(19) OFICINA ESPAÑOLA DE PATENTES Y MARCAS

ESPAÑA

(11) Número de publicación: **2 747 387**

(21) Número de solicitud: 201990066

(51) Int. CI.:

G06T 7/579 (2007.01)

(12) SOLICITUD DE PATENTE A2

(22) Fecha de presentación:

06.02.2017

(43) Fecha de publicación de la solicitud:

10.03.2020

(71) Solicitantes:

PHOTONIC SENSORS & ALGORITHMS S.L. (100.0%)
Calle Guardia Civil 23-2-38
46020 VALENCIA ES

(72) Inventor/es:

BLASCO CLARET, Jorge Vicente;
MONTOLIU ÁLVARO, Carles;
VIRGILIO PERINO, Iván y
MARTÍNEZ USÓ, Adolfo

(74) Agente/Representante:

CARVAJAL Y URQUIJO, Isabel

(54) Título: **DISPOSITIVO Y MÉTODO PARA OBTENER INFORMACIÓN DE PROFUNDIDAD A PARTIR DE UNA ESCENA.**

(57) Resumen:

Se da a conocer un método para obtener información de profundidad a partir de una escena, en el que el método comprende las etapas de: a) adquirir una pluralidad de imágenes de la escena por medio de al menos una cámara durante un tiempo de disparo en el que la pluralidad de imágenes ofrece al menos dos vistas diferentes de la escena; b) para cada una de las imágenes de la etapa a), simultáneamente adquirir datos sobre la posición de las imágenes con respecto a un sistema de referencia de seis ejes; c) seleccionar de las imágenes de la etapa b) al menos dos imágenes; d) rectificar las imágenes seleccionadas en la etapa c) generando de ese modo un conjunto de imágenes rectificadas; y e) generar un mapa de profundidad a partir de las imágenes rectificadas. Adicionalmente, se dan a conocer dispositivos para llevar a cabo el método.

ES 2 747 387 A2

DESCRIPCIÓN

DISPOSITIVO Y MÉTODO PARA OBTENER INFORMACIÓN DE PROFUNDIDAD A PARTIR DE UNA ESCENA

Campo técnico

La presente invención está comprendida en el campo del procesado de imágenes digital, y más particularmente a métodos y sistemas para estimar distancias y generar mapas de profundidad a partir de imágenes.

Técnica anterior

La recuperación de información en 3D a partir de imágenes es un problema extensamente investigado en visión artificial, que tiene importantes aplicaciones en robótica, comprensión de escenas y reconstrucción 3D. La estimación de mapas de profundidad se obtiene, en su mayoría, procesando más de una vista (habitualmente dos vistas) de una escena, o bien tomando varias imágenes de la escena con un dispositivo o bien tomando varias imágenes usando varios dispositivos (habitualmente dos cámaras en una configuración de cámara estéreo). Esto se conoce como multivista (o visión estéreo en el caso de dos cámaras o dos vistas) y se basa en técnicas de triangulación. Un enfoque general para extraer la información de profundidad de un punto objeto es la medición del desplazamiento de la imagen de este punto con respecto a las diversas imágenes captadas de la escena. El desplazamiento o disparidad está relacionado directamente con la profundidad real del objeto. Con el fin de obtener la disparidad de un punto, es necesario identificar la posición del mismo punto en el resto de las vistas (o al menos en dos vistas). Habitualmente, este problema se resuelve usando algoritmos de correspondencia, un procedimiento que se conoce bien en el campo de investigación del procesado de imágenes. Sin embargo, las técnicas de visión estéreo presentan dos flaquezas relevantes en comparación con la invención propuesta en este documento: en primer lugar, la necesidad de tener (al menos) dos cámaras resulta una limitación importante en muchos casos, y en segundo lugar, el hecho de que los enfoques estereoscópicos son mucho más costosos computacionalmente, dado que, habitualmente, requieren algoritmos de correspondencia de cálculo elevado (que emparejan patrones a partir de dos o más imágenes).

Una alternativa a tener múltiples dispositivos o a tomar múltiples fotografías de una escena sería el uso de una cámara plenóptica. Las cámaras plenópticas son dispositivos

de obtención de imágenes que pueden captar no solo la información espacial sino también la información angular de una escena en una estructura que se denomina campo claro. Habitualmente, las cámaras plenópticas están compuestas por una lente principal (o un conjunto de lentes equivalentes a dicha lente principal), una matriz de microlentes (MLA) y un sensor.

Las cámaras de tiempo de vuelo (ToF) producen un mapa de profundidad que puede usarse directamente para estimar la estructura 3D del mundo del objeto sin la ayuda de algoritmos de visión artificial tradicionales. Las cámaras ToF trabajan midiendo el retardo de fase de la luz infrarroja (IR) reflejada transmitida anteriormente por la propia cámara. Aunque ya está presente en algunos dispositivos móviles, esta tecnología todavía está lejos aceptarse como una capacidad habitual debido al hecho de que presenta un volumen mucho mayor y disipación de energía (la cámara de obtención de imágenes, el transmisor de IR y la cámara de IR, y el procedimiento para emparejar imágenes entre ambas cámaras), además, la distancia que puede discriminarse con transmisores de IR factibles en la técnica se limita bastante, y las condiciones al aire libre durante días de sol limita adicionalmente su uso, ya que un gran potencia de luz procedente de la luz del día enmascara los sensores de IR.

Habitualmente, los dispositivos móviles incorporan al menos una cámara para seguir tomando imágenes y videos. Las cámaras integradas en dispositivos móviles proporcionan muchas capacidades al usuario, sin embargo, entre estas capacidades, los fabricantes no pueden ofrecer un mapa de profundidad realista de una escena cuando solo se encuentra disponible una cámara.

Existen enfoques que toman en consideración la tarea de estimación de profundidad solo a partir de una única imagen fija como entrada, basándose la mayor parte de las veces en interpretaciones heurísticas de perspectiva y reducciones de tamaño para objetos que se conoce que tienen un tamaño constante. Sin embargo, estos enfoques formulan hipótesis que, a menudo, fallan al generalizar para todas las posibles situaciones de imagen, tal como asumir la perspectiva particular de la escena. También se basan en el uso de conocimiento anterior sobre la escena; lo que resulta, generalmente, una hipótesis altamente poco realista. Los mapas de profundidad obtenidos de esta manera, aunque son útiles para otras tareas, siempre resultarán inherentemente incompletos y no son lo suficientemente exactos como para producir imágenes 3D visualmente satisfactorias.

Otra metodología para obtener información en 3D a partir de imágenes es la obtención de imágenes integral de apertura sintética (SAII). Este método necesita un conjunto de cámaras (o un movimiento mecánico de una cámara que simule tomar fotografías

secuenciales que simule un conjunto de cámaras), que obtiene múltiples perspectivas en alta resolución con la cámara en diferentes puntos del conjunto.

La presente invención usa algunos de los conceptos procedentes de métodos usados por la técnica anterior en fotografía estéreo de manera novedosa: una primera etapa en fotografía estéreo es la “calibración” de las cámaras (etapa que puede evitarse en la presente invención debido al hecho de que se asume que la cámara ya está calibrada), una segunda etapa se denomina “rectificación” (en la que las imágenes procedentes de las dos cámaras en el par estereoscópico se procesan de manera adecuada para deducir las imágenes que se habrían grabado si las dos cámaras del par estereoscópico estuvieran completamente alineadas y fueras coplanares), la “rectificación de cámara” en la presente invención es muy diferente a lo que se realiza en obtención de imágenes estéreo y se describe en detalle a continuación. La tercera etapa en fotografía estéreo es la “correspondencia”, procedimiento para identificar patrones en las dos imágenes del par estereoscópico ya “rectificadas”, para, a continuación, realizar triangulaciones para calcular las distancias hacia el mundo objeto y para componer las imágenes 3D. Las tres etapas descritas “calibración de cámaras”, “rectificación de las imágenes” y “correspondencia entre vistas” (habitualmente dos vistas) son comunes con respecto al “registro”. La invención usa la misma terminología, pero los procedimientos de “correspondencia” y “rectificación” (y, por tanto, de “registro”) son diferentes de los de la técnica anterior, es decir, diferentes con respecto a cámaras estéreos o cámaras multivista.

La invención propuesta asume una situación en la que el usuario quiere obtener un mapa de profundidad de alta resolución a partir de una cámara convencional con una adquisición de único disparo y en tiempo real. La invención se beneficia del movimiento que experimenta la cámara durante el tiempo de disparo, grabándose este movimiento a partir de los datos proporcionados por, por ejemplo, los dispositivos de acelerómetro y giroscopio (dispositivos que están presentes en casi cualquier teléfono móvil en el momento en que se redacta esta patente). El procesado de imagen propuesto en el presente documento mejora los enfoques de la técnica anterior en visión 3D en cuanto a número de imágenes (por tanto, número de cámaras) necesarias, eficacia computacional y requisitos de energía. Por otro lado, la invención mejora los enfoques basándose en cámaras plenópticas en cuanto a resolución espacial y fiabilidad para grandes profundidades en el mapa de profundidad resultante.

Sumario de la invención

El método de procesamiento descrito en el presente documento implementa un algoritmo de correspondencia extremadamente simplificado entre varias imágenes captadas por un dispositivo móvil con una única cámara convencional, varias imágenes que se captan secuencialmente, y la posición en la que se captado cada imagen pueden calcularse usando el acelerómetro, el giroscopio o cualquier otra capacidad de este tipo integrada en el dispositivo móvil, automóvil o cualquier objeto en movimiento. Una vez realizada la compatibilidad de correspondencia entre imágenes, las imágenes se usan para crear un mapa de profundidad denso de la escena. Las imágenes se toman en un único disparo mediante un dispositivo móvil portátil, el movimiento del dispositivo móvil puede detectarse y procesarse durante el lapso de tiempo durante el que tiene lugar el disparo. Este movimiento puede producirse por el movimiento inherente de manos (temblores de mano), mediante las vibraciones de llamadas entrantes (programadas de manera conveniente para vibrar al tiempo que se dispara una fotografía o un video) o porque la cámara está sobre un objeto en movimiento (por ejemplo, un vehículo o automóvil) o porque el usuario está moviéndose. Los métodos descritos en el presente documento pueden adaptarse de manera eficaz con el fin de implementarlos en procesadores paralelos y/o GPU (cada vez más extendidos) así como procesadores paralelos específicos para dispositivos móviles que se hacen funcionar mediante baterías. La invención proporciona procesamiento en tiempo real para grabación de video.

Para la descripción de la presente invención se tendrán en consideración las siguientes definiciones a continuación en el presente documento:

- **Cámara plenóptica**: un dispositivo que puede captar no solo la posición espacial sino también la dirección de llegada de los rayos de luz incidentes.
- **Campo claro**: estructura de cuatro dimensiones $LF(px, py, lx, ly)$ que contiene la información procedente de la luz captada por los pixeles (px, py) bajo las microlentes (lx, ly) en una cámara plenóptica o un sistema de obtención de imágenes integral de apertura sintética.
- **Profundidad**: distancia entre el plano de un punto objeto de una escena y el plano principal de la cámara, siendo ambos planos perpendiculares al eje óptico.
- **Imagen epipolar**: segmento bidimensional de la estructura de campo claro compuesta eligiendo un determinado valor de (px, lx) (epipolar vertical) o (py, ly) (epipolar horizontal) tal como se describe en la **figura** 3.
- **Línea epipolar**: conjunto de pixeles conectados dentro de una imagen epipolar correspondientes a bordes de imagen en el mundo objeto.

- **Vista plenóptica**: imagen bidimensional formada tomando un segmento de la estructura de campo claro eligiendo un determinado valor (px, py), el mismo (px, py) para cada una de las microlentes (lx, ly).

- **Mapa de profundidad**: imagen bidimensional en la que los valores de profundidad calculados del mundo objeto (dz) se añaden como un valor adicional a cada posición (dx, dy) de la imagen bidimensional, componiendo (dx, dy, dz). Cada pixel del mapa de profundidad codifica la distancia al punto correspondiente en la escena

- **Matriz de microlentes (MLA)**: matriz de pequeñas lentes (microlentes).

- **Microimagen**: imagen de la abertura principal producida por una microlente determinada con respecto al sensor.

- **Línea de referencia**: distancia entre el centro de las aberturas de dos imágenes (tomadas por cámaras plenópticas o convencionales o cualquier cámara).

- **Compatibilidad estéreo** (también denominados algoritmos de correspondencia): este término se refiere al procedimiento de, al recibir dos imágenes de la misma escena, conocer qué pixeles de una imagen representan los mismos puntos de la escena en los pixeles de la segunda imagen. Se puede realizar un paralelismo con los ojos humanos, el problema entonces es qué puntos observados por el ojo izquierdo corresponden a qué puntos observados por el ojo derecho.

- **Disparar**: acto de presionar el botón con el fin de tomar una fotografía. En última instancia, pueden adquirirse muchos fotogramas durante esta situación.

- **Disparo**: acto de haber presionado el botón con el fin de tomar una fotografía.

- **Exposición**: un sensor de cámara se expone a luz entrante si su abertura está abierta, permitiendo a la luz entrar en la cámara.

- **Acelerómetro**: dispositivo que registra la aceleración lineal de movimientos de la estructura a la que está unido (habitualmente en las direcciones x, y, z).

- **Giroscopio**: dispositivo que permite la aceleración de rotación angular (en contraposición a la aceleración lineal del acelerómetro) habitualmente con respecto al eje de tres rotaciones (cabeceo, balanceo y guiñada; en contraposición a x, y y z en acelerómetros).

- **IMU y AHRS**: las unidades de medición inercial (IMU) y sistemas de referencia de actitud y rumbo (AHRS) son dispositivos electrónicos que monitorizan e informan sobre la fuerza específica de un objeto, velocidad angular, y en ocasiones el campo magnético que rodea el cuerpo usando una combinación de acelerómetros y giroscopios, y en ocasiones también magnetómetros. Los IMU y AHRS se usan, normalmente, dentro de aeronaves, incluyendo vehículos aéreos no tripulados (UAV) y embarcaciones incluyendo submarinos y vehículos submarinos no tripulados (UUV). La principal diferencia entre una unidad de medición inercial (IMU) y un AHRS es la adición

de sistema de procesamiento integrado (que, por ejemplo, puede incluir microprocesadores y memorias) en un AHRS que proporciona información de actitud y rumbo, en comparación con IMU que solo proporcionan datos de sensor a un dispositivo adicional que calcula actitud y rumbo.

- **Velocímetro**: un instrumento que mide e indica el cambio de posición de un objeto a lo largo del tiempo (velocidad).

- **GPS**: el sistema de localización global (GPS) es un sistema de navegación global por medio del uso de satélites que proporcionan la geolocalización y la información de tiempo a un receptor de GPS.

- **Rectificación de imagen**: en el contexto de esta invención el procedimiento de aplicar homografías bidimensionales a imágenes adquiridas en diferentes momentos en el tiempo moviendo las cámaras cuya geometría tridimensional se conoce, de modo que líneas y patrones en las imágenes originales (con respecto a un sistema de referencia de seis ejes [x´, y´ ,z´, cabeceo´, balanceo´ y guiñada´] en el que la cámara en movimiento se dispara tras una determinada cantidad de tiempo t1) se mapean para alinear líneas y patrones en las imágenes transformadas (con respecto a un sistema de referencia de seis ejes [x, y ,z, cabeceo , balanceo y guiñada] en el que la cámara se encontraba en tiempo cero), dando como resultado dos imágenes (inicialmente adquiridas en tiempos t1 y cero) que son imágenes comparables como si se hubieran adquirido mediante cámaras coplanares con el mismo z, cabeceo, balanceo y guiñada, y con valores “rectificados” de x e y que dependen del movimiento a lo largo de esos dos ejes (líneas de referencia en x e y entre tiempo 0 y tiempo t1). Tras la rectificación de imagen, puede usarse el procedimiento de disparos en tiempo 0 y en tiempo t1 para componer vistas diferentes de “cámaras estéreos virtuales”, y/o vistas diferentes de “cámaras multivista virtuales” y/o vistas diferentes de “cámaras plenópticas virtuales”.

- **Dispositivo móvil**: pequeño dispositivo informático, generalmente lo suficientemente pequeño para hacerse funcionar manualmente. También tienen cámaras integradas y otras capacidades tales como GPS, acelerómetro, giroscopio, etc. pueden ser teléfonos móviles, tabletas, ordenadores portátiles, cámaras y otros dispositivos.

- **Cámara convencional:** dispositivo que solo capta la posición espacial de los rayos de luz incidentes en el sensor de imagen, de manera que cada pixel del sensor integra toda la luz que proviene de cualquier dirección a partir de la totalidad de la abertura del dispositivo.

- **Obtención de imagen integral de apertura sintética (SAII):** una matriz de sensores de imagen (cámaras) distribuidos en una red aleatoria homogénea o (alternativamente).

En esencia, la presente invención da a conocer un método para obtener información de profundidad a partir de una escena, que comprende las etapas de:

a) adquirir una pluralidad de imágenes de la escena por medio de al menos una cámara durante un tiempo de disparo en el que la pluralidad de imágenes ofrece al menos dos vistas diferentes de la escena;

b) para cada una de las imágenes de la etapa a), simultáneamente adquirir datos sobre la posición de las imágenes con respecto a un sistema de referencia de seis ejes;

c) seleccionar de las imágenes de la etapa b) al menos dos imágenes;

d) rectificar las imágenes seleccionadas en la etapa c) generando de ese modo un conjunto de imágenes rectificadas; y

e) generar un mapa de profundidad a partir de las imágenes rectificadas.

La posición de las imágenes durante el tiempo de disparo puede medirse a partir de un conjunto de datos de colocación adquiridos por medio de al menos un dispositivo de colocación, por ejemplo, un dispositivo seleccionado del grupo de: un acelerómetro, una unidad de medición inercial (IMU), un sistema de referencia de actitud y rumbo (AHRS), un GPS, un velocímetro y/o un giroscopio.

Las unidades de medición inercial (IMU) y los sistemas de referencia de actitud y rumbo (AHRS) son dispositivos electrónicos que monitorizan e informan sobre la fuerza específica de un objeto, la velocidad angular, y en ocasiones, el campo magnético que rodea el cuerpo usando una combinación de acelerómetros y giroscopios, y en ocasiones también magnetómetros. Normalmente, los IMU y AHRS se usan dentro de aeronaves, incluyendo vehículos aéreos no tripulados (UAV) y embarcaciones incluyendo submarinos y vehículos submarinos no tripulados (UUV). La principal diferencia entre una unidad de medición inercial (IMU) y un AHRS es la adición de un sistema de procesamiento integrado (que, por ejemplo, puede incluir microprocesadores y memorias) en un AHRS que proporciona información sobre actitud y rumbo, en comparación con IMU que solo proporcionan datos de sensor a un dispositivo adicional que calcula actitud y rumbo.

