

# CAPÍTULO 3

## Detección de Patrones Visuales

El método empleado para la extracción de las características básicas de la imagen, patrones visuales, consiste en un conjunto de filtros orientados situados en distintos puntos del dominio frecuencial. De esta forma, obtendremos un banco de filtros sintonizados desde la más baja frecuencia hasta la más alta y en distintas orientaciones. Así podremos realizar un análisis de todo el espectro de la imagen y obtener las características frecuenciales que componen nuestra imagen.

Este banco de filtros, como hemos comentado, cubrirá todo el espectro de frecuencias, sin embargo, es de suponer que cualquier imagen no tiene contenido en todo el espectro, por lo que podemos plantearnos un método para poder reducir el coste computacional haciendo que filtros situados en zonas sin información relevante sean descartados. Estos filtros serán desactivados, reduciendo notablemente la carga computacional.

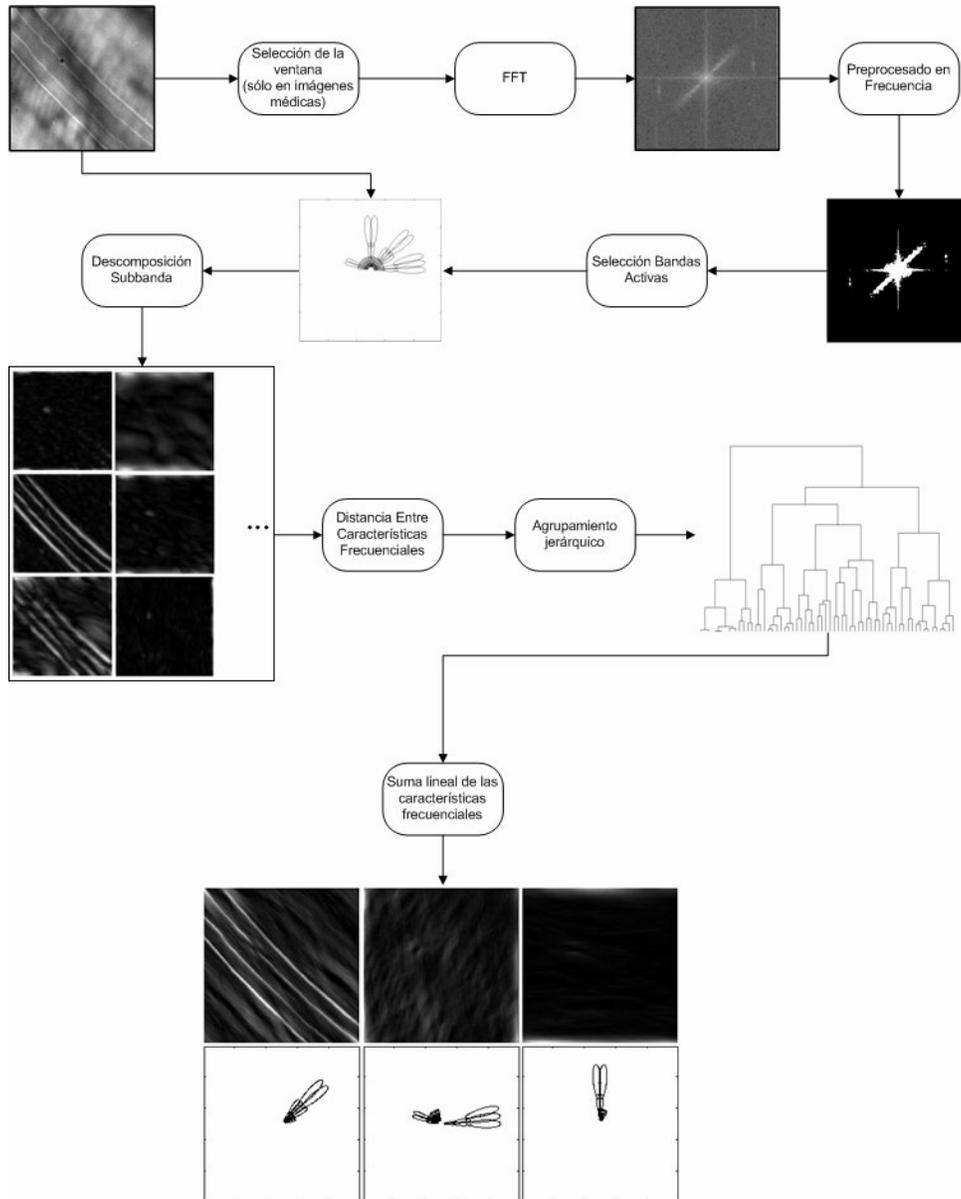
Una vez realizado el filtrado, obtendremos una serie de características frecuenciales o lo que es lo mismo, el contenido espectral de la imagen en todo el rango de frecuencias.

Para obtener los patrones visuales de nuestra imagen, tendremos que tener en cuenta la similitud de las distintas salidas de los filtros para agruparlas y finalmente obtener las estructuras que componen a la imagen original.

En breves palabras, los distintos pasos que componen el método empleado pueden describirse en los siguientes puntos.

1. Preprocesado de la imagen en el dominio de la frecuencia para conseguir una reducción del ruido y eliminación de elementos espurios que puedan activar filtros en zonas carentes de información.
2. En el caso de imágenes médicas, selección de la mejor ventana para extraer la estructura que queremos analizar, las cuales pueden ser estructuras blandas como tejido adiposo, muscular, tejidos duros como pueden ser las estructuras óseas, gases o líquidos.
3. Selección de un conjunto de filtros dentro del banco de filtros. Éstos, se activaran únicamente si existe información relevante, de lo contrario, permanecerán inactivos.
4. Generación de las características frecuenciales usando los filtros correspondientes.
5. Medida de la similitud entre pares de características frecuenciales.
6. Agrupamiento jerárquico de las distintas características frecuenciales usando una matriz, llamada matriz de distancias, la cual nos proporcionará la distancia existente entre pares de imágenes.
7. Reconstrucción de los patrones visuales correspondientes a cada grupo, mediante la suma lineal de las distintas características frecuenciales, teniendo en cuenta la mejor agrupación existente. Para ello nos basaremos en un índice que nos proporciona el agrupamiento óptimo y por tanto el número de patrones visuales existentes en la imagen.

Un esquema del algoritmo empleado se muestra en el siguiente esquema, donde se muestra el proceso desde la entrada de la imagen al sistema, su descomposición en características frecuenciales, posterior medida de similitud entre ellas y agrupamiento jerárquico para finalmente reconstruir los patrones visuales que componen la imagen.



**Ilustración 3.1** Diagrama de bloques del sistema de detección de patrones visuales implementado.

# 1. Banco de Filtros

Una de las técnicas más empleadas que modelan el sistema visual humano, es aquel que usa funciones de Gabor. Son funciones que se emplean en los bancos de filtros encargados del análisis y representación de imágenes multirresolución.

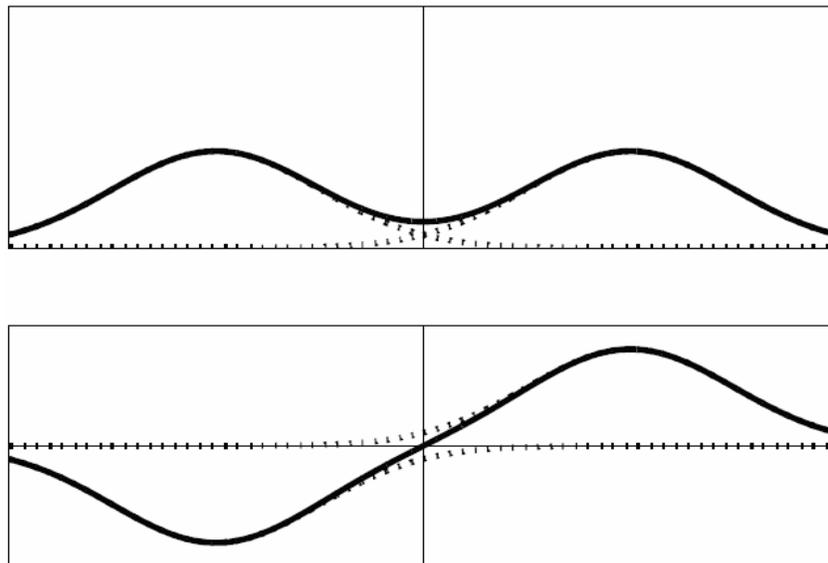
Una función Gabor está compuesta por una Gausiana modulada por una exponencial compleja. La versión unidimensional de su respuesta al impulso para una determinada frecuencia  $\rho_i$  y escala  $\sigma$  tiene la siguiente expresión [16].

$$h(x; \rho_i) = \exp\left(-\frac{x}{2\sigma}\right)^2 \cdot \exp(jx\rho_i) \tag{ec. 3.1.1}$$

donde se aprecia que tenemos una componente real e impar  $h^e$  y una componente imaginaria y par  $j \cdot h^o$ . De este modo, el filtro podrá, simultáneamente procesar las componentes pares e impares de la imagen, es decir podremos, perfectamente, determinar la energía y la fase local de la imagen.

En el dominio de la frecuencia, las funciones Gabor, presentan una Gausiana centrada en la frecuencia de modulación.

