

19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 989 245**

51 Int. Cl.:

G06T 7/11 (2007.01)

G06T 7/136 (2007.01)

G06T 7/174 (2007.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

86 Fecha de presentación y número de la solicitud internacional: **16.08.2021** **PCT/EP2021/072732**

87 Fecha y número de publicación internacional: **17.02.2022** **WO22034245**

96 Fecha de presentación y número de la solicitud europea: **16.08.2021** **E 21759316 (9)**

97 Fecha y número de publicación de la concesión europea: **09.10.2024** **EP 4172934**

54 Título: **Preprocesamiento basado en el movimiento de datos de imágenes bidimensionales previo al seguimiento de objetos tridimensionales con sincronización de tiempo virtual**

30 Prioridad:

14.08.2020 US 202063065872 P

45 Fecha de publicación y mención en BOPI de la traducción de la patente:
25.11.2024

73 Titular/es:

**TOPGOLF SWEDEN AB (100.0%)
Svärdvägen 3A
182 33 Danderyd, SE**

72 Inventor/es:

**FORSGREN, DANIEL y
NORHEIM, STEIN**

74 Agente/Representante:

IZQUIERDO BLANCO, María Alicia

ES 2 989 245 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

Preprocesamiento basado en el movimiento de datos de imágenes bidimensionales previo al seguimiento de objetos tridimensionales con sincronización de tiempo virtual

ANTECEDENTES

Esta memoria descriptiva se refiere al seguimiento de un objeto en movimiento, tal como una bola de golf en vuelo, utilizando datos obtenidos de diferentes sensores, que pueden emplear diferentes tecnologías de sensores.

Los sistemas y métodos para seguir el vuelo de un golpe de golf con sensores incluyen monitores de lanzamiento, seguimiento bidimensional (2D) de vuelo completo y seguimiento tridimensional (3D) de vuelo completo. Los tipos de sensores más utilizados son las cámaras, el radar Doppler y el radar de antena en fase. El método de monitorización del lanzamiento se basa en la medición de un conjunto de parámetros que pueden observarse durante el swing del palo de golf y las primeras pulgadas de vuelo de la bola después de que el palo haya impactado contra la bola. A continuación, los parámetros medidos se utilizan para extrapolar el vuelo previsto de la bola usando modelos matemáticos y físicos.

Por el contrario, los sistemas de seguimiento 3D de vuelo completo se caracterizan por un diseño que intenta seguir el vuelo completo del golpe de golf, en lugar de extrapolar los parámetros de lanzamiento. Además, los sistemas de seguimiento 2D de vuelo completo siguen la forma de un golpe de golf, visto desde un ángulo concreto, pero no producen información 3D y, por lo general, no pueden utilizarse para determinar parámetros clave, tal como la distancia recorrida por la bola. El seguimiento tridimensional de vuelo completo mediante una combinación de datos de cámara y radar Doppler se ha descrito en la patente de EE. UU. n.º 10.596.416. Por último, se ha descrito la posibilidad de utilizar en algunos contextos para el seguimiento tridimensional de objetos el seguimiento tridimensional de vuelo completo usando cámaras estereoscópicas que tienen sus adquisiciones de imágenes sincronizadas entre sí.

El artículo "Visualization of putting trajectories in live golf broadcasting" de MASAKI TAKAHASHI ET AL, publicado en SIGGRAPH'19, DOI 10.1145/3306307.3328148, divulga la superposición de trayectorias de putts en retransmisiones de golf en directo. Este sistema determina posiciones 2D de la bola que convierte en posiciones 3D con un modelo de ondulación del green, antes de dibujar las posiciones y trayectorias 3D de la bola en imágenes en directo.

SUMARIO

Esta memoria descriptiva describe tecnologías relacionadas con el seguimiento de un objeto en movimiento, tal como una bola de golf en vuelo, utilizando datos obtenidos de al menos una cámara.

En general, uno o más aspectos de la materia objeto descrita en la presente memoria descriptiva pueden materializarse en uno o más métodos que incluyen: recibir, en uno o más primeros ordenadores, fotogramas de imagen de una cámara a través de un primer canal de comunicaciones que acopla la cámara con el uno o más primeros ordenadores, el primer canal de comunicaciones teniendo un primer ancho de banda de datos; identificar, por el uno o más primeros ordenadores, ubicaciones de interés en los fotogramas de imagen; encontrar secuencias de las ubicaciones identificadas en los fotogramas de imagen, en donde cada una de las secuencias satisface un criterio de movimiento para ubicaciones identificadas en al menos tres fotogramas de imagen de la cámara; enviar datos de salida para las secuencias de las ubicaciones, en donde los datos de salida incluyen, para cada ubicación en cada secuencia, una posición bidimensional de la ubicación en un fotograma de imagen específico que tiene una marca de tiempo; recibir, en uno o más segundos ordenadores, los datos de salida del uno o más primeros ordenadores a través de un segundo canal de comunicaciones que acopla el uno o más primeros ordenadores con el uno o más segundos ordenadores, teniendo el segundo canal de comunicaciones un segundo ancho de banda de datos que es menor que el primer ancho de banda de datos; procesar, mediante el uno o más segundos ordenadores, al menos una de las secuencias en los datos de salida interpolando entre posiciones bidimensionales especificadas en fotogramas de imagen específicos para al menos una de las secuencias, utilizando las marcas de tiempo de los fotogramas de imagen específicos, para producir una posición bidimensional virtual en un punto predeterminado en el tiempo; y construir una pista tridimensional de una bola en movimiento en un espacio tridimensional utilizando la posición bidimensional virtual y la información de posición obtenida del al menos otro sensor para el punto predeterminado en el tiempo.

La detección y el envío pueden ser realizados por el uno o más segundos ordenadores, y las ubicaciones identificadas en los fotogramas de imagen pueden ser recibidas en el uno o más segundos ordenadores desde el uno o más primeros ordenadores a través del segundo canal de comunicaciones. Alternativamente, la detección y el envío pueden ser realizados por el uno o más primeros ordenadores, y los datos de salida pueden ser recibidos en el uno o más segundos ordenadores desde el uno o más primeros ordenadores a través del segundo canal de comunicaciones.

La detección puede incluir la formación de árboles enraizados a partir de las ubicaciones de interés, incluyendo: establecer nodos raíz de los árboles enraizados a partir de las respectivas primeras ubicaciones de interés

identificadas en respuesta a que cada una de las primeras ubicaciones de interés identificadas tenga valores de datos de imagen que satisfacen un criterio de inicio de árbol; añadir segundas ubicaciones de interés identificadas como subnodos de los árboles enraizados en respuesta a que al menos algunas respectivas de las segundas ubicaciones identificadas están dentro de un umbral de distancia de una ubicación identificada en un fotograma de imagen anterior que se ha añadido al por lo menos uno de los árboles enraizados; confirmar cada secuencia respectiva de ubicaciones identificadas para la salida cuando el árbol enraizado de la secuencia tiene una profundidad de árbol mayor que dos.

El envío puede incluir el retardo de la salida de datos para un fotograma de imagen dado y sus ubicaciones de interés encontradas en una o más de las secuencias, hasta que no se puedan incluir más ubicaciones de interés identificadas para el fotograma de imagen dado en ninguna de las secuencias basadas en ubicaciones de interés identificadas en fotogramas de imagen posteriores. El envío puede incluir: enviar datos para los fotogramas de imagen a medida que se completa la identificación para cada fotograma de imagen respectivo; y enviar datos para cada ubicación de interés sólo después de descubrir que una o más de las secuencias incluyen la ubicación de interés que se va a enviar.

La cámara puede incluir una cámara de obturador rodante, los datos de salida pueden incluir un valor de desfase de tiempo para cada ubicación de interés incluida en cada secuencia, y el procesamiento puede incluir: calcular un primer tiempo de observación para una primera ubicación que tenga una de las posiciones bidimensionales especificadas en los fotogramas de imagen específicos añadiendo un primer valor de desfase de tiempo para la primera ubicación a la marca de tiempo de uno de los fotogramas de imagen específicos; calcular un segundo tiempo de observación para una segunda ubicación que tenga otra de las posiciones bidimensionales especificadas en los fotogramas de imagen específicos añadiendo un segundo valor de desfase de tiempo para la segunda ubicación a la marca de tiempo de uno de los segundos fotogramas de imagen específicos; y realizar la interpolación utilizando el primer tiempo de observación y el segundo tiempo de observación.

La construcción puede incluir: combinar, mediante el uno o más segundos ordenadores, la posición bidimensional virtual con la información de posición obtenida del al menos otro sensor para formar una posición tridimensional de un objeto de interés; añadir, por el uno o más segundos ordenadores, la posición tridimensional del objeto de interés a otras posiciones tridimensionales de objetos de interés en una nube de posiciones tridimensionales de objetos de interés para el punto predeterminado en el tiempo; realizar, mediante el uno o más segundos ordenadores, un análisis de movimiento a través de múltiples nubes de posiciones tridimensionales para construir la pista tridimensional de la bola en movimiento en el espacio tridimensional, en donde cada una de las múltiples nubes es para un único punto en el tiempo, y las múltiples nubes incluyen la nube de posiciones tridimensionales de objetos de interés para el punto predeterminado en el tiempo; y mostrar la pista tridimensional de la bola en movimiento en el espacio tridimensional.

La cámara puede ser una primera cámara, el al menos otro sensor puede ser una segunda cámara, la información de posición puede incluir múltiples posiciones bidimensionales obtenidas de la segunda cámara, y la combinación puede incluir: excluir al menos una, pero no todas las múltiples posiciones bidimensionales obtenidas de la segunda cámara como no capaces de formar un punto tridimensional con la posición bidimensional virtual obtenida de la primera cámara; triangular al menos la posición tridimensional del objeto de interés utilizando la posición bidimensional virtual obtenida de la primera cámara, al menos una de las múltiples posiciones bidimensionales obtenidas de la segunda cámara, datos de calibración intrínsecos para la primera cámara y la segunda cámara, y datos de calibración extrínsecos para la primera y la segunda cámaras.

La exclusión puede incluir: determinar una región alrededor de al menos una porción de una línea epipolar en un plano de imagen de la segunda cámara utilizando la posición bidimensional virtual, un centro óptico de la primera cámara, un centro óptico de la segunda cámara, un valor de referencia entre la primera y segunda cámaras, y los datos de calibración extrínseca para la primera y la segunda cámaras; y rechazar emparejamientos de la posición bidimensional virtual obtenida de la primera cámara con las respectivas de las posiciones bidimensionales múltiples obtenidas de la segunda cámara en respuesta a que las respectivas posiciones bidimensionales múltiples estén fuera de la región alrededor de al menos una porción de la línea epipolar en el plano de imagen de la segunda cámara.

El uno o más primeros ordenadores pueden incluir una primera unidad de procesamiento y al menos una unidad de procesamiento adicional, el primer canal de comunicaciones acopla la cámara con la primera unidad de procesamiento, la recepción de los fotogramas de imagen puede incluir la recepción de los fotogramas de imagen en la primera unidad de procesamiento, la identificación de las ubicaciones de interés puede incluir la identificación de las ubicaciones de interés en la primera unidad de procesamiento, encontrar las secuencias puede incluir encontrar las secuencias en la al menos una unidad de procesamiento adicional en respuesta a la recepción de las ubicaciones desde la primera unidad de procesamiento a través de un tercer canal de comunicaciones que acopla la primera unidad de procesamiento con la al menos una unidad de procesamiento adicional, y enviar los datos de salida puede incluir enviar los datos de salida desde la al menos una unidad de procesamiento adicional, y en donde el tercer canal de comunicaciones tiene un tercer ancho de banda de datos que es menor que el primer ancho de banda de datos pero mayor que el segundo ancho de banda de datos.

Uno o más aspectos de la materia objeto descrita en la presente memoria descriptiva pueden materializarse en uno o más sistemas que incluyen: al menos un sensor que incluye una cámara y uno o más primeros ordenadores que incluyen un primer procesador de hardware y una primera memoria acoplada con el primer procesador de hardware, la primera memoria codificando instrucciones configuradas para hacer que el primer procesador de hardware realice primeras operaciones que incluyen la recepción de fotogramas de imagen, la identificación de ubicaciones de interés, la búsqueda de secuencias y el envío de datos de salida, de acuerdo con los métodos descritos en este documento; al menos otro sensor; y uno o más segundos ordenadores que incluyen un segundo procesador de hardware y una segunda memoria acoplada con el segundo procesador de hardware, la segunda memoria que codifica instrucciones configuradas para hacer que el segundo procesador de hardware realice segundas operaciones que incluyen la recepción de los datos de salida, el procesamiento de las secuencias y la construcción de una pista tridimensional, de acuerdo con los métodos descritos en este documento.

El al menos otro sensor puede incluir un dispositivo de radar. El al menos otro sensor puede incluir una segunda cámara. Además, uno o más aspectos de la materia objeto descrita en la presente memoria descriptiva pueden incorporarse en uno o más medios no transitorios legibles por ordenador que codifican instrucciones que hacen que el aparato de procesamiento de datos asociado con una cámara realice operaciones de acuerdo con los métodos descritos en este documento.

Varias realizaciones de la materia objeto descrita en esta memoria descriptiva pueden implementarse para obtener una o varias de las siguientes ventajas. La detección de objetos puede realizarse cerca de la cámara con datos de imagen en bruto (sin comprimir), lo que facilita el uso de una cámara de mayor resolución, potencialmente con una mayor frecuencia de fotogramas y/o una mayor profundidad de bits, lo que permite un seguimiento 3D de mayor calidad. Los datos de ubicación de la bola pueden comprimirse eficazmente para reducir los requisitos de ancho de banda para el envío de datos que se utilizarán en el seguimiento 3D, sin perder información relevante para el seguimiento 3D posterior de alta calidad. Las posibles trayectorias de un objeto pueden representarse mediante árboles enraizados (gráficos acíclicos conectados, cada uno de los cuales tiene un nodo raíz) y estos árboles enraizados pueden utilizarse para (en efecto) eliminar el ruido exportando únicamente los nodos de los árboles que pertenecen a ramas de una determinada profundidad.

Además, las restricciones utilizadas en el seguimiento 2D en datos de imagen pueden relajarse, tanto en términos de prefiltrado para identificar bolas candidatas como en términos de modelado del movimiento esperado de una bola en el espacio 2D, con el fin de pasar más datos utilizables al rastreador 3D sin saturar el ancho de banda de la conexión de comunicaciones con el ordenador del rastreador 3D. El componente de seguimiento 3D posterior puede diseñarse para manejar grandes cantidades de falsos positivos, proporcionando así una buena capacidad para filtrar el ruido y encontrar los objetos reales a seguir (p. ej., las bolas de golf). Con este filtrado de falsos positivos por parte del componente de seguimiento 3D, pueden simplificarse y relajarse sustancialmente las restricciones en el rastreador 2D, proporcionando el beneficio de hacer el rastreador 2D más fácil de escribir y mantener en comparación con un rastreador 2D que utiliza restricciones más estrictas que tienen que ser calibradas para producir pocos falsos positivos, pero aun así encontrar todos los verdaderos positivos.

Además, la separación de la tarea de seguimiento 3D en varios procesos secuenciales, en los que la comunicación principal entre procesos fluye en una sola dirección, y en los que cada etapa de procesamiento reduce el ancho de banda necesario para los componentes posteriores, y cada proceso puede ejecutarse en un ordenador independiente, proporciona una flexibilidad sustancial a la hora de diseñar y desplegar un sistema de seguimiento de movimiento de objetos 3D, especialmente si la distancia entre cámaras y/o recursos informáticos es considerable. Además, el procesamiento previo realizado para identificar las bolas candidatas y modelar el movimiento esperado de una bola en el espacio 2D permite la sincronización temporal virtual (posterior a la captura de imágenes) de las posiciones medidas del objeto en el tiempo y el espacio, evitando así la necesidad de sincronizar realmente las imágenes de la cámara con otro u otros sensores en el punto de captura de datos. Es posible la triangulación entre los puntos de diferentes sensores (p. ej., diferentes cámaras) aunque la captura original no fuera sincrónica. Por último, la sincronización temporal virtual está habilitada para las cámaras de obturador rodante, lo que permite una triangulación de alta calidad en un segundo ordenador (durante una fase posterior al procesamiento) utilizando datos tanto de la cámara o cámaras de obturador rodante como de la cámara o cámaras de obturador global.

Los detalles de una o más realizaciones de la materia objeto descrita en la presente memoria descriptiva se exponen en los dibujos adjuntos y en la descripción que figura a continuación. Otras características, aspectos y ventajas de la invención se desprenderán de la descripción, los dibujos y las reivindicaciones.

BREVE DESCRIPCIÓN DE LOS DIBUJOS

La figura 1A muestra un ejemplo de un sistema que realiza un procesamiento previo basado en el movimiento de datos de imágenes bidimensionales (2D) seguido de un seguimiento tridimensional (3D) de un objeto en movimiento a través de un espacio tridimensional.

Las figuras 1B-1D muestran ejemplos de diferentes configuraciones de sensores y ordenadores, como se pueden utilizar en el sistema de la figura 1A.

La figura 2 es un diagrama esquemático de un sistema de procesamiento de datos que incluye un aparato de procesamiento de datos.

La figura 3 muestra un ejemplo de procesos realizados en distintos ordenadores para detectar objetos, rastrearlos en 2D, producir posiciones virtuales en 2D para sincronizar el tiempo y construir seguimientos en 3D de los objetos en movimiento.

Las figuras 4A-4D muestran ejemplos de un proceso que encuentra secuencias de ubicaciones de objetos que satisfacen un criterio de movimiento.

La figura 5 muestra un ejemplo de un proceso que interpola entre posiciones 2D especificadas en fotogramas de imagen específicos obtenidos de una cámara con obturador rodante.

