(19) OFICINA ESPAÑOLA DE PATENTES Y MARCAS

ESPAÑA

(11) Número de publicación: **2 530 687**

(21) Número de solicitud: 201300816

(51) Int. CI.:

G06F 17/30 (2006.01)

(12) PATENTE DE INVENCIÓN B1

(22) Fecha de presentación:

04.09.2013

(43) Fecha de publicación de la solicitud:

04.03.2015

Fecha de modificación de las reivindicaciones:

14.07.2016

Fecha de la concesión:

11.08.2016

(45) Fecha de publicación de la concesión:

19.08.2016

(73) Titular/es:

SHOT & SHOP. S.L.
Avda. de los Prunos, 15, PE, 1º- B
28042 MADRID (Madrid) ES

(72) Inventor/es:

PÉREZ DE LA COBA, Sira

(74) Agente/Representante:

FORTEA LAGUNA, Juan José

(54) Título: **Método implementado por ordenador para recuperación de imágenes por contenido y programa de ordenador del mismo**

(57) Resumen:

Método implementado por ordenador y programa de ordenador para recuperación de imágenes por contenido.

El método comprende seleccionar y segmentar una imagen consultada; extraer características mediante el cálculo de al menos dos descriptores de características incluyendo color y textura; y determinar la similitud de la imagen consultada con una pluralidad de imágenes incluidas en una base de datos que contiene imágenes, dicha pluralidad de imágenes incluyendo también características extraídas calculadas por dichos al menos dos descriptores. Los citados descriptores de características de color y textura calculados se combinan en diferentes espacios de color, los cuales comprenden unas medidas estadísticas globales y locales en dichos espacios, proporcionando un descriptor semántico de alto nivel invariante.

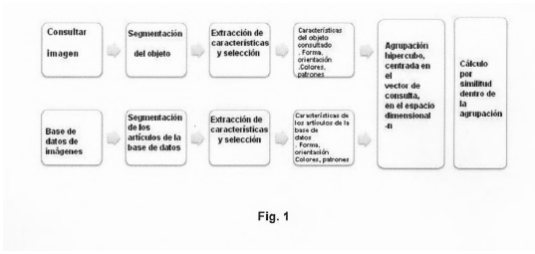


Fig. 1

ES 2 530 687 B1

Aviso: Se puede realizar consulta prevista por el art. 37.3.8 LP.

DESCRIPCIÓN

Método implementado por ordenador para recuperación de imágenes por contenido y programa de ordenador del mismo

Campo de la invención

Esta invención concierne al campo de las tecnologías para reconocimiento visual de objetos e imágenes. Particularmente, la invención concierne a un método implementado por ordenador y a un programa de ordenador para recuperación de imágenes por contenido, basado principalmente en descriptores visuales únicos. Asimismo también se utilizan herramientas basadas en una similitud subjetiva y objetiva, y en un cálculo optimizado de la distancia entre imagen consultada y potenciales similares.

Antecedentes

El reconocimiento visual es cada día más importante en nuestra sociedad debido a la implantación de los ordenadores y al ámbito virtual.

El campo de aplicación de los motores de búsqueda visual y de las tecnologías de visión artificial, reconocimiento de objetos y patrones es amplio y ya se ha extendido a una diversidad de diferentes funcionalidades y sectores tales como: visión industrial y por ordenador, navegación, control de procesos, seguridad nacional, comercio electrónico, diagnóstico médico, investigación biológica, identificación de personas y biometría, marketing, redes sociales, etc.

Especialmente el uso de la búsqueda visual para su uso en la identificación y similitud es un campo con múltiples intereses, donde las aplicaciones comerciales se han desarrollado en las últimas décadas debido al incremento de las imágenes y vídeos digitales, el uso de Internet y las últimas tecnologías de los teléfonos inteligentes, PDAs, etc. que incluyen cámaras cada vez más avanzadas.

La primera aproximación para solucionar el problema de la búsqueda visual fue la "recuperación por texto", donde las imágenes se indexan usando palabras clave, códigos de clasificación o títulos en el mensaje. Las principales limitaciones de estas tecnologías son dos; la primera, que se necesita indexar y etiquetar las imágenes, lo que supone emplear mucho tiempo y recursos, y la segunda, que no es un método estándar, ya que cada usuario puede interpretar, definir y describir subjetivamente las imágenes de forma diferente.

Una alternativa a la recuperación basada en el texto es la técnica de Recuperación de Imágenes Por Contenido (CBIR) que recupera imágenes relevantes

de forma semántica de una base de datos de imágenes basándose en características de imagen que se derivan automáticamente.

El procesamiento de imágenes es muy complejo, aparte del volumen que ocupan, existe el gran desafío de trasladar de forma eficaz percepciones de alto nivel a características de imagen de bajo nivel y solucionar el conocido intervalo semántico. Los objetivos a conseguir con estas tecnologías son:

- Debería tener un tiempo de respuesta menor
- Debería ser preciso
- Debería ser sencillo consultar la recuperación de una imagen
- Debería ser robusto e invariante a los distintos entornos, condiciones de captura de imagen y cambios de perspectiva
- Debería ser escalable a grandes bases de datos y flexible y extensible a otros tipos de objetos, imágenes y/o patrones

Dentro de los sistemas CBIR, uno de los puntos fundamentales para su buen funcionamiento es la definición y extracción de las características de las imágenes, es decir, la definición de los vectores óptimos y adecuados, también llamados descriptores, que describan de la forma más completa y precisa, la información visual de la imagen o región que representan, con la mínima cantidad de datos necesarios, y con el fin de reconocer, identificar, ordenar o clasificar la imagen u objeto de interés y/o sus similares, mediante métodos eficaces de búsqueda y comparación sobre grandes bases de datos de imágenes.

Algunas de las tecnologías desarrolladas se basan en comparaciones directas, coincidencias de patrones ("pattern matching") o métodos de correlación aplicados a las imágenes completas o a las ventanas de imágenes completas / regiones de interés (ROI). Dichos enfoques son precisos y están bien adaptados para estudiar la estructura global de un objeto concreto, previamente conocido, acotado y entrenado, o patrones estáticos fijos, pero no pueden afrontar la oclusión parcial, cambios importantes en la perspectiva u objetos deformables (de K. Grauman and B. Leibe *Chaper 3 Local Features: Detection and Description..Visual Object Recognition. Synthesis Lectures on Artificial Intelligence and Machine Learning, Morgan & Claypool (2011)*). Además, normalmente no resisten los cambios de iluminación o la presencia de ruido de elementos extraños, por lo que la flexibilidad y escalabilidad de estos sistemas es muy costosa, y por tanto, su aplicación en CBIR es bastante cuestionable.

Otro de los factores clave en la definición de los descriptores adecuados para los sistemas CBIR es que sean invariantes, lo que significa que no se vean afectados

por parámetros sensibles a las diferentes condiciones o entornos de captura de la imagen u objeto, como son la iluminación, rotación, escala, reversión, traslación, transformaciones afines, y otros efectos.

Como alternativa, se busca desarrollar sistemas CBIR donde se implementen descriptores basados en características simples (Low level features) que sean invariantes, para que, por un lado, describan de forma robusta las imágenes u objetos bajo distintos contextos y condiciones de captura, y por otro lado, eviten el uso y análisis de características de alto nivel, (High level features), que son más complejas y costosas, tanto a nivel de implementación como de consumo y procesamiento necesario.

El uso de los vectores formados por estas características de bajo nivel (Low level features) consiste en indexar propiedades visuales, usando valores numéricos para describir dichas características, representando después la imagen u objeto como un punto en un espacio N-dimensional. El proceso consiste en extraer las características del vector de la imagen u objeto consultado, y aplicar los sistemas de medición y métodos de clasificación para analizar la similitud conforme a la base de datos.

Actualmente hay métodos algorítmicos para extraer este tipo de características invariantes de las imágenes como son Scale-invariant feature transform (or SIFT), G-RIF: Generalized Robust Invariant Feature, SURF "Speeded Up Robust Features", PCA-SIFT, GLOH, etc. Sin embargo estos métodos describen la apariencia local concreta de objetos o regiones específicas de las imágenes, a partir de la selección de un conjunto de puntos de interés, obtenidos mediante métodos de entrenamiento aplicados sobre dichas imágenes, previamente conocidas y acotadas, y por tanto no siendo extensibles a otros objetos y categorías sin el entrenamiento previo correspondiente.

En este contexto, los principales desafíos afrontados, son la especificación de estructuras de indexación que permitan acelerar la recuperación de imágenes con métodos flexibles y escalables.

Así pues, otra alternativa de características de bajo nivel son los descriptores de características tipo color, forma, textura, etc., que permiten desarrollar vectores genéricos aplicables a imágenes y objetos de distinta naturaleza. Entre los métodos de optimización de dichos vectores/descriptores, se persigue obtener la máxima información con el mínimo número de parámetros o variables incluidos en los mismos, utilizando para ello métodos de selección para determinar las características más

importantes y sus combinaciones para describir y consultar artículos en grandes bases de datos, reduciendo la complejidad (de tiempo y procesamiento a nivel computacional) de búsqueda y recuperación, al tiempo que intentando mantener un alto rendimiento de precisión. Además, ayuda a los usuarios finales al asociar automáticamente las características y medidas adecuadas de una base de datos determinada (I.Guyon and A.Elisseff *.An Introduction to Variable and Feature Selection.*2003) Journal of Machine Learning Research 3 (1157-1182) .). Estos se pueden dividir en dos grupos:

- Los métodos de transformación de características tales como PCA e ICA hacen corresponder el espacio de características original con el espacio dimensional más bajo y construyen nuevos vectores de características. El problema de los algoritmos de transformación de características es su sensibilidad al ruido y las características resultantes no transmiten ningún significado para el usuario.
- Los esquemas de selección de características son sólidos contra el ruido y las características resultantes son altamente interpretables. El objetivo de la selección de características es elegir un subconjunto de características para reducir la longitud de los vectores de características con la menor pérdida de información. Los esquemas de selección de características de acuerdo con sus métodos de evaluación de subconjuntos se clasifican en dos grupos:
    - En métodos de filtrado, las características se evalúan en base a su efecto intrínseco y separación natural en clases o agrupaciones.
    - En métodos "wrapper" se aprovecha la precisión de los métodos de aprendizaje para evaluar subconjuntos de características.

