



19



OFICINA ESPAÑOLA DE  
PATENTES Y MARCAS

ESPAÑA

11 Número de publicación: **2 366 439**

51 Int. Cl.:  
**G06F 17/30** (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

96 Número de solicitud europea: **04805546 .1**

96 Fecha de presentación : **25.11.2004**

97 Número de publicación de la solicitud: **1697862**

97 Fecha de publicación de la solicitud: **06.09.2006**

54 Título: **Procedimiento de indexación y de identificación de documentos multimedia.**

30 Prioridad: **27.11.2003 FR 03 13907**

45 Fecha de publicación de la mención BOPI:  
**20.10.2011**

45 Fecha de la publicación del folleto de la patente:  
**20.10.2011**

73 Titular/es: **HOLOGRAM INDUSTRIES**  
**22 avenue de l'Europe**  
**Parc d'Activité Gustave Eiffel**  
**Bussy Saint Georges**  
**77607 Marne la Vallee Cédex, FR**

72 Inventor/es: **Essafi, Hassane;**  
**Guezouli, Larbi;**  
**Sayah, Salima;**  
**Behloul, Ali;**  
**Mandridake, Clarisse y**  
**Essafi, Louafi**

74 Agente: **Curell Aguilá, Marcelino**

ES 2 366 439 T3

Aviso: En el plazo de nueve meses a contar desde la fecha de publicación en el Boletín europeo de patentes, de la mención de concesión de la patente europea, cualquier persona podrá oponerse ante la Oficina Europea de Patentes a la patente concedida. La oposición deberá formularse por escrito y estar motivada; sólo se considerará como formulada una vez que se haya realizado el pago de la tasa de oposición (art. 99.1 del Convenio sobre concesión de Patentes Europeas).

## DESCRIPCIÓN

Procedimiento de indexación y de identificación de documentos multimedia.

5 La presente invención se refiere a los procedimientos de indexación y de identificación de documentos multimedia.

Desde un punto de vista general, la identificación de un documento multimedia comprende dos etapas:

- 10 • Una etapa denominada de indexación, en la que se intenta caracterizar cada documento de una base previamente registrada por un número finito de parámetros que pueden fácilmente almacenarse y manipularse más tarde.
- 15 • Una etapa denominada de búsqueda, en la que tras una consulta formulada por el usuario, por ejemplo la identificación de una imagen pregunta, se buscan todos los documentos multimedia similares o que responden a esta consulta.

20 Ya existen varios procedimientos de indexación de imágenes que ponen en práctica la extracción de los atributos de la forma de los objetos que componen la imagen si existen, así como los de la textura o del color de fondo de la imagen.

Sin embargo, los procedimientos conocidos se aplican a campos muy especializados o implican el tratamiento de una cantidad muy grande de información que conduce a una complejidad y una lentitud en el tratamiento de estas informaciones.

25 Se conoce en particular por el artículo de Y. Chang *et al.* titulado "Conceptual retrieval based on feature clustering of documents", publicado en Proceedings of ACM Sigir Workshop on Mathematical-Formal Methods in Information Retrieval, en agosto de 2002, páginas 1-10, un procedimiento de búsqueda de documentos con por lo menos una etapa de identificación y de extracción para cada documento de términos constituidos por vectores que caracterizan las propiedades textuales del documento, y una etapa de descomposición en partes representadas por conceptos.

30 La presente invención tiende a remediar los inconvenientes mencionados anteriormente y a proporcionar un procedimiento de indexación y de identificación de documentos multimedia de una aplicación general que racionaliza el procedimiento de tratamiento y conduce a tiempos de tratamiento más cortos a la vez que aumenta la calidad de los resultados y su fiabilidad, lo que permite en particular proceder a búsquedas eficaces por el contenido.

35 Estos objetivos se alcanzan según la invención, gracias a un procedimiento de indexación de documentos multimedia, que comprende por lo menos las siguientes etapas:

40 (a) identificar y extraer para cada documento términos  $t_i$  constituidos por vectores que caracterizan propiedades del documento multimedia que va a indexarse, tales como la forma, la textura, el color o la estructura de una imagen, la energía, la tasa de oscilación o informaciones frecuenciales de una señal de audio, o un grupo de caracteres de un texto,

45 (b) almacenar los términos  $t_i$  que caracterizan propiedades del documento multimedia en una base de términos que comprende P términos, caracterizado porque comprende además las siguientes etapas:

(c) determinar un número máximo N de conceptos deseados que reagrupan los términos  $t_i$  pertinentes, siendo N un número entero inferior a P, y estando previsto cada concepto  $C_i$  para reagrupar todos los términos próximos desde el punto de vista de sus características,

50 (d) calcular la matriz T de distancias entre los términos  $t_i$  de la base de términos,

55 (e) descomponer el conjunto P de términos  $t_i$  de la base de términos en N partes  $P_j$  ( $1 \leq j \leq N$ ) tales como  $P = P_1 \cup P_2 \dots \cup P_j \dots \cup P_N$ , comprendiendo cada parte  $P_j$  un conjunto de términos  $t_{ij}$  y estando representada por un concepto  $C_j$ , estando los términos  $t_i$  repartidos en la base de la matriz T, de tal manera que los términos más alejados unos de otros se encuentran en partes  $P_l$ ,  $P_m$  distintas y los términos próximos unos de otros se encuentran en la misma parte  $P_l$ ,

60 (f) estructurar el diccionario de conceptos de manera que constituye un árbol binario en el que cada hoja del árbol binario contiene un concepto  $c_i$  del diccionario y cada nudo del árbol contiene las informaciones necesarias para el escrutinio del árbol durante una etapa de identificación de un documento multimedia mediante comparación con los documentos indexados anteriormente, y

65 (g) construir una base de huellas constituida por el conjunto de conceptos  $c_i$  que representan los términos  $t_i$  de los documentos que van a indexarse, estando cada documento asociado con una huella que le es propia.

De manera más particular, se asocia a cada concepto  $c_i$  de la base de huellas un conjunto de informaciones que comprende el número NbT de términos en los documentos en los que el concepto  $c_i$  está presente.

5 Según un aspecto particular de la invención, para cada documento en el que está presente un concepto  $c_i$ , se registra una huella del concepto  $c_i$  en el documento, comprendiendo esta huella la frecuencia de aparición del concepto  $c_i$ , la identificación de conceptos que están próximos al concepto  $c_i$  en el documento y una puntuación que es un valor medio de medidas de similitudes entre el concepto  $c_i$  y los términos  $t_i$  del documento que son los más próximos al concepto  $c_i$ .

10 Ventajosamente, el procedimiento según la invención comprende una etapa de optimización de la división del conjunto P de términos de la base de términos para descomponer este conjunto P en M clases  $C_i$  ( $1 \leq i \leq M$ , con  $M \leq P$ ), de manera que reduce el error del reparto del conjunto P de términos de la base de términos en N partes ( $P_1, P_2, \dots, P_N$ ) en el que cada parte  $P_i$  se representa por el término  $t_i$  que se tomará como concepto  $C_i$ , siendo el error

cometido  $\varepsilon$  tal que 
$$\varepsilon = \sum_{i=1}^N \varepsilon_{t_i}$$
 en la que 
$$\varepsilon_{t_i} = \sum_{t_j \in P_i} d^2(t_i, t_j)$$
 es el error cometido cuando se sustituyen los términos  $t_j$  de una parte  $P_i$  por  $t_i$ .

En este caso, el procedimiento puede comprender las siguientes etapas:

- 20 (i) descomponer el conjunto P de términos en dos partes  $P_1$  y  $P_2$ ;
- (ii) determinar los dos términos más alejados  $t_i$  y  $t_j$  del conjunto P que corresponde a la distancia más grande  $D_{ij}$  de la matriz T de distancias;
- 25 (iii) para cada término  $t_k$  del conjunto P, examinar si la distancia  $D_{ki}$  entre el término  $t_k$  y el término  $t_i$  es más pequeña que la distancia  $D_{kj}$  entre el término  $t_k$  y el término  $t_j$ , si es así asignar el término  $t_k$  a la parte  $P_1$  y si no es así asignar el término  $t_k$  a la parte  $P_2$ ;
- (iv) repetir la etapa (i) hasta obtener el número N de partes  $P_i$  deseado y en cada iteración aplicar las etapas (ii) y (iii) en los términos de las partes  $P_1$  y  $P_2$ .

30 El procedimiento según la invención puede caracterizarse más particularmente porque comprende una optimización a partir de las N partes desunidas  $\{P_1, P_2, \dots, P_N\}$  del conjunto P así como N términos  $\{t_1, t_2, \dots, t_N\}$  que las representan para reducir el error de descomposición del conjunto P en N partes, y porque comprende las siguientes etapas:

- 35 (i) calcular los centros de gravedad  $C_i$  de las partes  $P_i$
- (ii) calcular los errores 
$$\varepsilon_{C_i} = \sum_{t_j \in P_i} d^2(C_i, t_j)$$
 y 
$$\varepsilon_{t_i} = \sum_{t_j \in P_i} d^2(t_i, t_j)$$
 cuando se sustituyen los términos  $t_j$  de la parte  $P_i$  respectivamente por  $C_i$  y por  $t_i$ ,

- 40 (iii) comparar  $\varepsilon_{t_i}$  y  $\varepsilon_{C_i}$  y sustituir  $t_i$  por  $C_i$  si  $\varepsilon_{C_i} \leq \varepsilon_{t_i}$ ,
- (iv) calcular la nueva matriz T de distancias entre los términos  $t_j$  de la base de términos y procedimiento de descomposición del conjunto P de términos de la base de términos en N partes, salvo si se cumple una condición de parada con 
$$\frac{\varepsilon_{C_i} - \varepsilon_{C_{i+1}}}{\varepsilon_{C_i}} < \text{umbral}$$
, en la que  $\varepsilon_{C_i}$  representa el error cometido en el instante t.

45 Con el fin de facilitar la búsqueda y la identificación de documentos, para efectuar una estructuración del diccionario de conceptos, se produce de manera iterativa en cada iteración una carta de navegación empezando por dividir el conjunto de conceptos en dos subconjuntos, y luego seleccionando un subconjunto en cada iteración hasta la obtención del número de grupo deseado o hasta que se satisface un criterio de parada.

50 El criterio de parada puede caracterizarse por el hecho de que los subconjuntos obtenidos son todos homogéneos con una pequeña desviación estándar.

De manera más particular, durante la estructuración del diccionario de conceptos, se determinan indicadores de navegación a partir de una matriz  $M = [c_1, c_2, \dots, c_N] \in \mathfrak{R}^{p \times N}$  del conjunto C de conceptos  $c_i \in \mathfrak{R}^p$  en la que  $c_i$  representa un concepto de p valores, según las siguientes etapas:

- (i) calcular un representante  $w$  de la matriz M,
- 60 (ii) calcular la matriz de covarianza  $\tilde{M}$  entre los elementos de la matriz M y el representante  $w$  de la matriz M,

(iii) calcular un eje de proyección  $u$  de los elementos de la matriz  $M$ ,

5 (iv) calcular el valor  $p_i = d(u, C_i) - d(u, w)$  y descomponer el conjunto de conceptos  $C$  en dos subconjuntos  $C1$  y  $C2$  de la siguiente manera:

$$\begin{cases} c_i \in C1 & \text{si } p_i \leq 0 \\ c_i \in C2 & \text{si } p_i > 0 \end{cases}$$

10 (v) almacenar en el nudo asociado con  $C$  las informaciones  $\{u, w, |p1|, p2\}$  en la que  $p1$  es el máximo de todos los  $p_i \leq 0$  y  $p2$  es el mínimo de todos los  $p_i > 0$ , constituyendo el conjunto de las informaciones  $\{u, w, |p1|, p2\}$  los indicadores de navegación en el diccionario de conceptos.

Según un modo particular de realización, se analizan a la vez las componentes estructurales y el complemento de estas componentes estructurales constituido por las componentes texturales de una imagen del documento, y:

15 (a) durante el análisis de las componentes estructurales de la imagen

(a1) se procede a un reparto de las zonas fronteras de las estructuras de la imagen en diferentes clases según la orientación de la variación local de intensidad de manera que define los elementos de soporte estructural (ESS) de la imagen, y

20 (a2) se procede por análisis estadístico a la construcción de términos constituidos por vectores que describen las propiedades locales y globales de los elementos de soporte estructurales,

25 (b) durante el análisis de las componentes texturales de la imagen

(b1) se procede a una detección y una caracterización paramétrica de una componente puramente aleatoria de la imagen,

30 (b2) se procede a una detección y una caracterización paramétrica de una componente periódica de la imagen,

(b3) se procede a una detección y una caracterización paramétrica de una componente direccional de la imagen,

35 (c) se reagrupa en un número limitado de conceptos el conjunto de los elementos descriptivos de la imagen constituidos por una parte, por los términos que describen las propiedades locales y globales de los elementos de soporte estructurales y por otra parte los parámetros de las caracterizaciones paramétricas de las componentes aleatoria, periódica y direccional que definen las componentes texturales de la imagen, y

40 (d) se define para cada documento una huella a partir de las apariciones, de las posiciones y de las frecuencias de dichos conceptos.

45 Ventajosamente, las propiedades locales de los elementos de soporte estructurales tenidas en cuenta para la construcción de términos que comprenden por lo menos el tipo de soporte seleccionado de entre una banda lineal o un arco de curva, las dimensiones de longitud y anchura del soporte, la dirección principal del soporte y la forma y las propiedades estadísticas de los píxeles que constituyen el soporte.

Las propiedades globales de los elementos de soporte estructurales tenidas en cuenta para la construcción de términos comprenden por lo menos el número de cada tipo de soportes y su disposición espacial.

50 Preferiblemente, durante el análisis de las componentes estructurales de la imagen se procede a una prueba previa de detección de la presencia de por lo menos una estructura en la imagen y, en caso de ausencia de estructura, se pasa directamente a la etapa del análisis de las componentes texturales de la imagen.

55 Ventajosamente, para proceder a una reparto de las zonas fronteras de las estructuras de la imagen en diferentes clases, a partir de la imagen digitalizada definida por el conjunto de los píxeles  $y(i,j)$  en la que  $(i,j) \in I \times J$ , designando  $I$  y  $J$  respectivamente el número de líneas y el número de columnas de la imagen, se calcula la imagen gradiente vertical  $g_v(i,j)$  con  $(i,j) \in I \times J$  y la imagen gradiente horizontal  $g_h$  con  $(i,j) \in I \times J$  y se procede a la división de la imagen según la orientación local de su gradiente en un número finito de clases equidistantes, definiéndose la imagen que contiene la orientación del gradiente por la fórmula

60

$$O(i,j) = \arctan \left[ \frac{gh(i,j)}{gv(i,j)} \right], \quad (1)$$

se identifican las clases que constituyen regiones de soporte susceptibles de contener los elementos de soporte significativos, y a partir de las regiones de soporte, se determinan los elementos de soporte significativos y se catalogan según criterios predeterminados.

5 Según un aspecto particular de la invención, se analiza las formas de una imagen de un documento según las siguientes etapas:

- 10 (a) proceder a una multirresolución seguida de una decimación de la imagen,
- (b) definir la imagen en el espacio logarítmico polar,
- (c) representar la imagen o la porción de la imagen afectada por su transformada de Fourier H,
- 15 (d) proceder a una caracterización de la transformada de Fourier H de la siguiente manera:
  - (d1) se proyecta H en varias direcciones para obtener un conjunto de vectores cuya dimensión es igual a la dimensión del movimiento de proyección,
  - 20 (d2) se calculan las propiedades estadísticas de cada vector de proyección, y
- (e) representar la forma de la imagen por un término  $t_i$  constituido por los valores de las propiedades estadísticas de cada vector de proyección.

25 Según un aspecto particular de la invención, durante la indexación de un documento multimedia que comprende señales de vídeo, se eligen términos  $t_i$  constituidos por imágenes clave que representan grupos de imágenes homogéneas consecutivas, y se determinan conceptos  $c_i$  mediante reagrupamiento de términos  $t_i$ .

30 Para determinar imágenes clave que constituyen términos  $t_i$ , se elabora en primer lugar un vector de puntuación VS que comprende un conjunto de elementos VS(i) que materializan la diferencia o la similitud entre el contenido de una imagen de índice i y el de una imagen de índice i-1, y se analiza el vector de puntuación VS con el fin de determinar las imágenes clave que corresponden a los máximos de los valores de los elementos VS(i) del vector de puntuación VS.

35 De manera más particular, una imagen de índice j se considera como una imagen clave si el valor VS(j) del elemento correspondiente del vector de puntuación VS es un máximo y el valor VS(j) se sitúa entre dos mínimos min G y min D y el mínimo M1 tal que  $M1 = (|VS_{(j)} - \min G|, |VS_{(j)} - \min D|)$  es superior a un umbral dado.

40 Se considerará de nuevo la indexación de un documento multimedia, que comprende componentes de audio, se toman muestras y se descompone el documento en tramas, que se reagrupan a continuación en clips de los que cada uno se caracteriza por un término  $t_i$  constituido por un vector de parámetro.

45 Una trama puede comprender por ejemplo entre aproximadamente 512 y 2048 muestras del documento de audio del que se han tomado muestras.

Ventajosamente, los parámetros tenidos en cuenta para la definición de los términos  $t_i$  comprenden informaciones temporales que corresponden a por lo menos uno de los siguientes parámetros: la energía de las tramas de la señal de audio, la desviación estándar de las energías de las tramas en los clips, la relación de las variaciones sonoras, la relación de baja energía, la tasa de oscilación alrededor de un valor predeterminado, la alta tasa de oscilación alrededor de un valor predeterminado, la diferencia entre el número de tasa de oscilación por encima y por debajo de la tasa de oscilación media de las tramas de clips, la varianza de la tasa de oscilación, la relación de las tramas silenciosas.

55 Sin embargo, de manera alternativa o complementaria, de manera ventajosa, los parámetros tenidos en cuenta para la definición de los términos  $t_i$  comprenden informaciones frecuenciales que corresponden a por lo menos uno de los siguientes parámetros: el centro de gravedad del espectro de frecuencia de la transformada de Fourier corta de la señal de audio, la anchura de banda de la señal de audio, la relación entre la energía en una banda de frecuencia y la energía total en toda la banda de frecuencia de la señal de audio de la que se han tomado muestras, el valor medio de la variación del espectro de dos tramas adyacentes en un clip, la frecuencia de corte de un clip.

60 De manera más particular, los parámetros tenidos en cuenta para la definición de los términos  $t_i$  pueden comprender por lo menos la modulación de energía a 4 Hz.

Otras características y ventajas de la invención se destacarán de la siguiente descripción de modos particulares de realización, facilitados a modo de ejemplo, haciendo referencia a los dibujos adjuntos, en los que:

- 5 - la figura 1 es un diagrama de bloques que muestra el procedimiento de producción de un diccionario de conceptos a partir de una base de documentos, según la invención,
- la figura 2 muestra el principio de construcción de una base de conceptos a partir de términos,
- 10 - la figura 3 es diagrama de bloques que muestra el procedimiento de estructuración de un diccionario de conceptos, según la invención,
- la figura 4 muestra la estructuración de una base de huellas puesta en práctica en el marco del procedimiento según la invención,
- 15 - la figura 5 es un organigrama que muestra las diferentes etapas de construcción de una base de huellas,
- la figura 6 es un organigrama que muestra las diferentes etapas de identificación de documentos,
- 20 - la figura 7 es un organigrama que muestra la selección de una primera lista de respuestas,
- la figura 8 es un organigrama que muestra las diferentes etapas de una etapa de indexación de documentos según el procedimiento según la invención,
- 25 - la figura 9 es un organigrama que muestra las diferentes etapas de extracción de términos en el caso del tratamiento de imágenes,
- la figura 10 es un esquema que resume el procedimiento de descomposición de una imagen regular y homogénea,
- 30 - las figuras 11 a 13 muestran tres ejemplos de imágenes que contienen diferentes tipos de elementos,
- las figuras 14a a 14f muestran respectivamente un ejemplo de imagen original, un ejemplo de imagen después de tratamiento teniendo en cuenta el módulo de gradiente, y cuatro ejemplos de imágenes tratadas con desmantelamiento de las zonas fronteras de la imagen,
- 35 - la figura 15a representa un primer ejemplo de imagen que contiene un elemento direccional,
- la figura 15a1 es una vista 3D del espectro de imagen de la figura 15a,
- 40 - la figura 15b representa un segundo ejemplo de imagen que contiene un elemento direccional,
- la figura 15b1 es una imagen módulo de Fourier de la imagen de la figura 15b,
- 45 - la figura 15c representa un tercer ejemplo de imagen que contiene dos elementos direccionales,
- la figura 15c1 es una imagen módulo de Fourier de la imagen de la figura 15c,
- 50 - la figura 16 ilustra direcciones de proyección para pares de enteros  $(\alpha, \beta)$  en el marco del cálculo de la transformada de Fourier discreta de una imagen,
- la figura 17 ilustra un ejemplo de mecanismo de proyección con el ejemplo de un par de entradas  $(\alpha_k, \beta_k) = (2, -1)$ ,
- 55 - la figura 18a1 representa un ejemplo de imagen que contiene componentes periódicas,
- la figura 18a2 representa la imagen módulo de la transformada de Fourier discreta de la imagen de la figura 18a1,
- 60 - la figura 18b1 representa un ejemplo de imagen sintética que contiene una componente periódica,
- la figura 18b2 representa una vista en 3D de la transformada de Fourier discreta de la imagen de la figura 18b1, que hace aparecer un par de picos simétricos,
- 65 - la figura 19 es un organigrama que muestra las diferentes etapas de tratamiento de una imagen con establecimiento de un vector que caracteriza la distribución espacial de las propiedades icónicas de la imagen,

- la figura 20 muestra un ejemplo de división de una imagen y de creación de un vector característico de esta imagen,
  - 5 - la figura 21 muestra una rotación de 90° de la imagen dividida de la figura 20 y la creación de un vector característico de esta imagen,
  - la figura 22 muestra la descomposición de una señal sonora en tramas en clips,
  - 10 - la figura 23a muestra la variación de energía de una señal de voz,
  - la figura 23b muestra la variación de energía de una señal de música,
  - 15 - la figura 24a muestra la tasa de paso por cero de una señal de voz,
  - la figura 24b muestra la tasa de paso por cero de una señal de música,
  - la figura 25a muestra el centro de gravedad del espectro de frecuencia de la transformada de Fourier corta de una señal de voz,
  - 20 - la figura 25b muestra el centro de gravedad del espectro de frecuencia de la transformada de Fourier corta de una señal de música,
  - la figura 26a muestra la anchura de banda de una señal de voz,
  - 25 - la figura 26b muestra la anchura de banda de una señal de música,
  - la figura 27a muestra para tres sub-bandas de frecuencia 1, 2, 3 la relación de energía en cada sub-banda de frecuencia con respecto a la energía total de toda la banda de frecuencia, para una señal de voz,
  - 30 - la figura 27b muestra para tres sub-bandas de frecuencia 1, 2, 3 la relación de energía en cada sub-banda de frecuencia con respecto a la energía total de toda la banda de frecuencia, para una señal de música,
  - la figura 28a muestra el flujo espectral de una señal de voz,
  - 35 - la figura 28b muestra el flujo espectral de una señal de música,
  - la figura 29 es un gráfico que ilustra la definición de la frecuencia de corte de un clip, y
  - 40 - la figura 30 ilustra, para una señal de audio, la modulación de la energía alrededor de 4 Hz.
- Se describirá en primer lugar haciendo referencia a las figuras 1 a 5, el principio general del procedimiento de indexación de documentos multimedia según la invención, que conduce a la construcción de una base de huellas, estando asociado cada documento indexado con una huella que le es propia.
- 45 A partir de una base de documentos multimedia 1, una primera etapa 2 consiste en la identificación y la extracción, para cada documento, de términos  $t_i$  constituidos por vectores que caracterizan propiedades del documento que va a indexarse.
- A modo de ejemplo, se describirá, haciendo referencia a las figuras 22 a 30, la manera en que es posible identificar y extraer términos  $t_i$  para un documento sonoro.
- 50 Un documento de audio 140 se descompone en primer lugar en tramas 160 que se reagrupan a continuación en clips 150 de los cuales cada uno va a caracterizarse por un término constituido por un vector de parámetros (figura 22). Un documento de audio 140 se caracterizará por tanto por un conjunto de términos  $t_i$  que se almacenarán en una base de términos 3 (figura 1).
- 55 Se pueden tomar muestras de los documentos audio de los cuales se ha extraído su vector característico, por ejemplo, a 22.050 Hz con el fin de evitar el efecto de solapamiento. El documento se divide a continuación en un conjunto de tramas cuyo número de muestras por trama se fija en función del tipo de archivo que va a analizarse.
- 60 Para un documento de audio rico en frecuencias y que contiene muchas variaciones, como las películas por ejemplo, los programas de variedades o incluso los programas deportivos, el número de muestras en una trama tiene que ser pequeño, del orden de 512 muestras por ejemplo. Sin embargo, para un documento de audio homogéneo que sólo contiene voz o música por ejemplo, este número tiene que ser importante, por ejemplo del
- 65 orden de 2.048 muestras.

