que los resultados obtenidos son independientes de la división entrenamiento/prueba que se haya hecho de los datos. Esta validación se hará con K iteraciones, de manera que los patrones de entrenamiento se dividan en K subconjuntos, utilizando uno de los subconjuntos como prueba y el resto como datos de entrenamiento. El proceso se repite iterativamente para cada uno de los subconjuntos de prueba. La media aritmética de los resultados de cada iteración permite obtener un resultado único. Este método, aunque computacionalmente es lento, es muy preciso, al evaluarse un clasificador a partir de K combinaciones de patrones de entrenamiento y prueba.

3.3. Análisis de resultados.

Una vez llevada a cabo la clasificación de los patrones, teniendo en cuenta las consideraciones realizadas en los apartados de este capítulo, se hará una comparación de los resultados obtenidos, en aras a determinar el que mejor comportamiento haya tenido.

En esta comparativa se tendrán en cuenta los siguientes parámetros:

- *Precisión*, o porcentaje de patrones clasificados correctamente.
- *Eficiencia*, o tiempo necesario para construir o usar el clasificador.
- *Complejidad*, o memoria necesaria para crear el esquema del modelo de clasificación.
- *Kappa estadística*. Refleja la concordancia entre varios observadores, pudiendo valer desde -1 (mayor grado de discordancia) a 1 (mayor grado de concordancia). En nuestro caso los observadores serían cada uno de los clasificadores generados durante la validación cruzada.
- *Área ROC*. Área bajo la curva ROC (*Receiver Operator Characteristics*). La curva ROC es una representación de los patrones clasificados correctamente en función de los mal clasificados. El área bajo la curva ROC se puede interpretar como la posibilidad de que un clasificador etiquete en su clase a un patrón elegido al azar.
- *Error cuadrático medio*. Promedio de los errores de estimación (diferencia entre la clase asignada y la real) al cuadrado.

Los resultados de esta comparativa servirán, en su caso, para determinar el conjunto selector de características-clasificador que mejor comportamiento tiene ante muestras de chapas de madera.

Capítulo 4

Resultados y discusión.

La metodología descrita en el capítulo 3 se puede resumir en una serie de pasos a seguir, orientados a la obtención de resultados que puedan ser analizados, como sigue:

1. Obtención de imágenes.
2. Construcción del vector de características.
3. Reducción del número de características.
4. Clasificación de patrones.
5. Comparativa de los clasificadores utilizados.

En este proceso se parte de las 42 especies de madera con las que se va a trabajar, que se distribuyen entre coníferas, frondosas y tropicales en los porcentajes mostrados en la figura 4.1.

Para aplicar la metodología descrita en el capítulo 3 se ha partido de la disponibilidad de 166 chapas de madera -4 por especie, salvo el *Sipo*-.

4.1. Obtención de imágenes.

La calibración realizada asegura una iluminación uniforme en la cama de adquisición y una corrección de los colores de la imagen obtenida. Además, la limpieza realizada antes

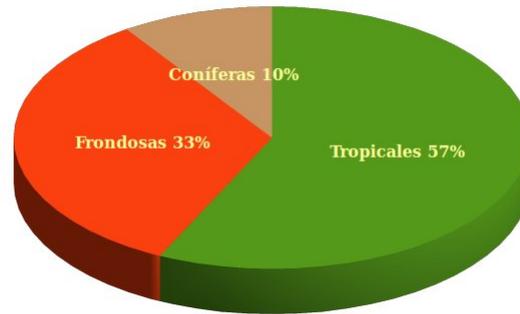


Figura 4.1: Porcentaje de especies por tipo de madera.

de cada toma de imagen minimiza el posible ruido de la imagen debido a la presencia de partículas de chapas escaneadas con anterioridad.

Tras la calibración del escáner, se digitalizan las chapas de madera, recortándolas a un tamaño de 4.096 x 4.096 píxeles.

Las imágenes escaneadas se han dividido en 16 partes, siguiendo el procedimiento explicado en el apartado 3.2.2, dando lugar a un número total de 2.656 imágenes -64 por especie, excepto el *Sipo* que tiene la mitad de patrones-.

Para cada una de estas imágenes de color se han generado las imágenes derivadas -componentes primarias de color, pseudotridimensional y patrón periódico subyacente-.

Para conseguir el conjunto de imágenes derivadas de la original en color se ha procedido como se indica en los apartados 3.2.3.1, 3.2.3.2 y 3.2.3.3.

Se dispone así de un primer patrón formado por 6 imágenes -original en color de 1.024 x 1.024 píxeles, componentes roja, verde y azul, pseudotridimensional y de patrones periódicos subyacentes-, de modo que se tiene una matriz de datos-imágenes de 15.936 imágenes (2.656 patrones de 6 imágenes cada uno).

4.2. Construcción del vector de características.

En principio habría que determinar las características fotométricas y de textura y la dimensión fractal a todos los elementos de esta matriz de datos-imágenes.

Ahora bien, antes de medir las características se ha calculado el coeficiente de correlación entre las componentes de color roja, verde y azul de cada imagen, para determinar la componente que menos información va a aportar al conjunto de datos a obtener.

Para ello se ha utilizado el plugin `Image CorrelationJ` 10 de *ImageJ*, obteniéndose los siguientes resultados:

- Coeficiente de correlación entre las componentes roja y verde, con valores comprendidos entre 0,854 para una muestra de *Palo rosa* y 0,992 para una muestra de *Castaño* y dos muestras de *Pino melis*.
- Coeficiente de correlación entre las componentes roja y azul, de valores comprendidos entre 0,526 para una muestra de *Palo rosa* y 0,97 para una muestra de *Nogal americano*.
- Coeficiente de correlación entre las componentes verde y azul, entre 0,798 para una muestra de *Palo rosa* y 0,989 para una muestra de *Nogal americano*.

Analizados todos y cada uno de los coeficientes de correlación entre pares rojo-verde, rojo-azul y verde-azul de cada una de las muestras, en todos los casos el menor valor se produce entre las componentes roja y azul. De otra forma, trabajar con la componente verde en el proceso de clasificación supondría tener una redundancia de información innecesaria. El resultado obtenido al realizar la correlación de componentes es consecuente con la anulación de las señales relacionadas con la información de la componente verde en la transmisión de señales de televisión, por ser la que menos información aporta, incluso en el caso de transmitir imágenes verdes [105].

Con esto, el número de datos-imágenes total se reduce en un 16,67%, pasando a ser de 5 imágenes por patrón o de 13.280 datos totales a los que medir las características.

Utilizando diferentes comandos de *ImageJ*, las medidas de intensidad -media, desviación típica, moda, valor máximo, valor mínimo, mediana, asimetría y curtosis- realizadas sobre cada una de las imágenes hacen que el vector de características sea un vector fila con 40 columnas.

Como las medidas de textura -contraste, correlación, energía, entropía, homogeneidad, prominencia, segundo momento angular, varianza y sombra- se realizan en pasos de 45° , desde 0 a 135° , a cada vector de características habrá que añadir las 180 resultantes de esta medida.

Se completa el vector de características con las medidas de la dimensión fractal, dando como resultado que el patrón de cada muestra es un vector de 225 características. Por tanto, la matriz de datos inicial es una matriz de 597.600 datos, con 2.656 filas (muestras) y 225 columnas (valores de las características del vector de características).

4.3. Reducción del número de características.

Sobre cada uno de los vectores de características se ha determinado el valor del coeficiente de variación de Pearson de cada una de las 5 imágenes que lo forman, valor que sustituirá al de la desviación típica. De igual forma, en función de los valores máximo y mínimo del nivel de gris de cada imagen se ha calculado el rango, comprobándose que en el caso de las imágenes pseudotridimensional y patrón periódico subyacente éste es de 255 en todos los casos. Por tanto, las características fotométricas que se van a considerar se mantienen en 8 en todas las imágenes de un patrón, excepto de las dos mencionadas, lo que reduce de 40 a 38 las características fotométricas, es decir, se produce una primera reducción de éstas del 17,50 %.

Con respecto a las características de textura, al hallar la media direccional se produce una reducción de las mismas del 80 %, pasando de 225 iniciales a 45 características.

En el cuadro 4.1 se muestran los resultados de esta primera fase de reducción de características, consistente en reconfigurar los datos, introduciendo nuevas características y eliminando otras, resultando una reducción de los datos del vector de características del 69,26 %.

Como se desprende del cuadro anterior, esta reducción del 69,26 % de los datos iniciales, deja una dimensión del vector de características de 83, que da lugar a una matriz de 220.448 datos.

CARACTERÍSTICAS	INICIALES	FINALES	VARIACIÓN
Fotométricas	40	33	17,50 %
Textura	225	45	80,00 %
Dimensión fractal	5	5	0,00 %
Totales	270	83	69,26 %

Cuadro 4.1: Reducción de características inicial.

Normalizando los valores de las diferentes características y aleatorizando la posición de los diferentes patrones se ha construido el fichero ARFF, necesario para trabajar con *Weka*, añadiendo como última característica de cada patrón la correspondiente al nombre de la clase (especie), coincidente con las siglas indicadas en el cuadro 3.1, si bien en la discusión de los resultados esta última sólo se contempla como una etiqueta de los procedimientos supervisados que se van a ejecutar.

4.3.1. Selección basada en la correlación.

En primer lugar se ha utilizado evaluador basado en la correlación con estrategia de búsqueda hacia adelante. Los parámetros de entrada han sido los siguientes:

- Máximas características de búsqueda: 50.
- Criterio de parada: 5 subconjuntos de expansión sin mejora.
- Uso de evaluación cruzada con 10 iteraciones.

Los resultados obtenidos han sido los siguientes:

- 1.262 subconjuntos evaluados.
- Mérito del mejor subconjunto encontrado: 0,864.
- Dimensión del vector de características: 37.
- Reducción sobre el patrón base: 55,42 %.

- Características seleccionadas: las dadas en el cuadro 4.2, para cada imagen, a las que se ha denominado IMAG (imagen en color), IM-R (componente roja), IM-B (componente azul), IP3D (imagen pseudotridimensional) e IPPS (imagen del patrón periódico subyacente).