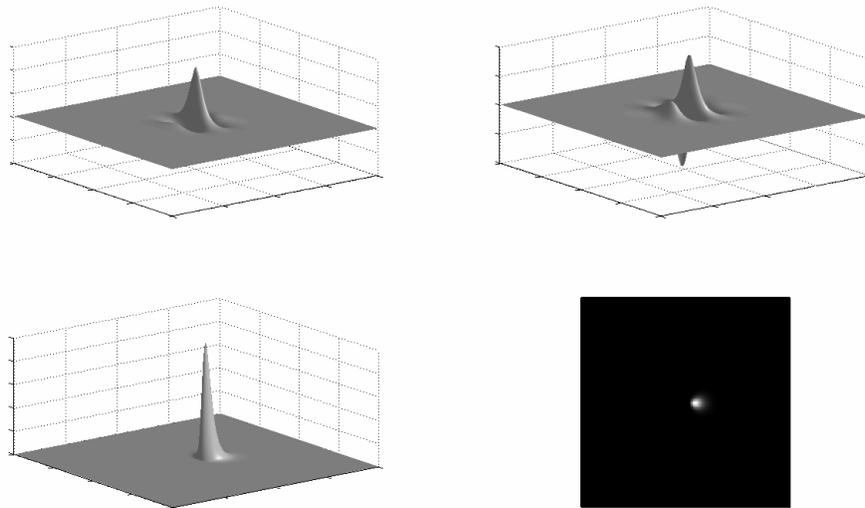
En el siguiente esquema podemos apreciar las partes par e impar de una función Gabor determinada.



**Ilustración 3.2** Función de Gabor en el dominio del tiempo. Superior: Componente par. Inferior: componente impar.

Para nuestro sistema emplearemos una versión modificada de las funciones Gabor. Esta modificación se basa fundamentalmente en emplear la versión logarítmica de dichas funciones, las cuales se conocen como funciones logGabor. La razón de emplear esta modificación se basa en que las funciones Gabor presentan una limitación, al conseguir únicamente filtros con máximo ancho de banda de 1 octava [5].

Para poder realizar el procesado de una imagen, deberemos extender la versión unidimensional de las funciones logGabor a dos dimensiones, de forma que presentaría el siguiente aspecto.



**Ilustración 3.3** Función de logGabor. Superior Izquierda: Componente par en el dominio del tiempo. Superior Derecha: Componente impar en el dominio del tiempo. Inferior Izquierda: Filtro de Gabor en el dominio de la frecuencia. Inferior Derecha: Planta del filtro en el dominio de la frecuencia.

Dicho sistema, el cual se expresa en coordenadas polares, se compone de diferentes filtros, cada uno de los cuales abarcará un espacio comprendido desde la frecuencia de ángulo 0 hasta la frecuencia de ángulo  $\pi$ , y en diferentes escalas que cubrirán la zona comprendida desde frecuencias bajas hasta las altas.

De esta forma conseguimos filtros orientados en todas las direcciones, de manera que la frecuencia de ángulo 0 corresponde con estructuras orientadas verticalmente, las de frecuencia  $\pi/2$  a estructuras orientadas horizontalmente y los ángulos comprendidos entre 0 y  $\pi$ , (éste último sin incluir, pues tiene la misma orientación que el ángulo 0) a estructuras diagonales en mayor o menor medida.

Cada uno de estos filtros, tendrá pues, dos componentes que modulen la orientación y la escala, las cuales denominaremos *componente angular* y *componente radial*, respectivamente.

La componente angular, como se ha comentado, barrerá la zona comprendida entre 0 y  $\pi$  y tendrá, en el dominio de la frecuencia, la siguiente expresión [5].

$$A(\theta) = \exp\left(\frac{-(\theta - \theta_0)}{2\sigma_\theta}\right) \quad (\text{ec. 3.1.2})$$

donde  $\theta_0$  es la dirección del filtro,  $\theta$  es la frecuencia expresada en coordenadas polares y  $\sigma_\theta$  es la desviación estándar, que nos dará el ancho de banda angular como sigue.

$$\Delta\Omega = 2\sigma_\theta\sqrt{2\ln 2} \quad (\text{ec. 3.1.3})$$

La componente radial del filtro barrera la zona comprendida desde la más baja frecuencia hasta la zona de altas frecuencias y tendrá, en el dominio de la frecuencia, la siguiente forma [5].

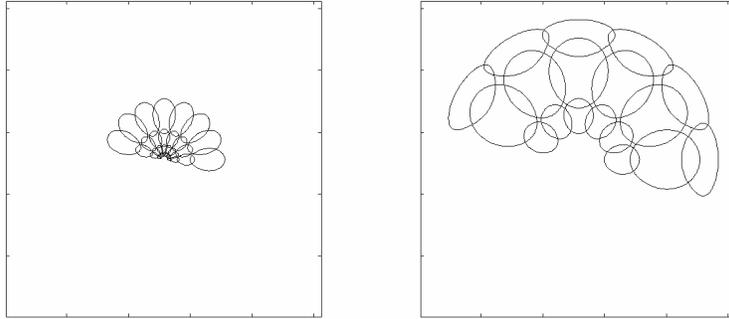
$$R(f) = \exp\left(\frac{-[\ln(f/f_0)]^2}{2[\ln(\sigma_f/f_0)]^2}\right) \quad (\text{ec. 3.1.4})$$

donde  $f_0$ , la frecuencia central del filtro,  $f$  la frecuencia y  $\sigma_f$  la desviación estándar que nos facilitará el ancho de banda radial, dado por la siguiente expresión.

$$B = 2\sqrt{2\ln 2} \cdot [\ln(\sigma_f/f_0)] \quad (\text{ec. 3.1.5})$$

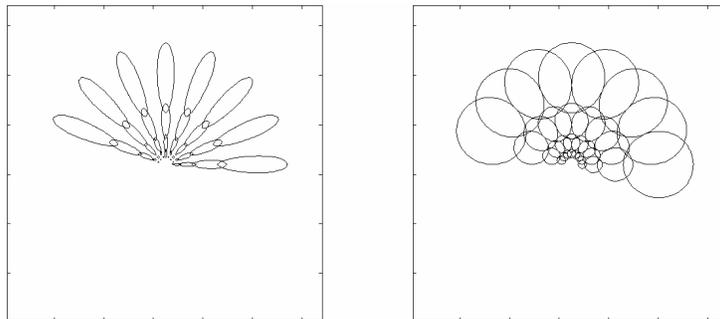
Para poder implementar nuestro banco de filtros, haremos una descripción de los parámetros que utiliza en cada ejecución:

- **Imagen de Entrada:** al banco se le debe introducir la imagen a la cual se le desea calcular sus patrones visuales.
- **Número de escalas y orientaciones:** nos darán el número total de filtros y la posición de los mismos en el espectro de frecuencias.
- **Mínima longitud de onda:** con este parámetro, podremos determinar si queremos centrarnos en las bajas y/o altas frecuencias. Ésta se encuentra relacionada con la inversa de la frecuencia, de modo que si tenemos una longitud de onda mínima pequeña, la frecuencia máxima será elevada y cubriremos sin problemas la zona de altas frecuencias, en cambio si la longitud de onda mínima es elevada, la frecuencia máxima será más pequeña y no tendremos tan cubierta la zona de altas frecuencias. Debemos tener cuidado con esto, pues según la elección de este parámetro estaremos o no definiendo el banco de manera adecuada. En el siguiente ejemplo, vemos que tanto por defecto, como por exceso, no estamos creando un banco con buenas características, pues en el primer caso no cubrimos la zona de alta frecuencia, y en el segundo, además de no cubrir la zona de bajas, en la de altas tenemos un filtro recortado al salirse de los márgenes permitidos.



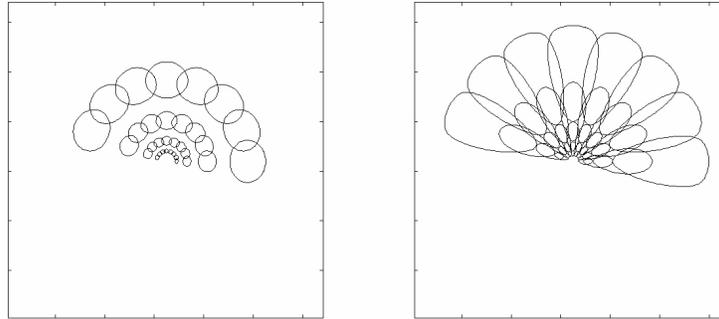
**Ilustración 3.4** Izquierda: longitud de onda mínima 8 píxeles. Derecha: longitud de onda mínima 1 píxel.

- **Factor multiplicativo:** es el modo que tenemos de obtener las siguientes longitudes de onda. En el caso de que sea igual a 2, tendremos las frecuencias separadas por octavas, de manera que  $\lambda_i = \lambda_{\min} \cdot 2^{(i-1)}$  con  $i = \{1..Ns\}$ , siendo Ns el número de escalas.
- **Desviaciones estándar de las componentes angulares y radiales:** como se ha comentado previamente, estos parámetros marcan la desviación respecto a dichas componentes, marcando el ancho de banda que va a poseer el filtro. Podemos manipular las desviaciones angulares y radiales del filtro, obteniendo así diferentes configuraciones, en concreto manipulando la desviación estándar de la componente angular de la frecuencia, obtenemos los siguientes bancos de filtros.



**Ilustración 3.5** Banco de filtros. Izquierda: Desviación estándar de la componente angular disminuyendo. Derecha: Desviación estándar de la componente angular aumentando.

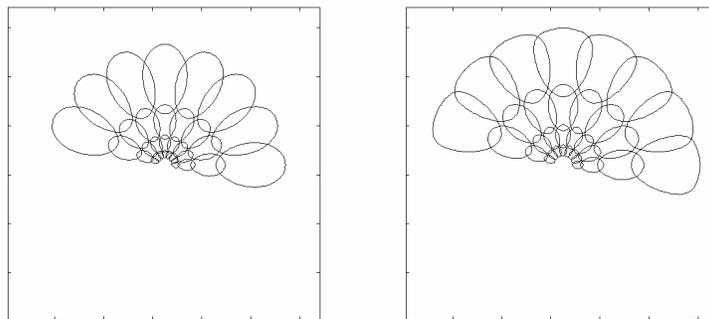
Igualmente, si manipulamos la desviación estándar de las componentes radiales, obtenemos los siguientes bancos.