Un clip de documento de audio puede caracterizarse por diferentes parámetros que sirven para constituir los términos y que caracterizan informaciones temporales o frecuenciales.

5 Es posible utilizar la totalidad o parte de los parámetros que se mencionarán más adelante para formar vectores de parámetros que constituyen los términos que identifican los clips sucesivos del documento de audio del que se han tomado muestras.

10 La energía de las tramas de la señal de audio constituye un primer parámetro que representa una información temporal.

15 La energía de la señal de audio varía mucho para la voz mientras que es bastante estable para la música. Por lo tanto, permite discriminar la voz de la música pero también detectar los silencios. La energía puede acoplarse con otro parámetro temporal tal como la tasa de oscilación (TO) alrededor de un valor, que puede corresponder por ejemplo a la tasa de paso por cero (TPZ). En efecto, una TO pequeña y una energía fuerte son sinónimos de un sonido vocal mientras que una TO elevada induce una zona no vocal.

La figura 25a representa una señal 141 que ilustra la variación de energía en el caso de una señal de voz.

20 La figura 23b representa una señal 142 que ilustra la variación de energía en el caso de una señal de música.

Sea N el número de muestras en una trama, el volumen o energía E(n) se define por:

$$E(n) = \frac{1}{N} \sum_{i=0}^{N-1} S_n^2(i). \quad (2)$$

25 en la que S<sub>n</sub>(i) representa el valor de la muestra i de la trama de índice n de una señal de audio.

Otros parámetros representativos de informaciones temporales se pueden deducir de la energía, como por ejemplo:

- 30 - la desviación estándar de las energías de las tramas en los clips (también denominado EEC o VSTD) que constituye un estado definido como la varianza de los volúmenes de las tramas en un clip normalizado por el máximo del volumen de las tramas del clip,
- 35 - la relación de las variaciones sonoras (RVS) que está constituida por la diferencia entre el máximo y el mínimo de los volúmenes de las tramas de un clip dividido por el máximo de los volúmenes de estas tramas,
- la relación de baja energía (o LER) que es el porcentaje de las tramas cuyo volumen es inferior a un umbral (que se fija por ejemplo al 95% del volumen medio de un clip).

40 Otros parámetros permiten caracterizar el aspecto temporal de un clip, en particular la tasa de oscilación alrededor de un valor predeterminado, que, cuando este valor predeterminado es cero, define una tasa de paso por cero (o TPZ).

La TPZ también puede definirse por el número de veces que la onda pasa por cero.

$$Z(n) = \frac{1}{2} \left( \sum_{i=0}^{N-1} | \text{Sign}(S_n(i)) - \text{Sign}(S_n(i-1)) | \right) \frac{f_s}{N} \quad (3)$$

45 S<sub>n</sub>(i): Valor de la muestra i, de la trama n.

50 N: número de muestras en una trama.

f<sub>s</sub>: frecuencia de toma de muestras.

55 Esta característica se utiliza frecuentemente para la clasificación voz/música. En efecto, las variaciones bruscas de la TPZ son significativas de la alternancia vocal/no vocal y por lo tanto de la presencia de voz. Para la voz, la TPZ es pequeña para las zonas vocales, y muy elevada para las zonas no vocales mientras que para la música, las variaciones de la TPZ son muy pequeñas.

La figura 24a muestra una curva 143 que ilustra un ejemplo de TPZ para una señal de voz.

60 La figura 24b muestra una curva 144 que ilustra un ejemplo de TPZ para una señal de música.

Otro parámetro que caracteriza el aspecto temporal de un clip puede estar constituido por la alta tasa de oscilación alrededor de un valor predeterminado que, cuando este valor predeterminado es cero, define una alta tasa de paso por cero (o HTPZ).

5 La HTPZ puede definirse como la relación del número de tramas cuya TPZ es un valor  $\alpha$ , por ejemplo 1,5 por encima de la TPZ media del clip (1s):

$$HTPZ = \frac{1}{2N} \sum_{n=0}^{N-1} [\text{sgn}(TPZ(n) - 1.5avTPZ) + 1] \quad (4)$$

$$avTPZ = \frac{1}{N} \sum_{n=0}^{N-1} TPZ(n) \quad (5)$$

10 tal que:

siendo:

n: índice de la trama.

15 N: número de tramas en un clip.

Para los segmentos de voz los clips son de 0 a 200 s con una HTPZ de alrededor de 0,15.

20 Sin embargo, para los segmentos de música, los clips son de 200 a 350 s y la HTPZ varía alrededor de 0,05 y es en general casi nula.

Para el sonido ambiente los segmentos que corresponden a los clips son de 351 a 450 s.

25 La HTPZ es pequeña para el ruido blanco y grande para un sonido ensordecedor (tambor por ejemplo).

También puede definirse el parámetro DTPZ que está constituido por la diferencia entre el número de TPZ por encima y por debajo de la TPZ media de las tramas de un clip, así como el parámetro VTPZ que está constituido por la varianza de la TPZ.

30 Otro parámetro que caracteriza el aspecto temporal de un clip es la relación de las tramas silenciosas (RFS) que es el porcentaje de las tramas no silenciosas en un clip.

35 Una trama es no silenciosa si su volumen sobrepasa un determinado umbral (10) y si el valor de la TPZ es inferior a una  $T_{pz}$  umbral.

Así la relación de tramas no silenciosas en un clip permite detectar el silencio.

40 Otras propiedades estadísticas de la TPZ puede utilizarse como parámetros característicos, tales como:

- i) instante del tercer orden de la media,
- ii) el número de TPZ que sobrepasan un determinado umbral.

45 Los parámetros tenidos en cuenta para la definición de los términos  $t_i$  también pueden comprender informaciones frecuenciales que tienen en cuenta el cálculo de la transformada de Fourier rápida (FFT) de la señal de audio.

Así, un parámetro denominado centroide espectral (CS) puede definirse como el centro de gravedad del espectro de frecuencia de la transformada de Fourier corta (STFT) de la señal de audio:

$$CS(n) = \frac{\sum_{i=0}^{N-1} i S_n(i)}{\sum_{i=0}^{N-1} S_n(i)} \quad (6)$$

50 tal que  $S_n(i)$ : Potencia espectral de la trama  $i$  del clip

El parámetro CS es elevado para la música ya que las alturas se reparten en una zona más extendida que la de la

voz (en general 6 octavas para la música y 3 para la voz). Tiene una relación con la sensación de la claridad del sonido que se escucha. Es un atributo perceptivo importante para la caracterización del timbre.

La figura 25a muestra una curva 145 que ilustra un ejemplo de CS para una señal de voz.

La figura 25b muestra una curva 146 que ilustra un ejemplo de CS para una señal de música.

Otro parámetro está constituido por la anchura de banda LB que se puede calcular a partir de la varianza del parámetro anterior CS(n).

$$LB^2(n) = \frac{\sum_{i=0}^{N-1} (i - CS(n))^2 S_n(i)}{\sum_{i=0}^{N-1} S_n(i)} \quad (7)$$

La anchura de banda LB es importante tanto en música como en voz.

La figura 26a muestra una curva 147 que ilustra un ejemplo de anchura de banda de una señal de voz.

La figura 26b muestra una curva 148 que ilustra un ejemplo de anchura de banda de una señal de música.

Otro parámetro útil está constituido por la relación ERSB entre la energía en una sub-banda de frecuencia i y la energía total en toda la banda de frecuencia de la señal de audio de la que se han tomado muestras.

Al considerar las propiedades perceptivas del oído humano, la banda de frecuencia se ha dividido en cuatro sub-bandas en las que estas últimas corresponden a los filtros de Cochlear. Cuando la frecuencia de toma de muestras es de 22025 Hz, las bandas de frecuencias son: 0-630Hz, 630-1720Hz, 1720-4400Hz y 4400-11025Hz. Para cada una de estas bandas se calcula su energía ERSBi, que corresponde a la relación de la energía de esta última con respecto a la energía en toda la banda de frecuencia.

La figura 27a muestra tres curvas 151, 152, 153 que ilustran para tres sub-bandas de frecuencia 1, 2, 3 la relación de energía en cada sub-banda de frecuencia con respecto a la energía total de toda la banda de frecuencia, para un ejemplo de señal de voz.

La figura 27b muestra tres curvas 154, 155, 156 que ilustran para tres sub-bandas de frecuencia 1, 2, 3 la relación de energía en cada sub-banda de frecuencia con respecto a la energía total de toda la banda de frecuencia, para un ejemplo de señal de música.

Otro parámetro está constituido por el flujo espectral que se define como el valor media de la variación del espectro de dos tramas adyacentes en un clip:

$$FS(n) = \frac{1}{N} \sum_{i=1}^N [\log(S_n(i) + \delta) - \log(S_n(i-1) + \delta)]^2 \quad (8)$$

en la que

$\delta$  Una constante de un valor pequeño,

$S_n(i)$ : Potencia espectral de la trama i del clip n.

El flujo espectral de la voz es en general más importante que el de la música, y el del sonido ambiental es el más grande. Varía considerablemente en comparación con las otras dos señales.

La figura 28a muestra una curva 157 que ilustra el flujo espectral de un ejemplo de señal de voz.

La figura 28b muestra una curva 158 que ilustra el flujo espectral de un ejemplo de señal de música.

Otro parámetro útil está constituido por la frecuencia de corte de un clip (FCC).

La figura 29 muestra una curva 149 que ilustra el espectro de amplitud en función de la frecuencia fe, y la frecuencia de corte fc que es la frecuencia por debajo de la cual se concentra el 95% de la energía del espectro (la potencia espectral).

Para determinar la frecuencia de corte del clip, se calcula la transformada de Fourier del clip  $DS(n)$

$$DS(n) = \sum_{i=0}^{N-1} S_n^2(i) \quad (9)$$

5

La frecuencia  $f_c$  está determinada por:

$$\sum_{i=0}^{f_c} S_n^2(i) \geq 0.95 \times DS \quad (10)$$

y

10

$$\sum_{i=0}^{f_c-1} S_n^2(i) < 0.95 \times DS \quad (11)$$

La FCC es más elevada para un sonido no vocal (sonido rico en altas frecuencias) que para un sonido vocal (presencia de voz en la que la potencia se concentra en las bajas frecuencias).

15

Esta medida permite caracterizar las alternancias vocales/no vocales de la voz ya que este valor es pequeño para los clips que contienen únicamente música.

También pueden tenerse en cuenta otros parámetros para la definición de los términos  $t_i$  de un documento de audio, tales como la modulación de energía alrededor de 4 Hz, que constituye un parámetro procedente a la vez de un análisis frecuencial y de un análisis temporal.

20

La modulación de energía a 4 Hz (4 ME) se calcula a partir del contorno del volumen, según la siguiente fórmula:

$$4ME = \frac{\sum_{i=0}^{N/T} (\sum_{j=0}^T W(j) S_n^2(j + i \times T)) / T}{\sum_{i=0}^N S_n^2(i)} \quad (12)$$

25

en la que

$S_n(i)$ : Potencia espectral de la trama  $i$  del clip  $n$ .

30

$W(j)$ : Ventana triangular centrada en 4Hz.

$T$ : Anchura de un clip.

La voz presenta una 4ME más importante que la música ya que, para la voz, los cambios de sílaba se sitúan alrededor de 4 Hz.

35

Una sílaba es en efecto una combinación de una zona de pequeña energía (consonante) y de una zona de fuerte energía (vocal).

La figura 30 muestra una curva 161 que ilustra un ejemplo de señal de audio y una curva 162 que muestra para esta señal la modulación de la energía alrededor de 4 Hz.

40

Se ha descrito anteriormente el caso de documentos multimedia que comprenden componentes de audio.

En el caso de la indexación de documentos multimedia que comprenden señales de vídeo, se pueden elegir términos  $t_i$  constituidos por imágenes clave que representan grupos de imágenes homogéneas consecutivas.

45

Los términos  $t_i$  pueden a su vez representar por ejemplo colores dominantes, propiedades texturales, estructuras de zonas dominantes de las imágenes clave del documento de vídeo.

50

De una manera general, en el caso de las imágenes que se desarrollará con mayor detalle a continuación, los términos pueden representar colores dominantes, propiedades texturales, estructuras de las zonas dominantes de la imagen. Pueden ponerse en práctica varios procedimientos de manera alternativa o acumulativa, tanto en la totalidad de la imagen como en porciones de la imagen, para determinarlos los términos  $t_i$  que deben caracterizar la

imagen.

En el caso de un documento que contiene texto, los términos  $t_i$  pueden estar constituidos por palabras del lenguaje hablado o escrito, por números y por otros identificadores constituidos por combinaciones de caracteres (por ejemplo de las combinaciones de letras y de cifras).

Se considerará de nuevo la indexación de un documento multimedia que comprende señales de vídeo, para el que se eligen los términos  $t_i$  constituidos por imágenes clave que representan grupos de imágenes homogéneas consecutivas, y se determinan conceptos  $c_i$  mediante reagrupamiento de términos  $t_i$ .

La detección de las imágenes clave se basa en el reagrupamiento de las imágenes de un documento de vídeo en grupos que contienen cada uno únicamente imágenes homogéneas. De cada uno de los grupos se extrae una o varias imágenes (denominadas imágenes clave) que representan el documento de vídeo.

El reagrupamiento de las imágenes del documento de vídeo se basa en la producción de un vector de puntuación denominado VS que representa el contenido del vídeo, caracteriza la variación de las imágenes consecutivas del vídeo (el elemento  $VS_i$  materializa la diferencia entre el contenido de la imagen de índice  $i$  y el de la imagen de índice  $i-1$ ), VS es igual a cero cuando los contenidos  $im_i$  y  $im_{i-1}$  son idénticos y es importante cuando la diferencia entre los dos contenidos es importante.

Para calcular la señal VS, las tres bandas de cada imagen  $im_i$  RGB de índice  $i$  de vídeo se adicionan para constituir una sola imagen denominada  $TR_i$ . A continuación, la imagen  $TR_i$  se descompone en varias bandas de frecuencia para conservar solo la componente de baja frecuencia  $TRBi$ . Se utiliza para ello dos filtros de espejo (un filtro de paso bajo PB y un filtro de paso alto PH) que se aplican sucesivamente en las líneas y en las columnas de la imagen. Se considerarán dos tipos de filtro: ondículas de Haar y el filtro cuyo algoritmo es el siguiente:

Barrido lineal

A partir de  $TR_k$  se produce la imagen Bajo

Para cada punto  $a_{2xi,j}$  de la imagen TR hacer

Calcular el punto  $b_{i,j}$  de la imagen baja frecuencia bajo,  $b_{i,j}$  adopta el valor mediano de  $a_{2xi,j-1}$ ,  $a_{2xi,j}$  y  $a_{2xi,j+1}$ .

Barrido en columnas

A partir de las dos imágenes Bajo se produce la imagen  $TRB_k$

Para cada punto  $b_{i,2xi}$  de la imagen TR hacer

Calcular el punto  $bb_{i,j}$  de la imagen baja frecuencia bajo,  $bb_{i,j}$  adopta el valor mediano de  $b_{i,2xi-1}$ ,  $b_{i,2xi}$  y  $b_{i,2xi+1}$

Los barridos lineal y en columnas se aplican tantas veces como se desee. El número de iteraciones depende de la resolución de las imágenes del vídeo. Para imágenes de tamaño 512x512 puede fijarse  $n$  a tres.

La imagen resultado  $TRBi$  se proyecta en varias direcciones para obtener un conjunto de vectores  $V_k$ ,  $k$  es el ángulo de proyección (el elemento  $j$  de VO, vector obtenido tras la proyección horizontal de la imagen, es igual a la suma de todos los puntos de la línea  $j$  de la imagen). Los vectores de direcciones de la imagen  $TRBi$  se comparan con los vectores de dirección de  $TRBi-1$  para obtener una puntuación  $i$  que mide la similitud entre estas dos imágenes. Esta puntuación se obtiene mediante la media de todas las distancias de los vectores de misma dirección: para cada  $k$  se calcula la distancia entre el vector  $V_k$  de la imagen  $i$  y el vector  $V_k$  de la imagen  $i-1$  y luego se calculan todas estas distancias.

El conjunto de todos los resultados constituye el vector de puntuación VS: el elemento  $i$  de VS mide la similitud entre la imagen  $TRBi$  y la imagen  $TRBi-1$ . El vector VS se alisa con el fin de eliminar las irregularidades debidas al ruido generado durante la manipulación del vídeo.

A continuación se describirá un ejemplo de reagrupamiento de las imágenes y de extracción de las imágenes clave.

El vector VS se analiza con el fin de determinar las imágenes clave que corresponden a los máximos de los valores de VS. Una imagen de índice  $j$  se considera como una imagen clave si el valor  $VS(j)$  es un máximo y si  $VS(j)$  se sitúa entre dos mínimos  $\min G$  (mínimo izquierdo) y  $\min D$  (mínimo derecho) y si el mínimo  $M1$  tal como  $M1 = \min(|VS(C_j) - \min G|, |VS(j) - \min D|)$  es superior a un umbral dado.

Para detectar las imágenes clave, se inicializa  $\min G$  con  $VS(0)$  y luego se recorre el vector VS de izquierda a derecha. En cada etapa, se determina el índice  $j$  que corresponde al valor máximo situado entre dos mínimos ( $\min G$  y  $\min D$ ) y luego en función del resultado de la ecuación que define  $M1$  se decide considerar  $j$  como un índice de una imagen clave o no. Es posible tomar un grupo de varias imágenes clave próximas, por ejemplo de las imágenes clave de índices  $j-1$ ,  $j$  y  $j+1$ .

Se presentan tres casos si el mínimo de las dos pendientes, definidas por los dos mínimos ( $\min G$  y  $\min D$ ) y el valor máximo, no es superior al umbral:

- i) Si  $|VS(j) - \min G|$  es inferior al umbral y  $\min G$  no corresponde a  $VS(o)$ , el máximo  $VS(j)$  se ignora y  $\min D$  se vuelve  $\min G$ ,
- ii) Si  $|VS(j) - \min G|$  es superior al umbral y si  $|VS(j) - \min D|$  es inferior al umbral, el  $\min D$  y el máximo  $VS(j)$  se conservan y  $\min D$  se ignora salvo si el máximo más próximo a la derecha de  $\min D$  es superior a un umbral. En este caso, también se conserva  $\min D$  y se declara  $j$  como un índice de una imagen clave. En el caso en el que  $\min D$  se ignora,  $\min D$  adoptará el valor más próximo del mínimo situado a la derecha de  $\min D$ .
- iii) Si las dos pendientes son inferiores al umbral,  $\min G$  se conserva y  $\min D$  y  $j$  se ignoran.

Tras la selección de una imagen clave, se itera el procedimiento. En cada iteración  $\min D$  se vuelve  $\min G$ .

Si se hace referencia de nuevo a la figura 1, a partir de una base de términos 3 que comprende  $P$  términos, se procede en una etapa 4 a un tratamiento de los términos  $t_i$  y a su reagrupamiento en conceptos  $C_i$  (figura 2) destinados a almacenarse en un diccionario de conceptos 5. Se trata en este caso de elaborar un conjunto de firmas que caracterizan una clase de documentos. Las firmas son descriptores que, por ejemplo en el caso de la imagen, representan el color, la forma y la textura. Un documento puede entonces caracterizarse y representarse por los conceptos del diccionario.