Con el fin de lograr una mejor precisión, el dispositivo de colocación puede unirse de manera rígida a al menos una de las cámaras.

En una realización, al menos una cámara se asocia con un dispositivo móvil. Tal dispositivo móvil puede ser, por ejemplo, un teléfono inteligente, una tableta, un ordenador portátil o una cámara compacta.

En una realización más preferida, en la etapa c), las imágenes se seleccionan basándose en sus posiciones en el sistema de referencia de seis ejes.

En una primera realización preferida, las imágenes se seleccionan de modo que sus distancias relativas son lo suficientemente pequeñas como para provocar una

disparidad máxima de, como máximo, un pixel. En este caso, en la etapa e) un sistema de obtención de imágenes integral de apertura sintética virtual puede generarse con las imágenes rectificadas generando de ese modo un conjunto de imágenes epipolares.

En una segunda realización preferida, las imágenes se seleccionan de modo que sus distancias relativas son lo suficientemente grandes como para provocar una disparidad de más de un pixel. En este caso, en la etapa e), un sistema plenóptico-estereoscópico virtual se genera con las imágenes rectificadas generando de ese modo un conjunto de imágenes epipolares extendidas.

Una vez se generan las imágenes epipolares mediante, por ejemplo, la primera realización preferida o la segunda realización preferida, la etapa e) puede comprender, además, calcular pendientes de líneas epipolares a partir de las imágenes epipolares. Con estas imágenes epipolares, puede generarse un mapa de profundidad de la escena convirtiendo las pendientes de las líneas epipolares en profundidades. Adicionalmente, las pendientes pueden obtenerse analizando las líneas epipolares horizontal y vertical combinadas en una matriz multidimensional.

El método de la presente invención puede comprender una etapa de generar una imagen tridimensional de la escena a partir del mapa de profundidad. En particular, las profundidades/pendientes pueden calcularse en las líneas epipolares horizontal y/o vertical combinadas directamente en un mapa de profundidad/pendiente disperso bidimensional. Además, el mapa de profundidad/pendiente disperso puede rellenarse aplicando técnicas de relleno de imagen para obtener valores de profundidad/pendiente para cada pixel. Preferiblemente, para la estimación de profundidad los cálculos se realizan solo para aquellos pixeles de los sensores en los que los bordes del mundo objeto se han detectado

En la etapa a), al menos una cámara se mueve, preferiblemente, durante el tiempo de disparo, por ejemplo, debido a movimientos aleatorios indeterminados producidos por temblores de mano de ser humano o uniendo al menos una cámara a una estructura que se mueve con respecto a la escena (por ejemplo la cámara está montada o colocada en una ubicación de automóvil con amplia visibilidad a zonas de interés fuera del automóvil, o para medir distancias dentro del automóvil para aplicaciones tales como reconocimiento de gestos).

También, la pluralidad de imágenes de la etapa a) se adquieren, preferiblemente, por al menos dos cámaras. En este caso, las al menos dos cámaras pueden alinearse con sus posiciones relativas que se conocen.

En una realización preferida, una secuencia de video está compuesta por al menos dos niveles de profundidad de primer plano, planos medios opcionales e imágenes bidimensionales anteriores (ubicadas en diferente profundidades en el mundo objeto) y

en el que dicha combinación de diferentes niveles de imágenes bidimensionales en fotogramas sucesivos y/o el cambio de oclusiones en imágenes bidimensionales más próximas a las anteriores y/o el cambio de perspectiva y el tamaño en imágenes bidimensionales más próximas al primer plano produce una percepción en 3D al usuario.

Además, en una realización a modo de ejemplo, solo algunas o todas las imágenes epipolares distribuidas a lo largo de la dimensión vertical/horizontal se tienen en consideración con el fin de reducir ruido estadístico.

Breve descripción de los dibujos

A continuación, se describe brevemente una serie de dibujos que ayudan a comprender mejor la invención y que están relacionados de manera expresa con realizaciones de dicha invención, presentados como ejemplos no limitativos de la misma.

La figura 1 muestra un esquema de la cámara 100 plenóptica, que incluye un sensor, una MLA (matriz de microlente), y una lente principal de la cámara. También muestra dos microimágenes.

La figura 2 muestra una realización (2A) de cámara plenóptica con el patrón producido sobre el sensor (2B) para un punto en el mundo objeto ubicado más lejos de la cámara que el plano conjugado de la MLA.

La figura 3 muestra el procedimiento de formación de imágenes epipolares centrales horizontal y vertical a partir de un campo claro para radiar puntos en el mundo objeto.

La figura 4 muestra una posible realización de un sistema de obtención de imágenes integral de apertura sintética multivista (SAII): una matriz bidimensional de MxN cámaras.

La figura 5A ilustra una comparación entre la línea de referencia de una cámara plenóptica (“línea de referencia estrecha”) y la línea de referencia entre dos cámaras en una configuración estéreo (“línea de referencia ancha”). La cámara en la parte superior es una cámara plenóptica y la cámara en la parte inferior es una cámara convencional, ambas dispuestas en una configuración de cámara plenóptica estéreo.

La figura 5B muestra un dispositivo móvil con una cámara plenóptica y dos cámaras convencionales adicionales (sin embargo, cualquiera de las dos cámaras adicionales puede ser o bien una cámara convencional o bien una cámara plenóptica).

Las figuras 6A y 6B ilustran el procedimiento de extensión de una imagen epipolar captada con una cámara plenóptica con una imagen bidimensional de la misma escena captada por una cámara convencional con ambas cámaras en una configuración estéreo como en la figura 5A.

La figura 7A muestra un sistema de referencia de seis ejes (x, y, z, cabeceo, balanceo y guiñada) que incluye todos los posibles movimientos que pueden grabarse por

acelerómetros y giroscopios en un teléfono móvil (o cualquier dispositivo móvil) que incluye una cámara.

La figura 7B muestra un ejemplo de los datos adquiridos a partir del acelerómetro de un dispositivo móvil (aceleraciones en las direcciones x, y y z).

La figura 8 muestra los procedimientos de “rectificación” y “correspondencia” en un sistema de par estereoscópico formado por dos cámaras.

La figura 9 ilustra el procedimiento de “rectificación” para una matriz de 4 cámaras.

La figura 10 muestra cómo el sistema de referencia de 6 ejes asociado con una cámara dada cambia si el movimiento de la cámara implica deltas positivos en la posición x, y y z, así como una rotación de guiñada negativa.

La figura 11 muestra un cambio en el sistema de referencia de la cámara para una traslación negativa en x e y, una traslación positiva en z, así como una rotación de balanceo positiva.

La figura 12 muestra un cambio en el sistema de referencia de la cámara para una traslación positiva en las direcciones x e y, una traslación negativa en z, así como una rotación de cabeceo positiva.

La figura 13 ilustra un sistema multivista con un ejemplo de trayectoria de una única cámara moviéndose a través de posiciones A, B, C y D a lo largo de una zona bidimensional del mismo tamaño que una matriz de MxN cámaras.

La figura 14 muestra una grabación de 2 segundos de los movimientos espaciales (en milímetros) detectados por acelerómetros a partir de teléfonos inteligentes fabricados en serie en las direcciones x e y al tiempo que el teléfono se sostiene por un ser humano para tomar una fotografía.

La figura 15A muestra un dispositivo móvil electrónico que incluye un sistema multivista que adquiere imágenes que se someten a tratamiento a través de un procesador que incluye un procesador de múltiples núcleos.

La figura 15B es tal como la figura 15A, pero con dos CPU (unidades de procesamiento central) en lugar de un procesador de múltiples núcleos.

La figura 15C es tal como la figura 15B, pero las CPU se sustituyen por una GPU (Unidad de procesamiento de gráficos) que incluye un gran número de procesadores paralelos.

La figura 16A muestra un procedimiento de rectificación de imagen y correspondencia de cámara estéreo.

La figura 16B muestra un método para calcular un mapa de profundidad según la invención en esta divulgación.

Descripción detallada

La presente invención se refiere a un dispositivo y método para generar un mapa de profundidad a partir de un campo claro. Un campo claro puede captarse puede captarse por múltiples tipos de dispositivos. Por motivos de simplicidad, se tendrá en consideración un primer ejemplo en el que una cámara convencional que está moviéndose al tiempo que toma varias imágenes. El método descrito en el presente documento crea un sistema de obtención de imágenes equivalente de esas imágenes captadas mediante un dispositivo en movimiento y aplica algoritmos plenópticos para generar un mapa de profundidad de una escena.

En un ejemplo adicional, el método se describe aplicándolo a sistemas formados por varias cámaras en movimiento con la posibilidad de incluir una o más cámaras plenópticas y una o más cámaras convencionales. No obstante, el método descrito en el presente documento puede aplicarse a campos claros captados por cualquier otro dispositivo, incluyendo otros dispositivos de obtención de imágenes integrales.

La figura 1 ilustra una realización de una cámara 100 plenóptica: un sensor 1, las microlentes 22, y el cilindro superior de componentes ópticos (o lente 3 principal de la cámara). La figura 1 muestra dos conjuntos de rayos que cruzan la abertura principal del sistema plenóptico y que alcanzan la parte central y próximo a las microlentes centrales. Las microimágenes 11, 12 no se solapan si el sistema óptico se diseña de manera adecuada.

La figura 2 muestra un punto 210 objeto que está más lejos que el plano conjugado de la matriz 22 de microlentes a través de la lente 3 principal. Por tanto, ilumina más de una microlente, de manera que el punto de enfoque es más próximo a la lente 3 principal que la posición de la matriz 22 de microlentes y, por tanto, el patrón captado por el sensor 206 de imagen se muestra en la figura 2B. Los niveles grises en algunas de las microimágenes 212 corresponden a pixeles parcialmente iluminados mientras que en los pixeles blancos la totalidad de la zona del pixel se ha visto impactada por la luz que proviene del punto 210 objeto en el mundo objeto.

La base de la obtención de imágenes plenópticas es que los objetos en el mundo a diferentes profundidades o distancias con respecto a la cámara producirán diferentes patrones de iluminación en el sensor de una cámara plenóptica. Los diversos patrones captados por el sensor pueden representarse en imágenes epipolares, que proporcionan información de profundidad implícita de objetos en el mundo.

La figura 3 muestra el procedimiento de formación de imágenes epipolares centrales horizontal 300 y vertical 302 a partir de un campo 206 claro para radiar puntos en el mundo 210 objeto ubicados a diferentes distancias con respecto a una cámara 100 plenóptica: en la distancia conjugada desde las microlentes 22 (figura 3A), más próximas

a la distancia conjugada (figura 3B), y más alejadas a la distancia conjugada (figura 3C), mostrando de ese modo la inherente capacidad de cámaras plenópticas para calcular distancias con respecto al mundo objeto. El caso de la figura 3C se visualiza en las figuras 2A y 2B, que muestran cómo la luz procedente del punto 210 de radiación en el mundo objeto se propaga en el interior de la cámara 100, cruzando las microlentes 22 e imprimiendo un patrón de luz sobre el sensor 206.

El procedimiento para transformar los patrones encontrados en imágenes epipolares para información de profundidad requiere la aplicación de algunas técnicas de procesado de imagen que se conocen bien en la técnica anterior. Las imágenes epipolares contienen líneas epipolares; que son pixeles conectados que forman una línea (correspondiendo varios pixeles de sensor al mismo punto en el mundo objeto), tal como se muestra claramente en las figuras 2B y 3C para fuentes de radiación de mundo más alejadas que el punto de enfoque de las microlentes (línea epipolar inclinada hacia la izquierda en la figura 3C), para fuentes de radiación de mundo más próximas que el foco de las microlentes (línea epipolar inclinada hacia la derecha en la figura 3B), y para fuentes de radiación de mundo enfocadas de manera precisa sobre la superficie de la microlente (línea epipolar vertical en la figura 3A). Las pendientes de estas líneas epipolares están relacionadas directamente con la forma del patrón iluminado sobre las microlentes y con la profundidad correspondiente de ese punto en el mundo objeto. Resumiendo, el procedimiento, los patrones encontrados en imágenes epipolares, las líneas epipolares, proporcionan información sobre la profundidad de los objetos en el mundo objeto real. Estas líneas pueden detectarse usando algoritmos de detección de borde y sus pendientes pueden medirse. Por tanto, la pendiente de cada línea epipolar aporta un valor que procesado de manera conveniente proporciona la profundidad real del punto en el mundo objeto que produjo tal patrón.

Aunque es una tecnología muy prometedora, la obtención de imágenes plenópticas también tiene un coste, dado que el rendimiento de una cámara plenóptica se ve limitado por la resolución de la matriz de microlentes, lo que da como resultado una resolución de imagen mucho menor que los dispositivos de obtención de imágenes tradicionales. Además, las cámaras plenópticas son una tecnología bastante nueva que todavía resulta difícil de encontrar en dispositivos móviles.

La figura 4 muestra una posible configuración SAII (obtención de imágenes integral de apertura sintética) de una matriz de cámaras. Esta matriz puede presentar MxN cámaras o una única cámara que se mueve a lo largo de la matriz (por ejemplo, comenzando en la posición 1, entonces 2, 3, etc., hasta la posición MN) que toma una imagen fija en cada posición de la matriz. El paralelismo con una cámara plenóptica es evidente y las

mismas imágenes epipolares anteriormente descritas para una cámara plenóptica can pueden obtenerse con una SAII, tal como se conoce bien, una cámara plenóptica tal como en la figura 2A con “OxP” pixeles por microlente y “TxS” microlentes es funcionalmente equivalente a “OxP” cámaras convencionales con “TxS” pixeles con las cámaras separadas de manera uniforme sobre la pupila de entrada de la cámara plenóptica. Del mismo modo, la matriz de MxN cámaras de la figura 4 (con QxR pixeles por cámara) es equivalente a un sistema plenóptico tal como en la figura 1 con MxN pixeles por microlente 22 y varios pixeles por cámara 51 equivalente que son los mismos que el número total de microlentes en la cámara plenóptica equivalente. La única diferencia práctica es que el tamaño de este número (QxR) en un sistema SAII, debido a las limitaciones de tecnología e implementación, es mucho mayor que el número de microlentes que pueden diseñarse en una cámara plenóptica. Mapas de profundidad calculados a partir de una SAII pueden beneficiarse de una línea de referencia más amplia que las cámaras plenópticas dado que la distancia entre nodos en la figura 4 (que puede ser tan elevada como varios cm o incluso superior) es mayor que la distancia entre las cámaras equivalentes “OxP” de las cámaras plenópticas (varios milímetros y en cámaras pequeñas de hasta una décima de un milímetro). Las figuras 5A y 5B (una vista bidimensional lateral de cámaras que puede extrapolarse de manera evidente a una configuración tridimensional en la que la tercera dimensión sería perpendicular al papel sin perder la generalidad de la discusión posterior) compara la línea de referencia de una cámara plenóptica (“línea de referencia estrecha” *d* que muestra la separación *d* entre las OxP cámaras equivalentes de una cámara plenóptica con TxS pixeles por microlente, con cada cámara 51 equivalente que tiene tantos pixeles como microlentes en la cámara plenóptica) y la “línea de referencia ancha” B entre las dos cámaras de una cámara estéreo o una línea de referencia de SAII más amplia: en un ejemplo práctico de una cámara estéreo o una SAII la “línea de referencia ancha” B puede ser de unos pocos centímetros, mientras que en una cámara plenóptica habitual la “línea de referencia estrecha” *d* puede alcanzar valores tan pequeños como milímetros o incluso una décima de un milímetro. La matriz de MxN cámaras de la figura 4 (con QxTR pixeles por cámara) es equivalente a un sistema plenóptico tal como en la figura 1 con MxN pixeles por microlente 22 y varios pixeles por cámara 51 equivalente que son los mismos que el número total de microlentes en la cámara equivalente plenóptica (Qx), el tamaño de este número (QxT) en este caso (un sistema SAII) es mucho mayor que el número de microlentes que pueden diseñarse en una cámara plenóptica. Obviamente, los sistemas SAII ofrece una mayor resolución que una cámara plenóptica y la línea de referencia más amplia hace que sea más preciso calcular la profundidad a grandes distancias de la cámara.

La invención propuesta obtiene un mapa de profundidad de alta resolución a partir de una cámara convencional con una adquisición de único disparo y, en caso de grabación de video, el mapa de profundidad se obtiene en tiempo real. La invención usa el movimiento y las vibraciones experimentadas por la cámara durante el tiempo en el que se realiza un disparo para obtener una secuencia de fotogramas, simulando de ese modo las diversas imágenes de una SAII (o las cámaras equivalentes de una cámara plenóptica) con la secuencia de fotogramas adquiridos por la cámara en movimiento. La presente invención usa las distancias de la cámara entre las adquisiciones elegidas como líneas de referencia (distancias entre vistas) de un sistema multivista que pueden usarse para estimar la profundidad de la escena. El objetivo principal de estos métodos es proporcionar la capacidad para crear un mapa de profundidad de alta resolución cuando solo se encuentra disponible una cámara convencional y en solo un disparo (implicando el disparo la adquisición de varios fotogramas). La presente invención es muy eficaz computacionalmente, tan eficaz que puede usarse para obtener mapas de profundidad en tiempo real en secuencias de video incluso en dispositivos móviles económicos (la mayor parte del tiempo con procesadores económicos que se hacen funcionar mediante baterías, en los que son necesarios cálculos eficaces para evitar vaciar las baterías rápidamente).

La invención propuesta tiene dos etapas principales tras la grabación de varios fotogramas consecutivos. En primer lugar, una etapa para “corregir” las imágenes adquiridas durante el tiempo de disparo (cada imagen adquiridas con la cámara en posiciones ligeramente diferente para x, y, z, guiñada, cabeceo y balanceo) para obtener un conjunto de “imágenes corregidas” que están relacionadas entre sí como si se hubieran realizado por una única cámara plenóptica o un único sistema SAII de obtención de imágenes (un procedimiento de “rectificación de imagen” tal como en la figura 9, o también produciendo una serie de imágenes A, B, C y D en la figura 13). Esta primera etapa realiza la rectificación entre imágenes (tal como en la figura 9) usando las grabaciones procedentes del acelerómetro y el giroscopio, o cualquier otra capacidad que pueda encontrarse en cualquier teléfono inteligente, automóvil u objeto en movimiento actual. Una segunda etapa se aplica para crear un mapa de profundidad usando algoritmos plenópticos. Esto consiste en calcular el valor de profundidad de cada punto en la escena detectando la pendiente de las líneas epipolares de una imagen epipolar. En una realización, este cálculo puede realizarse solo para los bordes detectados en la escena, en lugar de realizarlo para todos los pixeles de la escena. El método de la presente invención puede procesar imágenes de vídeo en tiempo real (aproximadamente 15 fotogramas por segundo y más) al tiempo que implementaciones

anteriores utilizan desde cientos de milisegundos hasta minutos solo para procesar un único fotograma.