Para conseguir la selección de características en sistemas CBIR, se han realizado muchos enfoques pero se basan en métodos de aprendizaje y entrenamiento por ordenador ("machine learning"), lo que supone su optimización y adaptación a un caso y universo de muestras específico, no generalmente extensible a otros casos no contemplados o categorías de imágenes y objetos de distinta naturaleza.

Entre todos los vectores, color y textura son dos de los más importantes y los que más se usan en la recuperación de imágenes y vídeos, por consiguiente, las empresas y los investigadores se han esforzado mucho por mejorarlos y basar los sistemas CBIR en ellos.

Los descriptores de color o característica de color es una característica global, que describe las propiedades de la superficie de la escena en imágenes, regiones u objetos de las mismas. Las diferentes formas de extraer las características del color se explican en Lulu Fan, Zhonghu Yuan, Xiaowei Han, WenwuHua "Overview of Content-based Image Feature Extraction Methods," International Conference on Computer, Networks and Communication Engineering. 2013.).

Los diferentes espacios de color son ampliamente conocidos por su aplicación en CBIR y por sus ventajas a la hora de identificar colores perceptivos. Ningún espacio de color puede considerarse como universal porque el color puede interpretarse y modelarse de formas diferentes. Con la gran variedad de espacios de colores disponibles (por ejemplo, RGB, CMY, Y IQ, YUV, XY Z, rg, CIE Lab, Luv, HSV, etc.) y con gran variedad de descriptores para definir los colores de las imágenes y objetos, no es obvio definir qué espacio de color y que características se deben medir para describir una imagen y poder identificar aquellas más similares. En este contexto, la inevitable pregunta que surge es cómo seleccionar el modelo de color que ofrece el mejor resultado para una tarea de visión artificial en concreto. Estas dificultades se explican en detalle en (H. Stokman and T. Gevers "*Selection and Fusion of Color Models for Image Feature Detection" IEEE transactions on pattern analysis and machine intelligence, vol. 29, no. 3, march 2007*) donde sugieren un modelo de selección genérico o modelos (invariantes).

La mayoría de los descriptores de este tipo desarrollados hasta la fecha tienen múltiples limitaciones, como se refleja en la reciente publicación *de Lulu Fan, Zhonghu Yuan, Xiaowei Han, WenwuHua "Overview of Content-based Image Feature Extraction Methods," International Conference on Computer, Networks and Communication Engineering. 2013.* Los existentes en color no suelen ser capaces de describir distribuciones locales, localizaciones espaciales y cambios de regiones en la imagen, y en definitiva no son suficientes para interpretar, reconocer, clasificar e identificar de forma unívoca objetos o imágenes concretas complejas, patrones y detalles específicos de alto nivel, así como encontrar otros cercanos y semánticamente similares. Con los de forma y textura, son necesarios procesos computacionales complejos o modelos acotados previamente entrenados. En resumen, existe pues un compromiso clave entre los objetivos perseguidos en la selección y extracción de los descriptores de los sistemas CBIR. Cuando se busca robustez, invariancia, flexibilidad y escalabilidad, se pierde precisión. Cuando se persigue precisión, se pierde robustez y extensibilidad a otros tipos de imágenes, productos o categorías.

Como solución y evolución de estos descriptores de características, surgen los denominados descriptores semánticos de alto nivel, que persiguen interpretar la información visual tal y como lo hacemos nosotros, es decir, de la forma más cercana a nuestra percepción subjetiva humana para así conseguir descriptores óptimos simultáneamente en precisión, invariancia, robustez, flexibilidad y escalabilidad como así hace nuestro cerebro al interpretar el mundo visual que nos rodea. Éstos descriptores pues, cuyo objetivo es acercarse cada vez más a la inteligencia humana, presentan barreras por su complejidad algorítmica, computacional y de almacenamiento, que hasta hoy no se han conseguido superar de forma convincente.

Análisis de similitud

La fase de análisis de similitud entre la imagen de búsqueda y recuperación y las imágenes en la base de datos es otro de los puntos críticos de los CBIR, sobre todo a nivel de escalabilidad y velocidad de procesamiento.

Obviamente este análisis de similitud dependerá en gran parte del tipo de descriptores/vectores que se estén analizando. La técnica principal y más utilizada es el análisis de la distancia entre la imagen consultada y el resto de las imágenes en la base de datos, midiendo la distancia para determinar la idéntica y sus similares.

El análisis de similitud cuantifica la semejanza en el contenido entre dos imágenes. Dependiendo del tipo de características, la formulación del análisis de similitud y su medida, varía de forma significativa.

- Algunas de las medidas de similitud empleadas para las características de color son: Medida de la Distancia Cuadrática del Histograma (*Histogram Quadratic Distance Measure,* HQDM), Coincidencia Binaria del Histograma Integrado (*Integrated Histogram Bin Matching*, IHBM), Intersección del Histograma, distancia euclídea del histograma, medición de Minkowski, distancia Manhattan, distancia Canberra, distancia Angular, coeficiente czekanonski, producto interno, coeficiente Dice, coeficiente Cosine, coeficiente Jaccard, distancia óptima de composición del color (*Optimal Color Composition Distance*, OCCD) y la distancia EMD.
- Algunas de las medidas de similitud para las características de textura son: distancia Kull back-leiber, transformación wavelet de árbol estructurado, densidad gaussiana generalizada (*Generalized Gaussian density*, GGD), método del histograma, transformación wavelet, transformación wavelet de pirámide estructurada (*Pyramid structured wavelet transform*, T'WT), modelo

autorregresivo simultáneo de multiresolución (*Multiresolution simultaneous autoregressive model*, MR-SAR), distancia euclídea ponderada, método Monte-Carlo y EMD. La distancia Kullback-Leibler proporciona una mayor precisión y escalabilidad en capturar la información de la textura.

- Algunas de las medidas de similitud para las características de forma son: distancia perceptiva, método de aproximación del polígono, método descriptivo de Fourier, Dynamic Time Wrapping (*Dynamic Time Wrapping*, DTW), distancia angular, producto interno, coeficiente Dice, distancia de Ray y correlación ordinal. DTW desarrolla un esquema de cálculo eficaz de la distancia que concuerda con el sistema visual humano en percibir la similitud de la forma.

Las principales limitaciones en el uso de métricas basadas en distancias para el análisis de similitudes son:

- Escalabilidad, dado que cuantos más descriptores tengamos, más sistemas de métricas de similitud habremos de llevar a cabo y cuanto más grande sea la recopilación de imágenes más lento será el análisis.
- Intervalo semántico. Cuando se utilizan características de bajo nivel que no representan de forma adecuada la imagen completa o el objeto final, y el sistema de análisis de similitud podría no alcanzar la verdadera interpretación visual de la imagen.

Para solucionar la limitación de los sistemas de análisis de similitud anteriores, se han propuesto métodos alternativos basados en métodos de aprendizaje combinados con análisis de similitudes para mejorar las búsquedas por semejanza y lograr un CBIR escalable. Como se explica en (G. Rafiee, S.S. Dlay, and W.L. Woo "*A Review of Content-Based Image Retrieval" CSNDSP 2010 SIP 8 (775-779) de),* existen tres marcos comunes para establecer un enlace entre las características de imagen y un concepto: aprendizaje supervisado, aprendizaje no supervisado y modelos interactivos. Los problemas de estos métodos, tanto los supervisados como los no supervisados, es que precisan aprendizaje de una forma u otra, por lo que no pueden asegurar un comportamiento preciso para todas las posibilidades que se puedan dar en espacios ilimitados y en constante cambio y crecimiento.

En resumen, a pesar de los muchos avances que se han llevado a cabo en el campo, todavía en la actualidad el principal reto en las tecnologías CBIR es lograr el objetivo de diseñar nuevos métodos para buscar y reconocer imágenes/objetos idénticos y similares, aplicables a cualquier categoría u objeto, de manera invariante y

robusta, y por consiguiente, ampliable y escalable a cualquier cambio en bases de datos ilimitadas y en continuo crecimiento, puesto que la falta de información estará siempre presente dado que se capturan objetos en 3D con cámaras individuales y no estereográficas, generando imágenes digitales en 2D, cuando además los objetos del mundo real siguen cambiando o surgen nuevos y diferentes.

Es por tanto crucial encontrar métodos alternativos a las técnicas de entrenamiento y aprendizaje por ordenador hasta ahora empleadas, ya que éstas tienen la limitación de ser válidas únicamente para aquellas bases de datos previamente conocidas y acotadas, cuando la escalabilidad y robustez es algo que se requiere cada vez más hoy en día para solucionar el problema de búsqueda de cualquier tipo de imagen u objeto e incluyendo no solo todas las posibilidades y condiciones presentes de búsqueda, sino debiendo estar preparados y adaptados a las que irán apareciendo en el futuro.