Entonces puede formarse una huella de un documento por los vectores firmas de cada concepto del diccionario 5. El vector firma está constituido por los documentos en los que el concepto  $C_i$  está presente así como por las posiciones y el peso de este concepto en el documento.

Los términos  $t_i$  extraídos de una base de documentos 1 se almacenan en una base de términos 3 y se tratan en un módulo 4 de extracción de conceptos  $c_i$  que se reagrupan a su vez en un diccionario de conceptos 5. La figura 2 ilustra el procedimiento de construcción de una base de conceptos  $c_i$  ( $1 \leq i \leq m$ ) a partir de términos  $t_j$  ( $1 \leq j \leq n$ ) que presentan resultados de similitud  $w_{ij}$ .

El módulo de la producción del diccionario de conceptos recibe en la entrada el conjunto  $P$  de los términos de la base 3 y el número máximo  $N$  de conceptos deseado se fija por el usuario: cada concepto  $c_i$  está previsto para reagrupar todos los términos próximos desde el punto de vista de sus características.

Para producir el diccionario de conceptos, se empieza por calcular la matriz de distancia  $T$  entre los términos de la base 3, esta matriz se utiliza para crear una división cuyo cardinal es igual al número  $N$  de conceptos deseado.

La creación del diccionario de conceptos se efectúa en dos etapas: descomposición de  $P$  en  $N$  partes  $P = P_1 \cup P_2 \dots \cup P_N$

Procedimiento de optimización de la división que descompone  $P$  en  $M$  clases  $P = C_1 \cup C_2 \dots \cup C_M$  siendo  $M$  inferior o igual a  $P$ .

El procedimiento de optimización tiene como objetivo de reducir el error del reparto de  $P$  en  $N$  partes  $\{P_1, P_2, \dots, P_N\}$  en la que cada parte  $P_i$  se representa por el término  $t_i$  que se tomará como concepto, el error cometido es entonces igual a la siguiente expresión:

$$\mathcal{E} = \sum_{i=1}^N \mathcal{E}_{t_i}, \quad \mathcal{E}_{t_i} = \sum_{t_j \in P_i} d^2(t_i, t_j)$$

es el error cometido cuando se sustituyen los términos  $t_j$  de  $P_i$  por  $t_i$ .

Se puede descomponer  $P$  en  $N$  partes de manera que se reparten los términos de tal manera que los términos más alejados se encuentran en partes distintas y los términos próximos se encuentran en la misma parte.

Se describirá en primer lugar la etapa 1 de descomposición del conjunto de términos  $P$  en dos partes  $P_1$  y  $P_2$ :

(a) Se determinan los dos términos más alejados  $t_i$  y  $t_j$  de  $P$  que corresponden a la distancia más grande  $D_{ij}$  de la matriz  $T$ .

(b) Para cada  $t_k$  de  $P$ ,  $t_k$  se asigna a  $P_1$  si la distancia  $D_{ki}$  es más pequeña que la distancia  $D_{kj}$  y si no a  $P_2$ .

Se itera la etapa 1 hasta la obtención del número de partes deseado y en cada iteración se aplican las etapas (a) y (b) en los términos del conjunto  $P_1$  y del conjunto  $P_2$ .

Ahora se describirá una etapa de optimización.

El procedimiento de optimización tiene como punto de partida las N partes desunidas de P  $\{P_1, P_2, \dots, P_N\}$  así como los N términos  $\{t_1, t_2, \dots, t_N\}$  que las representan y se utiliza con el fin de reducir el error de descomposición de P en  $\{P_1, P_2, \dots, P_N\}$  partes.

$$\epsilon c_i = \sum_{t_j \in P_i} d^2(t_i, t_j)$$

Se empieza por calcular los centros de gravedad  $C_i$  de  $P_i$ . A continuación se calcula el error que se compara con  $\epsilon c_i$  y se sustituye  $t_i$  por  $C_i$  si  $\epsilon c_i$  es inferior a  $\epsilon t_i$ . Luego, después de haber calculado la nueva matriz  $T_j$  y si no se alcanza la convergencia, se procede a una descomposición. La condición de parada se define por

$$\frac{(\epsilon c_t - \epsilon c_{t+1})}{\epsilon c_t}$$

$< \text{umbral}$  que es del orden de  $10^{-3}$ . Siendo  $\epsilon c_t$  el error cometido en el instante t que representa la iteración.

A continuación se presenta una matriz T de distancias entre los términos, en la que  $D_{ij}$  designa la distancia entre el término  $t_i$  y el término  $t_j$ .

	$t_0$		$t_i$		$t_k$		$t_j$		$t_n$
$t_0$	$D_{00}$		$D_{0i}$		$D_{0k}$		$D_{0j}$		$D_{0n}$
$t_i$	$D_{i0}$		$D_{ii}$		$D_{ik}$		$D_{ij}$		$D_{in}$
$t_k$	$D_{k0}$		$D_{ki}$		$D_{kk}$		$D_{kj}$		$D_{kn}$
$t_j$	$D_{j0}$		$D_{ji}$		$D_{jk}$		$D_{jj}$		$D_{jn}$
$t_n$	$D_{n0}$		$D_{ni}$		$D_{nk}$		$D_{nj}$		$D_{nn}$

La figura 3 ilustra, en el caso de documentos multimedia de contenido diverso, un ejemplo de estructuración del diccionario de conceptos 5.

Con el fin de facilitar la navegación dentro del diccionario 5 y de determinar rápidamente durante una etapa de identificación el concepto más próximo de un término dado, se analiza el diccionario 5 y se establece un mapa 9 de navegación dentro del diccionario.

La producción de la mapa 9 de navegación se efectúa de manera iterativa. En cada iteración, se empieza por dividir el conjunto de conceptos en dos subconjuntos, y luego en cada iteración se selecciona un subconjunto hasta la obtención del número de grupos deseado o bien hasta que se satisface el criterio de parada. Este criterio de parada puede ser por ejemplo que los subconjuntos obtenidos son todos homogéneos con una pequeña desviación estándar por ejemplo. El resultado final es un árbol binario en el que las hojas contienen los conceptos del diccionario y los nudos del árbol contienen las informaciones necesarias para el escrutinio del árbol durante la etapa de identificación de un documento.

Se describirá a continuación un ejemplo de módulo 6 de reparto de un conjunto de conceptos.

El conjunto de conceptos C se representa en forma de una matriz  $M = [c_1, c_2, \dots, c_N] \in \mathbb{R}^{p \times N}$ , con

$c_i \in \mathbb{R}^p$ , en la que  $c_i$  representa un concepto de p valores. Diferentes procedimientos son posibles para garantizar un reparto axial. En este caso, se empieza por calcular el centro de gravedad C así que el eje utilizado para descomponer el conjunto en dos subconjuntos.

Las etapas de tratamiento son las siguientes:

Etapas 1: calcular un representante de la matriz M tal como el centroide w de la matriz

$$M: w = \frac{1}{N} \sum_{i=1}^N c_i \quad (13)$$

Etapas 2: calcular la matriz de covarianza  $\tilde{M}$  entre los elementos de la matriz M y el representante de la matriz M

con, en el caso particular anterior

$$\tilde{M} = M - we, \text{ en la que } e = [1,1,1,\dots,1] \quad (14)$$

5 Etapa 3: calcular un eje de proyección de los elementos de la matriz M, por ejemplo el vector propio U asociado al valor propio más grande de la matriz de covarianza.

Etapa 4: calcular el valor  $p_i = u^T(c_i - w)$  y descomponer el conjunto de conceptos C en dos subconjuntos C1 y C2 de la siguiente manera:

10

$$\begin{cases} c_i \in C1 & \text{si } p_i \leq 0 \\ c_i \in C2 & \text{si } p_i > 0 \end{cases} \quad (15)$$

Las informaciones almacenadas en el nudo asociado a C son  $\{u, w, |p_1|, p_2\}$  siendo  $p_1$  el máximo de todos los  $p_i \leq 0$  y  $p_2$  el mínimo de todos los  $p_i > 0$ .

15

El conjunto  $\{u, w, |p_1|, p_2\}$  constituye los indicadores de navegación en el diccionario de conceptos. En efecto, para determinar, durante la etapa de identificación por ejemplo, el concepto más próximo de un término  $t_i$ , se calcula el valor  $p_{ti} = u^T(t_i - w)$  luego se selecciona el nudo asociado a C1 si  $|p_{ti}| - |p_1| < |p_{ti}| - |p_2|$  y si no se selecciona el nudo C2. Se itera el procedimiento hasta que se alcance una de las hojas del árbol.

20

Un módulo detector de singularidad 8 puede asociarse al módulo 6 de reparto de conceptos.

Este detector de singularidad permite seleccionar el conjunto  $C_i$  que va a descomponerse. Uno de los posibles procedimientos consiste en seleccionar el conjunto menos compacto.

25

Las figuras 4 y 5 ilustran la indexación de un documento o de una base de documentos y la construcción de una base de huellas 10.

30

La base de huellas 10 está constituida por el conjunto de conceptos que representa los términos de los documentos que van a protegerse. A cada concepto  $C_i$  de la base de huellas 10 se asocia una huella 11, 12, 13 constituida por un conjunto de informaciones tales como el número de términos en los documentos en los que el concepto está presente, y para cada uno de estos documentos se registra una huella 11a, 11b, 11c que comprende el índice del documento que remite a la dirección del documento, el número de términos, el número de apariciones del concepto (frecuencia), el resultado, así como los conceptos que le son próximos en el documento. La puntuación es un valor medio de las medidas de similitud entre el concepto y los términos del documento que son los más próximos al concepto. El índice de un documento dado que remite a la dirección de este documento se almacena en una base 14 de las direcciones de los documentos protegidos.

35

40

El procedimiento 20 de generación de las huellas o firmas de documentos que va a indexarse se ilustra en la figura 5.

45

Durante el registro de un documento, se extraen los términos pertinentes del documento (etapa 21) y se tiene en cuenta el diccionario de conceptos (etapa 22). Cada uno de los términos  $t_i$  del documento se proyecta en el espacio del diccionario de conceptos con el fin de determinar el concepto  $C_i$  que representa el término  $t_i$  (etapa 23).

50

A continuación, se actualiza la huella del concepto  $C_i$  (etapa 24). Esta actualización se efectúa según si el concepto ya se ha encontrado, es decir está presente en los documentos que ya se han registrado, o no.

55

Si el concepto  $C_i$  todavía no está presente en la base, se crea una nueva entrada en la base (una entrada en la base corresponde a un objeto cuyos elementos son objetos que contienen la firma del concepto en los documentos en los que este concepto está presente). Se inicializa la entrada creada con la firma del concepto. La firma de un concepto en un documento se materializa principalmente por las siguientes informaciones: dirección del documento, número de términos, frecuencia, conceptos próximos y puntuación.

60

Si el concepto  $C_i$  existe en la base, se añade a la entrada asociada al concepto su firma en el documento que está compuesta por (dirección del documento, número de términos, frecuencia, conceptos Próximos y puntuación).

Cuando se construye la base de huellas (etapa 25), se procede al registro de la base de huellas (etapa 26).

60

La figura 6 ilustra un procedimiento de identificación de un documento que se implementa en una plataforma 30 de búsqueda en línea.

La identificación de un documento tiene como objetivo determinar si un documento planteado como pregunta es la reutilización de un documento de la base. Se basa en la medida de similitud entre documentos. El objetivo es identificar los documentos que contienen elementos protegidos. La reutilización puede ser total o parcial. En este último caso, el elemento copia ha sufrido modificaciones tales como: supresión de frases en un texto, supresión de motivo en una imagen, supresión de plan o de secuencia en un documento de vídeo,... cambio de orden de los términos o sustitución de términos por otros términos en un texto.

Después de la presentación de un documento que va a identificarse (etapa 31), se procede a la extracción de los términos de este documento (etapa 32).

En conexión con una base de huellas (etapa 25), se ponen en correspondencia los conceptos calculados a partir de los términos extraídos de la pregunta, con los conceptos de base (etapa 33), con el fin de establecer una lista de documentos que tienen contenidos similares al contenido del documento pregunta.

El procedimiento de establecimiento de la lista es el siguiente:

Se indica  $p_{dj}$ : el grado de semejanza del documento  $d_j$  con el documento pregunta, con  $1 \leq j \leq N$ ,  $N$  es el número de documentos de la base de referencia

Se inicializan a cero todos los  $p_{dj}$

Para cada término  $t_i$  de la pregunta proporcionada en la etapa 331 (figura 7) se determina el concepto  $C_i$  que lo representa (etapa 332).

Para cada documento  $d_j$  en el que está presente el concepto se actualiza su  $p_{dj}$  de la siguiente manera:

$p_{dj} = p_{dj} + f(\text{frecuencia, puntuación})$ , pueden utilizarse varias funciones  $f$  por ejemplo  $f(\text{frecuencia, puntuación}) = \text{frecuencia} \times \text{puntuación}$ , frecuencia designa el número de apariciones del concepto  $C_i$  en el documento  $d_j$  y puntuación designa la media de las puntuaciones de semejanza de los términos del documento  $d_j$  con el concepto  $C_j$ .

Se ordenan los  $p_{dj}$  y se conservan los que son superiores a un umbral dado (etapa 333). Se procede a continuación a una confirmación y una validación de las respuestas (etapa 34).

Confirmación de las respuestas: se filtra la lista de las respuestas con el fin de conservar sólo las respuestas más pertinentes. El filtrado utilizado se basa en la correlación entre los términos de la pregunta y de cada una de las respuestas.

Validación: permite conservar solo las respuestas en las que hay una grande certidumbre de reutilización de contenido. En esta etapa se filtran las respuestas teniendo en cuenta las propiedades algebraicas y topológicas de los conceptos dentro de un documento: se exige que la proximidad en el documento pregunta se respete en los documentos respuesta, es decir que dos conceptos próximos en el documento pregunta tienen que estar próximos en el documento respuesta.

Entonces se proporciona la lista de los documentos respuesta (etapa 35).

Ahora se considerará más particularmente el caso de documentos multimedia que contienen imágenes.

Se describirán en particular, para la construcción de la base de huellas que servirá de herramienta para la identificación de un documento, procedimientos rápidos y eficaces de identificación de imágenes que tienen en cuenta todas las informaciones pertinentes contenidas en las imágenes desde la caracterización de las estructuras u objetos que la componen, hasta la de las zonas texturizadas y el color de fondo. Los objetos de la imagen se identifican por la producción de una tabla que resume diferentes estadísticas realizadas en informaciones de las zonas fronteras de los objetos así como informaciones en las proximidades de estas zonas fronteras. La caracterización de las zonas texturizadas puede efectuarse con la ayuda de una descripción muy fina a la vez espacial y espectral de la textura según tres características fundamentales que son su periodicidad, su orientación global y el aspecto aleatorio de su motivo. La textura se asimila en este caso a una realización de procedimiento aleatorio bidimensional. La caracterización del color es una parte importante del procedimiento. Puede utilizarse como una primera clasificación de las respuestas similares basadas en el color, o bien como una última decisión realizada para afinar la búsqueda.

En la primera parte de la etapa de construcción de huellas, se tienen en cuenta las informaciones clasificadas en forma de componentes que pertenecen a dos grandes categorías:

- las componentes denominadas estructurales que describen la percepción por el ojo de un objeto puede aislarse

o de un conjunto de objetos dispuesto según una disposición espacial (imágenes 81 y 82 de las figuras 11 y 12),

- las componentes denominadas texturales que son el complemento de las componentes estructurales y que traducen la regularidad o la homogeneidad de los motivos de textura (imágenes 82 y 83 de las figuras 12 y 13).

La figura 11 muestra así una imagen 81 que contiene elementos estructurales y que no presentan motivos de textura.

La figura 12 muestra una imagen 81 que contiene elementos estructurales y un fondo de textura.

La figura 13 muestra una imagen 83 sin elementos estructurales pero totalmente texturizada.

Tal como se ha indicado anteriormente, durante la etapa de construcción de huellas, cada documento de la base de documentos se analiza con el fin de extraer del mismo las informaciones pertinentes. Estas informaciones se catalogarán y analizarán a continuación. Este análisis se realiza según una cadena de procedimientos que se resume en tres etapas:

- Extraer para cada documento características predefinidas y almacenar estas informaciones en un vector denominado término.
- Reagrupar en un concepto todos los términos "próximos" desde el punto de vista de sus características, lo que permite hacer la búsqueda más concisa.
- Construir una huella que caracteriza este documento por un número reducido de entidades. Cada documento se asocia entonces a una huella que le es propia.

La figura 8 ilustra el caso de la indexación de un documento imagen 52 contenido en una base de imágenes 51 previamente registrada, para caracterizar esta imagen 52 por un número finito de parámetros que pueden almacenarse fácilmente y manipularse posteriormente. Se procede en la etapa 53 a la extracción de términos del documento que va a buscarse que se almacenan en una memoria intermedia (etapa 54).

Se efectúa en la etapa 55 una proyección en el espacio de los términos de la base de referencias.

En la etapa 56, se procede a una descripción vectorial que proporciona los valores de importancia de los términos en el documento que va a buscarse.

La etapa 57 consiste en un reparto de los términos en N grupos 58 de conceptos.

La etapa 59 consiste en una proyección en el espacio de conceptos de cada grupo 58 para obtener N divisiones 61.

Por último, una proyección ortogonal 62 conduce a N conjuntos 63 de descripciones vectoriales reducidas.

Durante una etapa posterior de búsqueda, después de una consulta formulada por un usuario, por ejemplo la identificación de una imagen pregunta, se buscan todos los documentos multimedia similares o que contestan a esta consulta. Para ello, tal como se ha indicado anteriormente, se calculan los términos del documento pregunta y se comparan con los conceptos de la base con el fin de deducir el o los documentos de la base que son similares al documento pregunta.

A continuación se describirá de manera más detallada la etapa de construcción de los términos de una imagen.

La etapa de construcción de los términos de una imagen pone en práctica de manera útil la caracterización de los soportes estructurales de la imagen. Los soportes estructurales son los elementos que componen la escena de la imagen. Los más significativos son los que delimitan los objetos de la escena ya que son ellos los que caracterizan las diferentes formas que se perciben cuando se observa una imagen cualquiera.

Esta etapa se refiere a la extracción de estos soportes estructurales. Consiste en un desmantelamiento de las zonas fronteras de los objetos de la imagen, que se caracterizan por sitios entre dos zonas en los que se observan fuertes variaciones de intensidad. Este desmantelamiento se efectúa por un procedimiento que consiste en repartir estas zonas fronteras entre diferentes "clases" según la orientación local del gradiente de la imagen (orientación de la variación local de intensidad). Se obtiene así una multitud de pequeños elementos denominados los "elementos de soporte estructurales" (ESS). Cada ESS que pertenece efectivamente a un contorno de una escena se caracteriza por una similitud a nivel de la orientación local de su gradiente. Esto es una primera etapa que tiende a catalogar todos los elementos de soporte estructurales de la imagen.

El siguiente enfoque se efectúa en adelante a partir de *estos* ESS, a saber la construcción de términos que describen las propiedades locales y globales de los ESS.

Las informaciones extraídas de cada soporte se consideran como propiedades locales. Pueden distinguirse dos tipos de soportes: los elementos de rectas rectilíneas (EDR) y los elementos de arcos de curvas (EAC).

5 Los elementos de rectas rectilíneas EDR se caracterizan por las propiedades locales que son:

- La dimensión (longitud, anchura)
- Dirección principal (pendiente)
- Propiedades estadísticas de los píxeles que constituyen el soporte (valor medio de energía, los instantes)
- Informaciones de la proximidad (transformada de Fourier local)

En cuanto a los elementos de arcos de curvas EAC se caracterizan de la misma manera que anteriormente, además de la curvatura de los arcos.

15 Las propiedades globales engloban las estadísticas tales como el número de cada tipo de soportes y sus disposiciones espaciales (asociaciones geométricas entre los soportes: conexiones, izquierda, derecha, medio...).

En resumen, para una imagen dada, las informaciones pertinentes extraídas de los objetos que la constituyen se reagrupan en la tabla 1.

Tabla 1

Soportes estructurales de los objetos de una imagen		Tipo		
		ESS	EDR	EAC
Propiedades globales	N.º total	n	n <sub>1</sub>	n <sub>2</sub>
	N.º largos (> umbral)	n <sub>l</sub>	n <sub>1l</sub>	n <sub>2l</sub>
	N.º cortos (< umbral)	n <sub>c</sub>	n <sub>1c</sub>	n <sub>2c</sub>
	N.º de soportes largos a una conexión izquierda o derecha	-	n <sub>1</sub> 1gdx	n <sub>2</sub> 1gdx
	N.º de una conexión en el medio	-	n <sub>1</sub> 1gdx	n <sub>2</sub> 1gdx
	N.º de soportes largos paralelos	-	n <sub>1</sub> pll	n <sub>2</sub> pll
Propiedades locales	Luminancia (>umbral)	-		
	Luminancia (<umbral)	-		
	Pendiente	-		
	Curvatura	-		
	Caracterización de la proximidad de los soportes	-		

25 La etapa de construcción de los términos de una imagen también pone en práctica la caracterización de las informaciones texturales pertinentes de la imagen. Las informaciones procedentes de la textura de la imagen se dividen según tres aspectos visuales de la imagen:

- el aspecto aleatorio (como una imagen de arena fina, o de hierba) en el que no puede detectarse ninguna disposición particular,
- el aspecto periódico (como un jersey de jacquard) en el que se observa una repetición de motivos (píxel o agrupamiento de píxeles) dominantes,
- y por último el aspecto direccional en el que los motivos tienden globalmente a orientarse hacia una o varias direcciones privilegiadas.

Estas informaciones se obtienen aproximándose a la imagen por modelos o representaciones paramétricas. Cada aspecto se tiene en cuenta por sus representaciones espacial y espectral que constituyen las informaciones pertinentes de esta parte de la imagen. La periodicidad y la orientación se caracterizan por los soportes espectrales mientras que el aspecto aleatorio se traduce por la estimación de los parámetros de un modelo autorregresivo bidimensional.