GRUPO	PARÁMETRO	IMAGEN UTILIZADA				
		IMAG	IM-R	IM-B	IP3D	IPPS
FOTOMETRÍA	M	X	X	X	X	
	MOD	X	X			
	MED		X	X		
	CVP	X	X	X	X	
	RAN					
	ASI			X		
	CUR	X	X	X	X	
DIMENSIÓN FRACTAL		X		X		X
TEXTURA	SMA					X
	CON		X		X	
	COR				X	
	HOM	X	X		X	
	ENT					
	ENE			X		
	PRO		X	X	X	
	VAR	X		X	X	
	SOM	X		X	X	

Cuadro 4.2: Parámetros seleccionados: CfsSubsetEval con estrategia LinnearForwardSelection.

De los resultados obtenidos se pueden extraer las siguientes conclusiones:

- La no influencia del parámetro fotométrico “rango”, dado que aunque es un parámetro utilizado normalmente, no es una buena medida de dispersión, ya que calcula la dispersión a partir de dos únicos valores.

- El parámetro de textura “entropía” tampoco es seleccionado, lo que implica que no ha influido en la reducción de parámetros el grado de complejidad que presenten las imágenes.
- La aportación de información útil de las imágenes de los patrones periódicos subyacentes (IPPS) es nula en las medidas fotométricas, lo que prácticamente se repite en las medidas de textura. Esto es indicativo de que la imagen de los patrones periódicos subyacentes no está correlada con las otras imágenes del conjunto de una muestra. Este hecho implica la imposibilidad de localizar espacialmente los elementos periódicos de la estructura, debido a que no hay ninguna asignación exacta entre las imágenes de autoconvolución y la original en el dominio espacial.

A continuación se ha utilizado el mismo selector cambiando la estrategia de búsqueda a la del mejor primero, con los parámetros de entrada siguientes:

- Búsqueda bidireccional.
- Criterio de parada: 5 subconjuntos de expansión sin mejora.
- Uso de evaluación cruzada con 10 iteraciones.

Se han obtenido en este caso los siguientes resultados:

- 3.735 subconjuntos evaluados.
- Mérito del mejor subconjunto encontrado: 0,891.
- Dimensión del vector de características: 41.
- Reducción sobre el patrón base: 50,60 %.
- Características seleccionadas: las dadas en el cuadro 4.3, para cada imagen, a las que se ha denominado IMAG (imagen en color), IM-R (componente roja), IM-B (componente azul), IP3D (imagen pseudotridimensional) e IPPS (imagen del patrón periódico subyacente).

Las conclusiones que se extraen en este caso son prácticamente las mismas que en el caso anterior.

GRUPO	PARÁMETRO	IMAGEN UTILIZADA				
		IMAG	IM-R	IM-B	IP3D	IPPS
FOTOMETRÍA	M	X	X	X	X	
	MOD	X	X			
	MED	X	X	X		
	CVP	X	X	X	X	
	RAN		X			
	ASI	X		X	X	
	CUR	X	X	X	X	
DIMENSIÓN FRACTAL		X		X		X
TEXTURA	SMA					X
	CON	X	X		X	
	COR				X	
	HOM		X	X	X	
	ENT				X	
	ENE			X		
	PRO			X	X	
	VAR	X		X	X	
	SOM			X	X	

Cuadro 4.3: Parámetros seleccionados: CfsSubsetEval con estrategia BestFirst.

4.3.2. Selección basada en el análisis de componentes principales.

Antes de realizar la selección de características por este método se ha comprobado el índice KMO, resultando ser de 0,089, valor que indica una inaceptable adecuación de las muestras para el uso del análisis de componentes principales.

Por ello se ha hecho necesario reestructurar el vector de características, llegándose a un mayor valor del índice de 0,858, que da una buena adecuación, haciendo aconsejable esta técnica de reducción de características cuando se mantienen las de la componente verde y se quitan las características correspondientes al patrón periódico subyacente.

Por tanto, en el caso de utilizar la reducción de características basada en el análisis de componentes principales, el vector de características reducido estará formado por los valores de los parámetros fotométricos y de textura y dimensión fractal de la imagen en color, las tres componentes de color y la imagen pseudotridimensional.

Con el selector basado en el análisis de componentes principales se han utilizado un porcentaje de varianza total cubierta del 95 %, obteniéndose 14 componentes principales, lo que da un vector de características que reduce la dimensión del patrón inicial un 83,13 %.

Analizando los autovalores, la proporción de varianza y la varianza acumulada que cubren, han de considerarse prácticamente las 14 componentes principales para cubrir la varianza indicada, si bien el gráfico de sedimentación mostrado en la figura 4.2, que coincide en la apreciación obtenida con el criterio de Kaiser, bastaría con utilizar las 12 primeras componentes principales.

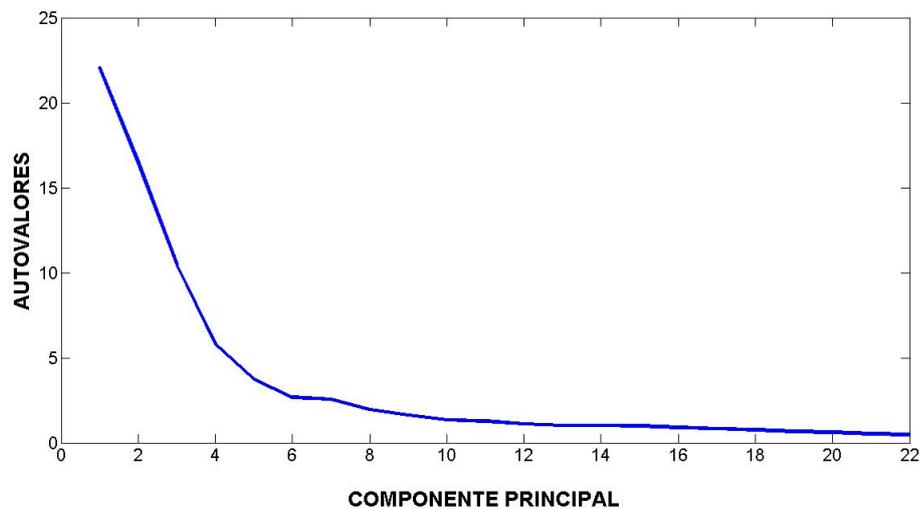


Figura 4.2: Gráfico de sedimentación

4.3.3. Comparativa de los algoritmos de selección.

En el cuadro 4.4 se muestra la comparativa de los resultados obtenidos con las diferentes técnicas de selección, con la denominación que se da en *Weka* a las mismas, referidos al patrón inicial de 83 características.

ALGORITMO	CARACTERÍSTICAS	REDUCCIÓN (%)
Patrón inicial	83	0
CfsSubsetEval-LinearForwardSelection	37	55,42
CfsSubsetEval-BestFirs	41	50,60
PrincipalComponents	14	83,13

Cuadro 4.4: Comparativa de algoritmos de selección.

La mayor reducción en la dimensión del vector de características la produce el análisis de componentes principales, aunque la comprobación de la reducción más efectiva se realizará una vez que se apliquen diferentes clasificadores a los distintos patrones obtenidos en esta fase de reducción de características.

4.4. Clasificación de patrones.

Se dispone hasta este momento de diferentes patrones, obtenidos tras el proceso de selección de características, tal y como se indica en el cuadro 4.4.

Para cada uno de ellos se van a probar los distintos clasificadores descritos en la sección 3.2.5. Antes de ello, se procede a establecer el conjunto de patrones de entrenamiento, formado por el 75% del total, y el conjunto de patrones de prueba, formado por el 25% restante. Así, para cada una de las pruebas a realizar se van a tener 1.992 vectores de características de prueba y 664 de entrenamiento.

Todos los clasificadores a utilizar están incluidos en el paquete *Weka*. Para cada uno de los pares selector-clasificador se va a optimizar el clasificador mediante la reiteración de entrenamientos sucesivos.

4.4.1. Clasificador J48.

El primer proceso de clasificación realizado con este clasificador se hizo con el conjunto de patrones de dimensión 37, correspondiente a la selección de características realizada

en base a la correlación con estrategia de búsqueda hacia adelante. Los parámetros de entrada de este clasificador han sido los siguientes:

- Clasificador: J48.
- Selección de características: basada en correlación con estrategia de búsqueda hacia adelante.
- Mínimo número de características por hoja: 2.
- Poda: sí.
- Factor de confianza: 0,25 (optimizado).
- Validación cruzada con $K = 10$.

Los resultados obtenidos con este clasificador, estructurado como árbol de 233 nodos con 117 hojas, han sido los que se recogen en el cuadro 4.5.

PARÁMETRO	ENTRENAMIENTO	PRUEBA
Precisión (%)	89,61	91,42
Eficiencia construcción (s)	120,06	–
Complejidad (bit/patrón)	85,30	–
Kappa estadística	0,894	0,912
Área ROC	0,959	0,963
Raíz del error cuadrático medio	0,0682	0,0629

Cuadro 4.5: Resultados del clasificador J48 con selección de características basada en correlación y estrategia hacia adelante.

El clasificador J48 se entrenó y probó a continuación con el conjunto de patrones de 41 características, seleccionadas en base a la correlación con estrategia de búsqueda el primer mejor, con los siguientes parámetros de entrada:

- Clasificador: J48.

- Selección de características: basada en correlación con estrategia de búsqueda el mejor primero.
- Mínimo número de características por hoja: 2.
- Poda: sí.
- Factor de confianza: 0,05 (optimizado).
- Validación cruzada con $K = 10$.

Con 225 nodos y 113 hojas, con este clasificador se han obtenido los resultados que se recogen en el cuadro 4.6.

PARÁMETRO	ENTRENAMIENTO	PRUEBA
Precisión (%)	89,51	89,31
Eficiencia construcción (s)	177,67	–
Complejidad (bit/patrón)	84,77	–
Kappa estadística	0,893	0,890
Área ROC	0,959	0,958
Raíz del error cuadrático medio	0,0684	0,0697

Cuadro 4.6: Resultados del clasificador J48 con selección de características basada en correlación y estrategia el mejor primero.

Una tercera prueba del clasificador J48 se hizo con el conjunto de patrones de 14 características, seleccionadas en base al análisis de componentes principales, y con los parámetros de entrada:

- Clasificador: J48.
- Selección de características: análisis de componentes principales.
- Mínimo número de características por hoja: 2.
- Poda: sí.
- Factor de confianza: 0,15 (optimizado).

