

19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 407 105**

21 Número de solicitud: 201101196

51 Int. Cl.:

G06K 9/00 (2006.01)

G06F 17/30 (2006.01)

G06K 9/62 (2006.01)

G06K 9/78 (2006.01)

12

PATENTE DE INVENCION

B1

22 Fecha de presentación:

07.11.2011

43 Fecha de publicación de la solicitud:

11.06.2013

Fecha de la concesión:

22.04.2014

45 Fecha de publicación de la concesión:

29.04.2014

56 Se remite a la solicitud internacional:

PCT/ES2012/070743

73 Titular/es:

**UNIVERSIDAD DE ALICANTE (100.0%)
CARRETERA SAN VICENTE DEL REASPEIG S/N
03690 SAN VICENTE DEL RASPEIG (Alicante) ES**

72 Inventor/es:

PERTUSA IBÁÑEZ, Antonio

54 Título: **METODO Y SISTEMA DE RECUPERACION DE INFORMACION DE IMAGENES EN DISPOSITIVOS MOVILES USANDO METADATOS.**

57 Resumen:

Método y sistema de recuperación de información de imágenes en dispositivos móviles usando metadatos que comprende el empleo de un teléfono móvil (100) y un servidor remoto (102), y en donde esencialmente, se emplean una pluralidad de metadatos como el giro del dispositivo, la aceleración en el momento de la toma, metadatos de audio, datos de red, precisión y cobertura GPS, tamaño de la región de interés, ángulo respecto a la horizontal, grado de luminosidad exterior, metadatos obtenidos mediante geocodificación inversa, altura sobre el suelo y condiciones climatológicas para la mejora interactiva de la identificación del objeto previamente seleccionado por el usuario.

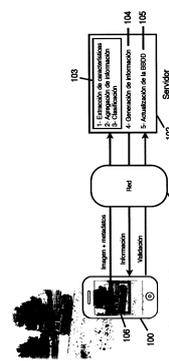


Figura 1

ES 2 407 105 B1

**MÉTODO Y SISTEMA DE RECUPERACIÓN DE INFORMACIÓN DE
IMÁGENES EN DISPOSITIVOS MÓVILES USANDO METADATOS**

DESCRIPCIÓN

5

La presente invención está referida a un método basado en una arquitectura cliente/servidor para la recuperación de información a partir de imágenes obtenidas mediante dispositivos móviles.

10 **Estado de la técnica anterior**

La búsqueda de información sobre un objeto a partir de su fotografía es útil en aquellos problemas donde es más sencillo describir la consulta mediante una imagen que mediante una descripción textual.

15

En la actualidad, existen multitud de sistemas de recuperación de información a partir de imágenes. Esta tarea, en la que destacan sistemas como Google images [<http://images.google.com>], es conocida como *content-based image retrieval*.

20

Para llevarla a cabo, normalmente se extraen una serie de características de la imagen (locales o globales), se procesan usando técnicas de agrupación de la información, y finalmente se obtiene un ranking de imágenes similares a la consulta presentes en una base de datos de referencia. Se puede identificar o clasificar el objeto consultando, por ejemplo, la categoría de la primera imagen del ranking (para ello es imprescindible que la base de datos esté etiquetada).

25

La principal ventaja de los dispositivos móviles avanzados es la toma de fotografías desde cualquier punto y su envío a través de internet. La recuperación de información de imágenes obtenidas mediante estos dispositivos es un problema de reciente estudio. Actualmente, la aplicación más conocida que aborda este problema es Google Goggles [descrito en US20100260373], que es capaz de reconocer marcas de vino, libros, obras de arte, monumentos y logos comerciales.

30

A pesar de que reconocer cualquier tipo de imagen en general es un problema extremadamente complicado, este puede simplificarse en parte usando dispositivos móviles, ya que estos proporcionan datos adicionales (metadatos) sobre la localización espacio/temporal, información sobre el usuario y datos aportados por los

35

sensores para complementar la información de la imagen.