Una vez extraídas todas las informaciones pertinentes, se puede proceder a la estructuración de los términos de las texturas.

Tabla 2

Soportes espectrales y parámetros autorregresivos de la textura de una imagen		
Componente periódico	N.º total de elementos periódicos	n <sub>p</sub>
	Frecuencias	par (ω <sub>p</sub> , ν <sub>p</sub> ), 0 < p ≤ n <sub>p</sub>
	Amplitudes	par (C <sub>p</sub> , D <sub>p</sub> ), 0 < p ≤ n <sub>p</sub>

Tabla 2 (continuación)

Soportes espectrales y parámetros autorregresivos de la textura de una imagen		
Componente direccional	N.º total de elementos direccionales	nd
	Orientaciones	par $(\alpha_i, \beta_i)$ , $0 < i \leq nd$
	Frecuencias	$v_i$ , $0 < i \leq nd$
Componentes aleatorios	Desviación estándar del ruido	$\sigma$
	Parámetros autorregresivos	$\{a_{i,j}\}$ , $(i, j) \in S_{N,M}$

5 Por último, la etapa de construcción de los términos de una imagen también puede poner en práctica la caracterización del color de la imagen.

El color se representa a menudo por los histogramas de color, estos últimos son invariables a la rotación y robustos contra la oclusión y los cambios de puntos de vista de la cámara.

10 La cuantificación de los colores puede hacerse en el espacio RVB (rojo, verde, azul), TSV (tinta saturación valor), o el espacio LUV pero el procedimiento de indexación por los histogramas de colores ha demostrado sus límites ya que da una información global de la imagen, y durante la indexación pueden encontrarse imágenes que presentan el mismo histograma de color, pero que son totalmente diferentes.

15 Muchos autores proponen histogramas de colores integrando la información espacial. Esto consiste por ejemplo en distinguir los píxeles coherentes de los píxeles incoherentes, un píxel es coherente si pertenece a una región bastante grande que reagrupa píxeles idénticos, se clasifica como incoherente si forma parte de una región de tamaño reducido.

20 Se describirá a continuación un procedimiento de caracterización de la distribución espacial de los constituyentes de la imagen (por ejemplo el color) que es menos costoso en tiempos de cálculo que los procedimientos citados anteriormente, y es robusto frente a las rotaciones y a la traslación.

25 Las diferentes características extraídas de los elementos de soporte estructurales así como los parámetros de las componentes periódica, direccional y aleatoria del campo de textura así como los parámetros de la distribución espacial de los constituyentes de la imagen constituyen los términos que pueden servir a la descripción del contenido de un documento. Estos términos se reagrupan en conceptos con el fin de reducir las informaciones útiles de un documento.

30 Las apariciones de estos conceptos así como sus posiciones y sus frecuencias constituyen lo que se llama la huella de un documento. Estas huellas servirán a continuación como rasgo de unión entre un documento pregunta y los documentos de una base, durante una etapa de búsqueda de documento.

35 Una imagen no contiene necesariamente todos los elementos y las características descritas anteriormente. Por consiguiente, identificar una imagen empieza por la detección de la presencia de sus elementos constituyentes.

40 La figura 9 muestra un ejemplo de organigrama de un procedimiento de extracción de los términos de una imagen con una primera etapa 71 de caracterización de los objetos de la imagen en soportes estructurales, que puede en su caso precederse de una prueba de detección de elementos estructurales que permite omitir esta etapa 71 en los casos en los que los elementos estructurales están ausentes.

45 La etapa 72 consiste en una prueba para determinar si existe un fondo de textura. Si es así, se pasa a una etapa 73 de caracterización del fondo texturizado en soportes espectrales y parámetros autorregresivos AR, y luego a una etapa 74 de caracterización del color de fondo.

Si no existe un fondo estructurado, se pasa directamente de la etapa 72 a la etapa 74.

50 Por último, una etapa 75 reside en el almacenamiento de los términos y la construcción de huellas.

Se volverá ahora con más detalle a la caracterización de los elementos de soporte estructurales de una imagen.

55 El principio de base de esta caracterización consiste en un desmantelamiento de las zonas fronteras de los objetos de la imagen en multitudes de pequeños elementos de base denominados elementos de soportes significativos (ESS) que transportan las informaciones útiles de las zonas fronteras que están compuestas por bandas lineales de tamaño variable, o codos de diferentes curvaturas. Entonces se analizarán estadísticas realizadas en estos objetos y se utilizarán para construir los términos de estos soportes estructuras.

Con el fin de describir más rigurosamente los principales procedimientos que componen este enfoque, se indicará

una imagen digitalizada por el conjunto  $\{y(i,j), (i,j) \in I \times J\}$ , en la que  $I$  y  $J$  son respectivamente el número de líneas y de columnas de la imagen.

5 A partir de las imágenes *gradiente vertical*  $\{g_v(i,j), (i,j) \in I \times J\}$  y *horizontal*  $\{g_h(i,j), (i,j) \in I \times J\}$  previamente calculadas, este enfoque consiste en dividir la imagen según la orientación local de su gradiente en un número finito de clases equidistantes. La imagen que contiene la orientación del gradiente se define por la fórmula:

$$O(i,j) = \arctan\left(\frac{g_h(i,j)}{g_v(i,j)}\right) \quad (1)$$

10 La división es sencillamente una subdivisión angular del plano 2D (de  $0^\circ$  a  $360^\circ$ ) mediante un paso de discretización bien definido. El hecho de utilizar la orientación local del gradiente como criterio de descomposición de las zonas fronteras permite un mejor agrupamiento de los píxeles que forman parte de una misma zona frontera. Con el fin de resolver el problema de los puntos fronteras que se pueden repartir entre dos clases yuxtapuestas, se utiliza una segunda división con el mismo número de clases que anteriormente, pero desplazadas  $\frac{1}{2}$  clase. A partir de las  
15 clases procedentes de las dos divisiones, un procedimiento sencillo consiste en elegir las que presentan en total el número más grande de píxeles. En efecto, cada píxel pertenece a dos clases cada una procedente de las dos divisiones. Sabiendo que cada píxel es un elemento potencial de un eventual ESS, vota por lo tanto para la clase que contiene más píxeles de las dos. Se trata de una región en la que la probabilidad de encontrar un ESS de tamaño más elevado es lo más grande posible. Después de los votos, se conservan únicamente las clases que  
20 presentan en total más del 50% de los sufragios. Estas son las regiones de soporte susceptibles de contener los ESS.

A partir de estas *regiones de soporte*, se determinan los ESS, se les catalogan según criterios determinados que pueden ser:

- 25 • La longitud (se determina para es un umbral  $l_0$  y se contabilizan los ESS inferiores y superiores a este umbral)
- La intensidad definida por la media del módulo del gradiente de los píxeles que componen cada ESS (por lo tanto se define un umbral indicado  $l_0$ , se catalogan los que son interiores y superiores a este umbral).
- 30 • El contraste definido por la diferencia entre el máximo y el mínimo de los píxeles.

En esta etapa del procedimiento, todos los elementos denominados estructurales se conocen y se catalogan según los tipos de soportes estructurales previamente identificados. Pueden extraerse de la imagen de origen para dejar  
35 lugar a la caracterización del campo de texturas.

A modo de ejemplo, se considera la imagen 81 de la figura 11, reutilizada como imagen 101 de la figura 14a, las zonas fronteras se ilustran en la imagen 102 de la figura 14b. Los elementos de estas zonas fronteras se  
40 desmantelan a continuación y se reparten según la orientación de su gradiente entre diferentes clases representadas por las imágenes 103 a 106 de las figuras 14c a 14f. Estos diferentes elementos constituyen los elementos de soportes significativos, y sus análisis estadísticos permiten construir los términos de la componente estructural.

En el caso de las figuras 14c a 14f, a modo de ejemplo, la imagen 103 corresponde a una clase 0 ( $0^\circ - 45^\circ$ ), la imagen 104 corresponde a una clase 1 ( $45^\circ - 90^\circ$ ), la imagen 105 corresponde a una clase 2 ( $90^\circ - 135^\circ$ ) y la imagen 106 corresponde a una clase 3 ( $135^\circ - 180^\circ$ ):

En ausencia de elementos estructurales, se supone que la imagen se textura con motivos más o menos regulares y se procede a una caracterización del campo de la textura. Para ello, se puede proceder a una descomposición de la  
50 imagen en tres componentes que son:

- una componente textural que contiene las informaciones anárquicas o aleatorias (como una imagen de arena fina, o de hierba) en la que no puede detectarse ninguna disposición particular,
- 55 • una componente periódica (como un jersey de jacquard) en la que se observa una repetición de motivos dominantes,
- y por último una componente direccional en la que los motivos tienden globalmente hacia una o varias direcciones privilegiadas.

60 Al ser el objetivo caracterizar perfectamente la textura de la imagen a partir de un conjunto de parámetros, estas tres componentes se representan por modelos paramétricos.

Así, la textura de la imagen 15 regular y homogénea indicada  $\{\tilde{y}(i,j), (i,j) \in I \times J\}$  se descompone en tres componentes 16, 17, 18 tal como se ilustra en la figura 10, según la siguiente relación:

$$\{\tilde{y}(i, j)\} = \{w(i, j)\} + \{h(i, j)\} + \{e(i, j)\}. \quad (16)$$

en la que  $\{w(i,j)\}$  es la componente puramente aleatoria 16,  $\{h(i,j)\}$  es la componente armónica 17 y  $\{e(i,j)\}$  la componente direccional 18. La estimación de los parámetros de estas tres componentes 16, 17, 18 termina esta etapa de extracción de informaciones de un documento. En los siguientes párrafos se describen procedimientos de estimación.

Se describirá en primer lugar un ejemplo de procedimiento de detección y caracterización de la componente direccional de la imagen.

Se trata en primer lugar de aplicar un modelo paramétrico a la componente direccional  $\{e(i,j)\}$ . Está constituida por una suma numerable de elementos direccionales en los que cada uno se asocia a un par de números enteros  $(\alpha, \beta)$  que define una orientación de ángulo  $\theta$  tal que  $\theta = \tan^{-1} \beta/\alpha$ . En otras palabras,  $e(i, j)$  se define por

$$e(i, j) = \sum_{(\alpha, \beta) \in O} e_{(\alpha, \beta)}(i, j)$$

en la que cada  $e_{(\alpha, \beta)}(i, j)$  se define por:

$$e_{(\alpha, \beta)}(i, j) = \sum_{k=1}^{N_e} [s_k^{\alpha, \beta}(i\alpha - j\beta) \times \cos(2\pi \frac{v_k}{\alpha^2 + \beta^2}(i\beta + j\alpha)) + t_k^{\alpha, \beta}(i\alpha - j\beta) \times \sin(2\pi \frac{v_k}{\alpha^2 + \beta^2}(i\beta + j\alpha))] \quad (17)$$

en la que

- $N_e$  es el número de elementos direccionales asociados a  $(\alpha, \beta)$ ,
- $v_k$  es la frecuencia del elemento  $k$ ,
- $\{s_k(i\alpha - j\beta)\}$  y  $\{t_k(i\alpha - j\beta)\}$  son las amplitudes.

La componente direccional  $\{e(i,j)\}$  está por tanto perfectamente definida por el conocimiento de los parámetros contenidos en el siguiente vector  $E$ :

$$E = \{\alpha_1, \beta_1, \{v_{1k}, s_{1k}(c), t_{1k}(c)\}_{k=1}^{N_e}\}_{(\alpha, \beta) \in O} \quad (18)$$

Para estimar estos parámetros, se utiliza el hecho que la componente direccional de una imagen se representa en el ámbito espectral por un conjunto de rectas de pendientes ortogonales a las definidas por los pares de números enteros  $(\alpha_i, \beta_i)$  del modelo que se indicarán  $(\alpha_i, \beta_i)^\perp$ . Estas rectas pueden descomponerse en un sub-conjunto de rectas de igual pendiente asociado cada uno a un elemento direccional.

A modo de ilustración, las figuras 15a y 15b muestran imágenes 84, 86 que contienen un elemento direccional y la figura 15c muestra una imagen 88 que contiene dos elementos direccionales.

La figura 15a1 muestra una vista 85 en tres dimensiones del espectro de la imagen 84 de la figura 15a.

Las figuras 15b1 y 15c1 muestran imágenes módulo de Fourier 87, 89, respectivamente imágenes 86 y 85 de las figuras 15b y 15c.

Para calcular los elementos del vector  $E$ , puede adoptarse un enfoque basado en la proyección de la imagen según diferentes direcciones. El procedimiento consiste en primer lugar en garantizar la presencia de la componente direccional antes de estimar estos parámetros.

La detección de la componente direccional de la imagen se basa en el conocimiento de las propiedades espectrales de ésta. Si se asimila el espectro de la imagen a una imagen 3D  $(X, Y, Z)$ , en la que  $(X, Y)$  representan las coordenadas de los píxeles y  $Z$  la amplitud, las rectas que intentan detectarse se representan por un conjunto de picos concentrados a lo largo de rectas cuyas pendientes se definen por los pares  $(\alpha_i, \beta_i)$  buscados (véase la figura 15a1). Para determinar la presencia de estas rectas, basta con contabilizar los picos predominantes. El número de

estos picos informa de la presencia o no de soportes direccionales o armónicos.

Ahora se describirá un ejemplo de procedimiento de caracterización de la componente direccional. Para ello, se procede al cálculo de los pares de dirección  $(\alpha_k, \beta_k)$  y a la determinación del número de elementos direccionales.

5 Se efectúa en primer lugar el cálculo de la transformada de Fourier discreta (TFD) de la imagen seguido de una estimación de las rectas de pendiente racional observadas en la imagen transformada  $\psi(i,j)$ .

10 Para ello, se define un conjunto de proyecciones que discretiza el ámbito frecuencial en diferentes ángulos de proyección  $\theta_k$ ,  $k$  finito. Este conjunto de proyección puede obtenerse de diferentes maneras. Por ejemplo se pueden buscar todos los pares de números enteros primeros entre ellos  $(\alpha_k, \beta_k)$  que definen un ángulo  $\theta_k$ , tal que

$$\theta_k = \tan^{-1} \frac{\alpha_k}{\beta_k} \quad \text{en la que} \quad 0 \leq \theta_k \leq \frac{\pi}{2}.$$
 Un orden  $r$  tal que  $0 \leq \alpha_k, \beta_k \leq r$  permite controlar el número de proyecciones. Las propiedades de simetría pueden utilizarse a continuación para obtener todos los pares hasta  $2\pi$ . Estos pares se ilustran en la figura 16 para  $0 \leq \alpha_k, \beta_k \leq 3$ .

15 Se efectúan proyecciones del módulo de la TFD de la imagen según siguiente los  $\theta_k$ . Cada proyección genera un vector de dimensión 1,  $V(\alpha_k, \beta_k)$ , indicado  $V_k$  para simplificar la notación, que contiene las informaciones direccionales buscadas.

20 Cada proyección  $V_k$  viene dada por la fórmula:

$$V_k(n) = \sum_{\tau} \Psi(i + \tau\beta_k, j + \tau\alpha_k), \quad 0 < i + \tau\beta_k < I - 1, 0 < j + \tau\alpha_k < J - 1 \quad (19)$$

25 siendo  $n = -i * \beta_k + j * \alpha_k$  y  $0 \leq |n| < N_k$  y  $N_k = |\alpha_k|(T-1) + |\beta_k|(L-1) + 1$ , en la que  $T * L$  es el tamaño de la imagen.  $\psi(i,j)$  es el módulo de transformada de Fourier de la imagen que va a caracterizarse.

Se seleccionan para cada  $V_k$  los elementos de fuertes energías así como sus posiciones espaciales. Estos elementos de fuerte energía son los que presentan un valor máximo con respecto a un umbral calculado según el tamaño de la imagen.

30 En esta etapa de cálculo, se conoce el número de rectas. De ello se deduce el número de componentes direccionales  $N_e$  utilizando las propiedades espectrales simples de la componente direccional de una imagen texturizada. Estas propiedades son:

- 35 1. Las rectas observadas en el ámbito espectral de una componente direccional son simétricas con respecto al origen. Por consiguiente se puede reducir el ámbito de investigación en solo una mitad del ámbito considerado.
- 40 2. Los máximos retenidos en el vector son candidatos para representar rectas que pertenecen a elementos direccionales. A partir de la conocimiento de las posiciones respectivas de las rectas en el módulo de transformada de Fourier discreta TFD, se deduce el número exacto de elementos direccionales. La posición del máximo recto corresponde al argumento del máximo del vector  $V_k$ , las otras rectas del mismo elemento se sitúan cada  $\min\{L/T\}$ .

45 El mecanismo de proyección se ilustra en la figura 17 para  $(\alpha_k, \beta_k) = (2, -1)$ .

Después del tratamiento de los vectores  $V_k$  y la producción de los pares de dirección  $(\hat{\alpha}_k, \hat{\beta}_k)$  se obtienen los números de rectas asociados a cada par.

50 Por tanto puede contarse el número total de elementos direccionales utilizando las dos propiedades mencionadas anteriormente y se identifican los pares de números enteros  $(\hat{\alpha}_k, \hat{\beta}_k)$  asociados a estos componentes que son las direcciones ortogonales a las que se han retenido.

55 Para todos estos pares  $(\hat{\alpha}_k, \hat{\beta}_k)$  la estimación de las frecuencias de cada elemento detectado es inmediata. En efecto, si se consideran únicamente los puntos de la imagen de origen en a lo largo de la recta de ecuación  $i\hat{\alpha}_k - j\hat{\beta}_k = c$ ,  $c$  es la posición del máximo en  $V_k$ , y estos puntos constituyen una señal monodimensional (1-D)

armónica de amplitud constante y que presenta como frecuencia  $\hat{v}_i^{(\alpha, \beta)}$ . Por lo tanto basta con estimar la frecuencia de esta señal 1-D por un procedimiento clásico (localización del valor máximo en la TFD 1-D de esta nueva señal).

En resumen, puede ponerse en práctica el procedimiento que comprende las siguientes etapas:

- 5 Determinar el máximo de cada proyección.
- Filtrar los máximos con el fin de conservar sólo los superiores a un umbral.
- 10 • Para cada máximo  $m_i$ , correspondiente a un par  $(\hat{\alpha}_k, \hat{\beta}_k)$
- Se determina el número de rectas asociadas a este par según las propiedades descritas anteriormente.
- 15 • Se calcula la frecuencia asociada a  $(\hat{\alpha}_k, \hat{\beta}_k)$  que corresponde a la intersección de la recta máxima (correspondiente al máximo de la proyección retenida) con el eje horizontal.

Ahora se describirá el cálculo de las amplitudes  $\{\hat{s}_k^{(\alpha, \beta)}(t)\}_y \{\hat{t}_k^{(\alpha, \beta)}(t)\}$  que son los otros parámetros contenidos en el vector E mencionado anteriormente.

20 Conociendo la dirección  $(\hat{\alpha}_k, \hat{\beta}_k)$  y la frecuencia  $V_k$ , pueden determinarse las amplitudes  $\hat{s}_k^{(\alpha, \beta)}(c)_y \hat{t}_k^{(\alpha, \beta)}(c)$ , verificando c la fórmula  $i\hat{\alpha}_k - j\hat{\beta}_k = c$ , utilizando un procedimiento de demodulación. En efecto,  $\hat{s}_k^{(\alpha, \beta)}(c)$  es igual a la media de los píxeles a lo largo de la recta de ecuación  $i\hat{\alpha}_k - j\hat{\beta}_k = c$  de la nueva imagen obtenida

multiplicando  $\tilde{y}(i, j)$  por  $\cos\left(\frac{\hat{v}_k^{(\alpha, \beta)}}{\hat{\alpha}_k^2 + \hat{\beta}_k^2}(i\hat{\beta}_k + j\hat{\alpha}_k)\right)$ . Esto se traduce por la ecuación

$$\hat{s}_k^{(\alpha, \beta)}(c) \equiv \frac{1}{N_s} \sum_{i\hat{\alpha}_k - j\hat{\beta}_k = c} \tilde{y}(i, j) \cos\left(\frac{\hat{v}_k^{(\alpha, \beta)}}{\hat{\alpha}_k^2 + \hat{\beta}_k^2}(i\hat{\beta}_k + j\hat{\alpha}_k)\right) \quad (20)$$

25 en la que  $N_s$  es simplemente el número de elementos de esta nueva señal. De la misma manera, se obtiene  $\hat{t}_k^{(\alpha, \beta)}(c)$  aplicando la ecuación:

$$\hat{t}_k^{(\alpha, \beta)}(c) \equiv \frac{1}{N_s} \sum_{i\hat{\alpha}_k - j\hat{\beta}_k = c} \tilde{y}(i, j) \sin\left(\frac{\hat{v}_k^{(\alpha, \beta)}}{\hat{\alpha}_k^2 + \hat{\beta}_k^2}(i\hat{\beta}_k + j\hat{\alpha}_k)\right) \quad (21)$$

30 El procedimiento descrito anteriormente puede resumirse por las siguientes etapas:

Para cualquier elemento direccional  $(\hat{\alpha}_k, \hat{\beta}_k)$

35 Para toda recta (d) calcular

1. la media de los puntos (i,j) ponderada por el  $\hat{s}_k^{(\alpha, \beta)}(d)$  Esta media corresponde a la estimación de la amplitud

$$\sin\left(\frac{v_k^{(\alpha,\beta)}}{\hat{\alpha}_k^2 + \hat{\beta}_k^2}(i\hat{\beta}_k + j\hat{\alpha}_k)\right).$$

2. la media de los puntos (i,j) ponderada por el  $\hat{f}_k^{(\alpha,\beta)}(d)$  Esta media corresponde a la estimación de la amplitud

5

La tabla 3 a continuación recapitula las etapas principales del procedimiento de proyección.

