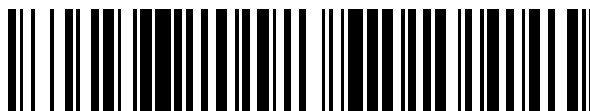


19



OFICINA ESPAÑOLA DE  
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 524 789**

51 Int. Cl.:

**G06F 17/30** (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

96 Fecha de presentación y número de la solicitud europea: **20.05.2009 E 09251342 (3)**

97 Fecha y número de publicación de la concesión europea: **03.09.2014 EP 2124159**

54 Título: **Aprendizaje de imagen, anotación automática, método de recuperación y dispositivo**

30 Prioridad:

**20.05.2008 CN 200810100721**

45 Fecha de publicación y mención en BOPI de la traducción de la patente:

**12.12.2014**

73 Titular/es:

**RICOH COMPANY, LIMITED (100.0%)  
3-6 NAKAMAGOME 1-CHOME OHTA-KU  
TOKYO 143-8555, JP**

72 Inventor/es:

**BILLOEUL, TIMOTHEE;  
ZHU, CAIZHI y  
XU, YINGHUI**

74 Agente/Representante:

**SUGRAÑES MOLINÉ, Pedro**

**ES 2 524 789 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**

Aprendizaje de imagen, anotación automática, método de recuperación y dispositivo

5 ANTECEDENTES DE LA INVENCION

1. Campo de la invención

10 La presente invención se refiere a tecnologías de procesamiento de imagen y, en particular, a aprendizaje de imagen, anotación automática, un método de recuperación y un dispositivo.

2. Descripción de la técnica relacionada

15 Con el rápido desarrollo de las redes y el uso extendido de los teléfonos móviles y cámaras digitales de uso doméstico que tienen sensores de imagen digital, se generan grandes cantidades de recursos de imágenes digitales.

20 Para manejar grandes cantidades de datos de imagen, se demandan sistemas de recuperación de imagen eficaces y prácticos. Por lo tanto, en un campo de recuperación de imagen basado en contenido (CBIR), se ha realizado un estudio basado en un contenido para extraer un contenido del significado de la palabra definitivo a partir de una imagen para acceder y recuperar imágenes de usuario.

25 El estudio anterior está basado en la recuperación en la que la CBIR usa similitud de imagen en una etapa inicial. En otras palabras, cuando se introducen imágenes, colores o vistas esquemáticas por el usuario, los resultados de la recuperación son también imágenes similares a las imágenes, los colores o las vistas esquemáticas de entrada. Sin embargo, una CBIR de este tipo apenas es práctica. Los siguientes dos puntos se consideran como razones principales.

30 En primer lugar, se requiere que el usuario tenga una imagen o se requiere que tenga una capacidad para seleccionar apropiadamente colores o describir vistas esquemáticas. Estos requisitos representan una barrera para el usuario, limitando por lo tanto el uso extendido del sistema anterior.

35 Adicionalmente, la recuperación de imagen basada en similitud de imagen depende de comparación en similitud entre imágenes de capa inferior. Sin embargo, hay una diferencia en significados de palabra entre las características de la capa inferior y la información de imagen de una capa superior. Por lo tanto, como resultado de la recuperación, es probable que las imágenes de capa inferior sean enormemente diferentes en significados de palabra aunque sean visualmente similares. Esto da como resultado un enorme impacto en la precisión en el resultado de la recuperación.

40 Para resolver los problemas anteriores en la CBIR, los investigadores han propuesto un sistema de recuperación basado en anotación de imagen. En este sistema de recuperación, las anotaciones se realizan en información de texto para datos de imagen para realizar recuperación de imagen basada en textos. Puesto que este método permite al usuario realizar recuperación únicamente con la entrada de una palabra clave, se reducen los requisitos anteriores sobre la capacidad del usuario.

45 Como anotación automática de imagen actual se conocen los siguientes métodos.

Método 1:

50 Se realiza anotación automática basada en datos originales (tal como tiempo, información de GPS, un nombre de imagen asociado con una imagen tomada mediante una cámara digital, un texto asociado con una imagen en un formato digital, etc.).

Método 2:

55 Después de la estimación de una asociación complicada entre un carácter y una imagen usando tecnología visual informática y tecnología de aprendizaje por máquina en una etapa de aprendizaje, se realiza anotación automática en imágenes que no aparecen en la etapa de aprendizaje basada en la asociación.

60 Sin embargo, el método 1 anterior basado en los datos originales tiene los siguientes problemas.

En otras palabras, los datos originales de una imagen pueden no asociarse con un contenido de imagen. Por lo tanto, la calidad de la anotación de imagen es pobre.

65 Puesto que el método anterior se aplica únicamente a una imagen con un texto, su alcance de aplicación es enormemente limitado.

El método 1 tiene un defecto inevitable. Por lo tanto, el método 2 se ha propuesto como una modificación del método 1. Los detalles del método 2 son como sigue.

El método 2 incluye las siguientes etapas.

Etapa A:

Una imagen se segmenta en regiones con un método de segmentación de región, y se calculan los vectores característicos de las respectivas regiones.

Etapa B:

En la etapa de aprendizaje, las respectivas regiones y las k regiones más cercanas a las respectivas regiones se enlazan entre sí, y la imagen y las anotaciones reales relacionadas se enlazan entre sí.

Etapa C:

En una etapa de anotación automática, se buscan todos los gráficos creados con un recorrido aleatorio con reinicio (RWR) para obtener anotaciones correspondientes.

El método anterior se denomina específicamente como "GCap: leyendas de imagen automáticas basadas en gráficos" en 20 Proc. Of the 4<sup>o</sup> International Workshop on Multimedia Data and Document Engineering (MDDE), junto con Computer Vision Pattern Recognition Conference (CVPR' 04), 2004 por J. Y. Pan, H. J. Yany, C. Faloutsos y P. Duygulu.

El algoritmo GCap está basado teóricamente en el hecho de que un tiempo de acceso al nodo de una imagen anotada (imagen medida) es mayor que un tiempo de acceso a otro nodo. Por lo tanto, confirmando el tiempo de acceso al nodo anotado, es posible encontrar la anotación que tiene la correlación más fuerte.

Sin embargo, existe una probabilidad de producir regiones de imagen que se enlazan erróneamente entre sí en el gráfico obtenido en la etapa de aprendizaje mediante el método anterior, que da como resultado precisión pobre en la anotación.

### SUMARIO DE LA INVENCION

La presente invención puede tener un objeto de proporcionar una anotación automática y un método de recuperación.

De acuerdo con la presente invención, se proporciona un método de anotación automática de imagen que comprende las características de la reivindicación 1.

Se definen realizaciones adicionales de la invención mediante las reivindicaciones dependientes.

Otros objetos, características y ventajas de la presente invención se harán más evidentes a partir de la siguiente descripción detallada cuando se lee junto con los dibujos adjuntos.

### BREVE DESCRIPCION DE LOS DIBUJOS

La Figura 1 es un diagrama que muestra la configuración de un dispositivo de aprendizaje de imagen de acuerdo con una realización de la presente invención;

La Figura 2 es un diagrama que muestra un estado en el que se segmenta una imagen original en diversos bloques rectangulares en la realización de la presente invención;

La Figura 3 es un diagrama que muestra un estado en el que se segmenta una imagen original en sub-bloques de árbol cuádruple de multi-resolución en la realización de la presente invención;

La Figura 4 es un diagrama que muestra un estado en el que se segmenta una imagen original en diversas regiones homogéneas en la realización de la presente invención;

La Figura 5 es un diagrama que muestra un ejemplo de una matriz característica de texto Y en la realización de la presente invención;

Las Figuras 6 a 9 son diagramas de flujo que muestran la implementación de un algoritmo para establecer enlaces y determinar los pesos de los enlaces en la realización de la presente invención;

La Figura 10 es un gráfico obtenido mediante el aprendizaje de imagen de acuerdo con la realización de la presente invención;

5 La Figura 11 es una matriz gráfica obtenida mediante el dispositivo de aprendizaje de imagen de acuerdo con la realización de la presente invención;

La Figura 12 es una matriz gráfica actualizada en la realización de la presente invención;

10 La Figura 13 es un diagrama que muestra un vector de reinicio de acuerdo con la realización de la presente invención;

La Figura 14 es una imagen de entrada para realizar anotaciones automáticas en la realización de la presente invención;

15 La Figura 15 es una ilustración que muestra resultados de recuperación obtenidos cuando "una persona y un animal" son textos de entrada en la realización de la presente invención;

20 La Figura 16 es una ilustración que muestra resultados de recuperación obtenidos cuando se introduce una imagen de la realización de la presente invención; y

La Figura 17 es una ilustración que muestra resultados de recuperación cuando se introducen un texto y una imagen en la realización de la presente invención.

DESCRIPCIÓN DETALLADA DE LAS REALIZACIONES PREFERIDAS

25 En el aprendizaje de imagen, la anotación automática, un método de recuperación y un dispositivo de acuerdo con una presente invención, los vectores característicos de imagen se proyectan en un sub-espacio de covarianza canónica en una etapa de aprendizaje para maximizar la covarianza entre una imagen y anotaciones, haciendo de esta manera posible crear un gráfico de manera más robusta, para reducir regiones de imagen que se enlazan erróneamente entre sí y finalmente realizar excelentes anotaciones de imagen.

30 Como se muestra en la Figura 1, un dispositivo de aprendizaje de imagen de acuerdo con la realización de la presente invención incluye un primer módulo de almacenamiento, un módulo de segmentación de imagen, un módulo de extracción de vector característico, un módulo de proyección de sub-espacio, un segundo módulo de almacenamiento y un módulo de creación de gráfico. El primer módulo de almacenamiento almacena una imagen que tiene anotaciones y un grupo de palabras clave de anotación, en el que la imagen tiene una o más anotaciones obtenidas a partir del grupo de palabras clave y que describen el contenido del significado de la palabra de la imagen. El módulo de segmentación de imagen realiza una operación de segmentación en la imagen almacenada en el primer módulo para segmentarla en una o más regiones de imagen. El módulo de extracción de vector característico extrae los vectores característicos de imagen de las regiones de imagen a partir de características de imagen y los vectores característicos de texto de las regiones de imagen a partir de características de texto, extrayendo de esta manera los vectores característicos que eliminan información y ruido innecesarios para la imagen y captan un contenido esencial. El módulo de proyección de sub-espacio proyecta los vectores característicos de imagen de las regiones de imagen en el sub-espacio para obtener vectores de proyección característicos de imagen y los vectores característicos de texto de las regiones de imagen en el sub-espacio para obtener vectores de proyección característicos de texto, maximizando de esta manera la covarianza entre las características de imagen y las características de texto. El segundo módulo de almacenamiento almacena por separado los vectores de proyección característicos de imagen y los vectores de proyección característicos de texto de las regiones de imagen. El módulo de creación de gráfico establece un enlace entre las regiones de imagen basándose en los vectores de proyección característicos de imagen, establece enlaces entre la imagen y las regiones de imagen basándose en la operación del módulo de segmentación de imagen, establece enlaces entre la imagen y las anotaciones de la imagen basándose en la imagen que tiene las anotaciones, establece enlaces entre las anotaciones basándose en los vectores de proyección característicos de texto, calcula todos los pesos de estos enlaces y obtiene un gráfico que muestra una relación triangular entre la primera imagen, las regiones de imagen y las anotaciones basándose en los enlaces y los pesos de los enlaces.

