

19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 675 171**

51 Int. Cl.:

G06T 17/00 (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

86 Fecha de presentación y número de la solicitud internacional: **27.01.2014 PCT/US2014/013211**

87 Fecha y número de publicación internacional: **07.08.2014 WO14120613**

96 Fecha de presentación y número de la solicitud europea: **27.01.2014 E 14703514 (1)**

97 Fecha y número de publicación de la concesión europea: **09.05.2018 EP 2951788**

54 Título: **Reconstrucción tridimensional en tiempo real con uso eficaz del sensor de profundidad en términos energéticos**

30 Prioridad:

30.01.2013 US 201361758699 P
26.02.2013 US 201361769465 P
22.01.2014 US 201414161140

45 Fecha de publicación y mención en BOPI de la traducción de la patente:
09.07.2018

73 Titular/es:

QUALCOMM INCORPORATED (100.0%)
5775 Morehouse Drive
San Diego, CA 92121-1714, US

72 Inventor/es:

WAGNER, DANIEL y
MULLONI, ALESSANDRO

74 Agente/Representante:

FORTEA LAGUNA, Juan José

ES 2 675 171 T3

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

DESCRIPCIÓN

Reconstrucción tridimensional en tiempo real con uso eficaz del sensor de profundidad en términos energéticos

5 REFERENCIA CRUZADA A SOLICITUDES RELACIONADAS

[1] Esta solicitud reivindica el beneficio y la prioridad ante la Solicitud Estadounidense N° 14/161,140 titulada "Real-Time 3D Reconstruction with Power Efficient Depth Sensor Usage" ["Reconstrucción tridimensional en tiempo real con uso eficaz del sensor de profundidad en términos energéticos"], presentada el 22 de enero de 2014. A su vez, la Solicitud Estadounidense N° 14/161,140 reivindica el beneficio y la prioridad ante la Solicitud Provisoria Estadounidense N° 61/758,699 titulada "Real-Time 3D Reconstruction with Power Efficient Depth Sensor Usage" ["Reconstrucción tridimensional en tiempo real con uso eficaz del sensor de profundidad en términos energéticos"] presentada el 30 de enero de 2013, y la Solicitud Provisoria Estadounidense N° 61/769,465 titulada "Real-Time 3D Reconstruction with Power Efficient Depth Sensor Usage" ["Reconstrucción tridimensional en tiempo real con uso eficaz del sensor de profundidad en términos energéticos"] presentada el 26 de febrero de 2013. Todas las solicitudes de patente estadounidense identificadas anteriormente se incorporan por referencia en este documento en su totalidad.

20 CAMPO

[2] Esta divulgación se refiere en general a aparatos y procedimientos para la eficacia energética en dispositivos móviles.

25 ANTECEDENTES

[3] En la visión por ordenador y los gráficos por ordenador, la reconstrucción tridimensional es el proceso de determinar la forma y/o apariencia de los objetos reales. La reconstrucción tridimensional se puede basar en datos y/o imágenes de un objeto, obtenidos de varios tipos de sensores. Por ejemplo, las cámaras pueden usarse para medir el resplandor o la luz reflejada por, o emitida desde, la superficie de un objeto y la estructura tridimensional o el modelo del objeto pueden deducirse de las imágenes del objeto capturado por la cámara y de la información proporcionada por otros sensores. En general, el término 'modelo tridimensional' se utiliza en este documento para referirse a una representación de un entorno tridimensional.

[4] Normalmente, en la reconstrucción tridimensional, un conjunto de imágenes digitales se procesa fuera de línea en la modalidad por lotes, junto con otra información sensorial, para obtener un modelo tridimensional, que puede adoptar la forma de una malla tridimensional del objeto. Sin embargo, debido a que la reconstrucción tridimensional ha sido tradicionalmente costosa en términos de cálculo, a menudo se ha realizado fuera de línea y los resultados de la reconstrucción tridimensional habitualmente han estado disponibles mucho más tarde. Por lo tanto, las aplicaciones prácticas en tiempo real que utilizan la reconstrucción tridimensional han sido hasta ahora limitadas.

[5] Más recientemente, la reconstrucción tridimensional en tiempo real, o casi en tiempo real, ha ganado fuerza debido a una combinación de factores que incluyen la disponibilidad de una creciente potencia de procesamiento, algoritmos avanzados, así como nuevas formas de datos de entrada. Los usuarios ahora pueden obtener respuesta sobre la reconstrucción tridimensional casi en tiempo real, ya que las imágenes capturadas se procesan rápidamente mediante dispositivos informáticos, incluidos los dispositivos móviles. Sin embargo, muchas técnicas utilizadas para la reconstrucción tridimensional consumen mucha energía y dan como resultado un consumo de energía relativamente alto. En los dispositivos móviles, por ejemplo, un creciente consumo de energía puede drenar la fuente de alimentación o la batería, limitando así la aplicabilidad práctica de la construcción tridimensional.

[6] Por lo tanto, existe la necesidad de aparatos, sistemas y procedimientos para facilitar la reconstrucción tridimensional en tiempo real, eficaz en términos energéticos, en dispositivos informáticos y móviles.

55 SUMARIO

[7] La invención está definida por las reivindicaciones independientes. De acuerdo a algunos aspectos, se divulga un procedimiento para la reconstrucción tridimensional en tiempo real, eficaz en términos energéticos, en un dispositivo móvil. En algunas realizaciones, la estación móvil (MS) puede comprender al menos una cámara y un sensor de profundidad y el procedimiento puede comprender: capturar una primera imagen con al menos una cámara, en donde la primera imagen comprende información de color para al menos una parte de un entorno modelado por la MS, y obtener información de pose de cámara para la primera imagen. En algunas realizaciones, el procedimiento puede comprender además: determinar, basándose, en parte, en la primera imagen capturada y en la información de pose de cámara para la primera imagen, si se extiende un primer modelo tridimensional (3D) del entorno; e inhabilitar el sensor de profundidad, cuando el primer modelo tridimensional no se extiende.

5 **[8]** En otro aspecto, una MS puede comprender: una cámara, que puede capturar una primera imagen que comprende información de color, un sensor de profundidad acoplado a la cámara y un procesador acoplado al sensor de profundidad y a la cámara. En algunas realizaciones, el procesador puede configurarse para: obtener información de pose de cámara para la primera imagen; determinar, basándose, en parte, en la primera imagen y en la información de pose de cámara para la primera imagen, si se extiende un primer modelo tridimensional (3D) de un entorno modelado por la MS; e inhabilitar el sensor de profundidad, si el primer modelo tridimensional no se extiende.

10 **[9]** En un aspecto adicional, las realizaciones divulgadas corresponden a un aparato que comprende: medios de formación de imágenes, siendo los medios de formación de imágenes para capturar una primera imagen que comprende información de color; medios de detección de profundidad acoplados a los medios de formación de imágenes, y medios de procesamiento acoplados a los medios de detección de profundidad y a los medios de formación de imágenes. En algunas realizaciones, los medios de procesamiento pueden comprender además: medios para obtener información de pose de los medios de formación de imágenes para la primera imagen; medios para determinar si se extiende un primer modelo tridimensional (3D) de un entorno modelado por el aparato, basándose, en parte, en la primera imagen y en la información de pose de los medios de formación de imágenes para la primera imagen; y medios para inhabilitar los medios sensores de profundidad, si el primer modelo tridimensional no se extiende.

20 **[10]** Las realizaciones divulgadas también corresponden a medios no transitorios legibles por ordenador que comprenden instrucciones que, cuando son ejecutadas por un procesador, realizan etapas en un procedimiento en una MS que comprende al menos una cámara y un sensor de profundidad. En algunas realizaciones, el procedimiento puede comprender: capturar una primera imagen con al menos una cámara, en donde la primera imagen comprende información de color para al menos una parte de un entorno modelado por la MS; y obtener información de pose de cámara para la primera imagen. En algunas realizaciones, el procedimiento puede comprender además: determinar, basándose, en parte, en la primera imagen capturada y en la información de pose de cámara para la primera imagen, si se extiende un primer modelo tridimensional (3D) del entorno; e inhabilitar el sensor de profundidad, cuando el primer modelo tridimensional no se extiende.

30 **[11]** En otro aspecto, un procedimiento en una estación móvil (MS) que comprende al menos una cámara y un sensor de profundidad puede comprender: capturar una primera imagen con al menos una cámara, comprendiendo la primera imagen información de color para al menos una parte de un entorno modelado por la MS; obtener información de pose de cámara para la primera imagen; determinar, basándose, en parte, en la primera imagen capturada y en la información de pose de cámara para la primera imagen, si se actualiza un primer modelo tridimensional (3D) del entorno; e inhabilitar el sensor de profundidad, cuando el primer modelo tridimensional no se actualiza. Además, la inhabilitación del sensor de profundidad puede comprender al menos uno entre: desconectar la energía suministrada al sensor de profundidad, o inhabilitar la funcionalidad relacionada con el cálculo de una imagen de profundidad. El procedimiento comprende además: habilitar el sensor de profundidad, cuando se actualiza el primer modelo tridimensional. Al habilitar el sensor de profundidad, el primer modelo tridimensional se puede actualizar: capturando una segunda imagen con la al menos una cámara; obteniendo un mapa de profundidad mediante el aumento de la segunda imagen con información de profundidad proporcionada, al menos en parte, por el sensor de profundidad; y actualizando un conjunto de datos volumétricos en el primer modelo tridimensional con la información de profundidad. En algunos ejemplos que no forman parte de la invención, el conjunto de datos volumétricos puede representarse mediante una función de distancia truncada con signo (TDSF) tridimensional. El conjunto de datos volumétricos se puede actualizar proyectando muestras en la TDSF tridimensional en el mapa de profundidad, basándose en la pose de la cámara; determinando una distancia relativa medida desde la ubicación de la muestra a una superficie definida por el conjunto de datos volumétricos; correlacionando la distancia relativa medida con la TDSF; y combinando el valor de la TDSF correlacionado con un valor de TDSF almacenado para la muestra.

50 **[12]** En otro aspecto, una MS puede comprender: una cámara, que puede capturar una primera imagen que comprende información de color, un sensor de profundidad acoplado a la cámara y un procesador acoplado al sensor de profundidad y a la cámara. En algunas realizaciones, el medio de procesamiento está configurado para: obtener información de pose de cámara para la primera imagen; determinar, basándose, en parte, en la primera imagen y en la información de pose de cámara para la primera imagen, si se actualiza un primer modelo tridimensional (3D) del entorno; e inhabilitar el sensor de profundidad, cuando el primer modelo tridimensional no se actualiza.

60 **[13]** En un aspecto adicional, las realizaciones divulgadas corresponden a un aparato que comprende: medios de formación de imágenes, siendo los medios de formación de imágenes para capturar una primera imagen que comprende información de color; medios de detección de profundidad acoplados a los medios de formación de imágenes, y medios de procesamiento acoplados a los medios de detección de profundidad y a los medios de formación de imágenes. En algunas realizaciones, los medios de procesamiento pueden comprender además: medios para obtener información de pose de cámara para la primera imagen; medios para determinar, basándose, en parte, en la primera imagen y en la información de pose de cámara para la primera imagen, si se actualiza un

primer modelo tridimensional (3D) del entorno; y medios para inhabilitar el sensor de profundidad, cuando el primer modelo tridimensional no se actualiza.

[14] Las realizaciones divulgadas también corresponden a medios no transitorios legibles por ordenador que comprenden instrucciones que, cuando son ejecutadas por un procesador, realizan etapas en un procedimiento en una MS que comprende al menos una cámara y un sensor de profundidad. En algunas realizaciones, el procedimiento puede comprender: capturar una primera imagen con al menos una cámara, comprendiendo la primera imagen información de color para al menos una parte de un entorno modelado por la MS; obtener información de pose de cámara para la primera imagen; determinar, basándose, en parte, en la primera imagen capturada y en la información de pose de cámara para la primera imagen, si se actualiza un primer modelo tridimensional (3D) del entorno; e inhabilitar el sensor de profundidad, cuando el primer modelo tridimensional no se actualiza.

[15] Las realizaciones divulgadas también se refieren al software, al firmware y a las instrucciones de programa creadas, almacenadas, objeto de acceso o modificadas por procesadores que usan medios legibles por ordenador o memoria legible por ordenador. Los procedimientos descritos se pueden realizar en procesadores y en varios dispositivos móviles.

[16] Estas y otras realizaciones se explican adicionalmente a continuación con respecto a las siguientes figuras. Se entiende que otros aspectos devendrán inmediatamente evidentes para los expertos en la materia a partir de la siguiente descripción detallada, en la que se muestran y describen diversos aspectos a modo de ilustración. Los dibujos y la descripción detallada han de considerarse como de naturaleza ilustrativa, y no restrictiva.

BREVE DESCRIPCIÓN DE LOS DIBUJOS

[17] Las realizaciones de la invención se describirán, sólo a modo de ejemplo, con referencia a los dibujos.

La Figura 1 muestra un diagrama de bloques de un dispositivo móvil ejemplar capaz de la reconstrucción tridimensional eficaz en términos energéticos.

Las Figuras 2A y 2B muestran diagramas de flujo para un procedimiento ejemplar para la reconstrucción tridimensional eficaz en términos energéticos, congruente con las realizaciones divulgadas.

La Figura 2C muestra un diagrama de flujo para la rutina o el procedimiento 275, para extender la reconstrucción usando una técnica de reconstrucción volumétrica congruente con las realizaciones divulgadas.

La Figura 2D muestra un diagrama de flujo que representa un procedimiento ejemplar para fusionar datos de profundidad del marco de la cámara en el conjunto de datos volumétrico congruente con las realizaciones divulgadas.

La Figura 3 muestra un diagrama de flujo del procedimiento ejemplar 400 para la reconstrucción tridimensional congruente con las realizaciones divulgadas.

DESCRIPCIÓN DETALLADA

[18] La descripción detallada que se expone a continuación en relación con los dibujos adjuntos está concebida como una descripción de diversos aspectos de la presente divulgación y no está concebida para representar los únicos aspectos en los que puede llevarse a la práctica la invención. Cada aspecto descrito en esta divulgación se proporciona meramente como un ejemplo o ilustración de la presente divulgación, y no debería interpretarse necesariamente como preferido o ventajoso sobre otros aspectos. La descripción detallada incluye detalles específicos con el fin de proporcionar una comprensión exhaustiva de la presente divulgación. Sin embargo, será evidente para los expertos en la técnica que la presente divulgación puede llevarse a la práctica sin estos detalles específicos. En algunos casos, estructuras y dispositivos bien conocidos se muestran en forma de diagrama de bloques para evitar el oscurecimiento de los conceptos de la presente divulgación. Los acrónimos y otra terminología descriptiva pueden usarse simplemente por conveniencia y claridad y no pretenden limitar el alcance de la divulgación.