Un temblor de mano normal (o temblor fisiológico) es un temblor pequeño, casi imperceptible que es difícil de percibir por el ojo humano y no interfiere con actividades. La frecuencia de las vibraciones es entre 8 y 13 ciclos por segundo y es un temblor normal en cualquier persona (no se considera que esté asociado con ningún proceso de enfermedad). Incluso estos pequeños temblores pueden usarse como una fuente para generar un movimiento en la cámara que puede crear una línea de referencia para la detección de profundidad.

Los sensores más habituales que determinan la posición y orientación de un objeto son el giroscopio y el acelerómetro. Ambos se encuentran presentes en la mayor parte de los dispositivos móviles actuales (teléfonos inteligentes y otros), y cuando la información se graba por ambos dispositivos simultáneamente al procedimiento de adquisición de imagen, es posible conocer para cada fotograma grabado el FOV exacto (campo de visión) (en cuanto a la posición x, y, z de la cámara en el mundo tridimensional, y la dirección a la que la cámara está dirigida en el momento del disparo, definida por los 3 ángulos fundamentales, cabeceo, balanceo y guiñada, tal como se describe en la figura 7A). Para registrar movimientos, la frecuencia de muestreo normal de la técnica anterior de acelerómetros y giroscopios es de aproximadamente 500Hz, esto significa que el acelerómetro y el giroscopio son lo suficientemente sensibles para registrar movimientos de temblor de mano (entre 8 y 13 ciclos por segundo). La figura 7B muestra una secuencia de movimientos registrada por el acelerómetro de un teléfono móvil en las direcciones X, Y Z. Comienza con el móvil colocado en la mano, en una posición tal como si fuéramos a tomar una fotografía. En determinado tiempo, se “presiona el botón” (se dispara el elemento de activación) para tomar una fotografía y, tras esto, se deja el dispositivo móvil sobre la mesa. Toda la secuencia lleva 10 segundos con una frecuencia de muestreo de 100Hz (dando como resultado aproximadamente 1000 muestras). Estos datos también pueden obtenerse para el dispositivo de giroscopio. Aunque los acelerómetros y giroscopios presentan algo de retardo con la información que proporcionan, sus mediciones tienen diferentes características. Los acelerómetros miden aceleraciones físicas triaxiales (X-Y-Z) mientras que los giroscopios miden aceleración angular triaxial (P-R-Y) a lo largo de cada eje de rotación, y la combinación de ambos dispositivos proporciona la detección de movimiento de 6 ejes, captando cualquier posible movimiento del dispositivo móvil para una determinación rápida y exacta de la posición y orientación relativas de la cámara. Estos parámetros de posición y orientación relativos se usan en la formación de “captación virtual” a partir de “sistemas SAII virtuales” (o “cámaras plenópticas virtuales”) y para componer imágenes epipolares

tal como se explicará a continuación. La figura 7A muestra un sistema de coordenadas de 6 ejes asociado con una cámara en un teléfono móvil que se usará para describir los movimientos registrados por el acelerómetro y el giroscopio.

Si se presupone una posición inicial determinada de un dispositivo móvil y un periodo de adquisición de imagen que comienza cuando el usuario presiona el botón para tomar una fotografía. Tal como se explica, al usar los datos procedentes del acelerómetro y el giroscopio la posición relativa del dispositivo móvil con respecto a esa posición inicial puede llevarse a cabo en cualquier tiempo durante el tiempo de exposición de las adquisiciones de secuencia de imagen que se producen tras presionar el botón de obturador. La figura 13 muestra un ejemplo de una trayectoria seguida por el dispositivo móvil durante determinado intervalo de tiempo. Durante este tiempo, el dispositivo móvil ha completado la trayectoria indicada por la línea discontinua, y también toma imágenes cuando se encontraba en las posiciones A, B, C y D. El ejemplo de la figura también muestra una MxN matriz como anterior con el fin de comparar el procedimiento secuencial de adquisición de imagen descrito con un sistema SAII virtual ubicado en un plano lo más próximo a la ubicación en donde se produjeron los disparos A-B-C-D. Por tanto, si el movimiento del dispositivo móvil se registra y procesa de manera apropiada, ambos sistemas (una SAII y la invención propuesta) son funcionalmente equivalentes.

Ahora va a describirse en detalle el intervalo de tiempo en el que la invención adquiere las imágenes y registra el movimiento del dispositivo móvil. La mayor parte de dispositivos móviles de hoy en día pueden adquirir imágenes con una frecuencia de velocidad de fotograma de aproximadamente 120 fotogramas por segundo (fps), lo que es significativamente mayor que lo que se considera como tiempo real (un valor subjetivo fijado por algunos entre 15 fps y 30 fps o un mayor número de fotogramas por segundo). Al presuponer que un dispositivo móvil de esa naturaleza incluye una cámara convencional y va a tomar una fotografía cuando se sostiene en una posición dada por una mano de humano (estos no están destinados a considerarse factores limitativos sino ejemplos). Si se graban imágenes durante 1 segundo, a 120 fps pueden elegirse, por ejemplo, cuatro imágenes dentro de este periodo con líneas de referencia dadas entre las mismas. Supongamos, también, que la trayectoria mostrada en la figura 13, se ha dibujado enfrente de una matriz de MxN posiciones para mantener un mejor paralelismo entre el método propuesto y un sistema SAII de MxN cámaras o una cámara plenóptica con MxN pixeles por microlente. A partir de esta trayectoria que se provoca de manera involuntaria por los temblores de mano se puede seleccionar, por ejemplo, aquellos puntos que maximizan la distancia total (tanto horizontal como verticalmente) dentro de la trayectoria. La resolución de los mapas de profundidad a largas distancias mejora con líneas de referencia más amplias y, por tanto, la selección de esas imágenes

que están tan separadas una con respecto a otra como sea posible, es la mejor solución para discriminar distancias con respecto al mundo objeto tanto como sea posible. Obsérvese que el ejemplo de trayectoria de la figura 13 es una simplificación 2D. para hacer que funcione la invención propuesta como un sistema SAII, las diferentes imágenes tomadas a lo largo de la trayectoria deben “rectificarse” según los parámetros de movimiento grabados por el acelerómetro, el giroscopio o cualquier otro dispositivo de este tipo, teniendo en cuenta los 6 grados de libertad (las seis posiciones x, y, z, P, R y Y).

Ahora va a definirse cómo se realiza la rectificación del procesamiento de imagen para la obtención de imágenes estéreo y las diferencias para la presente invención. La figura 8 muestra cómo se graba un patrón 81 por dos cámaras diferentes en una configuración estéreo. El patrón 81 se capta a partir de dos puntos de vista diferentes, grabando dos imágenes 82 y 83 planas. Estas dos imágenes estéreo se “rectifican” para obtener las imágenes que se habrían obtenido si las dos cámaras hubieran estado completamente alineadas, que es en la misma posición y y z en el espacio, con una distancia x fija conocida entre las mismas, estando ambas cámaras situadas en el mismo plano (habitualmente conocido condición coplanaria, lo que significa que su diferencia de balanceo y cabeceo es cero, o que sus ejes ópticos son paralelos), y con una diferencia de guiñada de cero entre los mismos (equivalente a afirmar que ambas imágenes 84 y 85 deben presentar el mismo grado de horizontalidad). La figura 9 muestra cómo una cámara movida por temblores de una mano humana que graba cuatro disparos diferentes (91 a 94) en cuatro instantes diferentes con cuatro posiciones diferentes de la cámara en un sistema de referencia de cinco ejes (x, y, cabeceo, balanceo y guiñada), es diferente de lo que habría grabado un sistema SAII con cuatro cámaras ubicadas en la posición de puntos (95 a 98). El procedimiento de “rectificación” para este sistema implica calcular un conjunto de imágenes 95-98 rectificadas procedentes de un conjunto de imágenes 91-94 adquiridas. Esta es una vista simplificada, ya que no implica movimientos en z y asume una buena superposición entre las imágenes 91 adquiridas y el lugar en el que se desea que estén las imágenes, o las imágenes 95-98 rectificadas. Obsérvese, sin embargo, que la rectificación en z es también muy importante cuando la cámara se coloca en una estructura móvil tal como un automóvil, siendo este valor directamente proporcional a su velocidad. Una realización más realista de la presente invención realiza una grabación secuencial de varios fotogramas de video (que, por ejemplo, pueden ser 120 fotogramas por segundo) y una grabación simultánea de la posición de cámara dentro de un sistema de referencia de seis ejes (x, y, z, cabeceo, balanceo y guiñada).

Esto se ejemplifica en la figura 10: en un tiempo dado la cámara capta un fotograma, momento en el que la cámara se ubica con sus 6 ejes asociados en una ubicación dada en el espacio (x, y, z, cabeceo, balanceo y guiñada), cuando la cámara capta el siguiente fotograma, el sistema de referencia de seis ejes se mueve a un nuevo lugar que se conoce porque su nueva posición (x´, y´, z´, cabeceo´, balanceo´ y guiñada´) se ha grabado por el acelerómetro y el giroscopio asociados con la cámara. En este ejemplo particular de la figura 10, había tres movimientos positivos en x, y y z, así como una rotación de guiñada negativa. La figura 11 muestra otro ejemplo en el que entre el primer fotograma y el segundo fotograma x e y existía un movimiento negativo, z un movimiento positivo, así como una rotación de guiñada positiva. La figura 12 es todavía otro ejemplo en el que entre los fotogramas primero y segundo los movimientos en x e y fueron positivos, z negativo, así como una rotación de cabeceo positiva.

Ahora van a comprarse los tiempos necesarios y qué es factible desde el punto de vista tecnológico para lograr el objetivo. Los temblores de mano muestran movimientos de baja frecuencia de 8 a 13 ciclos por segundo, durante los que pueden tomarse segundos 120 disparos mediante sistemas de cámara y fotosensores de la técnica anterior, y durante los que segundas 500 lecturas pueden muestrearse mediante acelerómetros y giroscopios de la técnica anterior. La figura 14 es una grabación durante 2 segundos de los movimientos espaciales detectados por acelerómetros a partir de teléfonos inteligentes fabricados en serie en las direcciones x e y (dirección z y guiñada, balanceo y cabeceo habituales también pueden grabarse y usarse en los cálculos), en este caso particular de la figura 14 el teléfono se sostiene por un usuario de rifle amateur (para una persona normal los movimientos son ligeramente mayores, para una persona que padece Parkinson los movimientos son mucho mayores), la figura muestra un intervalo de movimientos de casi 4 milímetros en los ejes x (vertical en la figura 14) y casi 2 milímetros en los ejes y (horizontal en la figura 14). Estos desplazamientos son mayores que la “línea de referencia estrecha” habitual *d* de separación entre cámaras equivalentes de una cámara plenóptica habitual (una pupila de entrada de 2 milímetros y 10 pixeles por microlente produce una línea de referencia mínima (“línea de referencia estrecha” *d* en la figura 5A de 0,2 mm); o si se compara una línea de referencia habitual *d* de una cámara plenóptica de 0,1 a 0,3 mm, la figura 14 muestra que es probable que la misma línea de referencia se produzca de cada 100 a 200 milisegundos si se produce por temblores de mano. Razón por la que la invención propuesta necesita aproximadamente 200ms para adquirir imágenes y datos suficientes para crear un mapa de profundidad de la escena. En una realización, al captar imágenes a una velocidad habitual de fotograma de 120 fps dentro de un intervalo de tiempo de 200ms la invención adquiere 24 fotogramas. Estos fotogramas se toman cuando el dispositivo móvil está en

movimiento debido a temblores de mano o a cualquier otra vibración. A partir de estos 24 fotogramas, lo 2 fotogramas con la mayor línea de referencia entre ellos puede elegirse, siendo esta línea de referencia lo suficientemente larga para mejorar la calidad del mapa de profundidad de una cámara multivista en cuanto a la precisión obtenida para distancias más largas.

Una vez se han captado varias imágenes y sus parámetros de movimiento correspondientes (posición x, y, z, P, R, Y) con una cámara convencional a medida que la cámara se mueve en el espacio 3D durante un determinado periodo de tiempo, se crea el SAII equivalente (o cámara plenóptica equivalente) rectificando todas estas imágenes según los parámetros de movimiento (nuevas posiciones en el sistema de referencia de 6 ejes). Entonces, se forman las imágenes 300 y 302 epipolares y se aplican los algoritmos plenópticos para generar un mapa de profundidad de una escena. Una característica de las cámaras plenópticas es que la disparidad máxima entre cámaras equivalentes consecutivas es +-1 pixel, lo que implica que los pixeles que forman una línea epipolar siempre están conectados entre sí. Por tanto, con el fin de aplicar algoritmos plenópticos de manera apropiada a la SAII equivalente creada (o cámara plenóptica equivalente), la línea de referencia entre imágenes consecutivas debe garantizar que no se crean huecos cuando se forman las imágenes epipolares. No obstante, esto no siempre resulta posible de garantizar dado que los movimientos de temblores humanos en la figura 14 se contaminan en ocasiones por movimientos anormalmente grandes que no pueden modelarse como sistemas SAII (o cámaras plenópticas) pero son extremadamente beneficiosos para aumentar la línea de referencia y, por tanto, beneficiosos para calcular grandes distancias en el mundo objeto con una fiabilidad muy elevada. Estos movimientos anormalmente grandes pueden producirse artificialmente por ejemplo grabando los fotogramas que se producen cuando el usuario está comenzando a alejar el teléfono, o por alguien que golpea accidentalmente el brazo de la persona que toma la fotografía, o por las grandes vibraciones de un palo para realizar autofotos (lo que obviamente produce movimientos mayores que en la figura 14); y se modelan mejor por un dispositivo novedoso que también forma parte de esta divulgación: un dispositivo 5200 plenóptico estéreo (figuras 5A y 5B) que incluye al menos una cámara 100 plenóptica y al menos una cámara convencional, pero en una realización preferida mostrada en la figura 5B se añadieron a la cámara 100 plenóptica dos cámaras convencionales o cámaras 1304 plenópticas (o una cámara convencional y una cámara plenóptica). Los prototipos de este dispositivo han demostrado la evidencia de que el dispositivo tiene utilidad por sí mismo (por ejemplo, en un teléfono móvil tal como en la figura 5B) y también para modelar movimientos de cámara anormalmente grandes que no pueden modelarse por una

cámara plenóptica o un sistema SAII, movimientos que son especialmente satisfactorios para calcular largas distancias con respecto a objetos muy distantes en el mundo. También merece la pena observar que los temblores de mano tal como en la figura 14 son habituales cuando el usuario está intentando sostener la cámara en su sitio tan fija como sea posible, sin embargo, los movimientos en el instante tras presionar el obturador son mucho mayores, aun así, beneficiosos, porque siguen estando orientados al mismos FOV, campo de visión, pero la línea de referencia puede aumentar varios centímetros, produciendo unas estimaciones de distancia mucho mejores. También, la distribución estadística del movimientos en las direcciones x e y en la figura 14 muestra, habitualmente, una gran relación pico con respecto a promedio (la mayor parte del tiempo el movimiento es de milímetros, pero cada cierto tiempo se produce una muestra o varias muestras que se mueven hacia arriba varios centímetros), lo que resulta beneficioso para mejorar la línea de referencia y se modela mejor a través de un dispositivo plenóptico estéreo tal como en la figura 5B dado que en este caso las imágenes epipolares vertical y/o horizontal tienen grandes huecos entre las varias hileras (imágenes captadas) tal como en las figuras 6A y 6B.

La realización en la figura 5B es una combinación novedosa de dos de las tecnologías recientemente mencionadas (plenóptica y estéreo) para crear un mapa de profundidad, que va mucho más allá que la técnica anterior (dado que incluye cámaras plenópticas mezcladas con cámaras convencionales o con otras cámaras plenópticas en una configuración multivista: un superconjunto que puede incluir más cámaras que en la figura 5B). La figura 5A muestra una configuración básica de un dispositivo plenóptico estéreo, un sistema multivista que mejora significativamente la precisión de estimación de profundidad de cámaras plenópticas para grandes distancias debido a la adición de una cámara convencional orientada hacia el mismo FOV (campo de visión) que la cámara plenóptica. Esta invención y sus métodos para estimación de profundidad en tiempo real están compuestos por al menos una cámara plenóptica de campo claro e incluyen cámaras plenópticas o convencionales adicionales. Tal sistema multivista, con los métodos de procesado de imagen apropiados, puede crear un mapa de profundidad de la escena con una resolución de muy alta calidad, superando las desventajas de las cámaras plenópticas (limitadas por mediciones de profundidad poco fiables para grandes profundidades) y de los sistemas multicámara (que necesitan mucha más energía de procesado). Esta invención de múltiples perspectivas es, al mismo tiempo, extremadamente eficaz en cuanto los requisitos computacionales. La figura 6A muestra una grabación de un dispositivo plenóptico (a la izquierda) en el que una línea 62 epipolar dentro de y una imagen epipolar del dispositivo plenóptico se combina con la imagen resultante de una cámara convencional (lado derecho) que tiene mucha más

resolución. Esta figura 6A también muestra cómo un punto de la cámara 61 convencional tal como, por ejemplo, la cámara inferior de la figura 5A (o las cámaras derecha o superior de la figura 5B) se usa para extender la línea de referencia de la cámara plenóptica con una imagen de una cámara convencional (tal como, por ejemplo, la cámara inferior de la figura 5A o la cámara 1304 en la figura 5B), lo que produce mejores capacidades y rendimientos de estimación de distancia para la combinación de ambas cámaras que la cámara plenóptica por sí misma. Una de las principales ventajas de esta realización es el uso de algoritmos plenópticos para la estimación de profundidad (mucho más eficaz computacionalmente que emparejamiento estéreo), que también se usan en la presente divulgación tal como se describe a continuación. Una ventaja adicional de este enfoque es que la resolución lateral del sistema multivista puede ser una resolución lateral de la cámara convencional (habitualmente mucho mayor que la resolución lateral de cámaras plenópticas), y que es posible calcular campos claros con tantos puntos como puntos haya en la(s) cámara(s) convencional(es).

