(19) OFICINA ESPAÑOLA DE PATENTES Y MARCAS

ESPAÑA

(11) Número de publicación: **2 731 327**

(51) Int. CI.:

G06T 7/174 (2007.01)
G06T 7/00 (2007.01)

(12) TRADUCCIÓN DE PATENTE EUROPEA T3

(86) Fecha de presentación y número de la solicitud internacional: **13.10.2014 PCT/EP2014/071928**

(87) Fecha y número de publicación internacional: **16.04.2015 WO15052351**

(96) Fecha de presentación y número de la solicitud europea: **13.10.2014 E 14795789 (8)**

(97) Fecha y número de publicación de la concesión europea: **20.03.2019 EP 3055836**

(54) Título: **Método para caracterizar imágenes adquiridas a través de un dispositivo médico de vídeo**

(30) Prioridad:

11.10.2013 US 201361889711 P
23.05.2014 US 201462002325 P

(45) Fecha de publicación y mención en BOPI de la traducción de la patente:
15.11.2019

(73) Titular/es:

MAUNA KEA TECHNOLOGIES (100.0%)
9, rue d'Enghien
75010 Paris, FR

(72) Inventor/es:

LINARD, NICOLAS;
ANDRE, BARBARA;
DAUGUET, JULIEN y
VERCAUTEREN, TOM

(74) Agente/Representante:

ELZABURU, S.L.P

DESCRIPCIÓN

Método para caracterizar imágenes adquiridas a través de un dispositivo médico de vídeo

Campo de la invención

La presente invención se refiere de manera general al procesamiento de imágenes y video y, en particular, a un sistema y método para caracterizar la capacidad de interpretación de imágenes adquiridas en secuencias y especialmente imágenes adquiridas a través de un dispositivo médico de video.

Antecedentes

Los dispositivos de adquisición de video generan cantidades masivas de datos. El uso eficiente de estos datos es de importancia para edición de video, resumen de video, visualización rápida y muchas otras aplicaciones relacionadas con la gestión y el análisis de video.

Como se ilustra en Koprinskaa et al., ("Temporal video segmentation: A survey.", Signal Processing: Image Communication, 16 (5), 477-500 (2001)), la segmentación de video temporal es un paso clave en la mayoría de las herramientas de gestión de vídeo existentes. Se han desarrollado muchos tipos diferentes de algoritmos para realizar la segmentación temporal.

Las primeras técnicas se centraron en la detección de límites de corte o la agrupación de imágenes usando diferencias de píxeles, comparaciones de histogramas, diferencias de bordes, análisis de movimiento y similares, mientras que métodos más recientes, tales como los presentados en los documentos US7783106B2 y US8363960B2, también han usado métricas de similitud, clasificación y agrupación de imágenes para lograr la misma meta.

En algunas aplicaciones como las de Sun, Z. et al. ("Removal of non-informative frames for wireless capsule endoscopy video segmentation", Proc. ICAL páginas 294-299 (2012)) y Oh, J.-H. et al. ("Informative frame classification for endoscopy video", Medical Image Analysis, 11 (2), 110-127 (2007)), el problema de segmentación de video temporal se puede reformular como un problema de clasificación que distingue entre imágenes informativas y de ruido.

En el documento US20070245242A1, se ha acoplado segmentación de video temporal con el cálculo de similitud entre escenas para producir resúmenes de video.

En el área de dispositivos médicos, y en particular en el campo de la endoscopia, la evaluación de patrones de movimiento ha jugado un papel importante en el análisis de videos largos.

En el documento US7200253B2, se describe un sistema para evaluar el movimiento de una cápsula de formación de imágenes digerible y para mostrar la información de movimiento frente al tiempo.

Se usó información de movimiento similar en el documento US20100194869A1 para segmentación de video temporal de videos de endoscopia. La proyección rápida del contenido del video se implementa mostrando solo la primera imagen de cada segmento temporal; por lo tanto omitiendo todas las demás imágenes.

Para abordar la misma meta de proyección rápida de video en endoscopia pero sin omitir imágenes, el documento US20100194869A1 se basa en evaluación de movimiento para calcular una velocidad de reproducción inversamente proporcional al movimiento estimado.

Basándose en herramientas de mosaico de video, se describe en el documento US8218901B2 una representación eficiente de videos endomicroscópicos en la que se han superpuesto imágenes consecutivas.

Para facilitar la interpretación de videos endomicroscópicos completos, André, B. et al. ("A Smart Atlas for Endomicroscopy using Automated Video Retrieval", Medical Image Analysis, 15 (4), 460-476 (2011)) propuso un método que se basa en similitud visual entre un video actual y videos de una base de datos externa para mostrar visualmente casos similares pero anotados en relación al video actual.

Un enfoque similar se describe en André, B. et al. ("Learning Semantic and Visual Similarity for Endomicroscopy Video Retrieval", IEEE Transactions on Medical Imaging, 31 (6), 1276-1288 (2012)) para complementar la similitud visual con la información semántica. En un tema relacionado (André, B. et al. "An image retrieval approach to setup difficulty levels in training systems for endomicroscopy diagnosis", MICCAI (páginas 480-487). Beijing: LNCS (2010)) presentó un medio de evaluación de un nivel de dificultad asociado con la interpretación de un video de endomicroscopia dado.

En escenarios clínicos, puede necesitar ser realizado un análisis de video durante el procedimiento. Para trabajar alrededor del problema del tiempo de cálculo (US20110274325A1) describe un método que aprovecha un almacenador temporal inmóvil de imágenes consecutivas para realizar tareas intensivas en cálculo mientras que se continúa con la adquisición de imágenes.

Como se ilustra en el trabajo mencionado anteriormente, la técnica anterior muestra que existe una necesidad real de un uso eficiente de los videos adquiridos con un dispositivo médico. Aunque el uso eficiente de los datos de video se ha abordado tanto en escenarios clínicos como no clínicos, ninguno de los enfoques anteriores enseña un método para caracterizar la capacidad de interpretación de las imágenes que componen un video adquirido con un dispositivo médico.

El documento de patente US2006/293558 describe un método para la medición automatizada de métricas que reflejan la calidad de un procedimiento colonoscópico. El documento de patente US2011/301447 describe un método para clasificar y anotar características clínicas en un video médico aplicando un análisis probabilístico de las relaciones dentro un cuadro o entre cuadros en las partes vecinas tanto espacial como temporalmente de los cuadros de video.

Compendio

Un objeto de la presente invención es mejorar la eficiencia del uso de los datos adquiridos con un dispositivo médico de video. La invención se refiere a un método según la reivindicación 1 y a un sistema según la reivindicación 8.

Esto permite al usuario de los datos de video médico centrar su atención en las partes más interpretables de la adquisición.

Los dispositivos médicos de video para adquirir imágenes pueden ser cualquier dispositivo conocido por los expertos en la técnica, incluyendo, pero no limitados a: endomicroscopios, dispositivos de tomografía de coherencia óptica, endoscopia clásica, endoscopia de alta definición, endoscopia de formación de imágenes de banda estrecha, endoscopia FICE®, enteroscopia de doble balón, endoscopia con zoom, endoscopia de fluorescencia, ecografía 2D/3D, ecoendoscopia o cualquier otra modalidad de formación de imágenes intervencionista.

Una salida del primer algoritmo puede ser un valor entre un conjunto de valores discretos. El valor puede ser típicamente un valor alfanumérico. En este caso, la línea de tiempo puede estar formada por regiones temporales correspondientes a imágenes consecutivas con igual salida. Estas regiones temporales pueden constituir una clasificación temporal o segmentación temporal del video de interés. En el caso particular de una salida binaria, estas regiones temporales pueden constituir una segmentación binaria temporal del video de interés.

Una salida del primer algoritmo también puede ser un valor escalar o vectorial continuo. En algunos casos, el algoritmo puede tener dos salidas diferentes, siendo una un valor discreto, siendo la otra un valor escalar o vectorial continuo. Un ejemplo que pertenece al diagnóstico sería como tal; la primera salida discreta indicaría una clase de diagnóstico predicha mientras que la otra salida continua indicaría la probabilidad de pertenecer a cada clase de diagnóstico predefinida.

Según una variante, los valores de la salida del primer algoritmo se representan por colores, siendo dichos colores superpuestos sobre la línea de tiempo mostrada. Los valores de la salida del primer algoritmo también se pueden mostrar además de la imagen mostrada actualmente.

Según una variante, cuando se definen regiones temporales correspondientes a imágenes consecutivas con igual salida, el método puede comprender además seleccionar al menos una región temporal y extraer del almacenador temporal las imágenes correspondientes a dichas regiones temporales. Las imágenes extraídas se pueden almacenar, por ejemplo, en un dispositivo de almacenamiento. Las imágenes extraídas también se pueden procesar usando un segundo algoritmo y la salida del segundo algoritmo mostrada. Por ejemplo, el segundo algoritmo puede ser un algoritmo de recuperación de imagen o video basado en contenido, un algoritmo de mosaico de imagen o video, un algoritmo de clasificación de imagen o video o similar.

La selección de al menos una región temporal se puede realizar o bien de manera totalmente automática o bien puede depender de alguna interacción del usuario. Por ejemplo, el segundo algoritmo puede utilizar el conjunto completo de imágenes para todas las regiones temporales segmentadas. También se puede basar en un algoritmo de selección simple o puede requerir la interacción del usuario para elegir las regiones seleccionadas.

Según una variante, el primer algoritmo puede generar resultados intermedios asociados con cada imagen del almacenador temporal. Por lo tanto, el método puede comprender almacenar dichos resultados intermedios en una base de datos interna. La base de datos interna se puede actualizar, por ejemplo, tras cada actualización del almacenador temporal. Según una variante, el primer algoritmo puede usar resultados intermedios de la base de datos interna.

Cuando las imágenes correspondientes a regiones temporales se extraen y procesan usando un segundo algoritmo, dicho segundo algoritmo podría usar los resultados intermedios de la base de datos interna.

Según una variante, un criterio de capacidad de interpretación puede ser estabilidad cinética.

Por ejemplo, la estabilidad cinemática se puede evaluar usando análisis de coincidencias de características. Las características se pueden situar en una cuadrícula regular o perturbada. Por ejemplo, la perturbación de la cuadrícula se acciona por la prominencia de imágenes locales.

Se puede usar un mapa de votos para seleccionar y contar el número de votos que determina la estabilidad cinemática.

La estabilidad cinemática se puede realizar inicialmente de una manera por pares sobre imágenes consecutivas, y se puede realizar un paso de procesamiento de señal sobre la señal de estabilidad cinemática inicial para proporcionar la salida de estabilidad cinemática.

Según la aplicación clínica de destino, el criterio de capacidad de interpretación puede ser al menos uno entre la lista no limitativa: estabilidad cinemática, similitud entre imágenes, por ejemplo, similitud entre imágenes dentro del almacenador temporal, probabilidad de pertenecer a una categoría, por ejemplo, probabilidad de pertenecer a una categoría dada de un conjunto predeterminado de categorías, calidad de imagen, dificultad para proponer un diagnóstico o una interpretación semántica, tipicidad o atipicidad de imágenes o ambigüedad de imágenes.

Además, un criterio de capacidad de interpretación puede usar la similitud entre imágenes dentro del almacenador temporal e imágenes dentro de una base de datos externa.

Los objetos, características, efectos operacionales y méritos anteriores y otros de la invención llegarán a ser evidentes a partir de la siguiente descripción y los dibujos adjuntos.

Breve descripción de los dibujos

La Figura 1 es una vista esquemática de un video adquirido con un dispositivo médico que se muestra en asociación con una línea de tiempo que resalta las regiones temporales de suficiente capacidad de interpretación.

La Figura 2 es una vista esquemática del video adquirido con un dispositivo médico que se muestra en asociación con una línea de tiempo que resalta las regiones temporales etiquetadas según valores discretos.

La Figura 3 es una vista esquemática del video adquirido con un dispositivo médico que se muestra en asociación con una línea de tiempo que presenta una evolución temporal de una salida continua.