**Ilustración 3.6** Banco de filtros. Izquierda: Desviación estándar de la componente radial disminuyendo. Derecha: Desviación estándar de la componente radial aumentando.

Como se puede apreciar hay que elegir bien estas desviaciones, pues, tanto por encima, como por debajo del valor correcto, estamos procesando de manera incorrecta, ya sea por defecto, al no cubrir todas las frecuencias, como por exceso, al realizar dobles procesados en ciertas zonas frecuenciales. Por ello, mantendremos las desviaciones a lo largo de las distintas ejecuciones que realicemos en el programa. Los valores que hemos escogido son 0.65 para la desviación estándar de la componente radial y 1.5 para la componente angular.

De este modo, como apreciamos en la siguiente ilustración, si nuestro banco estuviera compuesto por 32 filtros, distribuidos en 4 escalas, empezando por la de longitud de onda 4 píxeles (izquierda) y 3 píxeles (derecha), y en múltiplos de 2, es decir, 4, 8, 16 y 32 (izquierda) y 3, 6, 12 y 24 (derecha) y 8 orientaciones, equiespaciadas entre 0 y  $\pi$ , entonces nos encontraríamos con el siguiente sistema, cuya planta representamos a continuación, mediante la elección de un isocontorno de valor  $\exp(-1/2)$ .



**Ilustración 3.7** Banco de filtros. Planta del mismo tomada con un isocontorno de valor  $\exp(-1/2)$ . Izquierda: longitud de onda mínima 4 píxeles. Derecha: longitud de onda mínima 3 píxeles.

## 2. Selección de la Ventana

Como hemos comentado anteriormente, en el caso de imágenes médicas, debemos elegir previamente la mejor ventana para extraer el tipo de estructura corporal que queramos analizar. En este apartado comentaremos distintos conceptos que necesitaremos para entender el correcto funcionamiento de la selección de la ventana.

### 2.1. Formato DICOM

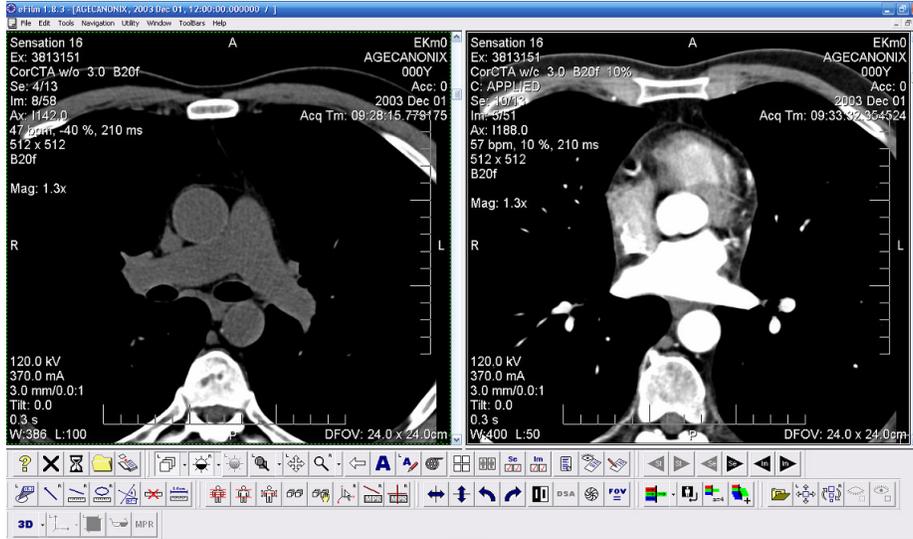
El formato DICOM (Imágenes y comunicaciones digitales en el ámbito de la medicina, del inglés Digital Imaging and Communications in Medicine) se utiliza habitualmente para la transferencia y almacenamiento de imágenes médicas, como ultrasonidos y escáneres. Los archivos DICOM contienen tanto datos de imagen, como cabeceras, que almacenan información sobre el paciente y la imagen médica.

Estos archivos suelen almacenarse en dispositivos ópticos administrados en diferentes carpetas, cada una de las cuales pueden contener información acerca de un determinado paciente y de los diferentes estudios que se le han hecho al mismo.

Toda esta información se encuentra gobernada por un archivo llamado DICOMDIR, el cual es una estructura, que proporciona un índice e información de resumen para cada uno de los ficheros DICOM del contenido. La estructura del DICOMDIR proporciona información sobre cada fichero que se encuentra presente en el estudio, de manera que hay menos necesidad de nombres de fichero con significado. Este archivo puede contener información tan diversa como puede ser el tamaño de la imagen, profundidad de bit usada para la codificación de la escala de grises, versión del estándar DICOM, modo de almacenamiento, tiempo de duración del estudio, modo de adquisición de las imágenes, identificación del paciente, orientación del mismo a la hora de realizar el estudio, así como otros tantos parámetros necesarios para la correcta interpretación de la imagen que se esta examinado.

Las imágenes DICOM tienen extensión .dcm y pueden ser interpretadas por cualquier visualizador de imágenes médicas. Estos programas se conocen por el nombre de *visualizador DICOM*, los cuales no sólo representan los estudios, sino que tienen numerosas herramientas entre las que destacan la capacidad de realizar mediciones y la transformación multiplanar, la cual permite obtener las vistas sagitales y coronales a través de la vista axial, forma natural que tienen los dispositivos captadores de adquirir la imagen.

Como ejemplo de visualizador tenemos el siguiente programa, llamado eFilm, el cual ha sido cargado con una imagen DICOM que representa un TAC de tórax. En ella, podemos apreciar la bifurcación del tronco pulmonar, a la izquierda y una vista de la aurícula izquierda, a la derecha.

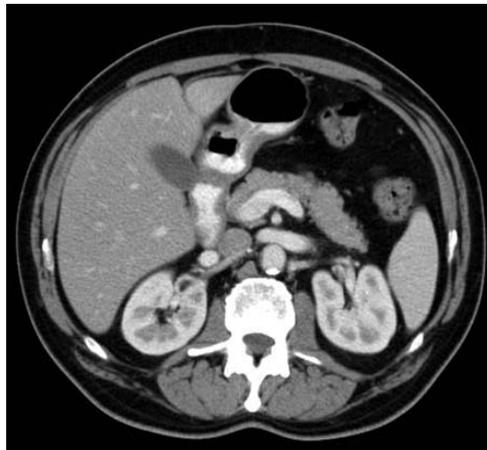


**Ilustración 3.8** Captura de pantalla del visualizador DICOM eFilm.

### 2.1.1. Tomografía Axial Computarizada.

Una Tomografía Axial Computarizada, o TAC, es una exploración de rayos X que produce imágenes detalladas de cortes axiales del cuerpo. En lugar de obtener una imagen como la radiografía convencional, el TAC obtiene múltiples imágenes al rotar alrededor del cuerpo. Una computadora combina todas estas imágenes en una imagen final que representa un corte del cuerpo como si fuera una rodaja o *slice*.

Se trata de una técnica de visualización por rayos X. Podríamos decir que es una radiografía de una fina rodaja obtenida tras cortar un objeto de manera axial.



**Ilustración 3.9** Imagen de TAC en donde podemos apreciar tejido blanco y estructuras óseas.

### 2.1.2. Imagen de Resonancia Magnética

La Imagen de Resonancia Magnética, o RMI, es una técnica de diagnóstico sofisticada que utiliza un campo magnético, ondas de radio y una computadora para producir imágenes de sección transversal de distintas partes del cuerpo. La RMI produce mejores exploraciones del tejido blando y se utiliza, en general, para producir imágenes del cerebro, la columna vertebral, los órganos en el tórax y el abdomen, el corazón y los vasos sanguíneos y estructuras próximos a los huesos, tales como los músculos y ligamentos.



**Ilustración 3.10** Imagen de RM en donde podemos apreciar tejido blanco y estructuras óseas.

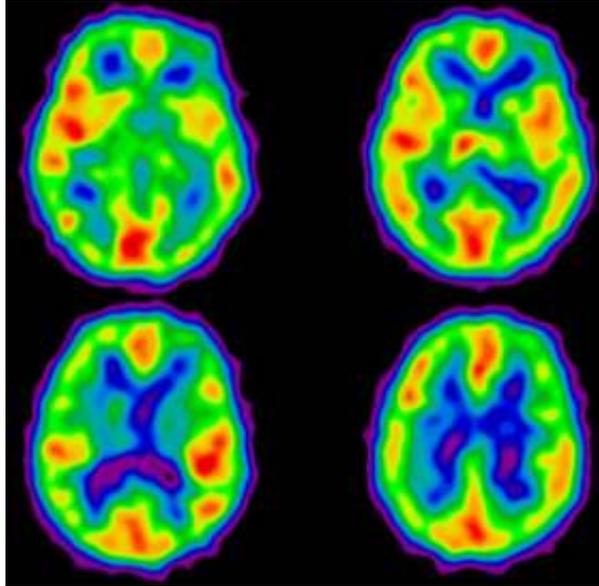
### 2.1.3. Tomografía por Emisión de Positrones

La Tomografía por Emisión de Positrones, o PET, es un examen de diagnóstico que consiste en obtener imágenes fisiológicas basadas en la detección de radiación emitida por positrones. Los positrones son pequeñas partículas emitidas por una sustancia radiactiva que se le administra al paciente. Las imágenes del cuerpo humano obtenidas con esta técnica se utilizan para evaluar diversas enfermedades.