La figura 6B es una realización de un método de cómo se realiza la “rectificación” de la(s) cámara(s) 1304 convencional(es) para emparejar sus imágenes con la cámara 100 plenóptica: se detecta una línea 1404 epipolar dentro de una imagen 400 epipolar de la cámara 100 plenóptica; la distancia B entre la vista 1516 central de la cámara 100 plenóptica y la(s) cámara(s) 1304 convencional(es) es evidente a partir de la cámara 100 plenóptica en la figura 5A y 5B, se obtiene basándose en la relación entre la “línea de referencia ancha” B entre la cámara 100 plenóptica y la(s) cámara(s) 1304 convencional(es) y las “líneas de referencia estrechas” *d* de la cámara 10 plenóptica en las figuras 5A, 5B y 6D; la distancia H se elige para coincidir con la parte común de los FOV, campos de visión, de la cámara 100 plenóptica y la(s) cámara(s) 1304 convencional(es); la línea 1404 epipolar de la cámara plenóptica (un conjunto de pixeles conectados en la imagen 400 epipolar de la cámara 100 plenóptica, que por definición marca un borde del mundo objeto) se dibuja linealmente (1506) para alcanzar la intersección con la hilera de pixeles 1406 de la(s) cámara(s) convencional(es), la intersección del sensor de la cámara plenóptica en el pixel 1504, sin embargo, en muchos de los casos el pixel 1504 (muestreado por la(s) cámara(s) convencional(es)) no coincide con los “patrones de borde” muestreados por la cámara 100 plenóptica, razón por la que la zona 1512 de búsqueda en la(s) cámara(s) convencional(es) se define para encontrar finalmente el pixel 61 de la(s) cámara(s) 1304 convencional(es) que coincide con los bordes detectados por la cámara 100 plenóptica. A través de este método se aumenta el número de vistas 1510 de la cámara 100 plenóptica captadas por las cámaras 51 equivalentes de la cámara 100 plenóptica con vistas(s) adicional(es) de

cámara(s) convencional(es) situadas a distancia(s) de la cámara plenóptica mucho mayores (centímetros o incluso más) que la separación habitual entre vistas de la cámara plenóptica (aproximadamente décimas de milímetros), lo que potencia extremadamente la línea de referencia (de *d* a B) y por tanto la precisión de las mediciones de profundidad para largas distancias desde la(s) cámara(s). Esto puede resumirse con la ayuda de la figura 6B de la siguiente manera: la separación estrecha "d" entre las vistas 1510 de una cámara 100 plenóptica necesitaría grandes aumentos de profundidad de patrones en el mundo objeto para producir variaciones muy pequeñas de pendiente de líneas 1404 epipolares, sin embargo, añadiendo vista(s) 1406 adicional(es) procedentes de cámara(s) convencional(es) o de cámaras 1304 plenópticas adicionales es posible ajustar "pendientes 1508 de líneas epipolares extendidas" muy precisas, lo que ofrece una mayor precisión de mediciones de profundidad para largas distancias.

La figura 5B muestra una realización de un dispositivo de esta invención dentro de un dispositivo móvil: una cámara 100 plenóptica asociada con dos cámaras convencionales (o asociada con una cámara convencional y una cámara plenóptica, o asociada con dos cámaras 1304 plenópticas adicionales), una alineada horizontalmente y la otra alineada verticalmente con el fin de mejorarlas líneas de referencia en ambas direcciones (x e y), ahorrando al mismo tiempo los elevados requisitos computacionales del emparejamiento de imagen estéreo y multivista usando una zona 1512 de búsqueda pequeña (que puede ser uni o bidimensional). Resulta evidente para un experto en la técnica cómo modificar/ampliar este dispositivo para tener varias opciones diferentes: solo una cámara plenóptica y una cámara convencional, solo dos cámaras plenópticas, tres cámaras plenópticas, cualquier matriz de cámaras que incluya al menos una cámara plenóptica, etc.

La situación ilustrada en las figuras 6A y 6B (imagen/imágenes 63 captada(s) por una cámara plenóptica, e imagen/imágenes 64 de la misma situación captada(s) por una cámara convencional) es equivalente a una única cámara convencional que ha captado varias imágenes en posiciones ligeramente diferentes a pequeñas distancias una con respecto a otra y una imagen captada adicional por la misma cámara convencional en una posición bastante distante del resto. Tal como se muestra en las figuras 6A y 6B la imagen epipolar formada tiene huecos *d* entre las imágenes captada (en donde *d* en una cámara plenóptica es el tamaño de la pupila de entrada dividida entre el número de pixeles por microlente en una dimensión [x o y]). Si el hueco B (entre la vista central de una cámara plenóptica virtual y la vista equivalente de la cámara 1304 convencional simulada por una cámara en movimiento) es mayor que la distancia D entre la vista central de la cámara plenóptica (o una cámara plenóptica virtual simulada por una

cámara en movimiento) y la vista de extremo de dicha cámara plenóptica (es decir cuatro tiempos *d* en los ejemplos de la figura 6B) es posible crear un sistema equivalente estéreo plenóptico virtual. Los criterios principales para crear o bien un SAII equivalente (o un sistema plenóptico equivalente) o un sistema plenóptico estereoscópico equivalente con una línea de referencia más amplia es tener al menos una gran línea de referencia (es decir entre las distancias entre imágenes adyacentes) que es mayor que *d*, si la línea de referencia es menor que la distancia *d* un sistema SAII equivalente se recomienda. También, un sistema SAII equivalente se seleccionará si la línea de referencia B es menor que la distancia D. Lo que debe observarse es si en las imágenes epipolares de la figura 6B existe al menos un gran hueco (B-D) (mayor que los huecos pequeños *d*), que requiera definir una región 1512 de búsqueda y encontrar el punto 61 de borde correspondiente. Por otro lado, en el caso que todas las líneas de referencia sean iguales o más pequeñas que *d*, las hileras de imágenes epipolares están en contacto de manera que se evitan los algoritmos de correspondencia (entre las diferentes hileras de imágenes epipolares) y se aplican algoritmos plenópticos habituales.

Obsérvese que en un dispositivo tal como en la figura 5B el número de microlentes en una cámara plenóptica es habitualmente más pequeño que el número de pixeles en la cámara convencional asociada, sin embargo, en la invención en donde las vistas 1510 plenópticas son vistas diferentes extraídas de una cámara en movimiento el número de pixeles de las vistas 1510 es igual al número de pixeles de la cámara equivalente en una línea B de referencia.

En una realización, la forma de determinar si el sistema equivalente para crear es un sistema plenóptico virtual (que también puede modelarse mediante un sistema SAII virtual) o un sistema plenóptico-estereoscópico virtual depende directamente de la mayor distancia entre imágenes captadas consecutivas (consecutivas en el dominio espacial o imágenes adyacentes), de manera que su mayor distancia es mayor que *d*, siendo *d* la distancia máxima entre “imágenes captadas elegidas” de una cámara plenóptica virtual que garantiza que la disparidad máxima entre dichas “imágenes captadas elegidas” es un pixel.

Las imágenes captadas se clasifican conectando en las dimensiones x e y cada una de estas imágenes con sus imágenes adyacentes formando una red. Si todas las distancias entre imágenes conectadas son iguales a o más pequeñas que *d* (disparidad más pequeña que un pixel) dichas imágenes pueden usarse para componer una SAII virtual (o del mismo modo, una cámara plenóptica virtual). Por otro lado, si se capta una o más imágenes a distancias en las direcciones x o y mayores que *d,* esas imágenes 64

pueden usarse para componer vistas 1406 adicionales de un sistema plenóptico-estereoscópico virtual tal como en las figuras 6A y 6B.

En una realización, con el fin de determinar qué imágenes entre todas las captadas son consecutivas entre sí, se usan las coordenadas *x* e *y* para crear una red como en la figura 13. Entonces, la "imagen consecutiva elegida" (en dominio espacial) de una imagen determinada es la que se ubica a la distancia mínima (en las direcciones x e y) desde dicha imagen determinada, pero siempre a una distancia más corta que *d*.

El procedimiento de "rectificación" descrito anteriormente para las cámaras convencionales frente a la cámara plenóptica, incluso aunque tenga sentido para el dispositivo en la figura 5B y dispositivos similares, es una simplificación excesiva de lo que ocurre cuando las cámaras 1304 no son cámaras físicas sino "cámaras virtuales" que disparan diferentes exposiciones desde diferentes puntos de vista desde una cámara en movimiento real. En la figura 6B se ha realizado una "rectificación" de imagen para la línea (B) de referencia y una "rectificación" H para hacer coincidir la parte común de los FOV de ambas cámaras (100 y 1304); si 1304 fuera una cámara virtual que se ha movido varias centímetros con movimientos mayores que milímetros debido a temblores humanos al tiempo que el usuario está intentando deliberadamente sostener la cámara tan quieto como sea posible (tal como en la figura 14), el procedimiento de "rectificación", en lugar de la línea B de referencia y el campo H de visión ha de considerar movimientos aleatorios en los 6 ejes (x, y, z, guiñada, cabeceo y balanceo) que puedan determinarse teniendo en consideración que ese acelerómetro, el giroscopio, o cualquier otro dispositivo de colocación asociado con la cámara que grabó la nueva posición (x´, y´, z´, guiñada´, cabeceo´ y balanceo´) fue la cámara 1304 virtual que captó la imagen una determinada una cantidad de tiempo después de que la cámara 100 plenóptica captó la primera imagen. En una realización diferente, la cámara 100 no es una cámara física sino una "cámara plenóptica virtual" (o sistema SAII virtual) que capta varios disparos (tal como en la figura 13: disparos A, B, C, D) debido a temblores de mano tal como en la figura 14.

La figura 16A muestra un primer procedimiento (1600) relacionado con obtención de imágenes de cámara estéreo. Esta figura muestra un procedimiento 1600 simplificado que asume una posición fija (conocida) de cámaras estéreo (que se conoce a partir del procedimiento de calibración de las dos cámaras estéreo). Este procedimiento comprende rectificación (1604) de imagen que es simple teniendo en cuenta la posición conocida (y la orientación) de las dos cámaras y una segunda etapa de correspondencia (1606) que implica hacer coincidir los patrones que son comunes a las dos imágenes adquiridas, obviamente el procedimiento de compatibilidad entre pixeles de las dos cámaras es diferente dependiendo de las distancias del mundo objeto de las fuentes de

luz que produjeron los patrones en ambas cámaras, o dicho de otro modo, un punto objeto en el mundo muy alejado de ambas cámaras producirá prácticamente cero disparidad entre sus dos imágenes en las dos cámaras, mientras que un punto objeto muy próximo a las cámaras producirá una disparidad muy grande en los sensores de las dos cámaras.

Un segundo procedimiento (1610) según la presente invención se describe en la figura 16B. Este procedimiento comprende: una etapa 1614 que graba fotogramas consecutivos (por ejemplo en 120 fps-fotogramas por segundo), simultáneamente graba la posición de 6 ejes de la cámara (x, y, z, P, R, Y) para cada uno de los "fotogramas grabados" (a 120 fotogramas por segundo y, por ejemplo, graba la posición de 6 ejes muestreada aproximadamente 5 posiciones por fotograma o 600 muestras por segundo); la siguiente etapa 1616 elige las posiciones con grandes líneas de referencia *d* de manera adecuada (como por ejemplo las posiciones A, B, C y D en la figura 13, posiciones que pueden ser diferentes para un "tirador de pistola olímpico" que para una persona que padece Parkinson) para componer un "sistema SAII virtual" (o una cámara plenóptica virtual) y en caso de que existan, también posiciones con "mayores líneas de referencia"-D de manera adecuada para componer un "sistema plenóptico estéreo virtual"; una tercera etapa 1618 rectifica los disparos o fotogramas elegidos tal como en las figuras 8 y 9 pero la rectificación depende de las posiciones de 6 ejes de la cámara (diferentes valores de x, y, z, cabeceo, balanceo y guiñada para cada uno de los disparos elegidos en la etapa 1616); una cuarta etapa 1620 crea el sistema SAII equivalente (o la cámara plenóptica equivalente) para los disparos que se han elegido y/o si algunos de los desplazamientos en las direcciones x y/o y son anormalmente grandes un sistema plenóptico-estereoscópico equivalente tal como en las figuras 5A-5B (pero de manera más probable con bastantes valores diferentes de z, cabeceo, balanceo y guiñada para la cámara equivalente con la "línea de referencia ancha" B, a medida que las cámaras en las figuras 5A y 5B se alinean y son coplanares, lo que no suele ser el caso con cámara(s) en movimiento). Una vez se crea el sistema equivalente (en la etapa 1620 de la figura 16B) es posible realizar una quinta etapa (1622) adicional destinada a calcular distancias hacia los objetos en el mundo tras el análisis de pendiente de línea epipolar tradicional (tal como en las figuras 3A-3C), o el análisis de línea epipolar ampliada (tal como en las figuras 6A y 6B) si las líneas de referencia son lo suficientemente grandes (al menos una de las imágenes está a una distancia en las direcciones x y/o y mayor que *d* con respecto al "conjunto de imágenes conectadas" [en donde la distancia de cada imagen dentro del "conjunto conectado" es igual a o menor que *d* desde sus imágenes más próximas dentro del "conjunto de imágenes conectadas"]), obteniendo una mapa de pendiente de las imágenes del FOV común,

campo de visión, de todas las cámaras del “sistema equivalente” de la etapa 1620. La pendiente de las líneas epipolares obtenidas anteriormente puede usarse además para obtener un mapa de profundidad a través de pendiente epipolar tradicional con respecto a conversiones de profundidad (etapa 1624), obteniendo un mapa de profundidad de las imágenes del FOV común, campo de visión, de todas las cámaras del “sistema equivalente” de la etapa 1620. Es posible crear imágenes 3D (etapa 1626) a partir de la pendiente y los mapas de profundidad a partir de imágenes 3D previamente calculadas que cumplen con cualquier formato 3D (imágenes estéreo, imágenes integrales, etc.)

La robustez del procedimiento propuesto se ha demostrado experimentalmente con diferentes usuarios y dispositivos y en tiempos diferentes del día. Además, toda la experimentación se ha repetido varias veces para evitar la aleatoriedad del procedimiento.

En una realización particular, la entrada de imágenes de la invención puede ser una secuencia de video: se supone una secuencia de video que se capta a 120 fps y se desea que la invención use 4 fotogramas (4 imágenes) para calcular los valores de profundidad de la escena. Esto significará que el sistema producirá mapas de profundidad (o imágenes 3D) a aproximadamente 30 fps (considerado por muchos como tiempo real). Los fotogramas seleccionados para calcular el mapa de profundidad (o para componer una imagen3D) son las que muestran una línea de referencia lo suficientemente ancha, no necesariamente fotogramas consecutivos.

Hasta el momento, se ha descrito el procedimiento de “registro” de dos o más imágenes tomadas por un dispositivo móvil usando los datos procedentes del acelerómetro, el giroscopio, o cualquier otro dispositivo de colocación. Debe recordarse que el procedimiento de registro implica “rectificación” de imagen (para garantizar que las 2 o más imágenes adquiridas se “recalculan” para volverse imágenes coplanarias comparables, tal como en las figuras 8 y 9) y la “correspondencia” o “compatibilidad de patrón” (mostrada a modo de ejemplo en la figura 8 buscando el patrón 86 común). La “correspondencia” o “compatibilidad de patrón” en SAII, en cámaras plenópticas y en una realización de esta invención se realiza identificando las líneas epipolares en las imágenes epipolares).

En otra realización, el procedimiento puede realizarse dentro de un intervalo de tiempo en el que el procedimiento descrito puede considerarse como un procedimiento en tiempo real.

Los movimientos registrados por el dispositivo móvil son lo suficientemente buenos para obtener un mapa de profundidad robusto. Para ello, se comparará de nuevo la línea de referencia obtenida mediante las subaberturas de una cámara plenóptica con la línea de referencia obtenida mediante la invención propuesta.

La línea de referencia de una cámara plenóptica es la distancia entre los centros de dos subaberturas consecutivas (la distancia *d* entre los centros de dos cámaras equivalentes 51 en la figura 5B), y el tamaño de la línea de referencia (así como el diámetro 2D máximo) está directamente relacionado con la distancia máxima al mundo objeto que puede estimar el dispositivo con precisiones aceptables; cuanto mayor sea la línea de referencia y el diámetro (*d* y D) mejor será el mapa de profundidad (obteniendo mejores estimaciones de grandes distancias con respecto al mundo objeto). Tal como se comentó anteriormente, una décima de un milímetro puede considerarse una línea de referencia normal en una cámara plenóptica (una abertura típica de la pupila de entrada de 1 o 2 mm y un número típico de 10-20 pixeles por microlente). La invención propuesta puede funcionar de manera similar a un sistema SAII (o una cámara plenóptica) pero solo con una cámara convencional tomando vistas secuenciales. La invención propuesta puede usar los mismos algoritmos basándose en el cálculo de pendientes a partir de las imágenes epipolares tal como una cámara plenóptica (o como un sistema SAII) para estimar un mapa de profundidad. Sin embargo, la invención puede funcionar con líneas de referencia mayores que la línea de referencia de una cámara plenóptica (aproximadamente 0,1 mm), dado que temblores de mano son normalmente mayores que esos, por tanto, la invención propuesta puede obtener mapas de profundidad de mayor calidad en cuanto una precisión para distancias mayores. Además de esta importante ventaja, es incluso más importante señalar que la invención propuesta puede obtener mapas de profundidad con mucha mayor resolución espacial que los obtenidos mediante una cámara plenóptica dado que el sistema tiene la totalidad de la resolución del sensor de cámara convencional, resolviendo la desventaja principal de las cámaras plenópticas (que tienen la misma resolución espacial pequeña que las microlentes, y con aproximadamente 100 pixeles por microlente cuadrada su resolución es aproximadamente 100 veces menor).

En una realización, el movimiento del dispositivo móvil debido a temblores de mano puede reforzarse o sustituirse por la vibración producida por un pequeño motor de vibración incluido en el dispositivo móvil (que pueden ser las vibraciones usadas como sustituto o complemento para los tonos de llamada) o colocando la cámara sobre un objeto en movimiento durante el tiempo de exposición (por ejemplo la cámara está montada o colocada en una ubicación de automóvil con amplia visibilidad a zonas de interés fuera al automóvil).