La figura 6A muestra un ejemplo de un proceso que construye una pista 3D de un objeto (p. ej., una bola) en movimiento.

La figura 6B muestra un ejemplo de proceso que excluye al menos una posición 2D obtenida de una segunda cámara.

Los números de referencia y designaciones similares en los distintos dibujos indican elementos similares.

DESCRIPCIÓN DETALLADA

La figura 1A muestra un ejemplo de un sistema 100 que realiza un procesamiento previo basado en el movimiento de datos de imágenes bidimensionales (2D) seguido de un seguimiento tridimensional (3D) de un objeto en movimiento a través de un espacio 3D 110. El objeto a seguir puede ser una bola de golf u otro tipo de objeto que se golpea, se patea o se lanza (por ejemplo, una bola de béisbol, de fútbol o de fútbol americano/rugby). En algunas implementaciones, el espacio 3D 110 es un campo de prácticas de gol, un campo de hierba u otra área abierta en la que se pueden lanzar objetos. Por ejemplo, el espacio 3D 110 puede ser parte de una instalación de entretenimiento de golf que incluye una o más dianas 114, un edificio que incluye pantallas de impacto de golf, cada una de las cuales incluye al menos una zona de tee 112 (más generalmente, una zona de lanzamiento 112), y potencialmente otros entretenimientos así como opciones de restauración.

En algunas implementaciones, el espacio 3D 110 es un área de juego para un deporte, tal como un campo de golf, donde el área de lanzamiento 112 puede ser el tee de golf para un hoyo particular en el campo de golf, o un punto de aterrizaje intermedio para una bola de golf en juego en el campo, y la diana 114 puede ser la copa al final del hoyo particular en el campo de golf o un punto de aterrizaje intermedio para una bola de golf en juego en el campo. también son posibles Otras implementaciones, como que el área de lanzamiento 112 sea una de las múltiples áreas de tee designadas a lo largo de una línea de tee donde los golfistas pueden golpear bolas de golf en un campo abierto 110, o que el área de lanzamiento 112 sea una de las múltiples áreas de tee designadas en las gradas de un estadio deportivo donde los golfistas pueden golpear bolas de golf sobre y hacia el campo de juego 110 del estadio deportivo.

El sistema 100 incluye dos o más sensores 130, incluyendo al menos una cámara 120 y su ordenador 125 asociado. Uno o más de los sensores 130 (incluyendo la al menos una cámara 120 y su ordenador asociado 125) pueden estar situados cerca de la zona de lanzamiento 112 del objeto a seguir, pero no es necesario que sea así. En algunas implementaciones, uno o más sensores 130 (incluyendo la cámara 120 y el ordenador 125) pueden estar situados a lo largo de uno o ambos lados del espacio 3D 110, y/o en el otro lado del espacio 3D 110 opuesto al área de lanzamiento 112. Por ejemplo, en un torneo de golf, la cámara 120 y el ordenador 125 pueden estar situados detrás del green, mirando hacia el golfista, asumiendo que los golpes se realizarán hacia el green. Así, en varias implementaciones, los sensores 130 pueden observar y seguir objetos que se alejan de un sensor 130, hacia un sensor 130, y/o a través del campo de visión de un sensor 130 (obsérvese que cada conjunto de tres puntos en secuencia en una figura indica que uno o más casos adicionales del sensor, ordenador, canal de comunicaciones, etc. que también pueden ser incluidos).

Los sensores 130 pueden incluir cámaras (p. ej., pares de cámaras estereoscópicas), dispositivos de radar (p. ej., dispositivos de radar Doppler de antena única), o combinaciones de los mismos, incluyendo potencialmente una unidad híbrida de sensor de cámara-radar, tal y como se describe en la Patente de EE.UU. N° 10.596.416. No obstante, al menos uno de los sensores 130 es una cámara 120 y su ordenador asociado 125, que están conectados por un canal de comunicaciones. Las figuras 1B-1D muestran ejemplos de diferentes configuraciones de sensores y ordenadores, como pueden utilizarse en el sistema de la figura 1A.

La figura 1B muestra un ejemplo de un par de cámaras 152, 156 que están conectadas a un primer ordenador 150 a través de primeros canales de comunicaciones 154, 158 que tienen un primer ancho de banda de datos que es mayor que el de al menos otro canal de comunicaciones utilizado en el sistema. Por ejemplo, los primeros canales de comunicaciones 154, 158 pueden emplear una o más tecnologías de comunicación de datos de gran ancho de banda y corta distancia, tales como Bus Serie Universal (USB) 3.0, Interfaz de Procesador Industrial Móvil (MIPI), Interconexión de Componentes Periféricos extendida (PCIe), etc. Como se describe con más detalle a continuación, el procesamiento previo de los datos de imagen de la cámara o cámaras 152, 156 puede realizarse cerca de la cámara en uno o más primeros ordenadores 150, y una vez que el procesamiento previo en el primer o primeros ordenadores 150 ha reducido el ancho de banda de datos, la salida de este procesamiento previo puede enviarse a través de un

segundo canal de comunicaciones 162 que tiene un segundo ancho de banda de datos que es menor que el primer ancho de banda de datos. Así, el segundo canal de comunicaciones 162 puede emplear uno o más anchos de banda inferiores, tecnologías de comunicación de datos de mayor distancia, tales como Ethernet de cobre o conexiones de datos inalámbricas (p. ej., utilizando WiFi y/o una o más tecnologías de comunicación de telefonía móvil).

Esto es importante porque permite que el sistema se implemente con cámara o cámaras de mayor resolución 120, 152, 156 y con un ordenador u ordenadores 125, 150 que funcionen con datos de imágenes en bruto (sin comprimir) de esta cámara o cámaras 120, 152, 156. Obsérvese que, tanto si se utiliza el seguimiento con cámara estereoscópica como el seguimiento híbrido cámara/radar, el uso de una cámara de mayor resolución con una frecuencia de imagen superior permite un seguimiento 3D de mayor calidad, pero sólo si los datos pueden procesarse de forma eficiente y eficaz. Además, si se pretende que el seguimiento de objetos funcione para objetos muy pequeños (p. ej., el objeto puede aparecer en un solo píxel incluso en una imagen de cámara de alta resolución), la detección de objetos puede necesitar tener acceso a datos de imagen sin procesar (sin comprimir), ya que el uso de técnicas tradicionales de compresión de vídeo con pérdida (MPEG y similares) puede eliminar de las imágenes información valiosa sobre objetos pequeños.

Para resolver estos problemas, el primer ordenador u ordenadores 150 pueden realizar un procesamiento previo de los datos de imagen (incluida la detección de objetos y, opcionalmente, el seguimiento 2D) cerca de la cámara o cámaras 152, 156 para reducir los requisitos de ancho de banda necesarios para enviar los datos de los sensores a uno o más segundos ordenadores 160 a través del segundo canal de comunicaciones 162. Además, el procesamiento previo (tal y como se describe en este documento) permite la sincronización temporal virtual (posterior a la captura de imágenes) de las posiciones de los objetos medidos en el tiempo y el espacio, lo que permite realizar un seguimiento 3D en el segundo ordenador 160 utilizando los datos recibidos a través de uno o más segundos canales de comunicación 162. Esto permite que el procesamiento posterior se realice fácilmente en un servidor remoto porque, después del procesamiento previo, el ancho de banda de los datos es tan bajo que es trivial enviar los datos a largas distancias.

Obsérvese que esto puede proporcionar ventajas significativas a la hora de configurar el sistema 100 debido a la flexibilidad que proporciona. Por ejemplo, en el caso de una competición de golf retransmitida por televisión (TV), donde el sistema 100 puede utilizarse para seguir bolas de golf a través del espacio 3D del campo de golf y superponer un trazo de la bola de golf en una señal de TV producida para su transmisión en directo, o para su grabación, los sensores 130 pueden desplegarse a una milla o más de las instalaciones de producción de TV (donde puede situarse el ordenador de seguimiento 3D 160). Obsérvese que la traslación de las posiciones de la bola (identificadas durante el seguimiento 3D) a las posiciones correspondientes en los datos de vídeo obtenidos por la cámara de TV (permitiendo la superposición del trazado de una representación gráfica de la trayectoria de vuelo de la bola sobre los datos de vídeo) puede realizarse utilizando técnicas de homografía conocidas. Como otro ejemplo, en el caso de una instalación de entretenimiento de golf, el ordenador de seguimiento 3D (p. ej., un ordenador servidor 140, 160) no necesita estar situado en la misma instalación, y el seguimiento 3D realizado por este ordenador (p. ej., para aumentar otros datos o medios, como mostrar la trayectoria de la bola de golf en una representación informática del entorno físico en el que se encuentra el golfista, o en un entorno virtual que sólo existe en el ordenador) puede transferirse fácilmente a otro ordenador (p. ej., proceso de conmutación por error).

Son posibles varias configuraciones de sensores y ordenadores. La figura 1C muestra un ejemplo en el que cada cámara 152, 156 tiene un primer ordenador dedicado 150A, 150B, y los ordenadores 150A, 150B comunican sus respectivos datos procesados previamente al segundo ordenador u ordenadores 160 a través de segundos canales de comunicación 162, 164 separados. Así, las cámaras (u otra tecnología de sensores) pueden compartir o no compartir los recursos del primer ordenador. Además, el procesamiento previo puede dividirse y realizarse en distintos ordenadores.

La figura 1D muestra un ejemplo en el que la cámara 152 está acoplada con el ordenador 150 sobre un primer canal de comunicaciones 154 que tiene un primer ancho de banda de datos, el primer ordenador 150 está acoplado con un tercer ordenador 166 sobre un tercer canal de comunicaciones 168 que tiene un tercer ancho de banda de datos, y el tercer ordenador 166 está acoplado con el segundo ordenador 160 sobre el segundo canal de comunicaciones 162 que tiene el segundo ancho de banda de datos, donde el segundo ancho de banda de datos es menor que el primer ancho de banda de datos, y el tercer ancho de banda de datos que es menor que el primer ancho de banda de datos pero mayor que el segundo ancho de banda de datos. El primer ordenador 150 realiza la detección del objeto, el tercer ordenador 166 realiza el seguimiento 2D del objeto, y el segundo ordenador 160 realiza la sincronización temporal virtual y el seguimiento 3D del objeto. Además, en algunas implementaciones, el primer ordenador 150 realiza la detección del objeto y el seguimiento previo en 2D (utilizando restricciones muy simples/flexibles), el tercer ordenador 166 realiza un seguimiento 2D más exhaustivo, y el segundo ordenador 160 realiza la sincronización de tiempo virtual y el seguimiento 3D del objeto.

También son posibles otras configuraciones de sensores y ordenadores, coherentes con la divulgación de este documento. Por ejemplo, el primer ordenador 150 puede realizar la detección del objeto (con seguimiento previo en 2D (utilizando restricciones muy simples/no estrictas) o sin seguimiento 2D del objeto), y un mismo segundo

ordenador 160 puede realizar el seguimiento 2D del objeto (seguimiento 2D más exhaustivo después de un seguimiento previo en 2D o todo el seguimiento 2D), la sincronización de tiempo virtual y el seguimiento 3D del objeto, en lugar de utilizar un tercer ordenador intermedio 166 para realizar el seguimiento 2D del objeto. Por el contrario, en algunas implementaciones pueden utilizarse uno o más ordenadores intermedios adicionales. Por ejemplo, el sistema puede emplear cuatro ordenadores separados para realizar cada una de las siguientes cuatro operaciones: detección del objeto, seguimiento 2D, sincronización de tiempo virtual y seguimiento 3D. Como otro ejemplo, el sistema puede emplear cinco ordenadores separados para realizar cada una de las siguientes cinco operaciones: detección de objetos, seguimiento previo en 2D (usando restricciones muy simples/flexibles), seguimiento 2D más completo, sincronización de tiempo virtual y seguimiento 3D. Son posibles otras configuraciones, siempre que al menos una de las operaciones se produzca en un primer ordenador acoplado comunicativamente con al menos una cámara a través de un primer canal de comunicaciones, y al menos otra de las operaciones se produzca en un segundo ordenador acoplado comunicativamente con el primer ordenador a través de un segundo canal de comunicaciones que tenga un ancho de banda de datos inferior al ancho de banda de datos del primer canal de comunicaciones.

En el sistema pueden utilizarse varios tipos de ordenadores. Los elementos esenciales de un ordenador son un procesador para ejecutar instrucciones y uno o más dispositivos de memoria para almacenar instrucciones y datos. Como se utiliza en este documento, un "ordenador" puede incluir un ordenador servidor, un ordenador cliente, un ordenador personal, un circuito programable integrado o un circuito lógico de propósito especial. La figura 2 es un diagrama esquemático de un sistema de procesamiento de datos que incluye un aparato de procesamiento de datos 200, que representa una implementación de un primer ordenador 150, un segundo ordenador 160, o un tercer ordenador 166. El aparato de procesamiento de datos 200 puede estar conectado con uno o más ordenadores 290 a través de una red 280.

El aparato de procesamiento de datos 200 puede incluir varios módulos de software, que pueden estar distribuidos entre una capa de aplicaciones y un sistema operativo. Estos pueden incluir programas de software ejecutables y/o interpretables o bibliotecas, incluyendo un programa 230 que funciona como un programa de detección de objetos (p. ej., en el primer ordenador 150), un programa de seguimiento 2D (p. ej., en el primer ordenador 150 y/o en el tercer ordenador 166), un programa de sincronización de tiempo virtual (p. ej., en el segundo ordenador 160), y/o un programa de seguimiento 3D (p. ej., en el segundo ordenador 160), como se describe en este documento. El número de módulos de software utilizados puede variar de una implementación a otra. Además, en algunos casos, p. ej., un programa de seguimiento 2D 230, el programa 230 puede implementarse en firmware integrado, y en otros casos, p. ej., un programa de sincronización temporal y seguimiento 3D 230, el programa 230 puede implementarse como módulos de software que se distribuyen en uno o más aparatos de procesamiento de datos conectados por una o más redes de ordenadores u otras redes de comunicación adecuadas.

El aparato de procesamiento de datos 200 puede incluir dispositivos de hardware o firmware que incluyen uno o más procesadores de hardware 212, uno o más dispositivos adicionales 214, un medio legible por ordenador no transitorio 216, una interfaz de comunicación 218, y uno o más dispositivos de interfaz de usuario 220. El procesador 212 es capaz de procesar instrucciones para su ejecución dentro del aparato de procesamiento de datos 200, tales como instrucciones almacenadas en el medio legible por ordenador no transitorio 216, que puede incluir un dispositivo de almacenamiento tal como uno de los dispositivos adicionales 214. En algunas implementaciones, el procesador 212 es un procesador de uno o varios núcleos, o dos o más unidades centrales de procesamiento (CPU). El aparato de procesamiento de datos 200 utiliza su interfaz de comunicación 218 para comunicarse con uno o más ordenadores 290, por ejemplo, a través de la red 280. Así, en diversas implementaciones, los procesos descritos pueden ejecutarse en paralelo o en serie, en un equipo informático de uno o varios núcleos, y/o en un clúster/nube de ordenadores, etc.

Ejemplos de dispositivos de interfaz de usuario 220 incluyen una pantalla, una pantalla táctil, un altavoz, un micrófono, un dispositivo de retroalimentación táctil, un teclado y un ratón. Además, no es necesario que el dispositivo o dispositivos de interfaz de usuario 220 sean dispositivos locales, sino que pueden ser remotos con respecto al aparato de procesamiento de datos 200, p. ej., dispositivo o dispositivos de interfaz de usuario 290 accesibles a través de una o más redes de comunicación 280. El aparato de procesamiento de datos 200 puede almacenar instrucciones que implementan operaciones como las descritas en este documento, por ejemplo, en el medio legible por ordenador no transitorio 216, que puede incluir uno o más dispositivos adicionales 214, por ejemplo, uno o más de un dispositivo de disquete, un dispositivo de disco duro, un dispositivo de disco óptico, un dispositivo de cinta y un dispositivo de memoria de estado sólido (por ejemplo, una unidad RAM). Además, las instrucciones que implementan las operaciones descritas en este documento pueden descargarse al medio legible por ordenador no transitorio 216 a través de la red 280 desde uno o más ordenadores 290 (por ejemplo, desde la nube), y en algunas implementaciones, la unidad RAM es un dispositivo de memoria volátil en el que se descargan las instrucciones cada vez que se enciende el ordenador.

La figura 3 muestra un ejemplo de procesos realizados en diferentes ordenadores para detectar objetos, seguir los objetos en 2D, producir posiciones virtuales en 2D para la sincronización temporal y construir seguimientos en 3D de los objetos en movimiento. Los procesos de la figura 3 incluyen operaciones de procesamiento previo 310-330 realizadas en uno o más primeros ordenadores (p. ej., ordenadores 125, 150, 166 en las figuras 1A-1D) y operaciones de procesamiento adicionales 360-375 realizadas en uno o más segundos ordenadores (p. ej.,

ordenadores 140, 160 en las figuras 1A-1D). Las operaciones de procesamiento previo pueden incluir la detección de objetos y el seguimiento 2D que comprime eficazmente los datos de ubicación de la bola (para reducir los requisitos de ancho de banda para el envío de datos que se utilizarán en el seguimiento 3D) de una manera que permite la sincronización de tiempo virtual de las posiciones de los objetos medidos durante el procesamiento adicional en el segundo o segundos ordenadores.