[19] La Figura 1 muestra un diagrama de bloques de una estación móvil (MS) 100 ejemplar, capaz de realizar una construcción tridimensional en tiempo real, o casi en tiempo real, de una manera congruente con las realizaciones divulgadas. Como se usa en el presente documento, el dispositivo móvil o la estación móvil (MS) 100 puede adoptar la forma de un teléfono celular, un teléfono móvil u otro dispositivo de comunicación inalámbrica, un dispositivo de sistema de comunicación personal (PCS), un dispositivo de navegación personal (PND), un Administrador de Información Personal (PIM) o un Asistente Digital Personal (PDA), un ordenador portátil, una tableta, un ordenador plegable y/o un ordenador de mano. Los términos dispositivo móvil o estación móvil se usan indistintamente en este documento. En algunas realizaciones, la MS 100 puede ser capaz de recibir señales de

comunicación y/o navegación inalámbricas. Los términos "dispositivo móvil" y "estación móvil" se usan indistintamente en este documento.

[20] Además, el término "estación móvil" también pretende incluir dispositivos que se comunican con un dispositivo de navegación personal (PND), tal como mediante una conexión de corto alcance, inalámbrica, infrarroja o cableada, u otras conexiones y/o mediante procesamiento relacionado con la posición que tiene lugar en el dispositivo o en el PND. Además, la "estación móvil" pretende incluir todos los dispositivos, incluidos varios dispositivos de comunicación inalámbrica, que son capaces de comunicarse con un servidor, independientemente de si la recepción de la señal inalámbrica, la recepción de datos de asistencia y/o el procesamiento relacionado se produce en el dispositivo, en un servidor o en otro dispositivo asociado a la red. Cualquier combinación operable de lo anterior también se considera una "estación móvil".

[21] El término "estación móvil" también pretende incluir juegos u otros dispositivos que pueden no estar configurados para conectarse a una red o comunicarse de otro modo, ya sea de forma inalámbrica o por una conexión por cable, con otro dispositivo. Por ejemplo, una "estación móvil" puede omitir elementos de comunicación y/o funcionalidad de red. Por ejemplo, las realizaciones descritas en este documento pueden implementarse en un dispositivo autónomo que no está configurado para conectarse, para formar redes cableadas o inalámbricas, con otro dispositivo.

[22] Como se muestra en la Figura 1, la MS 100 puede incluir las cámaras 110, la Unidad de Medición Inercial (IMU) 130, los procesadores 150, la memoria 160 y/o el transceptor 170, el visor / la pantalla 180 que pueden estar operativamente acoplados entre sí y con otras unidades funcionales (no mostradas) en la MS 110 mediante las conexiones 120. Las conexiones 120 pueden comprender buses, líneas, fibras, enlaces, etc., o alguna combinación de los mismos.

[23] El transceptor 170 puede, por ejemplo, incluir un transmisor habilitado para transmitir una o más señales por uno o más tipos de redes de comunicación inalámbrica y un receptor para recibir una o más señales transmitidas por los uno o más tipos de redes de comunicación inalámbrica. El transceptor 170 puede permitir la comunicación con redes inalámbricas basadas en varias tecnologías tales como, pero sin limitarse a, femtocélulas, redes de Wi-Fi o Redes de área local inalámbricas (WLAN), que pueden basarse en la familia de normas IEEE 802.11, Redes de área personal inalámbricas (WPAN) tales como Bluetooth, la Comunicación de Campo Cercano (NFC), redes basadas en la familia de normas IEEE 802.15x, etc., y/o redes de área amplia inalámbricas (WWAN) tales como LTE, WiMAX, etc. La MS 100 puede también incluir uno o más puertos para comunicarse por redes cableadas.

[24] En algunas realizaciones, la MS 100 puede comprender sensores de imagen tales como sensores CCD o CMOS y/o cámaras 110, que se denominan en lo sucesivo "cámaras 110". Las cámaras 110 pueden convertir una imagen óptica en una imagen electrónica o digital y pueden enviar imágenes capturadas al procesador 150.

[25] En general, las cámaras 110 pueden ser cámaras a color o de escala de grises, que proporcionan "información de color", mientras que la "información de profundidad" puede ser proporcionada por un sensor de profundidad. El término "información de color", tal como se usa en el presente documento, se refiere a información de color y/o de escala de grises. En general, como se usa en el presente documento, una imagen de color o la información de color puede visualizarse como comprendiendo de 1 a N canales, donde N es algún número entero que depende del espacio de colores que se usa para almacenar la imagen. Por ejemplo, una imagen RGB comprende tres canales, con cada canal para información roja, azul y verde.

[26] La información de profundidad se puede capturar de diversas maneras utilizando sensores de profundidad. El término "sensor de profundidad" se usa para referirse a las unidades funcionales que pueden usarse para obtener información de profundidad de manera independiente y/o conjuntamente con las cámaras 110. En algunas realizaciones, los sensores de profundidad pueden inhabilitarse cuando no se usan. Por ejemplo, el sensor de profundidad puede colocarse en modalidad de espera o apagarse cuando no se utiliza. En algunas realizaciones, los procesadores 150 pueden inhabilitar (o habilitar) la detección de profundidad en uno o más momentos en el tiempo de una manera congruente con las realizaciones divulgadas. El término "inhabilitar el sensor de profundidad" también se utiliza para referirse a inhabilitar sensores pasivos tales como sensores de visión estereoscópica y/o funcionalidad relacionada con el cálculo de imágenes de profundidad, incluyendo hardware, firmware y/o software asociado a dicha funcionalidad. Por ejemplo, en una realización, cuando se inhabilita un sensor de visión estereoscópica, las imágenes capturadas por la cámara 110 pueden ser monoculares. Además, el término "inhabilitar el sensor de profundidad" también se usa para referirse a la inhabilitación del cálculo asociado al procesamiento de imágenes estereoscópicas capturadas a partir de sensores pasivos de visión estereoscópica. Por ejemplo, en una realización, aunque las imágenes estereoscópicas pueden capturarse mediante un sensor pasivo de visión estereoscópica, los procesadores 150 pueden no procesar las imágenes estereoscópicas y, en su lugar, pueden seleccionar una única imagen del par estereoscópico para su procesamiento.

[27] En algunas realizaciones, el sensor de profundidad puede ser parte de las cámaras 110. Por ejemplo, en algunas realizaciones, la MS 100 puede comprender cámaras RGBD, que pueden capturar información de profundidad (D) por píxel cuando el sensor de profundidad está habilitado, además de imágenes de color (RGB).

- 5 [28] Como otro ejemplo, en algunas realizaciones, la cámara 110 puede adoptar la forma de una cámara tridimensional de Hora de Vuelo (3DToF). En realizaciones con cámaras 3DToF 110, el sensor de profundidad puede adoptar la forma de una luz estroboscópica acoplada a la cámara 3DToF 110, que puede iluminar objetos en una escena y la luz reflejada puede capturarse mediante un sensor CCD / CMOS en la cámara 110. La información de profundidad se puede obtener midiendo el tiempo que tardan los pulsos de luz en viajar a los objetos y volver al sensor.
- 10 [29] Como un ejemplo adicional, el sensor de profundidad puede adoptar la forma de una fuente de luz acoplada a las cámaras 110. En una realización, la fuente de luz puede proyectar un patrón de luz estructurado o texturizado, que puede consistir en una o más bandas estrechas de luz, sobre objetos en una escena. La información de profundidad puede obtenerse luego explotando las distorsiones geométricas del patrón proyectado, causadas por la forma de la superficie del objeto. En una realización, la información de profundidad puede obtenerse a partir de estereosensores tales como una combinación de un proyector de luz estructurado infrarrojo y una cámara de infrarrojos registrada para una cámara RGB.
- 15 [30] En algunas realizaciones, las cámaras 110 pueden adoptar la forma de cámaras estereoscópicas. Por ejemplo, un sensor de profundidad puede formar parte de un sensor pasivo de visión estereoscópica, que puede usar dos o más cámaras para obtener información de profundidad para una escena. Las coordenadas de píxeles de los puntos comunes a ambas cámaras en una escena capturada se pueden usar, junto con la información de la pose de la cámara y/o técnicas de triangulación, para obtener información de profundidad por píxel.
- 20 [31] En algunas realizaciones, la MS 100 puede comprender cámaras múltiples 110, tales como cámaras frontales duales y/o cámaras de visión frontal y trasera, que también pueden incorporar diversos sensores. En algunas realizaciones, las cámaras 110 pueden ser capaces de capturar imágenes tanto fijas como de vídeo. En algunas realizaciones, las cámaras 110 pueden ser cámaras de vídeo RGBD o estereoscópicas, capaces de capturar imágenes a 30 cuadros por segundo (fps). En una realización, las imágenes capturadas por las cámaras 110 pueden estar en un formato en bruto sin comprimir y pueden comprimirse antes de ser procesadas y/o almacenadas en la memoria 160. En algunas realizaciones, la compresión de imágenes puede ser realizada por procesadores 150 usando técnicas de compresión sin pérdida o con pérdidas.
- 25 [32] En algunas realizaciones, los procesadores 150 también pueden recibir información de entrada desde la IMU 130. En otras realizaciones, la IMU 130 puede comprender uno o más acelerómetros de 3 ejes, uno o más giroscopios de 3 ejes y/o uno o más magnetómetros. La IMU 130 puede proporcionar velocidad, orientación y / u otra información relacionada con la posición al procesador 150. En algunas realizaciones, la IMU 130 puede emitir información medida en sincronización con la captura de cada cuadro de imagen por las cámaras 130. En algunas realizaciones, la salida de la IMU 130 puede ser usada en parte por el procesador 150 para determinar una pose de la cámara 110 y/o la MS 100.
- 30 [33] Además, la MS 100 puede incluir una pantalla o visor 180 capaz de reproducir imágenes en color, incluidas imágenes tridimensionales. En algunas realizaciones, la pantalla 170 puede usarse para exhibir imágenes en vivo capturadas por la cámara 110, imágenes de Realidad Aumentada (AR), Interfaces Gráficas de Usuario (GUI), salida de programas, etc. En algunas realizaciones, la pantalla 180 puede comprender y/o albergarse con una pantalla táctil para permitir a los usuarios ingresar datos mediante alguna combinación de teclados virtuales, íconos, menús u otras interfaces gráficas de usuario (GUI), gestos de usuario y/o dispositivos de entrada tales como punzones y otros implementos de escritura. En algunas realizaciones, la pantalla 180 puede implementarse usando una Pantalla de Cristal Líquido (LCD) o una pantalla de Diodos Emisores de Luz (LED), tal como una pantalla de LED Orgánica (OLED). En otras realizaciones, la pantalla 180 puede ser una pantalla para llevar puesta, que puede estar operativamente acoplada a, pero alojada por separado de, otras unidades funcionales en la MS 100. En algunas realizaciones, la MS 100 puede comprender puertos para permitir la exhibición de las imágenes tridimensionales reconstruidas a través de un monitor independiente acoplado a la MS 100.
- 35 [34] La pose de la cámara 110 se refiere a la posición y orientación de la cámara 110 con respecto a un marco de referencia. En algunas realizaciones, la pose de cámara puede determinarse para 6 grados de libertad (6DOF), que se refiere a tres componentes de traslación (que pueden estar dados por coordenadas X, Y, Z de un marco de referencia) y tres componentes angulares (por ejemplo, balanceo, inclinación y guiñada en relación con el mismo marco de referencia).
- 40 [35] En algunas realizaciones, la pose de la cámara 110 y/o la MS 100 puede ser determinada y/o rastreada por el procesador 150 usando una solución de seguimiento visual basada en imágenes capturadas por la cámara 110. Por ejemplo, un Módulo de Visión por ordenador (CV) 155 que se ejecuta en el procesador 150 puede implementar y ejecutar un seguimiento basado en visión por ordenador, un seguimiento basado en modelos y/o procedimientos de Localización y Correlación Simultánea (SLAM). SLAM se refiere a una clase de técnicas donde se crea un mapa de un entorno, tal como un mapa de un entorno modelado por la MS 100, mientras se rastrea simultáneamente la pose de la cámara con respecto a ese mapa. En algunas realizaciones, los procedimientos implementados por el módulo de CV 155 pueden basarse en datos de imágenes en color o en escala de grises capturados por las cámaras 110 y
- 45
50
55
60
65

pueden usarse para generar estimaciones de mediciones de poses de 6DOF de la cámara. En algunas realizaciones, la salida de la IMU 130 puede usarse para estimar, corregir y/o ajustar de otro modo la pose estimada. Además, en algunas realizaciones, las imágenes capturadas por las cámaras 110 se pueden usar para recalibrar o realizar ajustes de sesgo para la IMU 130.

5 [36] No todas las unidades funcionales comprendidas en la MS 100 se han mostrado en la Figura 1. La MS 100 ejemplar también se puede modificar de varias formas, de una manera coherente con la divulgación, tal como mediante la adición, combinación u omisión de uno o más de los bloques funcionales mostrados. Por ejemplo, en algunas configuraciones, la MS 100 puede no incluir la IMU 130 o el Transceptor 170. Además, en ciertas implementaciones ejemplares, la MS 100 puede incluir varios otros sensores (no mostrados) tales como un sensor de luz ambiental, micrófonos, sensores acústicos, sensores ultrasónicos, telémetros láser, etc. En algunas realizaciones, partes de la MS 100 pueden adoptar la forma de uno o más conjuntos de chips, y/o similares.

15 [37] Los procesadores 150 pueden implementarse usando una combinación de hardware, firmware y software. Los procesadores 150 pueden representar uno o más circuitos configurables para realizar al menos una parte de un procedimiento informático o un proceso relacionado con reconstrucción tridimensional, SLAM, seguimiento, procesamiento de imágenes, etc., y pueden recuperar instrucciones y/o datos de la memoria 160. Los procesadores 150 pueden implementarse utilizando uno o más circuitos integrados específicos de la aplicación (ASIC), unidades de procesamiento central y/o gráfico (CPU y/o GPU), procesadores de señales digitales (DSP), dispositivos de procesamiento de señales digitales (DSPD), dispositivos lógicos programables (PLD), formaciones de compuertas programables en el terreno (FPGA), controladores, micro-controladores, microprocesadores, núcleos procesadores integrados, dispositivos electrónicos, otras unidades electrónicas diseñadas para realizar las funciones descritas en este documento, o una combinación de los mismos.