Los estudios de PET se usan más frecuentemente para detectar cáncer y estudiar el efecto del tratamiento mediante la caracterización bioquímica de los cambios producidos en dicha enfermedad, y se pueden hacer en todo el cuerpo. Los estudios de PET del corazón se pueden usar para determinar el flujo sanguíneo al mismo y evaluar los signos de enfermedad coronaria. Se usan también para determinar cuáles zonas de funcionamiento reducido siguen aún vivas, y cuáles son tejido cicatrizal causado por un infarto previo. Junto con el estudio de perfusión miocárdica, los PET permiten diferenciar entre el músculo cardíaco no funcional y el músculo cardíaco todavía viable que se podría beneficiar de un procedimiento como angioplastia o bypass (puente) coronario, que restablecen el flujo sanguíneo y mejoran el funcionamiento del corazón.

Los PET del cerebro se usan para evaluar a los pacientes con trastornos de la memoria de causa no determinada, tumores de cerebro sospechados o comprobados, o

convulsiones que no responden al tratamiento médico y por lo tanto son candidatas para cirugía.



**Ilustración 3.11** Imagen obtenida por PET, en donde podemos apreciar las diferentes radiaciones de positrones en contacto con el tejido cerebral.

## 2.2. Ventanas usada en Imágenes Médicas

Estas ventanas hacen referencia a la forma de visualización óptima que debe emplearse en una imagen DICOM según el tipo de estructura que se desea analizar. Se mide en unidades *Hounsfield*, en honor al ingeniero inglés del mismo nombre que consiguió el primer prototipo de Tomografía Axial Computarizada. El concepto de ventana puede prestar a confusión, puesto que no se realiza ningún tipo de recorte en la imagen a analizar. El modo que tiene de funcionar una ventana es mediante la manipulación de la escala de grises del histograma de la imagen DICOM. De este modo, podremos extraer diferentes texturas encontradas en el organismo, escogiendo *ventanas* en el histograma y realizando posteriores ecualizaciones si es necesario. Este modo de proceder lo llevan implantados todos los visualizadores DICOM y con un simple juego de ratón podemos cambiar la ventana según las necesidades.

Las principales ventanas que se usan en el sector hospitalario son:

- **Ventana de mediastino:** ventana utilizada para la visualización de tejidos blandos, principalmente tejido muscular.
- **Ventana de parénquima:** gracias a esta ventana podremos apreciar estructuras pulmonares. Es una ventana, principalmente, utilizada en el estudio de tórax, pero también empleada en los estudios abdominales para

detectar la presencia de gases fuera de las asas intestinales, diagnóstico de neumoperitoneo.

- **Ventana de hueso:** como su propio nombre indica, esta ventana se usa para la visualización de estructuras óseas, principalmente la cubierta de los huesos o zona cortical, y la cavidad medular cuyo contenido es la médula ósea amarilla.

### 2.3. Ventanas Implementadas en Nuestro Sistema

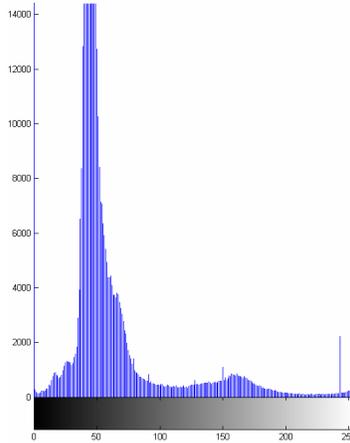
Una vez definidos estos conceptos, entraremos en detalle sobre el método de ventana que se ha implementado en nuestro sistema. Como se ha comentado, este sistema profundizará en la extracción de características en imágenes con estructuras orientadas. Esto ocurre con las estructuras vasculares, como las arterias y las venas, y las estructuras óseas.

En la siguiente imagen, podemos apreciar una imagen obtenida mediante una Tomografía Axial Computarizada a la que se le ha realizado una transformación multiplanar para obtener una vista coronal. En concreto nos encontramos con una vista anteroposterior en donde nos encontramos con estructuras vasculares, estructuras óseas y tejido blando.



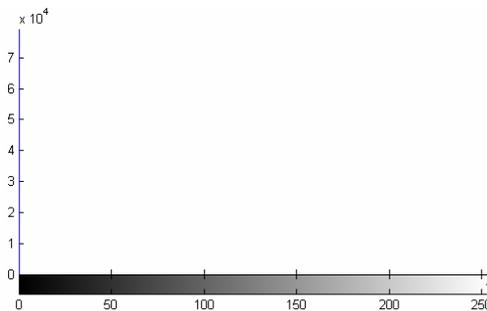
**Ilustración 3.12** Imagen de TAC. Vista coronal.

Esta imagen presenta el siguiente histograma, el cual se encuentra distribuido en una escala desde 0, que corresponde al nivel de gris negro, y 255, correspondiente al nivel de gris, blanco. Como era de esperar, este histograma se concentra en los niveles de gris más oscuros.



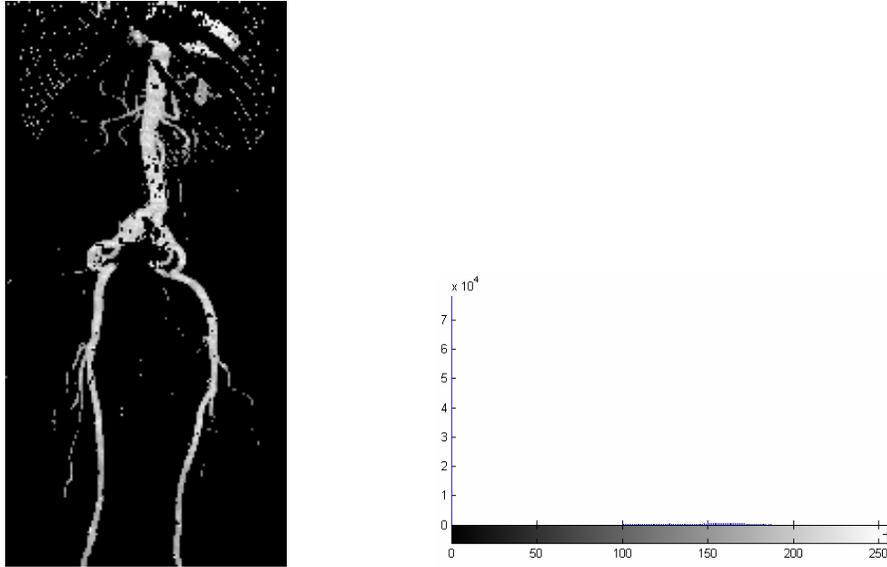
**Ilustración 3.13** Histograma de la imagen anterior.

Se puede apreciar que existe una gran ocurrencia de un nivel cercano al 255, el blanco, que se corresponde con las estructuras óseas. Por lo que si realizamos una extracción de este nivel de gris, escogiendo valores del histograma, podremos segmentar dichas estructuras.



**Ilustración 3.14** Izquierda: TAC en donde se han segmentado las estructuras óseas. Derecha: Histograma del mismo.

Igualmente, las estructuras vasculares se encuentran en las zonas intermedias del histograma, comprendidas entre 100 y 200, aproximadamente, por lo que escogiendo esos valores podremos conseguir extraer estos vasos sanguíneos.



**Ilustración 3.15** Izquierda: TAC en donde se han segmentado las estructuras vasculares. Derecha: Histograma del mismo.

De este modo, ya tenemos preparada la imagen para extraer según qué tipo de estructuras las distintas orientaciones y escalas.

### 3. Preprocesado en el Dominio de la Frecuencia

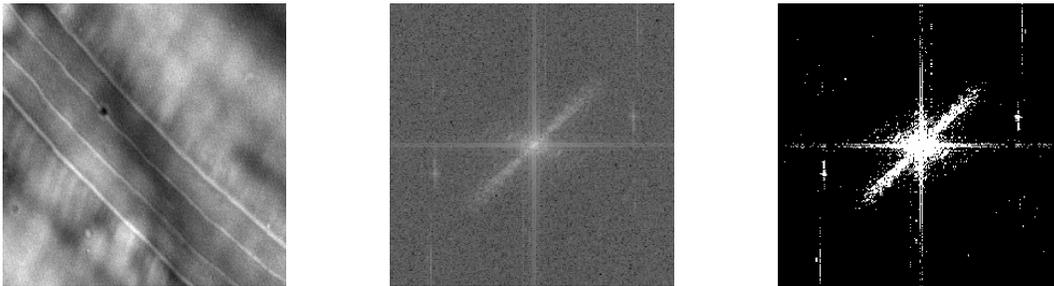
Una vez explicado el funcionamiento del banco de filtros y la elección de la ventana de interés en el caso de imágenes médicas, explicaremos la necesidad de un preprocesado en el dominio de la frecuencia. Dicho preprocesado tiene como finalidad obtener una imagen espectral pura, ausente de ruido y sin elementos espurios que hagan que el proceso no tenga una eficiencia óptima.

Esto es debido al hecho de que nuestro sistema tendrá en cuenta la información en el dominio de la frecuencia, es decir, activaremos los distintos filtros del banco de filtros únicamente si existe contenido frecuencial relevante. Si dicho contenido es ruido, la característica obtenida tras la salida del filtro, será precisamente ruido. Por ello, debemos obtener una imagen espectral lo más limpia posible.

Como es de suponer, una imagen no estará definida en todo el rango de frecuencias, por lo que debemos obtener un sistema que sólo contemple aquellos filtros que realmente son necesarios para extraer las características frecuenciales de una imagen, es decir, puede suceder que muchos de ellos extraigan elementos irrelevantes, esto es, estén situados en zonas donde la imagen no tiene ninguna información frecuencial. Es por esto, que necesitamos realizar el preprocesado del espectro y obtener

una imagen espectral lo más limpia posible, de forma que los filtros que caigan encima de zonas donde la imagen no tenga información sean descartados.