En otra realización, los métodos plenóptico y estéreo plenóptico descritos en el presente documento para resolver el problema de correspondencia usando el acelerómetro, el giroscopio, o cualquier otro dispositivo de colocación pueden sustituirse por algoritmos que hagan coincidir diferentes imágenes (compatibilidad estéreo o compatibilidad

multivista). En todavía otra realización, los objetos en el primer plano pueden identificarse, mientras que en una secuencia de video compuesta el fondo puede moverse frente a los objetos en primer plano quietos (o moviéndose los objetos en primer plano a una velocidad más reducida) creando efectos 3D combinando imágenes bidimensionales a diferentes distancias de la cámara: cuando en una secuencia de video las oclusiones de imagen de fondo cambian con tiempo, cuando los objetos en primer plano se mueven a una velocidad más reducida que los movimientos más rápidos en el fondo, cuando la perspectiva o el tamaño de los objetos en primer plano está cambiando lentamente (por ejemplo un tiburón nadando hacia la cámara, y ocluyendo cada vez más el fondo en los fotogramas sucesivos; o un tiburón nadando a una distancia constante del plano de cámara a lo largo del FOV, cambiando las oclusiones en fotogramas de video sucesivos); o justo lo contrario, el primer plano se mueve más rápido que el fondo y cambia las oclusiones. Como ejemplo, pero no de manera exclusiva, en los casos mencionados anteriormente, una combinación de secuencias de video de varios niveles diferentes de primer plano en 2D, planos medios e imágenes de nivel de fondo ubicadas a varias distancias diferentes de la cámara (niveles que pueden relacionarse con sus distancias calculadas debido a las técnicas mencionadas en esta divulgación permiten el cálculo de mapas de profundidad en tiempo real de imágenes de video), permiten una combinación de dos o más imágenes bidimensionales para producir una percepción 3D al observador.

Puede crearse un campo claro de muchas maneras, por ejemplo, con sistemas SAII que incluyen una matriz de cámaras o, de manera equivalente, una cámara que se mueve automáticamente para tomar imágenes de la escena a partir de ubicaciones bien definidas. Un campo claro también puede crearse usando una cámara plenóptica. La invención propuesta en el presente documento se implementa en un dispositivo móvil que adquiere varias imágenes dentro de un intervalo de tiempo y entonces rectifica estas imágenes usando datos procedentes del acelerómetro, el giroscopio o cualquier otra capacidad de este tipo integrado en el dispositivo, tal como se describió anteriormente. Este procedimiento también compone un campo claro de la escena. Varias realizaciones de procedimientos de procesamiento para producir un mapa de profundidad de una escena a partir de este campo claro se describen a continuación en detalle.

Una manera de obtener la información de profundidad de una escena a partir de un campo claro es analizar los patrones captados por el sensor en las imágenes epipolares. En la invención propuesta cada una de las imágenes adquiridas (rectificadas de manera conveniente) se trata como una vista plenóptica, y cada vista plenóptica se usa para crear las imágenes epipolares. Las figuras 3A-3B-3C muestran cómo se componen imágenes 300 y 302 epipolares horizontal y vertical a partir de un campo claro, y dentro

de esas imágenes es posible identificar pixeles conectados que forman líneas, las denominadas líneas epipolares. Todos los pixeles iluminados de líneas 62 epipolares corresponden al mismo punto en el mundo objeto. Adicionalmente, las pendientes de estas líneas están directamente relacionadas con el tamaño del patrón iluminado sobre las microlentes y a la profundidad correspondiendo del punto en el mundo objeto. Por tanto, al conocer este patrón es posible rastrear inversamente los patrones muestreados por lo pixeles a través de la cámara y obtener la profundidad exacta del punto en el mundo objeto que produjo tal patrón. Se conoce bien que en una cámara plenóptica la relación entre profundidad y pendiente depende de las dimensiones físicas y el diseño del dispositivo usado para captar el campo claro. En esta invención, la formación de los patrones en las imágenes epipolares depende del desplazamiento (línea de referencia) entre las diferentes imágenes adquiridas (vistas diferentes). Este desplazamiento también puede calcularse usando algoritmos de correspondencia (algoritmos de compatibilidad estéreo). Estos algoritmos buscan patrones que puedan aparecer en dos o más imágenes con el fin de establecer una relación de uno con respecto a uno entre los pixeles de dichas dos o más imágenes. Estos son algoritmos de cálculo intenso que pueden evitarse usando la presente invención. En la presente invención, el desplazamiento entre imágenes se calcula usando los datos del acelerómetro, giroscopio o cualquier otra capacidad de este tipo integrada en el dispositivo. Esto implica cálculos de movimientos de rotación y traslación continuos que tras “el procedimiento de rectificación de imagen” terminan con una relación de uno con respecto a uno entre los pixeles de ambas imágenes.

Objetos a diferentes profundidades o distancias de la cámara producirán diferentes patrones de iluminación sobre el sensor de una cámara plenóptica, así como sobre la composición propuesta de imágenes tomadas por una cámara en movimiento en un dispositivo móvil. Tal como ya se mencionó, de la misma manera que en una cámara plenóptica las denominadas vistas plenópticas (que componen un campo claro) pueden representarse en imágenes epipolares, en la presente invención las diversas “vistas rectificadas” que pueden obtenerse secuencialmente a partir de una única cámara en movimiento (que también compone un campo claro) también pueden representarse por imágenes epipolares, en ambos casos, las imágenes epipolares se componen tomando dos segmentos dimensionales del campo claro tal como se explica en la figura 3.

En una realización, los algoritmos plenópticos usados en esta invención para la estimación de profundidad pueden aplicar una técnica de regresión lineal a los puntos que forman una línea epipolar para obtener la pendiente de dicha línea epipolar. Cuando se analiza una línea epipolar en una imagen epipolar horizontal/vertical, todas las imágenes (tal como ocurre con las vistas plenópticas) distribuidas a lo largo de la

dimensión vertical/horizontal se tienen en consideración dado que el mismo punto objeto se ha captado por varias de estas vistas y las líneas epipolares producidas por el mismo punto en el mundo pueden aparecer en varias imágenes epipolares. Por tanto, esta técnica de regresión lineal y el uso de diferentes imágenes epipolares para calcular las distancias al mismo punto en el mundo objeto reducen el ruido estadístico beneficiándose de información redundante a lo largo de una dimensión.

En todavía otra realización, todas las líneas formadas en las imágenes epipolares horizontal y vertical se identifican y sus pendientes correspondientes se calculan. Entonces, la profundidad correspondiente del objeto se calcula a partir de la pendiente.

En otra realización, solo se calcula un valor de pendiente (y/o profundidad) por línea epipolar dado que se forma una línea epipolar por el mismo punto objeto captado a partir de varios puntos de vista. Por tanto, la cantidad de datos se reduce drásticamente debido a dos factores: (i) solo se detectan líneas en las líneas epipolares correspondientes a bordes en el mundo objeto (dado que las zonas del mundo objeto completamente uniformes, sin bordes, no producen ninguna línea epipolar) y, (ii) es posible calcular/almacenar solo un valor de pendiente por línea en lugar de calcular/almacenar un valor por cada pixel que forma la línea epipolar, tal como se realizó tradicionalmente en la técnica anterior. En al menos una realización, el resultado de este procedimiento de cálculo puede ser simplemente los valores de profundidad correspondientes de estas pendientes detectadas.

En otra realización posible, las pendientes obtenidas analizando las líneas epipolares horizontal y vertical se combinan en una matriz de múltiples dimensiones para reducir el ruido estadístico. Esta redundancia mejora el resultado de la invención dado que se tiene en consideración el mismo pixel de sensor cuando se analizan ambas, las imágenes epipolar vertical y horizontal y, por tanto, se producen varios valores de pendiente por el mismo punto del mundo objeto.

Las pendientes calculadas para las líneas epipolares se transforman en las profundidades de objeto correspondientes. En otra realización, esa etapa de transformación puede realizarse tras combinar todas las pendientes redundantes, reduciendo drásticamente el número de transformaciones de pendiente en profundidad.

En otra realización, las profundidades/pendientes calculadas en las líneas epipolares horizontal y vertical se combinan directamente en un mapa de profundidad/pendiente disperso bidimensional (disperso porque incluye cálculos de profundidad/pendiente solo para los puntos en las líneas epipolares, y no para cada punto en la imagen tal como en la técnica anterior), realizando por tanto una única etapa de combinación, que aumenta la eficacia computacional.

En otra realización, el mapa de profundidad/pendiente disperso puede rellenarse aplicando técnicas de relleno de imagen para obtener valores de profundidad/pendiente para cada pixel. Como resultado, la invención proporciona un mapa de profundidad denso en el que cada punto se asocia con la estimación de profundidad de ese punto en la escena.

En otra realización, los métodos descritos en el presente documento para estimar un mapa de profundidad pueden combinarse con o sustituirse por algoritmos estéreo de compatibilidad o algoritmos de compatibilidad multivista para mejorar el resultado final.

En al menos una realización, los métodos descritos en el presente documento pueden implementarse en dispositivos móviles equipados con una cámara plenóptica.

En una realización, las líneas epipolares pueden detectarse usando algoritmos de detección de borde y sus pendientes pueden medirse mediante técnicas de regresión lineal (ambas metodologías, detección de bordes y regresión lineal, pueden usarse con precisión subpixel).

En una realización, para la estimación de profundidad, todos los cálculos pueden realizarse solo para aquellos pixeles de los sensores en los que se han detectado los bordes del mundo objeto, evitando realizar cálculos en un gran número de pixeles de los sensores.

La disipación de energía en terminales móviles (que dependen de baterías) es extremadamente importante, razón por la que la eficacia de cálculo en algoritmos adquiere una importancia primordial. Es de dominio público que algunos teléfonos 3D (que usan 2 cámaras) desactivan la segunda cámara (y la función 3D) en condiciones de batería baja. Estos ejemplos aclaran que para obtener mapas de profundidad en tiempo real en dispositivos móviles es conveniente implementar los algoritmos de una manera extremadamente eficaz. La presente invención permitirá que cámaras convencionales proporcionen imágenes 3D en dispositivos móviles (teléfonos móviles, tabletas …) usando algoritmos extremadamente eficaces para calcular la profundidad solo para los bordes identificados.

Para ello, es posible beneficiarse de los múltiples núcleos incluidos hoy en día en los procesadores (incluso en procesadores de dispositivos móviles). La idea esencial es crear varios hilos de ejecución de algoritmo de tal manera que cada uno de ellos esté a cargo de realizar diferentes operaciones. Por ejemplo, en la figura 15A se muestra un dispositivo 1000 móvil electrónico que incluye el presente sistema 1001 multivista, que capta imágenes 1002, que se tratan a través de un procesador 1004, que puede ser un procesador 1006 de múltiples núcleos. El procesador 1004 puede estar compuesto por dos o más CPU (unidades de procesamiento central) 1008a y 1008b (figura 15B).

Pueden usarse técnicas computacionales más avanzadas para aumentar la eficacia computacional. Por ejemplo, los procesadores 1004 actuales pueden incluir unidades 1010 de procesador gráfico (GPU), incluso aquellas GPU diseñadas para dispositivos 1010 móviles, incluyen varios cientos o miles de núcleos que puede ejecutar operaciones simultáneamente. Por consiguiente, en al menos una realización, cada imagen epipolar se procesa simultáneamente en un núcleo diferente de una GPU para acelerar adicionalmente la ejecución del algoritmo.

REIVINDICACIONES

1. Método para obtener información de profundidad a partir de una escena, que comprende las etapas de:

a) adquirir una pluralidad de imágenes de la escena por medio de al menos una cámara durante un tiempo de disparo en el que la pluralidad de imágenes ofrece al menos dos vistas diferentes de la escena;

b) para cada una de las imágenes de la etapa a), simultáneamente adquirir datos sobre la posición de las imágenes con respecto a un sistema de referencia de seis ejes;

c) seleccionar de las imágenes de la etapa b) al menos dos imágenes;

d) rectificar las imágenes seleccionadas en la etapa c) generando de ese modo un conjunto de imágenes rectificadas; y

e) generar un mapa de profundidad a partir de las imágenes rectificadas.

2. Método según la reivindicación 1, en el que la posición de las imágenes durante el tiempo de disparo se mide a partir de un conjunto de datos de colocación adquiridos por medio de al menos un dispositivo de colocación seleccionado del grupo de: un acelerómetro, un IMU, un AHRS, un GPS, un velocímetro y/o un giroscopio.

3. Método según la reivindicación 2, en el que el dispositivo de colocación está unido de manera rígida a al menos una cámara.

4. Método según cualquiera de las reivindicaciones anteriores en el que al menos una cámara se asocia con un dispositivo móvil.

5. Método según la reivindicación 4, en el que el dispositivo móvil es un teléfono inteligente, una tableta, un ordenador portátil o una cámara compacta.

6. Método según cualquiera de las reivindicaciones anteriores en el que, en la etapa c) las imágenes se seleccionan basándose en sus posiciones en el sistema de referencia de seis ejes.

7. Método según la reivindicación 6, en el que las imágenes se seleccionan de modo que sus distancias relativas con respecto a imágenes adyacentes (*d*) provoca una disparidad entre imágenes adyacentes de, como máximo, un pixel.

8. Método según la reivindicación 7, en el que la etapa e) comprende generar un sistema (16200) de obtención de imágenes integral de apertura sintética virtual como una red de imágenes rectificadas generando de ese modo un conjunto de imágenes epipolares

9. Método según la reivindicación 6 en el que las imágenes se seleccionan de modo que al menos una imagen es de manera que su distancia relativa a sus imágenes adyacentes provoca una disparidad de más de un pixel.

10. Método según la reivindicación 9, en el que la etapa e) comprende generar un sistema plenóptico-estereoscópico virtual con una red de imágenes rectificadas y otra imagen rectificada más distante generando de ese modo un conjunto de imágenes epipolares.

11. Método según las reivindicaciones 8 o 10, en el que la etapa e) comprende, además, calcular al menos una pendiente de al menos una línea epipolar a partir del conjunto de imágenes epipolares.

12. Método, según la reivindicación 11, en el que las líneas epipolares se calculan usando algoritmos de detección de borde a nivel subpixel.

13. Método, según la reivindicación 11, en el que las pendientes de línea epipolar se calculan usando algoritmos de regresión lineal a nivel subpixel.

14. Método según la reivindicación 11 en el que la etapa e) comprende, además, obtener un mapa de profundidad de la escena convirtiendo las pendientes de las líneas epipolares en profundidades.

15. Método según cualquiera de las reivindicaciones anteriores, en el que el método comprende, además, una etapa de generar una imagen tridimensional de la escena a partir del mapa de profundidad.

16. Método según cualquiera de las reivindicaciones anteriores en el que en la etapa a) al menos una cámara se mueve durante el tiempo de disparo.

17. Método, según la reivindicación 16, en el que los movimientos de al menos una cámara son movimientos aleatorios indeterminados producidos por temblores de mano de ser humano.

18. Método según la reivindicación 16 en el que al menos una cámara se une a una estructura que se mueve con respecto a la escena, en el que la estructura en movimiento se selecciona al menos de un automóvil, un teléfono inteligente, una tableta, un ordenador portátil o una cámara compacta.

19. Método según cualquiera de las reivindicaciones anteriores, en el que la pluralidad de imágenes de la etapa a) se adquieren mediante al menos dos cámaras.

20. Método, según la reivindicación 19, en el que las al menos dos cámaras se alinean y sus posiciones relativas se conocen.

21. Método, según las reivindicaciones 20 o 19, en el que al menos una de las cámaras es una cámara plenóptica.

22. Método según cualquiera de las reivindicaciones anteriores en el que una secuencia de video está compuesta por al menos dos niveles de profundidad de primer plano, planos medios opcionales e imágenes bidimensionales de fondo ubicadas a diferentes profundidades en el mundo objeto y en el que dicha combinación de diferentes niveles de imágenes bidimensionales en fotogramas sucesivos y/o el cambio

de oclusiones en imágenes bidimensionales más próximas al fondo y/o el cambio de perspectiva y tamaño en imágenes bidimensionales más próximas al primer plano produce una percepción en 3D al usuario.

23. Método según cualquiera de las reivindicaciones anteriores en el que solo algunas o todas las imágenes epipolares distribuidas a lo largo de las dimensiones vertical/horizontal se tienen en consideración con el fin de reducir ruido estadístico.

24. Método según cualquiera de las reivindicaciones anteriores en el que las pendientes obtenidas analizando las líneas epipolares horizontal y vertical se combinan en una matriz de múltiples dimensiones.

25. Método según cualquiera de las reivindicaciones anteriores, en el que los valores de las profundidades/pendientes calculadas en las líneas epipolares horizontal y/o vertical se combinan directamente en un mapa de profundidad/pendiente disperso bidimensional [página 7, líneas 23-29].

26. Método según cualquiera de las reivindicaciones anteriores, en el que el mapa de profundidad/pendiente disperso se rellena aplicando técnicas de relleno de imagen para obtener valores de profundidad/pendiente para cada pixel.

27. Método según cualquiera de las reivindicaciones anteriores, en el que, para la estimación de profundidad, los cálculos se realizan solo para aquellos pixeles de los sensores en los que se han detectado los bordes del mundo objeto [página 7, líneas 23-29].

28. Dispositivo para obtener información de profundidad a partir de una escena que comprende al menos una cámara, al menos un dispositivo de colocación y medios de procesado configurados para ejecutar un método según cualquiera de las reivindicaciones 1-27.

29. Dispositivo, según la reivindicación 28, en el que el dispositivo comprende al menos dos cámaras.

30. Dispositivo, según la reivindicación 29, en el que al menos una de las cámaras es una cámara plenóptica.

31. Dispositivo, según la reivindicación 28, en el que las cámaras están alineadas y sus posiciones relativas se conocen.

32. Dispositivo, según cualquiera de las reivindicaciones 28 a 31, en el que el dispositivo comprende al menos una tercera cámara y en el que una de las cámaras está alineada horizontalmente con la cámara plenóptica, y al menos una de las cámaras está alineada verticalmente con dicha cámara plenóptica.

33. Dispositivo según cualquier reivindicación 28 a 32, en el que el método para obtener información de profundidad comprende las etapas de:

a) adquirir una pluralidad de imágenes de la escena durante un tiempo de disparo en el que la pluralidad de imágenes ofrece al menos dos vistas diferentes de la escena a partir de al menos dos cámaras;

b) rectificar las imágenes de la etapa a) generando de ese modo un conjunto de imágenes rectificadas;

c) generar un mapa de profundidad a partir de las imágenes rectificadas.

34. Dispositivo según cualquiera de las reivindicaciones 28-33, en el que una secuencia de video se compone por al menos dos niveles de profundidad de primer plano, planos medios opcional e imágenes bidimensionales de fondo (ubicadas a diferentes profundidades en el mundo objeto) y en el que dicha combinación de diferentes niveles de imágenes bidimensionales en fotogramas sucesivos y/o el cambio de oclusiones en imágenes bidimensionales más próximas al fondo y/o el cambio de perspectiva y tamaño en imágenes bidimensionales más próximas al primer plano produce una percepción en 3D al usuario.

35. Dispositivo según cualquiera de las reivindicaciones 28-34, en el que solo algunas o todas las imágenes epipolares distribuidas a lo largo de las dimensiones vertical/horizontal se tienen en consideración con el fin de reducir el ruido estadístico.

36. Dispositivo según cualquiera de las reivindicaciones 28-35, en el que las pendientes obtenidas analizando las líneas epipolares horizontal y vertical se combinan en una matriz de múltiples dimensiones.