25 [38] La memoria 160 puede implementarse dentro de los procesadores 150 y/o ser externa a los procesadores 150. Como se usa en este documento, el término "memoria" se refiere a cualquier tipo de memoria a largo plazo, a corto plazo, volátil, no volátil u otra memoria, y no está limitado a ningún tipo particular de memoria, o cantidad de memorias, o tipo de medios físicos sobre los cuales la memoria está almacenada. En algunas realizaciones, la memoria 160 puede contener código de programa que facilita uno o más procesamientos de imágenes, realizar tareas de SLAM, seguimiento, reconstrucción tridimensional y otras tareas realizadas por el Módulo de CV 155 en el procesador 150. Por ejemplo, la memoria 160 puede contener datos, imágenes fijas capturadas, información de profundidad, tramas de vídeo, resultados de programas, así como datos proporcionados por la IMU 130 y otros sensores. En general, la memoria 160 puede representar cualquier mecanismo de almacenamiento de datos. La memoria 160 puede incluir, por ejemplo, una memoria primaria y/o una memoria secundaria. La memoria primaria puede incluir, por ejemplo, una memoria de acceso aleatorio, memoria de solo lectura, etc. Mientras que en la Figura 1 se ilustra como independiente de los procesadores 150, debería entenderse que la totalidad o parte de una memoria primaria puede proporcionarse dentro de, o co-situada y/o acoplada de otra manera a, los procesadores 150.

40 [39] La memoria secundaria puede incluir, por ejemplo, igual o similar tipo de memoria que la memoria principal y/o uno o más dispositivos o sistemas de almacenamiento de datos, tales como, por ejemplo, unidades de memoria flash / USB, unidades de tarjeta de memoria, unidades de disco, unidades de disco óptico, unidades de cinta, unidades de estado sólido, unidades híbridas, etc. En ciertas implementaciones, la memoria secundaria puede ser operativamente receptiva de, o configurable de otro modo para acoplarse a, un medio no transitorio legible por ordenador en una unidad de medios extraíbles (no mostrada) acoplada a la MS 100. En algunas realizaciones, el medio no transitorio legible por ordenador puede formar parte de la memoria 160 y/o del procesador 150.

50 [40] En algunas realizaciones, el módulo de CV 155 puede implementar diversos procedimientos de visión por ordenador y/o procesar imágenes capturadas por la cámara 110. Por ejemplo, el módulo de CV 155 puede ser capaz de procesar una o más imágenes capturadas por la cámara 110 para realizar la reconstrucción tridimensional de un entorno que está siendo modelado utilizando la información de profundidad asociada a las imágenes capturadas. En algunas realizaciones, la información de profundidad puede obtenerse a partir de sensores de profundidad.

55 [41] En función de los datos de profundidad, durante la reconstrucción, a cada píxel se le puede asignar una coordenada tridimensional. En una realización, el módulo de CV 155 en el procesador 150 puede rastrear la posición de la cámara 110 usando un sistema SLAM visual monocular (de cámara única) para construir un mapa aproximado del entorno alrededor de la MS para un seguimiento de 6DOF preciso y robusto de la cámara 110. El término monocular se refiere al uso de una cámara única no estereoscópica para capturar imágenes, o a imágenes capturadas sin información de profundidad. Un módulo de reconstrucción tridimensional puede usar entonces la pose de la cámara y la información de profundidad por píxel para extruir la(s) imagen(es) capturada(s) a lo largo de una dirección de visualización y crear una malla tridimensional densa texturizada, que puede usarse para representar el entorno tridimensional modelado.

65 [42] En los procedimientos convencionales de reconstrucción tridimensional, todas las muestras de entrada son siempre de la misma modalidad, tales como los pares de color + profundidad. En consecuencia, la adquisición de

datos de profundidad utilizando sensores de profundidad tales como luz estructurada o luces estroboscópicas puede consumir significativamente más energía que una simple cámara RGB. Además, aunque la profundidad procedente de los sensores de visión estereoscópica puede usar detección pasiva, la sobrecarga de cálculo de las operaciones de fusión estereoscópica continua durante la reconstrucción tridimensional puede aumentar el consumo de energía de manera significativa. Por lo tanto, la reconstrucción tridimensional en tiempo real convencional en dispositivos móviles a menudo no es práctica debido a las restricciones de disponibilidad de energía. Por ejemplo, el uso repetido de luces estroboscópicas u otras fuentes de luz puede agotar rápidamente la batería en una MS, limitando así la aplicabilidad de la reconstrucción tridimensional.

[43] Las Figuras 2A a 2C muestran diagramas de flujo para un procedimiento ejemplar para la reconstrucción tridimensional eficaz en términos energéticos, coherente con las realizaciones divulgadas. En algunas realizaciones, el procedimiento 200 puede ser realizado por las unidades de procesamiento 150 en la MS 100, conjuntamente con otras una o más unidades funcionales. En algunas realizaciones, partes del procedimiento 200 pueden ser realizadas por el módulo de CV 155.

[44] En algunas realizaciones, el procedimiento 200 puede consistir en dos módulos, el Módulo de Seguimiento 210 y Módulo de Reconstrucción 250. Cualquier sistema de seguimiento en tiempo real coherente con las realizaciones divulgadas, tales como SLAM monocular (Mono-SLAM), seguimiento y correlación en paralelo (PTAM), etc., puede seleccionarse para su uso con el Módulo de Seguimiento 210. En algunas realizaciones, el Módulo de Seguimiento 210 y el Módulo de Reconstrucción 250 pueden ejecutarse simultáneamente. El Módulo de Seguimiento 210 puede rastrear o actualizar periódicamente, o continuamente, la pose de la cámara y / u otros datos relacionados del sensor, no de profundidad, y proporcionar la pose actualizada de la cámara, la información del sensor e imágenes en color sin información de profundidad, periódicamente o previa solicitud al Módulo de Reconstrucción 250. En algunas realizaciones, el Módulo de Seguimiento 210 puede usar una cámara única o monocular para capturar imágenes sin información de profundidad y obtener información de pose.

[45] El Módulo de Reconstrucción 250 que, en algunas realizaciones, puede ejecutarse simultáneamente con el Módulo de Seguimiento 210, puede recibir actualizaciones periódicas o continuas desde el Módulo de Seguimiento 210. Las actualizaciones recibidas desde el Módulo de Seguimiento 210 pueden proporcionar imágenes en color sin información de profundidad e información precisa de pose, tales como mediciones de pose de 6DOF para las cámaras 110. En algunas realizaciones, el Módulo de reconstrucción 250 puede usar la pose de cámara actualizada y la imagen asociada para determinar si se extiende un modelo tridimensional existente, que puede adoptar la forma de una malla tridimensional en color o una estructura de datos volumétricos.

[46] El término "extender" (y sus variantes), como se usa en el contexto de la reconstrucción de superficie, se refiere a la adición de datos recién adquiridos desde un modelo parcial del entorno a un modelo tridimensional existente del entorno. El término "actualización" (y sus variantes), tal como se utiliza en el contexto de la reconstrucción volumétrica, se refiere a la combinación de datos de profundidad recién adquiridos con un modelo tridimensional existente del entorno, que puede representarse mediante un conjunto de datos volumétricos.

[47] En algunas realizaciones, el Módulo de Reconstrucción 250 puede usar cualquier técnica de reconstrucción que no dependa de la introducción de una entrada de imagen de profundidad + color para cada cuadro. En algunas realizaciones, el Módulo de Reconstrucción 250 puede configurarse para usar una velocidad inferior fija para la entrada de datos de color + profundidad. En algunas realizaciones, la velocidad inferior fija para los datos de color + profundidad puede ser configurable basándose en los parámetros del sistema. Por ejemplo, en una realización, el Módulo de Reconstrucción 250 puede configurarse para recibir una entrada de imagen de color + profundidad después de cuatro cuadros de solo color (es decir, sin datos de profundidad). En otra realización, el Módulo de Reconstrucción 250 puede configurarse de modo que la imagen de profundidad adicional (o los datos de profundidad) puedan activarse por solicitud o proporcionarse a petición. En general, en función de los parámetros del sistema, la información de profundidad puede proporcionarse a una velocidad específica o bajo demanda. En consecuencia, el ahorro de energía puede resultar, en parte, de la menor frecuencia de uso del sensor de profundidad. Por ejemplo, con respecto a un sistema que funciona a una velocidad de cuadros de 30 cuadros por segundo, donde se proporcionan datos de profundidad con cada cuadro, si se proporciona información de profundidad al Módulo de Reconstrucción 250 una vez cada cinco cuadros, entonces, el uso del sensor de profundidad en el sistema 200 puede reducirse en un factor de 5. En algunas realizaciones, el Módulo de Reconstrucción 250 puede configurarse para usar una técnica basada en la reconstrucción de la superficie y/o una técnica basada en la reconstrucción volumétrica.

[48] El módulo de Reconstrucción 250, que puede tener imágenes previamente obtenidas con información de color y profundidad (por ejemplo, usando un sensor de profundidad en la MS 100), puede tener un modelo tridimensional existente basado en imágenes capturadas previamente con información de profundidad. Las imágenes capturadas con información de profundidad pueden representarse como una nube de puntos. Por ejemplo, la nube de puntos puede ser un conjunto de puntos de datos que representan la superficie externa de un objeto / entorno modelado. Los datos de datos pueden especificarse en algún sistema de coordenadas, por ejemplo, usando las coordenadas X, Y y Z. Una malla tridimensional, que puede consistir en triángulos, cuadriláteros u otros polígonos, se puede generar, por ejemplo, a partir de la nube de puntos.

[49] En realizaciones del Módulo de Reconstrucción 250, donde se aplica una técnica de reconstrucción de superficie, se puede obtener una malla tridimensional basada en la nube de puntos tridimensional, conectando algún subconjunto de puntos en la nube de puntos tridimensional para formar triángulos, cuadriláteros u otros polígonos basándose en varios criterios. Una malla tridimensional es una colección de vértices, bordes y caras que define la forma de un objeto poliédrico tridimensional. Las caras pueden ser triángulos, cuadriláteros u otros polígonos convexos simples, que pueden usarse para representar el objeto/entorno tridimensional. Se pueden usar varias técnicas bien conocidas y algoritmos de reconstrucción de malla para convertir nubes de puntos tridimensionales (que pueden obtenerse a partir de mapas de profundidad) en una malla tridimensional. En una realización, se puede usar una vecindad de mediciones de profundidad en el mapa de profundidad como un indicador para formar triángulos, conectando puntos vecinos con mediciones de profundidad similares.

[50] En realizaciones del Módulo de Reconstrucción 250, donde se aplica una técnica de reconstrucción volumétrica, la reconstrucción volumétrica fusiona los datos de profundidad del marco de la cámara en una estructura de datos volumétricos eficaz en términos de cálculo. En una realización, la reconstrucción volumétrica puede usar una representación implícita de la superficie usando una función tridimensional de distancia truncada con signo (TDSF). La TDSF tridimensional se puede representar como un conjunto de muestras regulares en el espacio tridimensional. En cada muestra, el valor de muestra da la distancia con signo a la superficie estimada. Las distancias positivas indican muestras fuera del objeto y las distancias negativas muestras dentro del objeto.

[51] En algunas realizaciones, el Módulo de Reconstrucción 250 puede usar la pose de cámara actualizada y la imagen asociada obtenida del Módulo de Seguimiento 210 para determinar si se extiende o actualiza el modelo tridimensional existente (malla tridimensional o estructura de datos volumétricos). En algunas realizaciones, se pueden usar técnicas de reconstrucción de superficie y/o volumétrica para el modelo tridimensional y las decisiones con respecto a las extensiones o actualizaciones del modelo se pueden tomar de una manera coherente con la técnica específica que se usa para el modelado.

[52] Por ejemplo, en un enfoque de reconstrucción de superficie, basado en la imagen y la pose de cámara recibida del Módulo de Seguimiento 120, el Módulo de Reconstrucción 250 puede determinar el grado en que la imagen incluye información que no está representada actualmente en un modelo existente o actual del entorno. En un enfoque de reconstrucción de superficie, si el modelo tridimensional del entorno que se está modelando ha de extenderse, por ejemplo, porque la imagen del Módulo de Seguimiento 120 incluye suficiente información nueva, entonces, se puede obtener información de imagen y profundidad encendiendo los sensores de profundidad adecuados conjuntamente con la captura de una imagen y la información de pose asociada. De lo contrario, los sensores de profundidad pueden ser inhabilitados. Por ejemplo, los sensores de profundidad pueden apagarse o ponerse en modalidad de espera.

[53] En algunas realizaciones que usan reconstrucción volumétrica, los datos de profundidad pueden adquirirse / medirse a alguna velocidad especificada (por ejemplo, una vez cada cinco cuadros). De lo contrario, los sensores de profundidad pueden ser inhabilitados. Por ejemplo, los sensores de profundidad pueden apagarse o ponerse en modalidad de espera.

[54] Si el modelo se va a actualizar, durante la reconstrucción volumétrica, se puede usar un mapa de profundidad basado en los datos de profundidad adquiridos/medidos para determinar las distancias a la superficie estimada desde un punto dado en el espacio, tal como la ubicación del centro de la cámara, que se puede determinar en función de la pose de la cámara. Estas distancias medidas pueden luego correlacionarse con la distancia con signo en el conjunto de datos volumétricos. Por ejemplo, la correlación puede ser realizada: (i) transformando cada muestra en el conjunto de datos volumétricos en el sistema de coordenadas de la cámara del mapa de profundidad y proyectando la muestra en el mapa de profundidad; (ii) obteniendo una medida de la distancia a la superficie a lo largo de un rayo como el valor de distancia en ese píxel proyectado (en el mapa de profundidad); (iii) obteniendo la distancia relativa medida desde el punto de muestra hasta la superficie, restando la distancia del punto de muestra en el rayo a la distancia medida obtenida en (ii) arriba; y (iv) actualizando la(s) estimación(es) de distancia proporcionada(s) por el punto de muestra en el rayo, combinando la(s) estimación(es) de distancia con la distancia relativa usando un filtro recursivo. Por ejemplo, una implementación puede usar la ponderación w con la estimación de la distancia s y actualizar la estimación de la distancia s basándose en una

$$s = \frac{(s * w) + m}{w + 1} \quad \text{y } w = w + 1.$$

nueva medición m usando una actualización ponderada, que puede estar dada por

[55] Para la reconstrucción volumétrica, en algunas realizaciones, se puede lograr una mayor calidad fusionando mediciones adicionales de puntos o áreas del entorno que ya están representadas en el modelo existente. Por consiguiente, en realizaciones en las que se usa la reconstrucción volumétrica, el modelo o malla tridimensional puede actualizarse incluso en situaciones en las que no hay nuevos puntos. En algunas realizaciones, se puede establecer un umbral u otros parámetros para limitar el número de veces que se actualiza el mismo elemento de superficie.

