(19) OFICINA ESPAÑOLA DE PATENTES Y MARCAS

ESPAÑA

(11) Número de publicación: **2 844 851**

(51) Int. CI.:

G06F 16/783 (2009.01)

(12) TRADUCCIÓN DE PATENTE EUROPEA T3

(86) Fecha de presentación y número de la solicitud internacional: **18.08.2016 PCT/ES2016/070605**

(87) Fecha y número de publicación internacional: **22.02.2018 WO18033652**

(96) Fecha de presentación y número de la solicitud europea: **18.08.2016 E 16805476 (5)**

(97) Fecha y número de publicación de la concesión europea: **21.10.2020 EP 3502924**

(54) Título: **Procedimiento generación de una base de datos vinculados con diferentes referencias temporales a un contenido audiovisual**

(45) Fecha de publicación y mención en BOPI de la traducción de la patente:
22.07.2021

(73) Titular/es:

TAGSONOMY, S.L. (100.0%)
Calle Peña Brava 21
33192 Llanera (Asturias), ES

(72) Inventor/es:

GARCÍA ESTRELLA, LUIS;
FIERRO IGLESIAS, PEDRO;
RAMEAU RODRÍGUEZ, MIGUEL;
BLANCO PÉREZ, BEATRIZ;
JULIANA MEDIO, JULIO ENRIQUE;
SÁNCHEZ LÓPEZ, SERGIO;
CASTRO MIGUEL, JONATHAN y
MARIÑO RUIZ, DANIEL

(74) Agente/Representante:

LÓPEZ CAMBA, María Emilia

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

DESCRIPCIÓN

Procedimiento generación de una base de datos vinculados con diferentes referencias temporales a un contenido audiovisual

OBJETO DE LA INVENCIÓN

Como especifica el título de la invención, el objeto de la presente invención es un procedimiento para la generación de una base de datos vinculada a un contenido audiovisual a través de diferentes referencias temporales, de manera que sea posible asignar metadatos a dicho contenido audiovisual; es decir, se refiere a un procedimiento que genera, de manera semiautomática, una serie de datos o información enriquecida, y los vincula automáticamente a un contenido audiovisual a través de diferentes referencias temporales.

Por lo tanto, la base de datos contiene una serie de datos enriquecidos conectados a una referencia temporal precisa del contenido audiovisual.

La presente invención se caracteriza por las características de cada una de las acciones realizadas durante cada etapa del procedimiento de generación de base de datos, de tal manera que, en su conjunto, se logra un procedimiento que asigna, de manera semiautomática, datos registrados enriquecidos que están conectados con precisión a una referencia temporal.

Por lo tanto, la presente invención se enmarca dentro del campo del contenido audiovisual e información adicional. La invención se define por las características de las reivindicaciones independientes. Otras realizaciones son la materia objeto de las reivindicaciones dependientes.

ANTECEDENTES DE LA INVENCIÓN

Actualmente, cuando se desea vincular información a un contenido audiovisual, se realiza de forma manual, identificando cada uno de los elementos a mostrar y estableciendo una marca de tiempo potencial para el inicio y el final de la difusión, así como la información adicional o los metadatos a mostrar.

En el estado de la técnica se conocen los siguientes documentos:

• Documento US2011113444, que describe un sistema para indexar objetos físicos, ubicaciones y personas, denominados colectivamente objetos de vídeo, que aparecen en vídeos. El sistema permite la identificación a nivel de objeto de vídeo de contenido de TV y vídeo, y hace que estos objetos de vídeo sean indexables, enlazables y consultables.

Documento US2007250901, un procedimiento y aparato para anotar flujos de medios. En una realización, el procedimiento facilita la creación de un flujo de medios anotado por un usuario que incluye recibir el flujo de medios del usuario, mapear al menos un elemento de contenido complementario con respecto al menos a una porción del flujo de medios para producir el flujo de medios anotado, y almacenar el flujo de medios anotado. En otra realización, el procedimiento distribuye un flujo de medios anotado proporcionado por un usuario que incluye recibir el flujo de medios anotado del usuario.