35 Adicionalmente, un método de aprendizaje de imagen de acuerdo con la realización de la presente invención incluye una primera etapa de almacenamiento, una etapa de segmentación de imagen, una etapa de extracción de vector característico, una etapa de proyección de sub-espacio, una segunda etapa de almacenamiento y una etapa de creación de gráfico. La primera etapa de almacenamiento incluye almacenar una imagen que tiene anotaciones y un grupo de palabras clave de anotación, en la que la imagen tiene una o más anotaciones obtenidas a partir del grupo de palabras clave de anotación y que describen el contenido del significado de la palabra de la imagen. La etapa de segmentación de imagen incluye realizar una operación de segmentación en la imagen almacenada para segmentarla en una o más regiones de imagen. La etapa de extracción de vector característico incluye extraer los vectores característicos de imagen de las regiones de imagen a partir de características de imagen y los vectores característicos de texto de las regiones de imagen a partir de características de texto, extrayendo de esta manera los

5 vectores característicos que eliminan información y ruido innecesarios para la imagen y captan un contenido esencial. La etapa de proyección de sub-espacio incluye proyectar los vectores característicos de imagen de las regiones de imagen en un sub-espacio para obtener vectores de proyección característicos de imagen y los vectores característicos de texto de las regiones de imagen en el sub-espacio para obtener vectores de proyección característicos de texto, maximizando de esta manera la covarianza entre las características de imagen y las características de texto. El segundo almacenamiento incluye almacenar por separado los vectores de proyección característicos de imagen y los vectores de proyección característicos de texto de las regiones de imagen. La etapa de creación de gráfico incluye establecer enlaces entre las regiones de imagen basándose en los vectores de proyección característicos de imagen, establecer enlaces entre la imagen y las regiones de imagen basándose en la operación en la etapa de segmentación de imagen, establecer enlaces entre la imagen y las anotaciones de la imagen basándose en la imagen que tiene las anotaciones, establecer enlaces entre las anotaciones basándose en los vectores de proyección característicos de texto, calcular todos los pesos de estos enlaces y obtener un gráfico que muestre una relación triangular entre la primera imagen, las regiones de imagen y las anotaciones basándose en los enlaces y los pesos de los enlaces.

15 El gráfico obtenido mediante el módulo de creación de gráfico se usa como una región para realizar anotaciones automáticamente en una imagen, una región para realizar recuperación de imagen, etc., que se describe en detalle a continuación.

20 A continuación, se realiza una descripción específicamente de las implementaciones de los respectivos módulos del dispositivo de aprendizaje de imagen y el método de aprendizaje de imagen.

25 El módulo de segmentación de imagen realiza la operación de segmentación en una imagen almacenada en el primer módulo de almacenamiento para segmentarla en una o más regiones de imagen. Específicamente, de acuerdo con la realización de la presente invención, la imagen puede segmentarse en una o dos o diversas regiones de imagen. En la siguiente descripción, se usan dos o más regiones de imagen.

30 Específicamente, en la realización de la presente invención, la segmentación de una imagen se realiza mediante los siguientes métodos.

Método 1:

35 Como se muestra en la Figura 2, una imagen original (la imagen que está almacenada en el primer módulo de almacenamiento y tiene anotaciones) se segmenta en diversos bloques rectangulares.

Método 2:

40 Como se muestra en la Figura 3, una imagen original se segmenta en sub-bloques de árbol cuádruple de multi-resolución.

Método 3:

45 Como se muestra en la Figura 4, una imagen original se segmenta en diversas regiones homogéneas no solapantes con un algoritmo de segmentación de imagen.

En la realización de la presente invención, el método de segmentación de imagen no está limitado a los métodos anteriores, sino que pueden emplearse también otras técnicas de segmentación de imagen convencionales.

50 Después de la segmentación de una imagen, se proporcionan las características de imagen que corresponden a las regiones de imagen para la imagen. El módulo de extracción de vector característico extrae vectores característicos de imagen y los vectores característicos de texto de toda la imagen original.

55 Todos los algoritmos convencionales se aplican también a la extracción de los vectores característicos de imagen de las regiones de imagen. En la realización de la presente invención, se emplea una característica LBP (Patrón Binario Local) que consiste de colores mezclados e información de patrón en consideración de la capacidad y la velocidad para extraer características de imagen importantes. Obsérvese que el LBP es una característica basada en un patrón, que es excelente en cálculo y contraste con colores mezclados.

60 Después de la extracción de los vectores característicos de imagen de todas las regiones de imagen (en este punto, la longitud de los vectores característicos de imagen se supone que es  $p$  por razones de conveniencia), los vectores característicos de imagen de las regiones de imagen se consideran como una cadena de vector. Por lo tanto, se obtiene una matriz característica de imagen  $X$ .

65 La extracción de los vectores característicos de texto incluye las siguientes etapas.

Etapa A1:

5 Se recoge una anotación que se almacena en el primer módulo de almacenamiento y que pertenece a un grupo de palabras clave asociadas con el contenido de una imagen. Todas las palabras clave se obtienen a partir de un grupo predeterminado cuyo tamaño es  $q$ .

Hay dos tipos de patrones de anotación como sigue.

Anotación reducida:

10 En este caso, la palabra clave de una imagen está asociada con toda la imagen. La imagen puede segmentarse en diversos bloques, pero un grupo de las palabras clave corresponde a toda la imagen.

Anotación local:

15 Se proporciona una palabra clave correspondiente para cada una de las regiones de imagen de una imagen.

En el caso de la anotación reducida, todas las regiones de imagen aceptan una anotación en toda la imagen.

20 Etapa A2:

Después de la recogida de las anotaciones, los vectores característicos de texto (longitud:  $q$ ) se extraen a partir de las regiones de imagen, y los vectores característicos de texto extraídos de las regiones de imagen se consideran como una cadena de vector. Por lo tanto, se obtiene una matriz característica de texto  $Y$ .

25 La Figura 5 muestra un ejemplo de la matriz característica de texto  $Y$ . En la Figura 5, los vectores característicos de texto de las regiones de imagen son  $q$ -dimensionales. Los elementos que corresponden a una palabra clave de imagen representan "1", mientras que los elementos restantes representan "0". En esta realización, la primera, la tercera y la quinta palabras clave en una tabla de palabra cuyo tamaño de grupo es  $q$  ( $q = 8$ ) constituyen una primera región.

30 Después de que se obtienen la matriz característica de imagen  $X$  y la matriz característica de texto  $Y$ , se proyecta el sub-espacio. Sin embargo, antes de la descripción detallada de la proyección de sub-espacio, se describen los efectos de funcionamiento de la proyección de sub-espacio.

35 La proyección de sub-espacio tiene dos objetos.

Un objeto es para reducir las dimensiones de la matriz característica de imagen y de la matriz característica de texto.

40 Por lo tanto, puede acelerarse la velocidad de creación de gráfico, la velocidad de anotación y la velocidad de recuperación posteriores.

45 El otro objeto es para mejorar la consistencia entre los datos de imagen y los datos de texto. Por lo tanto, puede crearse un gráfico más robustamente en el procesamiento posterior, que a su vez proporciona precisión de anotación de imagen y precisión de recuperación de imagen superiores.

En la realización de la presente invención, las operaciones del módulo de proyección de sub-espacio incluyen las siguientes etapas.

50 Etapa B1:

Se obtienen la matriz característica de imagen y la matriz característica de texto, es decir, un grupo de los vectores característicos de imagen de las regiones de imagen y un grupo de los vectores característicos de texto de las regiones de imagen.

55 Etapa B2:

Se maximiza la covarianza entre las características de imagen y las características de texto para obtener la matriz de proyección característica de imagen y la matriz de proyección característica de texto.

60 Etapa B3:

65 La matriz característica de imagen proyectada se obtiene y almacena mediante la matriz de proyección característica de imagen y la matriz característica de imagen. La matriz característica de texto proyectada se obtiene y almacena mediante la matriz de proyección característica de texto y la matriz característica de texto.

En la realización de la presente invención, las características de imagen se proyectan en el sub-espacio de covarianza canónica para maximizar la covarianza entre la imagen y las anotaciones, y se emplea el CCovA (análisis de covarianza canónica). La razón es como sigue.

5 En otras palabras, las regiones de imagen que tienen imágenes similares pueden relacionarse más estrechamente entre sí en el espacio de CCOvA, reduciendo por lo tanto un error en un enlace más próximo a las regiones.

Adicionalmente, se reducen las longitudes de los vectores característicos de imagen, acortando de esta manera un tiempo de cálculo.

10 En este punto, la covarianza entre las características de imagen y las características de texto se expresa mediante la siguiente fórmula.

$$\rho_{PLS} = W_x' C_{xy} W_y / \sqrt{W_x' W_x W_y' W_y}$$

15 En este punto,  $C_{xy}$  representa una matriz de correlación entre una imagen y un texto, que puede definirse como sigue.

$$\hat{C}_{xy} = \tilde{X} \tilde{Y}' / (N-1) = \hat{C}_{xy} \in R^{p \times q}$$

20 En este punto,  $\tilde{X}$  e  $\tilde{Y}$  son las pilas de varianza centrada y de unidad de las matrices características de imagen y de texto, respectivamente que pueden expresarse mediante las siguientes fórmulas.

$$\tilde{X} = \left( \frac{x_1 - \hat{m}_x}{\hat{\sigma}_x}, \dots, \frac{x_N - \hat{m}_x}{\hat{\sigma}_x} \right)$$

$$\tilde{Y} = \left( \frac{y_1 - \hat{m}_y}{\hat{\sigma}_y}, \dots, \frac{y_N - \hat{m}_y}{\hat{\sigma}_y} \right)$$

25 En este punto, se establecen las siguientes relaciones.

$$\hat{m}_x = \frac{1}{N} \sum_{i=1}^N x_i \in R^p$$

$$\hat{\sigma}_x^2 = \frac{1}{N-1} \sum_{i=1}^N (x_i - \hat{m}_x)^2 \in R^p$$

$$\hat{m}_y = \frac{1}{N} \sum_{i=1}^N y_i \in R^q$$

$$\hat{\sigma}_y^2 = \frac{1}{N-1} \sum_{i=1}^N (y_i - \hat{m}_y)^2 \in R^q$$

30 En este punto,  $N$  representa el número de regiones de imagen, y  $x_i$  e  $y_i$  representan las  $i$ ésimas columnas en las matrices  $X$  e  $Y$ , respectivamente.

SVD (Descomposición en Valor Singular) se aplica a una matriz de correlación  $C_{xy}$  para maximizar la covarianza. Se supone la SVD para aplicarse a  $C_{xy}$  como sigue.

$$35 \quad C_{xy} = U D V'$$

En este punto,  $D$  representa la matriz diagonal de  $m_0 \times m_0$  e incluye  $m_0$  piezas de valores de covarianza de imagen-texto notables más significativos dispuestos diagonalmente orden descendente.

40 Por consiguiente, una matriz de proyección característica de imagen  $W_x$  y una matriz de proyección característica de texto  $W_y$  se representan como sigue.

$$W_x = U$$

$$W_y = V$$

45 Después de que se obtienen la matriz de proyección característica de imagen  $W_x$  y la matriz de proyección

característica de texto  $W_y$ , los grupos de los vectores característicos de imagen y los vectores característicos de texto (es decir, las matrices apiladas  $X$  e  $Y$ ) se proyectan en el sub-espacio correspondiente. La matriz característica de imagen proyectada  $X_{proy}$  y la matriz característica de texto  $Y_{proy}$  pueden expresarse mediante las siguientes fórmulas.

5

$$X_{proy} = W_x^t \times X$$

$$Y_{proy} = W_y^t \times Y$$

$X_{proy}$  e  $Y_{proy}$  representan una matriz  $m_0 \times N$ , y se establecen las relaciones  $m_0 \leq p$  y  $m_0 \leq q_0$ .

10 La matriz característica de imagen proyectada  $X_{proy}$  está optimizada para realizar la búsqueda de vecino más cercano en la región de imagen, mientras que la matriz característica de texto proyectada  $Y_{proy}$  está optimizada para realizar la búsqueda de vecino más cercano en el sub-espacio de texto.

15 Después de que se obtienen la matriz característica de imagen proyectada  $X_{proy}$  y la matriz característica de texto  $Y_{proy}$ , el gráfico se crea basándose en estas matrices. A continuación, la estructura del gráfico se describe en detalle más adelante.

20 En la realización de la presente invención, la creación del gráfico incluye las etapas de establecer primeros enlaces entre las regiones de imagen basándose en los vectores de proyección característicos de imagen proyectados, establecer segundos enlaces entre la imagen y las regiones de imagen de la imagen basándose en la operación del módulo de segmentación de imagen; establecer terceros enlaces entre la imagen y las anotaciones de la imagen basándose en la imagen que tiene las anotaciones; establecer cuartos enlaces entre las anotaciones basándose en los vectores de proyección característicos de texto; calcular los pesos para todos los enlaces; y obtener el gráfico que muestra la relación triangular entre la primera imagen, las regiones de imagen y las anotaciones basándose en los enlaces y los pesos de los enlaces.