37. Dispositivo según cualquiera de las reivindicaciones 28-36, en el que los valores de las profundidades/pendientes calculadas en las líneas epipolares horizontal y/o vertical se combinan directamente en un mapa de profundidad/pendiente disperso bidimensional.

38. Dispositivo según cualquiera de las reivindicaciones 28-37, en el que el mapa de profundidad/pendiente disperso se rellena aplicando técnicas de relleno de imagen para obtener valores de profundidad/pendiente para cada pixel.

39. Dispositivo según cualquiera de las reivindicaciones 28-38, en el que para la estimación de profundidad, los cálculos se realizan solo para aquellos pixeles de los sensores en los que se han detectado los bordes del mundo objeto.

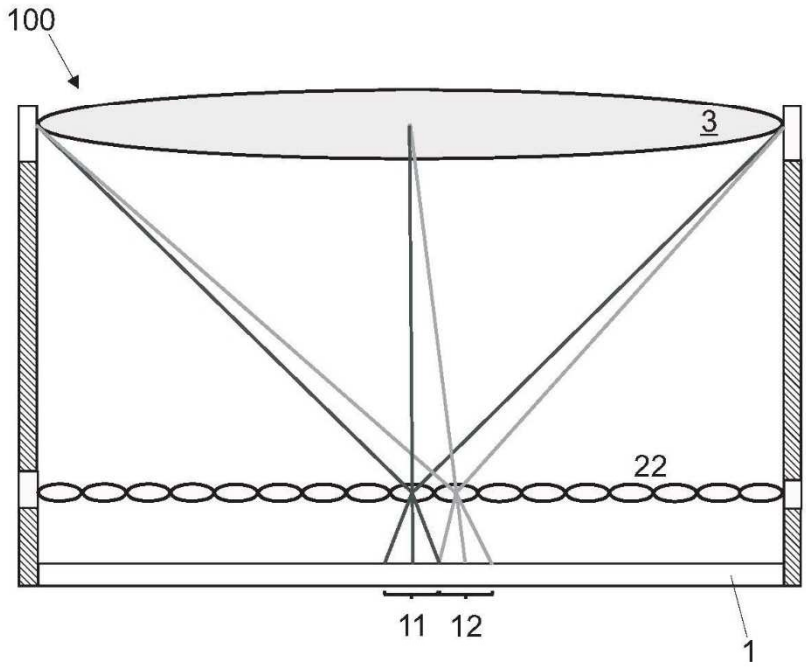

Fig. 1

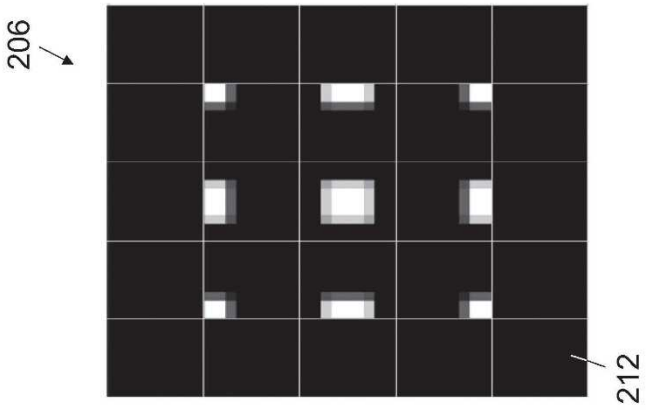

Fig. 2B

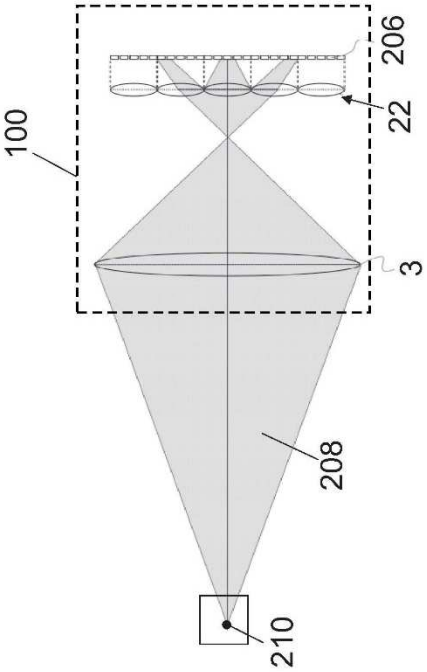

Fig. 2A

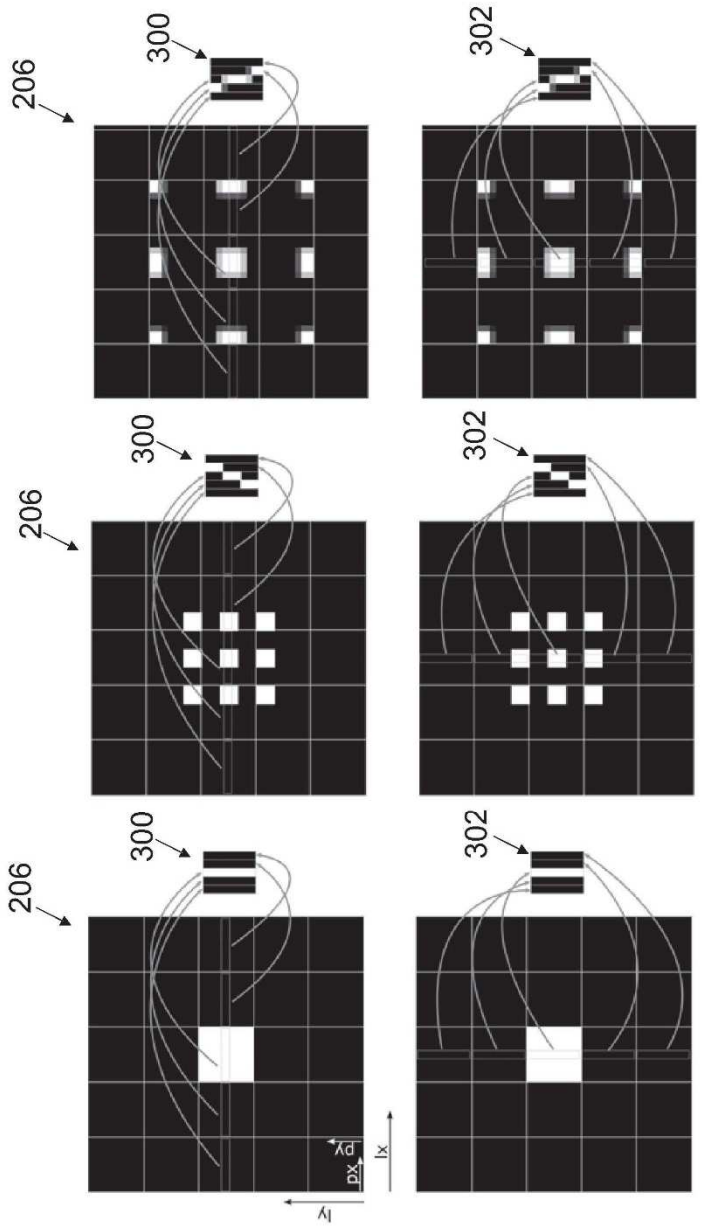

Fig. 3C

Fig. 3B

Fig. 3A

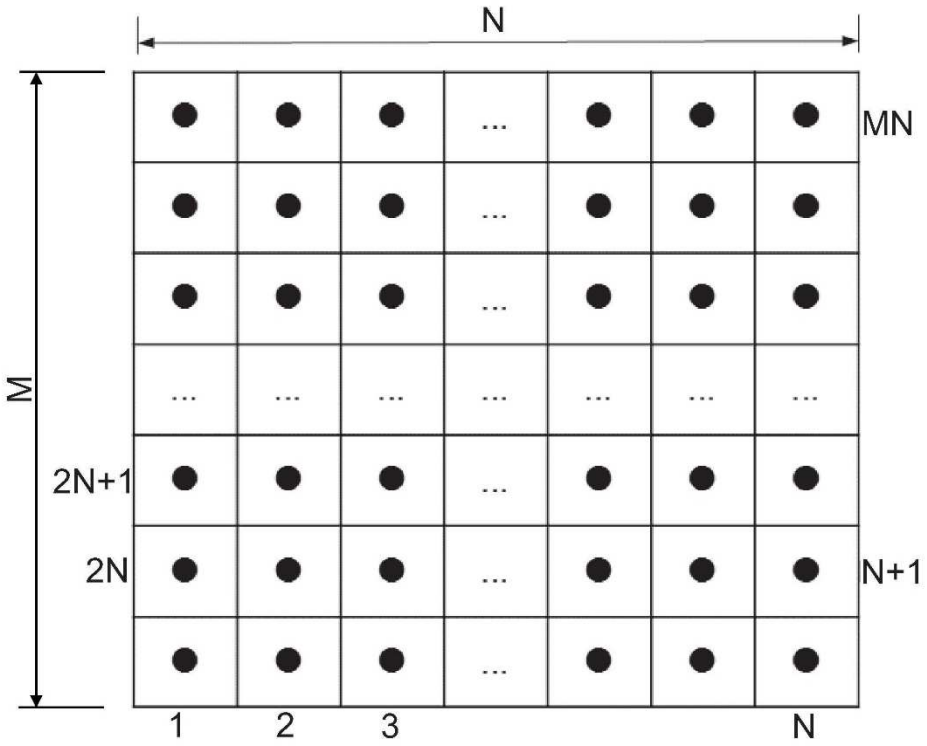

Fig. 4

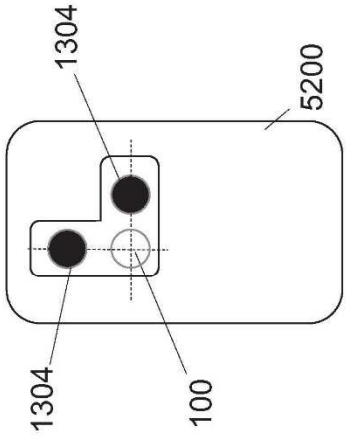

Fig. 5B

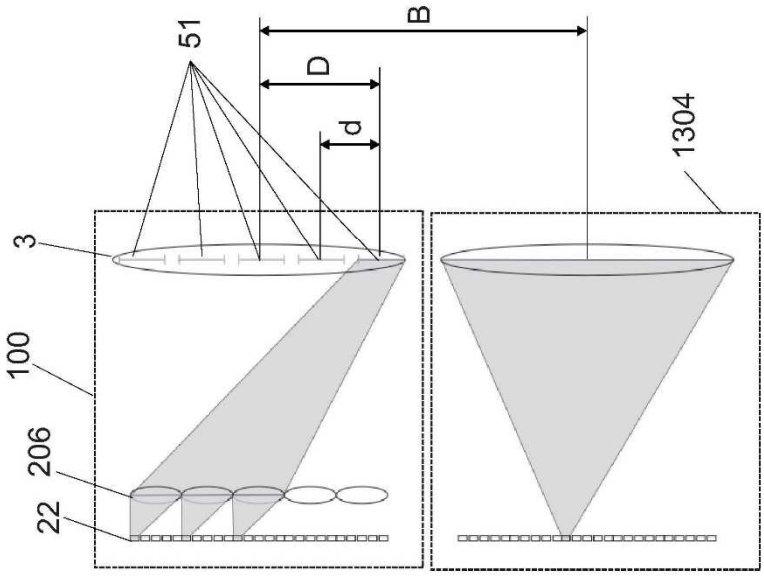

Fig. 5A

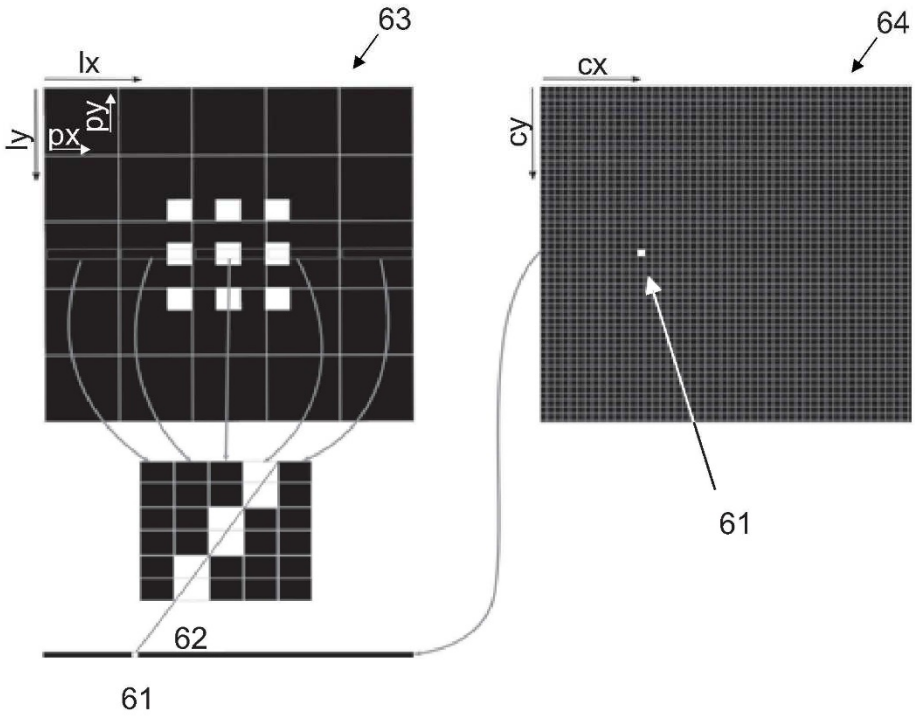

Fig. 6A

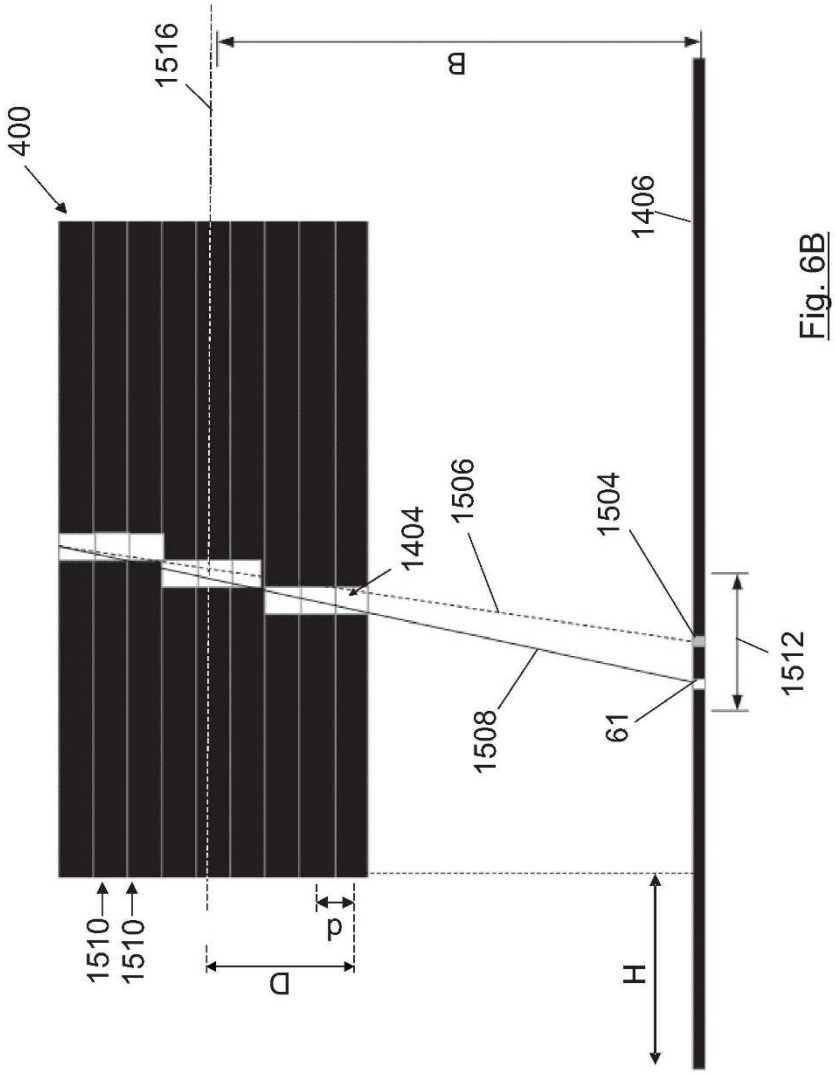

Fig. 6B

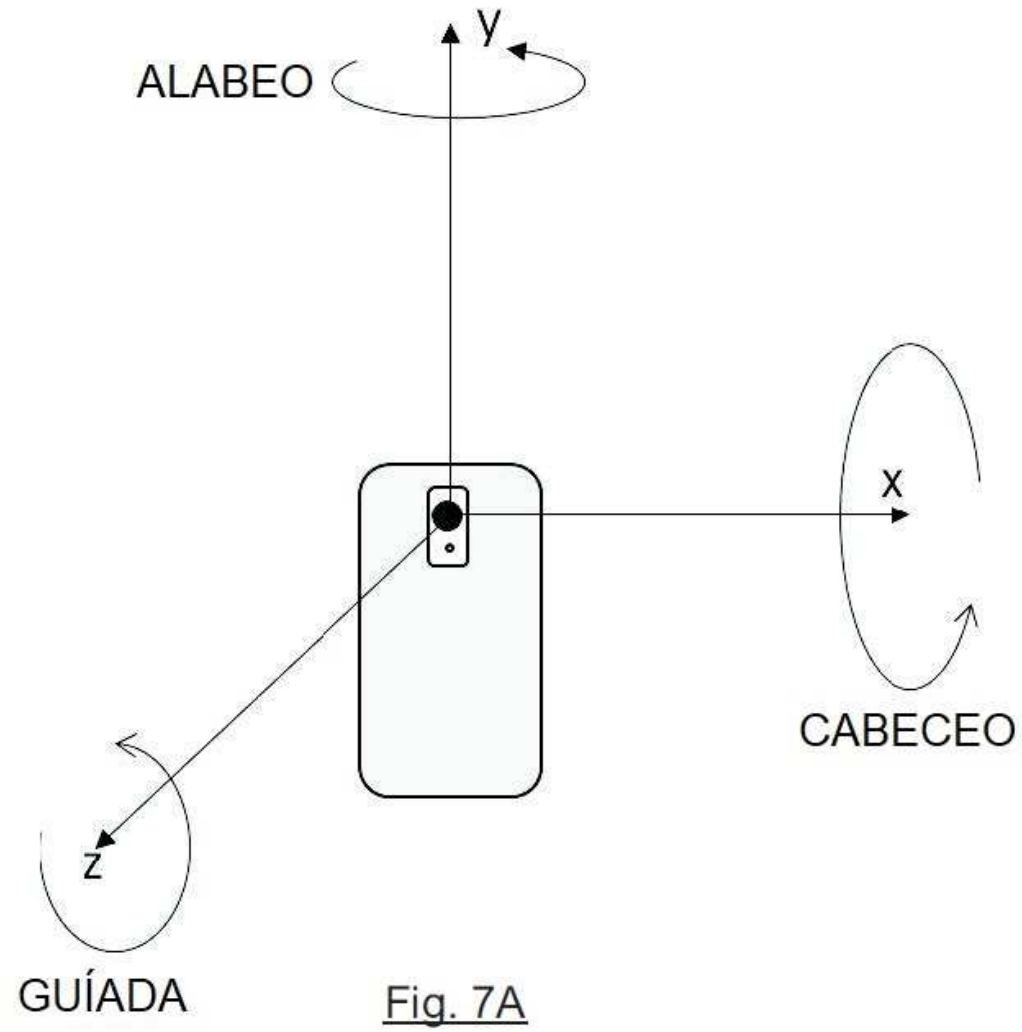

Fig. 7A

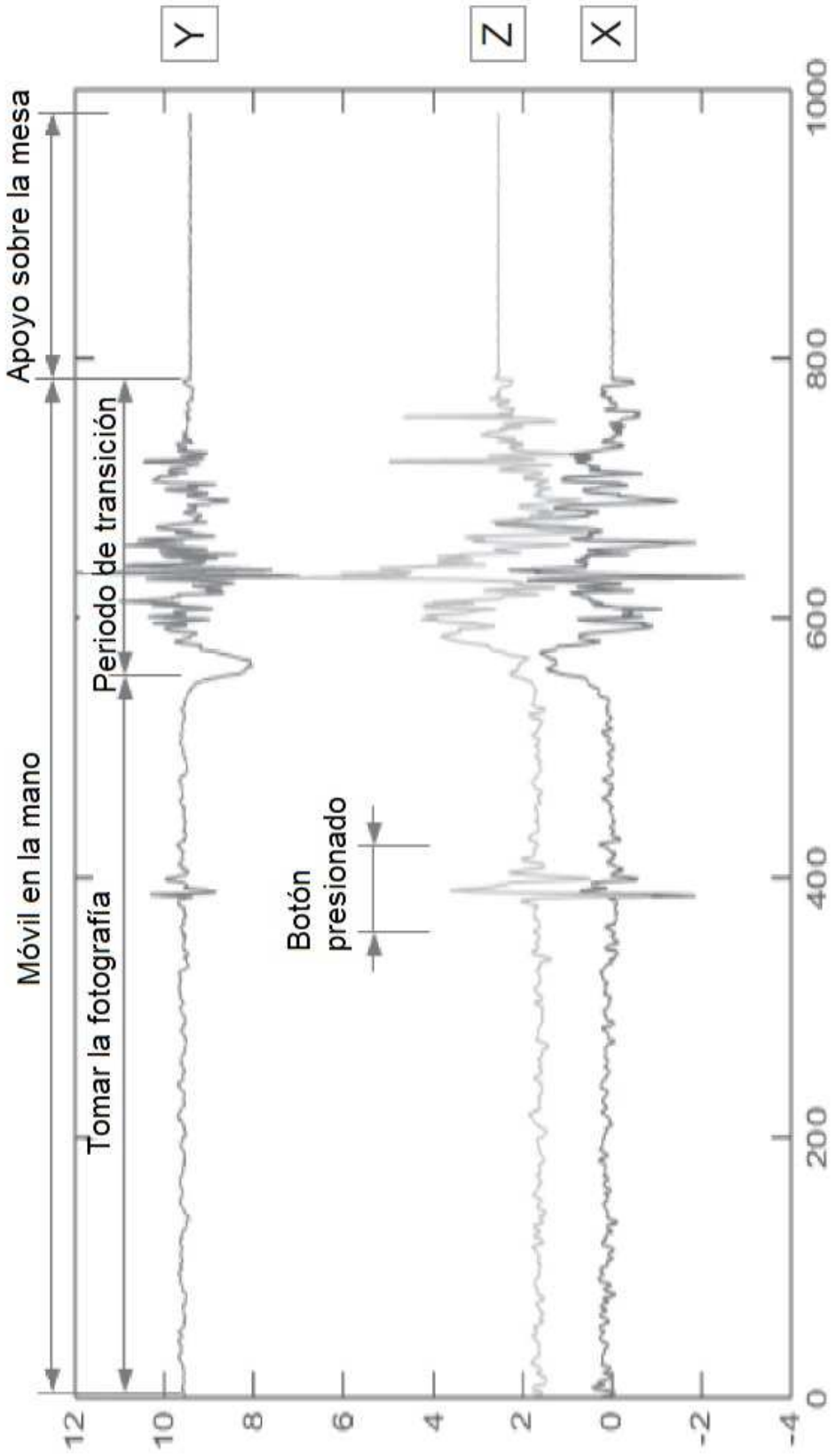
Y
Z
X
Apoyo sobre la mesa
Periodo de transición
Botón presionado
Móvil en la mano
Tomar la fotografía
12
10
8
6
4
2
0
-2
-4
0
200
400
600
800
1000