- Validación cruzada con $K = 10$.

Los resultados obtenidos con este clasificador, de 383 nodos y 192 hojas, han sido los que se recogen en el cuadro 4.7.

PARÁMETRO	ENTRENAMIENTO	PRUEBA
Precisión (%)	76,41	74,25
Eficiencia construcción (s)	309,59	–
Complejidad (bit/patrón)	309,59	–
Kappa estadística	0,758	0,736
Área ROC	0,900	0,887
Raíz del error cuadrático medio	0,1017	0,1063

Cuadro 4.7: Resultados del clasificador J48 con selección por análisis de componentes principales.

La cuarta prueba del clasificador J48 se hizo con el conjunto de patrones de 83 características, base de las distintas técnicas de selección, y los parámetros iniciales dados a continuación:

- Clasificador: J48.
- Mínimo número de características por hoja: 2.
- Poda: sí.
- Factor de confianza: 0,35 (optimizado).
- Validación cruzada con $K = 10$.

Los resultados obtenidos con este clasificador han sido los que se recogen en el cuadro 4.8, para 227 nodos y 114 hojas.

4.4.2. Clasificador IBk.

Siguiendo el mismo orden que en el caso del clasificador anterior, la primera prueba de clasificación se hizo utilizando los patrones de características seleccionadas en base a

PARÁMETRO	ENTRENAMIENTO	PRUEBA
Precisión (%)	87,90	88,85
Eficiencia construcción (s)	239,59	–
Complejidad (bit/patrón)	95,04	–
Kappa estadística	0,876	0,883
Área ROC	0,954	0,948
Raíz del error cuadrático medio	0,0735	0,0724

Cuadro 4.8: Resultados del clasificador J48 sin reducción de atributos.

la correlación con estrategia de búsqueda hacia adelante. Como parámetros de entrada se han elegido los siguientes:

- Clasificador: IBk.
- Selección de características: basada en correlación con estrategia de búsqueda hacia adelante.
- Número de vecinos a considerar, k : 3 (optimizado).
- Criterio de distancia: distancia euclídea ponderada por su inversa.
- Validación cruzada con $K = 10$.

Los resultados obtenidos con este clasificador, han sido los que se recogen en el cuadro 4.9.

La siguiente prueba se llevó a cabo con el conjunto de vectores de características seleccionadas en base a la correlación con estrategia de búsqueda el primer mejor, con los parámetros siguientes:

- Clasificador: IBk.
- Selección de características: basada en correlación con estrategia de búsqueda el primer mejor.
- Número de vecinos a considerar, k : 1 (optimizado).

PARÁMETRO	ENTRENAMIENTO	PRUEBA
Precisión (%)	97,34	98,19
Eficiencia construcción (s)	5,73	–
Complejidad (bit/patrón)	0,35	–
Kappa estadística	0,973	0,982
Área ROC	0,996	0,999
Raíz del error cuadrático medio	0,0350	0,0253

Cuadro 4.9: Resultados del clasificador IB3 con selección de características basada en correlación y estrategia hacia adelante.

- Criterio de distancia: distancia euclídea ponderada por su inversa.
- Validación cruzada con $K = 10$.

Con este clasificador se han obtenido los resultados que se recogen en el cuadro 4.10.

PARÁMETRO	ENTRENAMIENTO	PRUEBA
Precisión (%)	97,69	97,89
Eficiencia construcción (s)	6,65	–
Complejidad (bit/patrón)	0,34	–
Kappa estadística	0,976	0,978
Área ROC	0,995	0,998
Raíz del error cuadrático medio	0,0332	0,0317

Cuadro 4.10: Resultados del clasificador IB1 con selección de características basada en correlación y estrategia el mejor primero.

Una tercera prueba del clasificador IBk se hizo con el conjunto de patrones de 14 características, seleccionadas en base al análisis de componentes principales, con las siguientes entradas:

- Clasificador: IBk.
- Selección de características: análisis de componentes principales.

- Número de vecinos a considerar, k : 4 (optimizado).
- Criterio de distancia: distancia euclídea ponderada por su inversa.
- Validación cruzada con $K = 10$.

Los resultados obtenidos con este clasificador se muestran en el cuadro 4.11.

PARÁMETRO	ENTRENAMIENTO	PRUEBA
Precisión (%)	91,01	93,07
Eficiencia construcción (s)	3.81	–
Complejidad (bit/patrón)	0.95	–
Kappa estadística	0,908	0,929
Área ROC	0,989	0,995
Raíz del error cuadrático medio	0,0631	0,0570

Cuadro 4.11: Resultados del clasificador IB4 con selección por análisis de componentes principales.

La cuarta prueba de este clasificador se hizo con el conjunto de patrones de 83 características, base de las distintas técnicas de selección.

Los parámetros de entrada de este clasificador han sido los siguientes:

- Clasificador: IB k .
- Número de vecinos a considerar, k : 4 (optimizado).
- Criterio de distancia: distancia euclídea ponderada por su inversa.
- Validación cruzada con $K = 10$.

Los resultados obtenidos con este clasificador han sido los que se recogen en el cuadro 4.12.

PARÁMETRO	ENTRENAMIENTO	PRUEBA
Precisión (%)	87,50	88,25
Eficiencia construcción (s)	13,95	–
Complejidad (bit/patrón)	1,41	–
Kappa estadística	0,872	0,880
Área ROC	0,977	0,994
Raíz del error cuadrático medio	0,0734	0,0676

Cuadro 4.12: Resultados del clasificador IB4 sin reducción de características.

4.4.3. Clasificador RandomForest.

Con la misma estrategia en las pruebas de este clasificador, en la primera se utilizaron los siguientes parámetros:

- Clasificador: RandomForest.
- Selección de características: basada en correlación con estrategia de búsqueda hacia adelante.
- Número de árboles: 10 (optimizado).
- Características aleatorias: 12 (optimizado).
- Validación cruzada con $K = 10$.

Los resultados obtenidos con este clasificador, para el que el error de las muestras extraídas del conjunto de entrenamiento para probar el clasificador resultó de 0,0897, han sido los que se recogen en el cuadro 4.13.

En la siguiente prueba los parámetros de entrada de este clasificador han sido los siguientes:

- Clasificador: RandomForest.
- Selección de características: basada en correlación con estrategia de búsqueda el mejor primero.

PARÁMETRO	ENTRENAMIENTO	PRUEBA
Precisión (%)	95,43	96,84
Eficiencia construcción (s)	21,9	–
Complejidad (bit/patrón)	5,39	–
Kappa estadística	0,973	0,968
Área ROC	0,999	0,999
Raíz del error cuadrático medio	0,0489	0,0477

Cuadro 4.13: Resultados del clasificador RandomForest con selección de características basada en correlación y estrategia hacia adelante.

- Número de árboles: 10 (optimizado).
- Características aleatorias: 13 (optimizado).
- Validación cruzada con $K = 10$.

Con este clasificador se han obtenido los resultados que se recogen en el cuadro 4.14. El error para las muestras del conjunto de entrenamiento extraídas para prueba fue de 0,0912.

PARÁMETRO	ENTRENAMIENTO	PRUEBA
Precisión (%)	95,53	96,84
Eficiencia construcción (s)	23,2	–
Complejidad (bit/patrón)	4,62	–
Kappa estadística	0,954	0,968
Área ROC	0,998	0,998
Raíz del error cuadrático medio	0,0488	0,0474

Cuadro 4.14: Resultados del clasificador RandomForest con selección de características basada en correlación y estrategia el mejor primero.

En la tercera prueba los parámetros de entrada fueron los siguientes:

- Clasificador: RandomForest.
- Selección de características: análisis de componentes principales.
- Número de árboles: 10 (optimizado).
- Características aleatorias: 4 (optimizado).
- Validación cruzada con $K = 10$.

Al error de 0,2447 obtenido para las muestras extraídas del conjunto de entrenamiento para su uso como prueba se añaden los resultados mostrados en el cuadro 4.15.

PARÁMETRO	ENTRENAMIENTO	PRUEBA
Precisión (%)	86,90	86,45
Eficiencia construcción (s)	11,67	–
Complejidad (bit/patrón)	22,84	–
Kappa estadística	0,866	0,861
Área ROC	0,986	0,986
Raíz del error cuadrático medio	0,0796	0,0827

Cuadro 4.15: Resultados del clasificador RandomForest con selección por análisis de componentes principales.

La cuarta prueba de este clasificador se hizo con el conjunto de patrones de 83 características, base de las distintas técnicas de selección, con los siguientes parámetros de entrada:

- Clasificador: RandomForest.
- Número de árboles: 9 (optimizado).
- Características aleatorias: 20 (optimizado).
- Validación cruzada con $K = 10$.

Los resultados obtenidos para un error de 0,1088 en las muestras que se han extrído del conjunto de entrenamiento para probar el clasificador han sido los que se recogen en el cuadro 4.16.

PARÁMETRO	ENTRENAMIENTO	PRUEBA
Precisión (%)	96,13	95,93
Eficiencia construcción (s)	36,43	–
Complejidad (bit/patrón)	3,02	–
Kappa estadística	0,960	0,958
Área ROC	0,998	1,000
Raíz del error cuadrático medio	0,0502	0,0504

Cuadro 4.16: Resultados del clasificador RandomForest sin reducción de características.

4.4.4. Clasificador RotationForest.

Con este modelo de *Weka* se ha implementado un único caso de prueba con los parámetros de entrada siguientes:

- Clasificador: RotationForest.
- Características del patrón: 83.
- Filtro de datos: análisis de componentes principales para una varianza total cubierta del 95 %.
- Árbol utilizado: J48 con 2 características mínimas por hoja, poda y un factor de confianza de 0,15.
- Convalidación cruzada con $K = 10$.

Los resultados obtenidos con este clasificador se resumen en el cuadro 4.17.

4.4.5. Clasificador LibLINEAR.

También con este algoritmo se han realizado cuatro pruebas. La primera con los patrones de características seleccionadas en base a la correlación con estrategia de búsqueda hacia adelante y los siguientes datos de entrada:

- Clasificador: LibLINEAR.