Puesto que el primer, el segundo, el tercer y el cuarto enlaces cada uno incluye diversos enlaces, un único enlace del primer, del segundo, del tercer y del cuarto enlaces se denomina como un sub-enlace por razones de conveniencia.

30

La imagen de presentación, las regiones de imagen y las anotaciones se describen mediante nodos como se describe a continuación. Las relaciones entre los nodos se indican mediante líneas de enlace entre los nodos, y se representa un gráfico obtenido finalmente mediante una matriz de adyacencia  $A$ . El enlace entre los nodos  $i$  y  $j$  en el gráfico se expresa mediante la relación  $a_{ij} = w \neq 0$ . En este punto,  $w$  representa el peso del enlace. Si no hay enlace entre los nodos  $i$  y  $j$ , se establece la relación  $a_{ij} = 0$ . Puesto que no hay direccionalidad en el enlace, la matriz  $A$  es una matriz cuadrada en la que cualquier término diagonal representa 0 (el nodo y el enlace del propio nodo no se consideran).

35

El gráfico creado de la manera anterior tiene los siguientes tres tipos de nodos.

40

1. Todas las imágenes originales en el primer módulo de almacenamiento (nodos de imagen)
2. Todas las regiones de imagen después de segmentarse (nodos de región de imagen)
- 45 3. Palabras clave de anotaciones (nodos de anotaciones)

Existen dependencias entre el nodo de imagen y los nodos de región de imagen y entre el nodo de imagen y los nodos de anotación.

50

A continuación, se describen en detalle las respectivas partes más adelante.

(Establecimiento de enlaces entre nodos de región de imagen)

55 En los nodos de región de imagen, se busca el nodo de región de imagen más cercano basándose en la matriz característica de imagen proyectada  $X_{proy}$  para cada nodo de región de imagen (denominado para que sea el primer nodo de región de imagen por razones de conveniencia) de modo que se establece el enlace entre el primer nodo de región de imagen y  $K_1$  piezas de los nodos de región de imagen más cercanos al primer nodo de región de imagen.

60 En la realización de la presente invención, los pesos de los enlaces entre las regiones de imagen pueden confirmarse mediante los siguientes dos métodos.

Método de confirmación de peso 1:



Se aplica el mismo peso a todos los sub-enlaces. En otras palabras, se proporciona el mismo valor de peso para los respectivos enlaces.

Método de confirmación de peso 2:

Se aplican diferentes valores de peso a todos los sub-enlaces basándose en una diferencia en las imágenes de los sub-enlaces que corresponden a los nodos de región de imagen. Específicamente, cuanto mayor es la diferencia en las imágenes entre los sub-enlaces que corresponden a los nodos de regiones de imagen, más pequeños se hacen los valores de peso.

Cuando se emplea el método de confirmación de peso 2, el nodo de región de imagen que es relativamente enormemente diferente en la imagen a partir de un nodo de región de imagen actual tiene únicamente una pequeña influencia en el nodo de región de imagen actual puesto que el valor del peso del sub-enlace es pequeño. Por lo tanto, puede asegurarse un resultado de anotación y/o recuperación preciso incluso después de diversos accesos con el posterior algoritmo de RWR.

A continuación, se describe la implementación del método de confirmación de peso 2 basándose en un ejemplo.

Enlace ( $r_i \rightarrow r_j$ ) representa el peso del sub-enlace entre las regiones de imagen. En este punto, el valor de  $i$  es diferente del de  $j$ .

Enlace ( $r_i \rightarrow r_j$ ) se determina mediante similitud visual de imagen entre los nodos de región de imagen y se define como la función de disminución de una distancia característica entre los nodos de región de imagen  $i$  y  $j$  en el sub-espacio de imagen proyectado.

Por ejemplo, cuando se usa una norma L2 entre características como una distancia  $d_{im}$ , la función de disminución se define como sigue.

$$\text{enlace}(r_i, r_j) = A \exp\left(-\frac{d_{im}}{\bar{d}_{im}}\right)^2$$

En este punto,  $\bar{d}_{im}$  representa una distancia media entre un nodo de región de imagen  $r_i$  y  $K1$  piezas de los nodos de región de imagen más cercanos.  $A$  representa un parámetro de amplitud.

Cuando una diferencia visual en las regiones de imagen entre los nodos es grande, la distancia  $\bar{d}_{im}$  aumenta y el valor de peso disminuye, haciendo posible de esta manera evitar que los nodos de región de imagen se enlacen erróneamente entre sí en el gráfico y mejorar la calidad de hacer anotación y recuperación.

En la descripción anterior, se usa la norma L2. Sin embargo, otros métodos tales como chi cuadrado o medidas de entropía pueden usarse también.

(Enlaces entre nodo de imagen y nodos de región de imagen)

El enlace entre el nodo de imagen y todos los nodos de región de imagen de la imagen se establece para cada uno de los nodos de imagen. Cuando la imagen no está segmentada, únicamente existe un nodo de región de imagen. En este caso, el nodo de imagen corresponde al nodo de región de imagen.

En este punto, se establece el enlace mediante la operación del módulo de segmentación de imagen.

El peso (Enlace ( $l_{mi} \rightarrow r_j$ )) del sub-enlace entre el nodo de imagen y el nodo de región de imagen se determina mediante la relación padre-hijo entre el nodo de imagen y el nodo de región de imagen. Si el nodo de región de imagen  $r_j$  es el nodo de región de sub-imagen de un nodo de imagen  $l_{mi}$ , (Enlace ( $l_{mi} \rightarrow r_j$ )) se hace una constante distinta de 0. Si no, (Enlace ( $l_{mi} \rightarrow r_j$ )) se hace 0.

(Enlaces entre nodo de imagen y anotaciones del nodo de imagen)

El enlace entre el nodo de imagen y todas las anotaciones del nodo de imagen se establece para cada uno de los nodos de imagen.

En este punto, se establece el enlace basándose en la información de imagen del primer módulo de almacenamiento. En la realización de la presente invención, el peso del enlace entre el nodo de imagen y las

anotaciones del nodo de imagen puede determinarse mediante los siguientes dos métodos.

Método de determinación de peso 1:

- 5      Cuanto mayor es el número de apariciones del nodo de anotación que corresponden al sub-enlace en las anotaciones de la imagen almacenada en el primer módulo de almacenamiento, más grande se hace el valor de peso del sub-enlace.

Método de determinación de peso 2:

- 10     Cuanto mayor es el número de apariciones del nodo de anotación que corresponden al sub-enlace en las anotaciones de la imagen almacenada en el primer módulo de almacenamiento, más pequeño se hace el valor de peso del sub-enlace.

- 15     En el caso del método de determinación de peso 1, es natural que el algoritmo de RWR busque una palabra de anotación que aparezca a una frecuencia superior en las colecciones de palabras. Por lo tanto, la aparición de palabra a la frecuencia superior es siempre posible que se emita como el resultado de anotación de una imagen de recuperación. Sin embargo, este método puede ser impreciso en algunos casos. Por consiguiente, puede evitarse la desviación a la palabra de aparición con frecuencia con el método de determinación de peso 2.

- 20     Un método de determinación de peso específico es como sigue.

$$\text{enlace} (Im_i, w_j) = (B - \beta) + \beta \left( \frac{\text{máx\_ocu} - w_{j\_ocu}}{\text{máx\_ocu} - \text{mín\_ocu}} \right)$$

- 25     En este punto,  $B$  representa un parámetro de amplitud, y  $\beta$  es un parámetro que ajusta el equilibrio de peso entre palabras usadas populares y raramente. Adicionalmente,  $\text{máx\_ocu}$  representa el número de veces que aparece la palabra más popular, y  $\text{mín\_ocu}$  es el número de veces que aparece la palabra más raramente. Adicionalmente,  $w_{j\_ocu}$  es el número de veces que aparece la palabra considerada  $w_j$ .

- 30     (Enlaces entre nodos de anotación)

El enlace entre el nodo de anotación y  $K_2$  piezas de nodos convencionales más cercanos al nodo de anotación se establece para cada uno de los nodos de anotación. En este punto, se buscan los nodos convencionales más cercanos mediante la matriz característica de texto proyectada  $Y_{\text{proy}}$ .

- 35     Enlace ( $w_i \rightarrow w_j$ ) representa el peso del sub-enlace entre los nodos de anotación.

- 40     El peso del enlace entre  $w_i$  y  $w_j$  se determina mediante similitud textual entre las palabras clave de  $w_i$  y  $w_j$ , y se define como la función de disminución de la distancia entre las características de texto proyectadas  $i$  y  $j$  en el sub-espacio de texto. Por ejemplo, cuando se usa la norma L2 entre las características como una distancia  $d_{\text{txt}}$ , la función de disminución se define como sigue.

$$\text{enlace} (w_i, w_j) = C \exp \left( -\frac{d_{\text{txt}}}{\bar{d}_{\text{txt}}} \right)^2$$

- 45     En este punto,  $\bar{d}_{\text{txt}}$  representa una distancia media entre un nodo de región de imagen  $r_i$  y  $K_1$  piezas de nodos de región de imagen más cercanos al nodo de región de imagen  $r_i$ . Adicionalmente,  $C$  representa un parámetro de amplitud.

- 50     Si hay una diferencia relativamente grande en significado entre palabras, el valor de peso del sub-enlace se disminuye a medida que la distancia  $d_{\text{txt}}$  aumenta. Por lo tanto, se mejora el enlace entre las palabras que tienen los mismos significados, haciendo posible de esta manera hacer anotaciones más excelentes en correlación y consistencia en la anotación de imagen

- 55     Como se muestra en las Figuras 6 a 9, se establecen los enlaces anteriores y se realiza el algoritmo para determinar los pesos de los enlaces. En este punto,  $\text{ImageIndexDB.size}$  e  $\text{IMIndexDB.size}$  representan el número de nodos de región de imagen,  $\text{AnniDB.size}$  representa el número de imágenes almacenadas en el primer módulo de almacenamiento y  $\text{VOCA.size}$  representa el número de palabras clave en un grupo de palabras clave de anotación.

- 60     La Figura 10 muestra el gráfico creado mediante el método anterior.

En la Figura 10, únicamente se muestran dos nodos de imagen. Un primer nodo de imagen I1 tiene cuatro nodos de región de imagen R1 a R4, y un segundo nodo de imagen I2 tiene cuatro nodos de región de imagen R5 a R8. Además, existen nueve nodos de anotación, que se representan como W1 a W9. La Figura 10 muestra los enlaces entre los nodos de región de imagen, los enlaces entre los nodos de imagen y los nodos de región de imagen, los enlaces entre los nodos de imagen y los nodos de anotación y los enlaces entre los nodos de anotación.

La Figura 11 muestra la matriz de adyacencia A del gráfico obtenido finalmente.

En la realización de la presente invención, después de que se obtiene la matriz de adyacencia A, se realiza la anotación y la recuperación usando la matriz de adyacencia A, que se describe a continuación.