Breve exposición de la invención

Esta divulgación generalmente se refiere a un nuevo mecanismo o sistema CBIR basado en una combinación de sistemas de visión por computador y métodos estadísticos para resolver las últimas limitaciones identificadas en el estado de la técnica alcanzando los objetivos de precisión, escalabilidad y velocidad perseguidos.

En particular, la invención propone un CBIR cuyo objetivo es hallar y buscar de forma rápida y precisa imágenes/objetos idénticos y similares, siendo aplicable a cualquier categoría, tipo de imagen u objeto de distinta naturaleza, y de manera consistente e invariante a cualquier condición de captura. Se trata por tanto de una solución extensible, ampliable y fuertemente escalable a todo tipo de catálogos, clases o bases de datos en continuo crecimiento y cambio, pudiendo emplearse a través de todo tipo de dispositivos de captura, móviles, ordenadores, etc.

A tal efecto, la invención aporta un método para recuperación de imágenes por contenido que comprende, según técnicas conocidas:

a) seleccionar una imagen consultada;

b) segmentar, mediante la aplicación de una técnica de segmentación, dicha imagen consultada;

c) extraer características de dicha imagen consultada segmentada mediante el cálculo de al menos dos descriptores de características incluyendo color y textura; y

d) determinar la similitud de la imagen consultada con una pluralidad de imágenes incluidas en una base de datos, las cuales incluyen también características extraídas calculadas por dichos al menos dos descriptores.

Según la propuesta de esta invención, los citados descriptores de color y textura calculados incluyen la combinación de diferentes espacios de color, así como una medida local y global sobre los mismos, permitiendo de este modo describir no solo la precisión en el color y texturas de bajo nivel, sino además regiones diferentes, distribuciones locales, localizaciones espaciales, tejidos, entramados, estampados y hasta objetos específicos e incluso detalles precisos de alto nivel. Proporcionando un descriptor semántico de alto nivel con las ventajas de los de bajo nivel, manteniendo en todos los casos casi total invariancia a la rotación 3D, transformación afín (perspectiva, deformación, etc.), reversión, escala e iluminación, preferiblemente usando únicamente un vector de 36 valores numéricos para su representación.

Los citados descriptores de características de color y textura se pueden combinar con al menos unos descriptores de forma/geometría y orientación proporcionando un descriptor avanzado de forma/geometría y orientación también invariante a la rotación, escala, reversión, traslación e iluminación, agrupados cualitativamente para su optimización semántica.

De acuerdo a un ejemplo de realización, se puede realizar también un análisis y función de similitud objetiva con métricas n-dimensionales, empleando preferiblemente distancias Euclídea y de Mahalanobis, combinadas.

De acuerdo a otro ejemplo de realización, se pueden aplicar técnicas de transformación y métodos de filtrado para seleccionar las características optimas y eliminar redundancias. De manera preferida, se utilizan análisis estadísticos entrópicos con árboles de decisión y probabilidad, y regresiones logísticas multivariantes, evitando así el uso de métodos de aprendizaje o *wrappers*.

Finalmente, como mejora de la invención, de acuerdo a otro ejemplo de realización, se puede llevar a cabo además un análisis y definición de la función de similitud semántica basado en métricas espaciales y ponderaciones objetivas y subjetivas, y sin la necesidad de aprendizaje, entrenamiento o clasificaciones basadas en modelos predefinidos ("clústeres"). Por ejemplo, sin ser limitativo, sobre un hipercubo en tiempo real centrado en la imagen/objeto de consulta.

Breve descripción de los dibujos

La anterior y otras ventajas y características se entenderá mejor por referencia a la siguiente descripción detallada en conjunción con los siguientes dibujos adjuntos, en el cual:

Figura 1 es un diagrama de bloques que describe el método propuesto de acuerdo a varios ejemplos de realización de la presente invención.

Figura 2a ilustra algunos ejemplos obtenidos de las funciones de densidad de las variables de características para la forma y la orientación;

Figura 2b ilustra algunos ejemplos obtenidos de las las funciones de densidad de las variables de características para el color y la textura.

Figura 3 ilustra un ejemplo de un árbol de partición regresiva basado en entropía utilizado en un ejemplo de realización de la presente invención.

Figura 4 ilustra un ejemplo de clúster tipo hipercubo adaptado definido según un ejemplo de realización de la presente invención.

Descripción detallada de la invención

Con referencia a la Figura 2 se muestra el proceso completo, según varios ejemplos de realización, del mecanismo CBIR propuesto, que incluye análisis, diseño e implementación. El citado mecanismo puede dividirse principalmente en tres etapas: Tratamiento de imagen y algoritmos de visión artificial para la extracción de características definidas por uno mismo sin utilizar métodos automatizados o de aprendizaje; Análisis estadístico avanzado y métodos de selección y agrupación cualitativa entrópica para una evaluación, selección, extracción y compresión optimizada de las características a incluir en los descriptores semánticos planteados; e implementación de la función de similitud objetiva basada en distancias Euclídea y/o Mahalanobis.

Tratamiento de imagen y algoritmos de visión artificial

El objetivo es definir un método lo más genérico y escalable posible con el propósito de que sea aplicable, y extensible a todo tipo de imagen, objeto o Región de interés (ROI), desde la detección, la segmentación y el reconocimiento, manteniendo así la flexibilidad, interoperabilidad y escalabilidad del sistema en todo momento.

En primer lugar, se realiza una segmentación usando técnicas conocidas en el estado del arte. Con el objeto segmentado resultante se aplica una normalización del área del objeto simple hasta el tamaño mínimo necesario, para mantener todos los objetos con tamaños similares con el objetivo de mantener la escala invariante todo lo

que sea posible, sabiendo que cualquier función continua vs discreta no puede ser invariante 100% a escala debido a los principios teóricos de la información y a las señales físicas análogas vs su representación digitalizada.

Seguidamente, para la extracción de características, se calculan unos descriptores de características locales/globales iniciales (preferiblemente, forma, orientación, color y textura), identificando los parámetros y características más invariantes a partir de métodos de transformación de imágenes, a la vez que los que mejor representan y describen las características objeto. El primer enfoque es la optimización en la invariancia a la escala, iluminación, rotación, traslación y reversión. La máxima información provista en cada descriptor es el segundo objetivo. Para la forma, se extraen variables geométricas matemáticas tanto para la descripción de la forma local como de la global incluyendo: momentos estadísticos invariantes, excentricidades, convexidades, áreas, perímetros y sus relaciones, etc. hasta alcanzar un total de alrededor 30-40 variables y medidas incluidas inicialmente en el descriptor. Para la orientación, se incluyen el cálculo del PCA y de otros ratios basados en formas evolventes del objeto. Para el color y la textura se definen igualmente diversos momentos estadísticos, tanto medidas globales como locales, incluyendo al menos la media y la variancia (desviación típica) de diversos espacios de color invariantes, hasta alcanzar también unas 30-40 variables.

Para la extracción del descriptor de características de la forma se miden todas las características transformadas invariantes geométricas que puedan contribuir de forma eficaz a la descripción de la forma tanto para descripciones de bajo como de alto nivel.

Para medir la réplica más significante y fiel del mundo real tanto como sea posible, y saber que los objetos físicos/en 3D se proyectan a representaciones de imágenes digitales en 2D y, por consiguiente, siempre se pierde gran parte de la información principal del objeto, se buscan parámetros con el objetivo de mantener invariante la rotación en 3D todo lo que sea posible así como la reversión y el giro. Por consiguiente, se empieza utilizando los momentos de Hu invariantes y de bajo nivel y otros parámetros lineales y adimensionales relacionados con el volumen del objeto, basados en envolventes convexas, excentricidades, otras formas derivadas y sus ratios relacionados con el objeto.

Para los contornos internos y otras variables de alto nivel, se usan perímetros equivalentes basados en ratios y formas locales detalladas. Para bordes internos y

externos, convexidades, excentricidades, etc., se calculan las áreas de los defectos y entrantes de convexidad y los ratios lineales relacionados.

Para la extracción del descriptor de orientación la invención se basa en el valor del ángulo del primer componente del PCA y/o otros cálculos de ratios de ángulos relacionados a partir de las longitudes de rectángulos derivados vs ratios de altura, entre otras formas, ángulos y ratios.

De forma conjunta, las extracciones de características iniciales de forma y orientación preferiblemente hacen un total de 35 variables.

Para la extracción del descriptor semántico de color, textura, patrón y detalles se miden las características transformadas que puedan describir de forma eficaz colores y texturas dentro de objetos, manteniendo todo lo que sea posible la iluminación y la invariancia del color que no solo está presente debido a condiciones diferentes de captura de imagen sino debido también a proyecciones de 3D a 2D, incluyendo sombras, brillos, efectos de perspectiva del volumen, etc. Por consiguiente, la invención trabaja con diversos espacios de color diferentes: HSV; CieLAB y CieXYZ, debido a sus atributos adecuados para la percepción basados en la invariancia del color y la iluminación y para compensar sus limitaciones en la medición del color entre sí.

Además, con el fin de medir no solo las características de color/textura de bajo nivel sino las de alto nivel y, por consiguiente, combinar medidas globales con las de espacio local en el mismo descriptor, también se pueden medir los momentos estadísticos principales para el área completa del objeto así como para ROIs locales de menor tamaño. Para medir de este modo las representaciones de color y textura, inicialmente se calculan los momentos estadísticos generales del área global, incluyendo la media, varianza y desviación de los tres espacios de color invariantes en los que es más fácil medir el color predominante (con la media) y las texturas predominantes (con la desviación). Con el fin asimismo de poder diferenciar diversas regiones diferentes de color/textura en un objeto, también se puede incluir una pequeña ROI específica local de 9x9 píxeles en el objeto completo de la imagen, cuyos valores de varianza y media se aproximan a los valores globales, con una distancia Euclídea.