Estos metadatos pueden ser útiles para restringir la búsqueda. Por ejemplo, si tomamos una foto de un objeto en un exterior y conocemos la hora de la toma, 5 podríamos mejorar el reconocimiento seleccionando sólo las imágenes que se han tomado alrededor de dicha hora, ya que las condiciones de iluminación serán similares.

Podemos encontrar en la literatura diversas técnicas de extracción de características, 10 tanto locales (SIFT, SURF, etc.) como globales (color, textura, etc.). Tras obtener las características de la imagen, se pueden aplicar métricas basadas en distancias, agrupamiento en clusters o histogramas, o usar técnicas de aprendizaje, como vecinos más cercanos, support vector machines, o boosting, para obtener la clase de la imagen, o la imagen con mayor similitud.

15

Para mejorar los resultados en la búsqueda de imágenes, conviene restringir el número de objetos con los que comparar la imagen de la consulta. Metadatos como la fecha o la localización GPS pueden facilitar la identificación de la escena o del objeto reduciendo el espacio de búsqueda. Por ejemplo, usando únicamente metadatos de la 20 cámara se pueden obtener buenos resultados para discriminar entre escenas de interior y de exterior [US2005/0105776].

En la literatura se ha estudiado el aporte de los metadatos aportados por las cámaras fotográficas [US2005/0105776], o por los dispositivos móviles, aunque de una 25 manera básica [US20100260373].

Explicación de la invención

La presente invención tiene una serie de características diferenciadoras. La primera 30 es la inclusión de una pluralidad de nuevos metadatos para mejorar la búsqueda, dentro de un esquema multimodal. Otra característica diferenciadora es que se trata de un sistema interactivo, en el que el usuario selecciona la región de interés alrededor del objeto y además valida la respuesta dada por el sistema, de tal forma que pueda mejorar en futuras consultas.

35

Más concretamente, el método de recuperación de información de imágenes en dispositivos móviles usando metadatos emplea dispositivos móviles que comprenden

medios para la captura de imágenes, así como medios de transmisión de dichas imágenes capturadas junto con una pluralidad de metadatos a un servidor remoto.

5 El método se caracteriza porque comprende una etapa de captura de una imagen mediante el dispositivo móvil, así como la selección de una región dentro de la propia imagen por parte del usuario de dicho dispositivo móvil; en donde en dicha región se encuentra un objeto del cual se quiere obtener alguna información.

10 Posteriormente, se produce el envío de la región seleccionada de la imagen junto con una pluralidad de metadatos. Tras ello, se identifica en el servidor remoto y se clasifica el objeto identificado, remitiendo dicha información al dispositivo móvil.

15 Por último, se valida la respuesta del servidor remoto en el dispositivo móvil por parte del usuario, indicando cuando esa información es correcta y, en caso contrario, etiquetando el objeto para su devolución y almacenamiento en una base de datos del servidor remoto.

20 En un segundo aspecto de la invención, se reivindica el sistema que implementa el método anteriormente descrito.

25 Como se puede observar, se trata de un sistema y un método personalizable, en el sentido de que se pretende que el usuario decida qué tipos de objeto desea clasificar, y colaborativo, en el sentido de que las imágenes que vaya añadiendo un usuario se podrán usar para mejorar las consultas de otros usuarios.

30 A lo largo de la descripción y las reivindicaciones la palabra "comprende" y sus variantes no pretenden excluir otras características técnicas, aditivos, componentes o pasos. Para los expertos en la materia, otros objetos, ventajas y características de la invención se desprenderán en parte de la descripción y en parte de la práctica de la invención. Los siguientes ejemplos y dibujos se proporcionan a modo de ilustración, y no se pretende que sean limitativos de la presente invención. Además, la presente invención cubre todas las posibles combinaciones de realizaciones particulares y preferidas aquí indicadas.

35 **Breve descripción de los dibujos**

FIG 1. Muestra de forma esquemática el sistema y método de recuperación de

información de imágenes en dispositivos móviles usando metadatos.