Así, los fotogramas de imagen 300 se reciben 310 (por ejemplo, por un ordenador 125, 150) desde una cámara a través de un primer canal de comunicaciones 305 que acopla la cámara con el primer ordenador u ordenadores, donde el primer canal de comunicaciones 305 a través del cual se reciben los fotogramas de imagen 300 tiene un primer ancho de banda de datos. Por ejemplo, el primer canal de comunicaciones 305 puede ser un canal de comunicaciones USB 3.0, MIPI, oPClx, p. ej., el canal o canales de comunicaciones 154, 158. Obsérvese que el requisito de ancho de banda entre la cámara y el ordenador puede superar fácilmente 1 Gigabits por segundo (Gbps), p. ej., una cámara de 12 megapíxeles (MP) funcionando a 60 fotogramas por segundo (FPS) y 12 bits por píxel necesita un ancho de banda de más de 8 Gbps.

Además, el uso combinado de varias cámaras de este tipo puede requerir un ancho de banda total de 10-100 Gbps, lo que supondría una gran carga incluso para el hardware de comunicación Ethernet. Además, las configuraciones estereoscópicas (por ejemplo, las cámaras estereoscópicas 152, 156 de la figura 1B) a veces requieren una distancia significativa entre las cámaras, o entre las cámaras y la infraestructura informática, como salas de servidores o computación basada en la nube, lo que hace que la comunicación de gran ancho de banda sea aún más difícil cuando se requieren cables largos y/o comunicación a través de Internet. Como se ha señalado anteriormente, las técnicas tradicionales de compresión de vídeo, como la tecnología MPEG, pueden no ser una forma adecuada de reducir el ancho de banda, especialmente cuando se van a seguir objetos diminutos (por ejemplo, una bola de golf lejana), ya que los objetos que se van a seguir corren el riesgo de ser eliminados por la compresión de vídeo tradicional. Por lo tanto, como entrada para el proceso de detección de objetos se utiliza un canal de comunicaciones de gran ancho de banda 305 (para fotogramas de vídeo de una o más cámaras) que permite recibir datos de imagen de alta resolución, gran profundidad de bits y/o sin comprimir 310.

Las ubicaciones de interés se identifican 315 (p. ej., mediante un ordenador 125, 150) en los fotogramas de imagen recibidos. Por ejemplo, esto puede implicar el uso de técnicas de diferenciación de imágenes para identificar cada ubicación en un fotograma de imagen que tenga uno o más valores de datos de imagen que cambien en más de una cantidad umbral con respecto a un fotograma de imagen anterior. También son posibles otros enfoques. Por ejemplo, el proceso puede buscar grupos de píxeles de una determinada luminancia o color (p. ej., blanco para las bolas de golf), buscar formas que coincidan con la forma de los objetos a seguir (p. ej., una forma redonda o al menos elíptica para encontrar una bola de golf redonda), y/o utilizar la coincidencia de plantillas para buscar el objeto (p. ej., una bola de golf) en la imagen.

Además, la búsqueda de ubicaciones que tienen uno o más valores de datos de imagen que cambian en más de una cantidad umbral de un fotograma de imagen a otro fotograma de imagen puede incluir la aplicación de diferenciación de imágenes para encontrar píxeles o grupos de píxeles que cambian en más de la cantidad umbral. Por ejemplo, puede aplicarse la diferenciación de imágenes para encontrar píxeles que cambian en más de un determinado valor umbral en cada imagen, y pueden encontrarse grupos de tales píxeles cambiantes que son adyacentes entre sí, por ejemplo, utilizando técnicas conocidas de etiquetado de componentes conectados (CCL) y/o análisis de componentes conectados (CCA). Un grupo de tales píxeles (y potencialmente también un único píxel) que satisfaga los criterios de detección de objetos se denomina "mancha", la ubicación y el tamaño de cada una de tales manchas puede almacenarse en una lista, y la lista de todas las manchas de cada imagen puede enviarse al componente de seguimiento 2D. Convertir una imagen en una lista de ubicaciones de objetos (o manchas) tiene un efecto de reducción del ancho de banda. En algunos casos, la reducción del ancho de banda de esta operación puede ser de 10:1 o más. Pero se puede conseguir una mayor reducción del ancho de banda, como se describe en este documento, lo que puede proporcionar un beneficio significativo cuando se van a seguir objetos diminutos.

En el caso del seguimiento de objetos diminutos, existe un problema importante con las falsas detecciones, ya que es difícil discriminar objetos diminutos (posiblemente un solo píxel en una imagen) basándose en las características del objeto. Así, la identificación 315 (para detectar objetos de interés en ubicaciones específicas de las imágenes de la cámara) puede implementarse con un umbral bajo para favorecer cero falsos negativos, al tiempo que se permiten bastantes falsos positivos. Es de apreciar que este enfoque es generalmente contra intuitivo en el sentido de que los falsos positivos en el seguimiento de objetos se ven a menudo desfavorecidos, estableciendo así una competencia entre minimizar tanto los falsos positivos como los falsos negativos. Sin embargo, el presente enfoque del seguimiento de objetos admite fácilmente los falsos positivos, ya que el procesamiento posterior está diseñado para gestionar grandes cantidades de falsos positivos. No obstante, dado que la detección de objetos está diseñada para permitir muchos falsos positivos, se identificarán más objetos 315 en cada fotograma de la imagen 300, incluyendo muchos "objetos" que no son más que ruido en los datos de la imagen, compensando así parcialmente el efecto de reducción de ancho de banda de convertir una imagen en una lista de objetos.

Se encuentran las secuencias de las ubicaciones identificadas en los fotogramas de la imagen 320 (p. ej.,

mediante un ordenador 125, 150, 160, 166). Obsérvese que los procesos mostrados en la figura 3 (y en las demás figuras) se presentan como operaciones secuenciales para facilitar su comprensión, pero en la práctica, las operaciones pueden realizarse en paralelo o concurrentemente, por ejemplo, utilizando multitarea basada en hardware y/o sistema operativo, y/o utilizando técnicas de canalización. La canalización puede utilizarse para la concurrencia, p. ej., la identificación de objetos 315 puede empezar a procesar el fotograma $n+1$, si está disponible, justo después de entregar el fotograma n al seguimiento 2D 320, sin tener que esperar a que los componentes posteriores terminen primero. Por lo tanto, la divulgación presentada en este documento en relación con las figuras no se limita a la realización secuencial de las operaciones, tal como se representa en las figuras, excepto cuando los procesos realizados en los respectivos ordenadores se describen como procesos secuenciales, es decir, el proceso de identificación de objetos, el proceso o los procesos de seguimiento 2D, y la sincronización de tiempo virtual y el proceso o los procesos de seguimiento 3D se producen en secuencia porque cada etapa de procesamiento de identificación de objetos y de seguimiento 2D reduce el ancho de banda de los datos enviados a los componentes posteriores.

Cada una de las secuencias encontradas 320 satisface un criterio de movimiento para ubicaciones identificadas en al menos tres fotogramas de imagen de la cámara. En algunas implementaciones, el criterio se mide en relación con más de tres fotogramas y/o se utilizan uno o más criterios (p. ej., el criterio de inicio de árbol que se describe a continuación). En general, el seguimiento 2D trata de encontrar secuencias de objetos (o manchas) a lo largo de tres o más fotogramas que indiquen un movimiento del objeto consistente con el de un objeto en movimiento newtoniano, no afectado por fuerzas distintas de la gravedad, el rebote, el viento, la resistencia del aire o la fricción.

El criterio para este movimiento del objeto puede definirse para incluir el desplazamiento, la velocidad y/o la aceleración en cada dimensión (x e y en la imagen) dentro de un intervalo predefinido de valores. Este intervalo de valores se establece de modo que el movimiento 2D y la aceleración de un objeto en movimiento (p. ej., una bola de golf volando) representado por una cámara 2D estén dentro de los límites especificados, mientras que se rechazan los movimientos más bruscos (en ausencia de un objeto conocido sobre el que pueda rebotar el objeto a seguir). Además, dado que el sistema más amplio empleará una etapa de seguimiento secundaria en el procesamiento posterior, que puede realizar un filtrado más detallado de lo que constituye un objeto real a seguir, p. ej., golpes de golf, la detección 320 no tiene por qué ser un filtro perfecto (o ni siquiera cercano a la perfección) que sólo acepte el movimiento real del objeto, como el de una bola de golf después de ser golpeada desde una zona de tee de salida 112.

Más bien, el filtrado realizado en 320 está hecho intencionadamente para ser menos que perfecto, permitiendo que se incluyan en las secuencias encontradas objetos distintos de los objetos en movimiento, incluyendo potencialmente secuencias de ruido que son identificadas incorrectamente 315 como un objeto de interés y luego detectadas incorrectamente 320 para formar una secuencia. En otras palabras, la detección 320 puede implementar un filtro flojo que aumente los falsos positivos para minimizar los falsos negativos, p. ej., en 320 todas o casi todas las bolas de golf en movimiento serán aceptadas como formando una secuencia válida.

Este enfoque más laxo (beneficio de la duda) permite utilizar un algoritmo de seguimiento mucho más sencillo en 320, sabiendo que no necesita ser perfecto a la hora de discriminar los objetos deseados (p. ej., las bolas de golf) de los no deseados (p. ej., bolas que no son de golf). El conjunto de reglas que definen el seguimiento puede reducirse al mínimo, y cualquier error cometido por el seguimiento 2D (tal como dejar pasar a través una bola que no sea de golf) puede ser filtrado por los componentes y el procesamiento posteriores. En lugar de emitir trayectorias enteras, cada una de las cuales tiene un punto inicial y un punto final, las secuencias detectadas 320 pueden representarse mediante un "árbol enraizado" en el que cada vértice (nodo del árbol) es una mancha observada (en x , y y tiempo t) y cada arista es un posible movimiento entre ubicaciones de un objeto cuyo movimiento se está siguiendo. Cada una de estas ramas puede tener también algunos metadatos, tal como la profundidad total del árbol, como se describe con más detalle en relación con la figura 4A.

Sin embargo, incluso con este enfoque más laxo (beneficio de la duda/umbral bajo), es posible que se produzcan detecciones de objetos faltantes. Por lo tanto, se pueden utilizar observaciones ficticias para tener en cuenta los objetos que deberían estar en los datos de la imagen pero que no se identifican. En algunas implementaciones, si no se encuentra una mancha suficientemente buena que pueda extender una trayectoria, el rastreador 2D puede añadir una observación ficticia en la ubicación prevista. Las observaciones ficticias pueden ser implementadas con una penalización significativa y, en algunas implementaciones, no se permitirán las observaciones ficticias a menos que el gráfico esté ya a una cierta profundidad. Dado que hay límites en la penalización que puede tener una rama, en la práctica hay límites en el número de observaciones ficticias que puede tener una trayectoria.

Como se ha indicado anteriormente, la detección 320 puede implicar la formación de árboles enraizados a partir de las ubicaciones de interés, donde cada árbol enraizado es un gráfico acíclico conectado con un nodo raíz, que es la raíz del árbol, y cada arista del gráfico acíclico conectado se origina directa o indirectamente a partir de la raíz. La figura 4A muestra un ejemplo de un proceso que encuentra secuencias de ubicaciones de objetos que satisfacen un criterio de movimiento mediante la formación de árboles enraizados. En 400, se obtiene un siguiente conjunto de ubicaciones identificadas para un fotograma de imagen para su procesamiento, y mientras las ubicaciones de interés permanezcan 405 en el conjunto para el fotograma actual, este procesamiento continúa. Por ejemplo,

cuando se va a procesar un fotograma de manchas, todas las manchas del nuevo fotograma se pueden cotejar con todos los nodos del árbol que se añadieron durante el procesamiento del fotograma anterior para ver si la mancha puede ser una posible continuación de esa trayectoria, dependiendo de cuánto se parezca el punto de esta rama al movimiento deseado, definido por el criterio de movimiento.

Se recupera una siguiente ubicación de interés 410 del conjunto, se realiza una comprobación 415 para determinar si esta ubicación de interés satisface un criterio de iniciación de árbol. Si es así, se establece un nodo raíz de un nuevo árbol 420 utilizando esta ubicación de interés. Por ejemplo, si los valores de los datos de la imagen en la ubicación de interés son mayores que un tamaño mínimo de objeto, esto puede utilizarse para indicar que una bola está cerca de la cámara, y debe establecerse un nuevo árbol. La figura 4B muestra un ejemplo visual de esto, en el que se observan seis manchas 460, pero sólo cuatro de estas manchas 460 son lo suficientemente grandes como para ser utilizadas para establecer nuevos nodos raíz 465. Obsérvese que una observación de una mancha puede añadirse a varios árboles, y cada mancha observada podría en teoría ser el inicio de una nueva observación de un objeto. Esto puede llevar a una explosión combinatoria en entornos ruidosos, por lo que en algunas implementaciones se impone alguna restricción adicional (tal como una restricción de tamaño mínimo de las manchas) antes de establecer un nuevo árbol.

En este ejemplo, todas las manchas que superan un determinado tamaño mínimo son promovidas a nuevos árboles de profundidad 0, donde el límite de tamaño mínimo puede establecerse en función de la resolución de la cámara y del nivel de ruido del vídeo entrante. Por ejemplo, el criterio puede ser que una mancha debe tener al menos 2 píxeles (y no un solo píxel) para establecer un nuevo árbol. Obsérvese que el ruido verdaderamente aleatorio afecta a los píxeles individualmente y muy raramente genera agrupaciones más grandes en una imagen. También son posibles otros enfoques para evitar una explosión combinatoria. Por ejemplo, los umbrales de ruido en el proceso de generación de manchas (es decir, de identificación de la ubicación) pueden ajustarse de forma adaptativa.

Además, se puede aumentar la profundidad mínima del árbol necesaria para exportar una trayectoria, lo que garantiza que la cantidad de datos exportados sea limitada, ya que las secuencias de ruido aleatorio rara vez consiguen construir secuencias más largas por casualidad. Se generarán muchos gráficos con profundidad 0 y 1, ya que las restricciones son muy generosas, pero la mayoría de estos gráficos nunca alcanzarán la profundidad 2, ya que la predicción es mejor entonces, y aún menos los gráficos alcanzarán una profundidad de 3. Así, la estadística funciona en beneficio de los sistemas y técnicas de esta divulgación, ya que los árboles o ramas se descartan cuando no hay más nodos que se les puedan añadir.

Además, en el caso de que la cámara esté situada cerca de la diana, mirando hacia atrás, hacia la zona de lanzamiento del objeto, es decir, al detectar objetos entrantes, en distintas partes de la imagen se pueden aplicar umbrales de tamaño diferentes. Por ejemplo, se puede utilizar un tamaño mínimo de un píxel en la parte (p. ej., un rectángulo) en la que se detectaría una bola lejana, y en otros lugares puede utilizarse un tamaño mínimo de dos píxeles. Así, los criterios utilizados para formar los árboles de posibles trayectorias de objetos pueden determinarse en función de la ubicación de la cámara con respecto a la zona de lanzamiento.

Por ejemplo, si se supone que el suelo es algo plano y se conoce la posición y la orientación de la cámara, todas las bolas detectadas por debajo del "horizonte" del suelo pueden compararse con una estimación de la distancia máxima (y, por tanto, del tamaño mínimo) a la que puede encontrarse la bola. Un rayo desde la cámara a través de la detección de la bola interseca el plano del suelo a cierta distancia D. La bola debe estar a la distancia D o más cerca, o de lo contrario estaría bajo tierra. Para una detección de bola aleatoria por encima del "horizonte", no existe una heurística de distancia simple, ya que no hay intersección con el plano del suelo. Sin embargo, si se conocen los límites del espacio 3D, por ejemplo, el campo de juego, se pueden utilizar algunas restricciones generales y/o dependientes del ángulo (distancia máxima/tamaño mínimo del objeto), dado que el objeto debe estar dentro de la región 3D de interés.

Volviendo a la figura 4A, se realiza una comprobación 425 para determinar si la ubicación de interés actual está dentro de un umbral de distancia de un nodo raíz establecido para una ubicación identificada en un fotograma de imagen anterior. Si es así, la ubicación de interés se añade 430 como un primer subnodo de profundidad a este árbol, bajo el nodo raíz del árbol. Obsérvese que, si el árbol se compone de un solo vértice (profundidad=0), no hay forma de estimar aún la velocidad de esa trayectoria, por lo que es necesario emparejarla con cualquier mancha que se encuentre a una distancia razonable del vértice. La figura 4C muestra un ejemplo visual de esto, como continuación de la figura 4B. De las cuatro manchas 470 observadas, sólo dos están dentro del límite de distancia 472 de un nodo raíz 474 de un fotograma previo. Así, sólo dos de las manchas 470 se añaden como subnodos de primera profundidad 476 al árbol enraizado. En algunas implementaciones, el límite de distancia depende de la resolución y de la velocidad de fotogramas del sensor de entrada, así como de la velocidad máxima esperada (perpendicular a la dirección en la que apunta la cámara) del objeto que se está siguiendo. Si el tamaño de los objetos que se están siguiendo se conoce de antemano, como es el caso de las bolas de golf, esto puede tenerse en cuenta a la hora de determinar el límite/umbral de distancia 472. Además, se puede suponer que una mancha más grande está más cerca de la cámara que una mancha más pequeña, lo que significa que se puede tolerar un mayor movimiento entre fotogramas cuando hay una mancha más grande (ya sea en el fotograma anterior o en el fotograma actual, dependiendo de la colocación

de la cámara con respecto a la dirección prevista de los objetos que se están siguiendo).