La ROI elegida en este ejemplo de realización es de 9x9 ROI y es la que cumple la distancia más cercana a los valores globales, distancia (ROI, Objeto), es decir:

$$ROI_{(xi,yi)} \mid \min[\![distance([\mu_{ROI}, \sigma_{ROI}], [\mu_{object}, \sigma_{object}])]\!]$$

$$= \min \left[\sqrt{(\mu_{ROI} - \mu_{object})^2 + (\sigma_{ROI} - \sigma_{object})^2} \right]$$

En conjunto, la extracción de características iniciales de color y textura se compone preferiblemente de 36 variables.

Análisis estadístico y métodos de clasificación para la compresión y selección optimizadas de características

Una vez obtenidas las características iniciales de extracción de variables a incluir en el vector, se analizan a fondo desde un punto de vista estadístico y entrópico para obtener información teórica útil y optimizada a seleccionar, con el fin de decidir cuales representan una buena descripción de las características, y eliminar aquellas que no aportan información sino ruido o información redundante. Por otro lado, el objetivo es evitar preparar un determinado conjunto de imágenes o artículos, ya que el objetivo es que estas características puedan describir cualquier tipo de objeto y dentro de cualquier tipo de categoría de diversa naturaleza.

Bajo estas premisas iniciales, se analizan los datos de las características. Entonces se obtienen los vectores definidos de las características para una base de datos grande de calzado, por ejemplo, sin ser limitativo a este tipo de bases de datos, (más de 30.000 ejemplos) ya que representan de forma adecuada otras categorías de objetos y una gran variedad de formas complejas así como colores, texturas y además, no sólo para determinadas regiones uniformes sino también para diferentes regiones y tejidos dentro del mismo objeto.

Para los descriptores de las características extraídas inicialmente en la etapa anterior, se seleccionan las características aplicando prácticamente los mismos pasos estadísticos, aunque obviamente se generaran unos resultados y unas conclusiones diferentes para cada uno de ellos, y se revisan sus planos de densidades. Las Figuras 2a y 2b muestran algunas de las funciones de densidad obtenidas.

Después se aplican técnicas de filtrado y transformación basadas en métodos estadísticos de clasificación con árboles de partición basados en entropía, utilizando sólo estos métodos con el fin de evaluar si las variables extraídas son suficientes y adecuadas para describir cualitativamente y de forma completa las características semánticas definidas.

En base a métodos de modelización en "clústeres", clasificación en conjuntos y métodos de estimación de densidad, se busca el número óptimo de clases en las que

se puede dividir una base de datos de imágenes genérica, manteniendo su efecto intrínseco original, y sin ningún método de inducción o aprendizaje.

Para el número óptimo de clases resultantes obtenido, se clasifica un subconjunto representativo adecuado de toda la base de datos, para obtener varios subgrupos representativos clasificados aleatoriamente de todo el conjunto de imágenes, a los que aplicar los métodos de partición regresiva y clasificación entrópica y extraer el vector descriptivo más adecuado.

Preferiblemente, se emplean dos árboles principales de partición regresiva cómo métodos basados en la entropía:

- Método R-part con podado: árbol de regresión con partición recursiva (Leo Breiman 1984)
- J48 – C4.5 con árboles de clasificación talados o no talados (Quinlan, 1993).

La Figura 3 muestra un ejemplo de los resultados de los métodos de clasificación. De estos primeros resultados, se pueden extraer algunas conclusiones estadísticas:

• Para el descriptor de características de orientación y forma, se alcanza un 99,61% de la información entrópica a partir de las variables definidas. Esto quiere decir que se han podido medir y describir completamente los objetos por su forma externa y sus características de orientación basadas en este descriptor.

• Para el descriptor de características de color y textura, el porcentaje de entropía alcanzado es de un 98,92 % del cual se puede considerar que esta característica describe completamente, como mínimo los colores y texturas de dos regiones, para todo tipo de objetos o regiones de interés.

Una vez analizado entrópicamente los citados vectores, se realiza un análisis cualitativo de las características para su compresión y agrupación semántica cualitativa.

De modo que se analizan las variables resultantes incluidas en cada descriptor desde un punto de vista cualitativo y detectando además la información redundante que se pueda reducir, descartar, agrupar y/o comprimir.

En base a un análisis factorial y análisis de correlaciones de partida aplicado a ambos descriptores se obtiene que se puede reducir los vectores actuales a vectores

de 15-20 variables sin perder más de un 2% de información. Para hacer esto, se aplica primero un análisis factorial para obtener los correspondientes conjuntos ortogonales y entender las agrupaciones cualitativas por variables que se pueden definir.

Así, para poder reducir el número de variables sin aplicar un método de aprendizaje a un caso particular y que dicha reducción sea extensible a cualquier otro conjunto o base de datos, categoría de distinta naturaleza y de escalabilidad ilimitada, se debe analizar cada grupo de variables incluida en cada factor, de forma independiente, descartando aquellas variables que proporcionen menos información, y cuyo impacto en la pérdida de información total no supere el 1%. Además, para entender la percepción humana, se interpretan cualitativamente los grupos de factores resultantes obtenidos con el fin de separarlos como factores semánticos propios, y si fuera necesario en diferentes descriptores de características de cara a su adaptación y ponderación subjetiva posterior:

Para el descriptor de forma y orientación, las características cualitativas independientes definidas en base a los factores obtenidos y su interpretación cualitativa fueron:

- Volumen, Orientación, Entrantes y Convexidades:
  Mientras que el volumen y la orientación son características globales, los entrantes y las convexidades son características locales, proporcionando al descriptor completo un enriquecimiento en precisión a la vez que en flexibilidad, debido a la mezcla de las descripciones globales y locales en un mismo vector. Desde el punto de vista de la percepción humana, se podría dividir los descriptores originales en dos: orientación y forma geométrica (incluyendo volumen, entrantes y convexidades).

Para el descriptor de color y textura, las características cualitativas independientes extraídas fueron:

- Colores, Texturas, Estampados y Detalles:
  Observando en este caso que mientras los colores y las texturas son características globales, los estampados y detalles concretos son locales. Además, debido a su equilibrio global/local, se reconoce y describe más de un color, textura o entramados incluidos en la imagen u objeto. Para la percepción semántica humana, se podría dividir los descriptores originales en dos: colores y estampados/entramados (incluyendo texturas, estampados, entramados y detalles).

Por lo tanto, de esta etapa se deduce que los descriptores originales pueden descomponerse en otros descriptores de características independientes y, por lo tanto, tratarse por separado para mejorar la precisión semántica visual basada en la percepción subjetiva humana o usarse para la selección independiente de otros fines y aplicaciones de los usuarios. Además, y en base a estas agrupaciones alcanzadas y a la reducción de información redundante a partir de dichos métodos combinados con regresiones logísticas multivariantes, se consigue:

- Un descriptor avanzado semántico de alto nivel para la descripción de forma, geometrías y orientación, compuesto por 22 variables, robusto, invariante y preciso, mejorando así la interpretación y descripción inicial obtenida de las 35 características originales.
- Un descriptor avanzado semántico de alto nivel para la descripción de colores, texturas, estampados y detalles, compuesto por 36 variables, robusto, invariante y preciso, mejorando la interpretación y descripción inicial esperada, y sin necesidad de ninguna agrupación semántica subjetiva específica.

Definición de función de similitud escalable

Tal como se ha indicado anteriormente, la precisión es uno de los principales problemas de los sistemas de semántica CBIR y similitudes inducidas. Esto ocurre debido la diferencia entre los conceptos de alto nivel percibidos por el ojo humano y los descriptores de bajo nivel usados frecuentemente. Además, debido al uso de métricas espaciales y distancias, se está reduciendo un espacio de información n-dimensional a una variable unidimensional (la distancia) para solicitar los resultados más similares a la imagen de consulta y, por lo tanto, ocurre con frecuencia que muchos resultados distintos se cruzan con los similares, generando así "ruido" entre los resultados cercanos y similar, ya que, aunque desde un cálculo matemático de bajo nivel sí lo son, desde la percepción subjetiva humana de alto nivel, no.

Por otro lado, cuando se calculan las distancias 1:1 entre la imagen de consulta y cada imagen del conjunto de imágenes, la escalabilidad no es lineal sino exponencial y por tanto la recuperación de grandes bases de datos llega a ser un problema de limitación de velocidad y procesamiento computacional. Esto se resuelve normalmente mediante métodos entrenados de clasificación y modelizado en "clústeres", realizando sólo cálculos de distancia 1:1 de la imagen de consulta con los centroides correspondientes de cada clase. La limitación en este caso es que la

clasificación por "clústeres" se define de forma fija y optimizada a partir de métodos de aprendizaje y entrenamiento, únicamente válidos para una base de datos conocida específica. Además las posibilidades de error de clasificación son mayores al tener que agrupar por clústeres, representados únicamente por sus modelos centrales, una gran variedad, generalmente dispersa, de muestras, casos y puntos del universo n-dimensional. En cambio, para los universos de bases de datos ilimitados y en continuo crecimiento, estos métodos de clasificación no son válidos ya que tendrían que actualizarse y volver a entrenarse continuamente para mantener los niveles esperados de precisión, generando así costes y tareas adicionales, que impiden el crecimiento y escalabilidad lineal

Para resolver o mejorar todas las limitaciones mencionadas, se define, de acuerdo a un ejemplo de realización, una nueva función de similitud, que no está basada en métodos de clasificación predefinidos ni en los sistemas de métricas espaciales optimizados mediante aprendizaje por ordenador, sino que está basado en cuatro pilares principales:

• Análisis de similitud basada en al menos 4 descriptores semánticos de alto nivel con características combinadas globales y locales.