Un dispositivo de anotación automática de imagen de acuerdo con la realización de la presente invención está compuesto de los siguientes módulos. En otras palabras, el dispositivo de anotación automática de imagen incluye un primer módulo de almacenamiento, un módulo de recepción, un módulo de segmentación de imagen, un módulo de estación de vector característico, un módulo de proyección de sub-espacio, un segundo módulo de almacenamiento, un módulo de creación de gráfico, un módulo de actualización de imagen, un módulo de adquisición de nodo de anotación y un módulo de anotación. El primer módulo de almacenamiento almacena una primera imagen que tiene anotaciones y un grupo de palabras clave de anotación, en el que la primera imagen tiene una o más anotaciones obtenidas a partir del grupo de palabras clave y que describen el contenido del significado de la palabra de la primera imagen. El módulo de recepción recibe una segunda imagen esperando para la recuperación. El módulo de segmentación de imagen realiza una operación de segmentación en la primera y segunda imágenes para segmentarlas en una o más regiones de imagen. El módulo de extracción de vector característico extrae los vectores característicos de imagen de las regiones de imagen en la primera imagen a partir de las características de imagen, los vectores característicos de texto de las regiones de imagen en la primera imagen a partir de características de texto y los vectores característicos de imagen de las regiones de imagen en la segunda imagen a partir de las características de imagen, extrayendo de esta manera los vectores característicos que eliminan información y ruido innecesarios para las imágenes y captan un contenido esencial. El módulo de proyección de sub-espacio proyecta los vectores característicos de imagen de las regiones de imagen en la primera imagen en un sub-espacio para obtener primeros vectores de proyección característicos de imagen y los vectores característicos de texto de las regiones de imagen en la primera imagen en el sub-espacio para obtener vectores de proyección característicos de texto, maximizando de esta manera la covarianza entre las características de imagen y las características de texto y los vectores característicos de imagen de las regiones de imagen en la segunda imagen en el sub-espacio para obtener segundos vectores de proyección característicos de imagen. El segundo módulo de almacenamiento almacena los primeros vectores de proyección característicos de imagen y los vectores de proyección característicos de texto de las regiones de imagen. El módulo de creación de gráfico establece primeros enlaces entre las regiones de imagen basándose en los primeros vectores de proyección característicos de imagen, establece segundos enlaces entre la imagen y las regiones de imagen de la imagen basándose en la operación del módulo de segmentación de imagen, establece terceros enlaces entre la imagen y las anotaciones de la imagen basándose en la imagen que tiene las anotaciones, y establece cuartos enlaces entre las anotaciones basándose en los vectores de proyección característicos de texto, calcula los pesos del primer, del segundo, del tercer y del cuarto enlaces, y obtiene un gráfico que muestra una relación triangular entre la primera imagen, las regiones de imagen y las anotaciones, que se obtiene mediante el primer, el segundo, el tercer y el cuarto enlaces y los pesos de los enlaces que corresponden a estos enlaces y se expresa mediante una matriz. El módulo de actualización de imagen añade el nodo de imagen de la segunda imagen y los nodos de región de imagen de la segunda imagen al gráfico creado mediante el módulo de creación de gráfico, establece quintos enlaces entre los nodos de región de imagen de la segunda imagen y los nodos de región de imagen en el gráfico creado mediante el módulo de creación de gráfico basándose en los primeros vectores de proyección característicos de imagen y los segundos vectores de proyección característicos de imagen, establece sextos enlaces entre el nodo de imagen de la segunda imagen y los nodos de región de imagen de la segunda imagen, determina los pesos de los quintos y de los sextos enlaces, y actualiza el gráfico expresado mediante la matriz basándose en los quintos y los sextos enlaces y los pesos de los enlaces que corresponden a los quintos y a los sextos enlaces. El módulo de adquisición de nodo de anotación genera un vector de reinicio que corresponde a la segunda imagen y obtiene un número predeterminado de nodos de anotación más estrechamente relacionados con la segunda imagen con el algoritmo de RWR basándose en el gráfico expresado mediante la matriz actualizada. El módulo de anotación hace anotaciones en la segunda imagen usando la palabra clave que corresponde al número predeterminado de nodos de anotación.

En la realización de la presente invención, el dispositivo de anotación automática de imagen aplica procesamiento preliminar (la operación de división, la extracción de los vectores característicos de imagen de las regiones de imagen y la proyección de los vectores característicos de imagen) a la segunda imagen esperando para anotaciones basándose en el gráfico obtenido mediante el dispositivo de aprendizaje de imagen. Puesto que el procesamiento preliminar usa el mismo método que el dispositivo de aprendizaje de imagen, se omite la descripción detallada del método en este punto.

A continuación, se describen en detalle el módulo de creación de gráfico y el módulo de adquisición de nodo de anotación del dispositivo de anotación automática de imagen más adelante.

Al igual que la función del módulo de creación de gráfico, se añaden el nodo de imagen y el nodo de región de imagen de una imagen esperando convencional (es decir, datos de recuperación) gráfico creado. El gráfico se expresa mediante una matriz, y la matriz es una matriz de forma simétricamente cuadrada. Por lo tanto, cuando se añaden el nodo de imagen y el nodo de región de imagen a la matriz, se añaden una fila y una columna al gráfico basándose en la matriz mostrada en la Figura 11. En otras palabras, el nodo de imagen y el nodo de región de imagen de la segunda imagen se añaden a la matriz original A como la fila y la columna de la matriz. Específicamente, el nodo de imagen y el nodo de región de imagen de la segunda imagen se añaden recientemente a las filas y a las columnas de la matriz original A como los nuevos nodos, se establecen los enlaces entre el nodo de región de imagen recientemente añadido y K1 piezas de los nodos de región de imagen más cercanos en el gráfico original, añadiendo de esta manera los pesos de los nuevos enlaces.

Los valores de elemento de la matriz actualizada se determinan mediante los enlaces y los pesos de los enlaces.

Puesto que el establecimiento de los enlaces y la determinación de los pesos de los enlaces se han descrito anteriormente, se omiten sus descripciones duplicadas en este punto.

La Figura 12 muestra la matriz actualizada.

En la Figura 12, no se muestran los valores de la matriz antes de actualizarse, pero  $R_n$  representa un nuevo nodo de región de imagen e  $l_{mn}$  representa un nodo esperando para la recuperación.

El vector de reinicio que corresponde a la segunda imagen se genera inmediatamente después de la actualización de la matriz. Se obtiene un número predeterminado de nodos de anotación más estrechamente relacionados con la segunda imagen basándose en el gráfico expresado mediante la matriz actualizada con el algoritmo de RWR, y se emite un número predeterminado de palabras clave que corresponden a los nodos de anotación.

A continuación, se describe la generación del vector de reinicio.

Cuando se usa una anotación de imagen como una cantidad de recuperación en la recuperación de imagen, el elemento de un nodo de recuperación de imagen en el vector de reinicio representa 1 y los elementos restantes del mismo representan 0.

Obsérvese que se describen las etapas de procesamiento del algoritmo de RWR que especifican un número predeterminado de palabras clave más correlacionadas como anotaciones de imagen objetivo a continuación.

En una etapa de inicialización,  $U_0$  se establece como el vector de reinicio.

En una etapa de repetición, se realiza  $U_{n+1} = (1 - c) A U_n + cV$  repetidamente hasta que se cumplen las condiciones de parada.

En este punto,  $n$  representa un índice de repetición,  $c$  representa una constante que refleja una probabilidad de reinicio y  $A$  representa una matriz de adyacencia en el gráfico actualizado. Obsérvese que  $V$  representa el vector de reinicio ajustado.  $U_n$  representa un vector de estado que corresponde a una etapa  $n$ , y el valor del  $i$ -ésimo elemento en este vector refleja la probabilidad de alcanzar desde un nodo objetivo (nodo de imagen añadido recientemente) a un nodo  $i$  a  $n$  veces de repeticiones en el gráfico. En otras palabras, refleja un grado de correlación entre el nodo que corresponde al  $i$ -ésimo elemento y el nodo objetivo, es decir, una extensión de cuánto están estrechamente relacionados entre sí.

En una etapa de extracción, los elementos que corresponden a nodos de anotación en el vector de estado  $U_n$  se extraen y disponen en orden descendente.

En una etapa de salida, se emite un número predeterminado de palabras clave que corresponden a los nodos de anotación una a una.

En las etapas de procesamiento anteriormente descritas, se normaliza una matriz de adyacencia actualizada con antelación para converger después de que el RWR se repite varias veces.

La normalización se realiza de la siguiente manera.

Método 1:

En la matriz actualizada  $A$ , las filas de elementos respectivos se dividen por la suma de las filas de elementos. Por lo tanto, la matriz de adyacencia normalizada  $\tilde{A}$  se expresa como sigue.

$$\tilde{\mathbf{A}} = \mathbf{A}\mathbf{D}^{-1/2}, \quad \mathbf{D}_{ij} = \sum_j \mathbf{A}_{ij}, \quad \text{y} \quad \mathbf{D}_{ij} = 0 \quad \text{donde} \quad i \neq j$$

Método 2:

- 5 La matriz de adyacencia  $\tilde{\mathbf{A}}$  sometida a la normalización de Laplace se expresa como sigue.

$$\tilde{\mathbf{A}} = \mathbf{D}^{-1/2} \mathbf{A} \mathbf{D}^{-1/2}, \quad \mathbf{D}_{ij} = \sum_j \mathbf{A}_{ij}, \quad \text{y} \quad \mathbf{D}_{ij} = 0 \quad \text{donde} \quad i \neq j$$

10 La normalización no está limitada a los métodos anteriores, sino que otros métodos pueden emplearse también.

Obsérvese que el algoritmo de RWR después de la operación de normalización realiza las operaciones posteriores usando la matriz de adyacencia normalizada  $\tilde{\mathbf{A}}$ , pero se omiten las descripciones detalladas de las operaciones en este punto.

- 15 Un dispositivo de recuperación de imagen de acuerdo con la realización de la presente invención está compuesto de los siguientes módulos cuando se usa para realizar recuperación de imagen en la entrada de una segunda imagen. En otras palabras, el dispositivo de recuperación de imagen incluye un primer módulo de almacenamiento, un módulo de recepción, un módulo de segmentación de imagen, un módulo de extracción de vector característico, un módulo de proyección de sub-espacio, un segundo módulo de almacenamiento, un módulo de creación de gráfico, un módulo de actualización de imagen, un módulo de recuperación y un módulo de salida. El primer módulo de almacenamiento almacena una primera imagen que tiene anotaciones y un grupo de palabras clave de anotación, en el que la primera imagen tiene una o más anotaciones obtenidas a partir del grupo de palabras clave de anotación y que describe el contenido del significado de la palabra de la primera imagen. El módulo de recepción recibe una segunda imagen esperando para la recuperación. El módulo de segmentación de imagen realiza una operación de segmentación en la primera y segunda imágenes para segmentarlas en una o más regiones de imagen. El módulo de extracción de vector característico extrae los vectores característicos de imagen de las regiones de imagen en la primera imagen a partir de características de imagen, los vectores característicos de texto de las regiones de imagen en la primera imagen a partir de características de texto y los vectores característicos de imagen de las regiones de imagen en la segunda imagen a partir de características de imagen, extrayendo de esta manera los vectores característicos que eliminan información y ruido innecesarios para la imagen y captan un contenido esencial. El módulo de proyección de sub-espacio proyecta los vectores característicos de imagen de las regiones de imagen en la primera imagen en un sub-espacio para obtener primeros vectores de proyección característicos de imagen y los vectores característicos de texto de las regiones de imagen en la primera imagen en el sub-espacio para obtener vectores de proyección característicos de texto, maximizando de esta manera la covarianza entre las características de imagen y las características de texto y los vectores característicos de imagen de las regiones de imagen en la segunda imagen en el sub-espacio para obtener segundos vectores de proyección característicos de imagen. El segundo módulo de almacenamiento almacena los primeros vectores de proyección característicos de imagen y los vectores de proyección característicos de texto de las regiones de imagen. El módulo de creación de gráfico establece primeros enlaces entre las regiones de imagen basándose en los primeros vectores de proyección característicos de imagen, establece segundos enlaces entre la imagen y las regiones de imagen de la imagen basándose en la operación del módulo de segmentación de imagen, establece terceros enlaces entre la imagen y las anotaciones de la imagen basándose en la imagen que tiene las anotaciones y establece cuartos enlaces entre las anotaciones basándose en los vectores de proyección característicos de texto, calcula los pesos del primer, del segundo, del tercer y del cuarto enlaces y obtiene un gráfico que muestra una relación triangular entre la primera imagen, las regiones de imagen y las anotaciones, que se obtienen mediante el primer, el segundo, el tercer y el cuarto enlaces y los pesos de los enlaces que corresponden a estos enlaces y se expresa mediante una matriz. El módulo de actualización de imagen añade el nodo de imagen de la segunda imagen y los nodos de región de imagen de la segunda imagen al gráfico creado mediante el módulo de creación de gráfico, establece quintos enlaces entre los nodos de región de imagen de la segunda imagen y los nodos de región de imagen en el gráfico creado mediante el módulo de creación de gráfico basándose en los primeros vectores de proyección característicos de imagen y los segundos vectores de proyección característicos de imagen, establece sextos enlaces entre el nodo de imagen de la segunda imagen y los nodos de región de imagen de la segunda imagen, determina los pesos de los quintos y sextos enlaces, y actualiza el gráfico expresado mediante la matriz basándose en los quintos y los sextos enlaces y los pesos de los enlaces que corresponden a los quintos y los sextos enlaces. El módulo de recuperación genera un vector de reinicio que corresponde a la segunda imagen y obtiene un número predeterminado de nodos de anotación más estrechamente relacionados con la segunda imagen con el algoritmo de RWR basándose en el gráfico expresado mediante la matriz actualizada. El módulo de salida emite la primera imagen que corresponde al número predeterminado de nodos de imagen.