Por lo tanto, el objeto de la presente invención es desarrollar un procedimiento para obtener, de manera rápida, semiautomática y escalable, una base de datos formada por un conjunto de archivos y lapsos de tiempo, es decir, datos relevantes (sobre actores, personajes, datos de interés, vestimenta, música, lugares, objetos, etc.) relacionados con diversos contenidos audiovisuales (películas, series de televisión, documentales, vídeos musicales, etc.) y los intervalos de tiempo de interés, que son obviamente dependientes de y diferentes para cada uno de los diferentes contenidos, con las siguientes características que lo hacen único:

DESCRIPCIÓN DE LA INVENCIÓN

El objeto de la presente invención es un procedimiento de asignación de metadatos a un contenido audiovisual a través de diferentes referencias temporales, que comprende las siguientes etapas:

• **Reordenación óptima del contenido audiovisual en subescenas:** En su formato original, un contenido audiovisual puede verse como una sucesión de escenas, planos o incluso fotogramas dispuestos de tal manera que narran una historia o, más generalmente, dispuestos en la forma en que se van a consumir por los espectadores. Esta ordenación puede no ser la óptima cuando el contenido audiovisual está destinado a servir como entrada para diferentes procedimientos automáticos; por ejemplo, para la identificación automática de objetos en pantalla, es mucho más eficiente agrupar los mismos planos de una escena, incluso si no son consecutivos en el tiempo. Por esta razón, la primera etapa de este procedimiento implica un reordenamiento de las subescenas basándose en métricas de similitud entre imágenes y la detección de cambios repentinos en estas similitudes.

• Identificación de elementos en cada una de las subescenas y anotación temporal e informativa de los mismos: Los elementos de entrada utilizados en esta parte del procedimiento son el contenido audiovisual previamente reordenado y una base de datos con información relevante obtenida automáticamente de Internet en base a diferentes categorías (descripciones de objetos, fotografías de lugares, datos técnicos sobre vehículos, premios obtenidos por las películas, etc.), de tal forma que todas las coincidencias localizadas se proponen semiautomáticamente a la persona que dirige el procedimiento.
Dos datos muy importantes que se añaden a los elementos durante esta parte del procedimiento son:

- por un lado, el lapso de tiempo, o marca de tiempo con respecto a la ordenación original del contenido audiovisual, que indica el momento exacto en el que cada pieza de información se identifica en pantalla y deja de ser visible en pantalla; esta operación se realiza de manera automática;
- y, por otro lado, las diferentes relaciones (coincidencia, contenido o continente, lugar, pertenencia, parentesco, relación por compartir un tercer elemento, etc.) entre los diferentes elementos o diferentes piezas de información; esta acción se realiza de manera semiautomática.

• Búsqueda de correspondencias comerciales para los elementos identificados: De manera similar a la búsqueda de información adicional sobre todos los elementos identificados en el contenido, la siguiente etapa del procedimiento utiliza la información identificada hasta ese momento para buscar correspondencias comerciales.
Las correspondencias comerciales se buscan en una base de datos compuesta por todos los catálogos disponibles gracias a acuerdos con diferentes tipos de tiendas y distribuidores. A diferencia de las búsquedas informativas, en las que es necesario encontrar correspondencias exactas, en las búsquedas comerciales es interesante encontrar, además de estas coincidencias exactas, coincidencias similares o coincidencias con objetos que pueden resultar comercialmente interesantes para los consumidores del contenido, aunque estos no sean exactamente iguales; por ejemplo, una prenda similar, otro álbum de música del mismo artista que se está escuchando, otra película con el mismo actor, etc.
La búsqueda de correspondencias comerciales se realiza en base a diferentes tipos de sujetos, por ejemplo, personas, escenarios, música, vehículos, contexto histórico y vestimenta, de manera que, cuando se localiza un tipo de sujeto utilizando diferentes medios informáticos de reconocimiento automático, es posible identificar el nombre del actor, la marca del vehículo, la música que se está reproduciendo, etc., para cargar la información adicional en los elementos de las subescenas en la base de datos; esta información puede ser recuperada automáticamente de sitios web preasignados o de una base de datos previamente creada diseñada para otorgar un contenido uniforme a todos los contenidos audiovisuales con los mismos elementos.
La información relacionada con un elemento se denomina archivo. Un archivo contiene diferentes tipos de información relacionada con el elemento que lo define. Tendrá un título, y puede incluir fotografías del elemento, texto descriptivo o definiciones del mismo, tablas con características o datos, contenidos comerciales que puedan ser de interés para una persona interesada en dicho elemento, mapas de lugares, enlaces a sitios de interés. etc.