PARÁMETRO	ENTRENAMIENTO	PRUEBA
Precisión (%)	97,39	98,04
Eficiencia construcción (s)	9,42	–
Complejidad (bit/patrón)	1,90	–
Kappa estadística	0,973	0,980
Área ROC	0,999	1,000
Raíz del error cuadrático medio	0,0461	0,0449

Cuadro 4.17: Resultados del clasificador RotationForest con selección por análisis de componentes principales y árbol J48.

- Selección de características: basada en correlación con estrategia de búsqueda hacia adelante.
- SVM: L2-SVM.
- Coste: 30 (optimizado).
- Tolerancia del criterio de parada: 0,01.
- Validación cruzada con $K = 10$.

Los resultados obtenidos con este clasificador, han sido los que se recogen en el cuadro 4.18.

PARÁMETRO	ENTRENAMIENTO	PRUEBA
Precisión (%)	90,96	91,42
Eficiencia construcción (s)	91,92	–
Complejidad (bit/patrón)	97,05	–
Kappa estadística	0,907	0,912
Área ROC	0,954	0,956
Raíz del error cuadrático medio	0,0656	0,0639

Cuadro 4.18: Resultados del clasificador LibLINEAR con selección de características basada en correlación y estrategia hacia adelante.

En la segunda prueba se utilizó el conjunto de vectores de características seleccionadas en base a la correlación con estrategia de búsqueda el primer mejor, con los siguientes parámetros de entrada:

- Clasificador: LibLINEAR.
- Selección de características: basada en correlación con estrategia de búsqueda el primer mejor.
- SVM: L2-SVM.
- Coste: 30 (optimizado).
- Tolerancia del criterio de parada: 0,01.
- Validación cruzada con $K = 10$.

Con este clasificador se han obtenido los resultados que se recogen en el cuadro 4.19.

PARÁMETRO	ENTRENAMIENTO	PRUEBA
Precisión (%)	92,62	91,87
Eficiencia construcción (s)	82,50	–
Complejidad (bit/patrón)	79,26	–
Kappa estadística	0,924	0,917
Área ROC	0,962	0,958
Raíz del error cuadrático medio	0,0593	0,0622

Cuadro 4.19: Resultados del clasificador LibLINEAR con selección de características basada en correlación y estrategia el mejor primero.

Una tercera prueba del clasificador LibLINEAR se hizo con el conjunto de patrones de 14 características, seleccionadas en base al análisis de componentes principales y las entradas siguientes:

- Clasificador: LibLINEAR.
- Selección de características: análisis de componentes principales.

- SVM: L2-SVM.
- Coste: 3,5 (optimizado).
- Tolerancia del criterio de parada: 0,01.
- Validación cruzada con $K = 10$.

Los resultados obtenidos con este clasificador se muestran en el cuadro 4.20.

PARÁMETRO	ENTRENAMIENTO	PRUEBA
Precisión (%)	90,01	88,86
Eficiencia construcción (s)	182,6	–
Complejidad (bit/patrón)	107,29	–
Kappa estadística	0,898	0,886
Área ROC	0,949	0,943
Raíz del error cuadrático medio	0,0690	0,0728

Cuadro 4.20: Resultados del clasificador LibLINEAR con selección por análisis de componentes principales.

Por último, se entrenó y probó este algoritmo con el conjunto de patrones de 83 características, base de las distintas técnicas de selección y los siguientes parámetros de entrada:

- Clasificador: LibLINEAR.
- SVM: L2-SVM.
- Coste: 25 (optimizado).
- Tolerancia del criterio de parada: 0,01.
- Validación cruzada con $K = 10$.

Los resultados obtenidos con este clasificador han sido los que se recogen en el cuadro 4.21.

PARÁMETRO	ENTRENAMIENTO	PRUEBA
Precisión (%)	93,37	93,83
Eficiencia construcción (s)	79,47	–
Complejidad (bit/patrón)	71,17	–
Kappa estadística	0,932	0,937
Área ROC	0,966	0,968
Raíz del error cuadrático medio	0,0562	0,0542

Cuadro 4.21: Resultados del clasificador LibLINEAR sin reducción de características.

4.5. Comparativa de clasificadores.

Analizando los resultados obtenidos con todos los clasificadores entrenados y probados, se puede llegar a las conclusiones detalladas a continuación.

1. Se han entrenado y probado 17 sistemas de clasificación a partir de 5 clasificadores y tres técnicas de selección de atributos junto con los patrones sin reducir.
2. El objetivo de que los errores de clasificación no superen el 10 % lo cumplen en la fase de entrenamiento 11 de ellos, 2 están a menos del 0,5 % de cumplirlo -algoritmo J48 con selección de características por correlación- y el resto no cumple lo especificado -algoritmo J48 y RandomForest con selección basada en análisis de componentes principales y algoritmos J48 e IB4 sin reducción de características-. Con el número de muestras de entrenamiento utilizadas un 0,5 % corresponde a menos de 10 muestras mal clasificadas. En los puntos siguientes se hace referencia únicamente a los sistemas que cumplen o están a menos del 1 % del objetivo perseguido.
3. Para los clasificadores J48, RandomForest y LibLINEAR los peores resultados se obtienen cuando la selección de características se hace por análisis de componentes principales. Para el clasificador IBk los peores resultados se tienen cuando se utilizan los patrones sin reducir. Las diferencias de precisión llegan a ser de hasta un 11,49 %.
4. Con los mismos clasificadores del punto anterior se comprueba que los resultados de clasificación son peores cuando los patrones tienen menor dimensión -14 en el caso

de selección de características con análisis de componentes principales- o mayor dimensión -83 características cuando no se ha hecho un proceso de reducción-.

5. La excepción a las consideraciones anteriores la presenta el clasificador RotationForest, implementado en *Weka* para hacer una reducción de características mediante análisis de componentes principales, por lo que a su entrada se dispone de los vectores de características sin reducir.
6. Desde el punto de vista de la precisión, en la fase de entrenamiento el mejor comportamiento corresponde a los algoritmos IB1 con selección por correlación y RotationForest con selección por el método de análisis de componentes principales y árbol J48, cuya tasa de aciertos de clasificación está por encima del 97 %. En el caso peor -algoritmo J48 con selección por correlación y estrategia el mejor primero- el número de muestras mal clasificadas es de 209 de 1.992.
7. También en la fase de entrenamiento los mejores clasificadores se han construido en tiempos inferiores a los 10 s. Se puede comprobar que en el caso del clasificador J48 la eficiencia es del orden de 2-3 minutos, siendo del orden de 4-7 s para el IBk, de 20-40 s para el RandomForest, de unos 10 s para el clasificador RotationForest y del orden de 1-4 minutos para los clasificadores LibLINEAR. Se aprecia que, lógicamente, a mayor número de patrones mayor es el tiempo de construcción del entrenador.
8. La complejidad basada en el uso de memoria está por debajo de 1 bit/patrón en los IBk, subiendo a poco más de 1 bit/patrón en el RotationForest, de 3 a 5 bits/patrón en el RandomForest, de alrededor de 85 bits/patrón en el caso del J48 y de entre 70 y 108 bits/patrón en los LibLINEAR. Recordando que el número de patrones de entrenamiento es de 1.992, la complejidad máxima encontrada es de 215.136 bits -menos de 27 kB-.
9. La concordancia entre observadores, entendiendo que cada iteración de la convalidación cruzada es un observador, en todos los casos está por encima del 89 %, lo que significa que diferentes observadores coincidirían en la clasificación correcta de ese porcentaje de patrones.

10. La posibilidad de que un clasificador etiquete en su clase un patrón elegido al azar, como mínimo es del 94,5 %, según los resultados obtenidos para el área de la curva ROC.
11. En cuanto al valor de la raíz del error cuadrático medio su valor máximo es de 0,0734 correspondiente al clasificador IB4 sin reducción de características y su valor mínimo de 0,0332 para el IB1 con selección por correlación con estrategia el mejor primero.
12. Al utilizar los clasificadores con los patrones de prueba se aprecian diferencias en la precisión que van de un aumento de un 2,06 % (IB4 y reducción con análisis de componentes principales) a una disminución de 1,15 % (algoritmo LibLINEAR y reducción con análisis de componentes principales). Recuérdese que en el caso de las pruebas el conjunto de patrones es de 664, lo que hace que un sólo patrón que cambie de bien a mal clasificado, o a la inversa, equivale a una diferencia en uno u otro sentido del 0,15 % en la precisión, por lo que se está hablando de mejora de la clasificación correcta de unos 14 patrones o aumentar los patrones mal clasificados en unos 8 patrones.

A la vista de todos los resultados obtenidos y considerando, como ya se ha comentado, que el objetivo marcado lo cumplen 11 de los 17 sistemas probados, teniendo en cuenta los valores de todos los parámetros resultantes del proceso de clasificación -entrenamiento y prueba- se puede considerar que la decisión sobre cuál es el mejor clasificador ha de tomarse entre el **RotationForest** con filtrado de características por el método de análisis de componentes principales y que utiliza el árbol de decisión J48, el IB3 con selección por correlación y estrategia el mejor primero y el IB1 con selección por correlación y estrategia hacia adelante, si bien entre ellos no existen diferencias significativas.

Capítulo 5

Conclusiones

En este trabajo de investigación se presentan las bases para el diseño de un sistema automático de clasificación estadística de chapas de madera basado en las propiedades de luminosidad y textura de sus imágenes digitalizadas y en la determinación de la dimensión fractal de las mismas.

Partiendo de muestras que, visualmente, estaban libres de defectos, dado que la detección de estos quedaba fuera de los objetivos del trabajo, se han utilizado programas de dominio público para el procesado y clasificación de las imágenes se han conseguido tiempos de procesado muy competitivos.

Utilizando diferentes procedimientos de selección de características y algoritmos de clasificación, se han ensayado 17 combinaciones posibles, de las cuales 11 superan los objetivos marcados (error medio de clasificación inferior al 10 %), y de ellos 6 dan lugar a un error medio de clasificación por debajo del 5 %.