- 5 [56] La extracción de superficie también se puede realizar a partir de la representación de datos volumétricos, si se utiliza la reconstrucción volumétrica. Por ejemplo, un conjunto de datos volumétricos, eficaz en términos de cálculo, que comprende vóxeles puede usarse para representar una imagen volumétrica, donde los vóxeles pueden representar un valor en una cuadrícula en un espacio tridimensional. Varias técnicas bien conocidas, tales como los Cubos en Marcha, se pueden usar para extraer superficies (o isosuperficies) a partir del conjunto de datos volumétricos. Una isosuperficie es una representación tridimensional de la superficie de puntos con valores iguales en una distribución de datos tridimensional. En algunas realizaciones, la reconstrucción volumétrica puede usar una velocidad de datos fija / especificada (por ejemplo, 5 cuadros de color + profundidad por segundo), que puede configurarse dinámicamente. Por ejemplo, la frecuencia de las actualizaciones provenientes de la cámara de profundidad puede variar (aumentar o disminuir) en función de las solicitudes de la aplicación (por ejemplo, menor o mayor precisión del modelo solicitado) o en función de indicaciones contextuales (por ejemplo, nivel bajo o alto de batería en el dispositivo). En algunas realizaciones, la etapa de extracción de superficie puede realizarse independientemente de la fusión.
- 10
- 15 [57] Por consiguiente, en ambos enfoques anteriores, superficial y volumétrico, el módulo de reconstrucción tridimensional adquiere datos de profundidad esporádicamente, conservando así la potencia. En algunas realizaciones, el ahorro de potencia puede acumularse desde la inhabilitación, el apagado o la colocación de los sensores en un estado de espera en el que no consumen energía, o una energía muy mínima. Por ejemplo, si los sensores de profundidad permiten una reactivación rápida o casi instantánea, entonces, en una realización, la energía para los sensores activos puede apagarse. En otra realización, si los sensores pueden ponerse en una modalidad de espera donde consumen energía mínima para permitir una reactivación rápida y/o instantánea. Por ejemplo, las fuentes de luz estructuradas y/o luces estroboscópicas se pueden apagar cuando los sensores de profundidad se colocan en la modalidad de espera.
- 20
- 25 [58] En otras realizaciones, tales como cuando se usan cámaras estereoscópicas, se puede acumular el ahorro de energía a partir de una disminución en la frecuencia de las operaciones de estereofusión. Por ejemplo, en realizaciones en las que los sensores de profundidad adoptan la forma de sensores de visión estereoscópica, los sensores de visión estéreo pueden apagarse y el Módulo de seguimiento 210 puede continuar proporcionando imágenes monoculares de cámara e información de pose asociada. En algunas realizaciones, cuando los sensores de visión estereoscópica se apagan, los cálculos asociados al procesamiento de imágenes estereoscópicas, capturadas desde los sensores pasivos de visión estereoscópica, pueden desactivarse. Por ejemplo, en una realización, aunque las imágenes estereoscópicas pueden capturarse mediante un sensor pasivo de visión estereoscópica, los procesadores 150 pueden no procesar las imágenes estereoscópicas y, en su lugar, pueden seleccionar una única imagen del par estereoscópico para su procesamiento.
- 30
- 35 [59] La Figura 2A muestra las etapas en un proceso de reconstrucción tridimensional en ejecución 200. En algunas realizaciones, en la rutina 225, la pose de la cámara puede actualizarse. Por ejemplo, un algoritmo monocular de SLAM que se ejecuta en la MS 100 se puede usar para actualizar la pose de la cámara 110. Mediante el uso de SLAM monocular, la MS 100 puede construir una representación de su entorno a la vez que estima la pose o el movimiento de la cámara en relación con el entorno.
- 40
- 45 [60] En la rutina 225, se puede usar una sola cámara para obtener/actualizar información de pose, mientras que los sensores de profundidad, tales como luces estroboscópicas, sensores de visión estereoscópica y/o fuentes de luz estructuradas se pueden apagar, inhabilitar o poner en modalidad de espera, de modo que no consuman energía, o que consuman una energía mínima.
- 50 [61] En la etapa 230, pueden adquirirse nuevos datos de sensores que incluyen datos de imágenes en color sin información de profundidad e información de pose asociada, y el Módulo de Seguimiento 210 puede regresar a la etapa 225 para comenzar otra iteración y actualizar la pose de cámara. En general, el Módulo de Seguimiento puede usar cualquier procedimiento de seguimiento para determinar la pose. Por ejemplo, la entrada desde la IMU 130 se puede usar para determinar la pose.
- 55 [62] A continuación se describe un procedimiento ejemplar para implementar la rutina de actualización de pose de cámara 225, aunque, en la práctica, se puede usar cualquier procedimiento adecuado de seguimiento y actualización de pose. Para el seguimiento entre cuadros, el flujo óptico entre una imagen I inmediatamente precedente y una imagen actual J adquirida en la etapa 230 se puede calcular, por ejemplo, usando un algoritmo de Lucas-Kanade o variantes del mismo. El procedimiento de Lucas Kanade es un procedimiento diferencial de dos cuadros para la estimación de flujo óptico que puede usar gradientes de imagen y un enfoque iterativo para determinar los parámetros de movimiento.
- 60 [63] Si I y J son imágenes consecutivas de tamaño $n_x \times n_y$, donde $0 \leq x \leq n_x-1$ y $0 \leq y \leq n_y-1$, y $I(x, y)$ y $J(x, y)$ son los valores de píxel (por ejemplo, valores de escala de grises) de píxeles en el punto (x, y) ; luego, el seguimiento de características se puede usar para calcular la ubicación de un punto v en J , que corresponde al punto $u = [u_x, u_y]^T$ en I , donde $v = u + d = [u_x + d_x, u_y + d_y]^T$. El vector de desplazamiento $d = [d_x, d_y]^T$ se denomina el flujo óptico en u y el superíndice "T" se usa para referirse a la traspuesta de una matriz o vector. Un objetivo de Lucas-Kanade es encontrar el vector de desplazamiento d que minimice la función de error $\epsilon(d)$, donde,
- 65

$$\varepsilon(d) = \varepsilon(d_x, d_y) = \sum_{\chi = u_x - w}^{\chi = u_x + w} \sum_{\gamma = u_y - w}^{\gamma = u_y + w} [I(x, y) - J(x + v_x, y + v_y)]^2 \quad (1)$$

w es un número entero y representa una ventana de w x w píxeles de tamaño.

[64] En algunas realizaciones, se puede usar una implementación piramidal en tiempo real modificada del algoritmo de flujo óptico de Lucas-Kanade o variantes de la misma para calcular el flujo óptico bidimensional. Por ejemplo, se puede usar la variante de Alineamiento Inverso de Imágenes Compuestas del algoritmo de Lucas-Kanade.

[65] En la implementación piramidal, una imagen puede ser sometida sucesivamente a reducciones de tasa de muestreo para obtener una pirámide de imágenes de tasas reducidas de muestreo de diferentes resoluciones. Por ejemplo, el nivel más alto (nivel 0) de la pirámide puede tener la imagen en bruto o de máxima resolución y cada nivel inferior puede reducir la tasa de muestreo de la imagen en relación con el nivel inmediatamente superior, en algún factor. Por ejemplo, para una imagen I_0 de tamaño 640 X 480 (en el nivel 0), las imágenes I_1 , I_2 , I_3 e I_4 son de tamaños 320x240, 160x120, 80x60 y 40x30, respectivamente, donde el subíndice indica el nivel de imagen en la pirámide de imágenes.

[66] En el flujo óptico piramidal de Lucas-Kanade, el flujo óptico se calcula en el nivel piramidal más bajo L . Entonces, el resultado de ese cálculo se propaga al nivel superior $L-1$ en forma de una estimación inicial de desplazamiento de píxeles en el nivel $L-1$. Dada esa conjetura inicial, el flujo óptico refinado se calcula en el nivel $L-1$, y el resultado se propaga al nivel $L-2$ y así sucesivamente hasta el nivel 0 (la imagen original). Normalmente, el punto v en J correspondiente al punto u en I se busca dentro de una ventana (por ejemplo, de $w \times w$ píxeles de tamaño) alrededor de la ubicación $[u_x, u_y]$ en la imagen J . El flujo óptico basado en pirámide permite buscar en una ventana efectiva más grande y facilita el establecimiento de correspondencias de rasgos incluso frente a grandes movimientos de píxeles.

[67] En algunas realizaciones, el número de rasgos comparados entre imágenes durante el seguimiento se puede determinar basándose en las limitaciones del sistema. Por ejemplo, las resoluciones de las imágenes, la velocidad de los procesadores 150 y los criterios de rendimiento se pueden usar para determinar el número de rasgos comparados.

[68] En algunas realizaciones, en función de la ubicación de los rasgos coincidentes, se puede calcular una pose de cámara actualizada en la rutina 225. En algunas realizaciones, la pose y los datos de sensor adquiridos que incluyen las imágenes en color se pueden transferir periódicamente al Módulo de Reconstrucción 250, que espera actualizaciones de datos de seguimiento, proporcionadas por el Módulo de Seguimiento 210 en la etapa 255. En algunas realizaciones, el Módulo de Seguimiento 210 puede funcionar continuamente a una velocidad de cuadros de alrededor de 30 cuadros para capturar datos de sensor que incluyen imágenes monoculares y actualizar información de pose, que puede proporcionarse al Módulo de Reconstrucción 250.

[69] En algunas realizaciones, en la etapa 260, el Módulo de Reconstrucción 250 puede determinar si el modelo de reconstrucción existente ha de extenderse o actualizarse basándose en la imagen capturada actualizada y la información de pose asociada recibida en la etapa 255. En algunas realizaciones, el Módulo de Reconstrucción 250 puede usar cualquier técnica de reconstrucción que no dependa de la introducción de una entrada de imagen de profundidad + color para cada cuadro. Por ejemplo, las técnicas de superficie y/o volumétricas pueden ser utilizadas por el Módulo de Reconstrucción 250.

[70] En algunas realizaciones, para una reconstrucción de superficie, el Módulo de Reconstrucción 250 puede analizar cada cuadro proporcionado por el Módulo de Seguimiento 210 para determinar si el modelo ha de extenderse. En otra realización, el Módulo de Reconstrucción 250 puede seleccionar un subconjunto de cuadros entre los proporcionados por el Módulo de Seguimiento 210 para el análisis, en función de los ajustes de configuración, los parámetros del sistema y/o una frecuencia de cuadros actual del Módulo de Seguimiento 210 para determinar si el modelo ha de extenderse. En algunas realizaciones, la configuración del Módulo de Seguimiento 210 puede ser modificada por el Módulo de Reconstrucción 250 para fijar una velocidad de cuadros adecuada para la captura de imágenes.

[71] Además, para un enfoque de reconstrucción de superficie, en algunas realizaciones, la determinación de si el modelo de reconstrucción existente ha de extenderse o actualizarse puede basarse, en parte, en una imagen actual y en el grado en que la imagen actual incluye información que no está actualmente representada en el modelo existente.