• Desduplicación, anotación y edición experta:
Una vez que el contenido ha sido reordenado de manera eficiente y anotado semiautomáticamente con información enriquecida, se han identificado las relaciones entre las piezas de información y se han encontrado correspondencias comerciales, toda la información acumulada es revisada por un experto. Los expertos revisan las piezas de información relacionada con su campo de conocimiento y pueden desduplicar o evitar que una pieza de información sea identificada varias veces cuando realmente es la misma, así como corregir aquellos aspectos que no hayan sido identificados automáticamente de una manera totalmente correcta, y añadir o explicar con más profundidad aquellos contenidos que decidan. Esta última parte permite obtener una base de datos con un nivel único de detalle y calidad sin aumentar significativamente la duración total del procedimiento.

Salvo que se especifique lo contrario, todos los elementos técnicos y científicos usados en la presente memoria descriptiva tienen el significado que habitualmente entienden los expertos habituales en la técnica a la que pertenece esta invención. En la implementación de la presente invención, pueden usarse procedimientos y materiales similares o equivalentes a los descritos en la memoria descriptiva.

A lo largo de la descripción y las reivindicaciones, la palabra «comprende» y variantes de la misma no pretenden excluir otras características técnicas, aditivos, componentes o etapas. Para los expertos en la técnica, surgirán otros objetos, ventajas y características de la invención, en parte de la descripción y en parte de la implementación de la invención.

DESCRIPCIÓN DE LAS FIGURAS

Con el fin de complementar la descripción realizada, y contribuir a una mejor comprensión de las características de la invención, según una realización preferida de la misma, se adjunta a dicha descripción como parte integrante de la misma un conjunto de dibujos, donde se representa lo siguiente con fines ilustrativos, no limitativos:

En la figura 1, se puede observar una representación simplificada de un contenido audiovisual.

La figura 2 muestra un posible registro de la base de datos y una posible configuración de los campos en cada

registro.

REALIZACIÓN PREFERIDA DE LA INVENCIÓN

A la luz de las figuras, a continuación, se describe una realización preferida de la invención propuesta.

En la figura 1 se puede observar un posible contenido audiovisual simplificado (1), en el que se han seleccionado las subescenas (2), posteriormente se han identificado los elementos (3), teniendo cada uno de los elementos (3) una marca de tiempo para el inicio de la difusión (t1) y otra marca de tiempo para el final de la difusión (t2), y se ha asignado un archivo de información (4) a cada elemento relacionado con el elemento que lo define. Tendrá un título, y puede incluir fotografías del elemento, texto descriptivo o definiciones del mismo, tablas con características o datos, contenidos comerciales que puedan ser de interés para una persona interesada en dicho elemento, mapas de lugares, y enlaces a sitios de interés.

Las técnicas utilizadas para la separación automática del contenido audiovisual en subescenas en base a la sucesión de escenas, planos o incluso fotogramas, comprende dos partes:

- La primera parte implica una separación de planos que detecta cambios repentinos en las imágenes que se considerarán cambios de plano. Para ello, se comparan varios histogramas bloque por bloque.
- La segunda parte supone la separación en escenas e identifica aquellos planos extraídos que, por su similitud, pueden considerarse pertenecientes a una misma escena o, al menos, puede considerarse una buena idea hacerlos interactivos para aprovechar al máximo el trabajo. Para ello, se agrupan juntos el momento inicial y final de cada plano identificado.