Sin embargo, al hacer coincidir las manchas con un vértice en profundidad uno o mayor, es posible predecir dónde debería estar el siguiente punto de una trayectoria. En algunas implementaciones, la región esperada de la siguiente mancha se calcula suponiendo que la velocidad es la misma desde el último fotograma. La figura 4D muestra un ejemplo visual de esto. Basándose en la ubicación del nodo raíz 480 y la ubicación del primer subnodo de profundidad 482, más la velocidad de fotogramas conocida de la cámara, se calcula una ubicación prevista 484. A continuación, se determina un área de búsqueda 486 (o región) alrededor de la ubicación prevista 484 para buscar manchas que añadir al árbol. De forma similar a los detalles proporcionados anteriormente para determinar el límite/umbral de distancia 472, el tamaño del área/región 486 alrededor de la ubicación prevista 484 puede determinarse basándose en la resolución de la cámara y la velocidad máxima esperada (perpendicular a la dirección en la que apunta la cámara) del objeto que se está rastreando, potencialmente ajustada de acuerdo con el tamaño conocido del objeto y el tamaño de las manchas en los diferentes fotogramas de la imagen. Además, en algunas implementaciones, se puede utilizar un filtro de Kalman para crear predicciones, p. ej., cuando el orden del gráfico es suficientemente grande.

Cualquier mancha que esté lo suficientemente cerca de la ubicación prevista puede convertirse en un vértice y añadirse al árbol. Así, como se muestra en la figura 4D, una mancha se encuentra dentro de la región 486 de la ubicación prevista 484, por lo que esta mancha se añade como un nuevo subnodo 488. Además, en algunas implementaciones, las penalizaciones se pueden acumular en las ramas del árbol, donde las penalizaciones se suman a una puntuación que se utiliza para clasificar las ramas, lo que puede ayudar a decidir las mejores ramas para exportar. Además, en algunas implementaciones, cuanto más se desvíe la ubicación de una mancha de las coordenadas esperadas, mayor será la penalización para esta rama.

En algunas implementaciones, las penalizaciones se acumulan en las ramas del árbol y pueden utilizarse tanto para limitar la magnitud de la penalización permitida como para determinar cuándo descartar ramas. En el primer caso, si una rama ya tiene una penalización alta, no se le permitirá añadir nuevos nodos que harían que la rama extendida superase el límite. Cuando la penalización se calcula a partir de la discrepancia entre la ubicación prevista y la ubicación real, se trata esencialmente de una forma de asegurarse de que la aceleración está dentro de los límites. Si no lo está, la penalización será demasiado elevada. En este último caso, cuando hay varias ramas que tienen las 3 últimas manchas en común, puede conservarse la rama que es más profunda y tiene la penalización más baja y las otras pueden descartarse. Obsérvese que este tipo de seguimiento de hipótesis basado en penalizaciones puede aplicarse al seguimiento 3D además de al 2D.

Volviendo a la figura 4A, se realiza una comprobación 435 para determinar si la ubicación de interés actual está dentro de una región determinada utilizando una velocidad estimada de una ubicación utilizada para un subnodo precursor de un árbol establecido. Si es así, la ubicación de interés se añade 440 como un segundo subnodo o subnodo de mayor profundidad a este árbol, bajo el subnodo precursor, tal como se ha descrito anteriormente en relación con la figura 4D. Además, en algunas implementaciones, cuando la profundidad del árbol aumenta, se pueden utilizar algoritmos de predicción más sofisticados para obtener mejores predicciones. Ajustando los nodos más recientes de la rama a una función polinómica de segundo o tercer orden, se puede obtener tanto una buena predicción de dónde estaría el siguiente punto como descartar trayectorias con cambios de aceleración claramente irrazonables. Además, en algunas implementaciones, pueden proporcionarse las observaciones a un filtro de Kalman, y puede utilizarse el modelo del filtro de Kalman para producir nuevas predicciones.

Además, cabe señalar que la misma ubicación de interés en un único fotograma de imagen puede añadirse como subnodos respectivos (de primera o mayor profundidad) a dos o más árboles enraizados que estén realizando un seguimiento de posibles trayectorias de objetos a través de los fotogramas de imagen, así como utilizarse potencialmente para establecer un nuevo nodo raíz de un nuevo árbol. Además, puede añadirse la misma ubicación de interés en un único fotograma de imagen como subnodo en un árbol con más de un enlace de vuelta a los respectivos nodos precursores del árbol, incluyendo el mismo subnodo con diferentes profundidades dependiendo del enlace que se siga de vuelta al nodo raíz.

Por lo tanto, se producirán muchas trayectorias falsas, y aunque el procesamiento posterior pueda filtrar estas trayectorias falsas, debería existir una limitación sobre cuándo se emitirán las trayectorias potenciales. En algunas implementaciones, no se considerará ninguna secuencia para la salida hasta que el árbol enraizado que representa la secuencia exceda 445 una profundidad de árbol predeterminada. Así, no se emite una secuencia de ubicaciones identificadas hasta que la secuencia se confirma 450 para la salida basándose en la profundidad del árbol para la secuencia que excede 445 la profundidad predeterminada del árbol, como se especifica para una implementación dada, o como se determina sobre la marcha basándose en uno o más factores.

Obsérvese que sólo las porciones del árbol enraizado que tienen la profundidad de árbol predeterminada necesitan ser confirmadas 450 para su salida como secuencias de ubicaciones identificadas. Esto ayuda a evitar que el ruido, que produce falsas trayectorias en los árboles enraizados, se propague desde el seguimiento 2D en la etapa de procesamiento previo en un primer ordenador a la fase posterior al procesamiento en un segundo ordenador.

Además, el umbral mínimo de profundidad del árbol dependerá generalmente del entorno y del retardo aceptable para producir un seguimiento 3D del objeto. En algunas implementaciones, el árbol enraizado de la secuencia debe tener una profundidad de árbol superior a dos antes de que la secuencia se confirme 450 para la salida. En algunas implementaciones, la profundidad del árbol debe ser superior a tres, cuatro, cinco, seis o siete, antes de que la secuencia sea confirmada 450 para la salida.

Por ejemplo, en algunas implementaciones, los vértices de un árbol (correspondientes a las manchas/ubicaciones) sólo se exportan si se encuentran a una determinada profundidad (D) en el árbol, o si tienen nodos derivados que se exportan debido a que los nodos derivados tienen esa determinada profundidad (D) en el árbol. Un vértice con $D = 5$ tiene cinco aristas entre sí mismo y la raíz. Esto significa que la salida puede retrasarse el mismo número D de fotogramas, ya que a menudo no puede saberse de antemano si una mancha formará parte de una rama suficientemente larga. Puede demostrarse que el filtrado descrito (mediante la limitación de la exportación de nodos de árbol) reduce drásticamente el ancho de banda necesario para comunicar los candidatos a objeto, ya que las únicas manchas exportadas son las que constituyen posibles trayectorias. La siguiente tabla muestra un ejemplo de reducción del ancho de banda para una aplicación de golf basada en la profundidad mínima de rama requerida.

Tabla 1

Profundidad requerida	Reducción de datos
1	73%
2	84%
3	93%
4	97%
5	98%
6	99%

Como queda claro, la gran mayoría de las manchas/ubicaciones de interés no están conectadas por ninguna trayectoria o no están conectadas por una trayectoria de suficiente profundidad y, por lo tanto, la mayoría de las manchas/ubicaciones de interés son rechazadas por el filtrado efectivo realizado en la etapa de procesamiento previo. Volviendo a la figura 3, en 325, se realiza una comprobación para determinar si los datos de las secuencias están listos para la salida, es decir, para cualquier secuencia de nodos en un árbol que incluya al menos un nodo con una profundidad de árbol mayor que la profundidad de árbol predeterminada, como se ha descrito anteriormente, los nodos de esa secuencia están listos para la salida. En tal caso, los datos de salida para una o más secuencias de ubicaciones se envían 330 (p. ej., mediante un ordenador 125, 150, 166) a uno o más segundos ordenadores (p. ej., un ordenador 140, 160). Sin embargo, en algunas implementaciones, las ubicaciones identificadas 315 se envían desde un primer ordenador (p. ej., por un ordenador 125, 150) a un segundo ordenador (p. ej., un ordenador 140, 160, 166) que encuentra 320 las secuencias de las ubicaciones identificadas en los fotogramas de la imagen y, a continuación, emite los datos de las secuencias a otro proceso en el mismo ordenador o en un ordenador diferente. En cualquier caso, los datos de salida de la secuencia incluyen al menos, para cada ubicación de cada secuencia, una posición bidimensional de la ubicación en un fotograma de imagen específico con una marca de tiempo, que es necesaria para la sincronización virtual en el segundo ordenador.

Uno de los fundamentos del seguimiento 3D es la triangulación. Sin embargo, triangular la posición de un objeto visto en dos cámaras requiere disponer de observaciones de la misma instancia temporal. Una forma habitual de conseguirlo es utilizar una señal de disparo síncrona común para todas las cámaras (p. ej., transmitida a través de un cable). Pero esto sólo funciona si las cámaras tienen la misma frecuencia de imagen de captura. En algunas configuraciones, es difícil o imposible garantizar que las dos cámaras estén sincronizadas (disparando las capturas de imagen en tándem), lo que dificulta o imposibilita la triangulación de las observaciones de las distintas cámaras. Para resolver este problema, en lugar de intentar sincronizar realmente las imágenes de la cámara con otro u otros sensores en el punto de captura de datos, se utiliza información de marca de tiempo (en combinación con información sobre el movimiento representado por la trayectoria 2D) para la sincronización virtual en el segundo ordenador.

Además, los sistemas y técnicas descritos en este documento pueden utilizarse tanto con cámaras de obturador global como con cámaras de obturador rodante. Un obturador rodante significa que la sincronización de la captura de cada fila en la imagen de la cámara es diferente. Por lo tanto, cuando el sensor capta un objeto, se puede determinar la posición del objeto, pero el momento de la medición depende del lugar de la imagen en el que se encuentre el objeto. En algunas implementaciones, puede haber una diferencia de tiempo de aproximadamente 20 milisegundos entre las filas superior e inferior de la imagen. Esto también supone un problema para la triangulación, ya que puede no cumplirse el requisito de medición simultánea en ambas cámaras. Para resolver este problema, también se tiene en cuenta la información del obturador rodante considerando el tiempo relativo del desfase temporal concreto de la "ubicación" dentro del fotograma, además de la marca de tiempo del fotograma, de forma que también se puedan utilizar las posiciones del objeto medidas a partir de cámaras con obturador rodante para una triangulación de alta calidad en un segundo ordenador durante la fase posterior al procesamiento. Para la captura con obturador

global, este desplazamiento es siempre cero.

Dado que pueden ser necesarios varios fotogramas de datos para que una pista cumpla el requisito de longitud mínima de trayectoria (profundidad mínima del árbol), tal como se ha descrito anteriormente en relación con la Tabla 1, existe un compromiso entre introducir un retardo para todas las manchas o implementar un protocolo más complejo que permita comunicar las manchas ya cualificadas tan pronto como se conozcan. Así, en algunas implementaciones, el envío 330 implica un envío retardado de datos de fotograma y de mancha, y en algunas implementaciones, el envío 330 implica un envío incremental de datos de fotograma y de mancha.

En el caso del envío diferido, el envío 330 implica retrasar la salida de datos para un fotograma de imagen dado y sus ubicaciones de interés encontradas en una o más de las secuencias, hasta que no se puedan incluir más ubicaciones de interés identificadas para el fotograma de imagen dado en ninguna de las secuencias basadas en ubicaciones de interés identificadas en fotogramas de imagen posteriores. Así, la exportación de nodos para el fotograma puede realizarse cuando el número de fotogramas posteriores que se han procesado excluye que se alcance la profundidad de árbol predefinida para más manchas del fotograma actual. En otras palabras, la transmisión de un fotograma se retrasa hasta que se sabe que no se incluirán más manchas identificadas en el fotograma en una trayectoria en un árbol enraizado que tenga la profundidad de árbol mínima requerida.

Además, en algunas implementaciones, se puede utilizar una profundidad de árbol mínima dinámica. La profundidad mínima del árbol puede establecerse inicialmente a un nivel bajo para minimizar la latencia cuando hay pocas trayectorias, y luego incrementarse dinámicamente si la carga (salida total de trayectorias cualificadas por unidad de tiempo) excede algún umbral máximo de carga. Esto puede proporcionar una especie de efecto de estrangulamiento que mejore el rendimiento ajustando el procesamiento en respuesta a los datos de entrada actuales.

Además, la estructura de datos del fotograma puede incluir una lista de manchas que se detectaron en este fotograma y que pasaron los criterios de filtrado. Por ejemplo, la estructura de datos del fotograma puede incluir los campos de datos que se muestran en la Tabla 2.

Tabla 2

Nombre	Tipo de datos	Descripción
fn	int64	Número de fotogramas según la fuente de vídeo
fn_prev	int64	Número de fotograma del fotograma anterior. Será fn+1 a menos que se hayan perdido fotogramas en la subida.
t_sync	int64	Marca de tiempo en microsegundos. Esta marca de tiempo es la que debe utilizarse para correlacionar fotogramas entre cámaras.
t_wall (pared)	int64	Hora del reloj de pared expresada en microsegundos transcurridos desde <i>unix epoch</i>
blobs (manchas)	Blob []	Lista de manchas detectadas en este fotograma y que han superado los criterios de filtrado.

Además, la estructura de datos de Blob (mancha) incluida en la estructura de datos de fotograma puede incluir los campos de datos que se muestran en la Tabla 3 a continuación.

Tabla 3

Nombre	Tipo de datos	Descripción
fn	int64	Número de fotograma en el que se observó esta mancha.
idx	int	Índice basado en cero de la mancha en el fotograma. La primera mancha transmitida recibe el n.º 0, la siguiente el n.º 1, etc...
x	float (flotante)	Coordenada x de la mancha
y	float (flotante)	Coordenada y de la mancha
size (tamaño)	int	Número de píxeles de la mancha
c	int8	Contraste de la mancha. Indica si una mancha es más oscura o más brillante que el modelo de fondo.
dt	int	Desplazamiento temporal (en microsegundos) de la mancha dentro del fotograma, para controlar los efectos del obturador. El tiempo real en el que se observó la mancha es t_sync+dt. Debe ser 0 cuando se utiliza el obturador global.
previous (anterior)	int[]	Lista los índices de todas las manchas del fotograma anterior a las que está conectada esta mancha. Vacía si es la raíz del árbol de la trayectoria.

Obsérvese que estas estructuras de datos pueden codificarse en varios formatos, como por ejemplo en formato JSON, como se indica a continuación:

```

{"fn":25281,
"fn_prev":25280,
"t_sync":7430127868,
"t_wall":1587541719527486,
5  "blobs":[
    {"idx":0,"c":255,"dt":0,"prev":[0],"size":6,"x":939.5,"y":577.0},
    {"idx":1,"c":202,"dt":0,"prev":[1],"size":106,"x":973.625,"y":568.375},
    {"idx":2,"c":0,"dt":0,"prev":[2],"size":1,"x":681.0,"y":469.0},
    {"idx":3,"c":0,"dt":0,"prev":[4],"size":1,"x":2.5,"y":938.0},
10  {"idx":4,"c":0,"dt":0,"prev":[5],"size":3,"x":0.875,"y":939.375},
    {"idx":5,"c":255,"dt":0,"prev":[6],"size":4,"x":959.75,"y":558.625},
    {"idx":6,"c":0,"dt":0,"prev":[6,7],"size":2,"x":949.5,"y":579.0},
    {"idx":7,"c":0,"dt":0,"prev":[8],"size":1,"x":678.75,"y":475.625},
    {"idx":8,"c":0,"dt":0,"prev":[8],"size":2,"x":677.625,"y":476.75},
15  {"idx":9,"c":255,"dt":0,"prev":[9],"size":1,"x":433.0,"y":975.0},
    {"idx":10,"c":255,"dt":0,"prev":[10],"size":1,"x":410.5,"y":997.0},
    {"idx":11,"c":113,"dt":0,"prev":[11],"size":5,"x":411.75,"y":991.75},
    {"idx":12,"c":0,"dt":0,"prev":[12],"size":1,"x":407.5,"y":999.0},
    {"idx":13,"c":0,"dt":0,"prev":[13],"size":1,"x":410.5,"y":1012.0},
20  {"idx":14,"c":0,"dt":0,"prev":[14],"size":1,"x":793.5,"y":599.0}
  ]}

```

Para transmisiones con limitaciones de ancho de banda, puede utilizarse una codificación más compacta, como el formato de búferes de protocolo de Google.

Cuando las manchas no filtradas se transmiten entre nodos de procesamiento, por ejemplo si la detección de manchas tiene lugar en el dispositivo de borde, mientras que el seguimiento previo y/o el seguimiento se llevan a cabo en un ordenador diferente, se puede utilizar un formato simplificado, ya que de todos modos no hay información sobre la relación entre manchas de diferentes fotografías. Por ejemplo, la estructura de datos del fotograma puede incluir los campos de datos que se muestran en la Tabla 4 a continuación.

Tabla 4

Nombre	Tipo de datos	Descripción
fn	int64	Número de fotogramas según la fuente de vídeo
fn_prev	int64	Número de fotograma del fotograma anterior. Será fn+1 a menos que se hayan perdido fotogramas en la subida.
t_sync	int64	Marca de tiempo en microsegundos. Esta marca de tiempo es la que debe utilizarse para correlacionar fotogramas entre cámaras.
t_wall (pared)	int64	Hora del reloj de pared expresada en microsegundos transcurridos desde <i>unix epoch</i>
readout_duration (duración de lectura)	int32	Número de microsegundos entre el inicio de la captura de la línea de fotograma superior y la finalización de la línea de fotograma inferior. Para sensores de obturación global, es cero. Puede ser negativo si la lectura se realizó de abajo a arriba en lugar de arriba a abajo.
n_rows (filas)	int16	Número de filas (líneas de exploración) del fotograma.
n_columns (columnas)	int16	Número de columnas del fotograma
blobs (manchas)	SimpleBlob []	Una lista de todas las manchas detectadas en este fotograma.
bitmap (mapa de bits)	byte []	Contiene un mapa de bits monocromo con n_filas x n_columnas de píxeles. Para cada píxel, un 1 indica que este píxel se desvió del modelo de ruido lo suficiente como para generar una mancha, y un 0 significa que el píxel estaba dentro de los límites de ruido esperados. Se puede utilizar cualquier formato de imagen estándar que admita el almacenamiento sin pérdidas de imágenes monocromas, por ejemplo Portable Network Graphics (PNG) Specification (Segunda Edición). También puede tratarse de una matriz empaquetada en bits en bruto, posiblemente comprimida mediante un algoritmo de compresión sin pérdidas razonablemente rápido, como DEFLATE Compressed Data Format Specification versión 1.3. El remitente y el destinatario deben ponerse de acuerdo sobre el formato que se utiliza en cada implementación concreta.