Tabla 3

Etapa 1. Calcular el conjunto de pares de proyección $(\alpha_k, \beta_k) \in P_r$
Etapa 2. Calcular el módulo de la TFD de la imagen $\tilde{y}(i,j): \psi(\omega, \varpi) =  \text{TFD}(y(i,j)) $
Etapa 3-Para todo $((\alpha_k, \beta_k) \in P_r$ calcular el vector $V_k$ : la proyección de $\psi(\omega, v)$ según $(\alpha_k, \beta_k)$ según la fórmula (19)
Etapa 4- Detección de rectas:  Para todo $(\alpha_k, \beta_k) \in P_r$  <ul style="list-style-type: none"> <li>determinar:</li> <li>calcular <math>n_k</math>: el número de píxeles de valores significativos encontrados a lo largo de la proyección</li> <li>guardar <math>n_k</math> y <math>j_{\max}</math> el índice del máximo en <math>V_k</math>.</li> <li>seleccionar las direcciones que justifican el criterio</li> </ul> $\frac{M_k}{n_k} > s_e$ <p>en la que <math>s_e</math> es un umbral que va a definirse, dependiendo del tamaño de la imagen. Las direcciones retenidas se consideran como las de rectas buscadas.</p>
Etapa 5- Guardar los pares $(\hat{\alpha}_k, \hat{\beta}_k)$ buscados que son los ortogonales de los pares $(\alpha_k, \beta_k)$ retenidos en la etapa 4.

10

A continuación se describirá la detección y la caracterización de las informaciones texturales periódicas de una imagen, que están contenidas en la componente armónica  $\{h(i,j)\}$ . Este componente puede representarse por una suma finita de sinusoides 2-D:

$$h(i, j) = \sum_{p=1}^P C_p \cos 2\pi(i\omega_p + j\nu_p) + D_p \sin 2\pi(i\omega_p + j\nu_p), \quad (22)$$

15 en la que

- $C_p$  y  $D_p$  son las amplitudes.
- $(\omega_p, \nu_p)$  es la frecuencia espacial p.

20 Se ha representado en la figura 18a1 una imagen 91 que contiene componentes periódicas y, en la figura 18b1, una imagen sintética que contiene una componente periódica.

La figura 18a2 representa una imagen 92 que es una imagen módulo de la TFD que presenta un conjunto de picos.

25 La figura 18b2 representa una vista 3D, 94, de la TFD que muestra la presencia de un par de picos simétricos 95, 96.

30 En el ámbito espectral, la componente armónica aparece así como pares de picos aislados simétricos con respecto al origen (véase la figura 18 (a2)-(b2)). Este componente es el reflejo de la existencia de periodicidades en la imagen.

Las informaciones que se busca determinar son los elementos del vector:

$$H = \left\{ P, \{C_p, D_p, \omega_p, \nu_p\}_{p=1}^P \right\} \quad (23)$$

35

Para ello se empieza por detectar la presencia de esta componente periódica en la imagen módulo de Fourier y se estiman a continuación sus parámetros.

5 La detección de la componente periódica consiste en determinar la presencia de picos aislados en la imagen módulo de la TFD. Se actúa de la misma manera que en el caso de la determinación de la componente direccional. Según el procedimiento descrito en la tabla 1, si el valor  $n_k$  obtenido en la etapa 4 del procedimiento descrito en la tabla 1 es inferior a un umbral, entonces se trata de picos aislados que caracterizan la presencia de componente armónica en vez de picos que forman una recta continua.

10 La caracterización de la componente periódica corresponde a localizar los picos aislados en la imagen módulo de la TFD.

Estas frecuencias espaciales  $(\hat{\omega}_p, \hat{\nu}_p)$  corresponden a la posición de estos picos:

$$(\hat{\omega}_p, \hat{\nu}_p) = \arg \max_{(\omega, \nu)} \Psi(\omega, \nu) \quad (24)$$

15 Para el cálculo de las amplitudes  $(\hat{C}_p, \hat{D}_p)$ , se utiliza un procedimiento de demodulación como para estimar las amplitudes de la componente direccional.

20 Para cada elemento periódico de frecuencia  $(\hat{\omega}_p, \hat{\nu}_p)$ , la amplitud correspondiente es idéntica a la media de los píxeles de la nueva imagen obtenida multiplicando la imagen  $\{\tilde{y}(i,j)\}$  por  $\cos(i\hat{\omega}_p, j\hat{\nu}_p)$ . Esto se traduce por las siguientes fórmulas:

$$\hat{C}_p = \frac{1}{L \times T} \sum_{n=0}^{L-1} \sum_{m=0}^{T-1} y(n,m) \cos(n\hat{\omega}_p + m\hat{\nu}_p). \quad (25)$$

$$\hat{D}_p = \frac{1}{L \times T} \sum_{n=0}^{L-1} \sum_{m=0}^{T-1} y(n,m) \cos(n\hat{\omega}_p - m\hat{\nu}_p) \quad (26)$$

En resumen, un procedimiento de estimación de la componente periódica comprende las siguientes etapas:

Etapa 1. Localizar los picos aislados en la segunda mitad de la imagen módulo de Fourier y contar su número
Etapa 2. Para cualquier pico detectado: <ul style="list-style-type: none"> <li>• Calcular su frecuencia con la ayuda de la formula (24)</li> <li>• Calcular su amplitud con la ayuda de la formula (25-26)</li> </ul>

30 Las últimas informaciones que van a extraerse están contenidas en la componente puramente aleatoria  $\{w(i,j)\}$ . Esta componente puede representarse por un modelo autorregresivo 2-D de soporte semiplano no simétrico (DPNS) finito definido por la ecuación con las siguientes diferencias:

$$w(i,j) = - \sum_{(k,l) \in S_{N,M}} a_{k,l} w(i-k, j-l) + u(i,j) \quad (27)$$

35 en la que  $\{a_{(k,l)}\}_{(k,l) \in S_{N,M}}$  son los parámetros que van a determinarse para todo  $(k,l)$  que pertenece a  $S_{N,M} = \{(k,l)/k=0, 1 \leq l \leq M\} \cup \{(k,l)/1 \leq k \leq N, -M \leq l \leq M\}$ . El par  $(N,M)$  se denomina el orden del modelo.

40 •  $\{u(i,j)\}$  es un ruido blanco gaussiano de varianza terminada  $\sigma_u^2$ .

Los parámetros del modelo vienen dados por:

$$\mathcal{W} = \{(N, M), \sigma_u^2, \{a_{k,l}\}_{(k,l) \in S_{N,M}}\} \quad (28)$$

Los procedimientos de estimación de los elementos de W son numerosos como por ejemplo el algoritmo de Levinson 2D o los procedimientos de adaptación del tipo mínimos cuadrados (MCR).

5 Ahora se describirá un procedimiento de caracterización del color de una imagen de la que se quieren extraer términos  $t_i$  que representan características icónicas de esta imagen, siendo el color un ejemplo particular de estas características que pueden comprender otras características tales como los instantes algebraicos o geométricos, las propiedades estadísticas, las propiedades espectrales de los instantes de pseudo-Zernicke.

10 El procedimiento se basa en la caracterización perceptiva del color. En primer lugar se efectúa una transformación de las componentes del color de la imagen del espacio RVB (rojo, verde, azul) también denominado RGB, hacia el espacio TSV (tinta saturación valor) también denominado HSV. Se obtienen así tres componentes: tinta, saturación, valor. A partir de estas tres componentes se determinan N colores o componentes icónicas de la imagen. Cada componente icónica  $C_i$  se representa por un vector de M valores. Estos valores representan la distribución angular y anular de los puntos que representan cada componente así que el número de puntos de la componente en cuestión.

El procedimiento desarrollado se ilustra por la figura 19 con, a modo de ejemplo, N =16 y M= 17.

20 En una primera etapa principal 110, a partir de una imagen 11 del espacio RGB, se procede a una transformación de la imagen 111 del espacio (R,G,B) al espacio HSV (etapa 112) para obtener una imagen en el espacio HSV.

El modelo HSV puede definirse de la siguiente manera.

25 Tinta (H): varía de [0 360], y cada ángulo representa un tinta.

Saturación (S): varía de [0 1], mide la pureza de los colores, y permite distinguir los colores “vivos”, “pasteles”, o “descoloridos”.

30 Valor (V): Adopta valores de [0 1], indica si un color es claro u oscuro y en qué medida se acerca al blanco o al negro.

El modelo HSV es una transformación no lineal del modelo del espacio (R,G,B). El ojo humano puede distinguir 128 tintas, 130 saturaciones, y 23 sombras.

35 Para el Blanco V=1 y S=0, el negro presenta una valor V=0 mientras que la tinta H y la saturación S son indeterminadas. Cuando V=1 y S=1 se presenta un color puro.

Cada color se obtiene añadiendo blanco o negro a un color puro.

40 Para tener colores más claros se reduce S y se conservan H y V, sin embargo, para colores oscuros se añade negro reduciendo V y se conservan H y S.

El paso de la imagen en color expresada en las coordenadas (R,G,B) a una imagen expresada en el espacio (H,S,V) (tinta, saturación, valor) se efectúa de la siguiente manera:

45 Para cualquier punto de coordenada (i,j) y de valor ( $R_k, B_k, G_k$ ) se produce un punto de coordenadas (i,j) y valor ( $H_k, S_k, V_k$ ) con:

$$V_k = \max(R_k, B_k, G_k)$$

50

$$S_k = \frac{V_k - \min(R_k, G_k, B_k)}{V_k}$$

$$H_k = \begin{cases} \frac{G_k - B_k}{V_k - \min(R_k, G_k, B_k)} & \text{si } V_k \text{ es igual a } R_k \\ 2 + \frac{B_k - R_k}{V_k - \min(R_k, G_k, B_k)} & \text{si } V_k \text{ es igual a } G_k \\ 4 + \frac{R_k - G_k}{V_k - \min(R_k, G_k, B_k)} & \text{si } V_k \text{ es igual a } B_k \end{cases}$$

Se procede a continuación a una división del espacio HSV (etapa 113).

A partir de los valores de tinta, saturación, valor, se han definido N colores. En el caso en el que N es igual a 16, se presenta: negro, blanco, gris claro, gris oscuro, gris medio, rojo, rosa, naranja, marrón, oliva, amarillo, verde, azul celeste, azul verdoso, azul, púrpura, magenta.

Para cada píxel se evalúa a qué color pertenece. A continuación se calcula el número de puntos de cada color.

En una segunda etapa principal 120, se procede a una caracterización de las divisiones obtenidas durante la primera etapa principal 110.

En esta etapa 120, se busca caracterizar cada división  $C_i$  obtenida anteriormente. Una división se define por su componente icónica y por las coordenadas de los píxeles que la constituyen. La descripción de una división se basa en la caracterización del reparto espacial de estos píxeles (nube de puntos). El procedimiento empieza por el cálculo del centro de gravedad, el eje principal de la nube de puntos y el eje perpendicular a este eje. La nueva marca se utiliza como referencia en la descomposición de la división  $C_i$  en varias subdivisiones que se representan por el porcentaje de los puntos que constituyen cada una de las subdivisiones. El procedimiento de caracterización de una división  $C_i$  es el siguiente:

- calcular el centro de gravedad y el ángulo de orientación de las componentes  $C_i$  que definen la marca de división,
- calcular la distribución angular de los puntos de la división  $C_i$  en las N direcciones en el sentido antihorario, en N subdivisiones definidas por:

$$\left(0^\circ, \frac{360}{N}, \frac{2 \times 360}{N}, \dots, \frac{i \times 360}{N}, \dots, \frac{(N-1) \times 360}{N}\right)$$

- dividir el espacio de la imagen en cuadros de radios concéntricos, con cálculo en cada radio del número de puntos que corresponde a cada componente icónica.

El vector característico se obtiene a partir del número de puntos de cada reparto de color  $C_i$ , del número de puntos en los 08 subrepartos angulares así que del número de puntos de la imagen.

Por tanto, el vector característico se representa por 17 valores en el ejemplo considerado.

En la figura 19, se ha ilustrado la segunda etapa 120 de tratamiento a partir de las componentes icónicas  $C_0$  a  $C_{15}$  mostrando para las componentes  $C_0$  (módulo 121) y  $C_{15}$  (módulo 131) las diferentes etapas efectuadas, a saber la división angular 122, 132 que conduce a un número de puntos en las 8 orientaciones consideradas (etapa 123, 133) y la división anular 124, 134 que conduce a un número de puntos en los 8 radios considerados (etapa 125, 135), así como teniendo en cuenta el número de píxeles de componente  $C_0$ , respectivamente  $C_{15}$ , en la imagen (etapa 126, respectivamente 136).

Las etapas 123, 125, 126 conducen a la producción de 17 valores para el componente  $C_0$  (etapa 127) mientras que las etapas 133, 135, 136 conducen a la producción de 17 valores para la componente  $C_{15}$  (etapa 137).

Naturalmente, el procedimiento es análogo para las otras componentes  $C_1$  a  $C_{14}$ .

Las figuras 20 y 21 ilustran el hecho de que el procedimiento descrito anteriormente es invariable a la rotación.

Así, en el ejemplo de la figura 20, la imagen se divide en dos subconjuntos, uno que contiene las cruces x, otro los círculos 0. Después del cálculo del centro de gravedad así que del ángulo de orientación  $\theta$ , se obtiene la marca de orientación que permitirá de obtener los 04 subrepartos angulares ( $0^\circ, 90^\circ, 180^\circ, 270^\circ$ ).

A continuación, se efectúa un reparto anular, se calcula el número de puntos en un radio igual a 1 y luego 2. Se obtiene el vector  $V_0$  característico de la imagen de la figura 20: 19; 6; 5; 4; 4; 8; 11.

La imagen de la figura 21 se obtiene aplicando una rotación de  $90^\circ$  a la imagen de la figura 20. Al aplicar el procedimiento anterior a la imagen de la figura 21, se obtiene un vector  $V_1$  que caracteriza esta última que muestra que la rotación no influye en el vector característico. Esto permite concluir que el procedimiento es invariable a la rotación.

Tal como se ha indicado anteriormente, los procedimientos que permiten obtener para una imagen los términos que representan los colores dominantes, las propiedades texturales o las estructuras de las zonas dominantes de la

imagen, pueden aplicarse tanto en la totalidad de la imagen como en partes de la imagen.

A continuación se describirá brevemente el procedimiento de segmentación de un documento que permite producir las partes de la imagen que van a caracterizarse.

5 Según una primera técnica posible, se procede a una descomposición estática. La imagen se descompone en bloques con recubrimiento o sin recubrimiento.

10 Según una segunda técnica posible, se procede a una descomposición dinámica. En este caso, la descomposición de la imagen en partes es función del contenido de la imagen.

15 Según un primer ejemplo de técnica de descomposición dinámica, las partes se producen a partir de las semillas que son los puntos de singularidad de la imagen (los puntos de inflexiones). Se empieza por calcular las semillas, que se fusionan a continuación para que permanezca sólo un número reducido y por último los puntos de las imágenes se fusionan con las semillas que presentan las mismas propiedades visuales (estadísticas) para producir las partes o los segmentos de la imagen que van a caracterizarse.

20 Según otra técnica que se refiere a una segmentación jerárquica, los puntos de la imagen se fusionan para formar las  $n$  primeras clases. Luego se descomponen los puntos de cada una de las clases en  $m$  clases y así sucesivamente hasta alcanzar el número de clases deseado. Durante la fusión, los puntos se asignan a la clase más próxima. Una clase se representa por el centro de gravedad y/o un delimitador (caja envolvente, segmento, curva,...).

25 Ahora se describirán las etapas principales de un procedimiento de caracterización de las formas de una imagen.

La caracterización de la forma se efectúa en varias etapas:

30 Para una supresión de efecto de zoom o variación debida a los movimientos de los elementos no rígidos de la imagen (movimiento de los labios, de las hojas de árbol,...), se procede por una multirresolución seguida de una decimación de la imagen.

Para una reducción del efecto de traslación, la imagen o la parte de la imagen se representa por su Transformada de Fourier.

35 Para una reducción del efecto de zoom, la imagen se define en el espacio logarítmico polar.

Pueden ponerse en práctica las siguientes etapas:

40 a/ multirresolución  $f = \text{ondícula}(l,n)$ ; en la que  $l$  es la imagen de salida y  $n$  es el número de descomposiciones

b/ proyección de la imagen en el espacio logPolar:

$$g(l,m) = f(i,j) \text{ con } i = l \cdot \cos(m) \text{ y } j = l \cdot \sin(m)$$

45 c/ cálculo de la transformada de Fourier de  $g$ :  $H = \text{FFT}(g)$ ;

d/ caracterización de  $H$ :

50 d1/ proyección de  $H$  en varias direcciones (0, 45, 90,...): el resultado es un conjunto de vectores cuya dimensión es igual a la dimensión del segmento de proyección

d2/ cálculo de las propiedades estadísticas de cada vector de proyección (media, varianza, los instantes).

55 El término que representa la forma está constituido por los valores de las propiedades estadísticas de cada vector de proyección.

**REIVINDICACIONES**

1. Procedimiento de indexación de documentos multimedia, que comprende por lo menos las siguientes etapas:

- 5 (a) identificar y extraer para cada documento términos  $t_i$  constituidos por vectores que caracterizan propiedades del documento multimedia que debe indexarse, tales como la forma, la textura, el color o la estructura de una imagen, la energía, la tasa de oscilación o informaciones frecuenciales de una señal de audio, o un grupo de caracteres de un texto,
- 10 (b) almacenar los términos  $t_i$  que caracterizan propiedades del documento multimedia en una base de términos (3) que comprende P términos, caracterizado porque comprende además las siguientes etapas:
- (c) determinar un número máximo N de conceptos deseados que reagrupan los términos  $t_i$  pertinentes, siendo N un número entero inferior a P, y estando previsto cada concepto  $C_i$  para reagrupar todos los términos próximos desde el punto de vista de sus características,
- 15 (d) calcular la matriz T de distancias entre los términos  $t_i$  de la base de términos,
- (e) descomponer el conjunto P de términos  $t_i$  de la base de términos en N partes  $P_j$  ( $1 \leq j \leq N$ ) tales que  $P = P_1 \cup P_2 \dots \cup P_N$ , comprendiendo cada parte  $P_j$  un conjunto de términos  $t_{ij}$  y estando representada por un concepto  $C_j$ , estando los términos  $t_i$  repartidos en la base de la matriz T, de tal manera que los términos más alejados unos de otros se encuentran en partes  $P_i, P_m$  distintas y los términos próximos unos de otros se encuentran en la misma parte  $P_i$ ,
- 20 (f) estructurar el diccionario de conceptos (5) de manera que constituye un árbol binario en el que cada hoja del árbol binario contiene un concepto  $C_i$  del diccionario y cada nudo del árbol contiene las informaciones necesarias para el escrutinio del árbol durante una etapa de identificación de un documento multimedia mediante comparación con los documentos indexados anteriormente, y
- 25 (g) construir una base de huellas (25) constituida por el conjunto de los conceptos  $C_i$  que representan los términos  $t_i$  de los documentos que van a indexarse, estando cada documento asociado a una huella que le es propia.

2. Procedimiento de indexación según la reivindicación 1, caracterizado porque se asocia a cada concepto  $C_i$  de la base de huellas (25) un conjunto de informaciones que comprende el número NbT de términos en los documentos en los que el concepto  $C_i$  está presente.

3. Procedimiento de indexación según la reivindicación 1 ó 2, caracterizado porque para cada documento en el que está presente un concepto  $c_i$ , se registra una huella del concepto  $C_i$  en el documento, comprendiendo esta huella la frecuencia de aparición del concepto  $C_i$ , la identificación de los conceptos que están próximos al concepto  $C_i$  en el documento y una puntuación que es un valor medio de las mediciones de similitudes entre el concepto  $C_i$  y los términos  $t_i$  del documento que son los más próximos al concepto  $C_i$ .

4. Procedimiento de indexación según cualquiera de las reivindicaciones 1 a 3, caracterizado porque comprende una etapa de optimización de la división del conjunto P de los términos de la base de términos para descomponer este conjunto P en M clases  $C_i$  ( $1 \leq i \leq M$ , siendo  $M \leq P$ ), de manera que reduce el error del reparto del conjunto P de los términos de la base de términos en N partes ( $P_1, P_2, \dots, P_N$ ) en el que cada parte  $P_i$  se representa por el término  $t_i$

que se tomará como concepto  $C_i$ , siendo el error cometido  $\epsilon$  tal que 
$$\epsilon = \sum_{i=1}^N \epsilon_{t_i}$$
 en la que

$$\epsilon_{t_i} = \sum_{t_j \in P_i} d^2(t_i, t_j)$$

es el error cometido cuando se sustituyen los términos  $t_j$  de una parte  $P_i$  por  $t_i$ .

5. Procedimiento de indexación según la reivindicación 4, caracterizado porque comprende las siguientes etapas:

- 50 (i) descomponer el conjunto P de términos en dos partes  $P_1$  y  $P_2$ ;
- 55 (ii) determinar los dos términos más alejados  $t_i$  y  $t_j$  del conjunto P que corresponde a la distancia más grande  $D_{ij}$  de la matriz T de distancias;
- (iii) para cada término  $t_k$  del conjunto P, examinar si la distancia  $D_{ki}$  entre el término  $t_k$  y el término  $t_i$  es más pequeña que la distancia  $D_{kj}$  entre el término  $t_k$  y el término  $t_j$ , si es así, asignar el término  $t_k$  a la parte  $P_1$  y si no es así, asignar el término  $t_k$  a la parte  $P_2$ ;
- 60 (iv) repetir la etapa (i) hasta obtener el número N de partes  $P_i$  deseado y en cada iteración se aplican las etapas (ii) y

(iii) en los términos de las partes  $P_1$  y  $P_2$ .

6. Procedimiento de indexación según la reivindicación 4 ó 5, caracterizado porque comprende una optimización a partir de las  $N$  partes desunidas  $\{P_1, P_2, \dots, P_N\}$  del conjunto  $P$  así como de los  $N$  términos  $\{t_1, t_2, t_N\}$  que las representan para reducir el error de descomposición del conjunto  $P$  en  $N$  partes, y porque comprende las siguientes etapas:

(i) calcular los centros de gravedad  $C_i$  de las partes  $P_i$

10 (ii) calcular los errores  $\epsilon C_i = \sum_{t_j \in P_i} d^2(C_i, t_j)$  y  $\epsilon t_i = \sum_{t_j \in P_i} d^2(t_i, t_j)$  cuando se sustituyen los términos  $t_j$  de la parte  $P_i$  respectivamente por  $C_i$  y por  $t_i$ ,

(iii) comparar  $\epsilon t_i$  y  $\epsilon C_i$  y sustitución de  $t_i$  por  $C_i$  si  $\epsilon C_i \leq \epsilon t_i$ ,

15 (iv) calcular la nueva matriz  $T$  de distancias entre los términos  $t_i$  de la base de términos y procedimiento de descomposición del conjunto  $P$  de los términos de la base de términos (3) en  $N$  partes, salvo si se cumple una

condición de parada con  $\frac{\epsilon C_i - \epsilon C_{i+1}}{\epsilon C_i} < \text{umbral}$ , en la que  $\epsilon C_i$  representa el error cometido en el instante  $t$ .