- 60 El dispositivo de recuperación de imagen de acuerdo con la realización de la presente invención es casi similar al dispositivo de anotación automática de imagen excepto por un resultado de salida. Por lo tanto, se omite la descripción del dispositivo de recuperación de imagen en este punto.

El dispositivo de recuperación de imagen de acuerdo con la realización de la presente invención está compuesto de los siguientes módulos cuando se usa para realizar recuperación de imagen en la entrada de una segunda imagen y una palabra clave. En otras palabras, el dispositivo de recuperación de imagen incluye un primer módulo de almacenamiento, un módulo de recepción, un módulo de segmentación de imagen, un módulo de extracción de vector característico, un módulo de proyección de sub-espacio, un segundo módulo de almacenamiento, un módulo de creación de gráfico, un módulo de actualización de imagen, un módulo de recuperación y un módulo de salida. El primer módulo de almacenamiento almacena una primera imagen que tiene anotaciones y un grupo de palabras clave de anotación, en el que la primera imagen tiene una o más anotaciones obtenidas a partir del grupo de palabras clave de anotación y que describen el contenido del significado de la palabra de la primera imagen. El módulo de recepción recibe una segunda imagen esperando para la recuperación y la palabra clave. El módulo de segmentación de imagen realiza una operación de segmentación en la primera y segunda imágenes para segmentarlas en una o más regiones de imagen. El módulo de extracción de vector característico extrae los vectores característicos de imagen de las regiones de imagen en la primera imagen a partir de características de imagen, los vectores característicos de texto de las regiones de imagen en la primera imagen a partir de características de texto y los vectores característicos de imagen de las regiones de imagen en la segunda imagen a partir de las características de imagen, extrayendo de esta manera los vectores característicos que eliminan información y ruido innecesarios para la imagen y captan un contenido esencial. El módulo de proyección de sub-espacio proyecta los vectores característicos de imagen de las regiones de imagen en la primera imagen en un sub-espacio para obtener primeros vectores de proyección característicos de imagen y los vectores característicos de texto de las regiones de imagen en la primera imagen en el sub-espacio para obtener vectores de proyección característicos de texto, maximizando de esta manera la covarianza entre las características de imagen y las características de texto y los vectores característicos de imagen de las regiones de imagen en la segunda imagen en el sub-espacio para obtener segundos vectores de proyección característicos de imagen. El segundo módulo de almacenamiento almacena los primeros vectores de proyección de característicos de imagen y los vectores de proyección característicos de texto de las regiones de imagen. El módulo de creación de gráfico establece primeros enlaces entre las regiones de imagen basándose en los primeros vectores de proyección característicos de imagen, establece segundos enlaces entre la imagen y las regiones de imagen de la imagen basándose en la operación del módulo de segmentación de imagen, establece terceros enlaces entre la imagen y las anotaciones de la imagen basándose en la imagen que tiene las anotaciones, y establece cuartos enlaces entre las anotaciones basándose en los vectores de proyección característicos de texto, calcula los pesos del primer, del segundo, del tercer y del cuarto enlaces, y obtiene un gráfico que muestra una relación triangular entre la primera imagen, las regiones de imagen y las anotaciones, que se obtienen mediante el primer, el segundo, el tercer y el cuarto enlaces y los pesos de los enlaces que corresponden a estos enlaces y se expresa mediante una matriz. El módulo de actualización de imagen añade el nodo de imagen de la segunda imagen y los nodos de región de imagen de la segunda imagen al gráfico creado mediante el módulo de creación de gráfico, establece quintos enlaces entre los nodos de región de imagen de la segunda imagen y los nodos de región de imagen en el gráfico creado mediante el módulo de creación de gráfico basándose en los primeros vectores de proyección característicos de imagen y los segundos vectores de proyección característicos de imagen, establece sextos enlaces entre el nodo de imagen de la segunda imagen y los nodos de región de imagen de la segunda imagen, determina los pesos de los quintos y los sextos enlaces, y actualiza el gráfico expresado mediante la matriz basándose en los quintos y los sextos enlaces y los pesos de los enlaces que corresponden a los quintos y a los sextos enlaces. El módulo de recuperación genera un vector de reinicio que corresponde a la segunda imagen y obtiene un número predeterminado de nodos de anotación más estrechamente relacionados con la segunda imagen y una palabra clave de recuperación con el algoritmo de RWR basándose en el gráfico expresado mediante la matriz actualizada. El módulo de salida emite la primera imagen que corresponde al número predeterminado de nodos de imagen.

El dispositivo de recuperación de imagen de acuerdo con la realización de la presente invención es casi similar al dispositivo de anotación automática de imagen excepto por un resultado de salida y un vector de reinicio.

El vector de reinicio tiene dos partes para una imagen y una palabra clave. Un elemento en  $k$  que corresponde a un nodo de recuperación de imagen y un nodo de anotación representa 1 y los elementos restantes representan 0.

El dispositivo de recuperación de imagen de acuerdo con la realización de la presente invención está compuesto de los siguientes módulos cuando se usa para realizar recuperación de imagen en la entrada de una palabra clave de recuperación.

En otras palabras, el dispositivo de recuperación de imagen de acuerdo con la realización de la presente invención incluye un primer módulo de almacenamiento, un módulo de recepción, un módulo de segmentación de imagen, un módulo de extracción de vector característico, un módulo de proyección de sub-espacio, un segundo módulo de almacenamiento, un módulo de creación de gráfico, un módulo de recuperación y un módulo de salida. El primer módulo de almacenamiento almacena una primera imagen que tiene anotaciones y un grupo de palabras clave de anotación, en el que la primera imagen tiene una o más anotaciones obtenidas a partir del grupo de palabras clave de anotación y que describe el contenido del significado de la palabra de la primera imagen. El módulo de recepción recibe las palabras clave para la recuperación. El módulo de segmentación de imagen realiza una operación de segmentación en la primera imagen para segmentarla en una o más regiones de imagen. El módulo de extracción de vector característico extrae los vectores característicos de imagen de las regiones de imagen en la primera

imagen a partir de características de imagen y los vectores característicos de texto de las regiones de imagen en la primera imagen a partir de características de texto, extrayendo de esta manera los vectores característicos que eliminan información y ruido innecesarios para la imagen captan un contenido esencial. El módulo de proyección de sub-espacio proyecta los vectores característicos de imagen de las regiones de imagen en la primera imagen en un sub-espacio para obtener primeros vectores de proyección característicos de imagen y los vectores característicos de texto de las regiones de imagen en la primera imagen en el sub-espacio para obtener vectores de proyección característicos de texto, maximizando de esta manera la covarianza entre las características de imagen y las características de texto. El segundo módulo de almacenamiento almacena los primeros vectores de proyección característicos de imagen y los vectores de proyección característicos de texto de las regiones de imagen. El módulo de creación de gráfico establece primeros enlaces entre las regiones de imagen basándose en los primeros vectores de proyección característicos de imagen proyectados, establece segundos enlaces entre la imagen y las regiones de imagen de la imagen basándose en la operación del módulo de segmentación de imagen, establece terceros enlaces entre la imagen y las anotaciones de la imagen basándose en la imagen que tiene las anotaciones y establece cuartos enlaces entre las anotaciones basándose en los vectores de proyección característicos de texto, calcula los pesos del primer, del segundo, del tercer y del cuarto enlaces, y obtiene un gráfico que muestra una relación triangular entre la primera imagen, las regiones de imagen y las anotaciones, que se obtienen mediante el primer, el segundo, el tercer y el cuarto enlaces y los pesos de los enlaces que corresponden a estos enlaces y se expresa mediante una matriz. El módulo de recuperación genera un vector de reinicio que corresponde a la palabra clave de recuperación y obtiene un número predeterminado de nodos de imagen más estrechamente relacionados con la palabra clave de recuperación con el algoritmo de RWR basándose en el gráfico expresado mediante la matriz actualizada. El módulo de salida emite la primera imagen que corresponde al número predeterminado de nodos de imagen.

En la realización de la presente invención, las salidas de un número predeterminado de palabras clave y la primera imagen con el dispositivo de anotación automática de imagen y el dispositivo de recuperación de imagen pueden realizarse mediante diversos métodos. A continuación, se describen los métodos en detalle más adelante.

Después de que se converja y establezca un vector U, los valores (probabilidad de alcanzar a partir de un nodo objetivo a un nodo) que corresponden a nodos respectivos en el gráfico se disponen uno a uno dependiendo del tipo de nodo (nodos de imagen, nodos de región de imagen y nodos de anotación).

En las anotaciones de imagen, todos los elementos en el vector U que corresponden a los nodos de anotación en el gráfico se toman y disponen descendientemente en el orden de probabilidad. Entre los elementos, se determinan las primeras p piezas de palabras clave para que sean anotaciones más correlacionadas a una imagen de recuperación.

En la recuperación de imagen, todos los elementos en el vector U que corresponden a los nodos de anotación en el gráfico se toman y disponen descendientemente en el orden de probabilidad. Entre los elementos, se determinan las primeras p piezas de imágenes para que sean imágenes más correlacionadas para la recuperación (un texto o una imagen, o ambos del texto y de la imagen).

En este punto, se determina p de la siguiente manera.

Concretamente, se determina p automáticamente cuando se establece una cierta constante, o se determina mediante probabilidad de correlación.

En el último caso, se calcula p como sigue.

Es decir, se obtiene p cuando la probabilidad acumulativa de las primeras p piezas de salidas (anotaciones o recuperación) es mayor que un umbral  $th_1$ , o cuando una diferencia entre la probabilidad de la p-ésima salida (anotación o recuperación) y la de la (p+1)-ésima salida es mayor que un umbral  $th_2$ .

Puede determinarse p mediante métodos distintos de los anteriores métodos, pero se omiten las descripciones detalladas de los métodos en este punto.

La Figura 14 es el dibujo de un hombre y una mujer en una habitación. La Figura 14 muestra un caso en el que se aplica la realización de la presente invención a una imagen de entrada en la que se realizan anotaciones automáticamente. Cuando se introduce la imagen mostrada en la Figura 14, se obtiene un interior, personas, una mujer y un hombre como anotación.

La Figura 15 tiene cuatro dibujos teniendo cada uno al menos una persona y un animal. La Figura 15 muestra resultados de recuperación obtenidos mediante el dispositivo de recuperación de imagen y el método de recuperación de imagen de acuerdo con la realización de la presente invención cuando se introduce "una persona y un animal".

La Figura 16 tiene tres dibujos en los que tres niños juegan con ladrillos de madera y un dibujo de una persona esquiendo. La Figura 16 muestra resultados de recuperación obtenidos mediante el dispositivo de recuperación de

imagen y el método de recuperación de imagen de acuerdo con la realización de la presente invención cuando se introduce una imagen. En la Figura 16, la imagen más a la izquierda es la entrada y las tres imágenes restantes de la derecha son la salida.

5 La Figura 17 tiene un dibujo que incluye una flor y un texto y tres dibujos teniendo cada uno una flor y un automóvil. La Figura 17 muestra resultados de recuperación obtenidos mediante la realización de la presente invención cuando se introduce un texto y una imagen. En la Figura 17, la imagen más a la izquierda es una imagen de entrada y las tres imágenes restantes de la derecha son resultados de salida.

10 Cuando las descripciones con relación a las Figuras 14 a 16 y estas figuras se ponen juntas en las imágenes reales, el método y el dispositivo de acuerdo con la realización de la presente invención puede implementar satisfactoriamente anotación y recuperación de imágenes automática.

De acuerdo con la realización de la presente invención, pueden obtenerse las siguientes ventajas útiles.