Además, la estructura de datos SimpleBlob incluida en la estructura de datos de fotograma puede incluir los campos de datos que se muestran en la Tabla 5 a continuación.

Tabla 5

Nombre	Tipo de datos	Descripción
x	float (flotante)	coordenada x de la mancha
y	float (flotante)	coordenada y de la mancha
size (tamaño)	int	Número de píxeles de la mancha
c	int8	Contraste de la mancha. Indica si una mancha es más oscura o más brillante que el modelo de fondo.

En el caso del envío incremental de fotogramas, el envío 330 implica la salida de datos para los fotogramas de imagen a medida que se completa la identificación 315 para cada fotograma de imagen respectivo, y la salida 330 de datos para cada ubicación de interés sólo después de encontrar 325 que una o más de las secuencias incluyen la ubicación de interés que se va a enviar. Como se ha comentado anteriormente, la secuencia debe tener la profundidad de árbol necesaria para que un nodo sea exportado, y cuando un árbol alcanza el umbral de profundidad, todos los nodos precursores (del nodo del árbol que supera el umbral de profundidad) que no hayan sido exportados anteriormente (por alguna otra rama) se envían 330 retroactivamente. Se trata de un enfoque más complejo para enviar los datos que puede utilizarse si se requiere una baja latencia. El seguimiento 2D emite un flujo continuo con información sobre los fotogramas que se han procesado y listas de todas las manchas que han superado el filtrado de candidatos a bolas (es decir, todas las ubicaciones de interés identificadas 315) tan pronto como la información está disponible.

Por ejemplo, cuando se procesa un nuevo fotograma, la información del fotograma puede enviarse utilizando la estructura de datos de fotograma con los campos de datos que se muestran en la Tabla 6 a continuación.

Tabla 6

Nombre	Tipo de datos	Descripción
fn	int64	Número de fotograma según la fuente de vídeo.
fn_prev	int64	Número de fotograma del fotograma anterior. Será fn+1 a menos que se hayan perdido fotogramas en el flujo ascendente.
t_sync	int64	Marca de tiempo en microsegundos. Esta marca de tiempo es la que se debe utilizar para correlacionar fotogramas entre cámaras
t_wall (pared)	int64	Hora del reloj de pared expresada en microsegundos transcurridos desde <i>unix epoch</i> .

Además, la estructura de datos blob (mancha) puede incluir los campos de datos que se muestran en la Tabla 7 a continuación.

Tabla 7

Nombre	Tipo de datos	Descripción
fn	int64	Número de fotograma en el que se observó esta mancha.
idx	int	Índice basado en cero de la mancha en el fotograma. La primera mancha transmitida recibe el n.º 0, la siguiente el n.º 1, etc...
x	float (flotante)	Coordenada x de la mancha
y	float (flotante)	coordenada y de la mancha
size (tamaño)	int	Número de píxeles de la mancha
c	int8	Contraste de la mancha. Indica si una mancha es más oscura o más brillante que el modelo de fondo.
dt	int	Desplazamiento temporal (en microsegundos) de la mancha dentro del fotograma, para controlar los efectos del obturador. El tiempo real en el que se observó la mancha es t_sync+dt. Debe ser 0 cuando se utiliza el obturador global.

Además, cada conexión entre manchas puede representarse mediante una estructura de datos de conexión que incluya los campos de datos que se muestran en la Tabla 8 a continuación.

Tabla 8

Nombre	Tipo de datos	Descripción
fn	int64	Número de fotograma de esta mancha.
idx	int	n.º de índice "in" de la mancha cuyas conexiones se enumeran.
prev	Array [int]	Lista de los índices de todas las manchas del fotograma anterior con las que está conectada esta mancha en las trayectorias que cumplen los criterios.

Los registros generados utilizando las estructuras de datos anteriores pueden ponerse en el flujo (del primer ordenador al segundo ordenador) tan pronto como se establezca la información (los datos de fotogramas y manchas pueden enviarse 330 durante y después de la identificación 315, y los datos de conexiones de manchas pueden enviarse 330 después de que una secuencia encontrada 320 formada por estas conexiones de manchas haya sido confirmada 325 para la salida). Está garantizado que cada mancha y fotograma referenciado por un registro se pondrá en el flujo antes del registro que lo referencia. Sin embargo, en caso de que el consumidor comience a leer a mitad del flujo podría haberse perdido la declaración original, por lo que esta cuestión debería tratarse en el segundo ordenador.

Independientemente de las estructuras de datos o la temporización de salida utilizadas, los datos de salida 350 se reciben 360 (por ejemplo, por un ordenador 140, 160) desde el uno o más primeros ordenadores (por ejemplo, desde un ordenador 125, 150) a través de un segundo canal de comunicaciones 355 que acopla el primer ordenador u ordenadores con el segundo ordenador u ordenadores, donde el segundo canal de comunicaciones 355 a través del que se reciben los datos de salida 350 tiene un segundo ancho de banda de datos que es menor que el primer ancho de banda de datos del primer canal de comunicaciones 305. Por ejemplo, el segundo canal de comunicaciones 355 puede ser un canal de comunicaciones Ethernet de cobre o inalámbrico, p. ej., canal o canales de comunicaciones 162, 164. Además, en implementaciones en las que parte del procesamiento previo (por ejemplo, la detección 320) se realiza en el mismo segundo ordenador (por ejemplo, por un ordenador 140, 160) que las operaciones de procesamiento adicionales 360-375, los datos recibidos incluyen las ubicaciones 315 identificadas y la información de marca de tiempo asociada.

En cualquier caso, los datos se siguen recibiendo 360 hasta que una comprobación 365 muestra que una secuencia está lista para ser procesada. Además, como se ha indicado anteriormente, las operaciones pueden realizarse en paralelo o concurrentemente, p. ej., utilizando multitarea basada en hardware y/o sistema operativo, y/o utilizando técnicas de procesamiento en cadena, p. ej., durante la construcción 375. Así, la recepción 360 puede continuar mientras el procesamiento 370 de una o más secuencias se produce en paralelo o concurrentemente; cada componente puede empezar a procesar el fotograma $n+1$, si está disponible, justo después de entregar el fotograma n al componente posterior, sin tener que esperar a que los componentes posteriores terminen primero, p. ej., puede realizarse la triangulación estereoscópica para el siguiente fotograma incluso si el rastreador 3D no ha terminado de procesar las ubicaciones de interés.

La secuencia o secuencias en los datos de salida 350 son procesadas 370 por uno o más segundos ordenadores (por ejemplo, por un ordenador 140, 160) interpolando entre posiciones 2D especificadas en fotogramas de imagen específicos para la secuencia o secuencias, utilizando las marcas de tiempo de los fotogramas de imagen específicos, para producir una posición 2D virtual en un punto predeterminado en el tiempo. El registro de cada posición de interés (p. ej., cada mancha encontrada en cada fotograma) en los datos de salida 350 incluye tanto la marca de tiempo del fotograma como una indicación de la anterior o anteriores posiciones de interés conectadas a esta posición de interés por el componente de seguimiento 2D (p. ej., cada mancha se describe con punteros a la anterior mancha o manchas a las que pertenece). Así, a partir de estos datos 350 es posible determinar la hora, la ubicación y la dirección de desplazamiento y velocidad de cada ubicación de interés/mancha.

De este modo, los datos 350 permiten el uso de la interpolación para producir ubicaciones/bloques "intermedios" virtuales en cualquier punto de tiempo de una trayectoria, siempre que exista al menos una ubicación de interés/bloque en el árbol con una marca de tiempo anterior, y al menos una ubicación de interés/bloque con una marca de tiempo posterior. Además, en algunas implementaciones, el punto predeterminado en el tiempo es uno de los múltiples puntos de tiempo de una frecuencia de imagen constante predefinida de un servidor de seguimiento 3D. Por ejemplo, un ordenador de servidor de seguimiento 3D 140, 160 puede funcionar con una frecuencia de imagen constante predefinida y utilizar interpolación para generar instantáneas virtuales de todas las manchas de las cámaras en estos puntos temporales. Dado que las coordenadas de las manchas ahora representan todas el mismo punto en el tiempo, es posible la triangulación entre los puntos aunque la captura original no fuera sincrónica. Además, en algunas implementaciones, el punto predeterminado en el tiempo es un tiempo especificado por otro sensor, tal como otro sensor de cámara o un sensor de radar.

Además, como se ha indicado anteriormente, la cámara puede ser una cámara de obturador rodante, en cuyo caso, los datos de salida 350 pueden incluir un valor de desfase temporal para cada ubicación de interés incluida en cada secuencia. Con estos datos en la mano, el procesamiento 370 también funciona para la sincronización temporal virtual con una cámara de obturador rodante. La figura 5 muestra un ejemplo de un proceso que interpola entre posiciones 2D especificadas en fotogramas de imagen específicos obtenidos de una cámara de obturador rodante.

Se calcula 500 un primer tiempo de observación para una primera ubicación que tiene una de las posiciones 2D especificadas en los fotogramas de imagen específicos añadiendo un primer valor de desfase de tiempo para la primera ubicación a la marca de tiempo de uno de los fotogramas de imagen específicos. Por ejemplo, puede añadirse el desfase de tiempo (dt) de una primera mancha dentro de un primer fotograma (como se ha detallado anteriormente en las Tablas 3 y 7) a la marca de tiempo (t_{sync}) de ese primer fotograma (como se ha detallado anteriormente en las Tablas 2, 3, 6 y 7). Se calcula 510 un segundo tiempo de observación para una segunda ubicación que tenga otra

de las posiciones 2D especificadas en los fotogramas de imagen específicos añadiendo un segundo valor de desfase de tiempo para la segunda ubicación a la marca de tiempo de un segundo de los fotogramas de imagen específicos. Por ejemplo, puede añadirse el desfase de tiempo (dt) de una segunda mancha dentro de un segundo fotograma (como se ha detallado anteriormente en las Tablas 3 y 7) a la marca de tiempo (t_{sync}) de ese segundo fotograma (como se ha detallado anteriormente en las Tablas 2, 3, 6 y 7). A continuación, se realiza la interpolación 520 utilizando el primer tiempo de observación y el segundo tiempo de observación, calculados a partir de los desfases temporales y las marcas de tiempo de los fotogramas.

Volviendo a la figura 3, con la posición 2D virtual producida, se construye una pista 3D del objeto (p. ej., una bola) en movimiento en el espacio 3D 375 (p. ej., por un ordenador 140, 160) utilizando la posición 2D virtual que se produjo, y la información de posición obtenida de al menos otro sensor para el punto predeterminado en el tiempo. La construcción 375 puede ser para la visualización (por ejemplo, visualización inmediata) de la pista 3D del objeto, o la construcción 375 puede producir la pista 3D para su uso como entrada a un procesamiento posterior antes de la visualización. Por ejemplo, la pista 3D puede ser procesada posteriormente para ser visualizada de forma efectiva mediante la superposición de un trazo de una bola de golf en una señal de TV producida para su transmisión en directo, o para su grabación. Como otro ejemplo, la pista 3D puede ser procesada posteriormente para ser mostrada eficazmente aumentando otros datos o medios, tal como mostrando la trayectoria de una bola de golf en una representación por ordenador de un entorno físico en el que se encuentra un golfista, o en un entorno virtual que sólo existe en el ordenador pero que se muestra a un usuario del sistema.

También son posibles otros tipos de procesamiento antes de la visualización. Por ejemplo, la trayectoria 3D puede procesarse para determinar la posición final de reposo de la bola deportiva, lo que puede ser útil para las aplicaciones de apuestas o la recopilación de estadísticas generales que alimentan los sitios web deportivos. Además de mostrar el trazo, la pista 3D puede procesarse adicionalmente para medir la velocidad, el efecto, el alcance y el ángulo de lanzamiento del tiro. La pista 3D puede procesarse adicionalmente para determinar si las bolas han traspasado la red de un campo y se han introducido en las instalaciones vecinas. La pista 3D puede procesarse adicionalmente para indicar al propietario del campo de prácticas cuáles son los intervalos en los que hay actividad y también para contar el número de bolas que se han lanzado desde cada intervalo.

La figura 6A muestra un ejemplo de un proceso que construye una pista 3D de un objeto (p. ej., una bola) en movimiento. La construcción de la pista 3D incluye combinar 600 (p. ej., por un ordenador 140, 160) la posición 2D virtual con la información de posición obtenida de al menos otro sensor para formar una posición 3D de un objeto de interés. En general, esto implica triangulación de las observaciones de los diferentes sensores, usando los datos de observación del objeto producidos usando esos diferentes sensores, y usando datos de calibración para esos diferentes sensores. Obsérvese que los datos de observación producidos son utilizables en la triangulación debido a la sincronización de tiempo virtual lograda al producir la posición 2D virtual.

Por ejemplo, el otro sensor puede ser una segunda cámara, que se utilizará como par estereoscópico con la primera cámara para la que se produjo la posición 2D virtual. Las técnicas de detección de objetos, seguimiento 2D y sincronización temporal virtual descritas anteriormente también pueden utilizarse con la segunda cámara. A la luz de esto, los datos de salida para la segunda cámara pueden producir múltiples objetos detectados (con múltiples posiciones 2D virtuales correspondientes) en el mismo punto compartido en el tiempo para el que se produjo la posición 2D virtual para la primera cámara. Por lo tanto, la información de posición de la segunda cámara puede consistir en dos o más posiciones 2D obtenidas de la segunda cámara, y la combinación 600 puede incluir la determinación de cuál de las dos o más posiciones 2D de la segunda cámara debe emparejarse con la posición 2D virtual producida para los datos de la primera cámara.

Además, la combinación 600 puede implicar excluir 602 al menos una, pero no todas las dos o más posiciones 2D obtenidas de la segunda cámara como no capaces de formar un punto 3D con la posición 2D virtual obtenida de la primera cámara. La figura 6B muestra un ejemplo de un proceso que excluye al menos una posición 2D obtenida de una segunda cámara para efectuar el filtrado de línea epipolar antes de la triangulación 604. Una región sobre al menos una porción de una línea epipolar en un plano de imagen de la segunda cámara se determina 602A utilizando la posición 2D virtual producida para la primera cámara, un centro óptico de la primera cámara, un centro óptico de la segunda cámara, una línea de base entre la primera y la segunda cámaras, y datos de calibración extrínsecos para la primera y la segunda cámaras. En algunas implementaciones, la región se determina 602A basándose en un margen de error para los datos que se utilizan.

En algunas implementaciones, el punto de referencia entre la primera y la segunda cámaras 120, 130 de un par de cámaras estereoscópicas es de treinta metros o menos, lo que proporciona una precisión estereoscópica razonable en combinación con la capacidad de detectar el disparo con suficiente antelación para poder extrapolar la trayectoria hasta el lugar desde el que se golpeó. Utilizar un objetivo con un campo de visión más amplio permite observar el disparo antes, pero a costa de una menor precisión cuando la bola está lejos. Las cámaras de mayor resolución pueden mitigar esto en cierta medida. Un punto de referencia más corto proporciona menos precisión de profundidad, pero permite observar los tiros antes. Obsérvese que utilizando la sincronización de tiempo virtual descrita en esta divulgación, pueden combinarse los datos de diferentes pares de cámaras 120, 130 según sea necesario, p.

ej., dinámicamente en tiempo real. Así, diferentes pares de cámaras 120, 130 en el sistema pueden formar un par estereoscópico, y los diferentes pares de cámaras estereoscópicas pueden así tener diferentes puntos de referencia y diferente precisión de profundidad.

Además, en algunas implementaciones, la región 602A determinada se reduce (o se reduce aún más) 602A basándose en porciones de la línea epipolar que no pueden utilizarse debido a límites conocidos de la distancia al objeto, p. ej., la distancia conocida entre la cámara y la zona de tee 112 y/o la diana 114. Además, se pueden utilizar otros criterios a la hora de emparejar las observaciones de las bolas candidatas desde diferentes cámaras utilizadas como uno o más pares estereoscópicos. Por ejemplo, un criterio adicional puede ser que el contraste de las manchas para las respectivas manchas observadas por dos cámaras (utilizadas como un par estereoscópico) sean similares.

A continuación, los emparejamientos de la posición 2D virtual obtenida de la primera cámara con las respectivas de las dos o más posiciones 2D obtenidas de la segunda cámara se rechazan 602B en respuesta a que las respectivas de las dos o más posiciones 2D están fuera de la región sobre la al menos una porción de la línea epipolar en el plano de imagen de la segunda cámara. En esencia, la línea en el espacio 3D que es vista por la primera cámara como un único punto en el plano 2D de la primera cámara (porque esta línea está directamente en línea con el centro óptico de la primera cámara) es vista por la segunda cámara como una línea en el plano de imagen de la segunda cámara, que se conoce como la línea epipolar. Dada una posición 2D conocida de un objeto observado por la primera cámara, sólo los objetos observados por la segunda cámara que caigan a lo largo de esta línea epipolar (dentro de algún valor de tolerancia) pueden ser el mismo objeto observado por ambas cámaras. Además, si el sistema está diseñado y configurado con límites conocidos sobre las distancias a los objetos que deben considerarse (p. ej., distancias basadas en la distancia a la zona de salida 112, la distancia a la diana 114, como la distancia a un green cuando se siguen bolas que entran en el green, y/o la distancia o distancias a los objetos que se están siguiendo en ese momento), el sistema puede colocar topes duros (límites) en la línea epipolar para su uso en el rechazo 602B. Así, los objetos que están claramente fuera de los límites (p. ej., aviones, pájaros, luces de tráfico distante, etc.) pueden ser fácilmente ignorados por el sistema.