7. Procedimiento de indexación según cualquiera de las reivindicaciones 1 a 6, caracterizado porque para efectuar una estructuración del diccionario de conceptos (5), se produce de manera iterativa en cada iteración una carta de navegación empezando por dividir el conjunto de los conceptos en dos subconjuntos, y luego seleccionando un subconjunto en cada iteración hasta la obtención del número de grupos deseado o hasta que se satisface un criterio de parada.

8. Procedimiento de indexación según la reivindicación 7, caracterizado porque el criterio de parada está constituido por el hecho de que los subconjuntos obtenidos son todos homogéneos con una pequeña desviación estándar.

9. Procedimiento de indexación según la reivindicación 7 u 8, caracterizado porque durante la estructuración del diccionario de conceptos (5), se determinan indicadores de navegación a partir de una matriz  $M = [C_1, C_2, \dots, C_N] \in \mathcal{R}^{p \times N}$  del conjunto  $C$  de los conceptos  $C_i \in \mathcal{R}^p$  en la que  $C_i$  representa un concepto de  $p$  valores, según las siguientes etapas:

(i) calcular un representante  $w$  de la matriz  $M$

35 (ii) calcular la matriz de covarianza  $\tilde{M}$  entre los elementos de la matriz  $M$  y el representante  $w$  de la matriz  $M$ ,

(iii) calcular un eje de proyección  $u$  de los elementos de la matriz  $M$ ,

40 (iv) calcular el valor  $p_i = d(u, c_i) - d(u, w)$  y descomponer el conjunto de conceptos  $C$  en dos subconjuntos  $C_1$  y  $C_2$  de la siguiente manera:

$$\begin{cases} c_i \in C_1 & \text{si } p_i \leq 0 \\ c_i \in C_2 & \text{si } p_i > 0 \end{cases}$$

45 (v) almacenar en el nudo asociado a  $C$  las informaciones  $\{u, w, |p_1|, p_2\}$  en la que  $p_1$  es el máximo de todos los  $p_i \leq 0$  y  $p_2$  es el mínimo de todos los  $p_i > 0$ , constituyendo el conjunto de las informaciones  $\{u, w, |p_1|, p_2\}$  los indicadores de navegación en el diccionario de conceptos (5).

10. Procedimiento de indexación según cualquiera de las reivindicaciones 1 a 9, caracterizado porque se analizan a la vez las componentes estructurales y el complemento de estas componentes estructurales constituido por las componentes textuales de una imagen del documento, y porque:

(a) durante el análisis de las componentes estructurales de la imagen

55 (a1) se procede a un reparto de las zonas fronteras de las estructuras de la imagen en diferentes clases según la orientación de la variación local de intensidad de manera que define los elementos de soporte

estructural (ESS) de la imagen, y

(a2) se procede por análisis estadístico a la construcción de términos constituidos por vectores que describen las propiedades locales y globales de los elementos de soporte estructurales,

(b) durante el análisis de las componentes texturales de la imagen

(b1) se procede a una detección y una caracterización paramétrica de una componente puramente aleatoria de la imagen,

(b2) se procede a una detección y una caracterización paramétrica de una componente periódica de la imagen,

(b3) se procede a una detección y una caracterización paramétrica de una componente direccional de la imagen,

(c) se reagrupa en un número limitado de conceptos el conjunto de los elementos descriptivos de la imagen constituidos por una parte, por los términos que describen las propiedades locales y globales de los elementos de soporte estructurales y por otra parte, por los parámetros de las caracterizaciones paramétricas de las componentes aleatoria, periódica y direccional que definen las componentes texturales de la imagen, y

(d) se define para cada documento una huella a partir de las apariciones, de las posiciones y de las frecuencias de dichos conceptos.

11. Procedimiento de indexación según la reivindicación 10, caracterizado porque las propiedades locales de los elementos de soporte estructurales tenidas en cuenta para la construcción de términos comprenden por lo menos el tipo de soporte seleccionado de entre una banda lineal o un arco de curva, las dimensiones de longitud y anchura del soporte, la dirección principal del soporte y la forma y las propiedades estadísticas de los píxeles que constituyen el soporte.

12. Procedimiento de indexación según la reivindicación 10 u 11, caracterizado porque las propiedades globales de los elementos de soporte estructurales tenidas en cuenta para la construcción de términos comprenden por lo menos el número de cada tipo de soportes y su disposición espacial.

13. Procedimiento de indexación según cualquiera de las reivindicaciones 10 a 12, caracterizado porque durante el análisis de las componentes estructurales de la imagen se procede a una prueba previa de detección de la presencia de por lo menos una estructura en la imagen y, en caso de ausencia de estructura, se pasa directamente a la etapa del análisis de las componentes texturales de la imagen.

14. Procedimiento de indexación según cualquiera de las reivindicaciones 10 a 13, caracterizado porque para proceder a un reparto de las zonas fronteras de las estructuras de la imagen en diferentes clases, a partir de la imagen digitalizada definida por el conjunto de los píxeles  $y(i,j)$  en la que  $(i,j) \in I \times J$ , designando  $I$  y  $J$  respectivamente el número de líneas y el número de columnas de la imagen, se calcula la imagen gradiente vertical  $g_v(i,j)$  con  $(i,j) \in I \times J$  y la imagen gradiente horizontal  $g_h$  con  $(i,j) \in I \times J$  y se procede a la división de la imagen según la orientación local de su gradiente en un número finito de clases equidistantes, definiéndose la imagen que contiene la orientación del gradiente por la fórmula

$$O(i,j) = \arctan \left[ \frac{g_h(i,j)}{g_v(i,j)} \right],$$

se identifican las clases que constituyen regiones de soporte susceptibles de contener elementos de soporte significativos, y a partir de las regiones de soporte, se determinan los elementos de soporte significativos y se catalogan según criterios predeterminados.

15. Procedimiento de indexación según cualquiera de las reivindicaciones 1 a 9, caracterizado porque durante la indexación de un documento multimedia que comprende señales de vídeo, se eligen términos  $t_i$  constituidos por imágenes clave que representan grupos de imágenes homogéneas consecutivas, y se determinen conceptos  $C_i$  mediante reagrupamiento de términos  $t_i$ .

16. Procedimiento de indexación según la reivindicación 15, caracterizado porque para determinar imágenes clave que constituyen términos  $t_i$ , se elabora en primer lugar un vector de puntuación VS que comprende un conjunto de elementos  $VS(i)$  que materializan la diferencia o la similitud entre el contenido de una imagen de índice  $i$  y el de una imagen de índice  $i-1$ , y se analiza el vector de puntuación VS con el fin de determinar las imágenes clave que corresponden a los máximos de los valores de los elementos  $VS(i)$  del vector de puntuación VS.

- 5 17. Procedimiento de indexación según la reivindicación 16, caracterizado porque una imagen de índice  $j$  se considera como una imagen clave si el valor  $VS(j)$  del elemento correspondiente del vector de puntuación  $VS$  es un máximo y el valor  $VS(j)$  se sitúa entre dos mínimos  $\min G$  y  $\min D$  y el mínimo  $M1$  tal que  $M1 = (|VS_{(j)} - \min G|, |VS_{(j)} - \min D|)$  es superior a un umbral dado.
- 10 18. Procedimiento de indexación según cualquiera de las reivindicaciones 1 a 9, caracterizado porque durante la indexación de un documento multimedia que comprende componentes de audio, se toman muestras y se descompone el documento en tramas, que se reagrupan a continuación en clips de los que cada uno se caracteriza por un término  $t_i$  constituido por un vector de parámetro.
- 15 19. Procedimiento de indexación según la reivindicación 18, caracterizado porque una trama comprende entre aproximadamente 512 y 2048 muestras del documento de audio del que se han tomado muestras.
- 20 20. Procedimiento de indexación según la reivindicación 18 ó 19, caracterizado porque los parámetros tenidos en cuenta para la definición de los términos  $t_i$  comprenden informaciones temporales que corresponden a por lo menos uno de los siguientes parámetros: la energía de las tramas de la señal de audio, la desviación estándar de las energías de las tramas en los clips, la relación de las variaciones sonoras, la relación de baja energía, la tasa de oscilación alrededor de un valor predeterminado, la alta tasa de oscilación alrededor de un valor predeterminado, la diferencia entre el número de tasa de oscilación por encima y por debajo de la tasa de oscilación media de las tramas de clips, la varianza de la tasa de oscilación, la relación de las tramas silenciosas.
- 25 21. Procedimiento de indexación según cualquiera de las reivindicaciones 18 a 20, caracterizado porque los parámetros tenidos en cuenta para la definición de los términos  $t_i$  comprenden informaciones frecuenciales que corresponden a por lo menos uno de los siguientes parámetros: el centro de gravedad del espectro de frecuencia de la transformada de Fourier corta de la señal de audio, la anchura de banda de la señal de audio, la relación entre la energía en una banda de frecuencia y la energía total en toda la banda de frecuencia de la señal de audio de la que se han tomado muestras, el valor medio de la variación del espectro de dos tramas adyacentes en un clip, la frecuencia de corte de un clip.
- 30 22. Procedimiento de indexación según cualquiera de las reivindicaciones 18 a 21, caracterizado porque los parámetros tenidos en cuenta para la definición de los términos  $t_i$  comprenden por lo menos la modulación de energía a 4 Hz.
- 35 23. Procedimiento de indexación según cualquiera de las reivindicaciones 1 a 14, caracterizado porque se analizan las formas de una imagen de un documento según las siguientes etapas:
- 40 (a) proceder a una multirresolución seguida por una decimación de la imagen,
- (b) definir la imagen en el espacio logarítmico polar,
- (c) representar la imagen o la porción de la imagen afectada por su transformada de Fourier  $H$ ,
- 45 (d) proceder a una caracterización de la transformada de Fourier  $H$  de la siguiente manera:
- (d1) se proyecta  $H$  en varias direcciones para obtener un conjunto de vectores cuya dimensión es igual a la dimensión del movimiento de proyección,
- 50 (d2) se calculan las propiedades estadísticas de cada vector de proyección, y
- (e) se representa la forma de la imagen por un término  $t_i$  constituido por los valores de las propiedades estadísticas de cada vector de proyección.