15 1) En el aprendizaje de imagen, la anotación automática, el método de recuperación y el dispositivo de acuerdo con la realización de la presente invención, los vectores característicos de imagen y los vectores característicos de texto se proyectan en el mismo sub-espacio en la etapa de aprendizaje para maximizar la covarianza entre la característica de imagen y la característica de texto, reduciendo de esta manera las dimensiones de la característica de imagen y de la característica de texto. Por lo tanto, puede acelerarse la velocidad de creación de gráfico, la velocidad de anotación y la velocidad de recuperación. También, puede mejorarse la consistencia entre los datos de imagen y datos de texto. Por consiguiente, puede crearse el gráfico más robustamente, que a su vez proporciona precisión de anotación de imagen y precisión de recuperación de imagen superiores.

25 2) En el aprendizaje de imagen, la anotación automática, el método de recuperación y el dispositivo de acuerdo con la realización de la presente invención, se usa el sub-espacio de covarianza canónica como un sub-espacio de modo que las regiones de imagen que tienen imágenes similares pueden relacionarse más estrechamente entre sí. Por lo tanto, puede reducirse un error en el enlace más cercano entre los nodos de región de imagen.

30 3) En el aprendizaje de imagen, la anotación automática, el método de recuperación y el dispositivo de acuerdo con la realización de la presente invención, cuanto mayor es una diferencia en las imágenes entre los nodos de región de imagen que corresponden al enlace, más pequeño se hace el valor de peso del enlace. Por lo tanto, en comparación con el nodo de región de imagen que tiene una diferencia de imagen relativamente más grande, se hace pequeña una influencia en el nodo de región de imagen actual. Por lo tanto, pueden emitirse fiablemente resultados de anotaciones y/o recuperación precisos incluso después de diversos accesos con el algoritmo de RWR.

35 4) En el aprendizaje de imagen, la anotación automática, el método de recuperación y el dispositivo de acuerdo con la realización de la presente invención, cuando los nodos de región de imagen se enlazan entre sí, cuanto mayores son las apariciones del nodo de región de imagen que corresponde al enlace, más pequeño se hace el peso del enlace. Por lo tanto, puede hacerse anotación automática más precisa evitando desviarse a palabras frecuentemente usadas.

La presente invención no está limitada a las realizaciones específicamente desveladas, y pueden realizarse variaciones y modificaciones sin alejarse del alcance de la presente invención.

45



**REIVINDICACIONES**

1. Un método de anotación automática de imagen para hacer una anotación en una segunda imagen de entrada, comprendiendo el método de anotación automática de imagen una etapa de aprendizaje, una etapa de procesamiento preliminar, una etapa de actualización de gráfico y una etapa de anotación;

en el que la etapa de aprendizaje incluye:

realizar una operación de segmentación en una primera imagen que tiene anotaciones para segmentar la primera imagen en una o más primeras regiones de imagen;

extraer primeros vectores característicos de imagen a partir de todas las primeras regiones de imagen para obtener una primera matriz característica de imagen;

establecer primeros enlaces entre las primeras regiones de imagen;

establecer segundos enlaces entre la primera imagen y las primeras regiones de imagen basándose en un resultado de la operación de segmentación;

establecer terceros enlaces entre la primera imagen y las anotaciones basándose en la primera imagen que tiene las anotaciones;

calcular pesos de todos los enlaces; y

obtener un gráfico que muestra una realización triangular entre la primera imagen, las primeras regiones de imagen y las anotaciones basándose en todos los enlaces y los pesos de los enlaces que corresponden a los enlaces;

en el que la etapa de procesamiento preliminar incluye:

recibir la segunda imagen;

realizar la operación de segmentación en la segunda imagen para segmentar la segunda imagen en una o más segundas regiones de imagen; y

extraer segundos vectores característicos de imagen a partir de todas las segundas regiones de imagen para obtener una segunda matriz característica de imagen de la segunda imagen;

en el que la etapa de actualización de gráfico incluye:

establecer quintos enlaces entre los segundos nodos de región de imagen de la segunda imagen y los primeros nodos de región de imagen en el gráfico;

establecer sextos enlaces entre la segunda imagen y los segundos nodos de región de imagen basándose en un resultado de la operación de segmentación;

determinar pesos de los enlaces de los quintos enlaces y los sextos enlaces; y

actualizar el gráfico basándose en los quintos enlaces y los sextos enlaces y los pesos de los enlaces que corresponden a los quintos enlaces y a los sextos enlaces, y

en el que la etapa de anotación incluye:

generar un vector de reinicio que corresponde a la segunda imagen;

obtener un número predeterminado de anotaciones más estrechamente relacionadas con la segunda imagen con un recorrido aleatorio con reinicio; y

hacer las anotaciones en la segunda imagen usando palabras clave que corresponden al número predeterminado de anotaciones;

**CARACTERIZADO POR QUE:**

la etapa de aprendizaje incluye adicionalmente:

extraer vectores característicos de texto a partir de todas las primeras regiones de imagen para obtener una matriz característica de texto;

proyectar la primera matriz característica de imagen y la matriz característica de texto en un sub-espacio maximizando una covarianza entre una característica de imagen y una característica de texto, obteniendo de esta manera la primera matriz característica de imagen proyectada y la matriz característica de texto proyectada;

5 almacenar la primera matriz característica de imagen proyectada y la matriz característica de texto proyectada;

en el que se establecen dichos primeros enlaces basándose en la primera matriz característica de imagen proyectada;

10 establecer cuartos enlaces entre las anotaciones basándose en la matriz característica de texto proyectada;

en el que la etapa de procesamiento preliminar incluye adicionalmente

15 proyectar la segunda matriz característica de imagen en el sub-espacio para obtener una segunda matriz característica de imagen proyectada de la segunda imagen;

en el que la etapa de actualización de gráfico incluye adicionalmente

20 establecer dichos quintos enlaces basándose en la primera matriz característica de imagen proyectada y la segunda matriz característica de imagen proyectada;

en el que los pesos  $w$  de los primeros enlaces se calculan de acuerdo con:

$$w = A \exp\left(-\frac{d_{im}}{\bar{d}_{im}}\right)^2$$

25 para cada uno de los primeros enlaces, en el que  $d_{im}$  es una distancia característica en el subespacio proyectado entre las regiones de imagen enlazadas mediante un enlace individual,  $\bar{d}_{im}$  es una distancia media entre una región de imagen enlazada mediante el enlace y un número predeterminado de regiones de imagen más cercanas de dicha región de imagen y  $A$  representa un parámetro de amplitud.

30 2. El método de acuerdo con la reivindicación 1, en el que en la operación de segmentación de la etapa de aprendizaje, la primera imagen se segmenta en bloques rectangulares, sub-bloques de árbol cuádruple de multi-resolución o regiones homogéneas no solapantes con un algoritmo de segmentación de imagen.

35 3. El método de acuerdo con la reivindicación 1 o 2, en el que en la etapa de aprendizaje los primeros vectores característicos de imagen de todas las primeras regiones de imagen se extraen mediante un algoritmo basándose en una característica de patrón binario local que comprende colores mezclados e información de patrón.

40 4. El método de acuerdo con la reivindicación 1, 2 o 3, en el que el sub-espacio es un sub-espacio de covarianza canónica.

45 5. El método de acuerdo con cualquier reivindicación anterior, en el que la primera imagen, las primeras regiones de imagen y las anotaciones se representan mediante nodos en el gráfico, el gráfico se representa mediante una matriz de adyacencia, los enlaces entre los nodos se representan mediante los pesos de los enlaces en el gráfico y un valor de un peso correspondiente representa 0 si no hay enlace entre los nodos.

6. El método de acuerdo con la reivindicación 5, **caracterizado por que,**

50 en los terceros enlaces, cuanto mayor es el número de apariciones del nodo de anotación que corresponde al sub-enlace, más pequeño se hace el valor de peso del sub-enlace, y/o

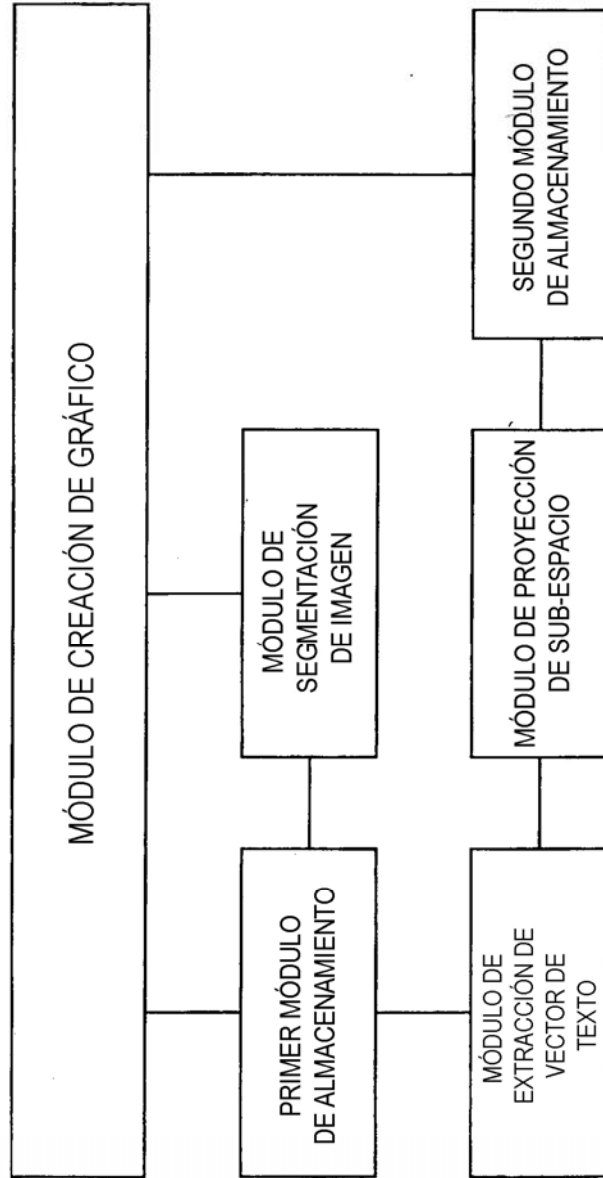
en los cuartos enlaces, cuanto mayor es una similitud en texto entre los nodos de anotación que corresponden al sub-enlace, más pequeño se hace el valor de peso del sub-enlace.

55 7. El método de acuerdo con cualquier reivindicación anterior, en el que en la operación de segmentación de la etapa de procesamiento preliminar, la primera imagen y la segunda imagen se segmentan en bloques rectangulares, sub-bloques de árbol cuádruple de multi-resolución, o regiones homogéneas no solapantes con un algoritmo de segmentación de imagen.

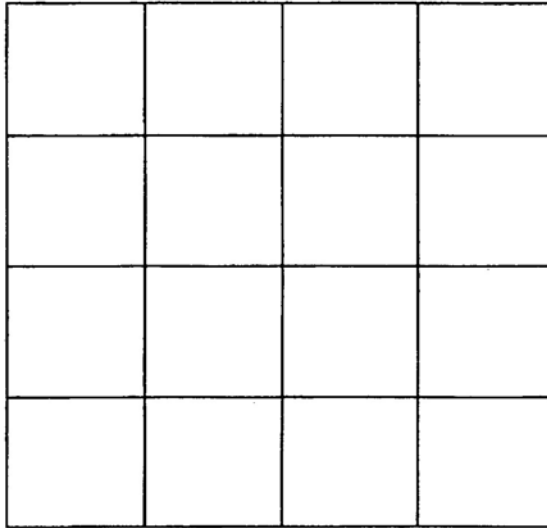
60 8. El método de acuerdo con cualquier reivindicación anterior, en el que en la etapa de procesamiento preliminar los vectores característicos de imagen de todas las regiones de imagen se extraen mediante un algoritmo basándose en una característica de patrón binario local que comprende colores mezclados e información de patrón.