Se han utilizado técnicas de procesado de imagen novedosas en el mundo de la madera, incluyendo en la clasificación parámetros derivados de la estructura de la madera y de su composición colorimétrica. Únicamente los patrones periódicos ocultos no han aportado información útil al proceso, lo que parece lógico debido a la imposibilidad de localizar espacialmente los elementos periódicos de la estructura, al no haber ninguna asignación exacta entre las imágenes de autoconvolución y la original en el dominio espacial. Los patrones están más bien relacionados con las características geométricas de las fibras

internas de la madera. Sin embargo, la experiencia adquirida en este trabajo hace que se plantee el uso de estos patrones en la determinación de acabados -lijados, lacados,...-.

Otro aspecto a destacar es la poca información que aporta la componente primaria de color verde de la imagen al hacer la selección de características por correlación, presentándose como una característica redundante, lo que la hace, por contra, útil en el caso del análisis de componentes principales, en el que se necesitan características correladas entre sí.

Con todo ello, se ha cumplido con el objetivo de que el sistema fuese de bajo coste. Los equipos utilizados, un escáner y un ordenador convencionales, suponen costes de menor orden de magnitud que los utilizados en las investigaciones revisadas en este trabajo. El resto de costes, derivados del uso de los programas informáticos, es muy parecido en nuestro caso y en los citados.

También hay que tener en cuenta que el número de especies clasificado es superior al de la mayoría de las investigaciones que, en este terreno, se han venido realizando.

Por tanto, se ha conseguido un sistema de caracterización de chapas de madera basado en la clasificación de sus imágenes, un sistema de bajo coste y con un alto rendimiento en la precisión de la clasificación.

5.1. Aportaciones originales.

Las aportaciones que el autor de esta tesis ha hecho como originales a la misma se pueden enumerar como sigue:

- Aportar un esquema de clasificación que obtiene menor error que los existentes en la literatura revisada, clasificando además un mayor número de especies.
- Proponer un sistema de bajo coste basado en los clasificadores estudiados que puede ser de utilidad para las empresas del sector.
- Proponer un algoritmo de reconstrucción pseudotridimensional de la superficie de la madera, realizada por ampliación de la imagen y su composición con estructuras

fractales que generan gradientes direccionales de la iluminación, simulando técnicas de microscopía.

- Utilizar junto con las imágenes originales las correspondientes a sus componentes de color, eliminando la verde por su redundancia de información.
- Destacar la estructura interna de la madera mediante los patrones periódicos subyacentes a la propia superficie, utilizando para ello la autoconvolución de la imagen previo filtrado con un fractal. A efectos de clasificación se ha comprobado su poca influencia.
- Adaptar los vectores de características recursivamente sobre la base de los resultados de las métricas de evaluación de los distintos algoritmos de selección.
- Comparar la eficacia de 17 combinaciones diferentes de algoritmos de clasificación con los de selección de características que junto con el número de especies de madera hace que este trabajo, en este sentido, sea uno de los más amplios desarrollados sobre este material en los últimos años.
- Establecer procesos de clasificación satisfactorios con tiempos de procesado suficientemente bajos.
- Introducir la necesidad de la calibración de los equipos de adquisición de imágenes, que se deberá realizar periódicamente para asegurar la uniformidad luminosa y la corrección de los colores obtenidos en el proceso de digitalización.

5.2. Líneas futuras de investigación.

A partir de este trabajo de investigación se pueden realizar nuevas aportaciones que mejoren la automatización y la eficiencia en la clasificación. Así, como continuidad del mismo se proponen las siguientes líneas para futuras investigaciones:

- Diseño de una aplicación integrada que se encargue de todos los procesos: control del escáner, obtención de parámetros, formación de las matrices de datos, generación del clasificador y optimización del mismo, y elaboración de informes.

- Estudio de nuevas formulaciones en la selección de características, entre las que podrían utilizarse métodos basados en algoritmos genéticos.
- Estudio de las ventajas e inconvenientes derivados de utilizar como captador una cámara fotográfica utilizando formatos de imagen con diferentes grados de compresión.
- Búsqueda de alternativas para corregir la clasificación de especies por debajo de unos límites preestablecidos cuando exista confusión clara con otra especie.
- Ampliación del entorno de clasificación, hacia la detección de defectos de la madera.
- Desarrollar las herramientas necesarias para, mediante procesado de imagen, poder detectar los acabados (lijado, barnizado...) de la madera y caracterizar cada uno de ellos.
- Utilizar equipos y dispositivos de microondas para parametrizar la madera.
- Crear una herramienta que “on line” permita determinar la especie a que pertenece una muestra concreta de madera subida a la red, para lo cual habrá de disponer previamente de una extensa xiloteca digital parametrizada.

5.3. Difusión de la investigación.

La investigación desarrollada ha dado lugar a una serie de publicaciones y presentaciones en congresos, nacionales e internacionales, así como la divulgación en revistas del sector de la madera y la participación en proyectos de I+D.

En resumen, el autor de esta tesis es coautor de cinco (5) artículos en revistas internacionales, cuatro (4) de ellas indexadas en el JCR, participa en un (1) proyecto de I+D de la convocatoria “Excelencia” de 2013, se han presentado comunicaciones en tres (3) congresos, dos (2) de ellos nacionales, y es coautor de un (1) artículo de divulgación.

Revistas internacionales.

- Rojas, J.A.M; **Alpuente, J.**, Rojas, I.M.; Vignote, S. “Fractal-based image enhancement techniques inspired by differential interference contrast microscopy”. *Journal of Optics A: Pure and Applied Optics*. 2009, vol.11, no. 2, 8 p. (Factor de impacto: 1,198).
- Rojas, J.A.M.; **Alpuente, J.**; Bolívar, E.; López-Espí, P.; Rojas, I.M.; Vignote, S. “Empirical characterization of wood surfaces by means of iterative autocorrelation of laser speckle patterns”. *Progress in Electromagnetic Researchs*.2008, no. 80, p. 295-306. (Factor de impacto: 4,375).
- Sánchez Montero, R.; López Espí, P.L.;**Alpuente, J.**; Rojas, J.A.M.; Vignote, S. “Polyspectral technique for the analysis of stress-waves characteristics and species recognition in wood veneers”. *Applied Acoustic*. 2014, vol. 86, p. 89-94. (Factor de impacto en 2013: 1,068).
- Martínez Rojas, J.A.; Vignote Peña, S.; **Alpuente Hermosilla, J.**; Sánchez Montero, R.; López Espí, P.L.; Martínez Rojas, Isaac. “Spectral biomimetic technique for wood classification inspired by human echolocation”. *Advances in Acoustics and Vibration*. 2012, vol. 2012, 8 p.
- Rojas, J.A.M.; **Alpuente, J.**; Postigo, D.; Rojas, I.M.; Vignote, S. “Wood Species identification using stress-wave analysis in the audible range”. *Applied Acoustics* . 2011, vol. 72, no.12, p.934-942. (Factor de impacto: 0,891).

Participación en proyectos de I+D.

Dentro de la convocatoria de ayudas a proyectos de I+D «EXCELENCIA», Subprograma de Generación del Conocimiento 2013, de la Subdirección General de Proyectos de Investigación del Ministerio de Ciencia y Tecnología,

- *Caracterización automática de la superficie de madera mediante medios ópticos no invasivos*. IP: Pablo Luis López Espí. No. investigadores: 7.

Participación en congresos.

Relacionado con el tema de la tesis, se ha participado en un congreso nacional:

- Rojas, J.A.M.; **Alpuente, J.**; Bolívar, E.; López-Espí, P.; Sánchez, R. “Aplicación de técnicas laser no invasivas para el estudio de maderas”. En: *CONAMA 9. Fondo Documental on line. Grupo de Trabajo: Teledetección y sensores ambientales (GT-TEL)*. Madrid, 1-5 de Diciembre de 2008.

Las técnicas fractales se han aplicado en el campo de la Medicina, dando lugar a dos ponencias en congresos nacionales:

- Sánchez Montero, R.; Núñez Martín, L.; Martínez Rojas, J.A.; **Alpuente Hermosilla, J.**; Castro Tejero, P.; López Espí, P.L. “Tratamiento mediante filtros fractales de las imágenes PET para su uso en RC y RT”. En *Libro de actas del XIV Congreso de la Sociedad Española de Radiocirugía*. Baiona, 24-26 de Septiembre de 2014.
- Sánchez Montero, R.; Núñez Martín, L.; Piferré Martínez, X.; Pérez Arangüena, R.; **Alpuente Hermosilla, J.**; López Espí, P.; Espinosa Esteban, M.M.; Martínez Rojas, J.A. “Mejora del contraste de microcalcificaciones en mamografías digitales mediante el empleo de filtros fractales direccionales”. En *III Congreso Conjunto de las Sociedades Españolas de Física Médica y Protección Radiológica SEMF 19-SEPR 14*. Cáceres, 18-21 de Junio de 2013. (Premio a la Mejor Comunicación en el Área de Física Médica).

También se han aplicado las mismas técnicas a la reducción de moteado de ruido en imágenes SAR, dando lugar a la presentación de la siguiente comunicación a congreso internacional:

- Martínez Rojas, J.A.; Sánchez Montero, R.; **Alpuente Hermosilla, J.**; López Espí, P. “Fractal convolution techniques for speckle noise reduction”. En *Proceedings of the 5th European Radar Conference*. Amsterdam, 27-31 de Octubre de 2010.

Divulgación en el entorno empresarial de la madera.

La divulgación de la tecnología en el entorno profesional es otro de los apartados que se han tratado de cubrir con publicaciones como la siguiente:

- Vignote, S.; Martínez Rojas, I.; Martínez Rojas, J.A.; Villasante, A.; **Alpuente, J.** “Automatización de la producción y su control. Análisis de imágenes” [en línea]. *Revista AITIM*. 2013, Boletín de Información Técnica no. 218, p. 66-76.

Agradecimientos.

Esta tesis ha sido desarrollada parcialmente dentro del proyecto de investigación *Caracterización automática de la superficie de madera mediante medios ópticos no invasivos*, con referencia AGL2013-44631-P.