Exposición detallada de un modo de realización

5 El esquema general se puede observar en la figura adjunta. El usuario puede tomar una fotografía mediante el dispositivo móvil (100) y seleccionar la región correspondiente al objeto del cual se quiere obtener información (106). La imagen comprendida en esta región se envía por red al servidor remoto (102) junto con los metadatos descritos a continuación. El servidor remoto (102) posteriormente
10 identifica o clasifica el objeto y devuelve la información al usuario, que puede validar la respuesta. El usuario puede validar la respuesta, indicando si es correcta y, en caso contrario, especificando el tipo o la clase de objeto.

15 Por ejemplo, si el sistema identifica que se le está haciendo una foto a una motocicleta pero se trata de una bicicleta, el usuario podrá indicar que la respuesta no es correcta para añadir la imagen etiquetada a la base de datos del servidor.

Se propone usar los metadatos que se describen a continuación, extraídos a partir de los sensores que poseen algunos dispositivos móviles:

20

- Giro del dispositivo. Un giroscopio puede proporcionar información sobre los ángulos de Tait-Bryan. Estos representan los ejes de cabeceo, dirección y alabeo, a partir de los cuales se pueden obtener más metadatos como el ángulo respecto a la horizontal.
- 25 - Orientación. Si el dispositivo móvil es rectangular, existen cuatro orientaciones posibles. Este dato es de relevancia para rotar la imagen al enviarla al servidor, y también para calcular el ángulo respecto a la horizontal.
- Posición GPS. Las coordenadas en el momento de la toma pueden usarse para, por ejemplo, escoger sólo aquellos objetos cercanos a la localización del
30 dispositivo. El verdadero potencial de la posición GPS se puede revelar mediante geo-codificación inversa.
- Datos de aceleración. Usando un acelerómetro se puede obtener la aceleración respecto cada uno de los ejes en el momento de la toma. Estos datos proporcionan una indicación sobre la estabilidad del dispositivo, y esta
35 información puede usarse para seleccionar las imágenes más estables como representativas de una clase.
- Fecha. Es más probable, por ejemplo, realizar una fotografía de una flor en

primavera, o de un muñeco de nieve en invierno, o de una piscina en verano. La información sobre la fecha puede ser de utilidad para discernir entre ciertos tipos de objetos o escenas.

- 5 - Hora. La hora en la que se ha realizado también puede proporcionar información de relevancia. Se puede usar, por ejemplo, para restringir la búsqueda en aquellos objetos que se han fotografiado a esa hora, comparando así imágenes con condiciones de iluminación similares.
- 10 - Parámetros de la cámara. Datos como la velocidad de obturación, la apertura, el tiempo de exposición, la distancia focal, la ISO, o el modo de flash son también relevantes para esta tarea. La distancia focal es un parámetro representativo de la distancia de los objetos fotografiados, pero actualmente las cámaras de los dispositivos móviles avanzados suelen tener distancias focales fijas.
- 15 - Metadatos de audio. El nivel de decibelios en el momento de la toma o en un intervalo de tiempo alrededor de la misma puede indicar si el usuario se encuentra en un entorno ruidoso (como un centro comercial) o silencioso (como una zona rural).
- 20 - Datos de red. Si existe una red WiFi cercana, lo más probable es que el dispositivo móvil se encuentre en una zona poblada. La densidad o cantidad de redes inalámbricas que se detectan en el momento de la toma puede usarse como un indicio para obtener datos sobre el entorno.
- 25 - Precisión y cobertura GPS. Por regla general, siempre se detecta al menos un satélite en localizaciones exteriores, pero en zonas interiores no se detectan debido a que techos y paredes impiden que la señal alcance el dispositivo. Por tanto, la precisión de la posición GPS (en metros) o la cobertura (en número de satélites detectados) también son datos relevantes para esta tarea, y pueden indicar si la escena es de interior o de exterior.
- 30 - Identificación del usuario. Un mismo usuario tiende a realizar fotografías de objetos similares (animales, objetos de oficina, etc.) y a repetir búsquedas sobre objetos ya fotografiados. Por tanto, es posible dar prioridad en la clasificación a las categorías de imágenes tomadas previamente por el usuario.
- Tamaño de la región de interés. La altura y anchura de la región de interés seleccionada por el usuario puede ayudar a conocer si el objeto es alargado o compacto.