- 5 9. El método de anotación automática de imagen de acuerdo con la reivindicación 1, 2, 3, 7 u 8, en el que la primera imagen, la segunda imagen, las regiones de imagen y las anotaciones se representan mediante nodos en el gráfico, el gráfico se representa mediante una matriz de adyacencia, los enlaces entre los nodos se representan mediante los pesos de los enlaces en el gráfico y un valor de un peso correspondiente representa 0 si no hay enlace entre los nodos.
10. El método de acuerdo con la reivindicación 9, en el que:
- 10 en los terceros enlaces, cuanto mayor es el número de apariciones de los nodos de anotación que corresponden al sub-enlace, más pequeño se hace el valor de peso del sub-enlace, y/o
- en los cuartos enlaces, cuanto mayor es la similitud en texto entre los nodos de anotación que corresponden al sub-enlace, más pequeño se hace el valor de peso de los sub-enlaces.
- 15 11. El método de acuerdo con la reivindicación 9 o 10, que comprende adicionalmente una etapa de aplicar normalización a la matriz de adyacencia actualizada antes de realizar la etapa de anotación basándose en la matriz de adyacencia sometida a la normalización.
- 20 12. Un método de recuperación de imagen basándose en una segunda imagen de entrada, comprendiendo el método de recuperación de imagen una etapa de aprendizaje, una etapa de procesamiento preliminar, una etapa de actualización de gráfico, una etapa de anotación y una etapa de recuperación en el que las etapas de aprendizaje, procesamiento preliminar, actualización de gráfico y anotación son de acuerdo con la reivindicación 1 **caracterizado por que**
- 25 la etapa de recuperación incluye:
- generar un vector de reinicio que corresponde a la segunda imagen; y
- 30 adquirir y emitir un número predeterminado de anotaciones más estrechamente relacionadas con la segunda imagen con un recorrido aleatorio con reinicio.
- 35 13. El método de recuperación de imagen de acuerdo con la reivindicación 12, **caracterizado por que** se introduce adicionalmente una palabra clave, y en la etapa de recuperación, se generan el vector de reinicio que corresponde a la segunda imagen y la palabra clave y se adquiere un número predeterminado de imágenes más estrechamente relacionadas con la segunda imagen y la palabra clave y se emiten basándose en el gráfico actualizado.