Bibliografía

- [1] UNECE AND FAO. *Forecast of the Committee on Forest and the Forest Industry: forest products production and trade 2012-2014*. Geneva: United Nations, 013. 35 p. Geneva Timber and Forest. Discussion Paper 61.
- [2] España. *Fichas sectoriales*. Ministerio de Industria, Energía y Turismo, 2014. Disponible en: www.minetur.gob.es/es-es/IndicadoresyEstadísticas/Sectores/. [Consulta: 28 de Junio de 2014].
- [3] CAPUZ, R.. *Materiales orgánicos. maderas*. Valencia: Universidad Politécnica de Valencia, 2005, 279 p.
- [4] VIGNOTE, S. Y MARTÍNEZ, I., *Tecnología de la madera*. 3a. edición. Madrid: Ediciones Mundi-Prensa, 2006, 687 p.
- [5] WIEDENHOEFT, A. “Structure and function of wood”. En: ROSS, R.J. (editor). *Wood handbook. Wood as an engineering material. General Technical Report FPL-GTR-190*. Madison: U.S. Department of Agriculture, Forest Service, Forest Products Laboratory, 2010, p. 3.1-3.18.
- [6] PERAZA, C. “La madera y sus propiedades”. Revista AITIM [en línea]. 1976, no. 80, 13 p. Disponible en:
http://www.infomadera.net/uploads/articulos/archivo_875_16757.pdf
- [7] GUTIÉRREZ, A. “La madera y sus propiedades (I)”. Revista AITIM [en línea]. 1981, no. 104, 5 p. Disponible en:
http://www.infomadera.net/uploads/articulos/archivo_1111_16991.pdf

- [8] KRETCHMANN, D.E. “Mechanical properties of wood”. En: ROSS, R.J. (editor). *Wood handbook. Wood as an engineering material. General Technical Report FPL-GTR-190*. Madison: U.S. Department of Agriculture, Forest Service, Forest Products Laboratory, 2010, p. 5.1-5.46.
- [9] SIMPSON, W.T. “Properties of wood related to drying”. En: SIMPSON, W.T. (editor). *Dry kiln operator’s manual*. Madison: U.S. Department of Agriculture, Forest Service, Forest Products Laboratory, 2003, 256 p.
- [10] GLASS, S.V. AND ZELINKA, S.L. “Moisture relations and physical properties of wood”. En: ROSS, R.J. (editor). *Wood handbook. Wood as an engineering material. General Technical Report FPL-GTR-190*. Madison: U.S. Department of Agriculture, Forest Service, Forest Products Laboratory, 2010, p. 4.1-4.19.
- [11] BUCUR, V. *Acoustic of wood*. 2nd edition. Berlin: Springer-Verlag, 2006. 399 p. Springer Series in Wood Science.
- [12] BUCUR, V. “Techniques for high resolution imaging of wood structure: a review”. *Measurement Science and Technology*. 2003, vol. 14, no. 12, p. R91-R98.
- [13] AENOR. *Alteraciones de la madera en pie y apeada. Terminología. Definiciones*. UNE 56510. Madrid: AENOR, 1964.
- [14] AENOR. *Defectos y anomalías de la madera aserrada. Nudos*. UNE 56521. Madrid: AENOR, 1972.
- [15] AENOR. *Defectos y anomalías de la madera aserrada. Fendas y acebolladuras*. UNE 56520. Madrid: AENOR, 1972.
- [16] FERNÁNDEZ-GOLFÍN, J.R.; DÍEZ, M.R.; HERMOSO, E. Y MIER, R. *Manual de clasificación de madera*. Madrid: AITIM, 2003, 38 p.
- [17] BRASHAW, B.K.; BUCUR, V.; DIVOS, F.; GONÇALVES, R.; LU, J.; MEDER, R.; PELLERIN, R.F.; POTTER, S.; ROSS, R.J.; WANG, X. AND YIN, Y. “Nondestructive testing and evaluation for wood: a worldwide research update”. *Forest Products Journal*. 2009, vol. 59, no. 3, p. 7-14.

- [18] GRÖNLUND, U. AND BROMAN, O. "Relations between wood properties and manual grading – an approach based on multivariate analysis". *European Journal of Wood and Wood Products*. 1995, vol. 53, no. 6, p. 413-420.
- [19] BOND, B.H. "Characterization of wood features using color, shape and density parameters". [Tesis Doctoral]. Tutor: Earl Kline. Faculty of the Virginia Polytechnic Institute and State University, Wood Science and Forest Products Department, 1998.
- [20] OJALA, T.; PIETIKÄINEN, M. AND MÄENPÄÄ, T. "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns". *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 2002, vol. 24, no. 7, p. 971-987.
- [21] VICKERS, A.L. AND MODESTINO, J.W. "A maximum likelihood approach to texture classification". *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 1982, vol. 4, no. 1, p. 61-68.
- [22] THYAGARAJAN, K.S.; NGUYEN, T. AND PERSONS, C.E. "A maximum likelihood approach to texture classification using wavelet transform". En: *Proceedings of the IEEE International Conference on Image Processing*. Austin, 13-16 de Noviembre de 1994, p. 640-644.
- [23] HALEY, G.M. AND MANJUNATH, B.S. "Rotation-invariant texture classification using a complete space-frequency model". *IEEE Transactions on Image Processing*. 1999, vol. 8, no. 2, p. 255-269.
- [24] AKHLOUFI, M.A.; BEN LARBI, W. AND MALDAGUE, X. "Framework for color-texture classification in machine vision inspection of industrial products". En: *Proceedings of the IEEE International Conference on Systems, Man and Cybernetics*. Quebec, 7-10 de Octubre de 2007, p. 1067-1071.
- [25] KASHYAP, R.L. AND KHOTANZAD, A. "A model-based method for rotation invariant texture classification". *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 1986, vol. 8, no. 4, p. 472-481.

- [26] CHELLAPPA, R. AND CHATTERJEE, S. “Classification of textures using Gaussian Markov random fields”. *IEEE Transactions on Acoustic, Speech and Signal Processing*. 1985, vol. 33, no. 4, p. 959-963.
- [27] PORTER, R. AND CANAGARAJAH, N. “Robust rotation-invariant texture classification: wavelet, Gabor filter and GMRF based schemes”. *IEEE Proceedings. Vision, Image and Signal Processing*. 1997, vol. 144, no. 3, p. 180-188.
- [28] ARIVAZHAGAN, S.; GANESAN, L. AND PRIYAL, S.P. “Texture classification using Gabor wavelets based rotation invariant features”. *Pattern Recognition Letters*. 2006, vol. 27, no. 16, p. 1976-1982.
- [29] PALM, C.; KEYSERS, D.; LEHMANN, T.M. AND SPITZER, K. “Gabor filtering of complex hue/saturation images for color texture classification”. En: *Proceedings of the International Conference on Computer Vision, Pattern Recognition & Image Processing*. Atlantic City, 27 de Febrero-3 de Marzo de 2000, p.45-49.
- [30] LIU, Z.; WANG, H. AND PENG, S. “Texture classification through directional empirical model decomposition”. En: *Proceedings of the 17th International Conference on Pattern Recognition*. Cambridge, 23-26 de Agosto de 2004, vol. 4, p. 803-806.
- [31] PYUN, K.; WON, C.S.; LIM, J. AND GRAY, R.M. “Texture classification based on multiple Gauss mixture vector quantizers”. En: *Proceedings of the 2002 IEEE International Conference on Multimedia and Expo*. Lausanne, 26-29 de Agosto de 2002, vol. 2, p. 501-504.
- [32] GREENSPAN, H.; BELONGIE, S.; GOODMAN, R. AND PERONA, P. “Rotation invariant texture recognition using a steerable pyramid”. En: *Proceedings of the 12th IAPR International Conference on Pattern Recognition. Conference B: Pattern Recognition and Neural Networks*. Jerusalem, 9-13 de Octubre de 1994, vol. 2, p. 162-167.
- [33] BOMBARDIER, V. AND SCHMITT, E. “Fuzzy rule classifier: capability for generalization in wood color recognition”. *Engineering Applications of Artificial Intelligence*. 2010, vol. 23, p. 978-988.

- [34] BREMANANTH, R.; NITHYA, B. AND SAIPRIYA, R. "Wood species recognition system". *World Academy of Science, Engineering and Technology*. 2009, vol. 52, p.873-879.
- [35] LIU, J. AND FURUNO, T. "The fractal evaluation of wood texture by the triangular prism surface area method". *Wood and Fiber Science*. 2001, vol. 33, no. 2, p. 213-222.
- [36] YOU, M. AND CAI, C. "Wood classification based on PCA, 2DPCA, (2D)2PCA and LDA". En: *Proceedings of the 2009 2nd International Symposium of Knowledge Acquisition and Modeling*. Wuhan, 30 de Noviembre-1 de Diciembre de 2009, p. 371-374.
- [37] LEBOW, P.K.; BRUNER, C.C.; MARISTANY, A.G. AND BUTLER, D.A. "Classification of wood surface features by spectral reflectance". *Wood and Fiber Science*. 1996, vol. 28, no. 1, p. 74-90.
- [38] SCHMITT, E.; BOMBARDIER, V. AND VOGRIG, R. "Fuzzy linguistic rules classifier for wooden board color sorting". En: BLANC-TALON, J.; PHILIPS, W.; POPESCU, D. AND SCHEUNDERS, P. (editores). *Advanced concepts for intelligence vision systems. Proceedings of the 7th International Conference, ACIVS 2005*. Antwerp, 20-23 de Septiembre de 2005, p. 154-161. Springer.
- [39] JOHANSSON, D. AND MORÉN, T. "The potential of colour measurement for strength prediction of thermally treated wood". *European Journal of Wood and Wood Products*. 2006, vol. 64, no. 2, p. 104-110.
- [40] FARIA, J.; MARTINS, T.; FERREIRA, M. AND SANTOS, C. "A computer vision system for color grading wood boards using fuzzy logic". En: *Proceedings of the IEEE International Symposium on Industrial Electronics*. Cambridge, 30 de Junio-2 de Julio de 2008, p. 1082-1087.
- [41] WANG, B.; WANG, H. AND QI, H. "Wood recognition based on grey-level co-occurrence matrix". En: *Proceedings of the International Conference on Computer Application and System Modeling*. Taiyuan, 22-24 de Octubre de 2010, vol. 1, p. 269-272.
- [42] TOU, J.Y.; TAY, Y.H. AND LAU, P.Y. "A comparative study for texture classification techniques on wood species recognition problem". En: *Proceedings of the 5th*