La Figura 4A es una vista esquemática de un video adquirido con un dispositivo médico que se muestra en asociación con una línea de tiempo que resalta las regiones temporales de suficiente capacidad de interpretación y la Figura 4B ilustra un conjunto de casos que comprenden video y metadatos adicionales que se han seleccionado de una base de datos externa según un criterio de similitud con respecto a la región temporal actual.

La Figura 5 es un diagrama que ilustra la coincidencia de imágenes consecutivas y el umbral basado en la calidad de coincidencia.

La Figura 6A y la Figura 6B son diagramas que ilustran una estrategia de refinamiento para colocar descriptores de imágenes locales.

La Figura 7A y la Figura 7B son diagramas que ilustran el procesamiento de una línea de tiempo de etiqueta de capacidad de interpretación inicial para la eliminación de valores atípicos.

Descripción detallada

En un modo básico de operación, un dispositivo de adquisición de video médico actúa como una entrada a nuestro sistema. El procesamiento de video en tiempo real se puede realizar durante la adquisición, y se pueden mostrar las imágenes. Mientras tanto, las imágenes se ponen en cola en un almacenador temporal finito de primero en entrar, primero en salir (FIFO), mientras que los resultados potenciales del cálculo en tiempo real se pueden almacenar en una base de datos interna.

En un segundo modo de operación, nuestro sistema puede usar un video que se grabó previamente por un dispositivo médico de video como entrada. En este caso, las imágenes que componen el video también se ponen en cola en un almacenador temporal FIFO. Si el cálculo en tiempo real se realizó durante la adquisición y se registró junto con las imágenes, los resultados de los cálculos se pueden cargar en una base de datos interna.

En ambos modos de operación, la base de datos interna se podría actualizar cada vez que se actualiza el almacenador temporal de imágenes.

Tras la revisión de las imágenes almacenadas en el almacenador temporal de entrada, nuestro sistema caracteriza automáticamente la capacidad de interpretación de las imágenes que componen el almacenador temporal y une su salida a una línea de tiempo correspondiente al contenido de las imágenes en el almacenador temporal. La caracterización de la capacidad de interpretación puede basarse en cálculos en tiempo real realizados previamente, así como en cálculos de procesamiento posterior.

Dependiendo de la aplicación clínica de destino, la capacidad de interpretación se puede caracterizar según diferentes criterios subyacentes. Estos criterios pueden estar relacionados con nociones diferentes tales como, pero no limitadas a:

- estabilidad cinemática,
- similitud de las imágenes dentro del almacenador temporal,
- cantidad de información nueva descubierta por una imagen con respecto a las anteriores,
- calidad de imagen,
- presencia e importancia de artefactos,
- naturaleza y tipo de artefactos de imágenes,
- probabilidad de pertenecer a una categoría dada, por ejemplo, clase de diagnóstico, dentro de un conjunto predefinido de categorías,
- tipicidad o atipicidad de imagen,
- ambigüedad de imagen, por ejemplo, ambigüedad visual con respecto a un conjunto de clases de diagnóstico,
- dificultad de proponer un diagnóstico o una interpretación semántica.

En la endomicroscopia, una sonda de formación de imágenes típicamente se pone en contacto con, o se pone cerca de, el tejido para adquirir imágenes. La adquisición en tiempo real se puede realizar gracias al escaneado de espejo a través del campo de visión. Debido al movimiento continuo de la sonda con respecto al tejido durante el escaneado de espejo, las imágenes se someten a artefactos de movimiento. La magnitud de estos artefactos se correlaciona típicamente con la capacidad de interpretación de las imágenes. De hecho, si se observa demasiado movimiento, la arquitectura celular del tejido se puede distorsionar fuertemente y puede conducir a imágenes que son difíciles de interpretar.

En la mayoría de los dispositivos médicos de video, un usuario guiará una sonda de formación de imágenes o un detector de formación de imágenes sobre o dentro del paciente y permanecerá en un área durante un tiempo que se correlaciona con el interés y la capacidad de interpretación del área.

Como tal, en algunas realizaciones de la presente invención, la capacidad de interpretación puede ser una función del movimiento de la sonda de formación de imágenes con respecto a su objeto. En otras palabras, en algunas realizaciones, la capacidad de interpretación se puede caracterizar en términos de estabilidad cinemática.

En otros escenarios, relacionar la capacidad de interpretación con características de cálculo basadas en modelo podría ser complejo de realizar. Sin embargo, podría darse el caso de que una base de datos externa de imágenes haya sido adquirida y anotada previamente según algunos criterios de capacidad de interpretación por usuarios expertos. En otras realizaciones de la invención, se pueden usar algoritmos de aprendizaje por máquina para inferir la capacidad de interpretación de nuevas imágenes aprendiendo de la base de datos externa anotada disponible.

Todavía en otros escenarios, la capacidad de interpretación de un video podría depender de la identificación de la variabilidad de las imágenes adquiridas con el dispositivo médico de video. En este caso, el usuario pudiera estar interesado en tener imágenes similares que se agrupan juntas. Las realizaciones de la invención pueden usar otras formas de aprendizaje por máquina para caracterizar la capacidad de interpretación agrupando imágenes según su similitud.

Se pueden usar varias técnicas de visualización para mostrar al menos una salida de caracterización de imágenes, mientras que el usuario está reproduciendo un video ya grabado, reproduciendo un video almacenado temporalmente, o visualizando la imagen que está siendo adquirida actualmente por el dispositivo médico de video. Para cada imagen almacenada en el almacenador temporal, el valor de salida calculado puede ser información discreta, tal como una letra o etiqueta, o un valor escalar o vectorial continuo.

Como se ilustra en las Figuras 1 a 4, los valores de salida se pueden unir a una línea de tiempo 11 del video, donde la línea de tiempo 11 comprende un cursor 15 temporal que indica el tiempo de una imagen 10 mostrada en la línea de tiempo 11. Según una realización de la invención, los colores que representan los valores de salida calculados para todas las imágenes del video se superponen directamente en la línea de tiempo del video, con el fin de proveer al usuario con una vista cronológica de la capacidad de interpretación de imágenes dentro del video. También se puede mostrar una leyenda que explica los valores de salida para facilitar la comprensión del usuario.

El valor de salida calculado para la imagen 10 mostrada actualmente, o un color que representa este valor, también se puede mostrar además de la imagen mostrada actualmente, con el fin de duplicar el valor de salida potencialmente oculto por el cursor 15 temporal actual, como se ilustra en la Figura 2 (elemento 29).

En el caso de una salida discretizada, cada valor de salida se puede representar por un color predefinido. La Figura 1 ilustra el caso de una salida binaria representada por un color gris 12 o blanco 14 predefinidos en la línea de tiempo 11. El color gris (respectivamente blanco) en una posición dada en la línea de tiempo puede indicar, por ejemplo, que la imagen en esta posición en el video es de suficiente (respectivamente, insuficiente) calidad, o que es estable (respectivamente, inestable) cinemáticamente con respecto a la imagen anterior.

La Figura 2 ilustra el caso de una salida discreta en cuatro valores distintos, siendo cada uno de ellos representado por un color distinto: blanco 24, gris claro (punto) 28, gris oscuro (onda) 22 o negro (rayado) 26. Si hay una relación de orden entre los valores de salida, este orden se puede mantener entre los niveles de gris a los que se asignan los valores. Si no se puede elegir un orden aleatorio. Estos cuatro valores de gris pueden indicar, por ejemplo, cuatro niveles de capacidad de interpretación ordenados como: no interpretable en absoluto, difícilmente interpretable, parcialmente interpretable, completamente interpretable. También pueden indicar, por ejemplo: no suficientemente interpretable, suficientemente interpretable y perteneciente al tipo de tejido A, suficientemente interpretable y perteneciente al tipo de tejido B, suficientemente interpretable y perteneciente al tipo de tejido C, donde no hay una relación de orden entre estos tres tipos de tejido.

En el caso de una salida continua (Figura 3), cada valor de salida todavía se puede representar por un color que se puede determinar automáticamente asignando el valor de salida, por ejemplo, a un valor de triplete RGB, HSL, HSV o YUV. Se puede usar una tabla de búsqueda para convertir salidas continuas en colores. Si la salida es un vector n-dimensional con $n \leq 3$, se puede adaptar el mismo proceso de asignación. Si la salida es un vector n-dimensional con $n > 3$, el proceso de asignación se puede calcular, por ejemplo, a partir de un Análisis Tridimensional de Componentes Principales. El valor 32 de color continuo puede indicar, por ejemplo, la calidad de la imagen o el porcentaje de regiones locales en la imagen que coinciden con regiones locales en la imagen anterior. La Figura 3 ilustra cómo tal visualización puede permitir al usuario apreciar la evolución temporal de un valor de capacidad de interpretación de imagen continua dentro del video.

En el caso particular donde el usuario solo está visualizando la imagen que se está adquiriendo actualmente, se puede calcular sobre la marcha para esta imagen al menos un valor de salida. Dicho valor de salida, o un color que representa este valor, se puede mostrar además de esta imagen actualmente adquirida.

En muchos casos, el usuario de los datos de video no es solo el médico directamente, sino que puede ser un segundo algoritmo de cálculo. Describimos una realización de la invención en la que la capacidad de interpretación caracterizada se usa para realizar cálculos adicionales únicamente en regiones temporales de capacidad de interpretación adecuada.

En el caso de una salida discreta unida a la línea de tiempo, las regiones temporales se pueden definir en la línea de tiempo como los segmentos más grandes correspondientes a imágenes consecutivas con el mismo valor de salida. La región temporal actual se define como la región temporal a la que pertenece el cursor temporal actual. Las interacciones del usuario se pueden definir entonces, permitiendo que el usuario potencialmente:

- Deshabilite o habilite la visualización de al menos una salida;
- Mueva el cursor temporal al siguiente punto de tiempo más cercano que pertenezca a una región temporal distinta de la región temporal actual;
- Mueva el cursor temporal al punto de tiempo anterior más cercano, el punto de tiempo que pertenece a una región temporal distinta de la región temporal actual;
- Seleccione al menos una región temporal;
- Refine y modifique las regiones temporales;
- Almacene las imágenes asociadas con la región temporal seleccionada en un dispositivo de almacenamiento, y potencialmente las anote;
- Lance al menos un segundo algoritmo en la región temporal actual, o en al menos una región temporal seleccionada por el usuario. Dicho segundo algoritmo usa como entrada la secuencia o secuencias secundarias de imagen asociadas con la región o regiones temporales. Un segundo algoritmo puede consistir, por ejemplo, en clasificar o crear un mosaico de esta secuencia o secuencias de imagen de entrada.
- Visualice al menos una salida creada por al menos un segundo algoritmo, dicho segundo algoritmo siendo lanzado de manera potencialmente automática en la región temporal actual. Ventajosamente, esta segunda salida se puede mostrar de manera automática, sin requerir ninguna interacción del usuario.

En este escenario con un segundo algoritmo, la capacidad de interpretación también se puede definir en términos de cómo se usan los datos por los cálculos posteriores. Se pueden usar técnicas de mosaico de video dedicadas para ampliar el campo de visión de un video alineando y fundiendo muchas imágenes consecutivas de una secuencia de video. Este proceso solo funciona si las imágenes consecutivas comparten una superposición suficiente y si se

observa algún movimiento entre el dispositivo de formación de imágenes y el objeto de interés. En una realización de la invención, la capacidad de interpretación se puede definir en términos de estabilidad cinemática y se pueden aplicar herramientas de mosaico de video sobre las regiones de suficiente capacidad de interpretación.

Según otra realización, si el mosaico de video se ha aplicado en al menos dos secuencias secundarias de video para producir imágenes de campo de visión más grandes, la técnica de mosaico de imagen se puede usar posteriormente para detectar y asociar mosaicos de imagen coincidentes, registrarlos espacialmente y fundirlos para crear imágenes de campo de visión incluso más grandes. La detección de mosaicos coincidentes también puede depender de la interacción del usuario.