• La creación de un "subconjunto personalizado" en tiempo real basado en el cálculo de un hipercubo adaptado. Centrado en la posición de la imagen de consulta en el espacio n-dimensional y calculando solo distancias 1:1 para los artículos incluidos dentro de este conjunto, con el propósito de aumentar la velocidad de búsqueda (sólo aplicada dentro de dicho espacio) y permitir la escalabilidad para bases de datos a gran escala en continuo cambio o crecimiento.

Con referencia a la Figura 4 se muestra el clúster de tipo hipercubo adaptado. En este caso, se define por un vector de radio n-dimensional del total de n-dimensiones de todos los descriptores juntos, con valores de radio independientes y diferentes para cada dimensión, y se basa en un porcentaje de distancia aceptada del valor de la posición de la imagen de consulta, limitada también por el cálculo de velocidad máxima permitida para calcular todas las distancias 1:1 necesarias dentro de la agrupación.

Además, y para realizar esta consulta n-dimensional, simultánea e independiente de cada dimensión de la manera más rápida, se hace un cambio en el factor de la escala para que todos los valores de los vectores del radio sean enteros (ya que los sistemas de bases de datos hacen consultas más

rápidas cuando trabajan solo con números enteros) para favorecer matemática y geométricamente consultas simultáneas de multidimensionales a unidimensionales para simplificar los procesos de las bases de datos y reducir parte del invariante inicial provocado por la distancia de proyección n-dimensional a 1-dimensional.

• La clasificación por similitud dentro del conjunto adaptado, basada en distancias combinadas de Mahalanobis y Euclídea, se aplica independientemente a cada descriptor de características semánticas por separado, y a su posterior ponderación en la función de semejanza visual subjetiva:

$$Similitud\ (distancia) = \alpha \cdot [d_{mah}(orientación + forma)] + (1-\alpha) \cdot [d_{mah}(colores + entramados)]$$

$$donde$$
$$\alpha = W_{forma}$$
$$1-\alpha = W_{entramados}$$

Para las distancias Mahalanobis y Euclídea, las variables tienen que estar normalizadas previamente en un intervalo, así se calculan los vectores máximo y mínimo del conjunto de toda la base de datos, y después, sus correspondientes matrices Mahalanobis (la inversa de la matriz de covarianza). Estas matrices son estándar para cualquier descripción de objeto y categoría ya que representan la intercorrelación entre las variables dentro de cada descriptor, donde son independientes del tipo de objeto y categoría que describen. Debido a esto, será posible la escalabilidad desde el punto de vista de la algoritmia.

Para el cálculo de α, se aplica un método de aprendizaje inicial supervisado apoyada en una regresión logística multivariante para adecuarse a la percepción visual humana basándose en dos criterios:

- α| Se maximiza la posición de ordenación de la función de similitud para los resultados idénticos a la imagen de consulta
- α | Se minimiza la invariancia percibida en la ordenación de la función de similitud semántica subjetiva obtenida

Para mejorar el presente descriptor de patrones, la invención permite la inclusión de nuevos espacios-color invariantes adicionales (por ejemplo, RGB, CMYK, otros CIEs, etc.), cálculos adicionales de ROIs espaciales (por ejemplo, ROI 50%, ROI

25%, etc.) y otras dependencias de similitud (por ejemplo, la más cercana, la más lejana, en el medio).

Además, y para probar los beneficios reales de la metodología propuesta en términos de requisitos de escalabilidad estimada, entre otros, se han realizado ensayos estadísticos detallados para probar las propiedades de invariancia requeridas, velocidad y solidez.

Ensayos y ejemplos realizados:

En cuanto a velocidad, todos los ensayos realizados consiguieron búsquedas de consulta de menos de 0,6 segundos/búsqueda y en un conjunto de base de datos de 30.000 imágenes.

En cuanto a la precisión, se midió la máxima información entrópica conseguida por cada descriptor:

Vector de características de orientación y forma (15 variables seleccionadas)

=== Resumen ===

Casos clasificados correctamente	99,6129 %
Casos clasificados incorrectamente	0,3871 %
Estadística Kappa	0,9956
Error aleatorio	0,0012
Error cuadrático medio	0,0241
Error absoluto relativo	0,6564 %
Error cuadrático medio	8,1031 %
Cobertura de casos (nivel 0,95)	100 %
Tamaño de la región media relativa (nivel 0,95)	10,2194 %

Vector de características de colores, texturas y patrones (tejido) (36 variables seleccionadas)

=== Resumen ===

Casos clasificados correctamente	98,9204 %
Casos clasificados incorrectamente	1,0796 %
Estadística Kappa	0,974
Error aleatorio	0,0053
Error cuadrático medio	0,0516
Error absoluto relativo	3,5393 %
Error cuadrático medio	18,8267 %
Cobertura de casos (nivel 0,95)	100 %
Tamaño de la región media relativa (nivel 0,95)	9,1797 %

Se calculó la α óptima para la función de similitud para optimizar:

- mayor % de apariencia del objeto consultado en los primeros resultados
- uniformidad máxima en la similitud de la percepción subjetiva en seres humanos para los resultados de un objeto consultado

Con los siguientes índices de acierto %:

Para un subconjunto aleatorio de 1033 elementos, con una función de similitud. Probado en más de 112 subconjuntos de muestras en diferentes vistas (imagen reverso/giro, girada, etc.)

TOP 1: 90,18%
TOP 2: 91,07%
TOP 3: 92,86%
TOP 4: 94,64%
TOP 5: 94,64%
TOP 6: 95,54%
TOP 7: 95,54%
TOP 8: 95,54%
TOP 9: 95,54%
TOP 10: 95,54%

Para un subconjunto aleatorio de 1033 elementos, sólo con las características optimizadas de estampado y color. Probado en más de 112 subconjuntos de muestras en diferentes vistas (imagen reverso/giro, girada, etc.)

TOP 1: 76,22%
TOP 2: 81,10%
TOP 3: 82,93%
TOP 4: 85,98%
TOP 5: 86,59%
TOP 6: 89,63%
TOP 7: 92,07%
TOP 8: 92,68%
TOP 9: 93,29%
TOP 10: 93,29%

Para un subconjunto aleatorio de 1033 elementos, sólo con las características optimizadas de forma y orientación. Probado en más de 112 subconjuntos de muestras en diferentes ángulos de orientación (imagen reverso/giro, girada en 2D)

TOP 1: 75,89%
TOP 2: 82,14%
TOP 3: 83,93%
TOP 4: 85,71%
TOP 5: 86,61%
TOP 6: 87,50%
TOP 7: 88,39%
TOP 8: 88,39%
TOP 9: 88,39%
TOP 10: 88,39%

Se puede concluir que ambas características contribuyen a la precisión de similitud global y tienen una precisión similar inicial para encontrar el objeto idéntico en el primer resultado, pero donde la característica de color y estampado es claramente

más fuerte ya que incrementa el % de índice de aciertos desde el top 1 al top 10, un 22% más mientras que para la forma y la orientación el incremento es del 16%.

Se calculó también asimismo el porcentaje de los índices de acierto, para otras condiciones de imágenes de consulta para verificar la solidez de la exactitud y la flexibilidad de la rotación en profundidad a escala 3D:

Para un subconjunto aleatorio de 1033 elementos, con una función de similitud. Probado en más de 112 subconjuntos de muestras a escala reducida (50% de la imagen original)

TOP 1: 94,64%
TOP 2: 97,32%
TOP 3: 97,32%
TOP 4: 97,32%
TOP 5: 97,32%
TOP 6: 97,32%
TOP 7: 97,32%
TOP 8: 97,32%
TOP 9: 97,32%
TOP 10: 97,32%

Para un subconjunto aleatorio de 1033 elementos, con una función de similitud. Probado en más de 112 subconjuntos de muestras a escala reducida (25% de la imagen original)

TOP 1: 93,75%
TOP 2: 94,64%
TOP 3: 94,64%
TOP 4: 95,54%
TOP 5: 95,54%
TOP 6: 95,54%
TOP 7: 95,54%
TOP 8: 96,43%
TOP 9: 96,43%
TOP 10: 96,43%

El CBIR propuesto es sólido para la variación de reducción de la escala a pesar del efecto de pérdida de muestras digitales.

Para un subconjunto aleatorio de 1033 elementos, con una función de similitud. Probado en más de 112 subconjuntos de muestras para una vista girada en profundidad en 3D.

TOP 1: 25,00%
TOP 2: 33,93%
TOP 3: 39,29%
TOP 4: 44,64%

TOP 5: 46,43%
TOP 6: 48,21%
TOP 7: 49,11%
TOP 8: 51,79%
TOP 9: 54,46%
TOP 10: 54,46%

El CBIR propuesto no es tan fuerte en la variación de la rotación profunda en 3D como se deseaba debido a la distorsión afín. Sin embargo, es una buena aproximación inicial si en los 10 primeros resultados, más de la mitad de los casos se identifican adecuadamente desde esta vista. De modo que, para por ejemplo mejorar esta variación de la rotación profunda en 3D y en conjunto mejorar y hacer que la invención propuesta sea más fuerte se realizarán nuevas acciones u etapas del método propuesto.