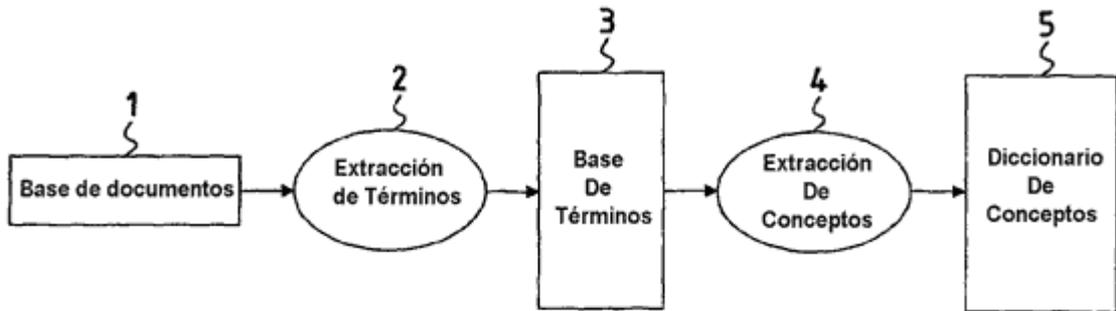


FIG.1

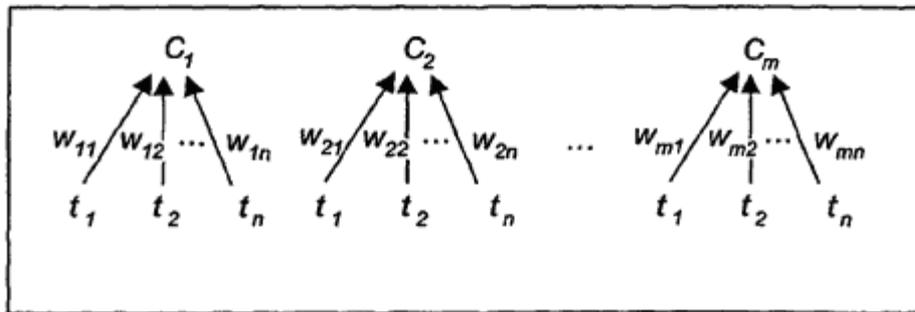


FIG.2

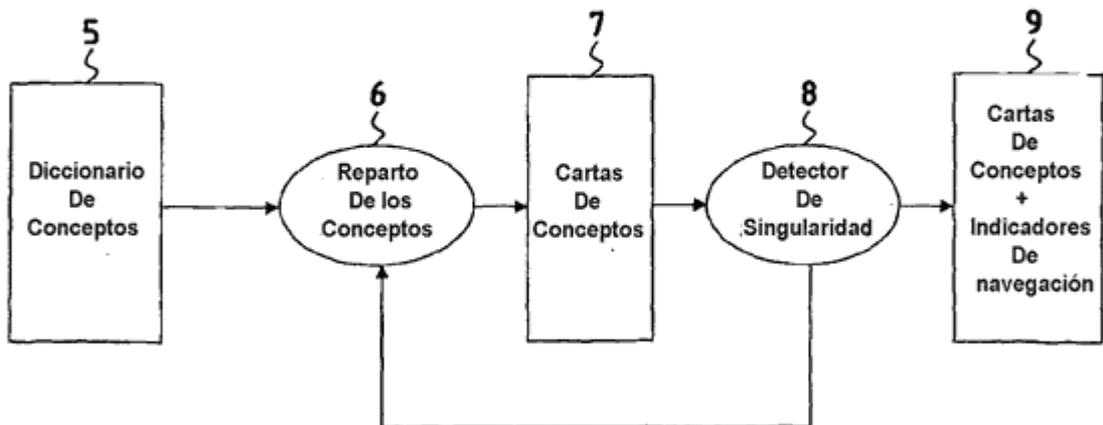


FIG.3

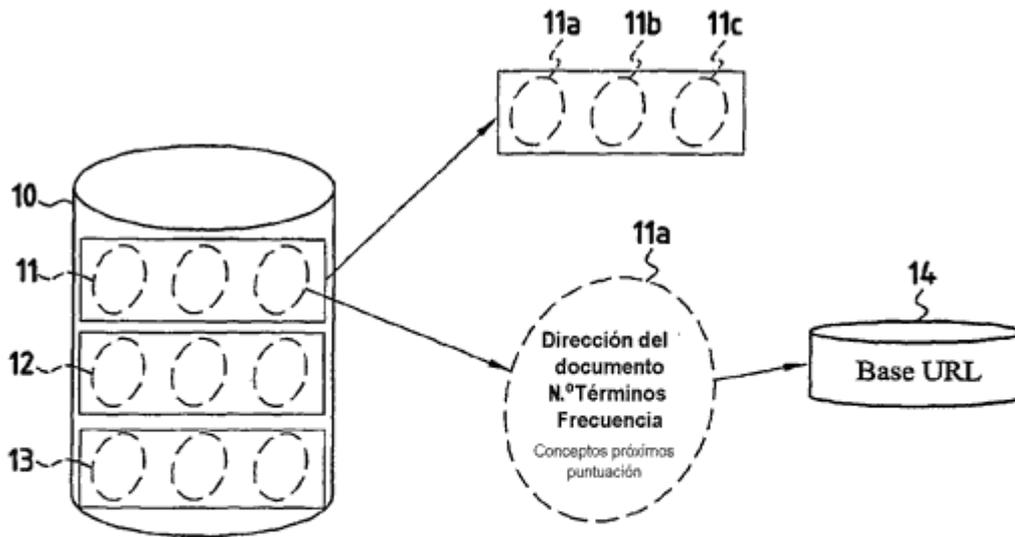


FIG.4

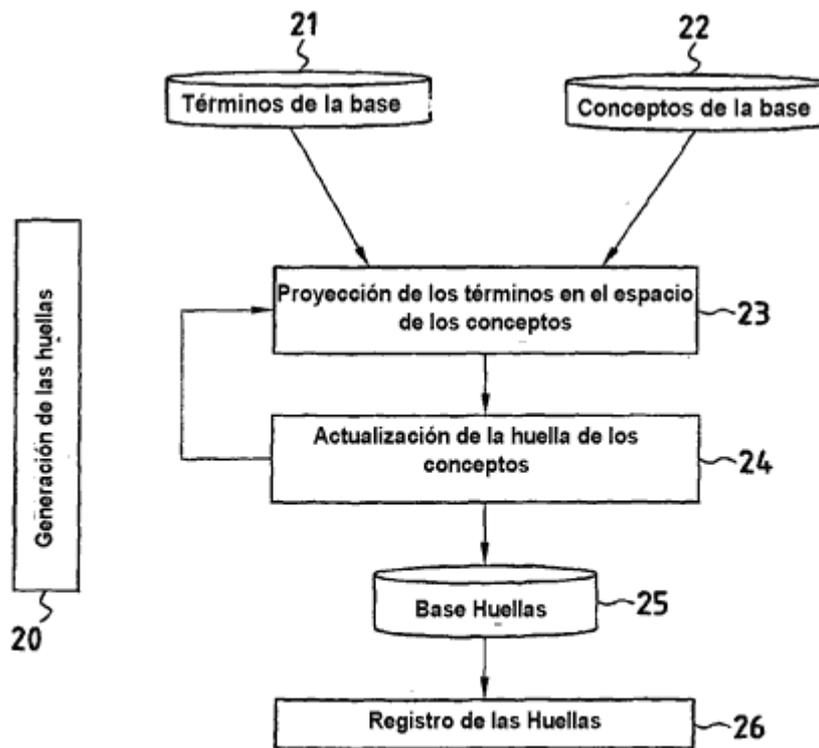


FIG.5

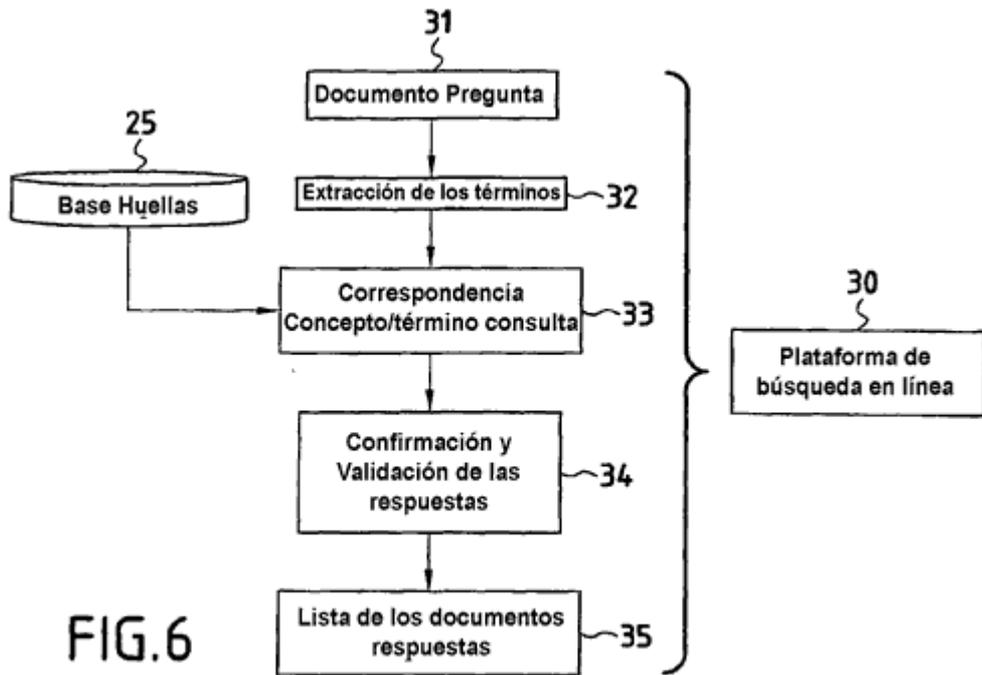


FIG. 6

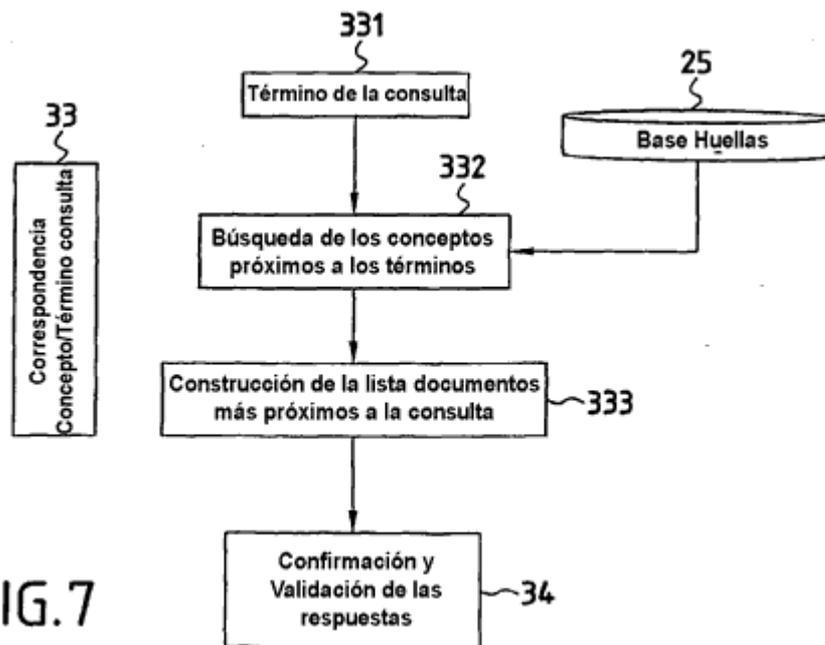


FIG. 7

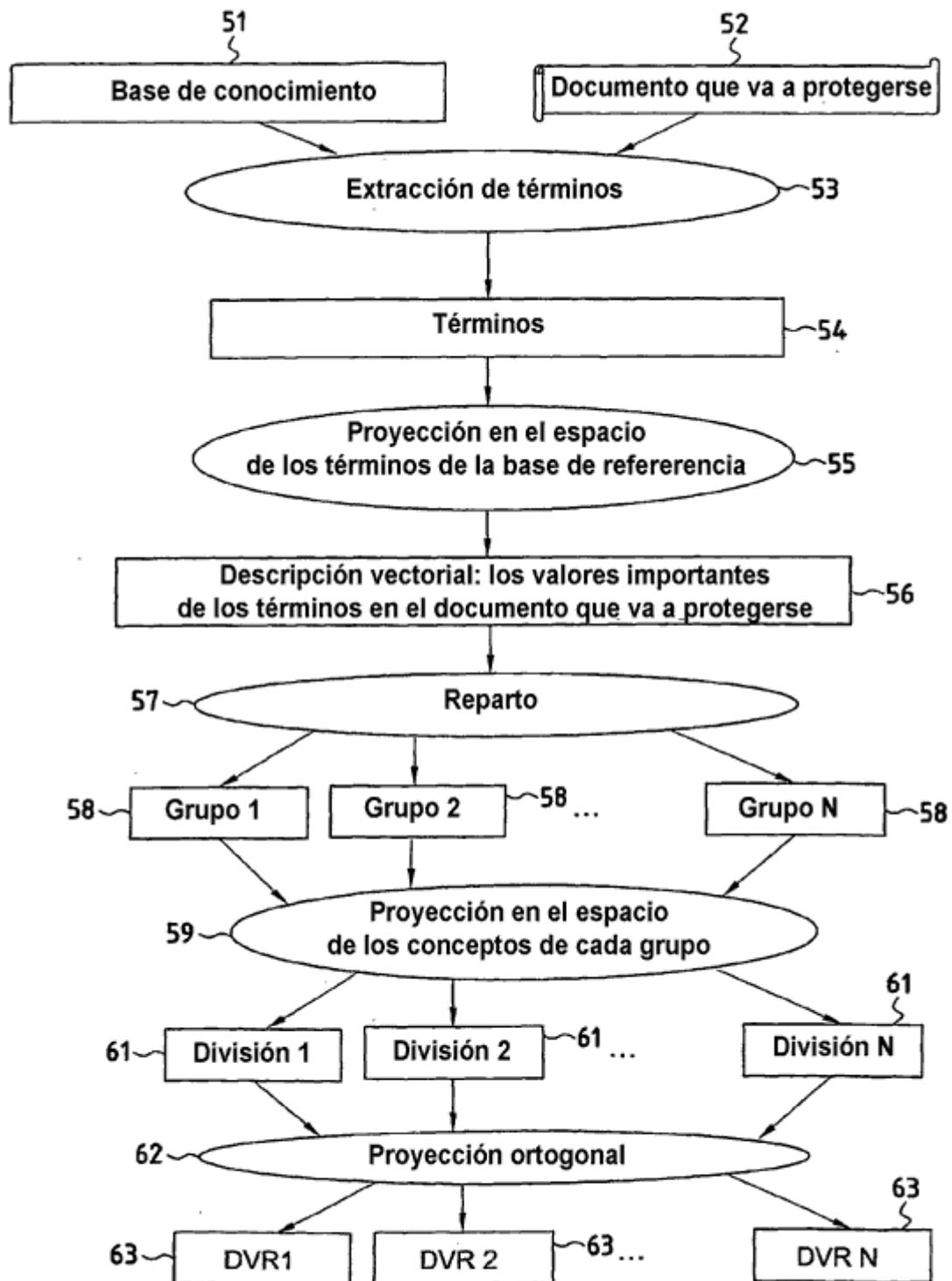
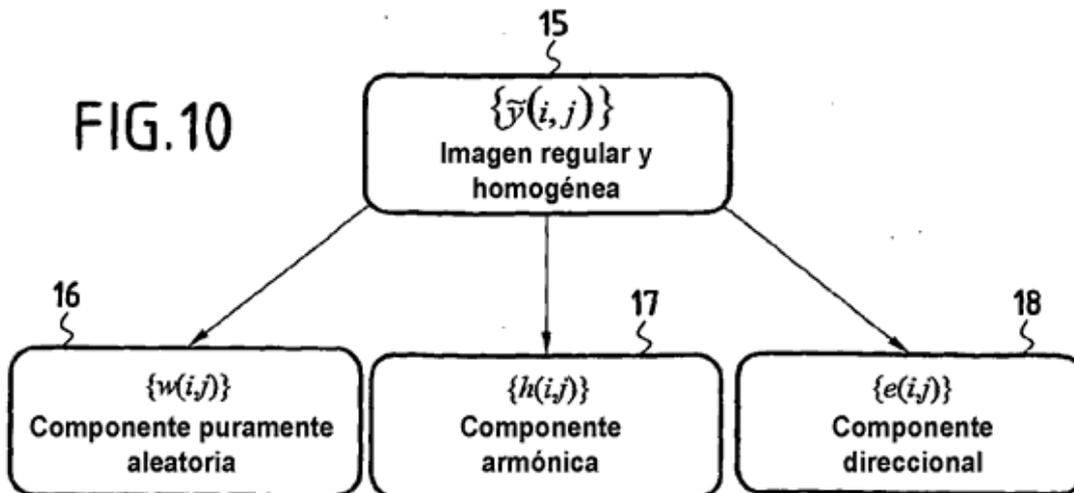
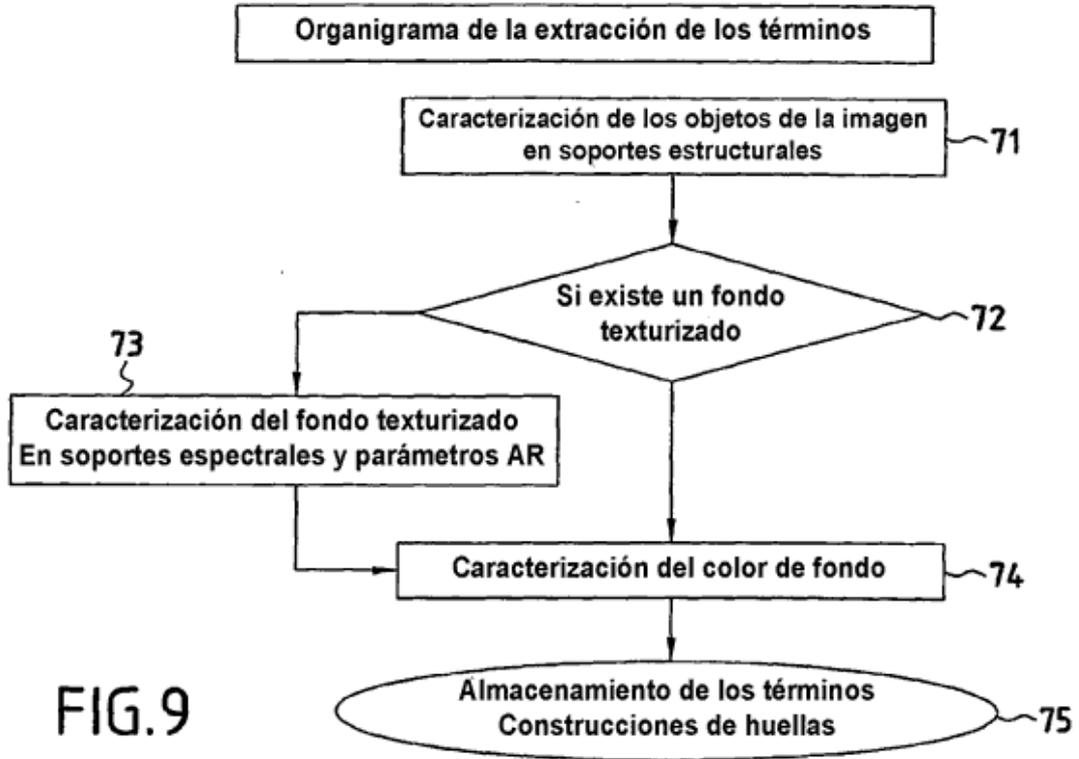


FIG.8

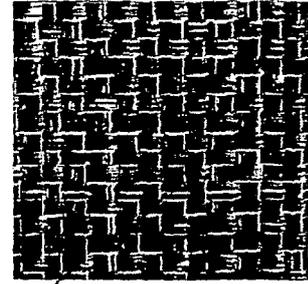




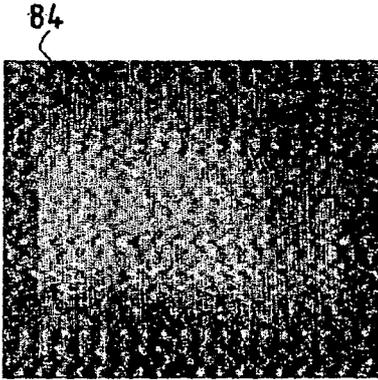
81 FIG. 11



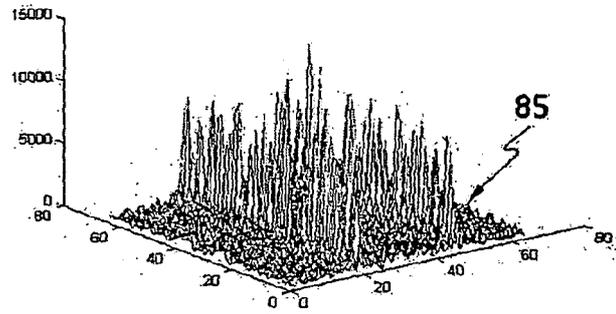
82 FIG. 12



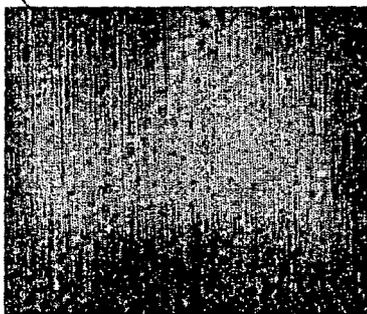
83 FIG. 13



84 FIG. 15a



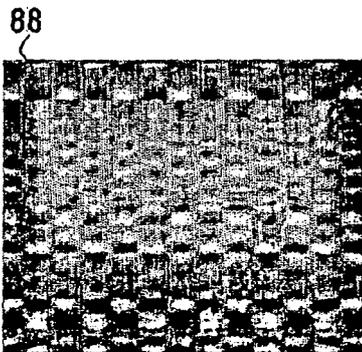
85 FIG. 15a1



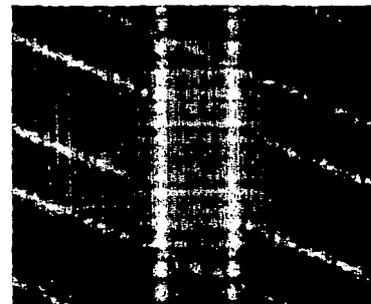
86 FIG. 15b



87 FIG. 15b1



88 FIG. 15c



89 FIG. 15c1

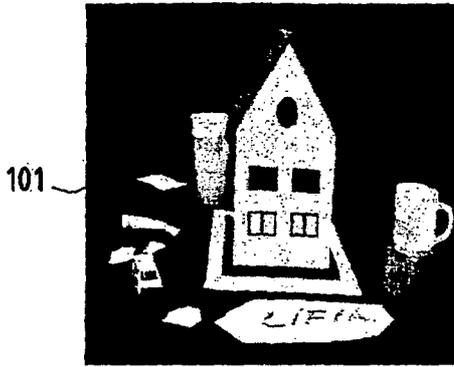


FIG. 14a

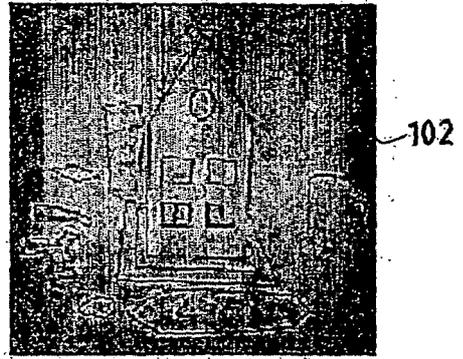


FIG. 14b

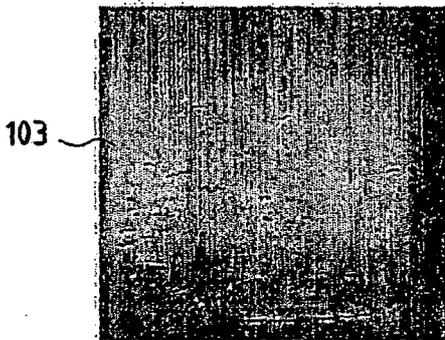


FIG. 14c

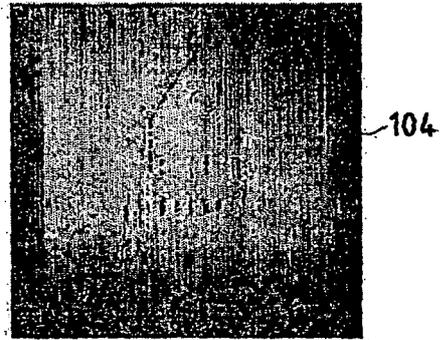


FIG. 14d

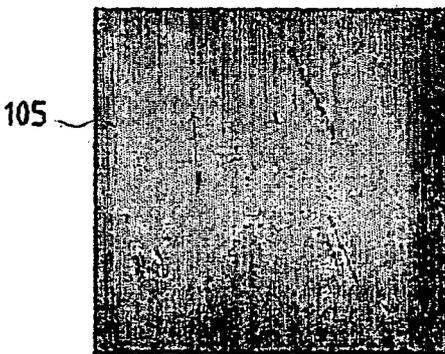


FIG. 14e

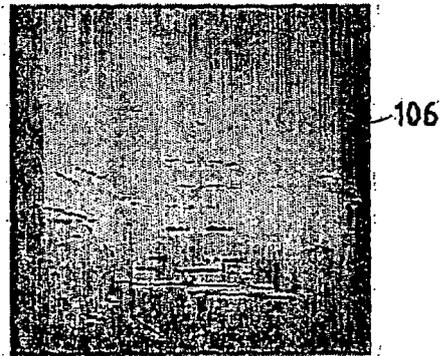


FIG. 14f

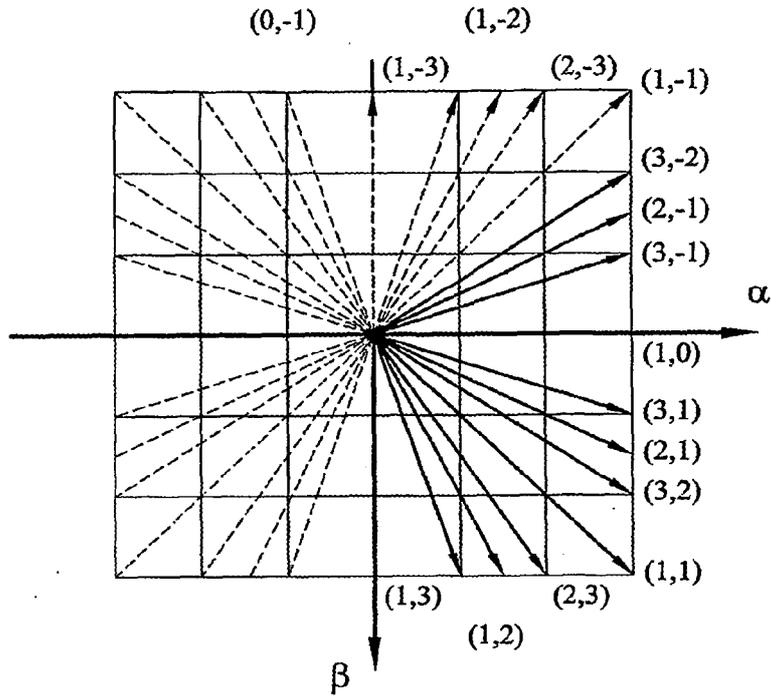


FIG.16

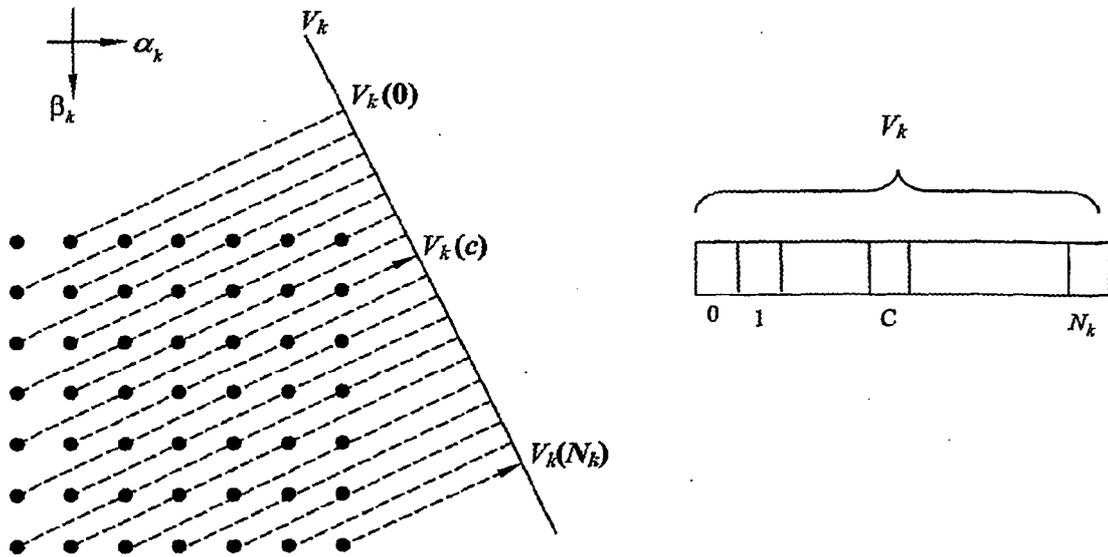


FIG.17

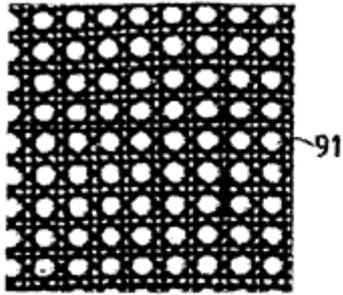


FIG.18a1

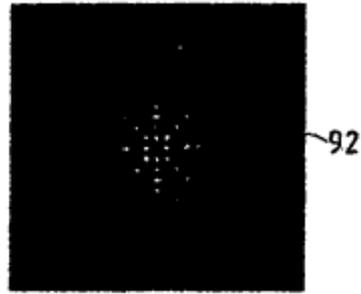


FIG.18a2

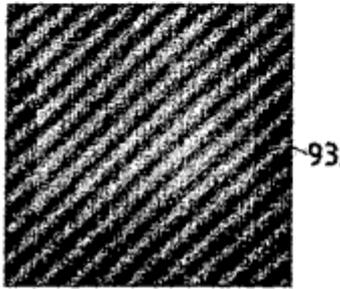


FIG.18b1

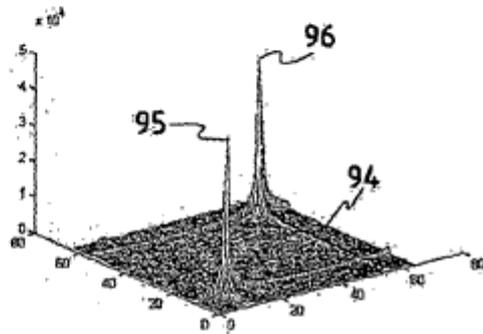
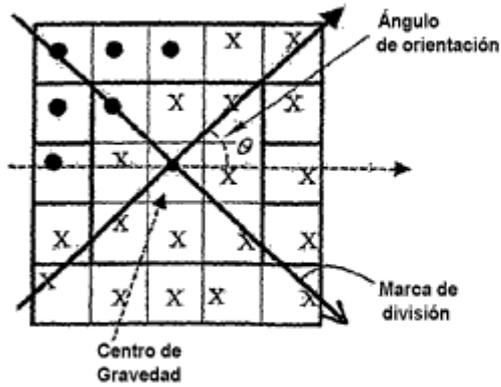