Para facilitar la capacidad de interpretación de las secuencias de video adquiridas con un dispositivo médico de video, se pueden usar herramientas de recuperación de video basadas en contenido como medio para aprovechar el razonamiento basado en similitud. Para una secuencia de video dada, se pueden presentar al médico, desde una base de datos externa, un conjunto de casos visualmente similares a la secuencia de video y previamente anotados por expertos. Las secuencias de video adquiridas con un dispositivo médico pueden contener partes de capacidad de interpretación variable, y pueden contener una mezcla de diferentes tipos de tejidos. Como tal, la relevancia de estas herramientas de recuperación de video basadas en contenido puede depender críticamente de elegir, según se solicite, una parte de un video que sea consistente en términos de capacidad de interpretación. En una realización de la invención, la caracterización de capacidad de interpretación se usa para dividir automáticamente un video de entrada en partes secundarias de suficiente capacidad de interpretación; dichas partes secundarias que se usan para construir al menos una consulta para un algoritmo de recuperación de video basado en contenido.

Según una variante, las partes secundarias se pueden usar de diferentes maneras para crear la consulta para el algoritmo de recuperación basado en contenido. Por ejemplo, cada parte secundaria se puede usar para crear una consulta independiente. Alternativamente, todo el conjunto de partes secundarias se puede usar para crear una sola consulta. Aún de manera alternativa, se le puede solicitar al usuario que seleccione un subconjunto de estas partes secundarias para crear una sola consulta.

Según otra variante, el usuario también tiene la capacidad de refinar la segmentación temporal proporcionada por el primer algoritmo antes de reanudar con el segundo algoritmo.

La Figura 4A y la Figura 4B ilustran el caso en el que el segundo algoritmo es un procesamiento de recuperación de video basado en contenido que se ha lanzado en la región temporal actual del video de interés. La salida creada por este segundo algoritmo y mostrada al usuario consta de tres videos (41, 42, 43) de referencia junto con sus anotaciones (44, 45, 46), donde las anotaciones incluyen, por ejemplo, la clase de diagnóstico del video de referencia. Estos videos de referencia se han extraído de una base de datos externa como los más similares visualmente al conjunto de imágenes contiguas asociadas con la región temporal actual seleccionada por el cursor 15 en la Figura 4A.

Según otra realización, en el caso de etiquetas discretas, la invención también permite ejecutar automáticamente un segundo algoritmo en cada una de las regiones.

Según otra realización, en el caso de etiquetas discretas, la invención también permite almacenar automáticamente el contenido de todas las regiones etiquetadas independientemente, o en el caso secundario de etiquetas binarias, almacenar en un dispositivo de almacenamiento la concatenación de todas las regiones temporales correspondientes a una etiqueta dada.

Estabilidad cinemática

Los enfoques basados en registro de imágenes se pueden usar para identificar de manera cinemática regiones temporales estables dentro de secuencias de video. Esto se puede hacer, por ejemplo, registrando realmente imágenes temporalmente consecutivas y luego analizando la calidad de la transformación espacial encontrada por el algoritmo de registro.

Otro ejemplo sería usar solo un subconjunto de los pasos de un algoritmo de registro de imágenes y analizar la calidad de los resultados proporcionados por este subconjunto. Esto se puede hacer en el caso de algoritmos basados en coincidencia de características, donde observar la coherencia de las coincidencias de características con un modelo de transformación espacial podría permitir a uno inferir información sobre la estabilidad cinemática.

Las mismas coincidencias de características también se pueden analizar en términos de coherencia local para obtener un resultado que sea más robusto al error de modelado para la transformación espacial.

Métodos más avanzados que registran imágenes múltiples al mismo tiempo, tales como el presentado en (Vercauteren, Perchant, Lacombe, y Savoire, 2011) también se pueden usar para inferir la estabilidad cinemática.

La Figura 5 ilustra con más detalle una posible realización para analizar la estabilidad cinemática dependiendo de una cuadrícula de características. Cada imagen 52 de una serie de imágenes 51 secuenciales almacenadas en el almacenador temporal en el almacenador temporal se asocia con una cuadrícula (57) de ubicaciones espaciales en

la imagen (paso I). Cada punto (58) de la cuadrícula (57) está asociado con una región espacial local con una escala dada alrededor de ese punto, estando cada región asociada a su vez con un descriptor o firma numérica. Haciendo coincidir cada descriptor de una imagen con un descriptor numéricamente similar de la imagen anterior (paso III), permite a uno hacer coincidir cada punto de una cuadrícula (59) en una imagen (54) con otro punto en una cuadrícula (57) de la imagen (53) anterior; dichos puntos coincidentes se asocian con regiones locales que son similares visualmente gracias al descriptor que es similar. El análisis de las coincidencias se realiza entonces para evaluar su coherencia local o su coherencia con respecto a un modelo de transformación espacial predefinido. Si se estima que la coherencia es demasiado baja, la imagen se considerará como inestable cinemáticamente con respecto a la anterior.

Representar una imagen como una cuadrícula de descriptores a menudo se denomina descripción de imagen local densa o descripción densa en breve. De manera intercambiable, también podemos usar el término basado en cuadrícula para estos enfoques. Cada punto de la cuadrícula también se puede denominar punto clave.

Una ventaja de confiar en la descripción de imagen local basada en cuadrícula es que se pueden usar los mismos descriptores tanto para caracterizar la estabilidad de las secuencias de video como para realizar una tarea de recuperación de video basada en contenido. Esto permitiría ahorrar tiempo de cálculo en el caso de que hayan de ser realizadas ambas tareas.

La descripción de imagen local, basada en cuadrícula o no, se usa ampliamente en visión por ordenador, reconocimiento de patrones y formación de imágenes médicas y ha servido para una variedad de propósitos. Ahora están disponibles muchos descriptores diferentes, incluyendo, pero no limitado a, LBP, SIFT, SURFT, HoG, GLOH y similares. Dependiendo de la aplicación exacta, diferentes requisitos de cálculo, requisitos de rendimiento, requisitos de facilidad de implementación, etc., pueden conducir a cada opción.

La localización de puntos clave es algunas veces crucial en la visión por ordenador. En la mayoría de los casos, una cuadrícula regular de puntos clave no es la elección más común. En algunos escenarios, es ventajoso tener puntos clave que se sitúen con precisión en los puntos más destacados.

Típicamente, la primera y segunda derivadas de la imagen se pueden usar para detectar los puntos más destacados en cuanto a estimar la escala de la región local correspondiente. El bien conocido detector de Harris, por ejemplo, usa la traza de la matriz hessiana para detectar esquinas. Otros detectores usan un estimador laplaciano que es el determinante de la matriz hessiana. Una vez que se detectan los puntos más destacados, los puntos clave se pueden establecer en las ubicaciones correspondientes con una escala proporcionada por el detector de prominencia.

Como en el caso de la cuadrícula, los puntos clave derivados de los puntos destacados se pueden usar entonces para calcular los descriptores de imagen locales. Una medición de discrepancia se puede calcular entonces entre descriptores, dando como resultado coincidencias de puntos clave, que se pueden analizar o regularizar mediante un modelo de transformación. Modelos de transformación de ejemplo incluyen, pero no se limitan a, modelos proyectivos muy adecuados para aplicaciones de cámara, modelos de traslado y modelos de transformación de cuerpo rígido, ambos muy adecuados para aplicaciones de microscopía y modelos deformables que pueden abarcar deformación del tejido.

Los métodos de coincidencia de puntos clave típicamente tienen varias restricciones. Por ejemplo, a menudo es el caso de que un buen rendimiento de coincidencia exige que los puntos clave se localicen en puntos suficientemente destacados, pero también estén bien distribuidos sobre el campo de imagen.

Tener los puntos clave situados en puntos suficientemente destacados típicamente hará la localización de los puntos clave más robusta con respecto al cambio de los parámetros de formación de imágenes. Por lo tanto, esto puede mejorar el rendimiento del algoritmo de registro haciendo la coincidencia de puntos clave más precisa.

Durante el proceso de coincidencia de puntos clave, a menudo es mejor tener una respuesta única mientras que se trata de asociar un punto clave con muchos otros. A menudo, también es deseable evitar tener regiones espaciales en la imagen sin puntos clave. Esto representa una buena distribución de los puntos clave.

También es a menudo ventajoso elegir descriptores que son invariantes bajo diferentes efectos de adquisición, incluyendo pero no limitados a:

- Cambios de intensidad. La señal de imagen observada, de hecho, puede cambiar dependiendo de la reflexión de la luz global y local, de la potencia de la iluminación, del efecto de fotoblanqueado, de los artefactos de formación de imágenes, etc.

- Distorsiones espaciales. La morfología observada del área descrita puede cambiar dependiendo del punto de vista; el tejido puede cambiar entre diferentes imágenes debido a la respiración, los latidos del corazón, el contacto con instrumentos; el usuario puede cambiar el zoom del instrumento; el dispositivo puede producir artefactos, etc.

En algunos escenarios, el proceso de medición de discrepancia y descripción puede beneficiarse de imitar la visión humana tan cerca como sea posible. Es al menos lo más ventajoso elegir a menudo una pareja de discrepancia-descripción suficientemente relevante para asociar correctamente la región de una imagen a otra la mayoría del tiempo.

Aunque la detección de puntos destacados seguida de una descripción estándar de región local responde a la mayoría de las restricciones en varias aplicaciones, se ha demostrado que falla encontrando regiones destacadas bien distribuidas en muchos problemas de formación de imágenes médicas diferentes. Las imágenes médicas son, de hecho, a menudo suaves pero texturizadas y carecen de los bordes de las esquinas que requieren muchas herramientas específicas de visión por ordenador.

Para responder a estas restricciones en el contexto de formación de imágenes médicas, la aplicación de una descripción basada en cuadrícula a escalas fijas sobre imágenes médicas a menudo es una elección interesante. La información se puede distribuir, de hecho, en todas partes en muchas imágenes médicas.

Confiar en la descripción basada en cuadrícula con el propósito de registro se considera a menudo una tarea desafiante. En comparación con los métodos basados en detección de prominencia, la elección de la pareja descripción-discrepancia tiene más impacto de la precisión correspondiente. También generó un número significativamente mayor de coincidencias atípicas que necesitan ser manejadas por el método.

Algunos dispositivos de escaneado de formación de imágenes que se usan en el campo clínico también pueden conducir a artefactos de movimiento bastante fuertes. Si el tejido está en contacto con una sonda de formación de imágenes, esto puede dar como resultado deformaciones complejas de predecir o impredecibles.

A continuación, nos centramos en un descriptor de ejemplo, el descriptor SIFT que se ha demostrado que es eficiente en algunos problemas de formación de imágenes médicas, para ilustrar algunos de los conceptos de los descriptores de imágenes locales. Se debería recordar que se puede usar cualquier otro descriptor de imagen local.

El algoritmo SIFT (Transformación de Característica Invariante de Escala) incluye tanto detección de puntos clave como descripción de imagen. Con el enfoque de la descripción basada en cuadrícula, la detección de puntos clave puede no ser requerida y solo se puede usar la parte del descriptor de SIFT.

La información de gradientes se puede usar para describir una región local de una imagen. Más específicamente, los histogramas de gradientes orientados han mostrado resultados eficientes. Dentro de una región de imagen local, se puede crear un histograma en diferentes regiones secundarias de la región local para resumir la magnitud de los gradientes en la región secundaria según algunos contenedores de orientación discretizados. Toda la región de imagen local se puede describir entonces mediante un histograma final que concatena todos los histogramas de regiones secundarias en un orden predefinido.

La noción de formación de ventanas también juega a menudo un papel importante para ponderar mejor la contribución de la magnitud del gradiente sobre el descriptor. La formación de ventanas se aplica típicamente en todo el descriptor. Los núcleos gaussianos son la elección de formación de ventanas más común, pero se puede usar cualquier otro tipo de ventana (Blackman, Hamming, Coseno ...).

Las ventanas gaussianas tienen un soporte infinito, una implementación práctica de ella puede depender de truncamiento o formas más complejas de aproximaciones, tales como filtrado recursivo. En muchos casos, puede ser ventajoso truncar el soporte de la ventana gaussiana después de una distancia que depende de la desviación estándar σ de la ventana gaussiana. Típicamente, la distancia de truncamiento r se puede elegir para ser proporcional a σ. Por ejemplo, es clásico usar r = σ /2, pero se podría usar cualquier otra relación.