También son posibles otros enfoques para excluir las posiciones 2D obtenidas de la segunda cámara. Por ejemplo, como se ha indicado anteriormente, los datos de salida pueden incluir, para cada posición después de una posición inicial en cada secuencia, una indicación de una posición anterior en la secuencia. Debido a que se incluyen estos datos (p. ej., porque el envío 330 transmite no sólo la posición de cada mancha sino también la posición de las manchas observadas anteriormente que han sido conectadas por el seguimiento 2D en una secuencia de trayectoria) se puede estimar la posible dirección de movimiento 2D y la velocidad de cada objeto de interés (mancha), y esta estimación se puede utilizar para hacer el filtrado de movimiento antes de la triangulación 604, p. ej., el algoritmo de triangulación estereoscópica puede usar esta información para rechazar emparejamientos de manchas con parámetros incompatibles, como por ejemplo usando los vectores de velocidad de los puntos estereoscópicos izquierdo y derecho para descartar falsos positivos. Por ejemplo, cuando se predice (basándose en la velocidad actual) la ubicación del objeto en el siguiente fotograma en ambas cámaras, las ubicaciones predichas (cuando se triangulan) también terminarán cerca de la siguiente observación real cuando las manchas sean efectivamente observaciones del mismo objeto desde dos ángulos diferentes.

Así, la exclusión 602 puede implicar: estimar una velocidad 2D para un objeto en la posición 2D virtual basándose en las posiciones 2D especificadas en los fotogramas de imagen específicos, donde se identifica al menos una de las posiciones 2D especificadas en los datos de salida usando una indicación de una ubicación previa al menos en una de las secuencias; estimar velocidades 2D para objetos en las dos o más posiciones 2D obtenidas de la segunda cámara; y rechazar emparejamientos de la posición 2D virtual obtenida de la primera cámara con las respectivas dos o más posiciones 2D obtenidas de la segunda cámara basándose en las velocidades 2D estimadas para el objeto en la posición 2D virtual y para los objetos en las dos o más posiciones 2D, y en los datos de calibración intrínseca y la calibración extrínseca.

Obsérvese que las coordenadas de la ubicación/mancha deben estar sin distorsión utilizando los datos de calibración intrínseca y también convertirse a sistemas de coordenadas compatibles utilizando los datos de calibración extrínseca, como apreciarán los expertos en la materia. Se puede realizar una calibración intrínseca que determine la distorsión óptica y la distancia focal real de cada cámara. Puede realizarse una calibración extrínseca que determine la dirección de las cámaras entre sí. Juntas, las calibraciones intrínseca y extrínseca se denominan "la calibración" o "los datos de calibración", y pueden utilizarse para triangular la posición de los objetos vistos por ambas cámaras.

Además, el otro sensor puede incluir un dispositivo de radar en lugar de una segunda cámara. Por ejemplo, uno o más dispositivos de radar pueden combinarse con una sola cámara, con un par de cámaras estereoscópicas, o con tres o más cámaras, para formar uno o más sensores híbridos de cámara-radar, de acuerdo con los sistemas y técnicas descritos en la patente de EE. UU. n.º 10.596.416, para proporcionar al menos una parte de los sensores de un sistema de detección de objetos que registran objetos de interés para el seguimiento de objetos en 3D. Las lecturas del radar o radares que pueden ser detecciones de un objeto en movimiento (por ejemplo, una bola de golf) pueden añadirse a una nube de puntos 3D de posiciones tridimensionales de los objetos de interés y, de este modo, aumentar la robustez potencial y el área de cobertura de los sensores combinados.

Las lecturas del radar pueden convertirse en puntos 3D mediante el siguiente enfoque. Puede combinarse una medición de distancia del radar con todos los candidatos a bola detectados por una de las cámaras en el mismo momento. El ángulo de cada candidato a bola, tal y como lo indica la cámara, y la distancia indicada por el radar determinarán una ubicación 3D en el espacio. Esto puede hacerse para cada candidato a bola visto por la cámara, lo que producirá una matriz de puntos 3D para añadir a la nube de puntos. Como máximo, uno de esos puntos será una asociación correcta de los datos del radar y una observación de la bola (lo que significa que ambos proceden del mismo objeto; un verdadero positivo). El resto de los puntos 3D serán falsas asociaciones (falsos positivos). Sin embargo, una vez más, el algoritmo de seguimiento de puntos 3D se elige para que sea robusto y no permita que la mayoría de los puntos 3D sean falsos positivos. Además, en algunas implementaciones, pueden utilizarse en el sistema uno o más sensores de radar capaces de determinar tanto el intervalo como el ángulo con respecto a cada objeto observado.

Pero independientemente de si se obtienen y/o excluyen datos de posición de un segundo sensor, en algunas implementaciones, puede triangularse una posición 3D del objeto de interés 604 utilizando la posición 2D virtual obtenida de la primera cámara, al menos una de las dos o más posiciones 2D obtenidas de la segunda cámara, datos de calibración intrínsecos para la primera cámara y la segunda cámara (la distorsión óptica determinada y la longitud focal real de cada cámara), y datos de calibración extrínsecos para la primera y la segunda cámaras (la dirección determinada de las cámaras entre sí). Obsérvese que, dado que las coordenadas de las ubicaciones/manchas representan todas el mismo punto en el tiempo, la triangulación 604 entre los puntos es posible aunque la captura original no fuera sincrónica.

A continuación, puede añadirse 610 la posición 3D del objeto de interés a otras posiciones 3D de objetos de interés en una nube de posiciones 3D de objetos de interés para el punto predeterminado en el tiempo. Obsérvese que pueden utilizarse varios tipos de estructuras de datos. En algunas implementaciones, las nubes de posiciones 3D se almacenan como árboles octales, y este tipo de estructura de datos permite representar cualquier punto del espacio 3D (hasta el límite de precisión de la representación informática de números escalares). Se realiza 620 el análisis de movimiento a través de múltiples nubes de posiciones 3D para construir la pista 3D de la bola en movimiento en el espacio 3D, donde cada una de las múltiples nubes es para un único punto en el tiempo. Obsérvese que pueden utilizarse varios tipos de análisis de movimiento. No obstante, debe apreciarse que, aunque el número de falsos positivos incluidos en los datos 350 puede ser mayor que el de una técnica tradicional de seguimiento de objetos, estos falsos positivos tienden a repartirse en el espacio 3D, que es grande en comparación con el plano de imagen 2D de la cámara desde el que se originan los falsos positivos. Debido a esto, una vez que el sistema de seguimiento 3D identifica un objeto en movimiento 3D, es sencillo para el sistema de seguimiento 3D descartar fácilmente cualquier falso positivo en el espacio 3D que no coincida con la pista 3D en curso que se está construyendo.

Además, el proceso puede incluir la salida 630 de la pista 3D para la generación de una representación de la pista 3D, o la generación y visualización 630 de la representación de la pista 3D en el espacio 3D. Como se ha indicado anteriormente, esto puede implicar un procesamiento adicional de la pista 3D para mostrarla de forma efectiva superponiendo un trazo de una bola de golf en una señal de TV producida para su transmisión en directo, o para su grabación, o esto puede implicar un procesamiento adicional de la pista 3D para mostrarla de forma efectiva aumentando otros datos o medios, tal como mostrando la trayectoria de una bola de golf en una representación por ordenador de un entorno físico en el que se encuentra un golfista, o en un entorno virtual que existe sólo en el ordenador pero que se muestra a un usuario del sistema, p. ej., un entorno virtual que comparten los participantes en un juego multijugador (ya sea localmente en el mismo intervalo o dispersados por todo el mundo).

Las realizaciones de la materia objeto y las operaciones funcionales descritas en la presente memoria descriptiva pueden implementarse en circuitos electrónicos digitales, o en software, firmware o hardware informático, incluidas las estructuras divulgadas en la presente memoria descriptiva y sus equivalentes estructurales, o en combinaciones de uno o más de ellos. Las realizaciones de la materia objeto descrita en la presente memoria descriptiva pueden implementarse utilizando uno o más módulos de instrucciones de programa informático codificados en un medio no transitorio legible por ordenador para su ejecución por, o para controlar el funcionamiento de, aparatos de procesamiento de datos. El medio no transitorio legible por ordenador puede ser un producto manufacturado, tal como un disco duro en un sistema informático o un disco óptico vendido a través de canales minoristas, o un sistema integrado. El medio no transitorio legible por ordenador puede adquirirse por separado y codificarse posteriormente con uno o más módulos de instrucciones de programa informático, por ejemplo, mediante el suministro de uno o más módulos de instrucciones de programa informático a través de una red cableada o inalámbrica. El medio no transitorio legible por ordenador puede ser un dispositivo de almacenamiento legible por máquina, un sustrato de almacenamiento legible por máquina, un dispositivo de memoria o una combinación de uno o más de ellos.

La expresión "aparato de procesamiento de datos" abarca todos los aparatos, dispositivos y máquinas para procesar datos, incluyendo por ejemplo un procesador programable, un ordenador o múltiples procesadores u ordenadores. El aparato puede incluir, además de hardware, código que crea un entorno de ejecución para el programa informático en cuestión, por ejemplo, código que constituye el firmware del procesador, una pila de protocolos, un sistema de gestión de bases de datos, un sistema operativo, un entorno de ejecución o una combinación de uno o

varios de ellos. Además, el aparato puede emplear distintas infraestructuras de modelos informáticos, tales como servicios web, informática distribuida e infraestructuras informáticas de malla.

Un programa informático (también conocido como programa, software, aplicación de software, script o código) puede estar escrito en cualquier forma adecuada de lenguaje de programación, incluyendo lenguajes compilados o interpretados, lenguajes declarativos o procedimentales, y puede ser implementado en cualquier forma adecuada, incluyendo como un programa independiente o como un módulo, componente, subrutina u otra unidad adecuada para su uso en un entorno informático. Un programa informático no corresponde necesariamente a un fichero en un sistema de ficheros. Un programa puede almacenarse en una porción de un archivo que contenga otros programas o datos (p. ej., uno o más scripts almacenados en un documento de lenguaje de marcas), en un único archivo dedicado al programa en cuestión o en múltiples archivos coordinados (p. ej., archivos que almacenen uno o más módulos, subprogramas o porciones de código). Un programa informático puede implementarse para ser ejecutado en un ordenador o en múltiples ordenadores ubicados en un mismo sitio o distribuidos en múltiples sitios e interconectados por una red de comunicaciones.

Los procesos y flujos lógicos descritos en esta memoria descriptiva pueden ser realizados por uno o más procesadores programables que ejecuten uno o más programas informáticos para realizar funciones operando sobre los datos de entrada y generando resultados. Los procesos y flujos lógicos también pueden ser realizados por, y los aparatos también pueden ser implementados como, circuitos lógicos de propósito especial, p. ej., una FPGA (matriz de puertas programable en campo) o un ASIC (circuito integrado de aplicación específica).

Los procesadores adecuados para la ejecución de un programa informático incluyen, a modo de ejemplo, microprocesadores tanto de propósito general como especial, y uno o más procesadores de cualquier tipo adecuado de ordenador digital. Generalmente, un procesador recibirá instrucciones y datos de una memoria de sólo lectura, de una memoria de acceso aleatorio o de ambas. Los elementos esenciales de un ordenador son un procesador para ejecutar instrucciones y uno o más dispositivos de memoria para almacenar instrucciones y datos. Generalmente, un ordenador también incluirá, o estará acoplado operativamente para recibir datos de o transferir datos a, o ambos, uno o más dispositivos de almacenamiento masivo para almacenar datos, p. ej., discos magnéticos, magneto-ópticos u ópticos. Sin embargo, no es necesario que un ordenador disponga de tales dispositivos. Además, un ordenador puede estar integrado en otro dispositivo, p. ej., un teléfono móvil, un asistente digital personal (PDA), un reproductor de audio o vídeo móvil, una consola de juegos, un receptor del Sistema de Posicionamiento Global (GPS) o un dispositivo de almacenamiento portátil (p. ej., una unidad flash de bus serie universal (USB)), por nombrar sólo algunos. Los dispositivos adecuados para almacenar instrucciones y datos del programa informático incluyen todas las formas de memoria no volátil, medios y dispositivos de memoria, incluidos, por ejemplo, dispositivos de memoria semiconductores, p. ej., EPROM (memoria programable borrrable de solo lectura), EEPROM (memoria programable borrrable eléctricamente de solo lectura) y dispositivos de memoria flash; discos magnéticos, p. ej., discos duros internos o discos extraíbles; discos magneto-ópticos; CD-ROM y DVD-ROM; almacenamiento conectado a la red; y diversas formas de almacenamiento en la nube. El procesador y la memoria pueden complementarse con circuitos lógicos especiales o incorporarse a ellos.

Para proporcionar la interacción con el usuario, las realizaciones de la materia objeto descrita en esta memoria descriptiva pueden implementarse en un ordenador que tenga un dispositivo de visualización, por ejemplo, LCD (pantalla de cristal líquido), OLED (diodo orgánico emisor de luz) u otro monitor, para mostrar información al usuario y un teclado y un dispositivo señalador, por ejemplo, un ratón o un trackball, mediante los cuales el usuario pueda proporcionar información al ordenador. También pueden utilizarse otros tipos de dispositivos para proporcionar interacción con el usuario; por ejemplo, la retroalimentación proporcionada al usuario puede ser sensorial, por ejemplo, retroalimentación visual, auditiva o táctil; y la entrada del usuario puede recibirse de diversas formas, como acústica, verbal o táctil.

El sistema informático puede incluir clientes y servidores. Un cliente y un servidor suelen estar alejados el uno del otro y normalmente interactúan a través de una red de comunicación. La relación de cliente y servidor surge en virtud de los programas informáticos que se ejecutan en los respectivos ordenadores y que tienen una relación cliente-servidor entre sí. Las realizaciones de la materia objeto descrita en la presente memoria descriptiva pueden implementarse en un sistema informático que incluya un componente back-end, p. ej., como servidor de datos, o que incluya un componente middleware, p. ej., un servidor de aplicaciones, o que incluya un componente front-end, p. ej., un ordenador cliente que tenga una interfaz gráfica de usuario o un navegador web a través del cual un usuario pueda interactuar con una realización de la materia objeto descrita en la presente memoria descriptiva, o cualquier combinación adecuada de uno o más de dichos componentes back-end, middleware o front-end. Los componentes del sistema pueden interconectarse mediante cualquier forma o medio adecuado de comunicación de datos digitales, p. ej., una red de comunicación. Entre los ejemplos de redes de comunicación se incluyen una red de área local ("LAN") y una red de área amplia ("WAN"), una red interna (p. ej., Internet) y redes entre pares (p. ej., redes ad hoc entre pares).

Aunque esta memoria descriptiva contiene muchos detalles de implementación, no deben interpretarse como limitaciones del alcance de la invención o de lo que puede reivindicarse, sino más bien como descripciones de

características específicas de realizaciones particulares de la invención. Ciertas características que se describen en esta memoria descriptiva en el contexto de realizaciones separadas también pueden implementarse de forma combinada en una única realización. A la inversa, diversas características que se describen en el contexto de una única realización también pueden implementarse en múltiples realizaciones por separado o en cualquier subcombinación adecuada. Además, aunque puede haberse descrito anteriormente que las características actúan en ciertas combinaciones e incluso reivindicarse inicialmente como tales, en algunos casos pueden eliminarse de la combinación una o más características de una combinación reivindicada, y la combinación reivindicada puede dirigirse a una subcombinación o variación de una subcombinación. Por lo tanto, a menos que se indique explícitamente lo contrario, o a menos que los conocimientos de un experto en la materia indiquen claramente lo contrario, puede combinarse cualquiera de las características de la realización descrita anteriormente con cualquiera de las otras características de la realización descrita anteriormente.

De manera similar, aunque las operaciones se representan en los dibujos en un orden particular, esto no debe entenderse como una exigencia de que tales operaciones se realicen en el orden particular mostrado o en orden secuencial, o que se realicen todas las operaciones ilustradas, para lograr los resultados deseados. En determinadas circunstancias, la multitarea y/o el procesamiento en paralelo pueden resultar ventajosos. Además, la separación de varios componentes del sistema en las realizaciones descritas anteriormente no debe entenderse como una exigencia de dicha separación en todas las realizaciones, y debe entenderse que los componentes del programa y los sistemas descritos generalmente pueden integrarse juntos en un único producto de software o empaquetarse en múltiples productos de software.

De este modo, se han descrito realizaciones particulares de la invención. Otras realizaciones están dentro del ámbito de las siguientes reivindicaciones. Por ejemplo, la descripción anterior se centra en el seguimiento del movimiento de una bola de golf, pero los sistemas y técnicas descritos también son aplicables a otros tipos de seguimiento del movimiento de objetos, como el béisbol o el tiro al plato, así como a aplicaciones no deportivas. Además, en algunas implementaciones, el seguimiento de un objeto "en movimiento" puede incluir el seguimiento del objeto cuando rebota y/o rueda por el suelo.