Fig. 7B

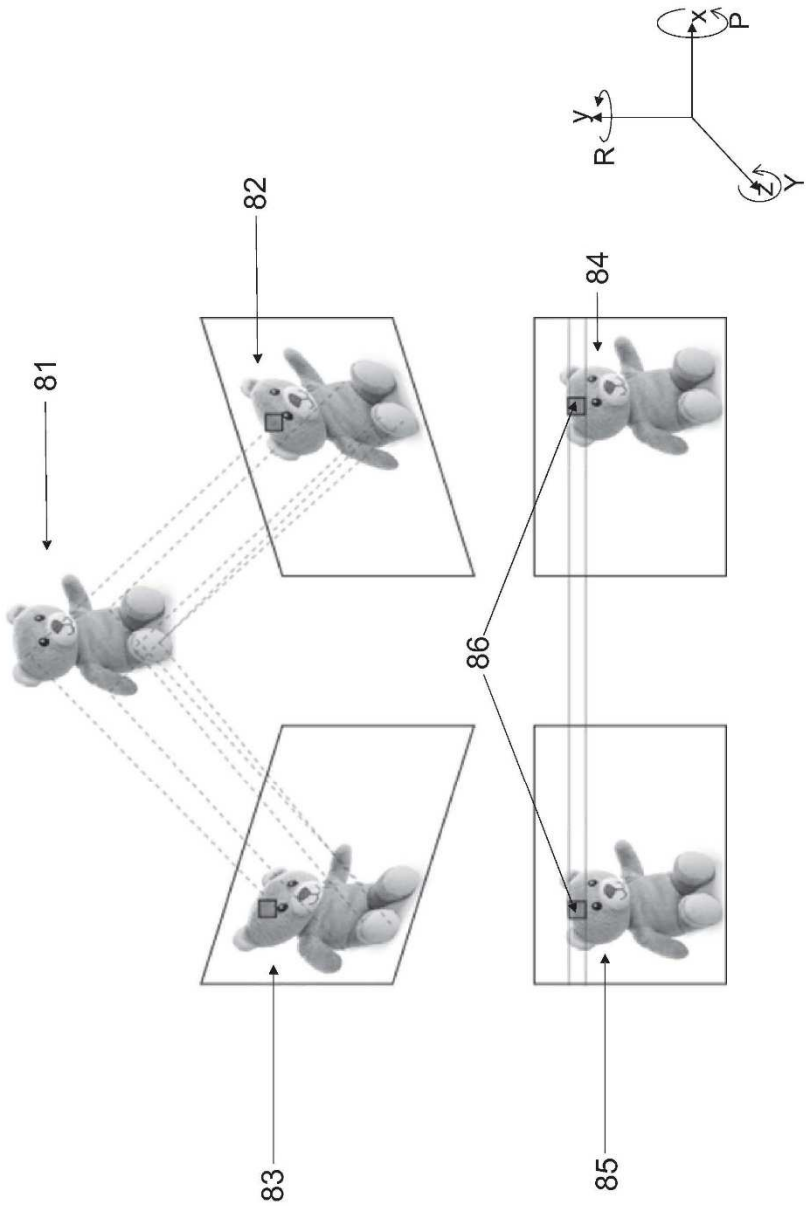

Fig. 8

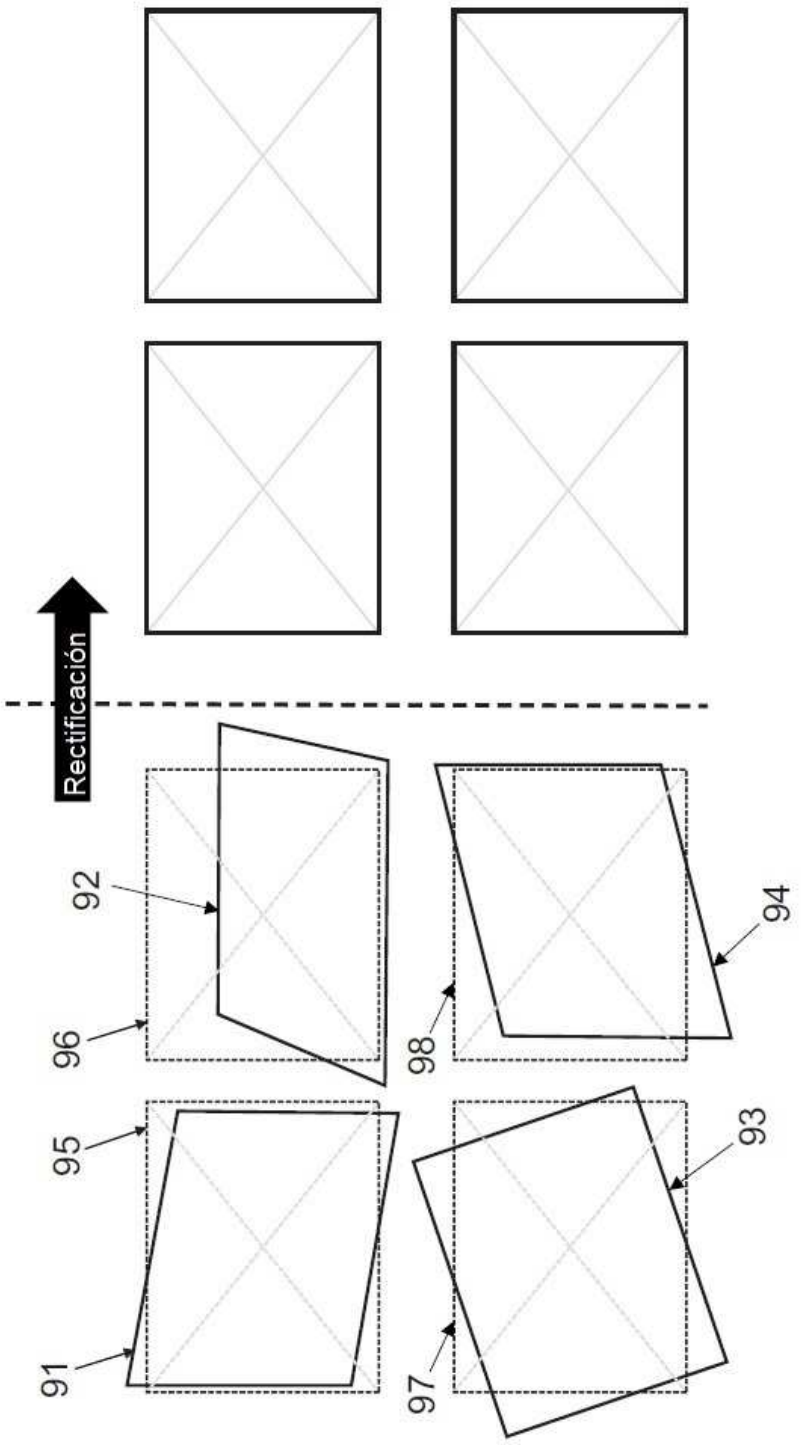

Fig. 9

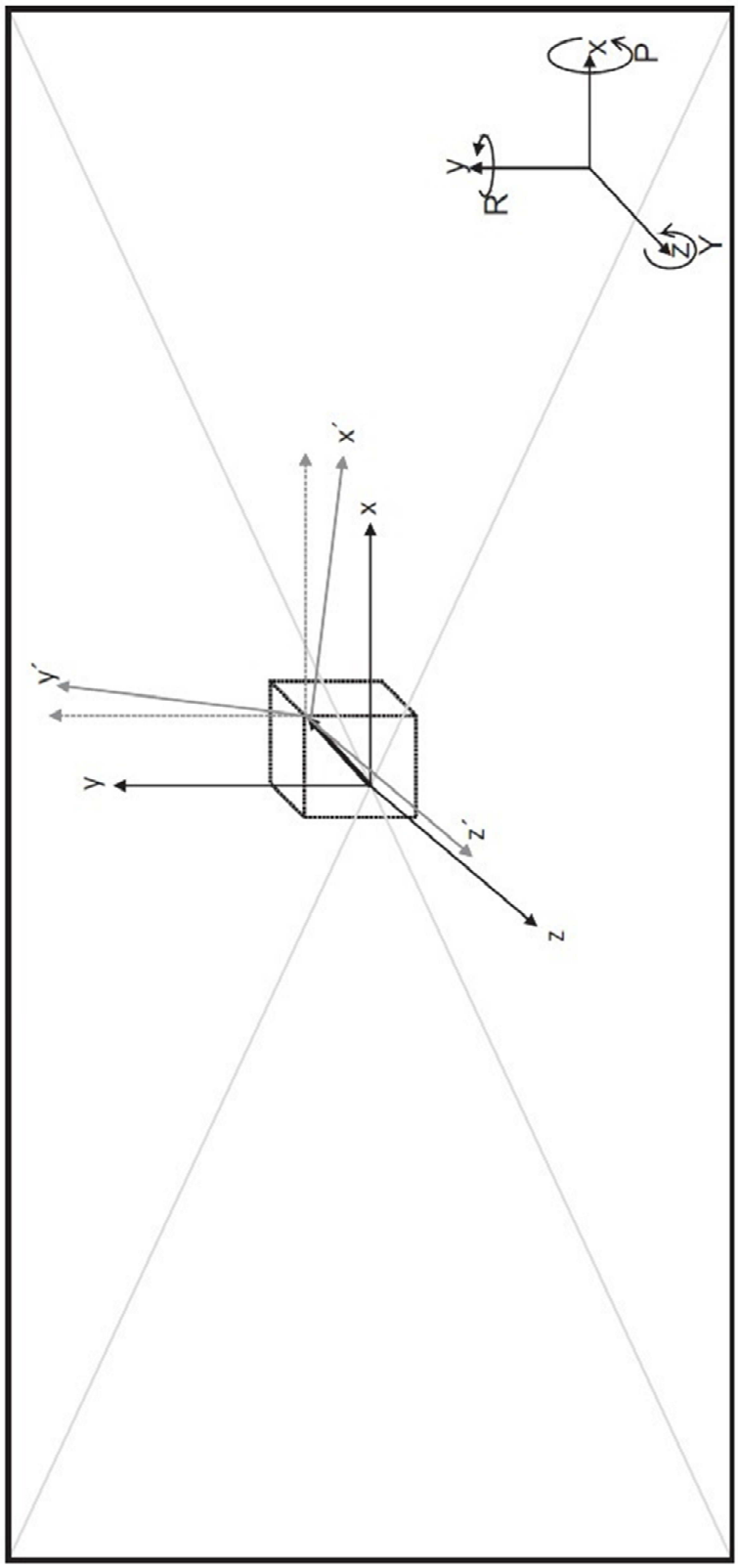

Fig. 10

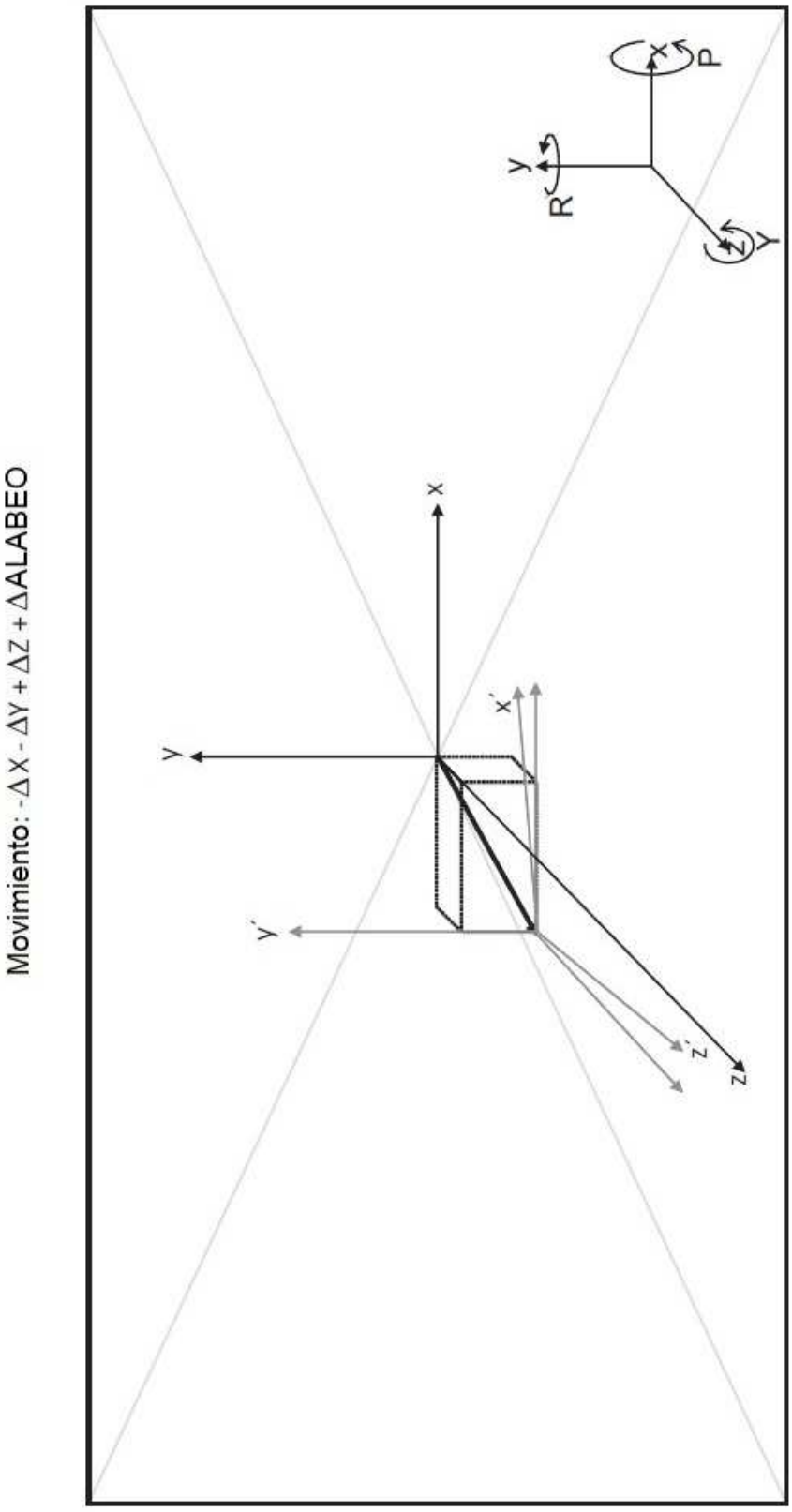

Fig. 11

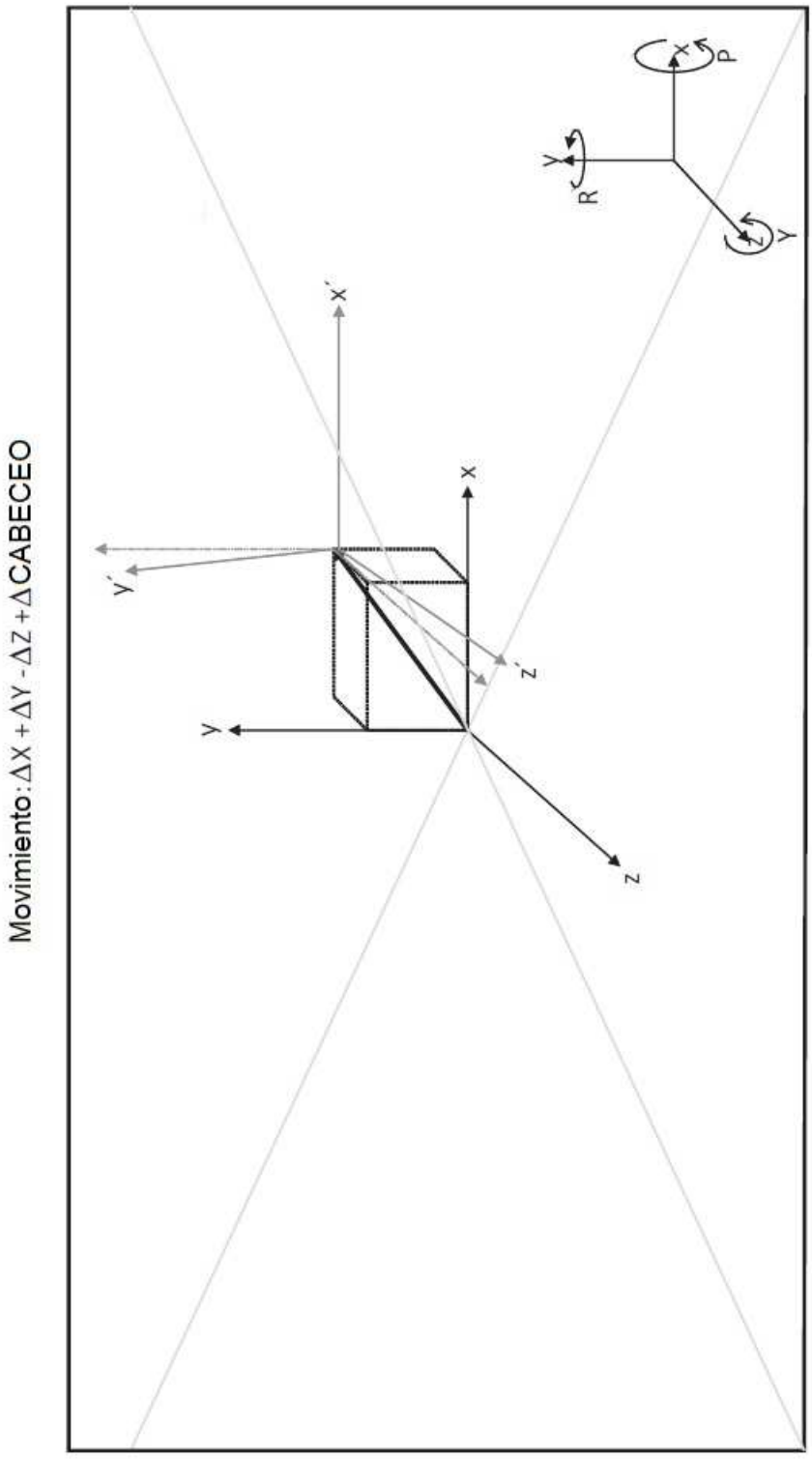

Fig. 12

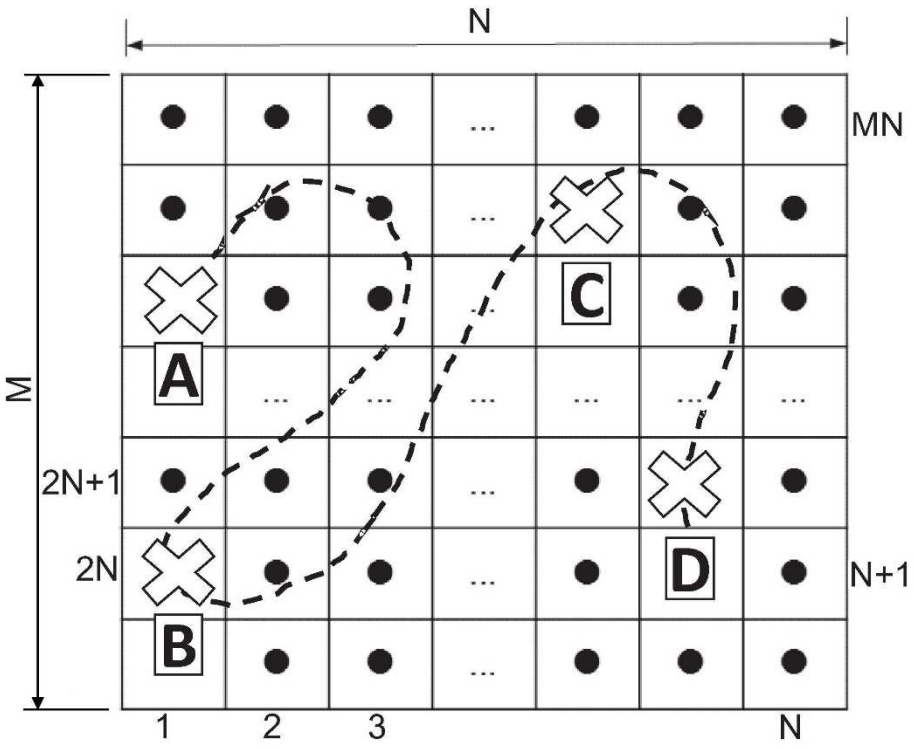

Fig. 13

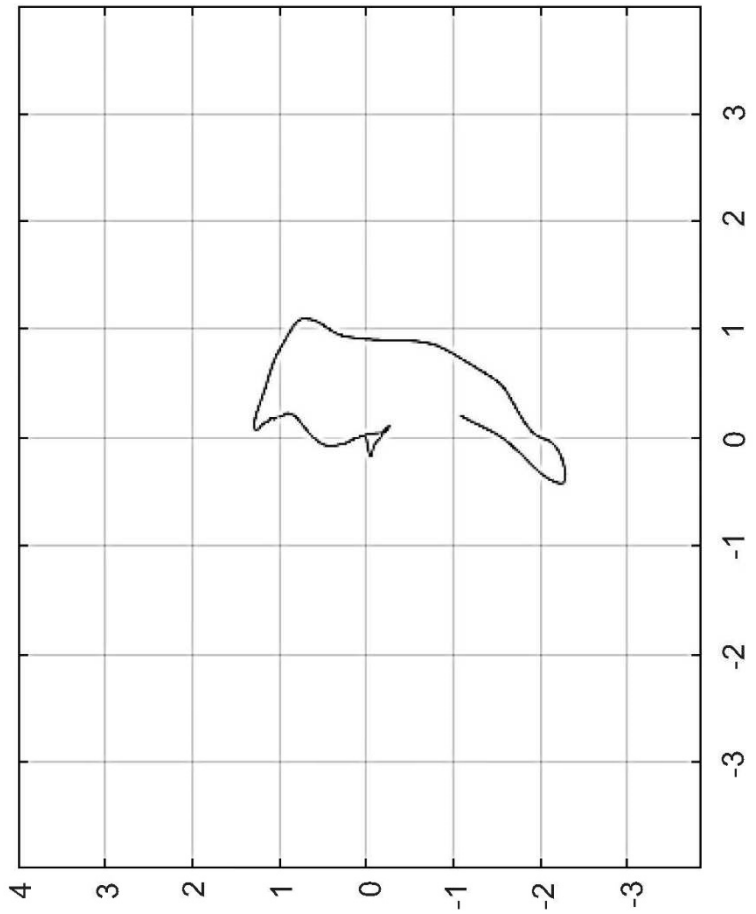

Fig. 14

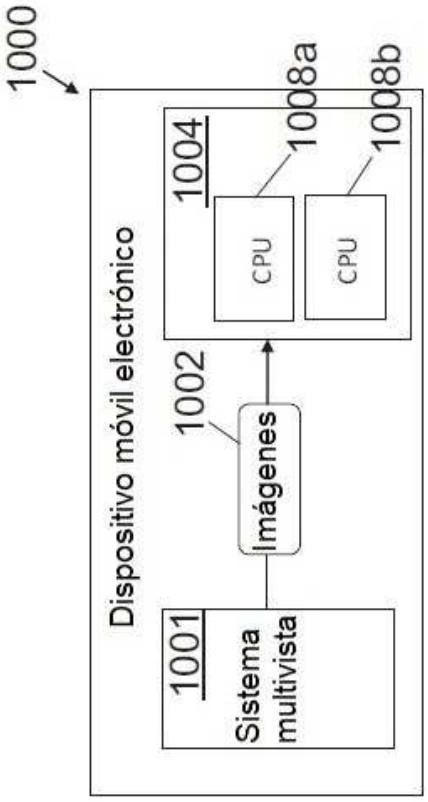

Fig. 15B

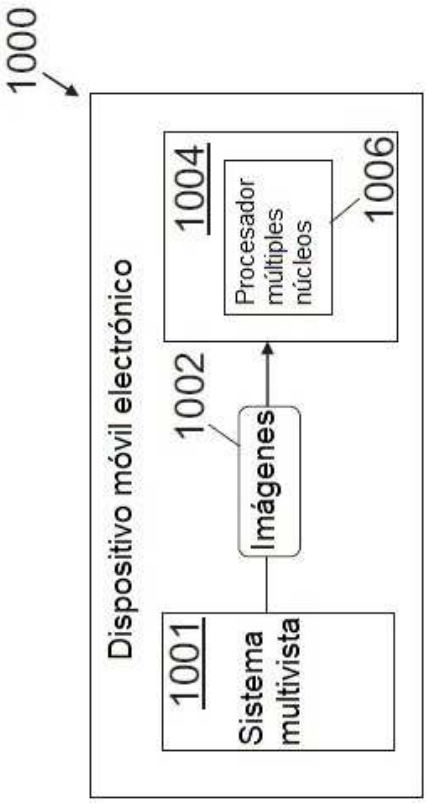

Fig. 15A

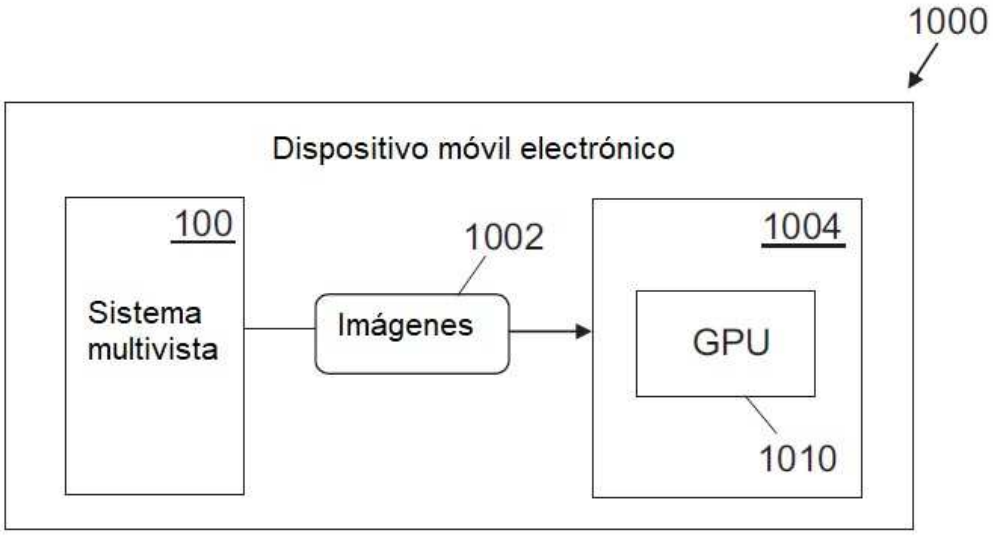

Fig. 15C

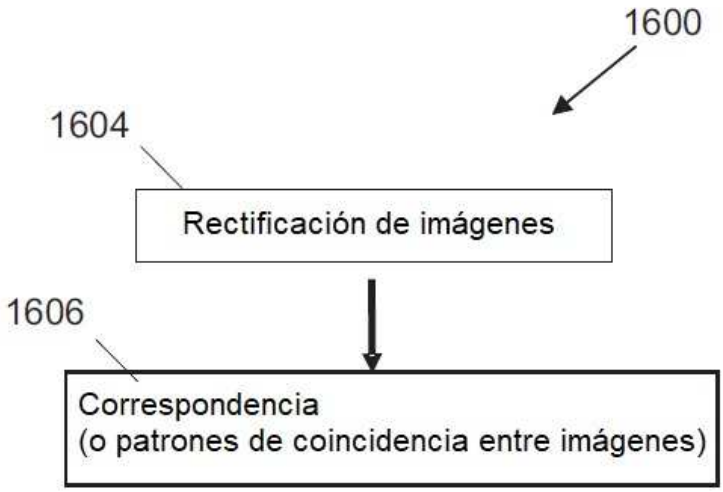

Fig. 16A

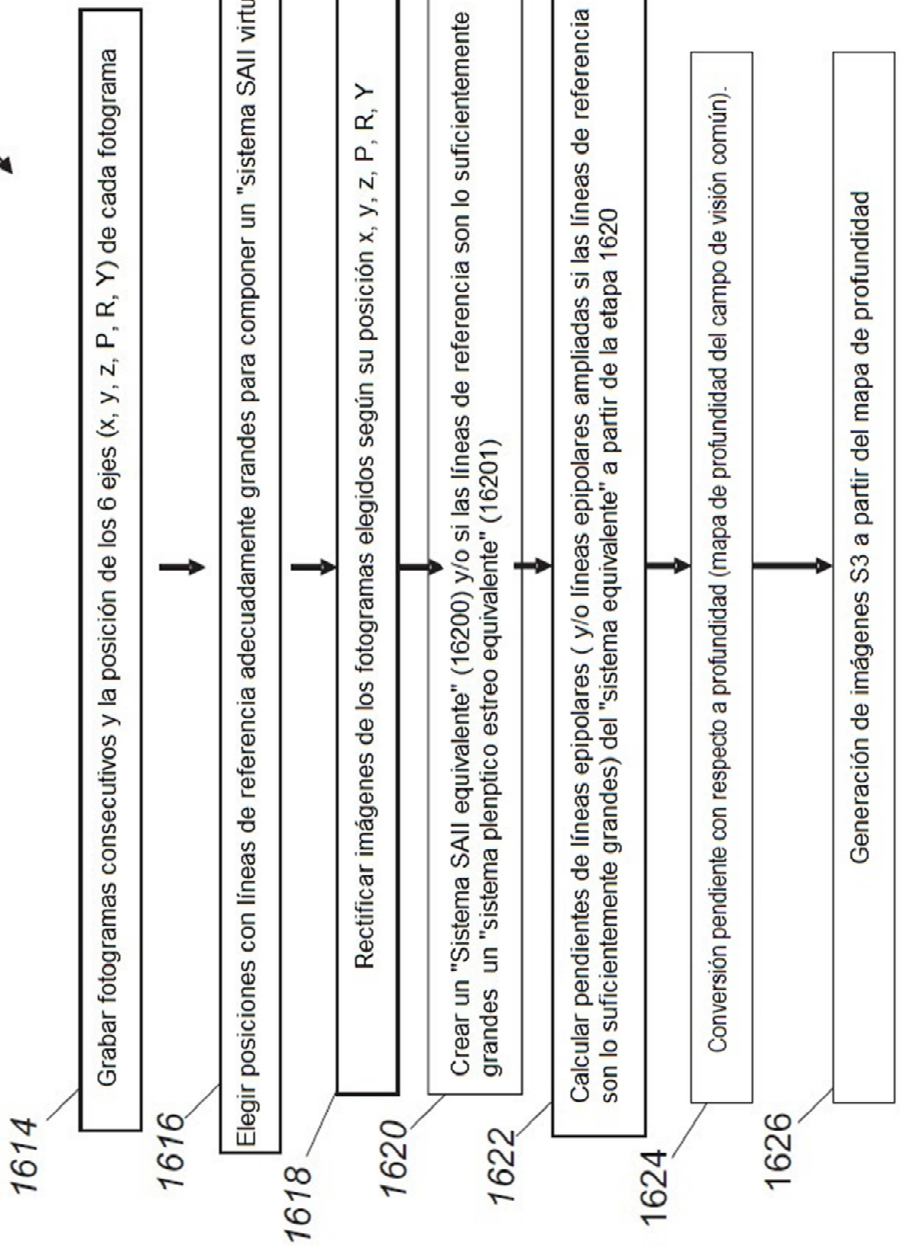
1610
1614
Grabar fotogramas consecutivos y la posición de los 6 ejes (x, y, z, P, R, Y) de cada fotograma
1616
Elegir posiciones con líneas de referencia adecuadamente grandes para componer un "sistema SAII virtual"
1618
Rectificar imágenes de los fotogramas elegidos según su posición x, y, z, P, R, Y
1620
Crear un "Sistema SAII equivalente" (16200) y/o si las líneas de referencia son lo suficientemente grandes un "sistema plenptico estreo equivalente" (16201)
1622
Calcular pendientes de líneas epipolares (y/o líneas epipolares ampliadas si las líneas de referencia son lo suficientemente grandes) del "sistema equivalente" a partir de la etapa 1620
1624
Conversión pendiente con respecto a profundidad (mapa de profundidad del campo de visión común).
1626
Generación de imágenes S3 a partir del mapa de profundidad

Fig. 16B