Una vez que se ha definido una estrategia de formación de ventanas, los valores de formación de ventanas se pueden usar en la creación del descriptor ponderando cada información de gradiente según la función de formación de ventanas durante la creación del histograma final.

En algunos casos, pudiera ser ventajoso obtener descriptores locales que son invariables bajo cualquier rotación de la imagen. Esto se puede lograr por muchos medios diferentes, incluyendo, pero no limitado a:

- encontrar un modo o una media de la orientación dentro de toda la región local y reorientar la región o los valores de gradiente según esta orientación principal

- usar bandas de forma circular para subdividir la región local en subregiones

- La definición de una orientación principal para la región de descriptores se puede hacer, por ejemplo, calculando un histograma de orientación de primer gradiente en toda la región local del descriptor. Este proceso de creación de histograma puede ser diferente al de creación de histograma de región secundaria, por ejemplo:

- el número de contenedores angulares usados para calcular la orientación principal puede ser ventajosamente mayor que el número de contenedores angulares usados para calcular el histograma de región secundaria. Esto

puede permitir tener una estrategia de reorientación más precisa que conduce potencialmente a una invariancia más alta con respecto a los cambios de rotación.

- se podría usar una función de formación de ventanas diferente para ponderar la contribución de cada muestra de gradiente.

Si la orientación principal se define como un modo del histograma de orientación de toda la región local, el pico más alto en este histograma de gradiente proporcionará el valor de esta orientación principal. De manera similar, se puede querer un valor medio, en cuyo caso el uso de una media de Fréchet en el histograma de orientación pudiera ser ventajoso para tener en cuenta la envoltura de ángulos a 360º. Encontrar el pico también puede beneficiarse del uso de cierta forma de regularización ajustando un modelo local, tal como uno de hendidura o uno gaussiano, para identificar la ubicación del pico con precisión de contenedor secundario y de una manera potencialmente más robusta.

Si se usa un modo para la definición, podemos querer también usar varios modos diferentes para crear varios descriptores, uno por modo seleccionado. La selección de varios modos se puede hacer, por ejemplo, sobre la base de una comparación entre el pico más alto y los picos secundarios. Si la altura del pico secundario es lo suficientemente cercana al más alto, por ejemplo, por encima de alguna fracción de él, pudiera ser interesante mantenerla. La determinación del umbral correspondiente se podría hacer a través de diferentes medios, incluyendo, pero no limitado a, regla de oro, ensayo y error, validación cruzada, optimización y similares.

Una vez que se da la orientación principal, los valores de orientación del gradiente de la muestra se pueden distribuir en los histogramas de gradiente de las regiones secundarias usando diferencia angular e interpolación trilineal. Como tal, la posición y el ángulo de las muestras se pueden tener en cuenta durante la interpolación.

Una ventaja de usar un truncamiento circular y una función de formación de ventanas simétrica circularmente es que puede ahorrar algo de tiempo de cálculo permitiendo evitar comprobar si una muestra está dentro o fuera de la región de truncamiento después de la reorientación.

Se debería señalar que la reorientación no siempre es una necesidad. Por ejemplo, si se puede suponer que si no se puede observar, o muy poca, rotación notable entre imágenes consecutivas del video, la invariancia de rotación puede ser inútil o incluso perjudicial, ya que puede conducir a requisitos de cálculo más altos. La ausencia de rotación notable en imágenes consecutivas es, por ejemplo, el caso estándar en videos endomicroscópicos. De hecho, las sondas de formación de imágenes típicamente tienen una alta resistencia al par. La rotación de la sonda de formación de imágenes con respecto al tejido, por lo tanto, a menudo se puede olvidar.

Una noción importante en los descriptores locales es la determinación de al menos una escala de observación. Esta escala se puede definir automáticamente o se puede fijar gracias al conocimiento específico de la aplicación. En el contexto de la detección de puntos clave, la escala se determina típicamente durante el proceso de detección. En el contexto de enfoques basados en cuadrícula, la fijación de una escala predefinida podría aparecer como una elección más natural. Sin embargo, se podrían hacer otras elecciones.

Como se ha mencionado anteriormente, la elección de una escala predefinida se puede hacer según el conocimiento específico de la aplicación. Por ejemplo, cuando se usa endomicroscopia, pudiera ser ventajoso usar una escala o escalas que esté o estén relacionadas con escalas anatómicamente significativas, tales como unas pocas micras para centrarse solo en unas pocas celdas, unas pocas decenas de micras para centrarse en patrones de arquitectura celular, etc.

Según otra realización de la invención, al menos una escala óptima también se puede detectar o bien en una base de datos de entrenamiento en todo el conjunto de imágenes optimizando alguna forma de energía dentro de la imagen en la escala dada o bien optimizando la prominencia promedio a través de toda la imagen a la escala dada.

Una vez que se da una escala, pudiera ser ventajoso volver a muestrear la región de imagen local a un parche de imagen con un tamaño de píxel fijo dado. Esto se puede hacer con enfoques estándar de espacio de escala. Una transformación de espacio de escala típica de una imagen I(x,y) se puede definir por $L(x,y,s)=G(x,y,s) \circ I(x,y)$ donde s es el factor de escala y $\circ$ es la operación de convolución en x e y, y G es una función gaussiana 2D. Este espacio de escala se usa para suavizar las regiones locales antes de muestrearlas de manera descendente al tamaño fijo deseado.

Pudiera ser ventajoso considerar que las imágenes de entrada ya se suavizan naturalmente por una cierta σ_0 que surge de algunos parámetros tales como la calidad de la óptica, el proceso de reconstrucción de imagen, etc. El valor de la desviación estándar usada para suavizar las imágenes antes de que la reducción de escala pueda representar este suavizado natural, por ejemplo, usando $\gamma(s-\sigma_0)$ en lugar de s directamente.

Cuando se toma un enfoque basado en cuadrícula y se proporciona una escala de observación fija, pudiera ser ventajoso elegir un paso de cuadrícula que sea lo suficientemente pequeño para capturar todas las estructuras posibles que realmente existen en la imagen, pero lo suficientemente grande para reducir los requerimientos de cálculo.

Una elección ventajosa puede ser elegir un paso de cuadrícula que sea proporcional al factor de escala. Para reducir el coste de cálculo, también pudiera ser ventajoso elegir un factor de proporcionalidad entero. De esta forma, los píxeles vueltos a muestrear y las muestras para el descriptor local se localizarán conjuntamente. Por lo tanto, se puede evitar un paso de interpolación de muestras.

Aunque un enfoque de cuadrícula a menudo muestra resultados precisos y eficientes, en algunos escenarios, pudiera ser ventajoso refinar los resultados de coincidencia de la cuadrícula. De hecho, la precisión de una coincidencia se limita al paso de la cuadrícula. Reducir el paso de la cuadrícula es una opción, pero esto es al precio de aumentar el coste de cálculo. En una realización de la invención, se puede usar una forma de vacilación en las posiciones de los puntos de cuadrícula para aleatorizar el error de cuantificación y, por lo tanto, disminuir su promedio.

Como se ilustra en la Figura 6, el ruido intencional se puede añadir a las posiciones 63 de punto de la cuadrícula (62) regular para crear una cuadrícula 64 perturbada. Preferiblemente, la desviación estándar de este ruido sería menos de un cuarto del paso de la cuadrícula 62 original para mantener las posiciones 65 de punto de la cuadrícula 64 perturbada lo suficientemente cerca de la original. Esto es potencialmente importante para asegurar una cobertura suficiente de toda la imagen.

En otra realización, los puntos originales se verían en los puntos iniciales, que podrían generar cada uno varios puntos con diferentes instancias de ruido. La elección de una instancia ruidosa por punto inicial conduciría a una cuadrícula perturbada simple, pero la elección de un número mayor de instancias podría ser beneficiosa.

En otra realización más, el ruido añadido a las ubicaciones de puntos de cuadrícula no se haría al azar, sino que se accionaría por el mapa de prominencia correspondiente a la imagen subyacente. Comenzando desde una cuadrícula de puntos regular original, cada punto de la cuadrícula sería atraído por puntos de imagen destacados cercanos, así como de ser atraído por la ubicación original. La competencia entre las dos atracciones definiría la posición final del punto de cuadrícula perturbado. De manera similar, también podríamos añadir un término de repulsión entre los puntos de cuadrícula. Con este enfoque, los descriptores estarían bien distribuidos sobre la imagen, pero también se centrarían en los puntos destacados dentro de la imagen, haciendo potencialmente la coincidencia más precisa.

Con más detalle, según una configuración de ejemplo, la atracción al punto de cuadrícula original podría ser binaria sin atracción siempre que el punto esté dentro de una región circular delimitada y la atracción infinita cuando el punto esté fuera de la región delimitada. Si no se usa el término de repulsión de puntos de cuadrícula, el punto de cuadrícula terminaría siendo localizado conjuntamente con el punto de imagen más destacado dentro de la región delimitada.

La derivación del mapa de prominencia de imagen se puede hacer usando un criterio de prominencia estándar, tal como, pero no limitado a, criterios basados en derivadas de segundo orden o criterios teóricos de información.

Como se ilustra en la Figura 5, una vez que está disponible una descripción 54 de imagen, los descriptores 59 de esta imagen se pueden hacer coincidir con los descriptores 57 de la imagen 53 anterior en el almacenador temporal. Ahora se puede analizar el conjunto de coincidencias (II) para evaluar si el movimiento fue estable o no entre estas dos imágenes.

Para encontrar buenas coincidencias de descriptores, una elección posible es confiar en los k descriptores más cercanos que se proporcionan por una medición de discrepancia. Se describen varios enfoques algorítmicos para aprovechar los puntos más cercanos.

Para medir la discrepancia entre dos descriptores, la distancia Euclidiana sería la elección más simple, produciendo a menudo suficientes resultados. Sin embargo, se pueden usar otras mediciones de discrepancia dependiendo de las distancias, las pseudo-distancias o más algoritmos ad hoc, incluyendo pero no limitado a χ^2, la distancia de Mahalanobis, Distancia del Movimiento de la Tierra (EMD), y similares. En algunos escenarios, el uso de tal medición de discrepancia podría conducir potencialmente a mejores resultados con propósitos de coincidencia de características.

La distancia Euclidiana se usa ampliamente para comparar cualquier punto de cualquier dimensión. Sin embargo, los descriptores se pueden normalizar y, por ejemplo, podrían representar la distribución local de gradientes dentro de una región de interés. En este escenario, la medición de discrepancia entre los descriptores podría beneficiarse de confiar en distancias relacionadas con la densidad de probabilidad, tales como la EMD.

Incluso en el caso anterior, la distancia Euclidiana o la distancia Euclidiana al cuadrado pueden ser de gran interés por razones de cálculo.

Dada una medida de discrepancia, podemos calcular cada posible discrepancia por pares entre dos conjuntos de descriptores. Esto permite la creación de una matriz de discrepancia D, donde D(i,j) = discrepancia (descriptor de orden i del conjunto 1º, descriptor de orden j del conjunto 2º). Esto plantea dos problemas potenciales. El primero es el de la complejidad de cálculo para crear la matriz D. La segunda es que este proceso puede generar un gran número de valores atípicos. Sería útil mejorar ambos aspectos. Para reducir el coste de cálculo, podemos, por ejemplo, tolerar algún error en el emparejamiento confiando en las herramientas aproximadas del vecino más cercano más que del vecino exacto más cercano. Para reducir el número de valores atípicos, es posible, por ejemplo, validar cada coincidencia antes de añadirla a la lista de coincidencias útiles. Tal paso puede requerir no solo centrarse en la coincidencia más cercana, sino también buscar las k coincidencias más cercanas.