- 5 **[72]** En algunas realizaciones, si hay puntos o áreas del entorno que no están representados actualmente en el modelo existente y que fueron incorporados a imágenes por primera vez en un marco de cámara recién capturado ("Y" en la etapa 260), entonces, el modelo o la malla tridimensional se puede extender o actualizar adquiriendo nuevos datos del sensor de profundidad o información de profundidad, en la etapa 265.
- 10 **[73]** Además, para un enfoque de reconstrucción de superficie, si hay puntos o áreas del entorno que ya están representados en el modelo existente, pero que se incorporaron a imágenes por primera vez desde un punto de vista diferente en un marco de cámara recién capturado ("Y" en la etapa 260), entonces, el modelo o malla tridimensional puede extenderse o actualizarse adquiriendo nuevos datos del sensor de profundidad o información de profundidad, en la etapa 265. En algunas realizaciones, la existencia de puntos nuevos puede determinarse calculando la diferencia entre la información incluida en la imagen actual y la información almacenada en el modelo existente, y determinando si tal diferencia no está vacía.
- 15 **[74]** Para el enfoque de reconstrucción de superficie, si el modelo de reconstrucción existente ha de extenderse ("SÍ" en la etapa 260), entonces, en la etapa 265, se puede activar el sensor de profundidad y se puede adquirir información de profundidad para aumentar el marco de imagen de cámara actualizada, recibido desde el Módulo de Seguimiento 210. Por ejemplo, si el marco de imagen de cámara actualizada, recibido desde el Módulo de Seguimiento 210 incluye puntos que fueron incorporados a imágenes por primera vez, entonces el algoritmo puede avanzar a la etapa 265 y los sensores de profundidad pueden habilitarse. En algunas realizaciones, se puede usar una cámara 3D TOF, luz estructurada o sensores estereoscópicos para obtener información de profundidad en la etapa 265.
- 20 **[75]** En el caso de que no haya puntos nuevos ("N" en la etapa 260), para la reconstrucción de superficie, el modelo no se puede extender y el algoritmo regresa a la etapa 255 para comenzar otra iteración. Por ejemplo, para la reconstrucción de superficie, si el nuevo marco de imagen de la cámara no incluye nuevos puntos, el algoritmo puede volver a la etapa 255 para esperar la próxima actualización y los sensores de profundidad pueden inhabilitarse o permanecer inhabilitados (si estaban previamente inhabilitados).
- 25 **[76]** En algunas realizaciones que usan reconstrucción volumétrica, se pueden adquirir datos de profundidad a alguna velocidad especificada (por ejemplo, una vez cada cinco marcos). De lo contrario, los sensores de profundidad pueden ser inhabilitados. Por ejemplo, los sensores de profundidad pueden apagarse o ponerse en modalidad de espera. Se puede usar un mapa de profundidad para determinar las distancias a la superficie estimada desde un punto dado en el espacio, tal como la ubicación del centro de la cámara. Por consiguiente, en algunas realizaciones que usan reconstrucción volumétrica, el sensor de profundidad se puede activar periódicamente (basándose en la velocidad especificada) y se puede adquirir información de profundidad para aumentar el cuadro de imagen de cámara actualizada recibido desde el Módulo de Seguimiento 210.
- 30 **[77]** Para la reconstrucción volumétrica, en algunas realizaciones, se puede lograr una mayor calidad fusionando mediciones adicionales de puntos o áreas del entorno que ya están representadas en el modelo existente. Por consiguiente, en realizaciones en las que se usa la reconstrucción volumétrica, el modelo o malla tridimensional puede actualizarse incluso en situaciones en las que no hay nuevos puntos. En algunas realizaciones, se puede establecer un umbral para limitar el número de veces que se actualiza el mismo elemento de superficie.
- 35 **[78]** En cualquiera de los enfoques, si la velocidad de cuadros del Módulo de Seguimiento 210 es suficientemente alta, entonces es poco probable que la pose de la cámara haya sufrido cambios (o puede haber cambiado mínimamente) desde el momento del cuadro actualizado más reciente (obtenido del Módulo de Seguimiento 210) y la captura de información de profundidad usando sensores de profundidad en la etapa 265. En consecuencia, la información de profundidad adquirida en la etapa 265, puede estar asociada al cuadro actualizado más reciente y a la pose asociada con una pérdida de precisión mínima o nula.
- 40 **[79]** En otra realización, se puede adquirir una nueva imagen que comprende tanto la imagen en color como la información de profundidad en la etapa 265 y se puede obtener una nueva pose de cámara para la nueva imagen de color + profundidad. En otra realización, puede adquirirse una nueva imagen que comprende tanto la imagen de color como la información de profundidad en la etapa 265, y la pose de cámara asociada al cuadro actualizado más reciente, obtenido desde el Módulo de Seguimiento 210, puede asociarse a la imagen de color + profundidad capturada recientemente. Como se ha indicado anteriormente, para una velocidad de cuadros suficientemente alta, el cambio de pose de cámara entre cuadros, desde el cuadro actualizado más reciente, obtenido del Módulo de Seguimiento 210, a la imagen de color + profundidad recién capturada puede ser mínimo, por lo que cualquier degradación de precisión puede ser mínima.
- 45 **[80]** En algunas realizaciones, se puede capturar una cámara 3D TOF, luz estructurada o imagen estereoscópica para obtener una imagen de color e información de profundidad en la etapa 265. En realizaciones en las que se utilizan sensores estereoscópicos pasivos, el par de imágenes estereoscópicas (por ejemplo, desde la imagen actualizada más reciente capturada por el Módulo de Seguimiento 210 usando los sensores estereoscópicos pasivos) ahora puede procesarse para obtener información de profundidad.
- 50 **[79]** En otra realización, se puede adquirir una nueva imagen que comprende tanto la imagen en color como la información de profundidad en la etapa 265 y se puede obtener una nueva pose de cámara para la nueva imagen de color + profundidad. En otra realización, puede adquirirse una nueva imagen que comprende tanto la imagen de color como la información de profundidad en la etapa 265, y la pose de cámara asociada al cuadro actualizado más reciente, obtenido desde el Módulo de Seguimiento 210, puede asociarse a la imagen de color + profundidad capturada recientemente. Como se ha indicado anteriormente, para una velocidad de cuadros suficientemente alta, el cambio de pose de cámara entre cuadros, desde el cuadro actualizado más reciente, obtenido del Módulo de Seguimiento 210, a la imagen de color + profundidad recién capturada puede ser mínimo, por lo que cualquier degradación de precisión puede ser mínima.
- 55 **[80]** En algunas realizaciones, se puede capturar una cámara 3D TOF, luz estructurada o imagen estereoscópica para obtener una imagen de color e información de profundidad en la etapa 265. En realizaciones en las que se utilizan sensores estereoscópicos pasivos, el par de imágenes estereoscópicas (por ejemplo, desde la imagen actualizada más reciente capturada por el Módulo de Seguimiento 210 usando los sensores estereoscópicos pasivos) ahora puede procesarse para obtener información de profundidad.
- 60 **[79]** En otra realización, se puede adquirir una nueva imagen que comprende tanto la imagen en color como la información de profundidad en la etapa 265 y se puede obtener una nueva pose de cámara para la nueva imagen de color + profundidad. En otra realización, puede adquirirse una nueva imagen que comprende tanto la imagen de color como la información de profundidad en la etapa 265, y la pose de cámara asociada al cuadro actualizado más reciente, obtenido desde el Módulo de Seguimiento 210, puede asociarse a la imagen de color + profundidad capturada recientemente. Como se ha indicado anteriormente, para una velocidad de cuadros suficientemente alta, el cambio de pose de cámara entre cuadros, desde el cuadro actualizado más reciente, obtenido del Módulo de Seguimiento 210, a la imagen de color + profundidad recién capturada puede ser mínimo, por lo que cualquier degradación de precisión puede ser mínima.
- 65 **[80]** En algunas realizaciones, se puede capturar una cámara 3D TOF, luz estructurada o imagen estereoscópica para obtener una imagen de color e información de profundidad en la etapa 265. En realizaciones en las que se utilizan sensores estereoscópicos pasivos, el par de imágenes estereoscópicas (por ejemplo, desde la imagen actualizada más reciente capturada por el Módulo de Seguimiento 210 usando los sensores estereoscópicos pasivos) ahora puede procesarse para obtener información de profundidad.

- 5 **[81]** Por consiguiente, si el cuadro de imagen de cámara actualizada, recibido desde el Módulo de Seguimiento 210 incluye puntos que se incorporaron a imágenes por primera vez (para la reconstrucción de superficie) o después del intervalo de cuadro adecuado, basándose en la velocidad de cuadros especificada (para reconstrucción volumétrica), entonces el algoritmo puede avanzar a la etapa 265 y los sensores de profundidad pueden habilitarse para obtener información de profundidad. Por lo tanto, en algunas realizaciones, si bien el Módulo de Seguimiento 210 puede funcionar continuamente, las operaciones intensivas en energía que implican la adquisición de profundidad se realizan solo esporádicamente en la etapa 265, conservando así la energía.
- 10 **[82]** El algoritmo puede entonces avanzar a la rutina 275, donde la reconstrucción tridimensional del entorno que se está modelando puede extenderse basándose en los datos de profundidad recién adquiridos.
- 15 **[83]** La figura 2B muestra un diagrama de flujo para la rutina o el procedimiento 275, para extender la reconstrucción usando una técnica de reconstrucción de superficie. En algunas realizaciones, la rutina 275 puede ser invocada por, o formar parte de, el Módulo de Reconstrucción 250. En algunas realizaciones, la rutina 275 puede aceptar, como entrada, la pose de cámara actualizada 304 y la imagen de color + profundidad 302 que comprende información de color y profundidad capturada en la etapa 265. En algunas realizaciones, la pose de cámara actualizada 304 puede ser la pose más reciente calculada por la rutina 225.
- 20 **[84]** En algunas realizaciones, en la etapa 305, la imagen de entrada de color + profundidad 302 puede ser filtrada optativamente para eliminar los componentes de ruido y sometida a reducción de frecuencia de muestreo. En algunas realizaciones, el nivel de disminución de la frecuencia de muestreo puede basarse, en parte, en la precisión del sensor de profundidad, la precisión de la estimación de pose de cámara 304 y / u otros parámetros del sistema. Por ejemplo, en una realización, una imagen de color + profundidad de 640x480 píxeles de color + profundidad
25 puede ser sometida a una reducción de frecuencia de muestreo de 320x240 píxeles de color + profundidad. En algunas realizaciones, las imágenes de color y profundidad se pueden someter a reducciones de frecuencia de muestreo de diferentes tasas. Por ejemplo, en una realización, la información de color de la imagen se puede mantener en una resolución completa mientras que la información de profundidad se puede muestrear a la baja.
- 30 **[85]** Por ejemplo, si una imagen de píxeles de color + profundidad de tamaño 640x 480 no fue muestreada a la baja en la etapa 305, entonces la malla parcial resultante, en la etapa 310, podría dar como resultado hasta $639 * 479 * 2 = 612.162$ triángulos. Si la imagen fue muestreada a la baja hasta 320x240, en la etapa 305, entonces, el número de triángulos no sería más que 152.482 triángulos. Por lo tanto, basándose en diversos parámetros del sistema, el muestreo a la baja puede usarse ventajosamente en la etapa 305, para reducir la complejidad de cálculo de la reconstrucción y acelerar el proceso con un impacto mínimo en la reconstrucción tridimensional, lo que puede
35 contribuir de tal modo a eficacias adicionales en el consumo de energía.
- 40 **[86]** En la etapa 310, para una reconstrucción de superficie, se puede generar una malla parcial basándose en la imagen de profundidad y la pose de cámara. En una realización, el módulo de reconstrucción de superficie puede usar la información de profundidad para extraer la imagen de color a lo largo de una dirección de visión y crear una malla tridimensional densa texturizada o coloreada. Por ejemplo, una malla puede estar compuesta de triángulos que pueden usarse para representar la imagen. El término "malla parcial" se refiere a una malla creada a partir de una única imagen de color + profundidad 302, que se obtuvo en la etapa 265. El término "malla existente" se refiere al resultado, en un momento en el tiempo, obtenido de la combinación o fusión de todas las mallas parciales entre sí.
- 45 **[87]** Cuando se crea la primera malla parcial, se usa automáticamente como la malla existente. Cuando se crea cualquier otra malla parcial posterior, la malla parcial posterior se fusiona con la malla existente. En algunas realizaciones, las mallas se crean a partir de cuadros clave con poses de cámara asociadas y puntos tridimensionales dispersos con descripciones que se originan a partir de los cuadros clave. Los cuadros clave, que incluyen información de profundidad, son imágenes seleccionadas que se capturan en poses de cámara que están relativamente distantes entre sí en términos de rotación y traslación de la cámara. En algunas realizaciones, un sistema de coordenadas global puede definirse basándose en la pose de cámara asociada al primer cuadro clave.
- 50 **[88]** En la etapa 320, se determina la superposición entre la malla existente y la malla parcial. En algunas realizaciones, la rutina 375 puede usar además la pose 6DOF 304 del sistema de seguimiento (calculada por la rutina 225) para alinear inicialmente una nueva malla parcial con la malla existente. En algunas realizaciones, en la etapa 325, según la precisión de la estimación de la pose del sistema de seguimiento, la rutina 375, optativamente, puede refinar adicionalmente la alineación utilizando procedimientos tales como el punto iterativo más cercano (ICP). El algoritmo ICP minimiza las diferencias entre la malla parcial y las secciones superpuestas de la malla existente.
60 En una realización, el algoritmo ICP puede realizar iterativamente las siguientes etapas para refinar la alineación de la malla parcial con partes superpuestas de la malla existente: asociar puntos vecinos más cercanos en las dos mallas, estimar parámetros de transformación usando una función de costo cuadrático medio, y luego transformar los puntos usando los parámetros estimados antes de comenzar la próxima iteración.
- 65 **[89]** A continuación, en la etapa 330, una vez que la malla parcial se ha alineado con la malla existente, el módulo de reconstrucción puede reducir cualquier detalle duplicado de la malla parcial que ya esté presente en la

5 malla existente. En algunas realizaciones, en la etapa 330, cualquier detalle duplicado puede eliminarse, o bien puede fundirse o fusionarse con la malla existente, por ejemplo, promediando todas las posiciones de vértices cercanas / correspondientes. En otra realización, los detalles duplicados pueden eliminarse, o bien pueden fundirse / fusionarse con el modelo existente basándose en otros varios métodos heurísticos. Por ejemplo, en una realización, la observación más reciente de un detalle podría mantenerse. En otra realización, los detalles duplicados pueden fundirse de una manera que preserve la tersura de la malla existente.

10 **[90]** En la etapa 335, se puede añadir información de textura o de color de vértices a la malla parcial, después de la eliminación de la información duplicada, usando la información en la imagen de color de entrada 302. En la etapa 340, la malla parcial se fusiona en la malla existente. En algunas realizaciones, la información duplicada puede eliminarse de la malla parcial antes de fusionar la malla parcial con la malla existente. En otra realización, como se ha descrito anteriormente, las mallas pueden fusionarse, sin eliminar información duplicada, por ejemplo, promediando todas las posiciones de vértices cercanas o correspondientes en las mallas parciales y existentes, o usando otros métodos heurísticos. En consecuencia, la malla existente puede crecer y volverse más completa. En algunas realizaciones, el enfoque de reconstrucción de superficie puede obtener información de profundidad a partir de los sensores de profundidad bajo demanda. Por ejemplo, un enfoque de reconstrucción de superficie puede adquirir información de profundidad bajo demanda, tal como una o dos veces por segundo al comienzo de una secuencia de correlación y solo cada pocos segundos hacia el final, dando como resultado un uso significativamente menor del sensor de profundidad y una disminución correspondiente en el consumo de energía.

20 **[91]** La Figura 2C muestra un diagrama de flujo para la rutina o el procedimiento 275 para extender la reconstrucción usando una técnica de reconstrucción volumétrica. En algunas realizaciones, la rutina 275 puede ser invocada por, o formar parte de, el Módulo de Reconstrucción 250. En algunas realizaciones, la rutina 275 puede aceptar, como entrada, la pose de cámara actualizada 304 y la imagen de color + profundidad 302 que comprende información de color y profundidad capturada en la etapa 265. En algunas realizaciones, la pose de cámara actualizada 304 puede ser la pose más reciente calculada por la rutina 225.

30 **[92]** En algunas realizaciones, en la etapa 305, la imagen de entrada de color + profundidad 302 puede ser filtrada optativamente para eliminar los componentes de ruido y sometida a reducción de frecuencia de muestreo. En algunas realizaciones, el nivel de disminución de la frecuencia de muestreo puede basarse, en parte, en la precisión del sensor de profundidad, la precisión de la estimación de pose de cámara 304 y / u otros parámetros del sistema. Por lo tanto, basándose en diversos parámetros del sistema, el muestreo a la baja puede usarse ventajosamente en la etapa 305, para reducir la complejidad de cálculo de la reconstrucción y acelerar el proceso con un impacto mínimo en la reconstrucción tridimensional, lo que puede contribuir de tal modo a eficacias adicionales en el consumo de energía.