REIVINDICACIONES

1. Un método que comprende:

5 recibir (310), en uno o más primeros ordenadores, fotogramas de imágenes procedentes de una cámara a través de un primer canal de comunicaciones que acopla la cámara con el uno o más primeros ordenadores, el primer canal de comunicaciones teniendo un primer ancho de banda de datos;
 10 identificar (315), mediante el uno o más primeros ordenadores, ubicaciones de interés en los fotogramas de imagen;
 10 detectar (320) secuencias de las ubicaciones identificadas en los fotogramas de imagen, en donde cada una de las secuencias satisface un criterio de movimiento para ubicaciones identificadas en al menos tres fotogramas de imagen de la cámara;
 15 enviar (330) datos de salida para las secuencias de las ubicaciones, en donde los datos de salida comprenden, para cada ubicación en cada secuencia, una posición bidimensional de la ubicación en un fotograma de imagen específico que tiene una marca de tiempo;
 15 recibir (360), en uno o más segundos ordenadores, los datos de salida, en donde el uno o más segundos ordenadores están acoplados con el uno o más primeros ordenadores a través de un segundo canal de comunicaciones que tiene un segundo ancho de banda de datos que es menor que el primer ancho de banda de datos;
 20 procesar (370), por el uno o más segundos ordenadores, al menos una de las secuencias en los datos de salida, interpolando entre posiciones bidimensionales especificadas en fotogramas de imagen específicos para al menos una de las secuencias, utilizando las marcas de tiempo de los fotogramas de imagen específicos, para producir una posición bidimensional virtual en un punto predeterminado en el tiempo; y
 25 construir (375) una pista tridimensional de una bola en movimiento en el espacio tridimensional utilizando la posición bidimensional virtual y la información de posición obtenida de al menos otro sensor para el punto predeterminado en el tiempo.

2. El método de la reivindicación 1, en donde la detección y el envío son realizados por el uno o más segundos ordenadores, y las ubicaciones identificadas en los fotogramas de imagen son recibidas en el uno o más segundos ordenadores desde el uno o más primeros ordenadores a través del segundo canal de comunicaciones.

3. El método de la reivindicación 1, en donde la detección y el envío son realizados por el uno o más primeros ordenadores, y los datos de salida son recibidos en el uno o más segundos ordenadores desde el uno o más primeros ordenadores a través del segundo canal de comunicaciones.

4. El método de la reivindicación 1, en donde la detección comprende formar árboles enraizados a partir de las ubicaciones de interés que comprende:

40 establecer nodos raíz de los árboles enraizados a partir de las primeras ubicaciones de interés identificadas respectivas en respuesta a cada una de las primeras ubicaciones de interés identificadas que tengan valores de datos de imagen que satisfagan un criterio de iniciación de árbol;
 45 añadir segundas ubicaciones de interés identificadas como subnodos de los árboles enraizados en respuesta a que al menos algunas respectivas de las segundas ubicaciones identificadas se encuentren dentro de un umbral de distancia de una ubicación identificada en un fotograma de imagen anterior que se ha añadido a al menos uno de los árboles enraizados;
 confirmar cada secuencia respectiva de ubicaciones identificadas para la salida cuándo el árbol enraizado de la secuencia tiene una profundidad de árbol superior a dos.

5. El método de la reivindicación 1, en donde el envío comprende retrasar la salida de datos para un fotograma de imagen dado y sus ubicaciones de interés encontradas en una o más de las secuencias, hasta que no puedan incluirse más ubicaciones de interés identificadas para el fotograma de imagen dado en ninguna de las secuencias basadas en ubicaciones de interés identificadas en fotogramas de imagen posteriores.

6. El método de la reivindicación 1, en donde el envío comprende

55 la salida de datos para los fotogramas de imagen a medida que se completa la identificación para cada fotograma de imagen respectivo; y
 la salida de datos para cada ubicación de interés sólo después de encontrar una o más de las secuencias que incluyen la ubicación de interés a emitir.

7. El método de la reivindicación 1, en donde la cámara comprende una cámara de obturador rodante, los datos de salida comprenden un valor de desfase temporal para cada ubicación de interés incluida en cada secuencia, y el procesamiento comprende:

65 calcular un primer momento de observación para una primera ubicación que tenga una de las posiciones

bidimensionales especificadas en los fotogramas de imagen específicos añadiendo un primer valor de desfase temporal para la primera ubicación a la marca de tiempo de uno de los fotogramas de imagen específicos; calcular un segundo momento de observación para una segunda ubicación que tenga otra de las posiciones bidimensionales especificadas en los fotogramas de imagen específicos añadiendo un segundo valor de desfase temporal para la segunda ubicación a la marca de tiempo de un segundo de los fotogramas de imagen específicos; y realizar la interpolación utilizando el primer momento de observación y el segundo momento de observación.

8. El método de la reivindicación 1, en donde la construcción comprende:

combinar, mediante el uno o más segundos ordenadores, la posición bidimensional virtual con la información de posición obtenida del al menos otro sensor para formar una posición tridimensional de un objeto de interés; añadir, mediante uno o más segundos ordenadores, la posición tridimensional del objeto de interés a otras posiciones tridimensionales de objetos de interés en una nube de posiciones tridimensionales de objetos de interés para el punto predeterminado en el tiempo; realizar, mediante uno o más segundos ordenadores, un análisis de movimiento a través de múltiples nubes de posiciones tridimensionales para construir la pista tridimensional de la bola en movimiento en el espacio tridimensional, en donde cada una de las múltiples nubes es para un único punto en el tiempo, y las múltiples nubes incluyen la nube de posiciones tridimensionales de objetos de interés para el punto predeterminado en el tiempo; y la salida para la visualización de la pista tridimensional de la bola en movimiento en el espacio tridimensional.

9. El método de la reivindicación 8, en donde la cámara es una primera cámara, el al menos otro sensor es una segunda cámara, la información de posición comprende múltiples posiciones bidimensionales obtenidas de la segunda cámara, y la combinación comprende:

excluir al menos una, pero no todas, de las múltiples posiciones bidimensionales obtenidas de la segunda cámara por no poder formar un punto tridimensional con la posición bidimensional virtual obtenida de la primera cámara; triangular al menos la posición tridimensional del objeto de interés utilizando la posición bidimensional virtual obtenida de la primera cámara, al menos una de las múltiples posiciones bidimensionales obtenidas de la segunda cámara, datos de calibración intrínsecos para la primera cámara y la segunda cámara, y datos de calibración extrínsecos para la primera y la segunda cámaras.

10. El método de la reivindicación 9, en donde la exclusión comprende:

determinar una región alrededor de al menos una porción de una línea epipolar en un plano de imagen de la segunda cámara utilizando la posición bidimensional virtual, un centro óptico de la primera cámara, un centro óptico de la segunda cámara, un punto de referencia entre la primera y segunda cámaras, y los datos de calibración extrínseca para la primera y segunda cámaras; y rechazar emparejamientos de la posición bidimensional virtual obtenida de la primera cámara con las respectivas posiciones bidimensionales múltiples obtenidas de la segunda cámara en respuesta a que las respectivas posiciones bidimensionales múltiples están fuera de la región alrededor de al menos una porción de la línea epipolar en el plano de imagen de la segunda cámara.

11. El método de la reivindicación 1, en donde el uno o más primeros ordenadores comprenden una primera unidad de procesamiento y al menos una unidad de procesamiento adicional, el primer canal de comunicaciones acopla la cámara con la primera unidad de procesamiento, recibir los fotogramas de imagen comprende recibir los fotogramas de imagen en la primera unidad de procesamiento, identificar las ubicaciones de interés comprende identificar las ubicaciones de interés en la primera unidad de procesamiento, detectar las secuencias comprende detectar las secuencias en la al menos una unidad de procesamiento adicional en respuesta a la recepción de las ubicaciones desde la primera unidad de procesamiento a través de un tercer canal de comunicaciones que acopla la primera unidad de procesamiento con la al menos una unidad de procesamiento adicional, y enviar los datos de salida comprende enviar los datos de salida desde la al menos una unidad de procesamiento adicional, y en donde el tercer canal de comunicaciones tiene un tercer ancho de banda de datos que es menor que el primer ancho de banda de datos pero mayor que el segundo ancho de banda de datos.

12. Un sistema que comprende:

al menos un sensor que comprende una cámara y uno o más primeros ordenadores que comprenden un primer procesador de hardware y una primera memoria acoplada al primer procesador de hardware, codificando la primera memoria instrucciones configuradas para hacer que el primer procesador de hardware realice las primeras operaciones; al menos otro sensor; y uno o más segundos ordenadores que comprenden un segundo procesador de hardware y una segunda memoria acoplada al segundo procesador de hardware, la segunda memoria codificando instrucciones configuradas para

hacer que el segundo procesador de hardware realice segundas operaciones;
en donde las primeras operaciones y las segundas operaciones comprenden la recepción de fotogramas de
imagen, la identificación de ubicaciones de interés, la detección de secuencias, el envío de datos de salida, la
recepción de los datos de salida, el procesamiento de las secuencias y la construcción de una pista tridimensional,
de acuerdo con cualquiera de las reivindicaciones 1-11.

13. El sistema de la reivindicación 12, en donde el al menos otro sensor comprende un dispositivo de radar.

14. El sistema de la reivindicación 12, en donde el al menos otro sensor comprende una segunda cámara.

15. Un medio no transitorio legible por ordenador que codifica instrucciones que, cuando son implementadas por un
aparato de procesamiento de datos asociado con una cámara, hacen que dicho aparato de procesamiento de datos
realice el método de cualquiera de las reivindicaciones 1-11.