Observando la complejidad de cálculo del enfoque de fuerza bruta, si consideramos buscar las k mejores coincidencias sobre dos conjuntos de N descriptores, cada descriptor que tiene el mismo tamaño n, la complejidad de la fuerza bruta del algoritmo de búsqueda del vecino más cercano k (k-NN) es exactamente $O((C(n)+k).N^2)$, siendo C(n) el coste de la medición de discrepancia. En el caso de la distancia Euclidiana, C(n) es aproximadamente igual a n. El coste de ordenar parcialmente cada fila con el fin de obtener los k mejores resultados es O(kN) en promedio. La complejidad de la búsqueda exacta es, por lo tanto, $O((n+k).N^2)$.

Para reducir la complejidad de cálculo, se pueden usar técnicas aproximadas del vecino más cercano. Esta reducción se puede lograr, por ejemplo, confiando en la partición de datos. Un árbol binario n-d se construye para separar puntos de dimensión n. Esto se hace recursivamente para cada hijo hasta que el cardinal del punto de una hoja alcanza uno. Construir este árbol mientras que se usa una división mediana como agrupación tiene una complejidad lineal de $\theta(nNlog_2(N))$. Se debería observar que cualquier método de agrupación se podría usar para dividir los datos en el árbol binario. Comúnmente, típicamente se usa una simple división mediana, pero también se usan ampliamente medios K jerárquicos u otros algoritmos de agrupación para esta aplicación específica.

Una vez que se construye el árbol n-d, el algoritmo de búsqueda va desde la parte superior del árbol a una hoja final para alcanzar el primer punto más cercano. La complejidad para buscar aproximadamente los k puntos más cercanos de N consultas es de alrededor O(kNlog(N)). La complejidad de la construcción del árbol n-d y la búsqueda aproximada en el árbol n-d es: O((n+k)Nlog(N)).

En el modo básico de operación, podríamos para que cada par de imágenes coincidan, construir el árbol n-d para la primera (respectivamente segunda) imagen y hacer coincidir cada descriptor de la segunda (respectivamente primero) con sus k descriptores más cercanos en este árbol n-d. Ambas órdenes también se pueden realizar concurrentemente si se requiere.

Para ahorrar aún más tiempo de cálculo, puede ser ventajoso construir un árbol n-d solo cada dos imágenes. Esto se puede lograr si podemos elegir cuál de las dos imágenes se usa para crear el árbol n-d. De hecho, podemos comenzar eligiendo la segunda imagen para la creación del árbol n-d, luego, cuando una tercera imagen se ha de hacer coincidir con la segunda, se usaría el árbol n-d para la segunda imagen en la medida que ya está disponible. Cuando la cuarta imagen ha de ser hecha coincidir con la tercera, se construirá un nuevo árbol n-d para la cuarta imagen y así sucesivamente.

Con el propósito de transferir un árbol n-d de un par de imágenes al siguiente, la invención puede hacer uso ventajosamente de la base de datos interna introducida anteriormente.

Dado el enfoque de fuerza bruta o unos más avanzados, cada descriptor en el primer conjunto se puede asociar con el descriptor más cercano en el segundo conjunto. Este emparejamiento no es necesario que sea simétrico. La característica de comprobación de simetría se puede usar ventajosamente para validar una coincidencia y, por lo tanto, para eliminar valores atípicos. Dada la mejor coincidencia, en el segundo conjunto, de un descriptor del primer conjunto, si el descriptor más cercano, en el primer conjunto, al descriptor del segundo conjunto es exactamente el mismo descriptor que el inicial del primer conjunto, entonces, el emparejamiento sería validado. Una implementación de la comprobación de simetría puede beneficiarse de la construcción y el almacenamiento de un árbol n-d por imagen.

Aunque la comprobación de simetría puede permitir eliminar muchos valores atípicos, puede ser beneficiosa en algunos casos para refinar aún más la eliminación de valores atípicos. Eliminar la mayoría de las asociaciones incorrectas permitiría producir un análisis más fácil, más preciso y más robusto de las coincidencias. Los casos típicos que conducen a coincidencias incorrectas incluyen, pero no se limitan a:

- Descriptores fuera de la superposición. Para cualquier transformación espacial no trivial que relacione dos imágenes consecutivas, aunque podría haber una superposición entre las imágenes consecutivas, en la mayoría de los casos habrá regiones espaciales en la primera imagen que no existen en la segunda imagen. Para esos descriptores en las regiones que no se superponen, no existen buenos descriptores en la otra imagen con los que ser asociados.

- Descriptores planos. Las regiones con muy poco contraste o regiones planas en la imagen no tienen ninguna información de gradiente fiable. La distribución del gradiente es homogénea, accionada por el ruido inherente del

sistema de formación de imágenes. Esto puede conducir a coincidencias aleatorias entre las regiones planas. El mismo problema puede aparecer de una forma menos estricta para las regiones que solo muestran contraste a lo largo de una sola dirección. Este es el llamado problema de apertura.

Se debería señalar que la simetrización descrita anteriormente puede ayudar a eliminar muchos valores atípicos en estas dos categorías. Sin embargo, hay casos para los que otros métodos pueden ser más beneficiosos. Algunos dispositivos de formación de imágenes pueden crear de hecho un patrón de ruido estático en la parte superior de sus imágenes debido a imprecisiones de calibración, efectos de viñeta, arañazos en la óptica, etc. En esta configuración de conjunto, las imágenes sin contraste útil todavía tienen un pequeño contraste que surge de cualquiera de los artefactos mencionados anteriormente. Por lo tanto, las regiones planas pueden no ser completamente planas. La información de gradiente débil de ese ruido estático se puede tener en cuenta entonces al tiempo de asociar los descriptores. Estas coincidencias erróneas potencialmente no se eliminarán mediante simetrización y desviarán la coincidencia hacia la identidad.

Para determinar si una coincidencia es fiable, se ha propuesto un análisis de la relación entre la discrepancia del descriptor actual con su descriptor más cercano en el otro conjunto y la discrepancia con su segundo descriptor más cercano. Si bien esto funciona bien en la práctica cuando se usa detección de puntos clave, esto deja de funcionar correctamente en el caso de la cuadrícula donde se pueden describir regiones superpuestas y, por lo tanto, pueden tener descriptores similares. La detección de puntos clave puede conducir a posiciones de descriptores, que aseguran que todas las regiones locales describan regiones casi no superpuestas dentro de la imagen de entrada. Cuando se usa un enfoque de descripción de imágenes basado en cuadrícula, las regiones cubiertas por descriptores pueden tener una superposición no despreciable. Por ejemplo, hay casos donde alrededor del 80% de la superposición parece ser beneficioso. Significaría entonces que los descriptores de dos regiones locales espacialmente vecinas podrían ser similares. Por lo tanto, el descriptor más cercano y el segundo más cercano a su vez podrían tener discrepancias muy similares con el descriptor actual.

Según una realización de la invención, se puede usar el análisis de la relación entre la discrepancia del descriptor actual con el descriptor más cercano en el otro conjunto y la discrepancia con el descriptor más cercano de orden k. La elección de k tiene que ser hecha teniendo en cuenta la estructura de la cuadrícula. Por ejemplo, elegir k=5 (respectivamente k=9) asegura que los puntos de la cuadrícula 4 conectados (respectivamente 8 conectados) directos con la mejor coincidencia no se tengan en cuenta. Un umbral en esta relación puede permitir eliminar muchos valores atípicos al tiempo que se mantiene en la mayoría de los valores típicos.

Tal análisis de la relación debería proporcionar resultados utilizables porque la comparación de una coincidencia correcta con la incorrecta más cercana debería conducir a una diferencia mucho más alta que la comparación de una coincidencia incorrecta y la otra coincidencia incorrecta más cercana. Un acercamiento estándar ha usado la primera coincidencia más cercana como un punto de comparación, mientras que describimos el uso de la de orden k para evitar tener en cuenta casi todas las coincidencias correctas de regiones que tienen una alta superposición con la coincidencia correcta. Como se ha mencionado anteriormente, es beneficioso adaptar el parámetro k dependiendo de la densidad de la cuadrícula de descriptores usada. Cuanto más densa es la cuadrícula, más necesitamos buscar el segundo descriptor usado en la relación.

Según otra realización de la invención, también es posible eliminar todas las coincidencias con una discrepancia por encima de un umbral dado. El umbral puede ser uno globalmente predefinido, se puede calcular globalmente para un par dado de imágenes en base a las estadísticas observadas de las discrepancias, o se puede calcular localmente en base a las discrepancias observadas en una vecindad local de un punto. También se pueden imaginar opciones más complejas que tengan en cuenta el contenido real de los descriptores de la región de la imagen local.

Dado un par de imágenes consecutivas y un conjunto de coincidencias filtradas, ahora podemos proceder con su análisis para evaluar la estabilidad cinemática de una imagen a otra.

Según una realización, el análisis de las coincidencias se realizaría como tal: las coincidencias votarían dentro de un conjunto de parámetros de transformación espacial discretizados, creando, por lo tanto, un mapa de votos. Los parámetros que tienen un número suficiente de votos serían considerados como votos consistentes. El porcentaje de votos consistentes frente a inconsistentes se podría usar entonces como una evaluación de confianza para la estabilidad cinemática.

Dado un par de imágenes consecutivas y un conjunto de coincidencias filtradas, también podemos querer estimar una transformación espacial que permita registrar o alinear las imágenes. Para imágenes médicas, tal registro a menudo es una tarea potencialmente desafiante debido a, pero no limitado a, algunas de las siguientes razones.

Cuando se forman imágenes de la misma región de tejido en diferentes puntos de tiempo, la señal de imagen observada puede variar debido a reflexión especular, fotoblanqueado, cambios en la vascularización u oxigenación, cambios en la cantidad de luz de excitación, etc.

Una oclusión podría ocurrir debido a la presencia de otros instrumentos, de sangre y otros líquidos biológicos, humo, heces, etc.

Las estructuras de tejido también se pueden deformar debido a la respiración, los latidos del corazón, el movimiento del paciente o el contacto entre tejido e instrumentos, incluyendo la sonda de formación de imágenes. Por lo tanto, las deformaciones locales pueden necesitar ser tenidas en cuenta mientras que se registran dos imágenes consecutivas.

El dispositivo de formación de imágenes también puede generar sus propios artefactos de movimiento que, en algunos casos, pueden ser demasiado complejos para ser modelados correctamente para la tarea de un registro de imágenes por pares. Por ejemplo, en el caso de un dispositivo de escaneado de formación de imágenes, el escaneado del campo de formación de imágenes para una imagen dada se puede realizar gracias a espejos. Esto implica que cada píxel puede ser adquirido en un momento diferente. Cuando la sonda de formación de imágenes se mueve con respecto al tejido, puede causar fuertes distorsiones que varían dentro del campo de visión. En algunos casos, si el movimiento de la sonda de formación de imágenes con respecto al tejido es constante mientras que adquiere una imagen, las distorsiones se pueden modelar y compensar. Sin embargo, en la mayoría de los casos, el movimiento de la sonda es más complejo y no se puede modelar fácilmente, especialmente si el movimiento evoluciona rápidamente.

En algunos escenarios, el dispositivo de formación de imágenes se basa en la reconstrucción de la imagen y la información de calibración para producir sus imágenes. La calibración puede tener imprecisiones e incluso puede cambiar con el tiempo. Esto puede conducir o bien a un patrón de ruido estático que puede sesgar el registro de imágenes o bien a un cambio en la apariencia visual que puede complicar la tarea de registro de imágenes.

En la mayoría de los casos, el dispositivo de formación de imágenes no tiene información de seguimiento que sería útil para guiar el proceso de registro de imágenes. Además, incluso cuando la información de seguimiento está disponible, la precisión podría ser bastante grande en comparación con el campo de visión. Esto sería especialmente cierto en el campo de la Endomicroscopia, pero también se mantendría para la mayoría de los dispositivos de formación de imágenes debido al movimiento del paciente.