Para realizar este preprocesado, nos valemos de la imagen en el espacio de la frecuencia, es decir, obtendremos la energía de la imagen y dicha representación lo trataremos como una imagen a procesar. Para ello, calculamos una imagen de energía con la siguiente expresión  $F = \log(|F(X)| + 1)$ , siendo  $F(X)$  la transformada de Fourier de la imagen. Este mapa de energía nos proporcionará las estructuras relevantes de la imagen más un cierto nivel de ruido de fondo. Para eliminar este ruido y así evitar activaciones de filtros innecesarias, consideraremos que dicho ruido es plano, distribuido igualmente por todo el espectro [6]. En la zona de altas frecuencias, no existirán componentes de nuestra imagen por lo que, únicamente habrá ruido. Por tanto, si nos desplazamos hasta dicha zona, podremos calcular ese umbral de ruido. Lo que haremos es hallar el valor medio de la energía de ruido en alta frecuencia y sustraeremos ese umbral de ruido más un margen de seguridad a la imagen original en frecuencia, como podemos observar en las siguientes figuras.



**Ilustración 3.16** Izquierda: Imagen microscópica de un organismo determinado. Centro: Transformada de Fourier de la misma. Derecha: Substracción del ruido y posterior umbralización de la imagen de la transformada anterior.

Es decir, a nuestro mapa de energía le sustraeremos una cantidad  $m + X\sigma$  y, posteriormente realizaremos una umbralización, colocando todos los píxeles de valor 0 o negativo en negro y los mayores que 0 en blanco.

En la anterior expresión nos encontramos con  $m$ , que representa el ruido medio en altas frecuencias y  $X\sigma$  el margen de seguridad, compuesto por un factor  $X$ , que dependerá de la imagen de entrada y  $\sigma$  la desviación estándar del ruido en alta frecuencia

Dicho factor  $X$  se ha calculado experimentalmente y tomará los siguientes valores:

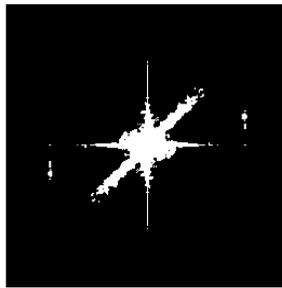
- $X=2$  para imágenes *estándar*, las cuales serán aquellas que apenas posean nivel de ruido y su espectro se concentre en las bajas frecuencias.
- $X < 2$  para imágenes sin ruido o para imágenes con gran contenido en zonas altas del espectro de frecuencia. De este modo, si la imagen tiene,

principalmente, contenido de alta frecuencia, con este margen menos agresivo conseguiremos mantener dichas componentes.

- $X > 2$  para imágenes muy ruidosas y que principalmente tenga contenido de baja frecuencia. Así, con este margen más agresivo, haremos que únicamente los filtros de baja frecuencia sean activados, ganando tiempo de computación.

Procediendo de este modo, tan sólo los filtros que caigan en zonas relevantes en cuanto a contenido de información serán los únicos en activarse, mientras que los demás serán descartados y no realizarán ningún tipo de filtrado, ahorrando así carga computacional.

Sin embargo, como podemos apreciar en la figura de la derecha de la ilustración 3.16, vemos que todavía existen elementos espurios que pueden activar ciertos filtros situados encima de ellos. Esto podemos solucionarlo con un filtrado de mediana sobre el mapa de energía al cual acabamos de substraerle el ruido de fondo, obteniendo así la siguiente imagen.



**Ilustración 3.17** Imagen obtenida tras la aplicación del filtrado de mediana en la imagen de la transformada de Fourier umbralizada.

Como podemos apreciar, se conservan las estructuras esenciales de la imagen, puesto que observamos que siguen permaneciendo componentes en baja frecuencia y una estructura diagonal, a frecuencias medias que corresponden con las formas lineales de la imagen original.

Así, ya hemos conseguido una imagen limpia y activar sólo los filtros que realmente son necesarios.

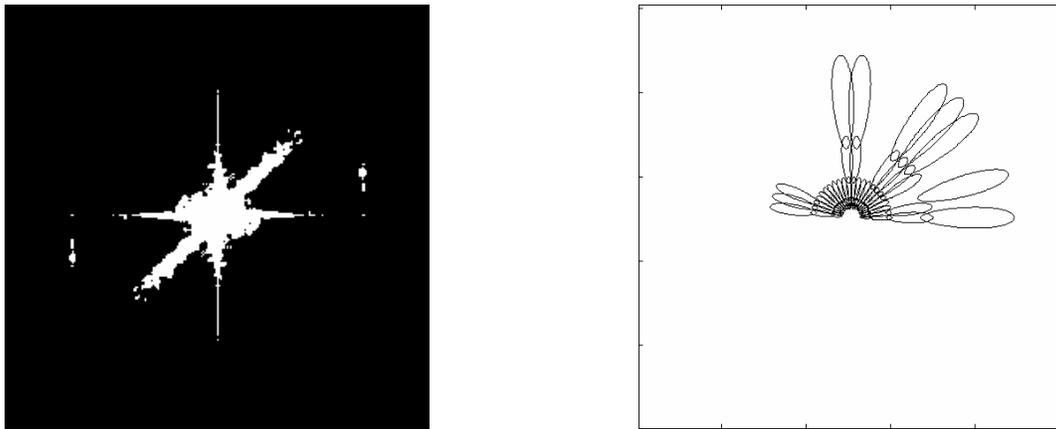
## 4. Funcionamiento del Sistema

En este momento, tras la selección de la ventana y el preprocesado espectral, nuestra imagen se encuentra preparada para ser procesada por el sistema de banco de filtros logGabor.

Como acabamos de comentar, este sistema usará la transformada de Fourier de la imagen para determinar qué filtros deben activarse y así reducir la carga computacional.

Esto lo conseguiremos activando los filtros que caigan encima de componentes espectrales superiores a un determinado valor, el cual se ha determinado que está situado en el valor numérico  $\exp(-1/2)$ . Se ha tomado este valor, pues aunque los filtros de Gabor tienen soporte infinito, la mayor parte de su energía se encuentra localizada en una banda cuya amplitud más baja toma valores cercanos al expuesto [16]. De este modo, si existe alguna componente en la imagen de la transformada de Fourier, previamente procesada para evitar elementos espurios, el filtro situado encima de ella permanecerá activo, de lo contrario se desactivará.

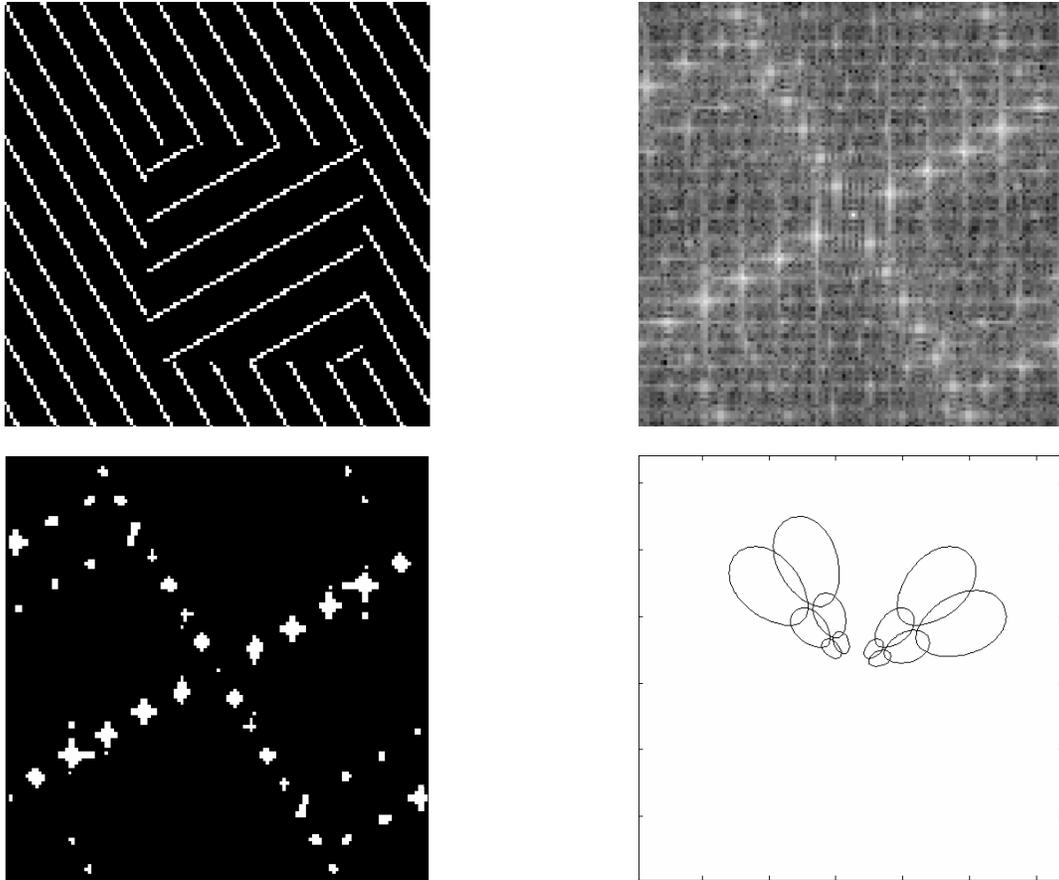
Como ejemplo, tenemos la imagen anterior, la cual, tras el preprocesado de ruido, umbralización y posterior filtrado de mediana, haría que se activaran los siguientes filtros.



**Ilustración 3.18** Izquierda: Espectro de la imagen que estamos analizando tras el filtrado de mediana y umbralización. Derecha: Filtros asociados a dicho espectro.