40 **[93]** En la etapa 350, un conjunto de datos volumétricos puede generarse y/o actualizarse basándose en la imagen de profundidad y la pose de cámara. En algunas realizaciones, en la etapa 350, los datos de profundidad del cuadro de la cámara se fusionan en una estructura de datos volumétrica, eficaz en términos de cálculo. En una realización, la reconstrucción volumétrica puede usar una representación implícita de la superficie usando una función de distancia tridimensional.

50 **[94]** En realizaciones que usan reconstrucción volumétrica, se pueden adquirir datos de profundidad a alguna velocidad especificada (por ejemplo, una vez cada cinco cuadros). De lo contrario, los sensores de profundidad pueden ser inhabilitados. Por ejemplo, los sensores de profundidad pueden apagarse o ponerse en modalidad de espera. Se puede usar un mapa de profundidad para determinar las distancias a la superficie estimada desde un punto dado en el espacio, tal como la ubicación del centro de la cámara. Por consiguiente, en algunas realizaciones que usan reconstrucción volumétrica, el sensor de profundidad se puede activar periódicamente (basándose en la velocidad especificada) y se puede adquirir información de profundidad para aumentar el cuadro de imagen de cámara actualizada recibido desde el Módulo de Seguimiento 210.

55 **[95]** Para la reconstrucción volumétrica, en algunas realizaciones, se puede lograr una mayor calidad fusionando mediciones adicionales de puntos o áreas del entorno que ya están representadas en el modelo existente. Por consiguiente, en realizaciones en las que se usa la reconstrucción volumétrica, el conjunto de datos volumétricos puede actualizarse incluso en situaciones en las que no hay nuevos puntos. En algunas realizaciones, se puede establecer un umbral para limitar el número de veces que se actualiza el mismo elemento de superficie.

60 **[96]** En la etapa 355, la extracción de superficie se puede realizar a partir del conjunto de datos volumétricos / malla tridimensional usando varias técnicas. Por ejemplo, el algoritmo de Cubos en Marcha se puede usar para extraer la superficie. En algunas realizaciones, para una reconstrucción volumétrica, la extracción de la superficie se realiza en una etapa autónoma que es independiente de la etapa de fusión.

65 **[97]** En algunas realizaciones, la reconstrucción volumétrica puede usar una velocidad de datos fija (por ejemplo, 5 cuadros de color + profundidad por segundo), que puede configurarse dinámicamente. Por ejemplo, la frecuencia de las actualizaciones provenientes de la cámara de profundidad puede variar (aumentar o disminuir) en función de las solicitudes de la aplicación (por ejemplo, menor o mayor precisión del modelo solicitado) o en función

de indicaciones contextuales (por ejemplo, nivel bajo o alto de batería en el dispositivo). En algunas realizaciones, la etapa de extracción de superficie puede realizarse independientemente de la fusión.

[98] La Figura 2D muestra un diagrama de flujo que representa un procedimiento ejemplar para fusionar datos de profundidad desde el cuadro de la cámara al conjunto de datos volumétricos. Se puede usar una función de distancia truncada con signo (TDSF) tridimensional para representar como un conjunto de muestras normales en el espacio tridimensional. En cada muestra, el valor (de muestra) de TDSF produce la distancia con signo a la superficie estimada, donde las distancias positivas indican muestras fuera del objeto, mientras que las distancias negativas indican muestras dentro del objeto.

[99] En algunas realizaciones que usan reconstrucción volumétrica, se pueden adquirir datos de profundidad a alguna velocidad especificada (por ejemplo, una vez cada cinco cuadros) y se puede usar un mapa de profundidad para determinar distancias a la superficie estimada desde un punto dado en el espacio, tal como la ubicación del centro de la cámara. Estas distancias pueden correlacionarse luego con la distancia con signo en el conjunto de datos volumétricos.

[100] En la etapa 351, cada muestra en el volumen puede transformarse en el sistema de coordenadas de cámara del mapa de profundidad y la muestra puede proyectarse en el mapa de profundidad basándose en la pose de la cámara de entrada para obtener una distancia medida a la superficie a lo largo de un rayo. El valor de distancia en ese píxel proyectado proporciona una estimación de distancia medida a la superficie a lo largo del rayo.

[101] A continuación, en la etapa 353, se calcula la diferencia entre la profundidad de la muestra y la estimación de distancia medida. La distancia relativa medida desde la ubicación de la muestra hasta la superficie se define como la diferencia entre la profundidad de la muestra y la distancia medida.

[102] En la etapa 355, la distancia relativa medida se correlaciona con la TDSF y, en la etapa 357, el valor de TDSF correlacionado se puede combinar con el valor almacenado. Por ejemplo, la(s) estimación(es) de distancia proporcionada(s) por el punto de muestra en el rayo se puede(n) actualizar combinando la(s) estimación(es) de distancia con la distancia relativa usando un filtro recursivo. Por ejemplo, una implementación puede usar la ponderación w con la estimación de distancia s y actualizar la estimación de distancia s basándose en una nueva

medición m usando una actualización ponderada, que puede estar dada por
$$s = \frac{(s * w) + m}{w + 1} \quad \text{y } w = w + 1.$$

[103] En la etapa 359, si se han de procesar muestras adicionales ("Y" en la etapa 359), entonces puede comenzarse el inicio de otra iteración para procesar la siguiente muestra. Si no hay más muestras ("N" en la etapa 359), el control puede devolverse a la rutina de llamada.

[104] En algunas realizaciones, el Módulo de Seguimiento 210 y el Módulo de Reconstrucción 250 pueden ejecutarse en paralelo. Por ejemplo: el Módulo de Seguimiento 210 y el Módulo de Reconstrucción 250 pueden ejecutarse como hilos o procesos independientes. En algunas realizaciones, el Módulo de Seguimiento 210 y el Módulo de Reconstrucción 250 pueden ejecutarse en núcleos procesadores independientes.

[105] En algunas realizaciones, el procedimiento 200 descrito anteriormente puede facilitar la reconstrucción tridimensional en tiempo real con un uso eficaz, en términos energéticos, del sensor de profundidad. En una realización, si bien el sistema de seguimiento visual puede funcionar continuamente a velocidades de cuadro adecuadas, el módulo de reconstrucción tridimensional puede adquirir datos de imagen + profundidad esporádicamente. Por ejemplo, la velocidad a la que pueden adquirirse pares de imágenes de color + profundidad, si bien depende de la aplicación, puede ser mayor al comienzo de una secuencia de correlación y puede disminuir significativamente hacia el final. En algunas realizaciones, con relación a los procedimientos convencionales donde la información de profundidad se adquiere continuamente, el procedimiento 200 anterior puede ser capaz de reducir el número de cuadros que requieren información de profundidad, en un orden de magnitud o más, dando como resultado por ello significativos ahorros de energía.

[106] En algunas implementaciones ejemplares, el sistema de seguimiento visual puede funcionar continuamente a una velocidad de cuadro de 30 fps, mientras que el módulo de reconstrucción tridimensional adquiere datos de imagen + profundidad esporádicamente. En general, la velocidad a la que se adquieren las imágenes de color + profundidad depende del uso específico. Sin embargo, en algunos casos de uso típicos, el módulo de reconstrucción tridimensional puede adquirir datos de imagen + profundidad una o dos veces por segundo al comienzo de una secuencia de correlación, y solo una vez cada pocos segundos hacia el final de la secuencia de correlación. Por lo tanto, con respecto a las implementaciones convencionales, que requieren información de color + profundidad en todo momento, las realizaciones descritas en este documento reducen el número de cuadros para los que se adquiere información de profundidad, en un factor de 10 o más en relación con las implementaciones convencionales. Por consiguiente, en casos donde el sensor de profundidad puede habilitarse e inhabilitarse rápidamente y con una sobrecarga mínima, el consumo de energía para la adquisición de profundidad en

aplicaciones de CV en las realizaciones divulgadas puede ser un 10% o menos de la energía consumida por procedimientos tradicionales.

5 **[107]** La Figura 3 muestra un diagrama de flujo del procedimiento ejemplar 400 para la reconstrucción tridimensional congruente con las realizaciones divulgadas. En algunas realizaciones, el procedimiento 400 puede ser realizado por los procesadores 150. En algunas realizaciones, partes del procedimiento 400 pueden ser ejecutadas por el módulo de CV 155.

10 **[108]** En la etapa 405, puede capturarse una primera (o siguiente) imagen usando al menos una cámara, en donde la primera imagen comprende información de color para al menos una parte de un entorno que se está modelando. En algunas realizaciones, la primera imagen puede capturarse usando una o más cámaras 110 en la MS 100.

15 **[109]** A continuación, en la etapa 410, la información de pose de cámara para la primera (o la siguiente) imagen capturada por la cámara puede obtenerse en la etapa 420. En algunas realizaciones, partes de las etapas 405 y 410 pueden ser realizadas por el Módulo de Seguimiento 210.

20 **[110]** En la etapa 420, se puede tomar una determinación basándose, en parte, en la primera (o siguiente) imagen capturada y la información de pose de cámara asociada, ya sea para extender o actualizar un primer modelo tridimensional (3D) de un entorno modelado por el dispositivo móvil. En algunas realizaciones, se pueden usar técnicas de reconstrucción de superficie y/o volumétrica para el modelo tridimensional y las decisiones con respecto a la extensión del modelo se pueden tomar de una manera coherente con la técnica específica que se usa para el modelado.

25 **[111]** Por ejemplo, para la reconstrucción de superficie, basada en la primera (siguiente) imagen y pose de cámara, se puede tomar una determinación del grado en que la imagen incluye información que no está actualmente representada en un modelo existente del entorno. Por ejemplo, para la reconstrucción de la superficie, una determinación en cuanto a ampliar el modelo existente del entorno puede basarse en si hay puntos o áreas del entorno que ya estén representados en el modelo existente pero que se incorporaron a imágenes por primera vez desde un punto de vista diferente en un cuadro de cámara recientemente capturado.

30 **[112]** En la etapa 430, si el modelo no ha de extenderse o actualizarse ("N" en la etapa 430), entonces, en la etapa 440, el sensor de profundidad en la MS 100 se puede inhabilitar. Para una reconstrucción volumétrica, por ejemplo, el sensor de profundidad puede inhabilitarse para los intervalos adecuados entre cuadros, según la frecuencia con la que se está adquiriendo la información de profundidad. Para una reconstrucción de superficie, por ejemplo, el sensor de profundidad puede inhabilitarse si la imagen no incluye nueva información. Para la reconstrucción volumétrica, donde la información de profundidad se puede adquirir periódicamente a una velocidad especificada, los sensores de profundidad pueden habilitarse a intervalos adecuados en función de la velocidad especificada. De lo contrario, los sensores de profundidad pueden ser inhabilitados. Por ejemplo, los sensores de profundidad pueden apagarse o ponerse en modalidad de espera.

35 **[113]** En algunas realizaciones, tanto para la reconstrucción de superficie como para los casos de reconstrucción volumétrica, la inhabilitación del sensor de profundidad puede comprender al menos uno entre: apagar la energía suministrada al sensor de profundidad, o inhabilitar la funcionalidad relacionada con el cálculo de una imagen de profundidad. Por ejemplo, los sensores de profundidad pueden apagarse o ponerse en modalidad de espera. En algunas realizaciones, por ejemplo, cuando se inhabilita un sensor de visión estereoscópica, las imágenes capturadas por la cámara 110 pueden ser monoculares. Además, en algunas realizaciones, al inhabilitar el sensor de profundidad, también puede inhabilitarse el cálculo asociado al procesamiento de imágenes estereoscópicas capturadas a partir de sensores pasivos de visión estereoscópica. Por ejemplo, en una realización, aunque las imágenes estereoscópicas pueden capturarse mediante un sensor pasivo de visión estereoscópica, los procesadores 150 pueden no procesar las imágenes estereoscópicas y, en su lugar, pueden seleccionar una única imagen del par estereoscópico para su procesamiento. Por consiguiente, en algunas realizaciones, el módulo de reconstrucción tridimensional adquiere/usa datos de profundidad esporádicamente conservando así la energía.

45 **[114]** Por otra parte, si el modelo ha de extenderse o actualizarse ("Y" en la etapa 430), entonces, en la etapa 450, se puede habilitar el sensor de profundidad en la MS 100. Por ejemplo, para una reconstrucción de superficie, si se ha de extender el modelo tridimensional del entorno que se está modelando, porque la primera (siguiente) imagen incluye suficiente información nueva, entonces se puede obtener información de imagen y profundidad encendiendo los sensores de profundidad adecuados conjuntamente con la captura de una imagen y la información de pose asociada.

50 **[115]** Como otro ejemplo, para una reconstrucción volumétrica donde la información de profundidad se está adquiriendo a alguna velocidad especificada, entonces, el sensor de profundidad puede habilitarse basándose en la velocidad especificada y el modelo tridimensional del entorno puede actualizarse con nueva información de profundidad. Para la reconstrucción volumétrica, en algunas realizaciones, se puede lograr una mayor calidad fusionando mediciones adicionales de puntos o áreas del entorno que ya están representadas en el modelo

65

existente. Por consiguiente, en realizaciones en las que se usa la reconstrucción volumétrica, el modelo o malla tridimensional puede actualizarse incluso en situaciones en las que no hay nuevos puntos. En algunas realizaciones, se puede establecer un umbral u otros parámetros para limitar el número de veces que se actualiza el mismo elemento de superficie.

5 **[116]** Por ejemplo, en algunas realizaciones, se puede capturar una segunda imagen con la al menos una cámara. Además, la segunda imagen puede ser aumentada con información de profundidad proporcionada, al menos en parte, por el sensor de profundidad.

10 **[117]** Para una reconstrucción de superficie, un modelo parcial del entorno modelado por la MS puede fusionarse con el primer modelo tridimensional del entorno. El modelo parcial puede basarse, en parte, en la segunda imagen, y en la profundidad y la información de pose de cámara para la segunda imagen. En algunas realizaciones, la información común al primer modelo tridimensional y al modelo tridimensional parcial puede eliminarse del modelo tridimensional parcial antes de fusionar el modelo tridimensional parcial con el primer modelo tridimensional. En algunas realizaciones, el modelo parcial puede basarse en una segunda imagen aumentada de resolución inferior muestreada a la baja. En algunas realizaciones, el modelo parcial del entorno modelado por la MS puede alinearse con el primer modelo tridimensional del entorno basándose en la información de pose de cámara para la segunda imagen. En algunas realizaciones, la alineación del modelo tridimensional parcial con el primer modelo tridimensional puede refinarse usando el Punto Iterativo Más Cercano (ICP).