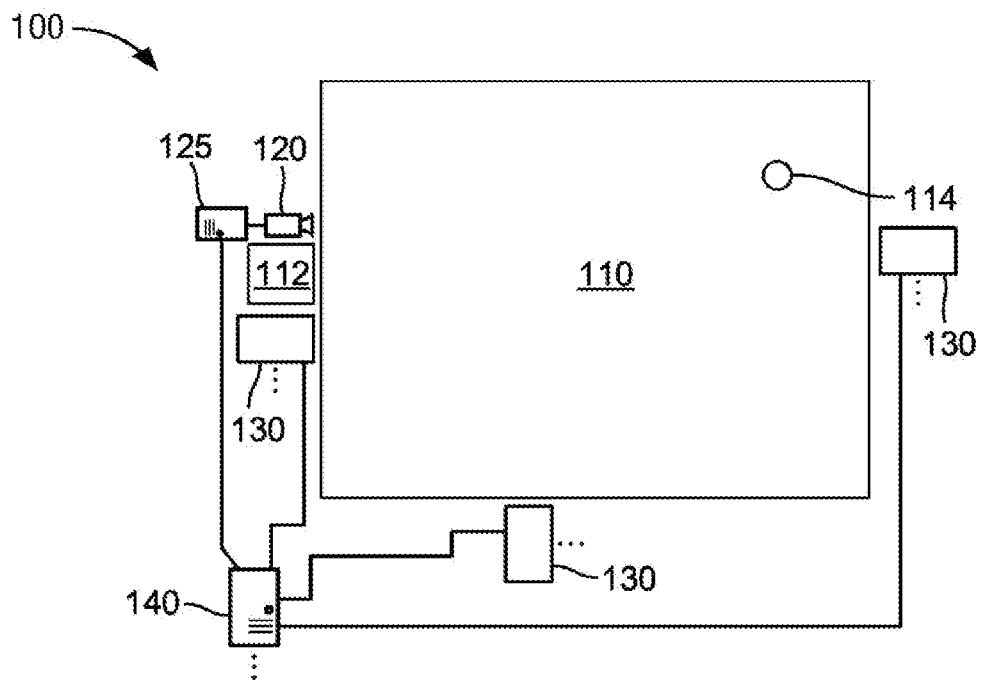


Figura 1A

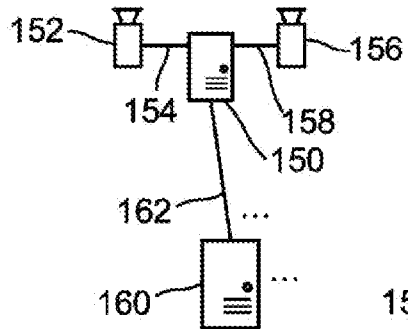


Figura 1B

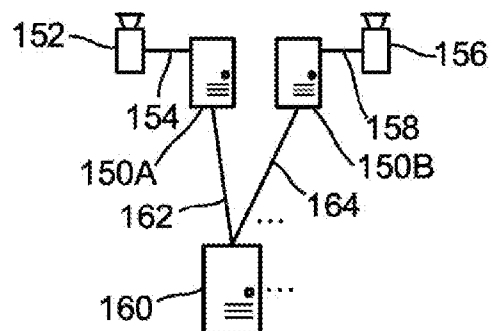


Figura 1C

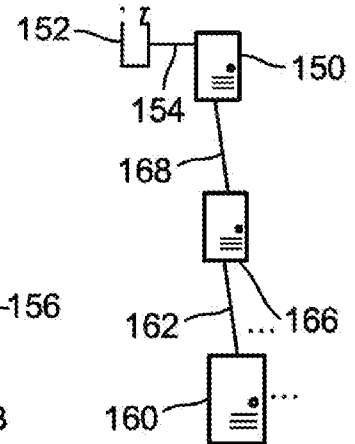


Figura 1D

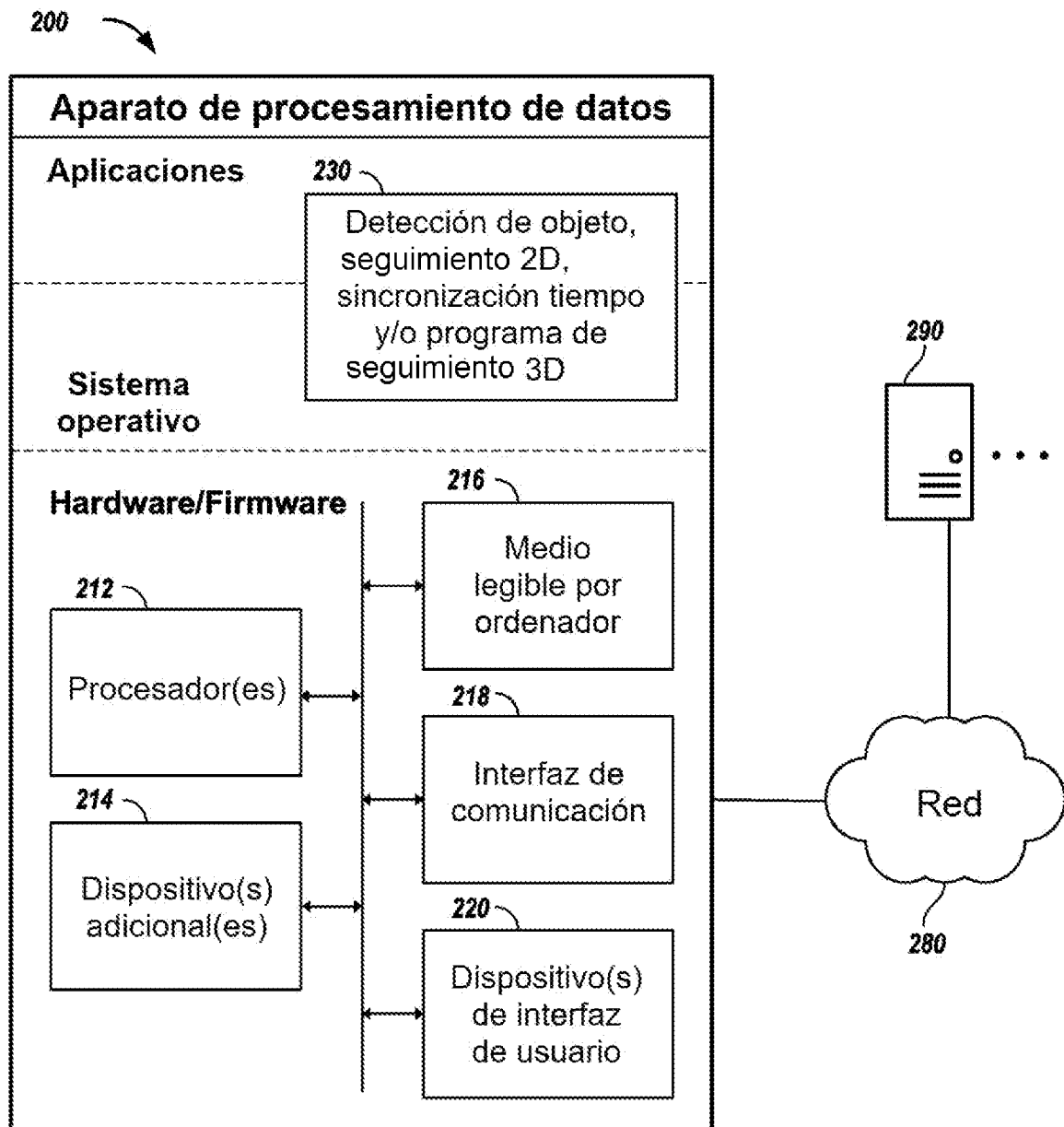
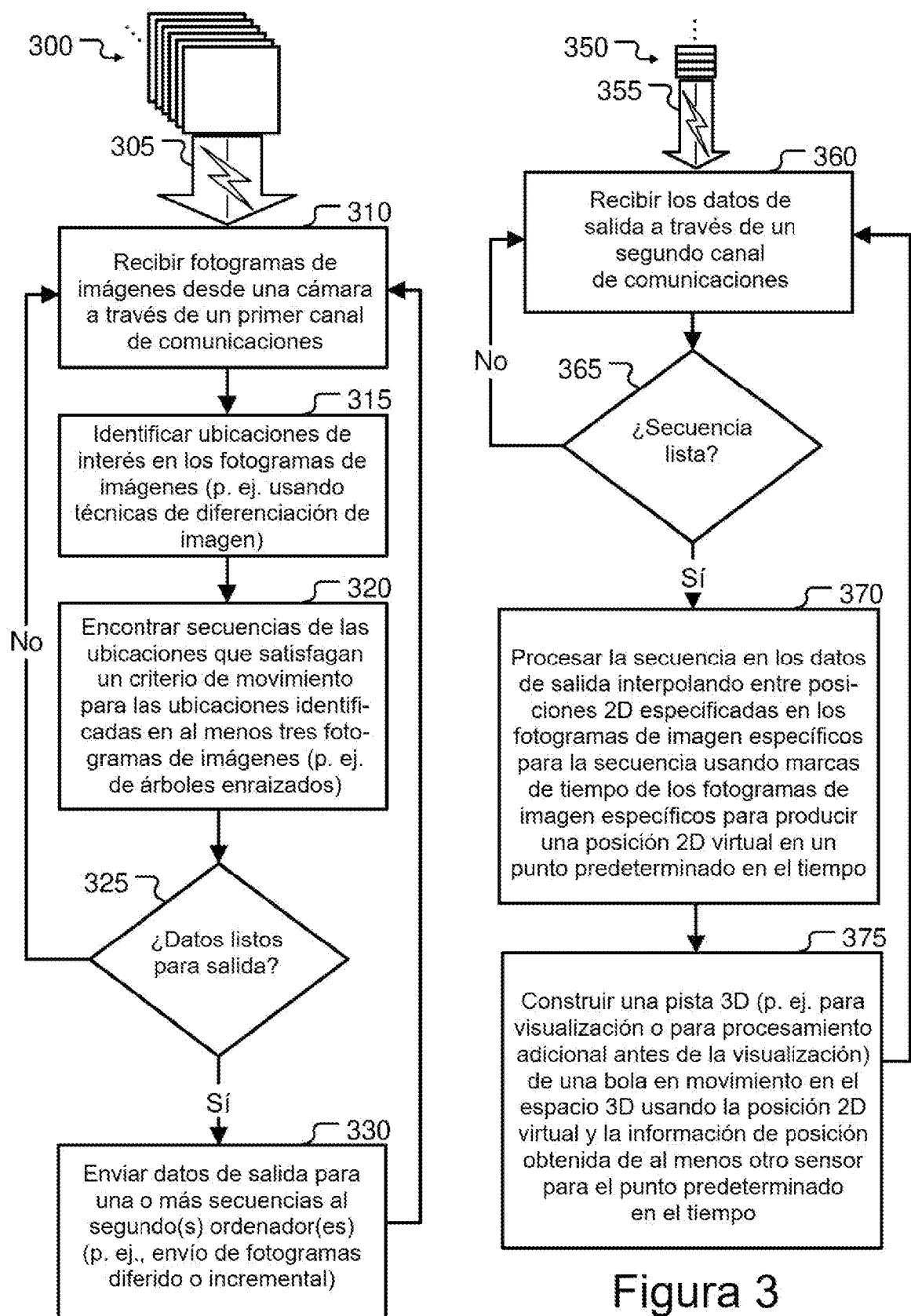
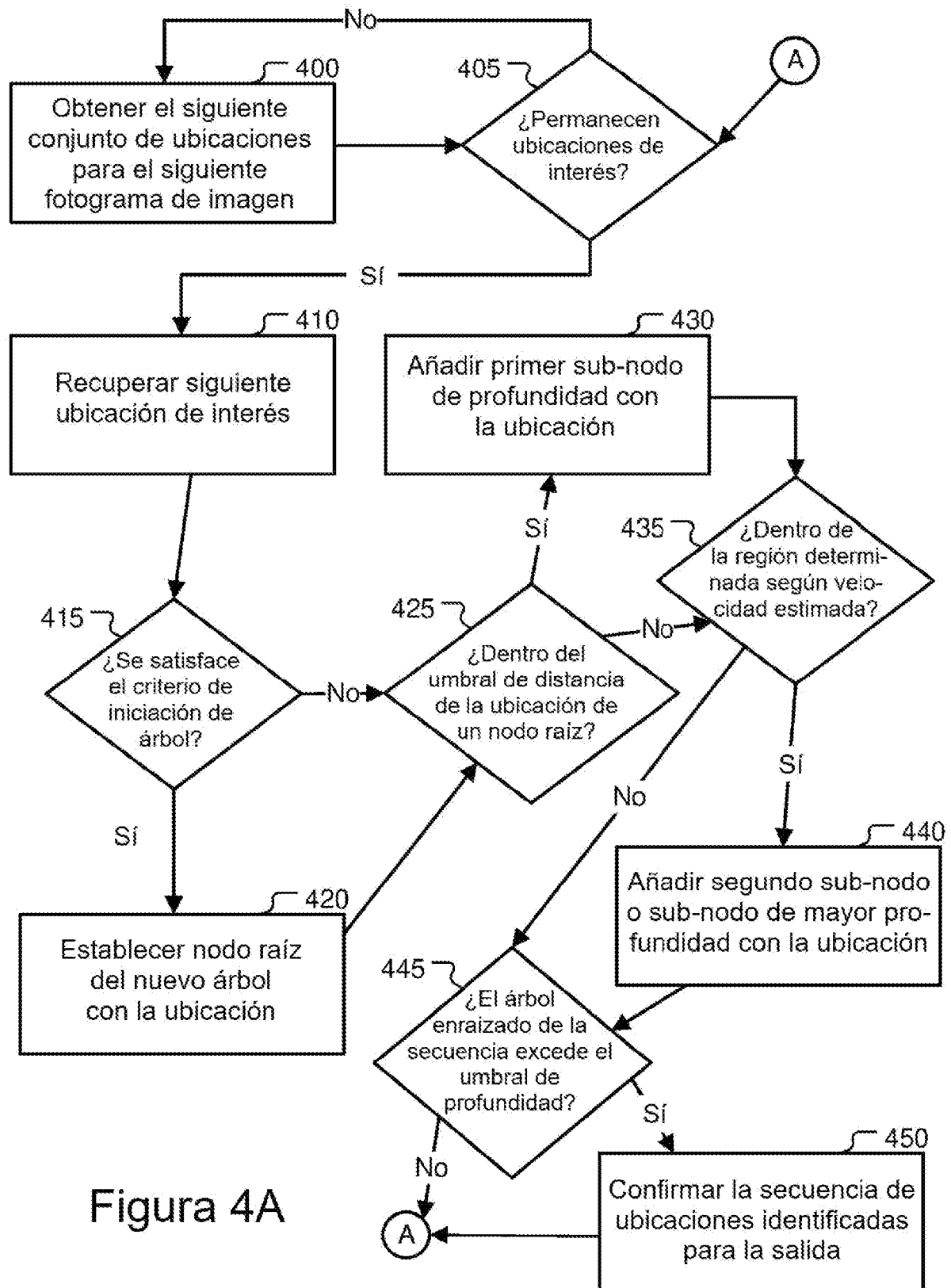
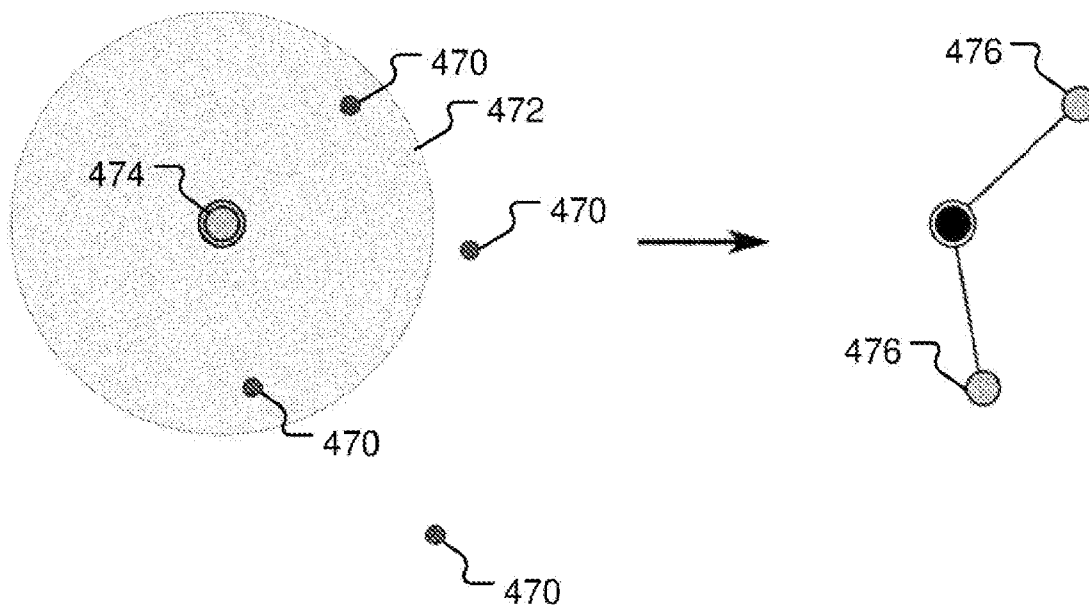
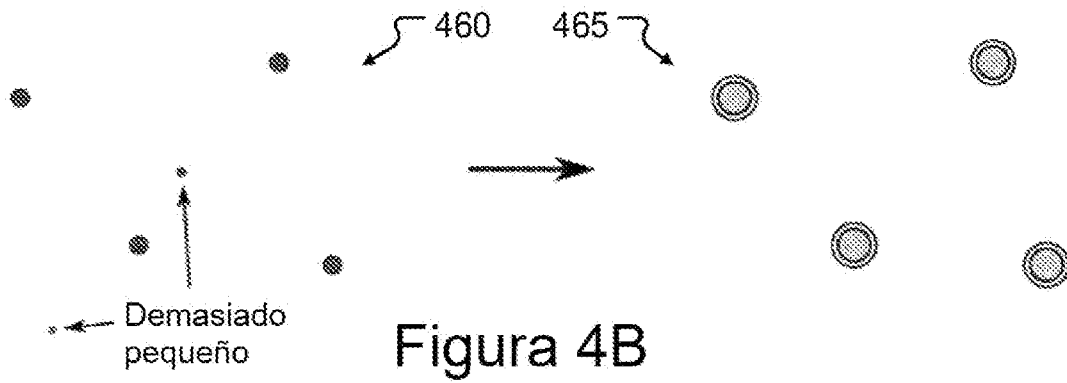


Figura 2







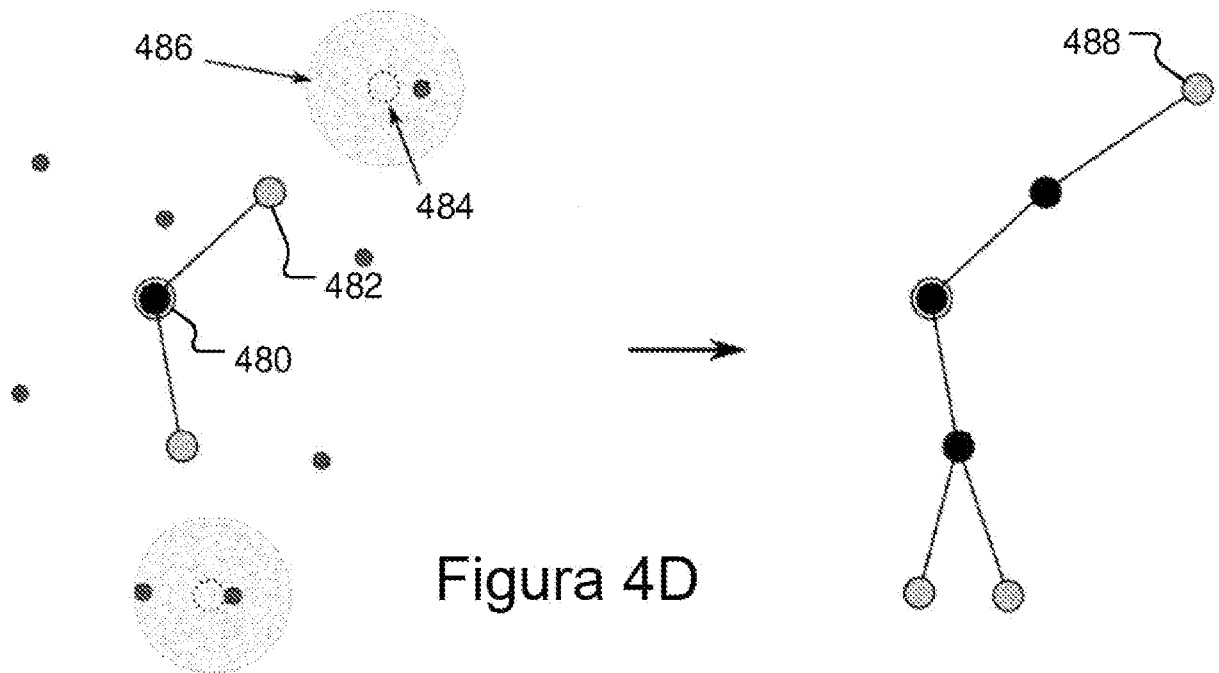
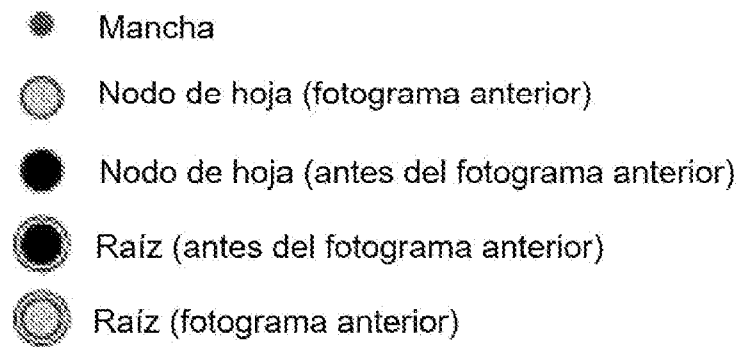


Figura 4D

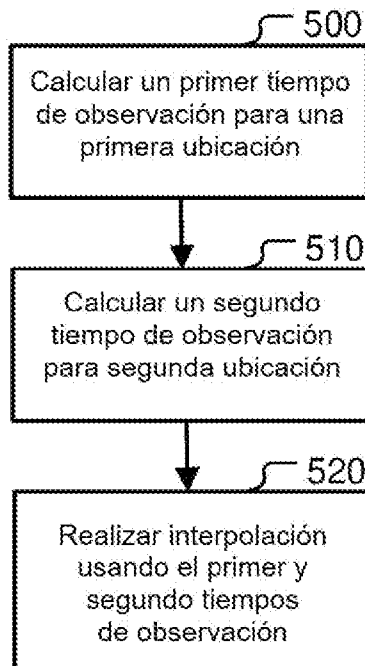


Figura 5

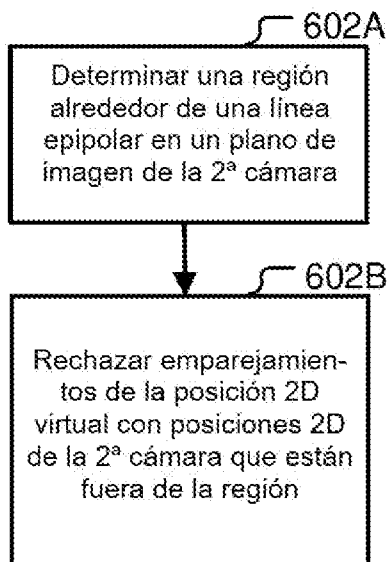


Figura 6B

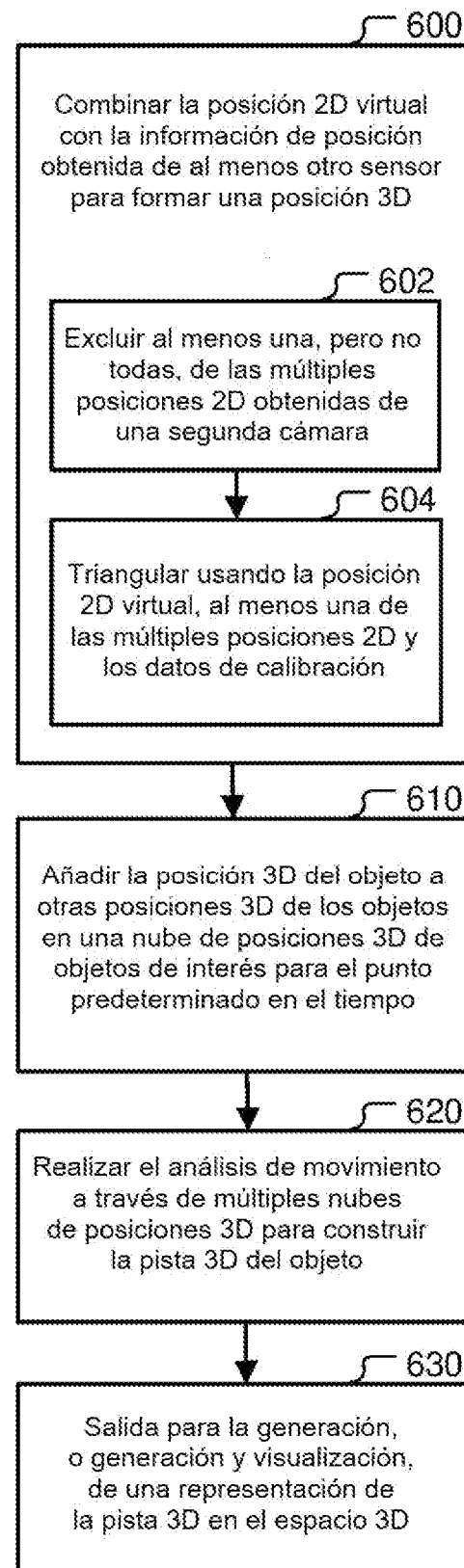


Figura 6A