Se puede apreciar, como el sistema funciona perfectamente, activando únicamente los filtros en donde existen componentes espectrales. Cabe aclarar, la simetría de la transformada de Fourier al ser la imagen una señal real, por lo que tenemos que tener en cuenta que los filtros situados en zonas alrededor de  $\pi/4$ , correspondientes a la estructura diagonal, también actuarán en la zona de  $5\pi/4$ , igual sucede con todas las orientaciones.

En este otro ejemplo, podemos apreciar dos texturas, cada una con diferente orientación. Debido a la simplicidad de la misma, vamos a procesar con 32 filtros, 4 para cada una de las escalas 4, 8 y 16 y 32, y 8 orientaciones. Vemos que nuevamente el sistema escoge los filtros necesarios para su tratamiento.



**Ilustración 3.19** Superior Izquierda: Imagen de entrada. Superior Izquierda: Espectro de la imagen anterior. Inferior Derecha: Imagen tras eliminación de ruido, umbralización y filtrado de mediana usada para la determinación de los filtros que han de activarse. Inferior Izquierda: Filtros asociados.

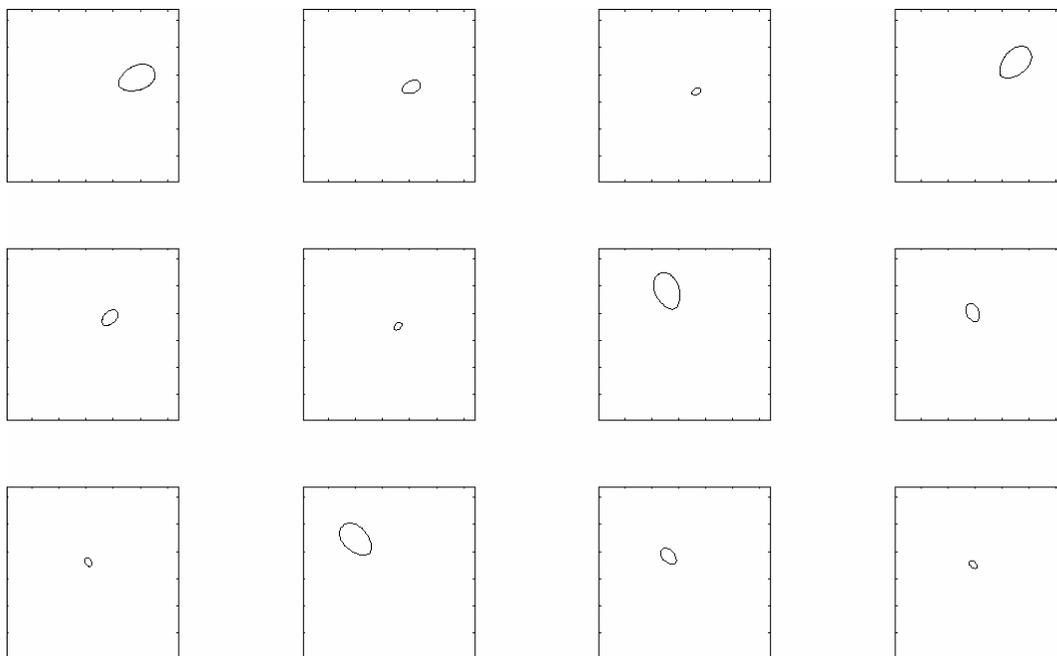
Una vez que hemos determinado cuáles son los filtros que deben ser activados, debemos introducir nuestra imagen en el sistema y encontrar todas las características frecuenciales que la componen.

Como ejemplo, escogeremos la imagen de la ilustración anterior y veremos cuáles son sus características frecuenciales.

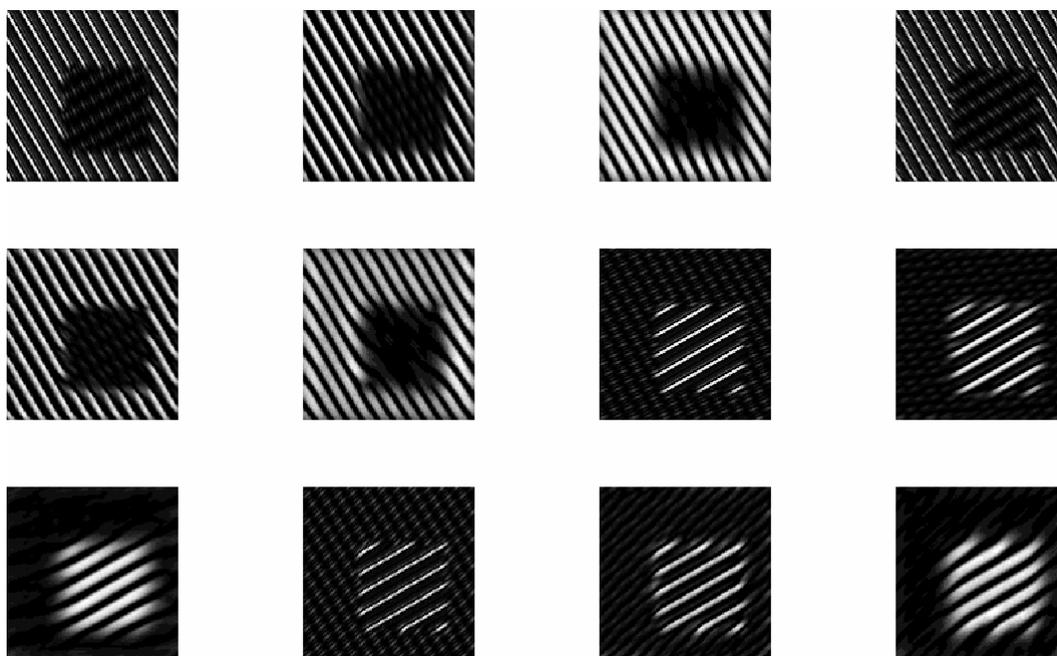
En la siguiente página podemos apreciar cuáles son los filtros activados y la salida de cada uno de ellos. Vemos que los filtros situados en las altas frecuencias obtienen imágenes con cambios más bruscos, mientras que los situados en las bajas, obtienen imágenes más difuminadas.

Debemos aclarar la orientación del banco de filtros. Las estructuras verticales son obtenidas por filtros cuya respuesta tiene lóbulos horizontales y viceversa. Con las estructuras diagonales sucede lo mismo.

Este giro de 90° antihorario es la forma natural que tienen de presentarse los bancos de filtros logGabor y no debe llevar a ningún tipo de confusión.



**Ilustración 3.20** Filtros obtenidos en el procesado de la imagen de la ilustración 3.17



**Ilustración 3.21** Características frecuenciales asociadas a la imagen de la ilustración 3.17.

## 5. Medida de la Similitud entre Características Frecuenciales

En este momento nos encontramos con todas las salidas de los filtros, previamente escogidos según nos indicaba el espectro de la imagen, y debemos agruparlas según su semejanza.

Es por ello que, una medida que refleje la dependencia estadística entre todos los mapas de energía o características frecuenciales extraídas de cada filtro podría representar una buena medida de similitud para el agrupamiento jerárquico posterior.

Vamos a analizar, a continuación, el método para obtener de dependencia estadística entre mapas de energía de las características frecuenciales obtenidas por cada uno de los filtros activos, el cual recibe el nombre de NMI o Información Mutua Normalizada. En dicho método, usaremos conceptos basados en la Teoría de la Información, por lo que haremos un breve repaso a los mismos.

### 5.1. Conceptos sobre la Teoría de la Información

En este punto, trataremos de relacionar las bases de la Teoría de la Información con las salidas que estamos obteniendo en nuestro sistema para poder tratarlas como variables aleatorias.

Definimos una variable  $X$  como *aleatoria* cuando conocemos los posibles valores, llamados eventos, que puede tomar dentro de un conjunto  $\Omega_X$ , pero no conocemos con certeza el número de ocurrencias de dichos valores. Precisamente, esta ocurrencia se mide con el concepto de *probabilidad*, la cual nos da una idea de la repetición de los distintos valores o eventos que puede tomar la variable aleatoria.

La forma que tenemos de representar los distintos eventos que puede tomar una variable aleatoria lo realizamos con el *histograma*, el cual es un vector que describe la probabilidad de tomar todos los valores de su espacio de eventos. De este modo, volviendo a nuestras características frecuenciales, podemos hacernos una idea de los distintos valores que puede tomar la imagen dentro del rango de la escala de grises, el cual con 8 bits se obtienen valores desde el 0, que correspondería al negro, hasta el 255, correspondiente al blanco.

Pero lo que realmente, tratamos de realizar es observar la similitud entre pares de variables aleatorias, para poder así obtener la matriz de distancias, que representaría el grado de semejanza entre las imágenes obtenidas, tomadas de dos en dos. Para ello, tenemos que definir la idea de *probabilidad conjunta* y con ella la *distribución de probabilidad conjunta*  $P(X,Y)$ , cuyo histograma nos ofrecerá una visión de la similitud entre dos variables, o como estamos tratando, entre dos imágenes.

### 5.1.1. Entropía e Información Mutua.

Una vez hechas las definiciones de variable aleatoria y probabilidad, ya sea para una variable o para un par de variables aleatorias, debemos cuantificar este grado de aleatoriedad existente entre ellas. Para ello, debemos hablar de de *entropía*, la cual nos marcará el grado de incertidumbre presente en una variable aleatoria. Esta se define del siguiente modo.

$$H(X) = \sum_i \log(P_X(x_i)) \frac{1}{P_X(x_i)} \quad (\text{ec. 3.5.1})$$

Puede apreciarse, gracias a la expresión, que si una variable X no es aleatoria, su incertidumbre o entropía es  $H(X) = 0$ . Por otro lado, si una variable aleatoria es equiprobable, esto es, todos sus eventos tienen igual probabilidad de suceder, la entropía toma el valor máximo, siendo éste  $\log(N)$ , donde N es el tamaño de muestras del dominio de eventos.