15 **[118]** Para una reconstrucción volumétrica, los datos de profundidad se pueden adquirir a alguna velocidad específica (por ejemplo, una vez cada cinco cuadros) y se puede usar un mapa de profundidad para determinar las distancias a la superficie estimada desde un punto dado del espacio, tal como la ubicación del centro de la cámara. Estas distancias pueden correlacionarse luego con la distancia con signo en el conjunto de datos volumétricos. Por consiguiente, cada muestra en el volumen puede transformarse en el sistema de coordenadas de cámara del mapa de profundidad y la muestra puede proyectarse en el mapa de profundidad basándose en la pose de cámara de entrada para obtener una distancia medida a la superficie a lo largo de un rayo. El valor de distancia en ese píxel proyectado proporciona una estimación de distancia medida a la superficie a lo largo del rayo.

20 **[119]** A continuación, se puede calcular la diferencia entre la profundidad de la muestra y la estimación de distancia medida. La distancia relativa medida desde la ubicación de la muestra hasta la superficie, que se define como la diferencia entre la profundidad de la muestra y la distancia medida, se puede correlacionar con la TDSF y el valor correlacionado de la TDSF se puede combinar con el valor almacenado. Por ejemplo, la(s) estimación(es) de distancia proporcionada(s) por el punto de muestra en el rayo se puede(n) actualizar combinando la(s) estimación(es) de distancia con la distancia relativa usando un filtro recursivo. Por ejemplo, una implementación puede usar la ponderación w con la estimación de la distancia s y actualizar la estimación de la distancia s basándose en una nueva medición m usando una actualización ponderada, que puede estar dada por

$$s = \frac{(s * w) + m}{w + 1} \quad y \quad w = w + 1.$$

30 **[120]** Si hay imágenes adicionales para procesar ("Y" en la etapa 460), entonces el proceso vuelve a la etapa 410 para comenzar otra iteración. Por ejemplo, si la reconstrucción tridimensional y/o la secuencia de correlación es incompleta, se puede tomar una determinación para adquirir imágenes adicionales y se inicia una nueva iteración en la etapa 405. De lo contrario, si no hay más imágenes ("N" en la etapa 460), el procedimiento puede finalizar.

35 **[121]** Las metodologías descritas en este documento pueden implementarse por diversos medios según la aplicación. Por ejemplo, para una implementación de firmware y/o software, las metodologías pueden implementarse con módulos (por ejemplo, procedimientos, funciones, etc.) que realizan las funciones descritas en el presente documento. Cualquier medio legible por máquina que realice instrucciones de forma tangible puede usarse en la implementación de las metodologías descritas en el presente documento.

40 **[122]** Por ejemplo, el código de software puede almacenarse en una memoria y ejecutarse mediante una unidad procesadora tal como los uno o más procesadores 150 en la MS 100. En algunas realizaciones, las funciones pueden almacenarse como una o más instrucciones o código en un medio legible por ordenador. Los ejemplos incluyen medios legibles por ordenador codificados con una estructura de datos y medios legibles por ordenador codificados con un programa de ordenador. Los medios legibles por ordenador incluyen medios físicos de almacenamiento por ordenador.

45 **[123]** Un medio de almacenamiento puede ser cualquier medio disponible al que se pueda acceder por un ordenador. A modo de ejemplo, y no de limitación, dichos medios legibles por ordenador pueden comprender RAM, ROM, EEPROM, CD-ROM u otro almacenamiento en disco óptico, almacenamiento en disco magnético u otros dispositivos de almacenamiento magnético, o cualquier otro medio que pueda usarse para almacenar código de programa deseado en forma de instrucciones o estructuras de datos y al que se puede acceder desde una ordenador; los discos, como se usan en el presente documento, incluyen el disco compacto (CD), el disco láser, el

disco óptico, el disco versátil digital (DVD), el disquete y el disco blu-ray, donde algunos discos usualmente reproducen datos magnéticamente, mientras que otros discos reproducen datos ópticamente con láser. Las combinaciones de lo que antecede también deberían incluirse dentro del alcance de los medios legibles por ordenador.

5
[124] Además del almacenamiento en un medio legible por ordenador, pueden proporcionarse instrucciones y/o datos como señales en medios de transmisión incluidos en un aparato de comunicación. Por ejemplo, un aparato de comunicación puede incluir un transceptor que tenga señales indicativas de instrucciones y datos. Las instrucciones y los datos están configurados para provocar que uno o más procesadores implementen las funciones esbozadas en las reivindicaciones. Es decir, el aparato de comunicación incluye medios de transmisión con señales indicativas de información para realizar las funciones divulgadas. En un primer momento, los medios de transmisión incluidos en el aparato de comunicación pueden incluir una primera parte de la información para realizar las funciones divulgadas, mientras que en un segundo momento los medios de transmisión incluidos en el aparato de comunicación pueden incluir una segunda parte de la información para realizar las funciones divulgadas.

10
15

REIVINDICACIONES

1. Un procedimiento en una estación móvil, MS, que comprende al menos una cámara y un sensor de profundidad, comprendiendo el procedimiento:
 - capturar una primera imagen con al menos una cámara, comprendiendo la primera imagen información de color sin información de profundidad para al menos una parte de un entorno modelado por la MS;
 - obtener información de pose de cámara para la primera imagen;
 - determinar, basándose, en parte, en la primera imagen y en la información de pose de cámara para la primera imagen, si se extiende, mediante un enfoque de reconstrucción de superficie, un primer modelo tridimensional (3D) del entorno mediante la fusión de un modelo parcial actual del entorno modelado por la MS con el primer modelo tridimensional del entorno; en donde la determinación se basa en el grado en que la primera imagen incluye información que no está actualmente representada en el primer modelo tridimensional y
 - habilitar el sensor de profundidad, cuando se extiende el primer modelo tridimensional, o
 - mantener la inhabilitación del sensor de profundidad, cuando el primer modelo tridimensional no se extiende.

2. El procedimiento de la Reivindicación 1, en el que habilitar el sensor de profundidad comprende al menos uno entre:
 - encender la energía suministrada al sensor de profundidad, o
 - habilitar la funcionalidad relacionada con el cálculo de una imagen de profundidad.

3. El procedimiento de la Reivindicación 1, en el que al habilitar el sensor de profundidad, el primer modelo tridimensional se extiende:
 - capturando una segunda imagen con la al menos una cámara;
 - aumentando la segunda imagen con información de profundidad proporcionada, al menos en parte, por el sensor de profundidad; y
 - obteniendo el modelo parcial actual basado, en parte, en la segunda imagen, y la información de profundidad y la pose de la cámara para la segunda imagen.

4. El procedimiento de la Reivindicación 3, en el que la información común para el primer modelo tridimensional y el modelo parcial se elimina del modelo parcial antes de fusionar el modelo parcial con el primer modelo tridimensional.

5. El procedimiento de la Reivindicación 3, en el que el modelo parcial se basa en una segunda imagen aumentada de resolución inferior muestreada a la baja.

6. El procedimiento de la Reivindicación 3, en el que fusionar el modelo parcial del entorno modelado por la MS con el primer modelo tridimensional del entorno, comprende además alinear el modelo parcial con el primer modelo tridimensional basándose en la información de pose de cámara para la segunda imagen.

7. El procedimiento de la Reivindicación 6, en el que la alineación del modelo parcial con el primer modelo tridimensional se refina usando el Punto Iterativo Más Cercano, ICP.

8. Un aparato que comprende:
 - medios de formación de imágenes, siendo los medios de formación de imágenes para capturar una primera imagen que comprende información de color sin información de profundidad,
 - medios de detección de profundidad acoplados a los medios de formación de imágenes, y
 - medios de procesamiento acoplados a los medios de detección de profundidad y a los medios de formación de imágenes, comprendiendo además los medios de procesamiento:
 - medios para obtener información de pose de medios de formación de imágenes para la primera imagen,

- 5 medios para determinar si se extiende, mediante un enfoque de reconstrucción de superficie, un primer modelo tridimensional (3D) de un entorno basándose, en parte, en la primera imagen y en la información de pose de medios de formación de imágenes para la primera imagen, en donde se extiende el primer modelo tridimensional fusionando un modelo parcial actual del entorno modelado por el aparato con el primer modelo tridimensional del entorno, en donde la determinación se basa en el grado en que la primera imagen incluye información que no está actualmente representada en el primer modelo tridimensional; y
- 10 medios para habilitar los medios de detección de profundidad, si se extiende el primer modelo tridimensional; o
- medios para mantener la inhabilitación del sensor de profundidad, cuando el primer modelo tridimensional no se extiende.
- 15 9. El aparato de la Reivindicación 8, en el que los medios para habilitar los medios de detección de profundidad comprenden además al menos uno entre:
- medios para encender la energía suministrada a los medios sensores de profundidad, o
- 20 medios para habilitar la funcionalidad relacionada con el cálculo de una imagen de profundidad.
10. El aparato de la Reivindicación 8, en el que, al habilitar los medios sensores de profundidad:
- 25 el medio de formación de imágenes captura adicionalmente una segunda imagen que comprende información de color;
- el medio de detección de profundidad aumenta adicionalmente la segunda imagen con información de profundidad; y
- 30 en el que el medio de procesamiento comprende además:
- medios para obtener la segunda imagen aumentada que comprende información de color y profundidad,
- 35 medios para obtener el modelo parcial actual basándose, en parte, en la segunda imagen aumentada, y en la información de pose de los medios de formación de imágenes para la segunda imagen.
11. El aparato de la Reivindicación 10, en el que los medios de procesamiento comprenden además:
- 40 medios para eliminar información, común al primer modelo tridimensional y al modelo parcial, del modelo parcial antes de fusionar el modelo parcial con el primer modelo tridimensional.
12. El aparato de la Reivindicación 10, en el que los medios para fusionar el modelo parcial del entorno modelado por el aparato con el primer modelo tridimensional del entorno comprenden además:
- 45 medios para alinear el modelo parcial con el primer modelo tridimensional basándose en la información de pose de los medios de formación de imágenes, asociada a la segunda imagen.
13. El aparato de la Reivindicación 10, en el que los medios para fusionar el modelo parcial del entorno con el primer modelo tridimensional del entorno, que comprenden además:
- 50 medios para reducir la frecuencia de muestreo de la segunda imagen aumentada hasta una segunda imagen aumentada de menor resolución, y
- 55 en el que los medios para la fusión operan sobre la segunda imagen aumentada de resolución inferior muestreada a la baja.
14. Un medio no transitorio legible por ordenador que comprende instrucciones que, cuando son ejecutadas por un procesador, realizan las etapas de acuerdo a cualquiera de las reivindicaciones 1 a 7.
- 60