FIG.1



**FIG.2**



**FIG.3**

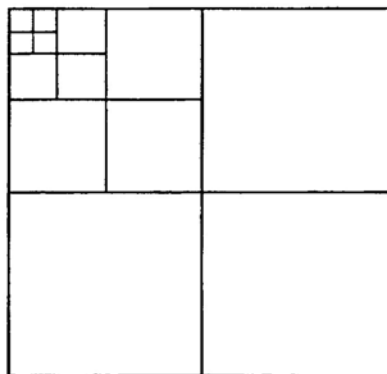


FIG.4

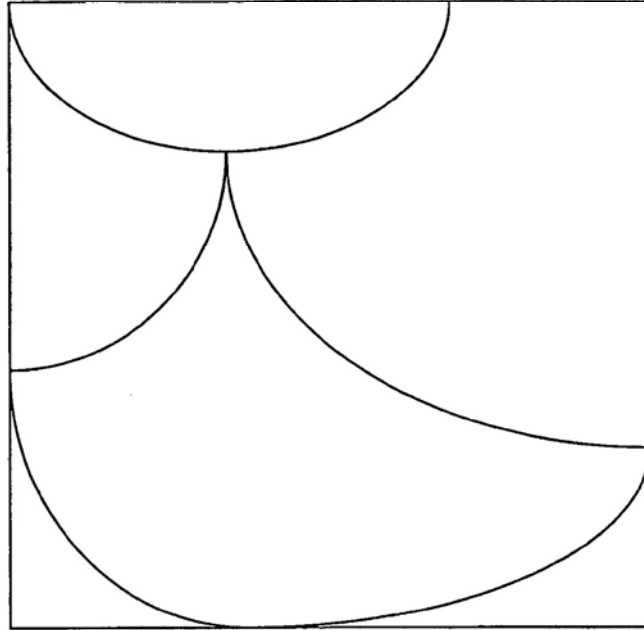


FIG.5

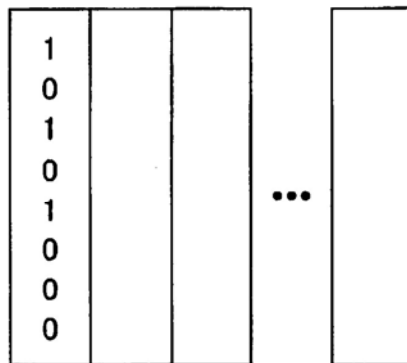


FIG.6

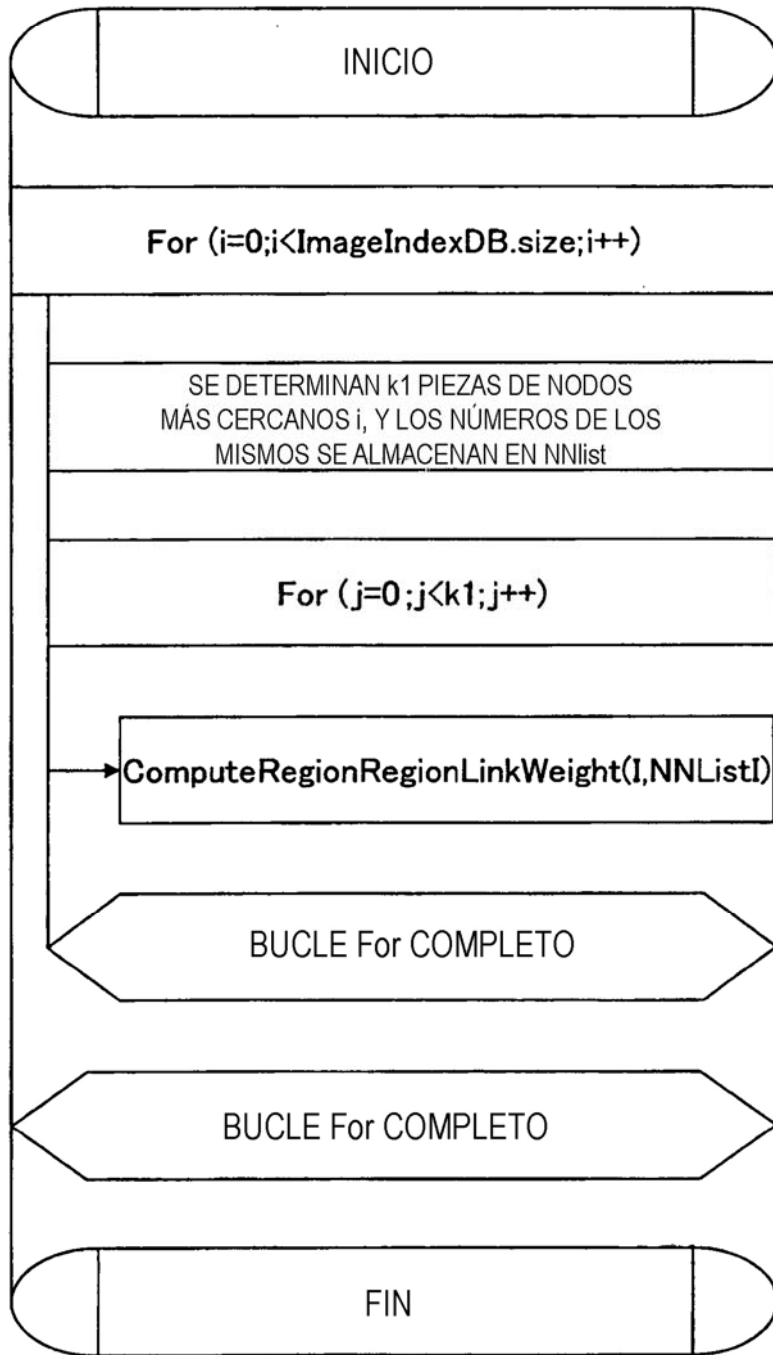


FIG.7

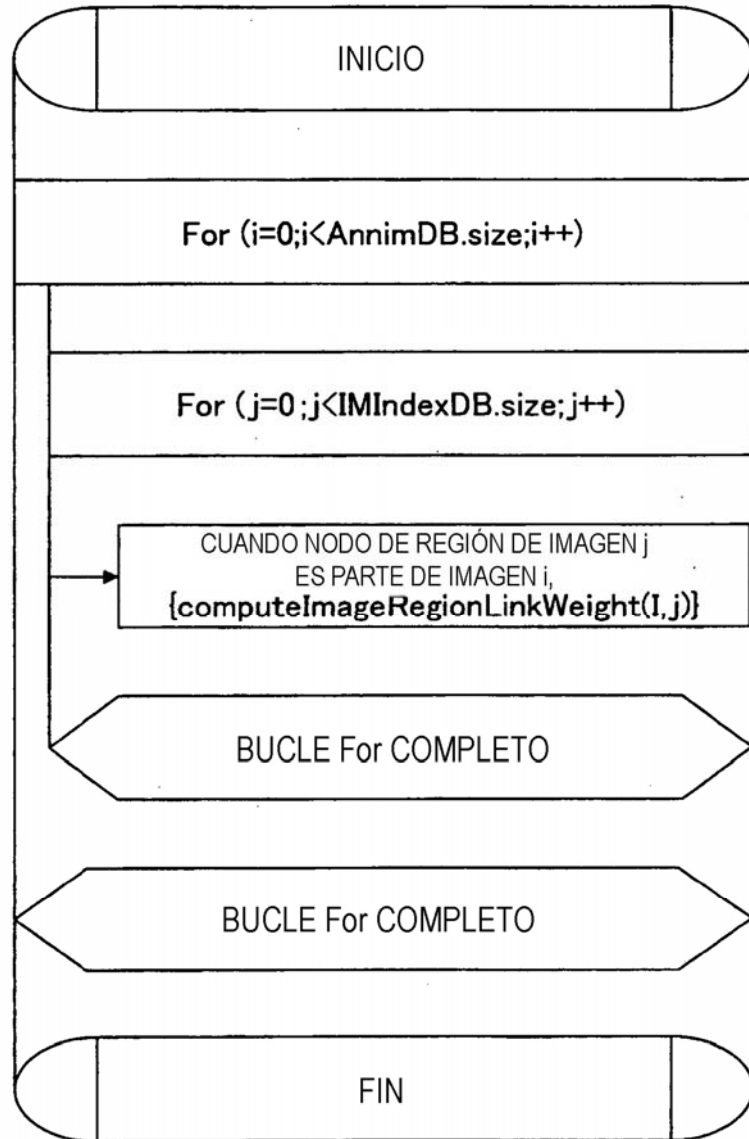




FIG.8

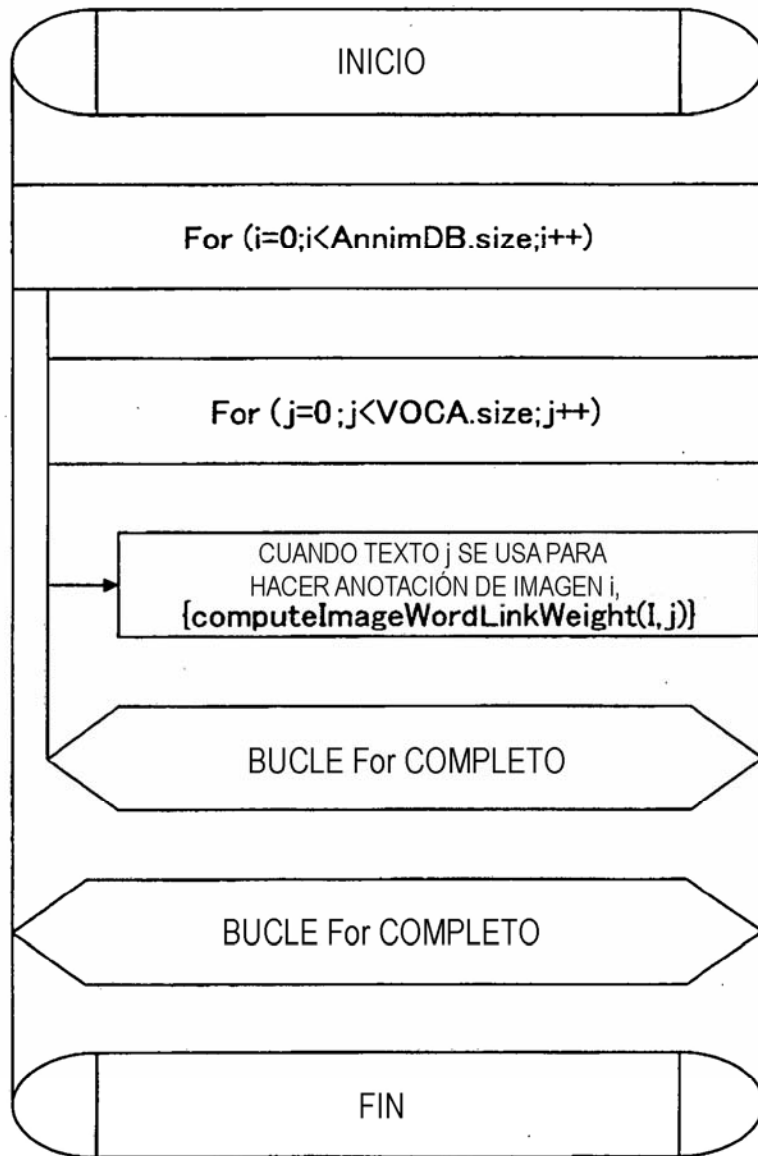


FIG.9

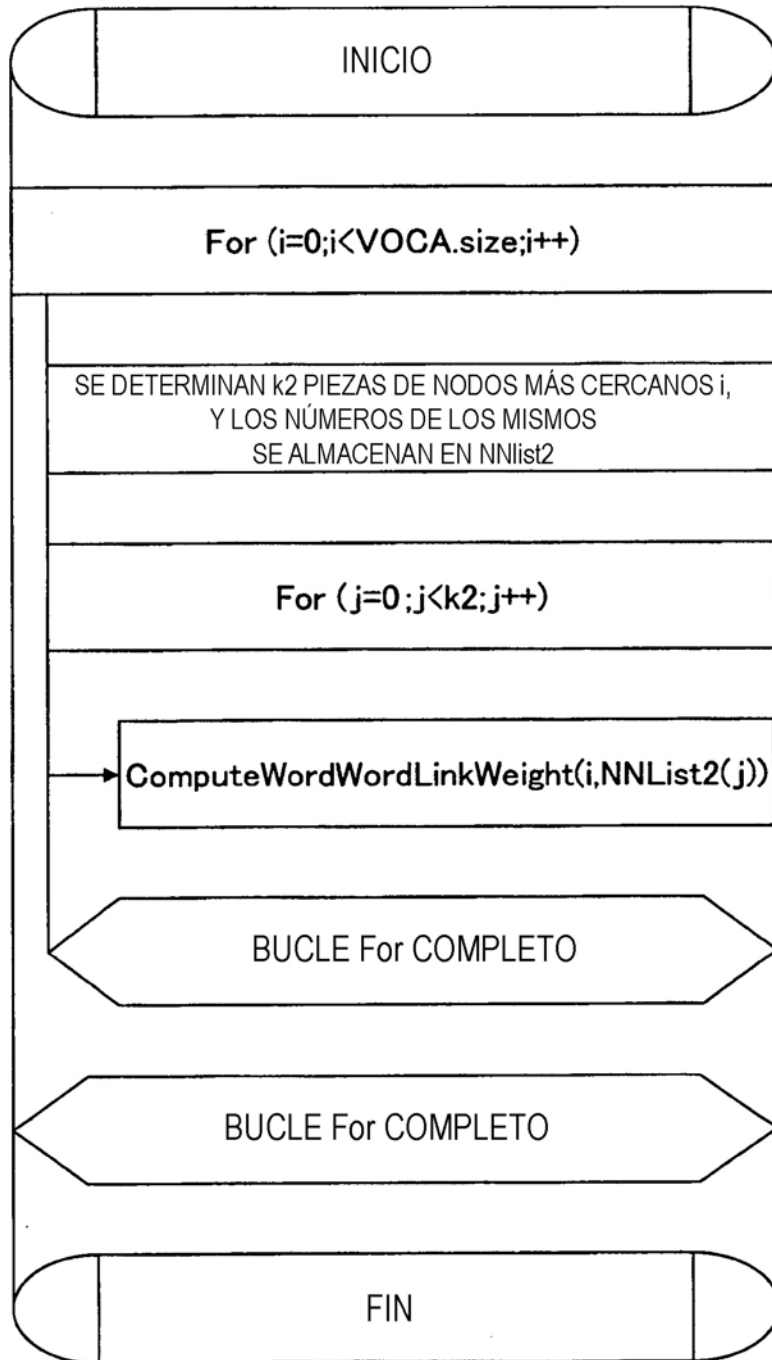


FIG.10

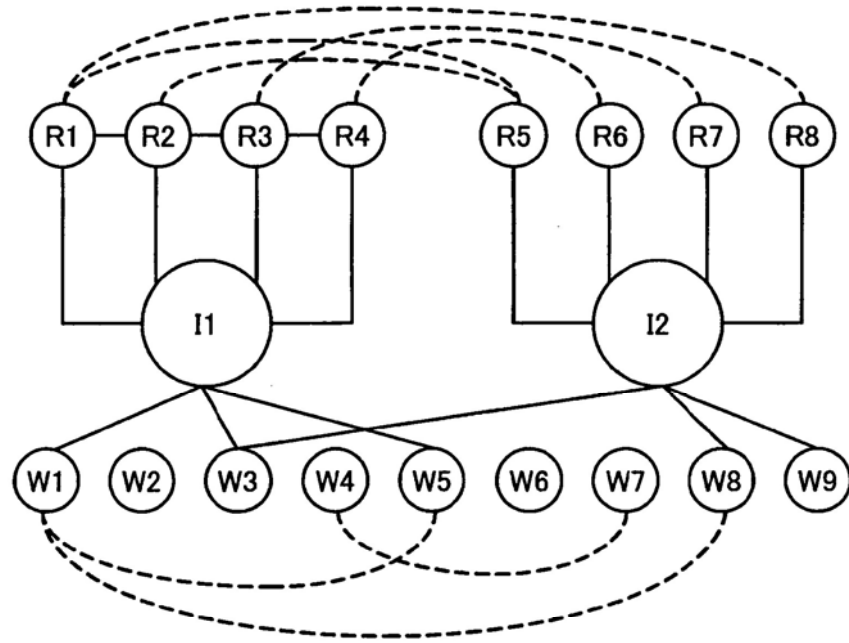
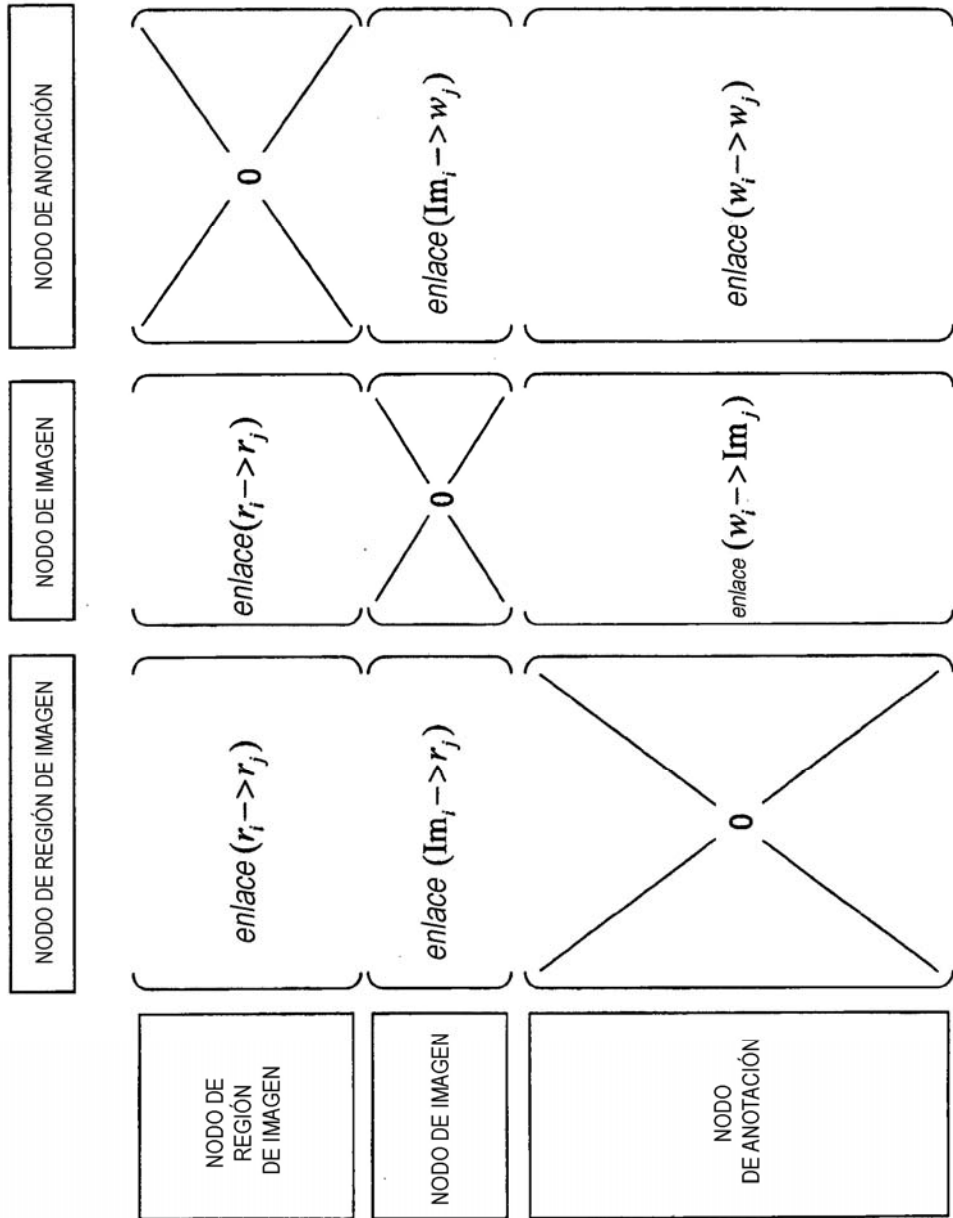


FIG.11



|     |    |    |    |    |    |    |    |    |    |    |     |
|-----|----|----|----|----|----|----|----|----|----|----|-----|
|     | R0 | R1 | R2 | I0 | I1 | I2 | W0 | W1 | W2 | Rn | Imn |
| R0  |    |    |    |    |    |    |    |    |    |    |     |
| R1  |    |    |    |    |    |    |    |    |    | 1  | 0   |
| R2  |    |    |    |    |    |    |    |    |    | 0  | 0   |
| I0  |    |    |    |    |    |    |    |    |    | 1  | 0   |
| I1  |    |    |    |    |    |    |    |    |    | 0  | 0   |
| I2  |    |    |    |    |    |    |    |    |    | 0  | 0   |
| W0  |    |    |    |    |    |    |    |    |    | 0  | 0   |
| W1  |    |    |    |    |    |    |    |    |    | 0  | 0   |
| W2  |    |    |    |    |    |    |    |    |    | 0  | 0   |
| Rn  | 1  | 0  | 1  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 1   |
| Imn | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 1  | 0   |

FIG.12

VECTOR  
DE REINICIO

|     | VECTOR<br>DE REINICIO | R0 | R1 | R2 | I0 | I1 | I2 | W0 | W1 | W2 | Rn | Imm |
|-----|-----------------------|----|----|----|----|----|----|----|----|----|----|-----|
| R0  |                       |    |    |    |    |    |    |    |    |    | 1  | 0   |
| R1  |                       |    |    |    |    |    |    |    |    |    | 0  | 0   |
| R2  |                       |    |    |    |    |    |    |    |    |    | 1  | 0   |
| I0  |                       |    |    |    |    |    |    |    |    |    | 0  | 0   |
| I1  |                       |    |    |    |    |    |    |    |    |    | 0  | 0   |
| I2  |                       |    |    |    |    |    |    |    |    |    | 0  | 0   |
| W0  |                       |    |    |    |    |    |    |    |    |    | 0  | 0   |
| W1  |                       |    |    |    |    |    |    |    |    |    | 0  | 0   |
| W2  |                       |    |    |    |    |    |    |    |    |    | 0  | 0   |
| Rn  |                       | 1  | 0  | 1  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 1   |
| Imm |                       | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 1  | 0   |

FIG.13

FIG.14



FIG.15



FIG.16



FIG.17