- International Conference on Natural Computation*. Tianjin, 14-16 de Agosto de 2009, vol. 5, p. 8-12.
- [43] WANG, K. AND BAI, X. “Research on classification of wood surface texture based on feature level data fusion”. En: *Proceedings of the 2nd IEEE Conference on Industrial Electronics and Applications*. Harbin, 23-25 de Mayo de 2007, p. 659-663.
- [44] YUSOF, R.; ROSLI, N.R. AND KHALID, M. “Using Gabor filters as image multiplier for tropical wood species recognition system”. En: *Proceedings of the 12th International Conference on Computer Modelling and Simulation*. Cambridge, 24-26 de Marzo de 2010, p. 289-294.
- [45] BAI, X. AND WANG, K. “Research on classification of wood surface texture based on Markov random field”. En: *Proceedings of the 2nd IEEE Conference on Industrial Electronics and Applications*. Harbin, 23-25 de Mayo de 2007, p. 664-668.
- [46] TOPALOVA, I. AND TZOKEV, A. “Modular structure of neural networks for classification of wooden surfaces with PLC industrial implementation”. En: *Proceedings of the 7th International Conference on Autonomic and Autonomous Systems*. Venecia, 22-27 de Mayo de 2011, p. 13-17.
- [47] GONZAGA, A.; FRANÇA, C.A. AND SLAETS, A.F.F. “Wood texture classification by fuzzy neural networks”. En: *Proceedings of the Applications of Artificial Neural Network in Image Processing IV*. San Jose, 28-29 de Enero de 1999, vol. 3647, p. 134-143.
- [48] WANG, Y.; WANG, H. AND MO, L. “Research on recognition of wood texture based on integrated neural network classifier”. En: *Proceedings of the International Conference on Intelligence Control and Information Processing*. Dalian, 13-15 de Agosto de 2010, p. 512-515.
- [49] KHALID, M.; LEE, E.L.Y.; YUSOF, R. AND NADARAJ, M. “Design of an intelligent wood species recognition system”. *International Journal of Simulation, Systems, Science and Technology*. 2008, vol. 9, no. 3, p. 9-19.
- [50] YI, T.J. “Texture classification on wood images for species recognition”. [Tesis de Máster]. Supervisor: Tay Yong Haur. Universiti Tunku Abdul Rahman, Faculty of *Tesis Doctoral: Caracterización automática de especies de madera...*

- Information and Communication Technology, Department of Computer Science and Information System, 2009.
- [51] KIM, C.W. AND KOIVO, A.J. “Hierarchical classification of surface defects on dusty wood boards”. *Pattern Recognition Letters*. 1994, vol. 15, no. 7, p. 713-721.
- [52] WIBOWO, H.I.A. “Rule-based wood knot defect image classification”. [Tesis de Máster]. Supervisor: Fabrice Meriadeau. Université de Bourgogne, Laboratoire Electronique, Informatique et Image (Le2i) UMR CNRS 5158, 2010.
- [53] JABO, S. “Machine vision for wood defect detection and classification”. [Tesis de Máster]. Supervisor: Irene Y.H. Gu. Chalmers University of Technology, Department of Signal and Systems, 2011.
- [54] ATHILAKSHMI, R.; WAHI, A. AND NAGARAJAN, B. “Defect identification of lumber trough correlation technique with statistical and textural feature extraction method”. En: *Proceedings of the International Conference on Communication and Computational Intelligence*. Erode, 27-29 de Diciembre de 2010, p. 524-528.
- [55] CAVDARLI, M. AND SEKE, E. “Measuring roughness on wood surfaces for detection of defects using multi-frame imaging”. En: *Proceedings of the International Symposium on Innovations in Intelligent Systems and Applications*. Kayseri, 21-23 de Junio de 2010, 5 p.
- [56] ALAPURANEN, P. AND WESTMAN, T. “Automatic visual inspection of wood surfaces”. En: *Proceedings of the 11th IAPR International Conference on Pattern Recognition. Conference A: Computer Vision and Applications*. La Haya, 30 de Agosto-3 de Septiembre de 1992, vol. I, p. 371-374.
- [57] HIETANIEMI, R.; HANNUKSELA, J. AND SILVÉN, O. “Camera based lumber strength classification system”. En: *Proceedings of the 12th IAPR Conference on Machine Vision Applications*. Nara, 13-15 de Junio de 2011, p. 251-254.
- [58] GU, I.Y.H.; ANDERSON, H. AND VICEN, R. “Automatic classification of good defects using support vector machines”. En: BOLC, L.; KULIKOWSKI, J. AND WOJCIE-

- CHOWSKI, K. (editores). *Computer Vision and Graphics. Lecture Notes in Computer Science*, vol. 5337. Berlín: Springer Berlin/Heidelberg, 2009, p. 356-367.
- [59] RUZ, G.A.; EESTÉVEZ, P.A. AND RAMÍREZ, P.A. “Automated vision inspection system for wood defect classification using computational intelligence techniques”. *International Journal of Systems Science*. 2009, vol. 40, no. 2, p. 163-172.
- [60] PACKIANATHER, M.S.; DRAKE, P.R. AND PHAM, D.T. “Feature selection method for neural network for the classification of wood veneer defects”. En: *Proceedings of the World Automation Congress*. Waikoloa, 28 de Septiembre-2 de Octubre de 2008, p. 1-6.
- [61] LI, M.X. AND WU, C.D. “A vision based inspection method for wood veneer classification”. En: *Proceedings of IEEE Conference on Robotics, Automation and Mechatronics*. Chengdu, 21-24 de Septiembre de 2008, p. 494-498.
- [62] LI, M.X. AND WU, C.D. “A VPRS and NN method for wood veneer surface inspection”. En: *Proceeding of the International Workshop on Intelligent System and Applications*. Wuhan, 23-24 de Mayo de 2009, p. 1-4.
- [63] DRAKE, P.R. AND PACKIANATHER, M.S. “A decision tree of neural networks for classifying images of wood veneer”. *International Journal of Advanced Manufacturing Technology*. 1998, vol. 14, p. 280-285.
- [64] MARCANO-CEDENO, A.; QUINTANILLA-DOMÍNGUEZ, J. AND ANDINA, D. “Wood defects classification using artificial metaplaticity neural network”. En: *Proceedings of the 35th Annual Conference of the IEEE Industrial Electronics Society*. Porto, 3-5 de Noviembre de 2009, p. 3422-3427.
- [65] RAMÍREZ, G.M.J. Y CHACÓN, M.I. “Clasificación de defectos en madera utilizando redes neuronales artificiales”. *Computación y Sistemas*. 2005, vol. 9, no. 1, p. 17-27.
- [66] ESTÉVEZ, P.A.; PÉREZ, C.A. AND GOLES, E. “Genetic input selection to a neural classifier for defect classification of radiata pine boards”. *Forest Products Journal*. 2003, vol. 53, no. 7/8, p. 87-94.

- [67] LAMPINEN, J. AND SMOLANDER, S. "Self-organizing feature extraction in recognition of wood surface defects and color images". *International Journal of Pattern Recognition and Artificial Intelligence*. 1996, vol. 10, no. 2, p. 97-113.
- [68] PAHM, D.T. AND ALCOCK, R.J. "Synergistic classification systems for wood dwfect identification". En: *Proceedings of the Institution of Mechanical Engineers. Part E, Journal of Process Mechanical Engineering*. 1999, vol. 213, no. 2, p. 127-133.
- [69] SHAHNORBANUN, S.; SITI NURUL HUDA, S.A.; HASLINA, A.; NAZLIA, O. AND ROSILAH, H. "A computational biological network for Wood defect classification". En: *Proceedings of the World Congress on Engineering and Computer Sciences*. San Francisco, 20-22 de Octubre de 2010, vol. I, p. 559-563.
- [70] CHACÓN, M.I. AND RAMÍREZ-ALONSO, G. "Wood defects classification using a SOM/FFP approach with minimum dimension feature vector". En: WANG, J.; YI, Z.; ZURADA, J.; LU, B. AND YIN, H. (editores). *Advances in neural networks. Lecture Notes in Computer Science*, vol. 3973. Berlín: Springer Berlin/Heidelberg, 2006, p. 1105-1110.
- [71] KAUPPINEN, H.; RAUTIO, H. AND SILVÉN, O. "Non-segmenting defect detection and SOM based classification for surface inspection using color vision". En: *Proceedings of the Conference on Polarization and Color Techniques in Industrial Inspection*. Munich, 17-18 de Junio de 1999, p. 270-280.
- [72] SILVÉN, O.; NISKANEN, M. AND KAUPPINEN, H. "Wood inspection with non-supervised clustering". *Machine Vision and Applications*. 2003, vol. 13, p. 275-285.
- [73] NISKANEN, M.; SILVÉN, O. AND KAUPPINEN, H. "Color and texture based wood inspection with non-supervised clustering". En: *Proceedings of the 12th Scandinavian Conference on Image Analysis*. Bergen, 11-14 de Junio de 2001, p. 336-342.
- [74] ROJAS, J.A.M.; ALPUENTE, J.; POSTIGO, D.; ROJAS, I.M. AND VIGNOTE, S. "Wood species identification using stress-wave analysis in the audible range". *Applied Acoustics*. 2011, vol. 72, p. 934-942.