La identificación de elementos en cada una de las subescenas se realiza en base a diferentes categorías, o tipos de archivos, que, en una realización potencial, no limitativa, pueden ser: Hombres, mujeres, niños, criaturas y robots, moda, transporte, comida y bebida, hogar, tecnología, deportes y ocio, salud y belleza, fauna y flora, armas, arte, negocio, lugar, tiempo, acción, música, referencias, curiosidades, miscelánea, citas, arquitectura, etc.

Para reconocer cada elemento particular dentro de cada categoría, se utilizan diferentes técnicas.

En el caso del reconocimiento automático de actores, el procedimiento se divide en dos fases:

- La primera es la detección de rostros en todos los fotogramas de la película. Para ello, se utiliza un algoritmo complejo que combina diversos procedimientos, tales como los gradientes orientados, para localizar determinadas formas; pirámides de imágenes, para encontrar estas formas a diferentes escalas; barrido espacial, para encontrarlas en diferentes lugares; y una toma de decisiones binaria final basada en máquinas de vectores de soporte. Dado que este procedimiento puede ser menos preciso para determinados fotogramas en los que las caras están cubiertas o giradas, se añade un seguimiento de los objetos detectados como caras y un agrupamiento de sus trayectorias, lo que ayuda a encontrarlas incluso en las condiciones más difíciles.
- La segunda parte implica la identificación específica de actores. Con este fin, se entrena un modelo de red neuronal convolucional que tiene un procedimiento de puntuación que es altamente paralelizable.

Por lo tanto, la base de datos obtenida contendrá una serie de registros, donde cada registro tendrá con los siguientes campos, como se muestra en la figura 2.

- Subescena (2)
- Elemento (3)
- Marca de tiempo para el comienzo de la difusión (t1)
- Marca de tiempo para el final de la difusión (t2)
- Un archivo

Cada archivo puede contener además una serie de campos, tales como, por ejemplo, título, fotografías, descripción, tablas de datos, contenidos comerciales, etc.

Habiendo descrito suficientemente la naturaleza de la presente invención, así como la forma de implementarla, se plantea que, dentro de su esencialidad, puede implementarse en otras realizaciones que pueden diferir en algunos detalles de la descrita a modo de ejemplo, y que igualmente recibirán la protección solicitada, siempre que estén dentro del alcance de las reivindicaciones adjuntas.

REIVINDICACIONES

1. Procedimiento de generación de una base de datos vinculada a un contenido audiovisual a través de diferentes referencias temporales, **caracterizado por que** comprende las siguientes etapas:

- reorganización del contenido audiovisual (1) en subescenas (2), en base a la sucesión de escenas, planos, fotogramas, que comprende dos partes:

  - La primera parte implica una separación de tomas, que detecta cambios repentinos en las imágenes,
  - La segunda parte implica una separación en escenas e identifica aquellos planos extraídos que, por ser muy similares, pueden considerarse pertenecientes a la misma escena,

- identificación de elementos (3) en cada una de las subescenas (2), que utiliza, como elementos de entrada (3), el contenido audiovisual previamente reordenado (1) y una base de datos con información relevante obtenida automáticamente de Internet en base a diferentes categorías, realizándose lo siguiente para cada uno de los elementos (3):

  ○ una anotación temporal e informativa de los mismos, en la que la anotación temporal indica el momento exacto en el que cada pieza de información se identifica en pantalla y deja de ser visible en pantalla, en base a la ordenación original del contenido audiovisual (1);
  ○ una anotación informativa de las diferentes relaciones (coincidencia, contenido o continente, lugar, pertenencia, parentesco, relación por compartir un tercer elemento, etc.) entre los diferentes elementos (3),