35

A partir de los metadatos anteriores, se pueden derivar una serie de metadatos adicionales que revelan el auténtico potencial de los dispositivos móviles para esta

tarea:

- 5 - Ángulo respecto a la horizontal. Esta información es de relevancia para conocer el tamaño del objeto fotografiado o, por ejemplo, para discernir entre los objetos situados en el suelo o colocados en una pared. Los ángulos de Tait-Bryan obtenidos mediante un giroscopio se pueden combinar con la orientación del dispositivo móvil para calcular el ángulo respecto al suelo en el momento de la toma.
- 10 - Grado de luminosidad exterior. A partir de la fecha, hora y latitud GPS, se puede calcular si es de día, de noche, o si está amaneciendo o anocheciendo. Este dato se puede usar, por ejemplo, para deducir que si la toma se ha hecho de noche y sin flash, es más probable que se trate una escena de interior que de exterior.
- 15 - Metadatos obtenidos mediante geocodificación inversa. Usando sistemas como simplegeo, google, geonames, etc., se pueden obtener datos relevantes a partir de una localización, tales como:
 - o El tipo de lugar donde se ha realizado la toma (zoo, universidad, bosque, ciudad, playa, etc.)
 - o El código postal de la zona
 - 20 o La ciudad, pueblo o región
 - o El país
 - o La densidad demográfica en la zona

Estas características son de interés para restringir la clasificación de determinados objetos. Por ejemplo, es muy improbable encontrar un elefante en una ciudad, pero sí se puede encontrar en un zoo.
- 25 - Altura sobre el suelo. Mediante geocodificación inversa se puede obtener la altitud sobre el mar en una determinada zona. Comparando esta altitud con la altitud obtenida mediante GPS, se puede deducir (con un cierto margen de error) si la toma se ha realizado desde, por ejemplo, un edificio de varias plantas o a nivel del suelo.
- 30 - Condiciones climatológicas. Los datos relativos a la temperatura, la humedad relativa o si está lloviendo o soleado también se pueden extraer a partir de la posición GPS, localizando la estación meteorológica más cercana y obteniendo su información mediante sistemas como *simplegeo* o *geonames*. Estos datos pueden ser de relevancia para conocer si la toma ha sido exterior o interior, si puede aparecer nieve en la imagen, etc.
- 35 - Orientación de la vista sobre el mapa. Los ángulos de Tait-Bryan en

combinación con la posición GPS permiten conocer el punto hacia el cual se ha hecho la fotografía. Por ejemplo, en una zona con cuatro puntos de interés, si la foto se ha tomado en el lugar central de los mismos se podrían usar estos ángulos para determinar a qué punto se está haciendo referencia.

- 5 - Metadatos adicionales a partir de los metadatos sobre la cámara. Por ejemplo, a partir del tiempo de exposición y del valor de apertura se puede obtener la energía de la escena, que está directamente correlacionada con el tipo de escena y los diferentes grados de luminosidad.

10 En la clasificación e identificación de la imagen (103) la invención permite el uso de cualquier sistema existente de clasificación o identificación de imágenes. Normalmente, esta tarea se realiza en tres pasos: extracción de características, agrupación de la información y clasificación o identificación. En función del problema, las características seleccionadas pueden ser tanto puntos de interés (SIFT, SURF, Harris, etc.) como descriptores de color, textura, forma, etc.

15

Posteriormente, se suelen agrupar las características empleando “bag of words”, histogramas, análisis de componentes principales, etc. para reducir el coste computacional y mejorar los resultados de la clasificación. Opcionalmente, se puede usar verificación geométrica para comprobar la coherencia espacial de las características.

20

Por último, se comparan las características agrupadas de una imagen con los de todas las imágenes de una base de datos etiquetada (o con un subconjunto de las misma) para obtener un ranking de imágenes similares. Las etiquetas de las imágenes más similares proporcionan la información sobre la clase del objeto. La comparación se puede realizar calculando los vecinos más cercanos, dada una distancia definida o usando técnicas de aprendizaje para clasificar la muestra, tales como “support vector machines”, redes de Bayes, redes neuronales, etc.