- [75] YELLA, S.; GRUPTA, N. AND DOUGHERTY, M. “Condition monitoring using pattern recognition techniques on data from acoustic emissions”. En: *Proceedings of the 5th International Conference on Machine Learning and Applications*. Orlando, 14-16 de Diciembre de 2006, p. 3-9.
- [76] MILLS, T.; JONES, A. AND LODDER, R.A. “Identification of wood species by acoustic-resonance spectrometry using multivariate subpopulation analysis”. *Applied Spectroscopy*. 1993, vol. 47, no. 11, p. 1880-1886.
- [77] SÁNCHEZ-MONTERO, R.; LÓPEZ ESPÍ, P.L.; ALPUENTE, J.; ROJAS, J.A.M. AND VIGNOTE, S. “Polyspectral technique for the analysis of stress-waves characteristics and species recognition in wood veneers”. *Applied Acoustics*. 2014, vol. 86, p. 89-94.
- [78] KAWAMOTO, S. AND WILLIAMS, R.S. “Acoustic emission and acousto-ultrasonic techniques for wood and wood-based composites. A review”. *General Technical Report FPL-GTR-134*. Madison: U.S. Department of Agriculture, Forest Service, Forest Products Laboratory, 2002, 16 p.
- [79] BEALL, F.C. “Overview of the use of ultrasonic technologies in research of wood properties”. *Wood Science and Technology*. 2002, vol. 36, p. 197-212.
- [80] KABIR, M.F.; LEININGER, T.D.; ARAMAN, P.A. AND WINN, M.F. “Detection of wetwood by ultrasonics”. *Forest Products Journal*. 2006, vol. 56, no. 3, p. 70-74.
- [81] JORDAN, R.; FEENEY, F.; NESBITT, N. AND EVERTSEN, J.A. “Classification of wood species by neural network analysis of ultrasonic signals”. *Ultrasonics*. 1998, vol. 36, p. 219-222.
- [82] MVOGO, J.K.; AYINA, O.; MORLIER, P. AND CASTERA, P. “A new approach of grouping wood species and a proposal of a grading system of timber in the Congo basin”. En: *Proceedings of 10th Conference on Timber Engineering*. Miyazaki, 2-5 de Junio de 2008, vol. 3, p. 1431-1438.
- [83] ÁL-MATTARNEH, H.M.A.; GHODGAONKAR, D.K. AND MAJID, W.M. “Microwave nondestructive testing for classification of Malaysian timber using free-space techni-

- ques”. En: *Proceedings of International Symposium on Signal Processing and its Applications*. Kuala Lumpur, 13-16 de Agosto de 2001, p. 450-453.
- [84] SONI, A.H. AND GURJAR, A.A. “Microwave based classification of material using neural network”. *International Journal of Engineering Science and Technology*. 2011, vol. 3, p. 5974-5978.
- [85] RODRÍGUEZ-LIÑÁN, C.; MORALES-CONDE, M.J.; RUBIO-DE HITA, P. Y PÉREZ-GÁLVEZ, F. “Análisis sobre la influencia de la densidad en la termografía de infrarrojos y el alcance de esta técnica en la detección de defectos internos de la madera”. *Materiales de Construcción*. 2012, vol. 62, no. 305, p. 99-113.
- [86] BRUNNER, M.; EUGSTER, R.; TRENKA, E. AND BERGAMIN-STROTZ, L. “FT-NIR spectroscopy and wood identification”. *Holzforschung*. 1996, vol. 50, no. 2, p. 130-134.
- [87] HAARTVEIT, E.Y. AND FLAETE, P.O. “Near infrared spectroscopy (NIRS) as a tool for effective classification of wood”. En: *Proceedings of the 51st International Convention of Society of Wood Science and Technology*. Concepción, 10-12 de Noviembre de 2008, 9 p.
- [88] TSUCHIKAWA, S. “A review of recent near infrared research for wood and paper”. *Applied Spectroscopy Reviews*. 2007, vol. 42, p. 43-71.
- [89] SALAMATI, N.; FREDEMBACH, C. AND SÜSSTRUNK, S. “Material classification using color and NIR images”. En: *Color and Image Conference*. Society for Imaging Science and Technology, 2009, p. 216-222.
- [90] BUTLER, D.A.; BRUNNER, C.C. AND FUNK, J.W. “Wood-surface feature classification using extended-color information”. *Holz als Rob- und Werkstoff*. 2001, vol. 59, p. 475-482.
- [91] ROJAS, J.A.M.; ALPUENTE, J.; BOLÍVAR, E. AND LÓPEZ-ESPÍ, P. “Empirical characterization of wood surfaces by means of iterative autocorrelation de laser speckles patterns”. *Progress in Electromagnetic Research*. 2008, vol. 80, p. 295-306.

- [92] YANG, H.; LEWIS, I.R. AND GRIFFITHS, P.R. “Raman spectrometry and neural networks for the classification of wood types. 2. Kohonen self-organizing maps”. *Spectrochimica Acta Part A*. 1999, vol. 55, p. 2783-2791.
- [93] PIURI, V. AND SCOTTI, F. “Design of an automatic wood types classification system by using fluorescence spectra”. *IEEE Transactions on System, Man and Cybernetics-Part C: Applications and Reviews*. 2010, vol. 40, no. 3, p. 358-366.
- [94] LABATI, R.D.; GAMASSI, M.; PIURI, V. AND SCOTTI, F. “A low-cost neural-based approach for wood types classification”. En: *Proceedings of International Conference on Computational Intelligence for Measurement Systems and Applications*. Hong Kong, 11-13 de Mayo de 2009, 5 p.
- [95] EPSON *Epson Perfection V700 Photo/V750 Pro. Manual de usuario* [en línea]. Disponible en: www.epson.es/es/es/corporatesite/products/mainunits/support/1760.
- [96] RASBAND, W. ImageJ (versión 1.47) [software]. National Institutes of Health. Disponible en: <http://imagej.nih.gov/ij/>
- [97] Waikato Environment for Knowledge Analysis-WEKA (versión 3.7.11) [software]. The University of Waikato. Disponible en: www.cs.waikato.ac.nz/ml/weka/
- [98] RASBAND, W.S. *ImageJ* [en línea]. Maryland: U.S. National Institutes of Health, 1997-2011. Disponible en: rsb.info.nih.gov/ij/
- [99] FERREIRA, T. AND RASBAND, W. *ImageJ users guide. IJ 1.46r* [en línea]. Maryland: U.S. National Institutes of Health, 198 p. Disponible en: rsb.info.nih.gov/ij/docs/guide/user-guide.pdf.
- [100] HALL, M.; FRANK, E.; HOLMES, G.; PFAHRINGER, B.; REUTEMANN, P. AND WITTEN, I.H. “The WEKA data mining software: an update”. *SIGKDD Explorations*. 2009, vol. 11, no. 1, p. 10-18.
- [101] BOUCKAERT, R.R.; FRANK, E.; HALL, M.; KIRKBY, R.; REUTEMANN, P.; SEEWALD, A. AND SCUSE, D. *Weka manual for version 3-7-11*. Hamilton: University of Waikato, 2014, 327 p.

- [102] ANSI. *Graphic technology - Color reflection target for input scanner calibration*. ANSI IT8.7/2-1993 (R2013). Washington: American National Standard Institute (ANSI), 2013.
- [103] RODRÍGUEZ, H. *Carta IT8 opacos. Instrucciones de uso*. Barcelona: EGM Laboratoris Color, 18 p.
- [104] ALPUENTE, J. “Clasificación automática de especies de madera basada en las características estadísticas de sus imágenes”. [Trabajo Fin de Máster]. Director: Juan Antonio Martínez Rojas. Universidad de Alcalá, Departamento de Teoría de la Señal y Comunicaciones, 2008.
- [105] ALPUENTE, J. *Televisión I. Sistemas de televisión*. Alcalá de Henares: Servicio de Publicaciones de la Universidad de Alcalá, 151 p.
- [106] ROJAS, J.A.M.; ALPUENTE, J.; ROJAS, I.M. AND VIGNOTE, S. “Fractal-based image enhancement techniques inspired by differential interference contrast microscopy”. *Journal of Optics A: Pure and Applied Optics*. 2009, vol. 11, 8 p.
- [107] WESZKA, J. S.; DYER, C. R. AND ROSENFELD, A. “A comparative study of textures measures for terrain classification”. *IEEE Transaction on Systems, Man. and Cybernetics*. 1976, vol. 6, no. 4, p. 269-285.
- [108] HARALICK, R.M.; SHANMUGAM, K. AND DISTEIN, I. “Textural features for image classification”. *IEEE Transactions on Systems, Man. and Cybernetics*. 1973, vol. SMC-3, no. 6, p. 610-3621.
- [109] HE. D.C.; WANG, L. AND GUIBERT, J. “Texture discrimination based on an optimal utilization of texture features”. *Pattern Recognition*. 1988, vol. 21, no. 2, p.141-146.
- [110] MANDELBROT, B.B. *Les objets fractals: forme, hasard et dimension*. París: Flammarion, 190 p.
- [111] HALL, M.A. “Correlation-based feature selection for machine learning”. [Tesis Doctoral]. Supervisor: Lloyd Smith. University of Waikato, Department of Computer Sciences, 1999.

- [112] HOTELLING, H. "Analysis of a complex of statistical variables into principal components". *Journal of Educational Psychology*. 1933, vo. 24, no. 6, p. 417-441.
- [113] BINGHAM, E. AND MANNILA, H. "Random projection in dimensionality reduction: Applications to image and text data". En: *Proceedings of the seventh ACM SIGKDD international conference on Knowledge discovery and data mining*. San Francisco, 26-29 de Agosto de 2001, p. 245-250.
- [114] DUDA, R.O.; HART, P.E. AND STORK, D.J. *Pattern classification*. Nueva York: John Wiley and Sons, 2001, 680 p.
- [115] KAISER, H.F. AND RICE, J. "Little jiffy, mark 4". *Educational and Psychological Measurement*. 1974, vol. 34, no. 1, p. 111-117.
- [116] QUINLAN, J.R. *C4.5: programs for machine learning*. San Francisco: Kaufmann Publisher Inc., 1993, 302 p.
- [117] COVER, T.M. AND HART, P.E. "Nearest neighbor pattern classification". *IEEE-Transactions on Information Theory*. 1967, vol. IT-13, p. 21-27.
- [118] BREIMAN, L. "Random forest". *Machine Learning*. 2001, vol. 45, no. 1, p. 5-32.
- [119] RODRÍGUEZ, J.J.; KUNCHEVA, L.I. AND ALONSO, C.J. "Rotation forest: a new classifier ensemble method. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 2006, vol. 28, no. 10, p. 1619-1630.
- [120] CORTES, C. AND VAPNIK, V.N. "Support-vector networks". *Machine Learning*. 1995, vol. 20, p. 273-297.
- [121] FAN, R.; CHANG, K.; HSSIEH, ; WANG, X. AND LIN, C. "LIBLINEAR: a library for large linear classification". *Journal of Machine Learning Research*. 2008, vol. 9, p. 1871-1874.