En algunos casos, incluso aunque las razones anteriores todavía existan, su impacto en las imágenes podría ser suficientemente pequeño que podamos estimar directamente una transformación espacial entre las imágenes y analizar el resultado para decidir la estabilidad cinemática. En otros casos donde las mismas razones tienen un impacto más alto en las imágenes, tal enfoque solo puede funcionar para un pequeño porcentaje de pares de imágenes. Por lo tanto, esto puede conducir a un sesgo hacia inestabilidad en la estimación de la estabilidad cinemática. De hecho, muchos pares de imágenes potencialmente podrían no estar correctamente registrados, aunque el movimiento general entre las imágenes se podría considerar como suave.

Según una realización de la invención, nos centramos en los casos para los cuales encontrar un modelo de transformación espacial es suficiente para estimar un análisis cinemático. La transformación espacial podría ser cualquiera de los modelos clásicos o menos clásicos, incluyendo, pero no limitado a, traslados, transformaciones de cuerpos rígidos, transformaciones afines, transformaciones proyectivas, traslados con cortes para representar distorsiones de movimiento, etc. En este escenario, las coincidencias pueden servir como datos de entrada para ajustarse al modelo de transformación. Esquemas basados en optimización tales como descenso del gradiente, recocido simulado y similares o esquemas de muestreo aleatorios tales como RANSAC, MSAC y similares, ajuste de mínimos cuadrados, cuadrados menos recortados, ajuste de mínimos cuadrados ponderados, ajuste L_1 y similares pueden ser todos usados. Enfoques de ajuste jerárquico, tales como los que refinan progresivamente el modelo de transformación espacial, también pueden ayudar a proporcionar resultados más robustos.

La estabilidad cinemática se puede evaluar entonces observando el número de valores típicos para el modelo de transformación espacial final y comparándolo con el número total de coincidencias o el número total de coincidencias mantenidas.

La estabilidad cinemática también se puede evaluar usando la transformación espacial final y calculando una puntuación de similitud en la región de superposición entre las imágenes después de deformar la de destino sobre la otra. La puntuación de similitud puede ser una de las puntuaciones de similitud estándar o menos estándar usadas en formación de imágenes médicas, incluyendo, pero no limitado a, suma de diferencias al cuadrado, correlación normalizada, información mutua, campo de gradiente normalizado y similares.

En este caso, la estabilidad cinemática se evalúa mediante una puntuación de similitud de registro. Se debería señalar que un enfoque directo para el registro que optimiza la puntuación de similitud también es posible y podría, en algunos casos, conducir a mejores resultados. En otros casos, incluso si la estabilidad cinemática se evalúa en términos de puntuación de similitud, ir a través de la ruta de coincidencia de características puede conducir a resultados más robustos que son menos propensos a ser atrapados en mínimos locales. Dependiendo de la implementación exacta, los costes de cálculo también podrían variar en gran medida dependiendo de la ruta elegida.

Aunque ajustar un modelo de transformación a los datos coincidentes en algunos casos puede ser realmente eficiente, pudiera haber casos donde definir el modelo sea demasiado complejo para ser utilizable en la práctica. Según otra realización de la invención, se puede usar un enfoque más local para analizar las coincidencias entre dos imágenes consecutivas para la estabilidad cinemática. De manera ventajosa, la invención permite no centrarse en el

modelo exacto de transformación espacial, sino evaluar la probabilidad de tener una transformación espacial bastante consistente espacialmente entre imágenes. Con este propósito, se propone una puntuación de similitud que se basa en los traslados locales proporcionados por las coincidencias de descriptores.

Según una realización de la invención, se puede crear una puntuación de similitud entre imágenes consecutivas a través de un mapa de votos. El mapa de votos es un histograma 2D que resume la contribución de cada traslado local encontrado por los descriptores coincidentes. La contribución puede ser ponderada por una función de la discrepancia entre los dos descriptores emparejados, por la calidad de la asociación o simplemente todos pueden tener una ponderación unitaria.

El mapa de votos usa contenedores de votación discretizados. De manera ventajosa, en el caso de una cuadrícula regular para la descripción de imagen, la resolución del mapa de votos se puede elegir que sea igual a la de la cuadrícula de descripción. En este caso, el tamaño del mapa de votos típicamente será dos veces el de la cuadrícula para permitir todos los traslados posibles desde un punto de la cuadrícula en la primera imagen a otro punto de la cuadrícula en la otra imagen.

En el caso de una cuadrícula perturbada o en el caso de detección de puntos clave, elegir la resolución del mapa de votos se puede hacer según la precisión requerida.

Se debería señalar que la superposición entre dos imágenes depende de la amplitud del traslado. Debido a eso, no todos los traslados pueden recibir el mismo número máximo de votos. En realidad, en una configuración simple, solo la transformación identidad puede recibir todos los votos. Si consideramos un traslado de la mitad del campo de visión en una dimensión y si usamos imágenes rectangulares, la superposición corresponderá a la mitad de una imagen, lo que significa que solo la mitad de las coincidencias pueden votar por el traslado correcto.

Para representar este sesgo potencial, el mapa de votos se puede ponderar además según el número máximo de votantes potenciales por contenedor de votación. De manera ventajosa, el número máximo de votantes potenciales para un traslado dado en el mapa de votos se puede calcular gracias a una convolución de dos imágenes de mascara que representan la organización espacial de las cuadrículas usadas para la descripción de imágenes.

En algunos dispositivos de formación de imágenes, el campo de visión de las imágenes no es cuadrado sino que puede ser típicamente de forma circular o de cualquier otra forma. Para calcular la normalización del mapa de votos, se puede crear una imagen de mascara donde todas las posiciones válidas del descriptor se llenan con unos y las inválidas con ceros.

Después de la convolución de las máscaras, obtenemos un mapa de contribución que contiene la relación de contribuyentes potenciales sobre el número máximo de contribuyentes para cada posible traslado. Los valores están entre 0 y 1. Según una realización de la invención, podemos querer considerar solo los traslados que se pueden votar por un número suficiente de coincidencias de descriptores.

El mapa de votos se puede normalizar como tal. Cada entrada en el mapa de votos se divide o bien por el valor del mapa de contribución si el valor del mapa de contribución está por encima de un umbral dado o bien se asigna a 0 de lo contrario (es decir, si el valor del mapa de contribución está por debajo del umbral).

Una vez que se calcula el mapa de votos normalizado, en el caso donde la transformación espacial pueda estar bien representada por un traslado, típicamente observaremos un pico principal en el mapa de votos alrededor del traslado esperado.

En el caso de transformaciones espaciales más complejas, incluyendo las no lineales, muchos picos aparecerán típicamente en el mapa de votos, normalizados o no. Según una realización de la invención, todos los picos se tienen en cuenta para evaluar la estabilidad cinemática. Para esto, se puede hacer un umbral simple en los valores del mapa de votos para seleccionar todos los votos que son suficientemente consistentes. Todos los valores en el mapa de votos que corresponden con votos consistentes seleccionados se pueden resumir entonces para evaluar una coherencia general relacionada con la estabilidad cinemática.

El enfoque anterior puede asegurar que solo se tengan en cuenta los traslados que se comparten en regiones de imagen local algo extendidas. Aunque esto puede cubrir la mayoría de las transformaciones importantes que necesitamos, en algunos casos, se podría requerir un enfoque más refinado. Según otra realización de la invención, se seleccionará una coincidencia según la siguiente regla. Dada una vecindad de coincidencias, se realiza una estimación robusta de un modelo de transformación simple. La coincidencia central para este vecindario se puede seleccionar dependiendo de su distancia a la transformación del modelo. De esta manera, solo se mantienen las coincidencias localmente consistentes para evaluar la coherencia general.

De manera ventajosa, dependiendo del modelo de transformación espacial local, tal selección se puede realizar dependiendo de un simple suavizado, filtrado o regularización del campo de desplazamiento producido por las coincidencias.

Una vez que se calcula la coherencia de la transformación espacial entre imágenes consecutivas, se puede usar un umbral simple de la coherencia como indicador de estabilidad cinemática.

Para reducir aún más la complejidad de cálculo, se puede emplear un enfoque de escala múltiple. Como primer paso, se puede usar una cuadrícula gruesa de descriptores. Mientras que la granularidad inferior significa que las estimaciones derivadas de esta cuadrícula son menos precisas, la disminución en el número de descriptores hace que el algoritmo se ejecute mucho más rápido. Dado el resultado encontrado usando la cuadrícula gruesa, ya podemos detectar pares de imágenes fáciles de hacer coincidir y pares de imágenes fáciles que no se pueden hacer coincidir. Para los pares de imágenes que no son fáciles, podemos ejecutar el algoritmo usando la cuadrícula fina. De manera ventajosa, podemos decidir usar reglas bastante conservadoras para distinguir pares de imágenes fáciles.

En lugar de usar una cuadrícula gruesa y luego una cuadrícula fina para comparación, la invención permite lograr una aceleración similar si la base de datos interna se usa para guardar los árboles n-d construidos a partir de las cuadrículas finas. Si se hace esto, una cuadrícula gruesa en una imagen se puede hacer coincidir eficientemente con la cuadrícula fina de la otra imagen. Esto es ventajoso porque el uso de dos cuadrículas gruesas de descriptores significa que se aumenta el error de discretización. Por lo tanto, podría haber una oportunidad de que las cuadrículas se desemparejen de manera demasiado grave y que el algoritmo pudiera no indicar correctamente un par de imágenes consecutivas estables. Sin embargo, usando una cuadrícula fina de descriptores, el error de discretización se mantiene similar al caso de la cuadrícula fina completa. Es solo el ruido en el mapa de votos el que será mayor cuando se use una cuadrícula gruesa.

Si se usan varias escalas de descripción, el mismo procedimiento también se puede aplicar de una forma de múltiples escalas estándar, pasando de la escala más gruesa a la más fina y deteniéndose siempre que una escala permita hacer una estimación segura de la estabilidad cinemática.

Según otra realización, se pueden usar concurrentemente varias escalas para crear un mapa de votos de múltiples escalas en el que se puede extender el análisis anterior trabajando en el análisis de múltiples valores.

Más allá de la estabilidad de imagen consecutiva, la noción de estabilidad cinemática también puede cubrir preferiblemente la idea de que secuencias secundarias estables no se deberían limitar a solo una o unas pocas imágenes aisladas y que secuencias secundarias estables separadas por una o pocas imágenes inestables se deberían unir.

Con este propósito, como se ilustra en la Figura 7A y la Figura 7B, según una realización de la invención, se pueden usar operaciones matemáticas morfológicas en el dominio temporal. Si los análisis de imágenes consecutivas condujeron a una línea de tiempo con una información (71, 72) binaria de estabilidad cinemática, se puede usar una operación de cierre morfológico (ilustrada en la Figura 7A) para llenar pequeños huecos (73) entre secuencias secundarias estables mientras que una operación de apertura morfológica (ilustrada en la Figura 7B) se puede usar para eliminar secuencias secundarias estables (75) pero demasiado cortas.

Como se ilustra en la Figura 7A y la Figura 7B, este enfoque puede permitir evitar algunos de los falsos negativos y falsos positivos en nuestra segmentación temporal inicial.

En lugar de binarizar el resultado del análisis de estabilidad cinemática de pares de imágenes antes de las operaciones de morfología matemática, la invención también permite usar herramientas de procesamiento de señales directamente en la salida continua del análisis cinemático. Con este propósito se pueden usar suavizamiento gaussiano simple, morfología matemática en escala de grises, métodos iterativos, cortes de gráficos y similares. Para el enfoque de corte de gráfico, una realización potencial usaría el análisis de estabilidad cinemática continua (o una función de transferencia de ella) entre dos imágenes consecutivas como factor de regularización (término de suavizado), y podría usar, como término de datos, un factor constante, un resultado de procesamiento previo o cualquier otra señal accionada por datos, tal como la desviación estándar de la imagen o las coincidencias y similares.

Uso de una base de datos interna avanzada

Según una realización, la invención puede procesar la imagen actualmente adquirida sobre la marcha. Una base de datos interna que inicialmente está vacía se puede crear y enriquecer progresivamente con el resultado del procesamiento sobre la marcha de las imágenes adquiridas previamente. Esta base de datos interna se puede actualizar tras cada actualización del almacenador temporal de imágenes.