25

30 Los metadatos pueden complementar la información de la imagen para mejorar la clasificación. La integración se puede realizar de distintas formas:

- Clasificación con características de imagen y filtrado posterior usando metadatos.
- 35 - Filtrado inicial con los metadatos y posterior clasificación con características de imagen.
- Clasificación conjunta uniendo características de imagen con metadatos.

5 En los tres casos se pueden emplear técnicas de aprendizaje automático, y en el caso de que haya interactividad con el usuario, estas deben ser incrementales. La información puede integrarse usando una red Bayesiana u otras técnicas de clasificación estadística.

10 En la generación de información (104), una vez el servidor remoto (102) haya identificado o clasificado el objeto, se envía información sobre el mismo al dispositivo móvil (100). Esta información incluye el tipo de objeto para que el usuario pueda validarlo (opcionalmente) e información adicional. Por ejemplo, si el sistema identifica un cuadro determinado de un artista, puede proporcionar información sobre el autor o sobre obras similares. Esta información puede ser textual, visual o sonora (por ejemplo, para una audioguía).

15 Finalmente, se incorpora la muestra etiquetada a la base de datos (105). El usuario valida la respuesta del servidor, indicando si la clase es correcta o no. En caso afirmativo, se añade la nueva imagen a la base de datos con la etiqueta obtenida. En caso de que la respuesta no sea correcta, el usuario puede indicar la respuesta correcta y esta se añadirá a la base de datos con su nueva etiqueta para mejorar
20 futuras consultas.

REIVINDICACIONES

1.- Método de recuperación de información de imágenes en dispositivos
5 móviles usando metadatos; en donde dichos dispositivos móviles (100) comprenden medios para la captura de imágenes y medios de transmisión de dichas imágenes capturadas más una pluralidad de metadatos a un servidor remoto (102);

que se caracteriza porque

(i) la captura de una imagen mediante un dispositivo móvil (100) y la
10 selección por parte del usuario de dicho dispositivo móvil (100) de una determinada región dentro de la propia imagen, en donde dicha región se corresponde con un objeto del cual se quiere obtener alguna información (106);

(ii) el envío de la región seleccionada de la imagen junto con una pluralidad de metadatos;

(iii) la identificación (103) en el servidor remoto (102) y clasificación del
15 objeto identificado, remitiendo dicha información (104) al dispositivo móvil (100); y

(iv) la validación de la respuesta del servidor remoto (102) en el dispositivo
móvil (100) por parte del usuario, indicando cuando esa información es correcta y, en caso contrario, etiquetando el objeto para su devolución y almacenamiento en una
20 base de datos del servidor remoto (102).

2.- Método de acuerdo con la reivindicación 1 en donde los metadatos comprenden al menos el giro del dispositivo, la aceleración en el momento de la
25 toma, metadatos de audio, datos de red, precisión y cobertura GPS, tamaño de la región de interés, ángulo respecto a la horizontal, grado de luminosidad exterior, metadatos obtenidos mediante geocodificación inversa, altura sobre el suelo y condiciones climatológicas.

3.- Método de acuerdo con las reivindicaciones 1 y 2 en donde la información
30 remitida por el servidor remoto (102) comprende el tipo de objeto e información adicional respecto a ese objeto.

4.- Sistema de recuperación de de información de imágenes en dispositivos
35 móviles usando metadatos; en donde dichos dispositivos móviles (100) comprenden medios para la captura de imágenes y medios de transmisión de dichas imágenes capturadas más una pluralidad de metadatos a un servidor remoto (102);

que se caracteriza porque

implementa medios configurados para la ejecución del método de las reivindicaciones 1 a 3 y que consisten en medios para la selección por parte del usuario de dicho dispositivo móvil (100) de una región dentro de la propia imagen, que se corresponde con un objeto del cual se quiere obtener alguna información (106); medios configurados para el envío de la región seleccionada de la imagen junto con una pluralidad de metadatos, comprendiendo al menos el giro del dispositivo, la aceleración en el momento de la toma, metadatos de audio, datos de red, precisión y cobertura GPS, tamaño de la región de interés, ángulo respecto a la horizontal, grado de luminosidad exterior, metadatos obtenidos mediante geocodificación inversa, altura sobre el suelo y condiciones climatológicas; y medios para la identificación en el servidor remoto (102) y clasificación del objeto identificado, remitiendo dicha información (104) al dispositivo móvil (100).