En resumen, la presente invención proporciona un nuevo método CBIR más rápido, preciso y escalable por las siguientes razones:

Selección propia de descriptores con apariencia inicial de bajo nivel en cuanto a su sencillez computacional y su reducido tamaño dimensional vectorial, siendo invariantes, robustos, flexibles y escalables debido a su implementación mediante métodos transformados y conservando su efecto intrínseco original, sin aplicar métodos de inducción o entrenamiento forzado, conservando así la extensibilidad y escalabilidad, y por tanto, el buen funcionamiento con cualquier tipo de objeto, categoría o imagen recuperada.

Aplicación de técnicas de transformación y métodos de filtrado para la selección de las características optimas y la eliminación de redundancias, mediante análisis estadísticos entrópicos con árboles de decisión y probabilidad, análisis factorial y de correlación, y regresiones logísticas multivariantes, evitando así el uso de métodos de aprendizaje o *wrappers*.

Obtención resultante de descriptores semánticos de alto nivel en cuanto a su precisión alcanzada y a su interpretación compleja y combinada de características globales y locales de forma simultánea, capaces de reconocer cosas mucho más complejas y detalladas que lo que inicialmente parecen describir, de forma precisa y eficaz, consiguiendo tasas de acierto altas que permiten interpretar conceptos semánticos cercanos a los humamos de forma directa y sencilla.

Análisis y función de similitud con métricas n-dimensionales, empleando distancias Euclídea y de Mahalanobis, combinadas, y cuya ponderación subjetiva, se extrae inicialmente mediante estimaciones orientativas generalizadas y el uso de regresiones lineales.

Además de la posibilidad de combinación y adaptación posterior de la función de semejanza subjetiva, mediante métodos de aprendizaje no supervisado. Y al incremento de la eficacia y rapidez debido a la mejora supuesta por la implementación del hipercubo adaptado y centrado en la imagen de consulta.

El alcance de la invención queda definido por el siguiente conjunto de reivindicaciones.

REIVINDICACIONES

1. Método implementado por ordenador para recuperación de imágenes por contenido, del tipo que comprende:

a) seleccionar una imagen consultada;

b) segmentar, mediante la aplicación de una técnica de segmentación, dicha imagen consultada;

c) extraer características de dicha imagen consultada segmentada mediante el cálculo de al menos dos descriptores de características incluyendo color y textura combinados con descriptores de forma y orientación, en el que el descriptor de forma utiliza los momentos de Hu invariantes; y

d) determinar la similitud de la imagen consultada con una pluralidad de imágenes incluidas en una base de datos, las cuales incluyen también características extraídas calculadas por dichos al menos dos descriptores,

en el que en dicha etapa c) dichos descriptores de características incluyen la combinación de diferentes espacios de color, los cuales comprenden unas medidas estadísticas globales y/o locales en dichos espacios, dando como resultado al menos dos descriptores semánticos de alto nivel, invariantes.

2. Método implementado por ordenador según la reivindicación 1, caracterizado porque comprende además calcular una región de interés (ROI) de un determinado tamaño de píxeles de la imagen consultada para diferenciar diferentes regiones de color o textura en la imagen consultada, en donde dicha ROI se calcula midiendo media y varianza donde la ROI cumple la distancia más cercana a los valores globales.

3. Método implementado por ordenador según la reivindicación 2, caracterizado porque dicho tamaño de píxeles es al menos 9x9.

4. Método implementado por ordenador según la reivindicación 1, caracterizado porque dichos espacios de color diferentes comprenden al menos HSV, CieLAB o CieXYZ.

5. Método implementado por ordenador según la reivindicación 1, caracterizado porque dicho descriptor de características de forma comprende utilizar los momentos de Hu invariantes y de bajo nivel y otros parámetros no lineales y adimensionales relacionados con el volumen del objeto, basados en áreas como las envolventes

convexas, excentricidades, otras formas derivadas y sus ratios relacionados con el objeto.

6. Método implementado por ordenador según la reivindicación 5, caracterizado porque comprende calcular:
- perímetros equivalentes basados en ratios y formas locales detalladas para los contornos internos y otras variables de alto nivel; y/o
- áreas defectuosas de convexidad y ratios lineales relacionados para los bordes internos y externos, convexidades y excentricidades.

7. Método implementado por ordenador según la reivindicación 1, caracterizado porque dicho descriptor de características de orientación se calcula al menos en base a un valor del ángulo del primer componente del PCA.

8. Método implementado por ordenador según la reivindicación 1, caracterizado porque dicha similitud se determina en base a una medición del espacio aplicada dentro de una agrupación hipercubo adaptada en tiempo real definida por un vector de radio n-dimensional del total de n- dimensiones de dichos dos descriptores juntos, con valores de radio independientes y diferentes para cada dimensión, en base a un porcentaje de distancia aceptada del valor de la posición de la imagen de consulta y limitada por el cálculo de velocidad máxima permitida para calcular todas las distancias 1:1 necesarias dentro de la agrupación, dicha agrupación hipercubo estando centrada en la posición del espacio de la imagen consultada.

9. Método implementado por ordenador según la reivindicación 8, caracterizado porque dicho vector de radio se calcula para cada dimensión basándose al menos en la posición de dicha imagen consultada pasando de forma matemática y geométrica de un espacio multidimensional a uno de una dimensión.

10. Método implementado por ordenador según la reivindicación 1, caracterizado porque los descriptores de dicha etapa c) se optimizan mediante la aplicación de técnicas de transformación y métodos de filtro basándose en métodos de clasificación estadísticos aplicados en un análisis de regresión logística multi-variante.

11. Método implementado por ordenador para recuperación de imágenes por contenido, del tipo que comprende:

a) seleccionar una imagen consultada;

b) segmentar, mediante la aplicación de una técnica de segmentación, dicha imagen consultada;

c) extraer características de dicha imagen consultada segmentada mediante el cálculo de al menos dos descriptores de características incluyendo color y textura; y

d) determinar la similitud de la imagen consultada con una pluralidad de imágenes incluidas en una base de datos, las cuales incluyen también características extraídas calculadas por dichos al menos dos descriptores,

en el que en dicha etapa c) dichos descriptores de características incluyen la combinación de diferentes espacios de color, los cuales comprenden unas medidas estadísticas globales y/o locales en dichos espacios diferentes que comprenden la media y la varianza de la imagen consultada para dichos espacios de color, dando como resultado un descriptor semántico de alto nivel, invariante.

12. Programa de ordenador que comprende código adaptado para realizar las etapas c) y d) del método según la reivindicación 1 u 11 cuando se ejecuta en uno cualquiera de un dispositivo de computación, o en un procesador digital de señal, o en un circuito específico integrado, o en un microcontrolador u otra forma de sistema de procesamiento de datos.


Consultar imagen
Segmentación del objeto
Extracción de características y selección
Características del objeto consultado . Forma, orientación .Colores, patrones
Agrupación hipercubo, centrada en el vector de consulta, en el espacio dimensional -n
Cálculo por similitud dentro de la agrupación
Base de datos de imágenes
Segmentación de los artículos de la base de datos
Extracción de características y selección
Características de los artículos de la base de datos . Forma, orientación Colores, patrones