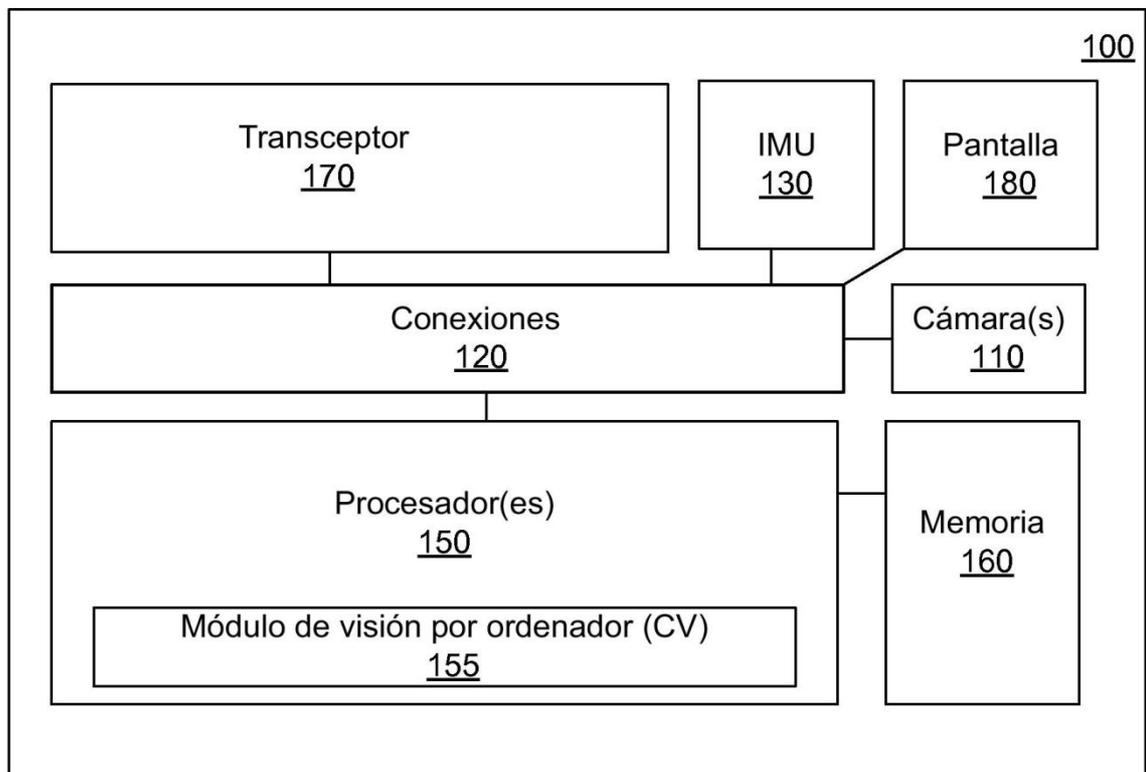


Fig. 1

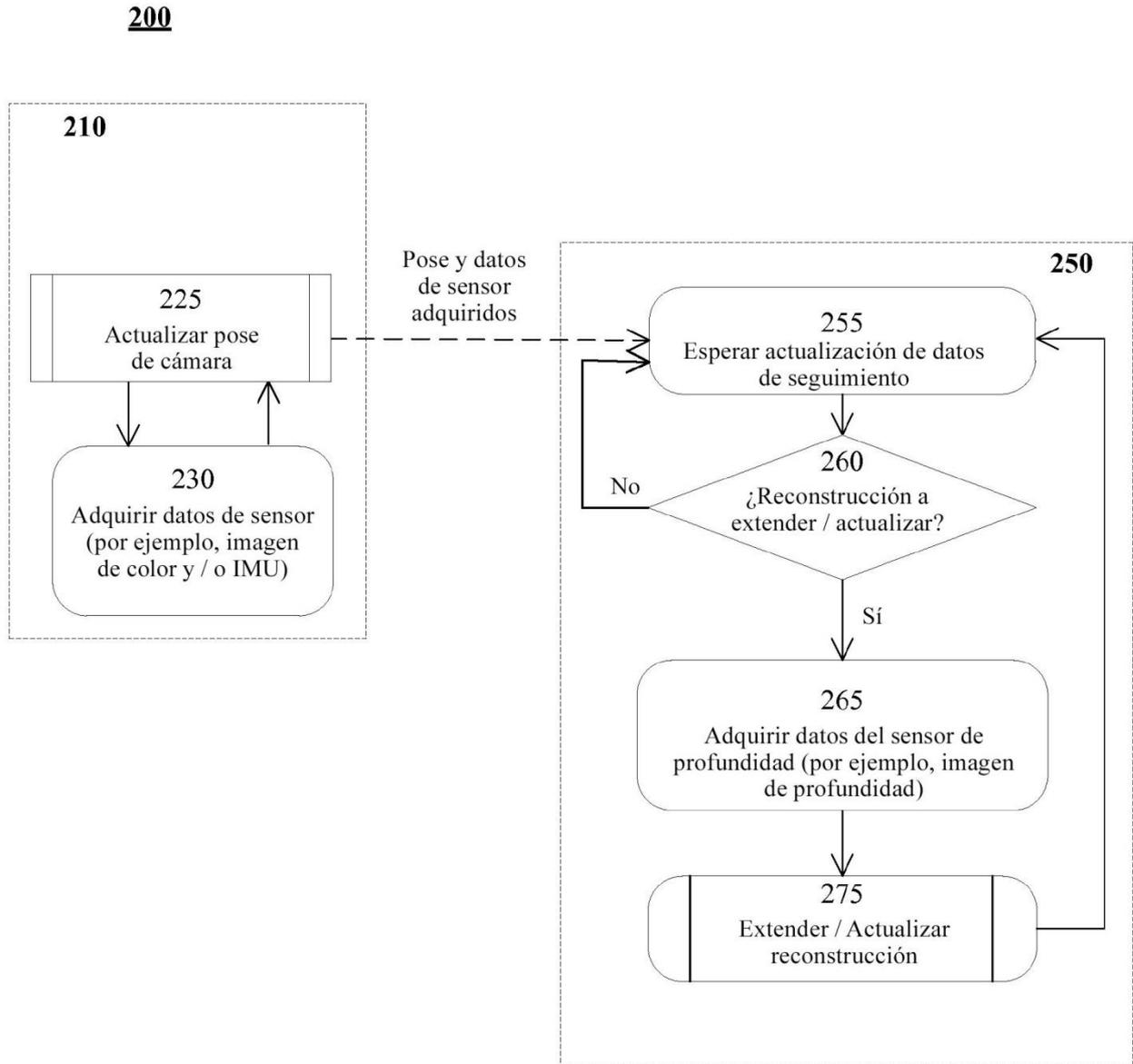


Fig. 2A

275

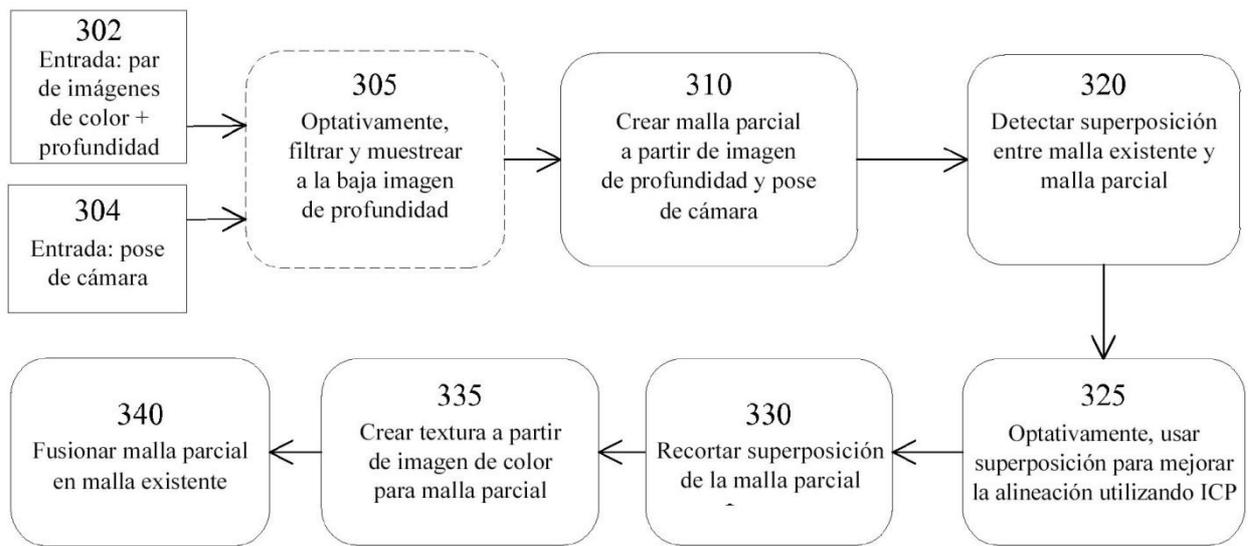


Fig. 2B

275

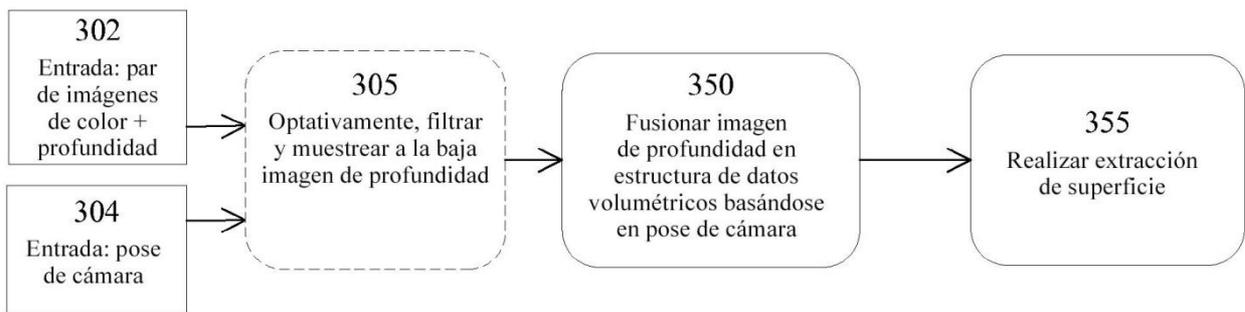


Fig. 2C

350

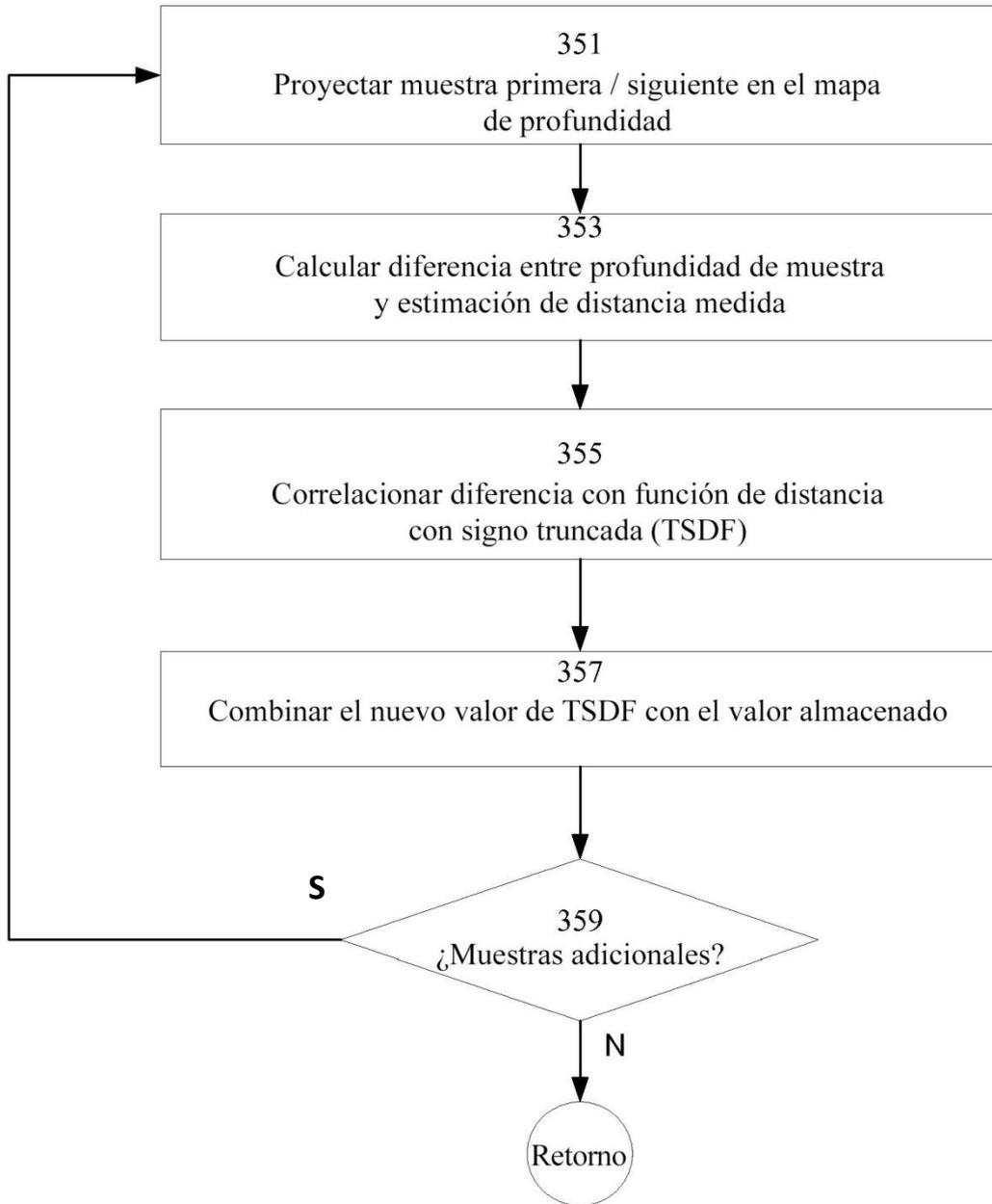


Fig. 2D

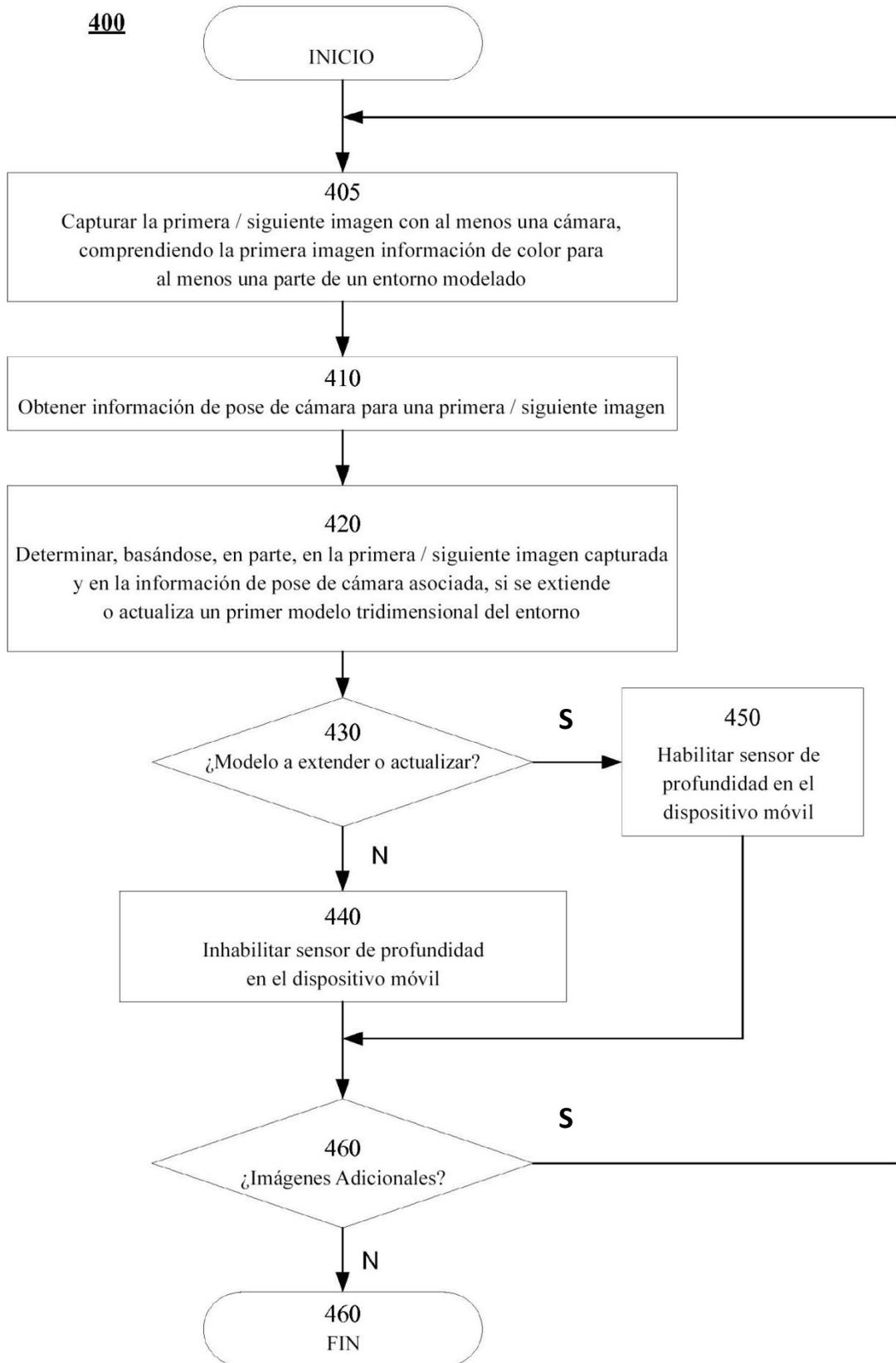


Fig. 3