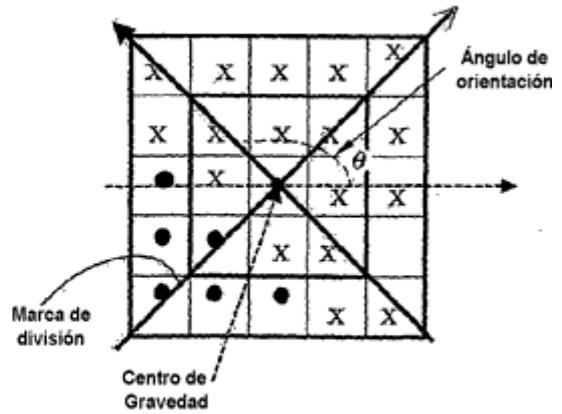
FIG.18b2



19	6	5	4	4	8	11
----	---	---	---	---	---	----

V0

FIG.20



19	6	5	4	4	8	11
----	---	---	---	---	---	----

VI

FIG.21

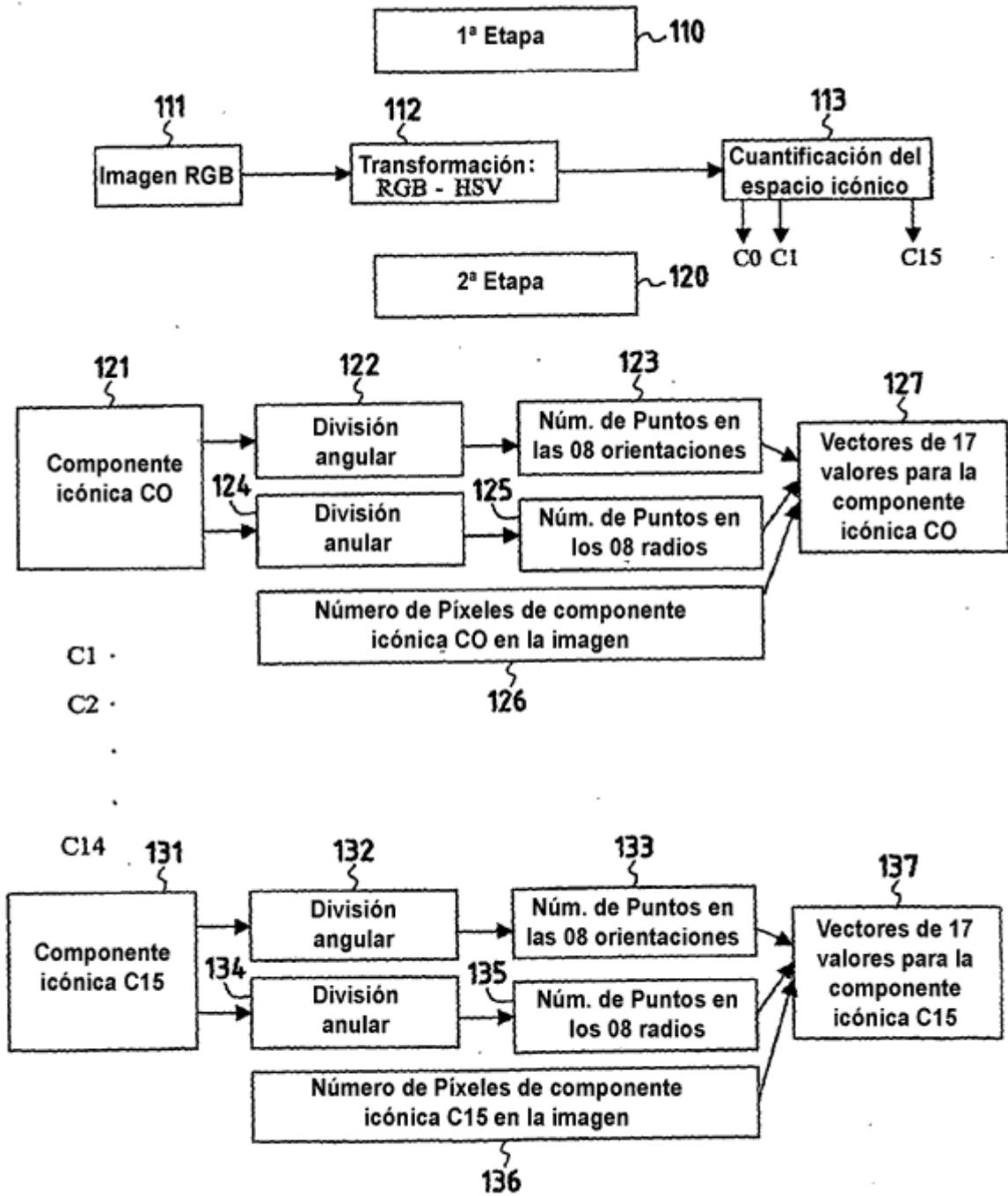


FIG.19

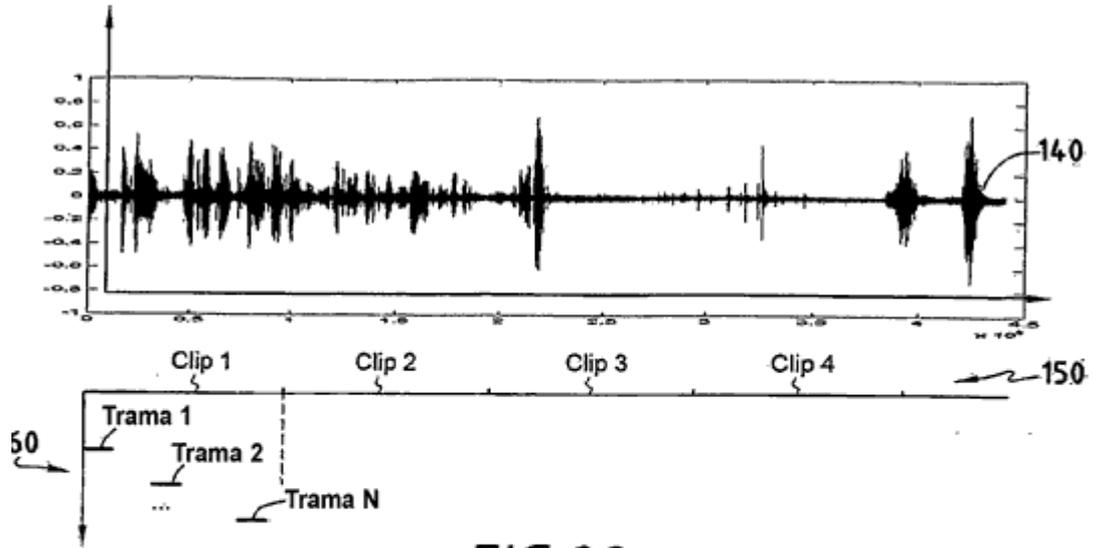


FIG.22

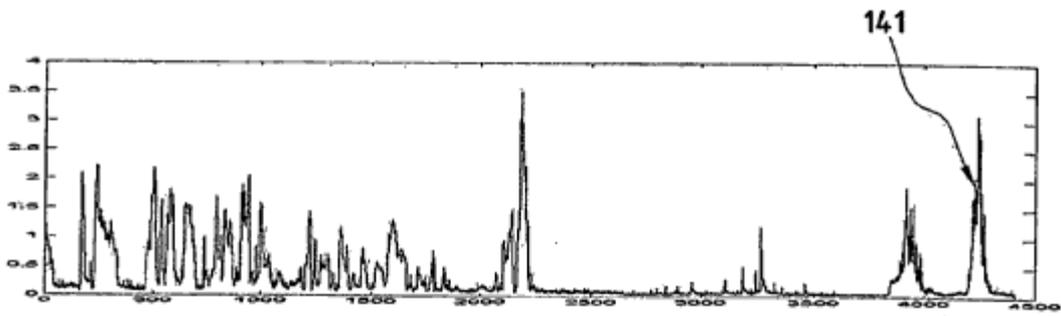


FIG.23a

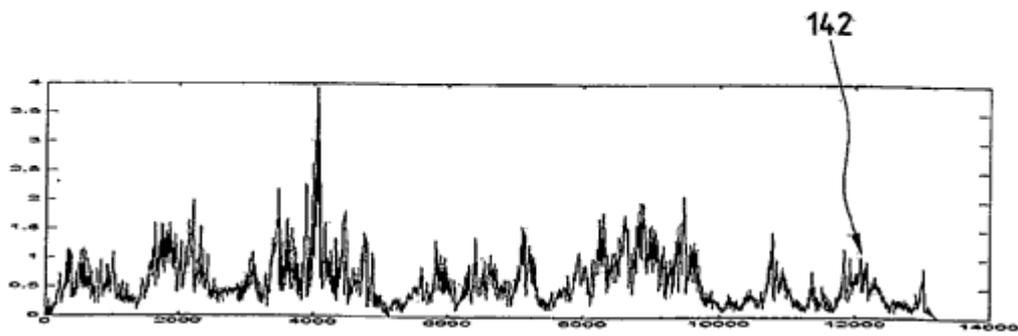


FIG.23b

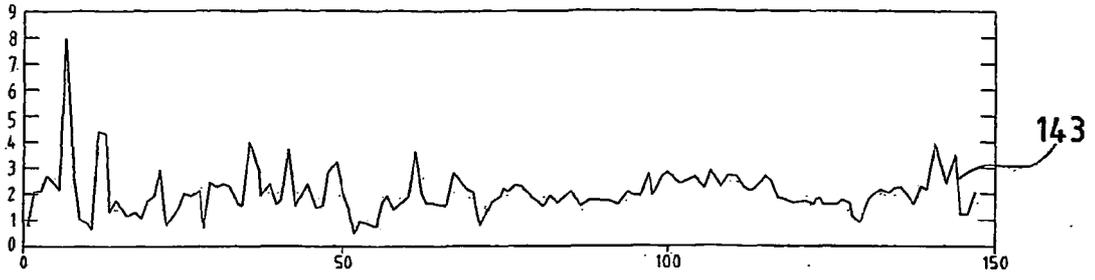


FIG.24a

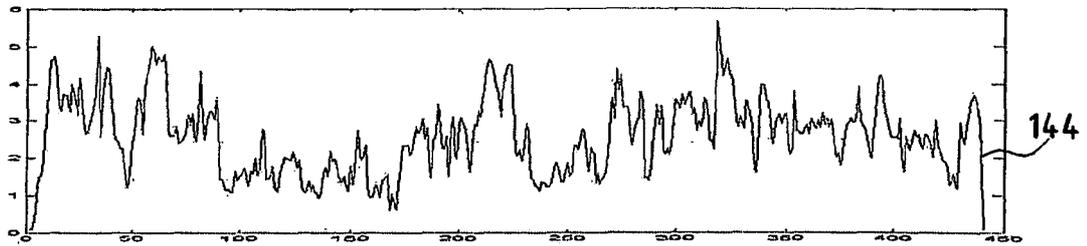


FIG.24b

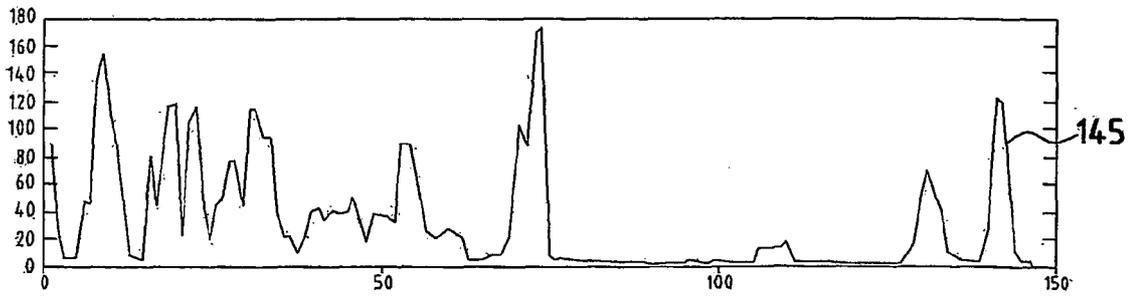


FIG.25a

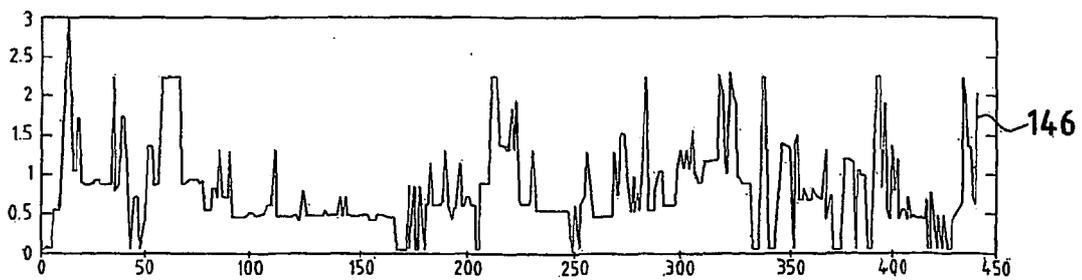


FIG.25b

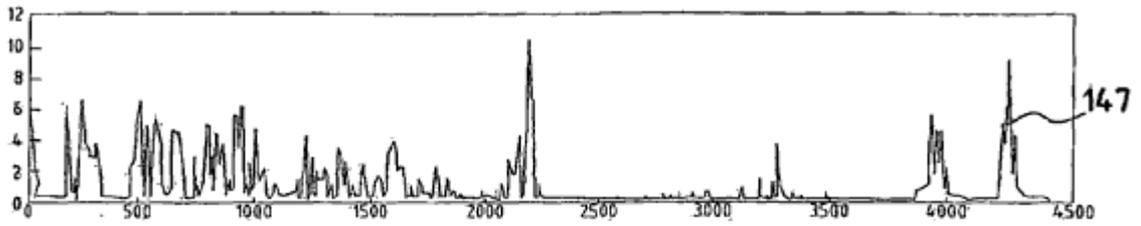


FIG.26a

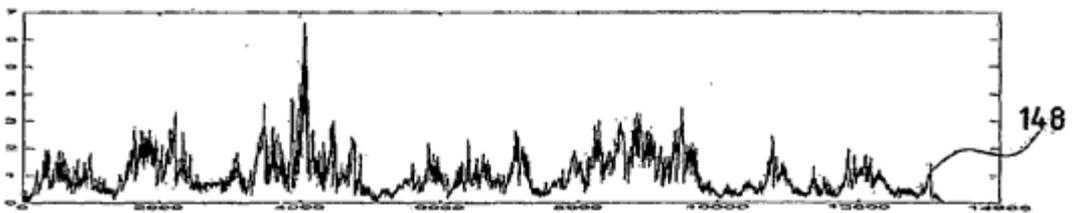


FIG.26b

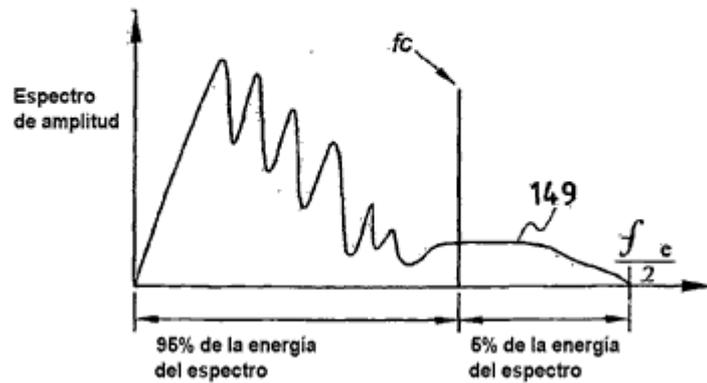


FIG.29

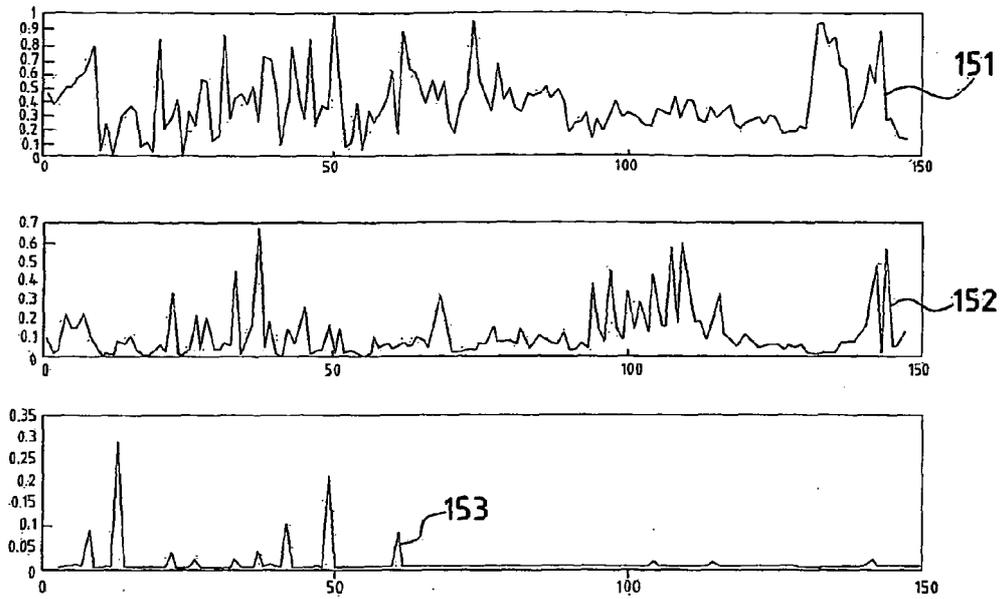


FIG.27a

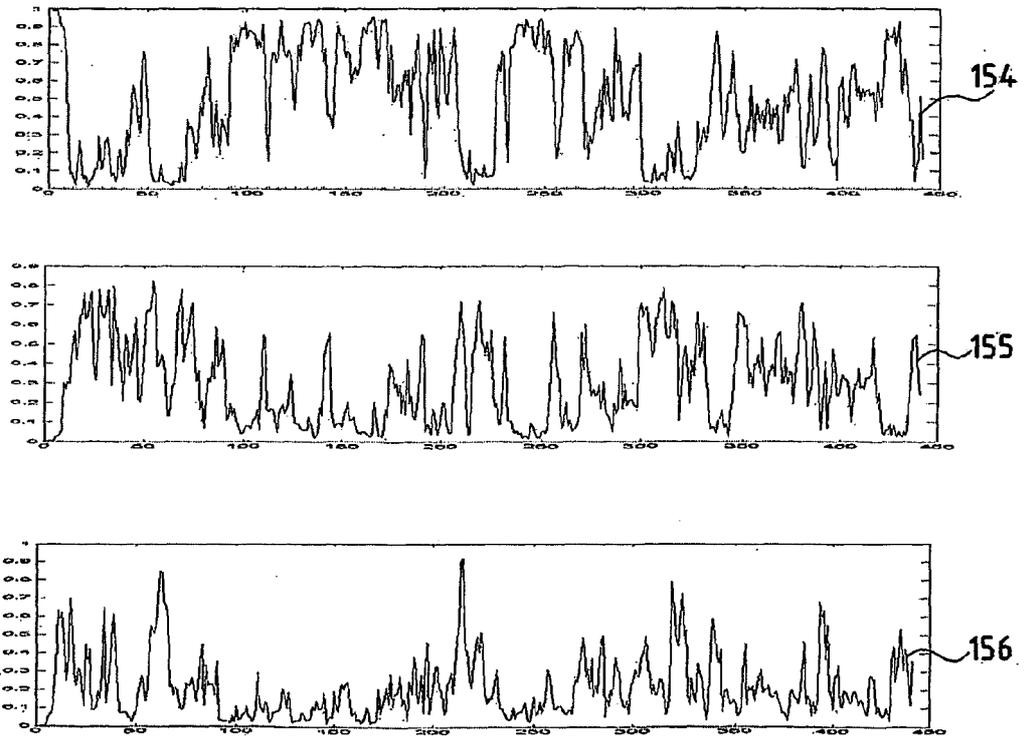


FIG.27b

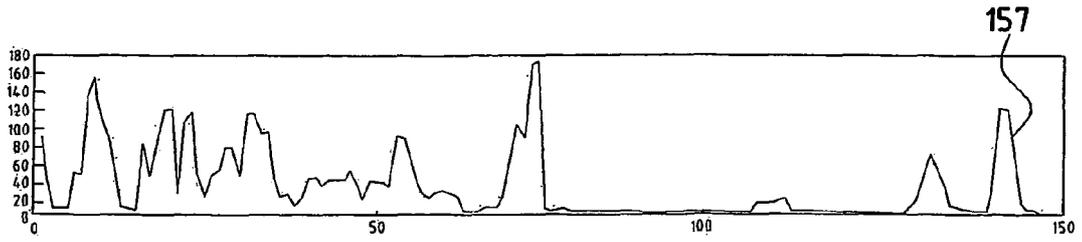


FIG.28a

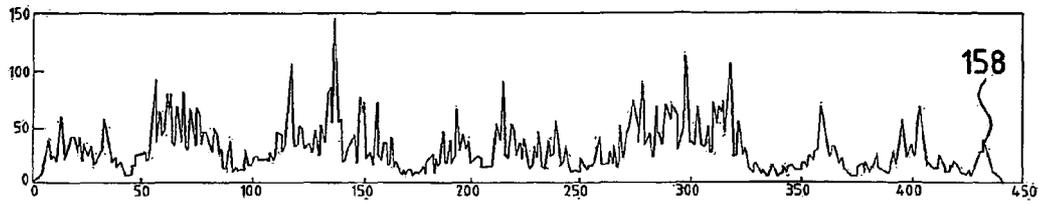


FIG.28b

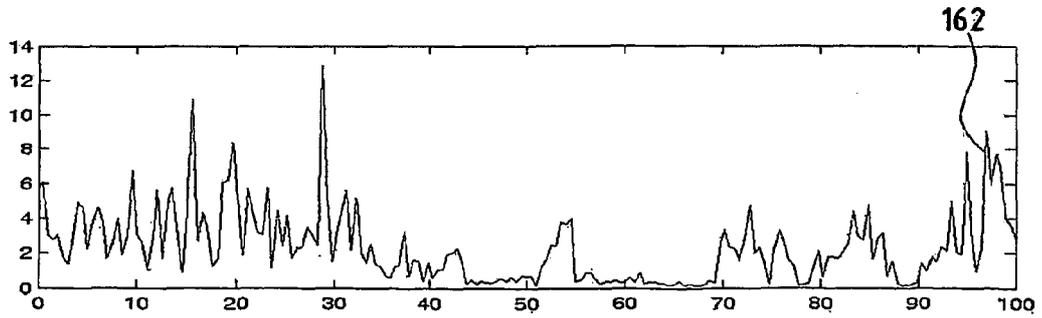
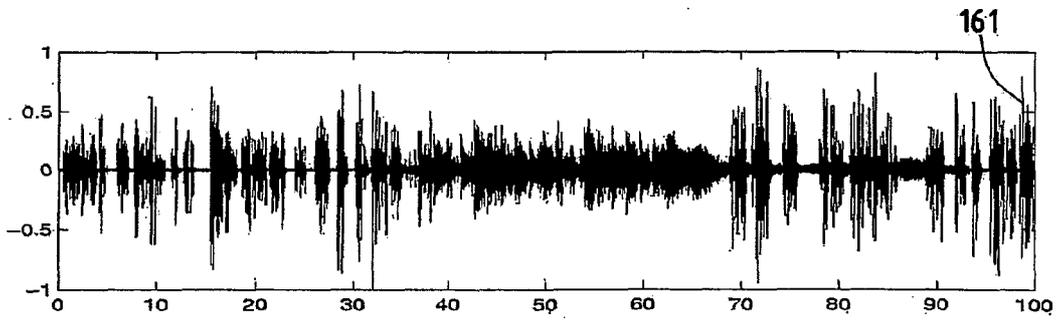


FIG.30