La base de datos interna puede almacenar resultados intermedios que se pueden calcular a nivel de región, a nivel de imagen o a nivel de secuencia secundaria de video. Dichos resultados intermedios pueden incluir, pero no se limitan a:

- características visuales globales y locales;
- distancias de similitud entre cuadros globales y locales;

- campos de desplazamiento global y local;
- palabras visuales construidas a partir de agrupación de características visuales;
- firmas visuales;
- distancias de similitud entre secuencias secundarias de video;
- distancias de similitud de secuencias secundarias de video del video de interés a videos de una base de datos externa;
- una información de conocimiento a posteriori extraída de una base de datos externa que contiene imágenes o videos ya adquiridos y anotados.

- La base de datos interna puede incluir, por ejemplo, una estructura basada en gráficos, tal como un árbol k-d o un bosque aleatorio, que soporta la generación y el tratamiento de los resultados intermedios almacenados. Dicho tratamiento incluye, pero no se limita a:

- agrupar características visuales;
- calcular una firma visual asociada con una secuencia secundaria de video;
- calcular distancias entre firmas visuales.

Esquemas basados en clasificación y basados en regresión

Según una realización, en el caso de salidas discretas, el primer algoritmo de la invención es capaz de usar un clasificador para estimar la etiqueta correspondiente a una imagen. El clasificador podría ser uno basado en reglas simples o puede depender del aprendizaje por máquina a ser entrenado desde una base de datos de capacitación externa donde un conjunto de imágenes se asocia con datos reales del terreno, tales como etiquetas o anotaciones.

Según otra realización, en el caso de salidas continuas, el primer algoritmo de la invención es capaz de usar un algoritmo de regresión para estimar la etiqueta o la salida continua correspondiente a una imagen. El algoritmo de regresión puede ser una regresión de mínimos cuadrados simple o puede depender de que el aprendizaje por máquina se entrene desde una base de datos de capacitación externa donde un conjunto de imágenes se asocie con datos continuos reales del terreno.

Las herramientas de aprendizaje por máquina usadas potencialmente por el primer algoritmo para propósitos de clasificación o regresión se pueden basar, por ejemplo, en Máquinas de Vectores de Soporte, Refuerzo Adaptativo, Recuperación de Imágenes Basada en Contenido seguidos por votación de Vecino más Cercano k, Redes Neuronales Artificiales, Bosques Aleatorios, y similares.

Evaluación y agrupación de similitud visual

Según una realización, la invención es capaz de operar de una manera completamente sin supervisión basándose solo en el contenido de la imagen del video de interés.

El primer algoritmo puede ser un algoritmo de agrupación completamente no supervisada que toma como entrada todas las imágenes almacenadas en el almacenador temporal y proporciona como salida una agrupación asociada con cada imagen. Según una realización, la agrupación asociada con una imagen se puede asignar a un color que se puede superponer sobre la línea de tiempo en la posición correspondiente a la imagen en el almacenador temporal de video.

El algoritmo de agrupación no supervisada se puede basar en agrupación de K Medios, agrupación jerárquica, agrupación de Desplazamiento de Medios, Cortes de Gráficos, la agrupación basada en Bosque Aleatorio, Helechos Aleatorios o cualquier otro algoritmo de agrupación estándar. El algoritmo de agrupación puede usar resultados intermedios almacenados en la base de datos interna.

Según una realización, se construye una firma visual para cada imagen almacenada en el almacenador temporal usando cualquier técnica adecuada, tal como la bolsa de palabras visuales, características de Haralick o redes invariantes de convolución de dispersión y basándose en la base de datos interna como base de datos de capacitación. Luego, la agrupación no supervisada de las imágenes se puede realizar en base a sus firmas visuales.

Acoplamiento de caracterización de la capacidad de interpretación y un segundo algoritmo

Si el primer algoritmo ha provisto al menos una salida discreta para cada imagen, se puede aplicar un segundo algoritmo a las secuencias secundarias de video hechas de imágenes consecutivas de igual salida, y proveer al menos una salida a ser mostrada. Como se ha mencionado anteriormente, tal salida discreta se puede denominar segmentación temporal del video de interés. El segundo algoritmo puede usar al menos una salida del primer

algoritmo, resultados intermedios almacenados en la base de datos interna, datos almacenados en la base de datos externa y similares.

Según una realización, el primer algoritmo proporciona un medio para detección, en el video de interés, de las secuencias secundarias de video que son consultas óptimas para el segundo algoritmo.

El segundo algoritmo incluye, pero no se limita a:

- mosaico de imagen o video para crear una imagen de campo de visión más grande a partir de al menos una secuencia secundaria de video;

- agrupación no supervisada de secuencias secundarias de video, por ejemplo, para agrupar el video de interés en escenas visuales;

- caracterización no supervisada de las secuencias secundarias de video, por ejemplo, para estimar la atipicidad visual de cada secuencia secundaria de video;

- clasificación supervisada de las secuencias secundarias de video, por ejemplo, para asociar una clase diagnóstica o patológica predicha y un nivel de confianza de predicción con una secuencia secundaria de video o con el video de interés completo;

- regresión supervisada de las secuencias secundarias de video, por ejemplo, para estimar una probabilidad de todo el video de interés de cada secuencia secundaria de video de pertenencia a una clase patológica dada;

- caracterización supervisada de las secuencias secundarias de video, por ejemplo, para estimar la ambigüedad visual de cada secuencia secundaria de video o de todo el video de interés con respecto a un conjunto de clases diagnósticas o patológicas;

- video basado en contenido o recuperación de imágenes, con al menos una secuencia secundaria de video como consulta, por ejemplo, para extraer de una base de datos externa videos ya anotados que son visualmente similares a la consulta.

Según una realización, cuando el segundo algoritmo es un algoritmo de recuperación basado en contenido, la invención permite a los usuarios, típicamente médicos, crear eficientemente, a partir de los resultados del primer algoritmo, consultas reproducibles para el segundo algoritmo de una forma semiautomatizada. Esto puede, en algunos escenarios, permitir aumentar el rendimiento de la recuperación cuando se compara con el uso de videos sin cortar como consultas o cuando se compara con la construcción de consultas totalmente automatizada. Tal enfoque semiautomatizado también puede permitirnos plantear el desempeño de consultas construidas cuidadosamente por un experto humano.

Para lograr esto, nuestro enfoque de construcción de consultas se puede descomponer en dos pasos. En un primer paso, se puede realizar una segmentación temporal automatizada del video original en un conjunto de secuencias secundarias de interés gracias a cualquiera de los métodos descritos anteriormente, tales como estabilidad cinemática o evaluación de calidad de imagen. Un segundo paso consiste en una selección de usuario rápida de un subconjunto de las secuencias secundarias segmentadas. Se le puede pedir simplemente al médico que guarde o descarte las secuencias secundarias proporcionadas por el primer paso. Aunque cada una de las posibles secuencias secundarias puede contener posiblemente imágenes de diferentes tipos de tejido, el paso de segmentación típicamente hará cada secuencia secundaria mucho más consistente por sí misma que el video original sin cortar. La interacción simplificada del usuario permite, por lo tanto, una construcción de consulta rápida y reproducible y le permite al médico construir una consulta con suficiente similitud visual dentro y entre las secuencias secundarias seleccionadas.

En una variante, se le pide al usuario que revise brevemente cada secuencia secundaria segmentada y haga clic en las que sean de interés para él. Debido a que todo esto puede ocurrir durante el procedimiento, la segmentación temporal puede ser de manera ventajosa compatible con tiempo real.

Dado el subconjunto de secuencias secundarias elegido por el usuario, la invención puede usar este subconjunto para crear una firma visual para el algoritmo de recuperación basado en contenido para consultar la base de datos externa. El caso más visualmente similar se puede presentar entonces al médico junto con cualquier anotación potencial que se pueda unir a ellos.

En una variante, el método de bolsa de palabras visuales, características de Haralick o cualquier otro método compatible se puede usar para calcular una firma por imagen para cada imagen en una secuencia secundaria de video seleccionada. Promediando estas firmas, cada secuencia secundaria y cada video se pueden asociar con una firma visual que se puede usar con propósitos de recuperación.

En otra variante, en lugar de calcular una firma por video, cada secuencia secundaria se puede asociar a una firma visual que luego se puede usar con propósitos de recuperación. Los casos recuperados para todas las secuencias secundarias entonces se pueden agrupar y reutilizar según su similitud visual con su consulta de secuencia secundaria inicial correspondiente.

Como debería quedar claro a partir de la descripción anterior, por razones de cálculo, cuando tanto el primer algoritmo como el segundo se basan en los mismos cálculos intermedios, tal cálculo se puede realizar solo una vez y compartir entre los dos algoritmos. Este es, por ejemplo, el caso cuando se basa en un conjunto común de descriptores de características, tal como una cuadrícula densa regular de SIFT, SURF y similares, tanto para la segmentación temporal basada en estabilidad cinemática como para recuperación basada en contenido basada en bolsa de palabras.

Mientras que la descripción escrita precedente de la invención permite a un experto hacer y usar lo que se considera actualmente que son los mejores modos de la misma, aquellos de los expertos entenderán y apreciarán la existencia de variaciones, combinaciones y equivalentes de las realizaciones, métodos y ejemplos específicos en la presente memoria. Por lo tanto, la invención no se debería limitar por las realizaciones, métodos y ejemplos descritos anteriormente, sino por todas las realizaciones y métodos dentro del alcance de la invención como se define por las reivindicaciones adjuntas.

REIVINDICACIONES

1. Un método para apoyar una decisión clínica caracterizando la capacidad de interpretación de imágenes adquiridas en secuencia a través de un dispositivo médico de video, en donde el método comprende:

- almacenar imágenes secuenciales en un almacenador temporal;

- para cada imagen en el almacenador temporal, determinar automáticamente al menos una salida de un primer algoritmo, dicha al menos una salida que se basa en al menos un criterio cuantitativo de imagen,

en donde una primera salida del primer algoritmo es un valor de salida entre un conjunto de valores discretos,

en donde dicho primer algoritmo calcula la medición de discrepancia entre descriptores de imágenes locales para asociar al menos una región de una imagen a una región de otra imagen,

en donde el primer algoritmo es un algoritmo de agrupación no supervisada que proporciona como la primera salida una agrupación asociada con cada imagen;

- realizar una segmentación temporal de las imágenes en un conjunto de secuencias secundarias de video hechas de imágenes consecutivas de igual primera salida;

- mostrar una línea de tiempo que indique posiciones de imágenes en el almacenador temporal y unir una salida del primer algoritmo a la línea de tiempo; y

- procesar al menos una de las secuencias secundarias de video usando un segundo algoritmo para proporcionar al menos una segunda salida, en donde el segundo algoritmo es un algoritmo de un conjunto de algoritmos que consiste en:

un algoritmo de clasificación supervisado configurado para asociar un clase diagnóstica o patológica predicha y un nivel de confianza de predicción con dicha al menos una de las secuencias secundarias de video;

un algoritmo de regresión supervisada configurado para estimar una probabilidad de que dicha al menos una de las secuencias secundarias de video pertenezca a una clase patológica determinada;

un algoritmo de caracterización supervisada configurado para estimar la ambigüedad visual de dicha al menos una de las secuencias secundarias de video con respecto a un conjunto de clases diagnósticas o patológicas; y

un algoritmo de caracterización no supervisado configurado para estimar la atipicidad visual de dicha al menos una de las secuencias secundarias de video.

2. El método según la reivindicación 1, en donde el algoritmo de agrupación no supervisada se realiza sobre la base de firmas visuales calculadas para cada imagen en el almacenador temporal.

3. El método según la reivindicación 1, en donde los descriptores de imágenes locales son invariables a cambios de intensidad y a distorsiones espaciales y/o rotaciones de imágenes.

4. El método según la reivindicación 1, que comprende además el paso de

- mostrar dicha segunda salida.

5. El método según cualquiera de las reivindicaciones precedentes, en donde

- al menos uno del primer o del segundo algoritmo usa una base de datos externa.

6. El método según cualquiera de las reivindicaciones precedentes, en donde

- al menos uno del primer o del segundo algoritmo se basa en aprendizaje por máquina.