Nuevamente, podemos extender el concepto a dos variables aleatorias, obteniendo así la *entropía conjunta*  $H(X,Y)$ , la cual definimos con la siguiente expresión.

$$H(X, Y) = \sum_i \sum_j P_{X,Y}(x_i, y_j) \frac{1}{P_{X,Y}(x_i, y_j)} \quad (\text{ec. 3.5.2})$$

Otro modo de obtener información conjunta sobre dos variables aleatorias es usando la *Información Mutua*, que representa la disminución de incertidumbre de la primera variable con el hecho del conocimiento de la segunda. Una forma de definirla, que es como nos interesa, sería haciendo uso de la entropía tanto para una variable, como para la conjunta.

$$I(X, Y) = H(X) + H(Y) - H(X, Y) \quad (\text{ec. 3.5.3})$$

### 5.1.2. Medidas Normalizadas de Información Mutua

En este momento nos encontramos con un problema, pues podemos encontrarnos con que las entropías conjuntas pueden darnos poca información tanto porque exista una fuerte dependencia entre variables o porque sean poco aleatorias [16]. Igualmente sucede con la información mutua, ya sea porque las dos variables presenten una relación muy débil o porque la entropía de cada una sea muy pequeña.

Por ello debemos definir otra forma de medir el grado de similitud entre variables aleatorias, o lo que es lo mismo, ver cuánto se asemejan nuestras características frecuenciales. Este nuevo concepto se denomina *entropía normalizada*, NE, [7], la cual se define con la siguiente expresión.

$$NE(X, Y) = \frac{H(X) + H(Y)}{H(X, Y)} = 1 + \frac{I(X, Y)}{H(X, Y)} \quad (\text{ec. 3.5.4})$$

Gracias a este nuevo concepto, pasaremos a describir la idea de *información mutua normalizada*, que nos ayudará enormemente a la hora de la descripción de la semejanza entre dos imágenes. Lo definimos así.

$$NI(X, Y) = 2 \left( 1 - \frac{1}{NE(X, Y)} \right) \quad (\text{ec. 3.5.5})$$

NI es una medida de la similitud entre las variables X e Y, la cual toma valores en el rango [0,1], correspondiendo los valores cercanos a 0 variables que ofrecen poca información mutua, y viceversa.

Pero lo que nosotros estamos buscando, es una matriz de distancias, una matriz que refleje el grado de similitud entre pares de imágenes. Para ello, debemos realizar una transformación [4] a la información mutua normalizada, la cual hace que cuando exista una gran distancia entre dos imágenes, ésta sea lo más grande posible y cuando haya poca distancia, esto es, sean muy semejantes, la reduzca.

De este modo, por fin definimos la expresión de distancia que vamos a emplear para describir el grado de semejanza entre dos imágenes.

$$D_{NI}(X, Y) = \left( 1 - \sqrt{NI(X, Y)} \right)^2 \quad (\text{ec. 3.5.6})$$

Esta distancia tomará valores comprendidos desde en el rango [0,1], correspondiendo al 1 a imágenes totalmente distintas (como en el caso de una imagen con todos sus píxeles de color negro, todos con valor 0 y otra con todos sus píxeles de color blanco, con valor 255) y siendo los valores cercanos a 0 para imágenes muy semejantes.

Gracias a esta distancia podremos saber el grado de similitud que existe entre las distintas características frecuenciales obtenidas en las salidas del banco de filtros y, de este modo, agruparlas hasta obtener los patrones visuales componentes de nuestra imagen inicial.

## 6. Agrupamiento de características

En este momento ya tenemos todas las características frecuenciales de nuestra imagen original y debemos agruparlas, basándonos en la matriz de distancias, para obtener unos patrones visuales concretos.

Para ello, debemos usar las ideas basadas en *algoritmos de clustering* o *agrupamiento* que haciendo uso de la matriz de distancias previamente obtenida, realizará el mejor agrupamiento de dichas características.

Un método bastante aplicado es aquel que intenta minimizar el error cuadrático medio, como por ejemplo, el algoritmo K-means que en cada iteración mueve las

observaciones de un cluster a otro con el objetivo de minimizar la distancia media de las observaciones al centroide del clúster. El problema que tienen estos métodos es que en cada iteración debemos recalcular los nuevos centroides como el vector de medias de todas las observaciones dentro de un clúster dado y por ello también deberíamos recalcular las nuevas distancias de las distintas observaciones al centroide del clúster. Estos cálculos iterativos suponen una gran carga computacional, además de no hacer uso de la matriz de distancias que, previamente hemos hallado, lo cual hace que no sean de gran utilidad para el realizar nuestro agrupamiento de características.

Por ello, podemos plantearnos un algoritmo que haga uso de la matriz de distancias y sólo basándose en ella, realice el mejor agrupamiento de las características frecuenciales para obtener los patrones visuales que componen nuestra imagen.

El método que vamos a usar para realizar el agrupamiento se denomina *agrupamiento jerárquico* y está basado únicamente en la matriz de distancias. Lo que crearemos con este método es un conjunto de estructuras anidadas de todas las características frecuenciales, donde el criterio de unión estará basado en las distancias entre características frecuenciales. De este modo, obtendremos una estructura piramidal cuya base será el agrupamiento de las distintas características frecuenciales iniciales, teniendo en cuenta las distancias que existan entre ellas y en cada nivel se irán creando nuevos clústeres anidando los del nivel anterior. Para realizar dicha anidación, nos basaremos en la idea de que un clúster  $A$  será anidado dentro de un clúster  $B$  si cada componente perteneciente al clúster  $A$  es un subconjunto de una componente en  $B$ .

Procediendo de esta manera, obtendremos una estructura cuya base estará formada por todos los datos completamente desagrupados, y en cada nivel, nuevos grupos formados por la animación de los grupos pertenecientes al nivel inmediatamente inferior, hasta obtener un único grupo formado por todo el conjunto de datos de partida.

Una característica muy importante de los algoritmos de agrupamiento jerárquico, es que únicamente necesitan la matriz de distancias entre datos de partida y una estrategia de unión y no necesitan trabajar con los datos en sí, evitando así, el continuo recálculo de distancias y medias de centroides ocurrentes en cada iteración.

El algoritmo que vamos a emplear se denomina Algoritmo de la Matriz Actualizable [8, 9, 10]. Es un algoritmo que partirá con el conjunto de datos sin ningún tipo de unión y en cada nueva iteración decidirá que clústeres de la partición anterior deben ser unidos para crear la nueva, los cuales serán aquellos con menor distancia interclúster. El método para hallar estas nuevas distancias será determinante para decidir la estrategia de unión.

Podemos describir el funcionamiento de este algoritmo con los siguientes pasos.

1. Partiremos del conjunto de datos iniciales, los cuales queremos agrupar. En este momento la matriz de distancias entre clústeres será nuestra matriz de distancias entre características frecuenciales.
2. Buscaremos los dos clústeres  $r$  y  $s$  que menos se distancian, esto es, que más se parecen entre ellos. Esto es equivalente a encontrar el mínimo en la matriz de distancias.

3. Uniremos los dos clústeres r y s en un único clúster, al cual llamaremos j, para crear la siguiente partición.
4. Actualizaremos la matriz de distancias eliminando las filas y las columnas pertenecientes a los clústeres r y s, y añadiremos las filas y columnas que nos proporcionarán las nuevas distancias entre el nuevo clúster j formado y los clústeres antiguos. La forma de esta nueva distancia  $D[k,j]$  es la que determinará la estrategia de unión, siendo k cada uno de los clústeres anteriores a la unión.
5. Si todos los datos de partida están unidos en un único cluster, habremos terminado, si no, tendremos que seguir iterando yendo al paso 2.

Como se ha comentado en el paso 4, la forma que tenga las nuevas distancias  $D[k,j]$  es la que determinará la forma que tendremos de unir los distintos clústeres. Las técnicas más usadas son las que se denominan single-link (o vecino más cercano) y complete-link (o vecino más lejano). Single-link elegirá como nueva distancia la distancia más pequeña existente entre los elementos de los que forman los dos clústeres, lo cual es equivalente a definir  $D[k,j]$  como la mínima distancia que hay entre los elementos de los clústeres k y j. Complete-link elegirá como nueva distancia la distancia más grande existente entre los elementos de los que forman los dos clústeres, lo cual es equivalente a definir  $D[k,j]$  como la máxima distancia que hay entre los elementos de los clústeres k y j.

Complete-link asegura que los elementos más distanciados de un clúster estén lo más cerca posible, lo cual hace que obtengamos clústeres bastante compactos. Esto no ocurre en single-link, con el cual se obtienen clústeres con mayor varianza intraclúster.

Existen otras técnicas que intentan conseguir que esta varianza dentro de los elementos de un clúster sea la menor posible, entre las cuales destaca [10,11] el método de unión de Ward, también conocido como método de la mínima varianza.

Este método se consigue haciendo que la nueva distancia entre el clúster j, formado por r y s, y los anteriores clústeres k tome la siguiente forma:

$$D^{ward}[k, j] = \frac{(n_k + n_r)D(k, r) + (n_k + n_s)D(k, s) - n_k D(r, s)}{n_r + n_s + n_k} \quad (\text{ec. 3.6.1})$$

estando j formado por los clústeres r y s, siendo k cada uno de los clústeres anteriores y siendo  $n_i$  el número de elementos que forman el clúster i.