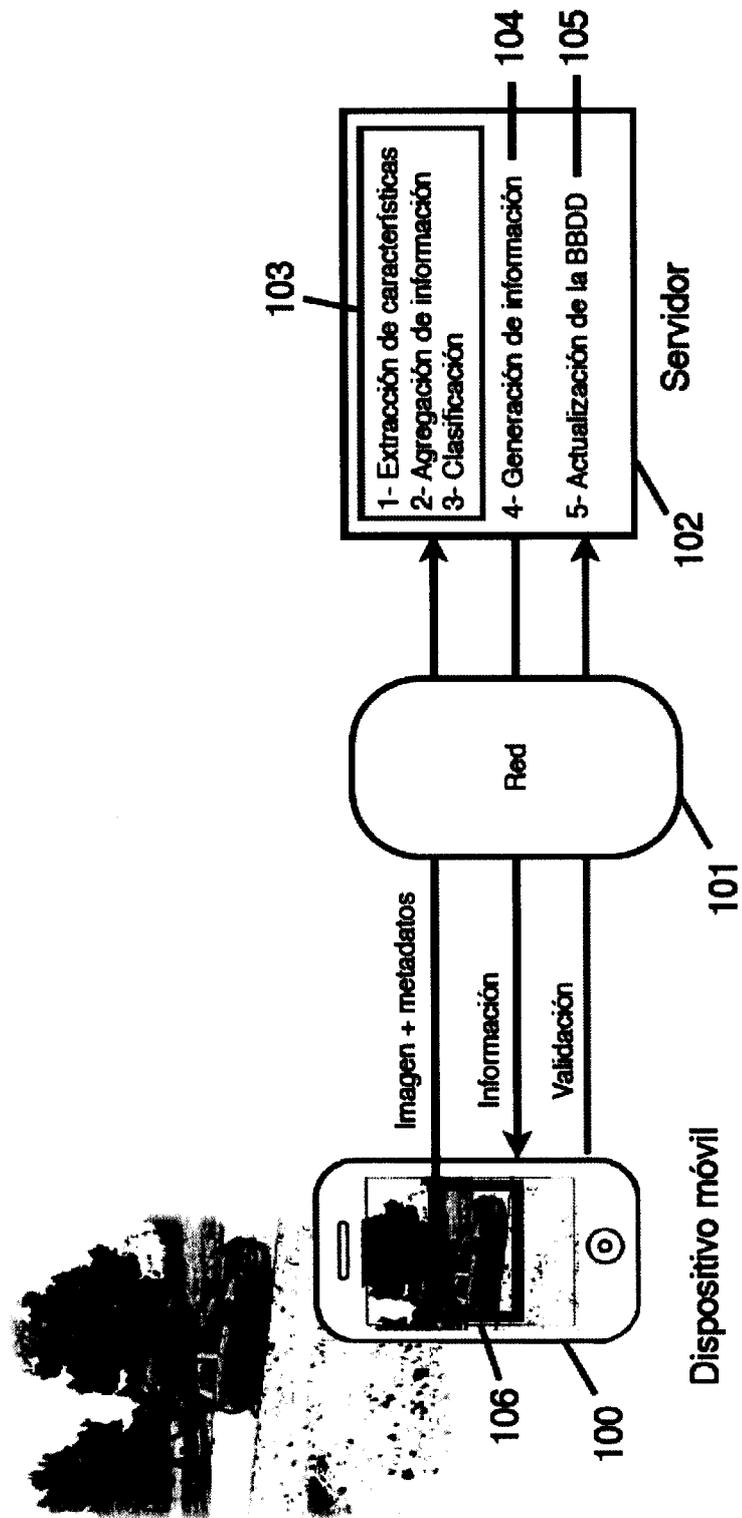


Figura 1