- búsqueda de correspondencias comerciales en los elementos (3) identificados,
- desduplicación, anotación y edición experta,

en las que las diferentes categorías en base a las cuales se identifican los elementos (3) de cada una de las subescenas (2) son: Hombres, mujeres, personajes, actores, niños, criaturas y robots, moda, transporte, comida y bebida, hogar, tecnología, deportes y ocio, salud y belleza, fauna y flora, armas, arte, negocio, lugar, tiempo, acción, música, referencias, curiosidades, miscelánea, citas, arquitectura,
caracterizado por que
el procedimiento de reconocimiento automático de actores se divide en dos fases:

- la primera fase es una detección de rostros en todos los fotogramas del contenido audiovisual, mediante el uso de un algoritmo complejo que combina; gradientes orientados, para localizar determinadas formas; pirámides de imágenes, para encontrar estas formas a diferentes escalas; barrido espacial, para encontrarlas en diferentes lugares; y una toma de decisiones binaria final basada en máquinas de vectores de soporte; se añade un seguimiento de los objetos detectados como caras y un agrupamiento de sus trayectorias,
- la segunda fase implica la identificación específica de actores, mediante el entrenamiento de un modelo de red neuronal convolucional que tiene un procedimiento de puntuación altamente paralelizable.

2. Procedimiento de generación de una base de datos vinculada a un contenido audiovisual a través de diferentes referencias temporales según la reivindicación 1, en el que la separación de planos que detecta cambios repentinos en las imágenes se realiza comparando varios histogramas bloque por bloque.

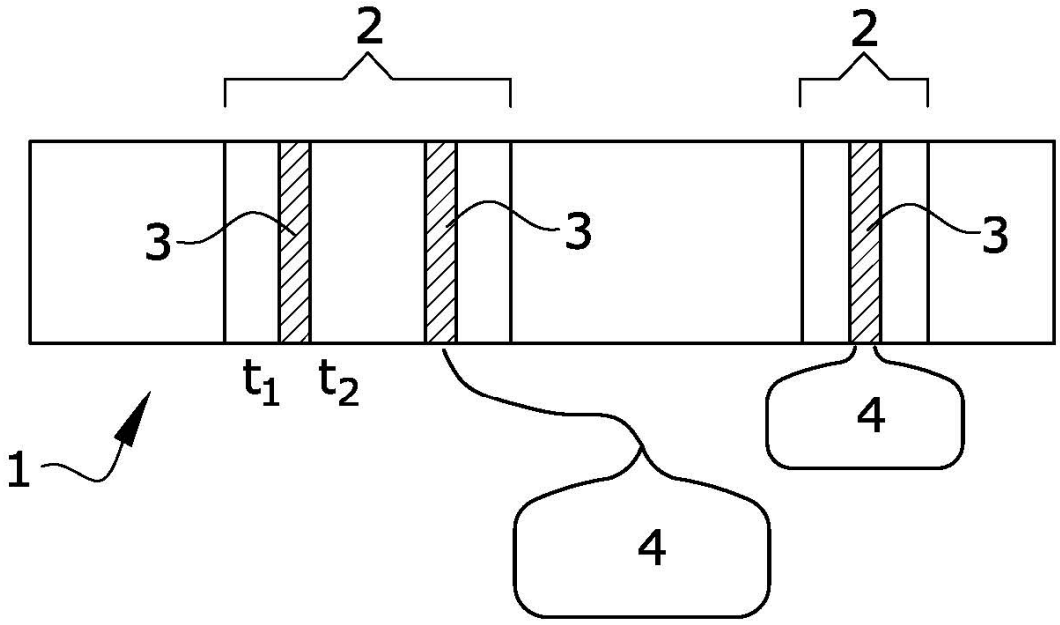

FIG.1

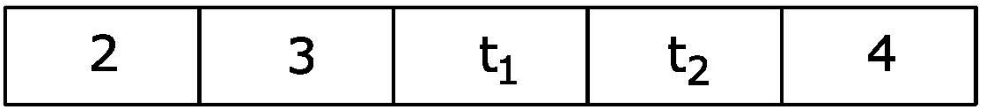

FIG.2