Podemos obtener una idea visual del funcionamiento del algoritmo de Ward, mediante la utilización de lo que se denomina *dendrograma*, el cual es un diagrama en forma de árbol que organiza los datos en subcategorías que se van agrupando en otros hasta llegar al nivel de detalle deseado. Este tipo de representación permite apreciar claramente las relaciones de agrupación entre los datos e incluso entre grupos de ellos. Observando las diferentes alturas de las ramas que se unen para formar un clúster determinado hacemos una idea de la similitud entre los elementos que lo forman, correspondiendo alturas pequeñas a un buen nivel de compactación. Cortando el árbol

en un determinado nivel, obtendremos un cierto número de patrones visuales, estando éstos formados por las características frecuenciales situadas en la base del diagrama.

En este momento, cabe preguntarse cuál es el número de patrones visuales óptimo que forman nuestra imagen de partida, o lo que es lo mismo, dónde situar la línea de corte en dendrograma. Para obtener dicho resultado, debemos evaluar la calidad de agrupamiento de los clústeres obtenidos y el grado de separación entre los mismos, es decir, debemos obtener clústeres con un alto grado de compactación dentro del clúster o mínima varianza entre los elementos que lo forman, y un alto grado de separación de los clústeres que hemos obtenido o máxima separación entre clústeres.

Existen varios índices que nos proporcionan cuál es el número óptimo de patrones que componen la imagen, entre los cuales podemos encontrar el índice de Davies-Bouldin [12, 13, 14] y que denotaremos por  $V_{DB}$ . Este índice será evaluado desde 1 hasta  $N_{max}$ , siendo  $N_{max}$ , el número de clústeres máximo que cabe pensar pueden existir en una única imagen. Debemos ajustar correctamente este  $N_{max}$ , pues si es muy pequeño, podemos encontrarnos con imágenes que contengan más patrones que este  $N_{max}$  y por ello no clasificarlos todos, pero tampoco debe ser muy grande, pues de otra forma, aumentaría considerablemente la carga computacional de forma innecesaria.

Una vez dicho esto, continuaremos con la definición de dicho índice. Dado un cierto nivel compuesto de  $N_c$  clústeres, este índice tiene la siguiente forma.

$$V_{DB}(N_c) = \frac{1}{N_c} \sum_{r=1}^{N_c} \max \left\{ \frac{\sigma_r + \sigma_s}{S_{rs}} \right\} \quad (\text{ec. 3.6.2})$$

donde  $\sigma_r$ , es una medida de dispersión entre los elementos que forman el clúster  $r$ , llamada *dispersión intraclúster*, igual sucede con el clúster  $s$ , y  $S_{rs}$  es una medida de la separación entre los clústeres  $r$  y  $s$ , o *dispersión interclúster*.

Para hacer una medida de esta dispersión intraclúster, podríamos realizar los algoritmos que se emplean en métodos como K-means, sin embargo nos encontraríamos con el problema de la carga computacional, pues en cada iteración, deberíamos hallar nuevas posiciones de centroides y nuevas distancias entre los mismos. Por ello, vamos a hacer uso de nuestra matriz de distancias que marca las diferencias entre todas las características frecuenciales obtenidas de la imagen que estamos procesando. Hemos obtenido esta medida usando una *teoría de grafos* [15] la cual usa el árbol o *Minimum Spanning Tree* (MST) obtenido en el algoritmo de agrupamiento jerárquico para cada uno de los clústeres obtenidos. Si el clúster  $r$  está formado por las características frecuenciales  $1, 2, \dots, N_r$ , entonces la máxima distancia que existe entre cada una de estas características se tomará como dispersión intraclúster del clúster  $r$ . Si dicho clúster se encuentra formado por sólo dos características, esta dispersión será la distancia entre ellas. Finalmente, si sólo posee una característica, dicha dispersión será 0.

Respecto a la medida de la separación entre clústeres, o dispersión interclúster, nuevamente haremos uso de la matriz de distancias, de forma que obtenemos que el parámetro  $S_{rs}$  toma la siguiente forma.

$$S_{rs} = \frac{1}{N_r N_s} \sum_{i \in r, j \in s} D_{i,j} \quad (\text{ec. 3.6.3})$$

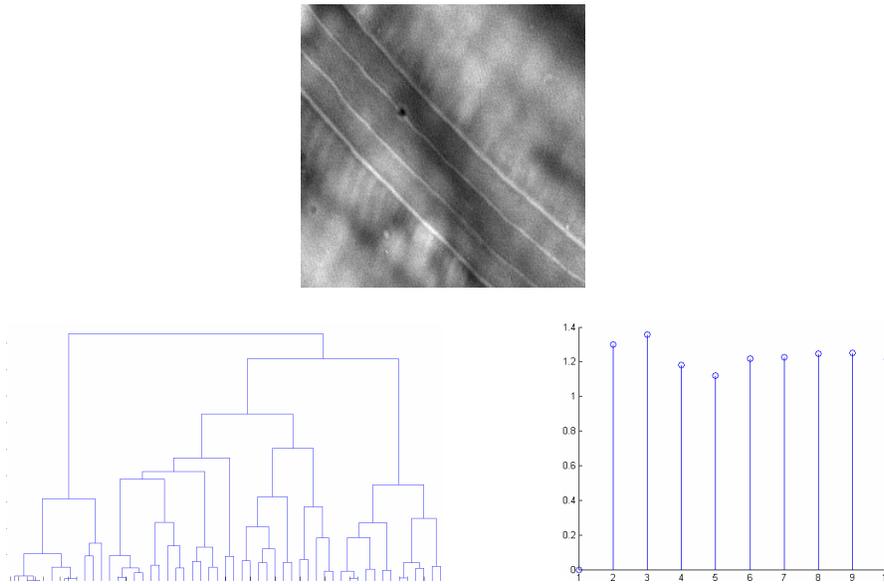
donde  $N_r$  y  $N_s$  representan, respectivamente, el número de elementos que forman los clústeres  $r$  y  $s$ , por lo que dividiendo la suma total de distancias entre los distintos elementos entre el número de elementos que hay, obtenemos una distancia media entre el clúster  $r$  y  $s$ .

De esta forma, el índice que nos guiará en la decisión de patrones visuales de la imagen, tomará el siguiente aspecto.

$$V_{DB}^{MST}(N_c) = \frac{1}{N_c} \sum_{r=1}^{N_c} \max_{s, s \neq r} \left\{ \frac{\sigma_r^{MST} + \sigma_s^{MST}}{S_{rs}} \right\} \quad (\text{ec. 3.6.4})$$

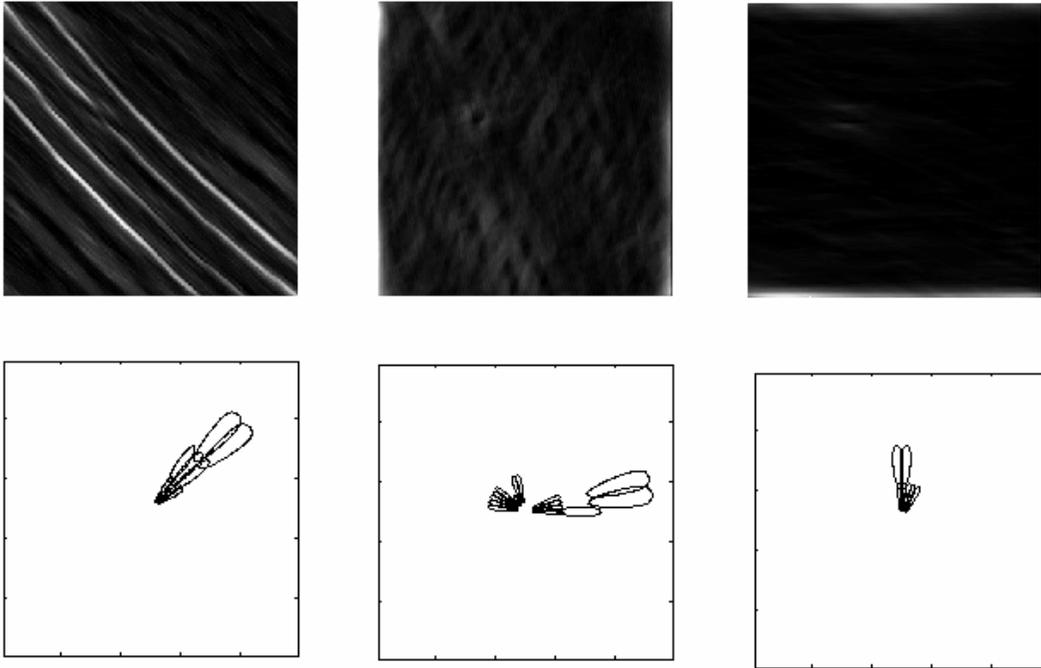
Aquel  $N_c$  que nos dé un máximo en este índice será el agrupamiento óptimo y por tanto el recorte que habrá que realizar en el árbol para obtener el número de patrones visuales.

En el siguiente ejemplo, planteamos cómo funciona este algoritmo.



**Ilustración 3.22** Superior: Imagen de entrada. Inferior Izquierda: Dendrograma obtenido tras agrupamiento de características. Inferior Derecha: Índice de Davies-Bouldin.

De este modo, agrupando cada una de las características como nos dice el árbol obtenido con el agrupamiento jerárquico y teniendo en cuenta el número obtenido por el índice de Davies-Bouldin modificado, obtenemos los siguientes patrones visuales.



**Ilustración 3.23** Patrones obtenidos.

Así, nuestro sistema ha segmentado los patrones que se observan en la imagen. El sistema extraerá las características frecuenciales que existen en la imagen, al introducir ésta en nuestro banco de filtros, previamente activados mediante el procesado de la imagen espectral y encontrará los patrones visuales que se encuentran en la imagen tras el agrupamiento jerárquico de cada una de las características.