7. El método según cualquiera de las reivindicaciones precedentes, en donde

- el criterio cuantitativo es uno entre: estabilidad cinemática, similitud entre imágenes, probabilidad de pertenencia a una categoría, tipicidad de imagen o video, atipicidad de imagen o video, calidad de imagen, presencia de artefactos.

8. Un sistema para apoyar una decisión clínica caracterizando imágenes adquiridas en secuencia a través de un dispositivo médico de video, en donde el sistema comprende medios para implementar los pasos de un método según cualquiera de las reivindicaciones 1 a 7.

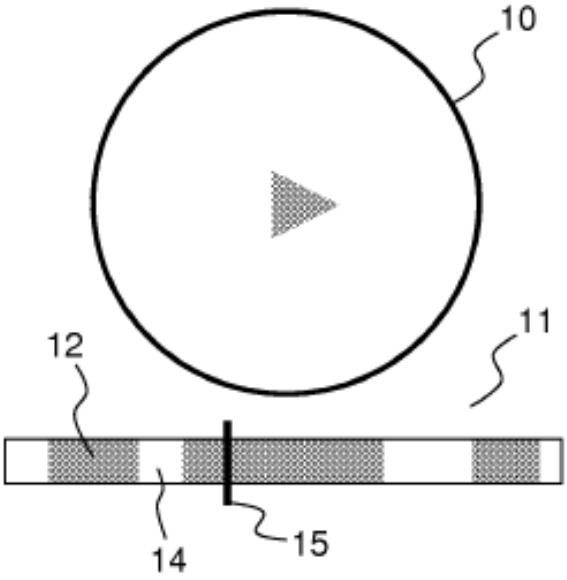

FIG.1

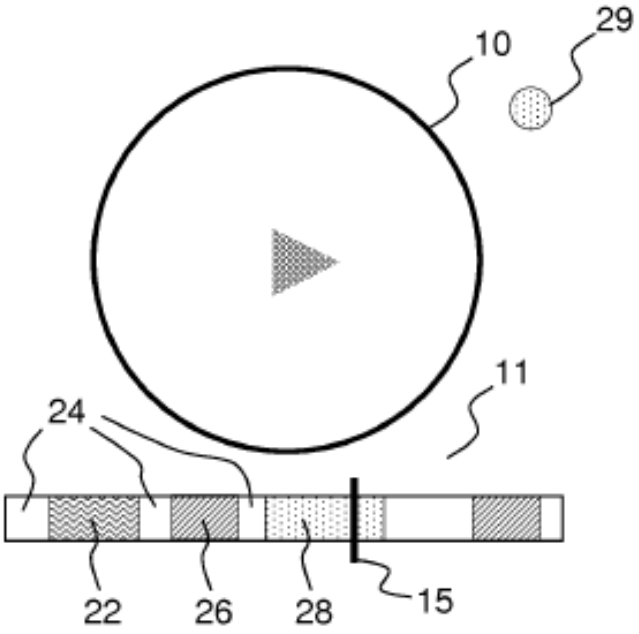

FIG.2

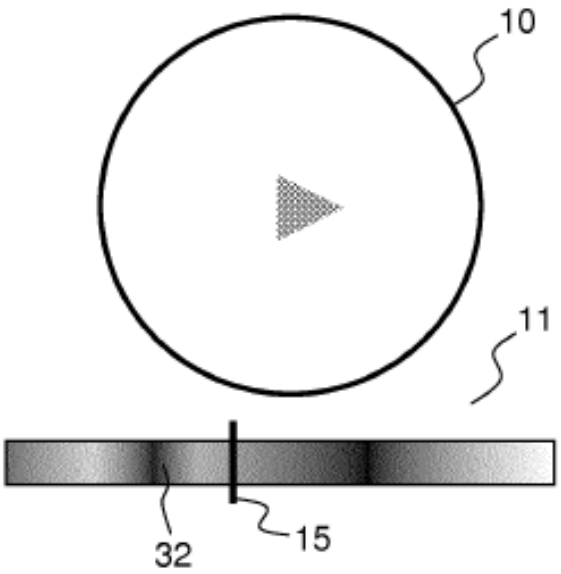

FIG.3

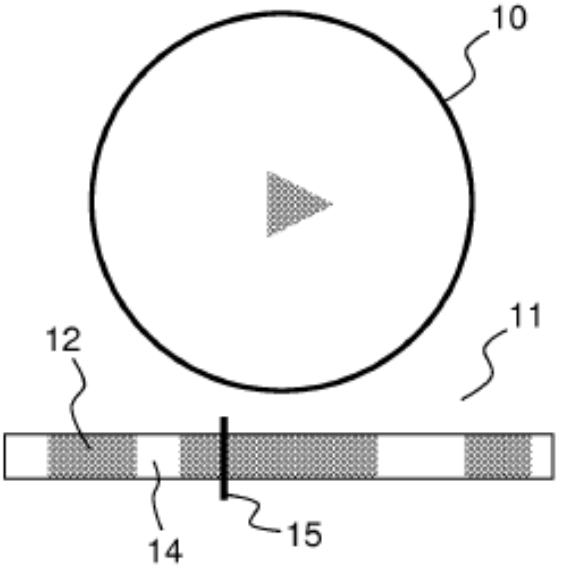

FIG.4A

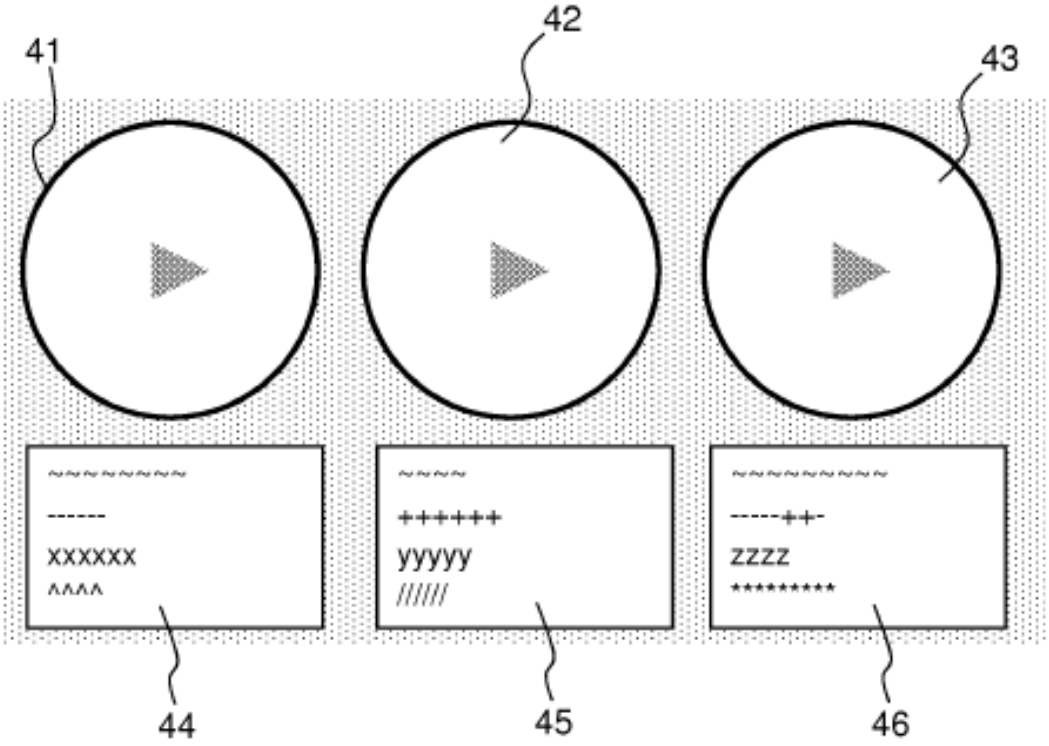

FIG.4B

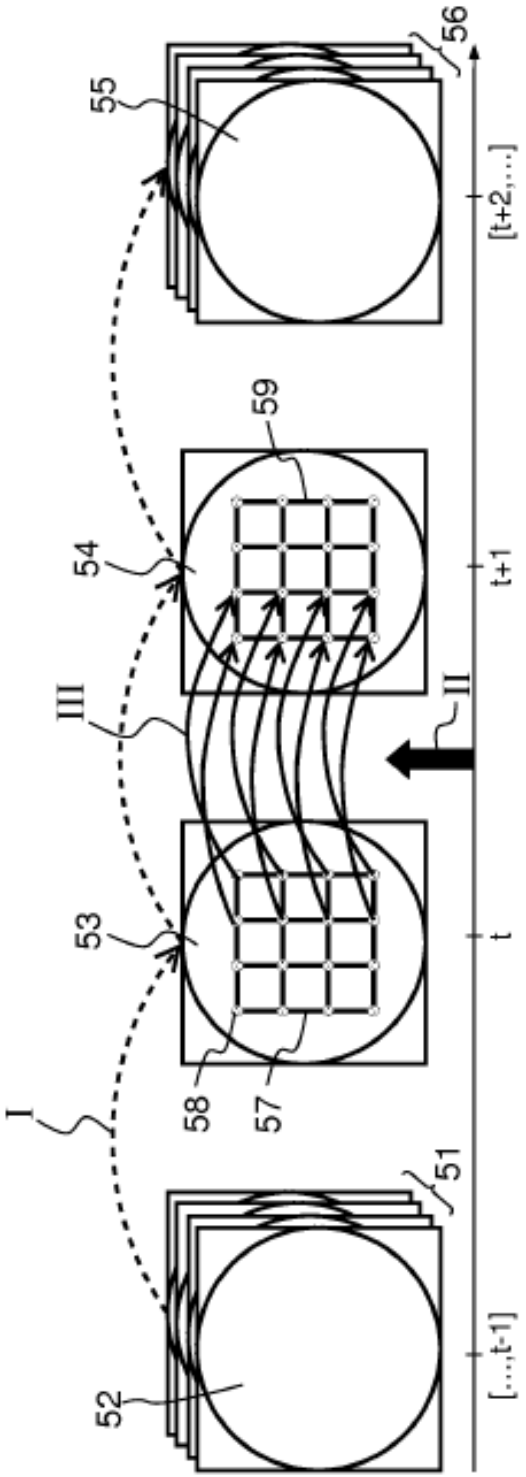

FIG.5

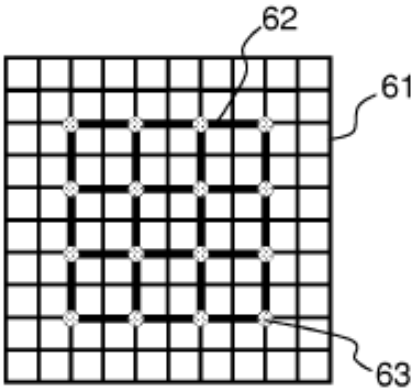

FIG.6A

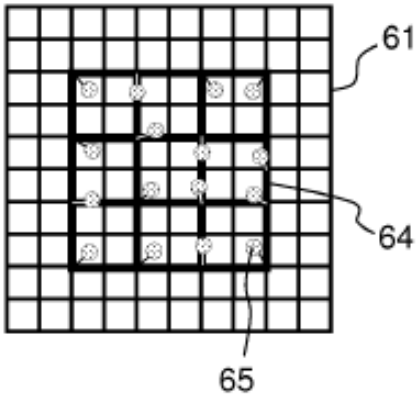

FIG.6B

FIG.7A

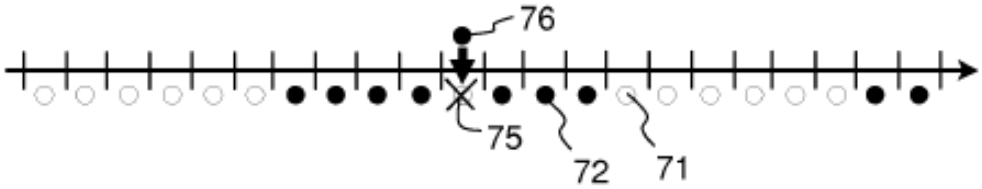

FIG.7B