Fig. 1

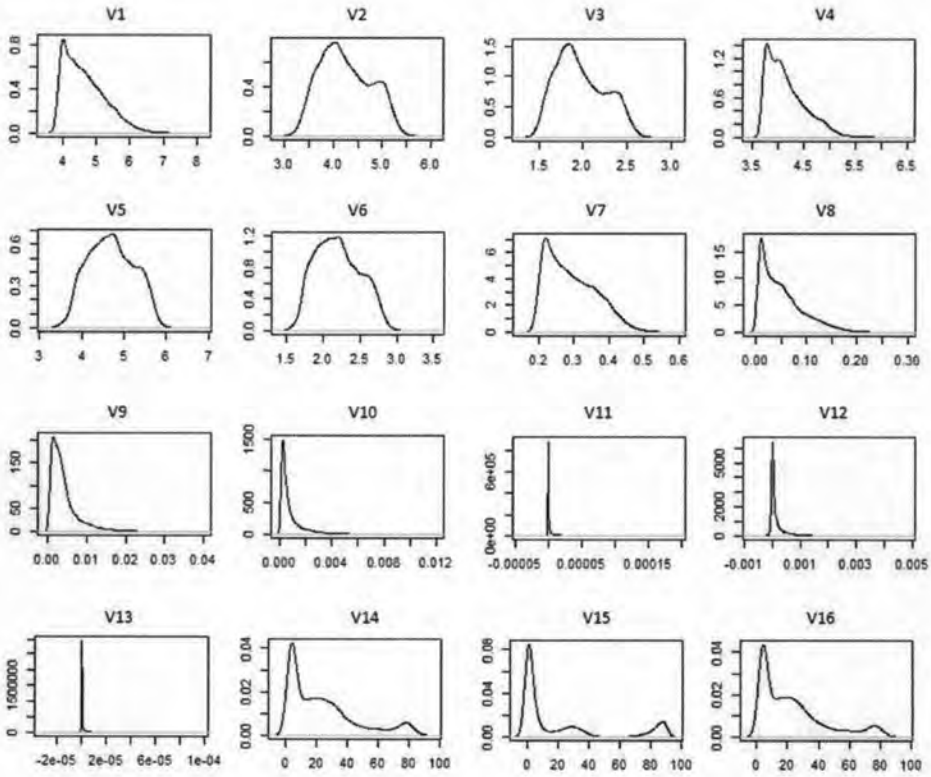

V1
V2
V3
V4
V5
V6
V7
V8
V9
V10
V11
V12
V13
V14
V15
V16

Fig. 2a

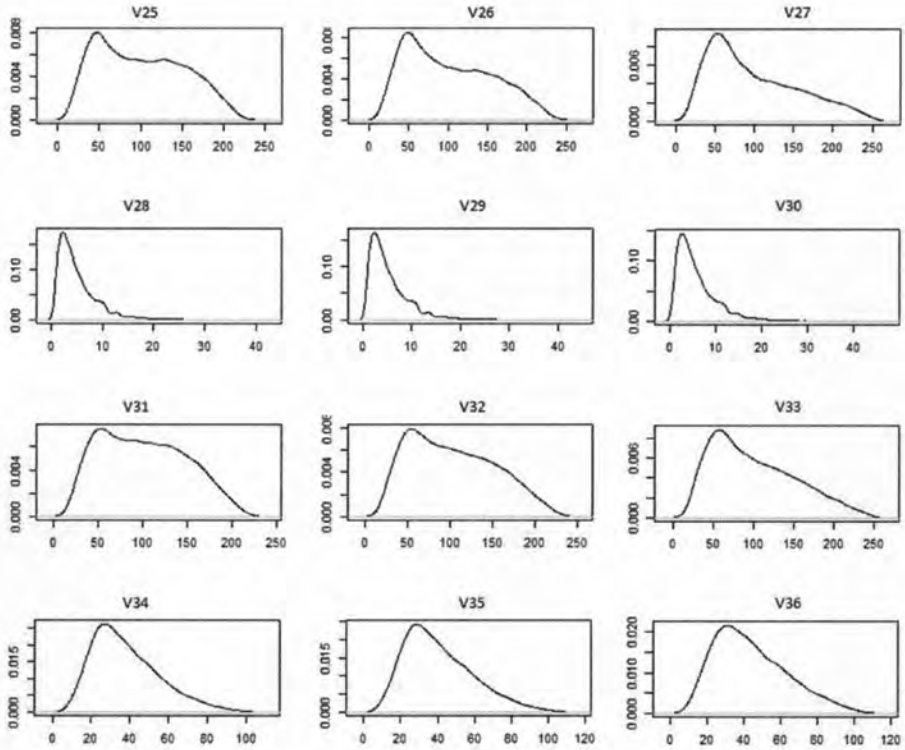

V25
V26
V27
V28
V29
V30
V31
V32
V33
V34
V35
V36

Fig. 2b

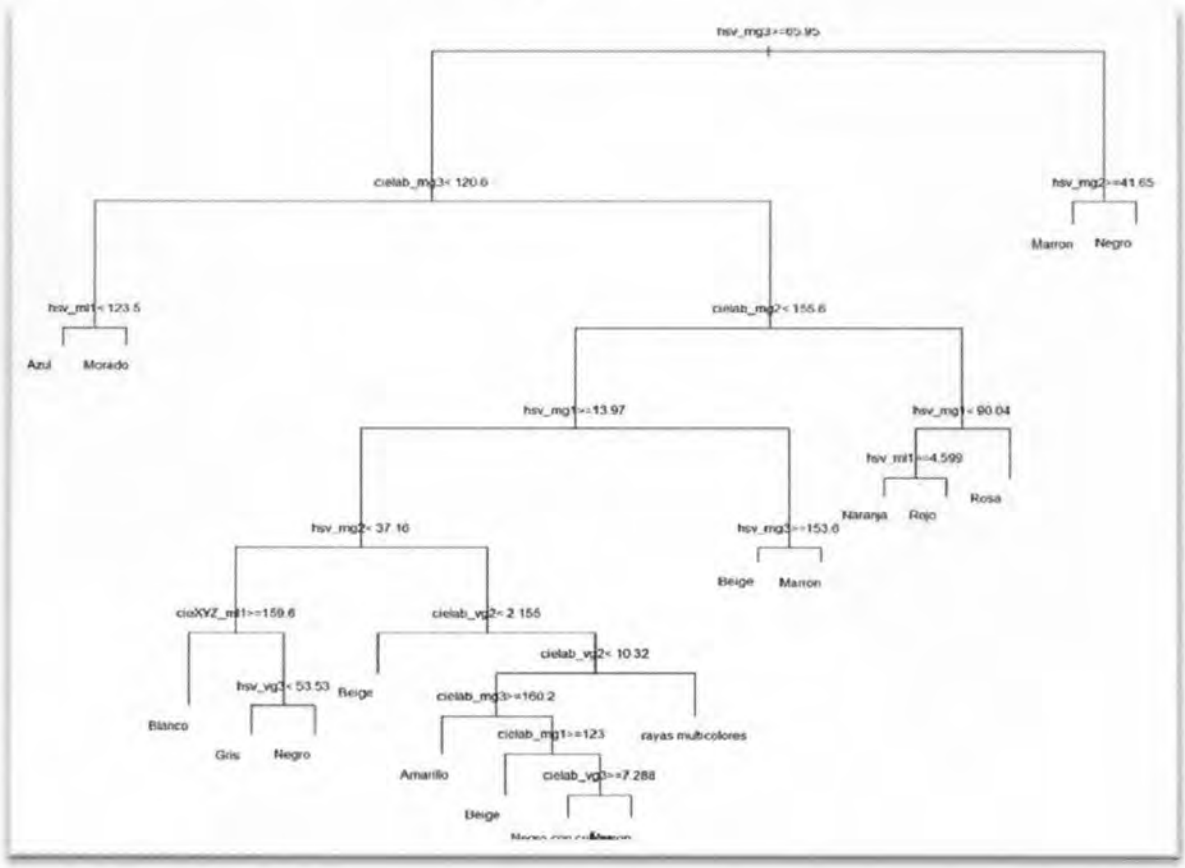

hsv_mg3>=60.95
cielab_mg3< 120.6
hsv_mg2>=41.65
Marron
Negro
hsv_ml1< 123.5
Azul
Morado
cielab_mg2< 155.6
hsv_mg1>=13.97
hsv_mg1< 90.04
hsv_ml1>=4.599
Rosa
Naranja
Rojo
hsv_mg2< 37.16
hsv_mg3>=153.6
Beige
Marron
cieXYZ_ml1>=159.6
cielab_vg2< 2.155
cielab_vg2< 10.32
hsv_vg3< 53.53
Beige
cielab_mg3>=160.2
Blanco
cielab_mg1>=123
rayas multicolores
Gris
Negro
Amarillo
cielab_vg3>=7.288
Beige

Fig. 3

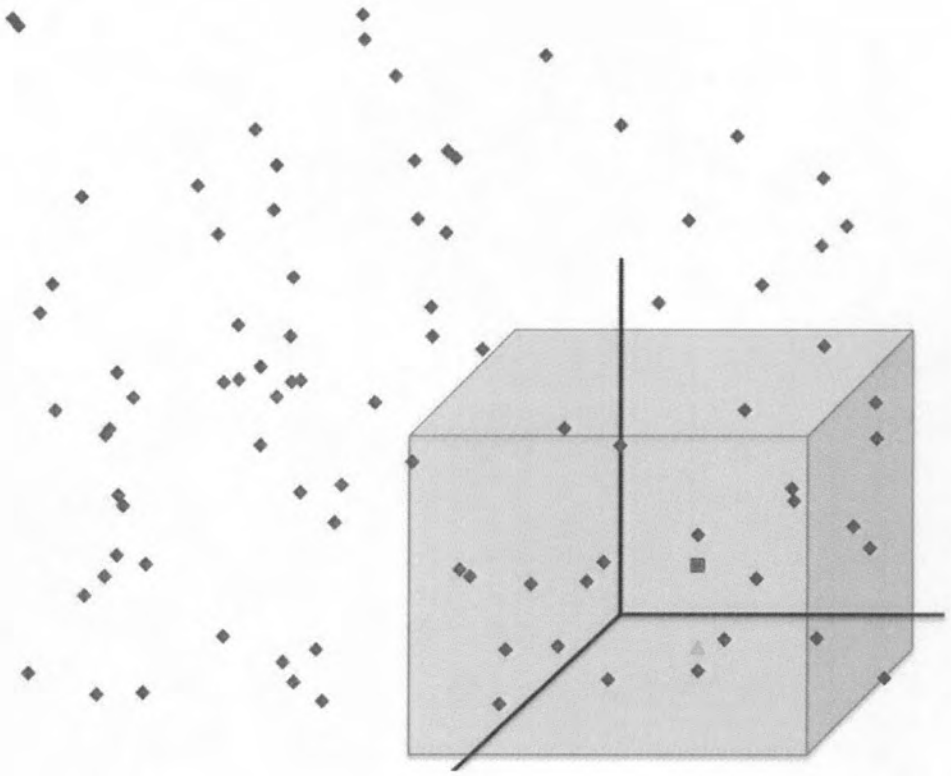


Fig. 4

OFICINA ESPAÑOLA
DE PATENTES Y MARCAS

ESPAÑA

(21) N.º solicitud: 201300816

(22) Fecha de presentación de la solicitud: 04.09.2013

(32) Fecha de prioridad:

INFORME SOBRE EL ESTADO DE LA TECNICA

(51) **Int. Cl. :** ***G06F17/30*** (2006.01)

DOCUMENTOS RELEVANTES

Categoría	(56) **Documentos citados**	**Reivindicaciones afectadas**
X	US 2003195883 A1 (MOJSILOVIC ALEKSANDRA et al.) 16/10/2003, figura 5, párrafos [0018 - 0110];	1-6, 9, 12-13
A		7-8, 10-11
X	US 2005271304 A1 (RETTERATH JAMIE E et al.) 08/12/2005, figura 4, párrafos [0002 - 0057];	1
A	US 5579471 A (BARBER RONALD J et al.) 26/11/1996, columna 2, línea 35 - columna 6, línea 34;	1-13
A	EP 0969388 A1 (EASTMAN KODAK CO) 05/01/2000, párrafos [0007 - 0035];	1-13

Categoría de los documentos citados

X: de particular relevancia
Y: de particular relevancia combinado con otro/s de la misma categoría
A: refleja el estado de la técnica
O: referido a divulgación no escrita
P: publicado entre la fecha de prioridad y la de presentación de la solicitud
E: documento anterior, pero publicado después de la fecha de presentación de la solicitud

El presente informe ha sido realizado

☒ para todas las reivindicaciones ☐ para las reivindicaciones nº:

Fecha de realización del informe	**Examinador**	**Página**
28.11.2014	**A. Casado Fernández**	1/6

Documentación mínima buscada (sistema de clasificación seguido de los símbolos de clasificación)

G06F

Bases de datos electrónicas consultadas durante la búsqueda (nombre de la base de datos y, si es posible, términos de búsqueda utilizados)

INVENES, EPODOC

Fecha de Realización de la Opinión Escrita: 28.11.2014

Declaración

Novedad (Art. 6.1 LP 11/1986)	Reivindicaciones	1-13	**SI**
	Reivindicaciones		**NO**
Actividad inventiva (Art. 8.1 LP11/1986)	Reivindicaciones	7-8, 10-11	**SI**
	Reivindicaciones	1-6, 9, 12-13	**NO**

Se considera que la solicitud cumple con el requisito de aplicación industrial. Este requisito fue evaluado durante la fase de examen formal y técnico de la solicitud (Artículo 31.2 Ley 11/1986).

Base de la Opinión.-

La presente opinión se ha realizado sobre la base de la solicitud de patente tal y como se publica.

1. Documentos considerados.-

A continuación se relacionan los documentos pertenecientes al estado de la técnica tomados en consideración para la realización de esta opinión.

Documento	Número Publicación o Identificación	Fecha Publicación
D01	US 2003195883 A1 (MOJSILOVIC ALEKSANDRA et al.)	16.10.2003
D02	US 2005271304 A1 (RETTERATH JAMIE E et al.)	08.12.2005
D03	US 5579471 A (BARBER RONALD J et al.)	26.11.1996
D04	EP 0969388 A1 (EASTMAN KODAK CO)	05.01.2000

2. Declaración motivada según los artículos 29.6 y 29.7 del Reglamento de ejecución de la Ley 11/1986, de 20 de marzo, de Patentes sobre la novedad y la actividad inventiva; citas y explicaciones en apoyo de esta declaración

El documento D01 se considera el más próximo del estado de la técnica a la invención solicitada.
Entre paréntesis se cita las partes relevantes del Documento D01.

Reivindicación 1:

Método implementado por ordenador para recuperación de imágenes por contenido ([0018]), del tipo que comprende:
a) seleccionar una imagen consultada ([0018], [0110]);
b) segmentar, mediante la aplicación de una técnica de segmentación ([0018]), dicha imagen consultada;
c) extraer características de dicha imagen consultada segmentada mediante el cálculo de al menos dos descriptores de características incluyendo color y textura ([0018], [0085]); y
d) determinar la similitud de la imagen consultada con una pluralidad de imágenes incluidas en una base de datos ([0110], figura 5), las cuales incluyen también características extraídas calculadas por dichos al menos dos descriptores, estando el método caracterizado porque en dicha etapa c) dichos descriptores de características de color y textura calculados incluyen la combinación de diferentes espacios de color, los cuales comprenden unas medidas estadísticas globales y locales en dichos espacios, dando como resultado un descriptor semántico de alto nivel, invariante.

La diferencia entre D01 y la reivindicación 1 es que en la reivindicación 1 los descriptores de color y textura calculados incluyen la combinación de diferentes espacios de color. El efecto técnico de esta diferencia es la obtención de un descriptor semántico de alto nivel.

La utilización de diferentes espacios de color en el tratamiento de imágenes digitales es ampliamente conocido en el estado de la técnica (ver a modo ilustrativo el documento D02 [0048]).

La combinación de diferentes espacios de color no dota a la reivindicación 1 de actividad inventiva.

Por tanto, la reivindicación 1 no implica actividad inventiva (Artículo 8 LP.).

Reivindicación 2:

Método implementado por ordenador según la reivindicación 1, caracterizado porque además dichos descriptores de características de color y textura están combinados con al menos unos descriptores de forma ([0085]-[0086]) y orientación proporcionando un descriptor de forma/geometría y orientación, invariante.

La diferencia entre D01 y la reivindicación 2 es que la reivindicación 2 combina los descriptores con un descriptor de orientación.

La utilización de un descriptor de orientación para caracterizar una imagen es conocida en el estado de la técnica (véase a modo ilustrativo el documento D03, columna 3 líneas 7-22).

Por lo tanto la utilización de un descriptor de orientación no dota a la reivindicación 2 de actividad inventiva ya que su utilización es una técnica habitual.

Por tanto, la reivindicación 2 no implica actividad inventiva (Artículo 8 LP.).

Reivindicación 3:

D01 no calcula la media y la varianza de la imagen consultada para dichos espacios de color diferentes.
El cálculo de la media y la varianza se consideran operaciones habituales en el tratamiento de imágenes digitales (véase D04 [0022]).

Por tanto, la reivindicación 3 no implica actividad inventiva (Artículo 8 LP.).

Reivindicación 4:

Método implementado por ordenador según la reivindicación 3, caracterizado porque comprende además calcular una región de interés (ROI) de un determinado tamaño de píxeles de la imagen consultada para diferenciar diferentes regiones de color o textura en la imagen consultada ([0110]), en donde dicha ROI calculada es aquella que cumple unos valores de media y varianza con una distancia Euclídea más cercana a dichos valores de media y varianza calculados de la imagen consultada.

Por tanto, la reivindicación 4 no implica actividad inventiva (Artículo 8 LP.).

Reivindicación 5:

La diferencia entre D01 y la reivindicación 5 es que en la reivindicación 5 se especifica el tamaño de la ROI calculada, de al menos 9x9 píxeles.

Esta diferencia es una mera opción de diseño que no confiere ningún elemento de significación inventiva con respecto al estado de la técnica conocido.

Consecuentemente, el objeto de la reivindicación 5 carece de actividad inventiva (Artículo 8 LP.).

Reivindicación 6:

La diferencia entre la reivindicación 6 y D01 es que la reivindicación 6 especifica los posibles espacios de color utilizados.

Esta diferencia no se considera que confiera ningún elemento de significación inventiva con respecto al estado de la técnica conocido (véase D02 [0048]).

Consecuentemente, el objeto de la reivindicación 6 carece de actividad inventiva (Artículo 8 LP.).

Reivindicación 7:

La invención reivindicada difiere principalmente de los documentos citados en que ninguno de los documentos citados muestra al descriptor de características de forma que comprenda utilizar los momentos de Hu invariantes y de bajo nivel y otros parámetros no lineales y adimensionales relacionados con el volumen del objeto, basados en áreas como las envolventes convexas, excentricidades, otras formas derivadas y sus ratios relacionados con el objeto.

No se considera que un experto en la materia obtenga la invención a partir de los documentos mencionados anteriormente.

Por lo tanto la invención reivindicada en la reivindicación 7 cumple con los requisitos de novedad y actividad inventiva (Art. 6 y 8 LP.)

Reivindicación 8:

La reivindicación 8, dependiente de la reivindicación 7, es por tanto nueva e implica actividad inventiva (Artículos 6 y 8 LP.).

Reivindicación 9:

La diferencia entre D01 y la reivindicación 9 es que el descriptor de características de orientación se calcula al menos en base a un valor del ángulo.

El cálculo de la orientación en base al ángulo es una mera opción constructiva y se considera una técnica ampliamente conocida por el experto en la materia.

Por lo tanto, la reivindicación 9 no implica actividad inventiva (Artículo 8 LP.).

Reivindicación 10:

La diferencia principal entre el documento D01 y la reivindicación 10 es que D01 no menciona que la similitud se determine en base a una medición en el espacio, en tiempo real y aplicada dentro de una agrupación hipercubo, en donde la agrupación hipercubo está centrada en la posición del espacio de la imagen de consulta, definida por un vector de radio n-dimensional del total de n-dimensiones de dichos dos descriptores juntos, con valores de radio independientes y diferentes para cada dimensión, basada en un porcentaje de distancia aceptado del valor de la posición de dicha imagen de consulta, y limitada al cálculo de distancias 1:1 de la imagen de consulta dentro de la agrupación.

Así, la invención reivindicada implica un efecto mejorado comparado con el estado de la técnica.

Por lo tanto la invención reivindicada cumple con los requisitos de novedad y actividad inventiva (Artículos 6 y 8 LP.).

Reivindicación 11:

La reivindicación 11, dependiente de la reivindicación 10, es por tanto nueva e implica actividad inventiva (Artículos 6 y 8 LP.).

Reivindicación 12:

La diferencia entre D01 y la reivindicación 12 es que en la reivindicación 12 los descriptores se optimizan mediante la aplicación de técnicas de transformación y métodos de filtrado.

La utilización de técnicas de transformación y métodos de filtrado no dota a la reivindicación 12 de actividad inventiva ya que su utilización es una técnica habitual (véase D02 [0048]-[0054]).

Por tanto, la reivindicación 12 no implica actividad inventiva (Artículo 8 LP.).

Reivindicación 13:

La reivindicación 13 es una reivindicación de tipo declarativo. No se considera inventiva en la medida en que la reivindicación 1 no implica actividad inventiva.

Consecuentemente, el objeto de la reivindicación 13 carece de actividad inventiva (Artículo 8 LP.).