

19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 599 858**

51 Int. Cl.:

H04N 19/597 (2014.01)

H04N 13/00 (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

86 Fecha de presentación y número de la solicitud internacional: **23.03.2007 PCT/IB2007/051031**

87 Fecha y número de publicación internacional: **11.10.2007 WO07113725**

96 Fecha de presentación y número de la solicitud europea: **23.03.2007 E 07735242 (5)**

97 Fecha y número de publicación de la concesión europea: **24.08.2016 EP 2005757**

54 Título: **Codificación eficaz de múltiples vistas**

30 Prioridad:

31.03.2006 EP 06112096

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

03.02.2017

73 Titular/es:

**KONINKLIJKE PHILIPS N.V. (100.0%)
HIGH TECH CAMPUS 5
5656 AE EINDHOVEN, NL**

72 Inventor/es:

**BRULS, WILHELMUS, H., A.;
VAREKAMP, CHRISTIAAN y
BRASPENNING, RALPH**

74 Agente/Representante:

ISERN JARA, Jorge

ES 2 599 858 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

Codificación eficaz de múltiples vistas

5 La invención se refiere a un método de codificación de múltiples vistas de una imagen en una señal de imagen, tal como por ejemplo una señal de televisión comprimida de acuerdo con una de las normas MPEG.

10 La invención también se refiere a: un aparato para generar una señal de este tipo, un receptor para recibir una señal de este tipo, un método de extracción de la información codificada de la señal, de tal manera que puede usarse para generar las múltiples vistas, y la propia señal codificada eficazmente.

15 En la actualidad se está trabajando en la estandarización de la codificación de la información de imágenes tridimensionales. Por ejemplo, Redert A et al: "ATTEST: advanced three-dimensional television system technologies" Visualización y transmisión de procesamiento de datos 3D, 2002. Actas. Primer Simposio internacional los días 19-21 de junio de 2002, Piscataway, NJ, US, IEEE, 19 de junio de 2002 (2002-06-19), páginas 313-319, XP010596672 ISBN: 0-7695-1521-4 trata los elementos de un sistema de televisión tridimensional para entornos de difusión. El artículo Christoph Fehn ET AL: "Study of some MPEG Tools Related to 3D-Video" ISO MPEG, documento M8423, 30 de Abril de 2002 (2002-04-30), páginas 1-5, XP030037386 Fairfax, US ofrece una breve visión general de algunas herramientas definidas dentro de la norma MPEG y que son pertinentes para aplicaciones de vídeo tridimensional.

20 Existen varias formas de representar un objeto tridimensional, por ejemplo, como un conjunto de vóxeles (por ejemplo, popular en una pantalla de datos médicos o una inspección de componentes industriales), o como una serie de imágenes de vistas capturadas desde diferentes direcciones y destinadas a verse desde diferentes direcciones, por ejemplo, por los dos ojos de un solo espectador o por diversos espectadores, o un espectador en movimiento, etc.

25 Un formato popular es el formato de izquierda/derecha, en el que se captura un fotograma por una cámara de la izquierda y se captura un fotograma por la cámara de la derecha. Estos fotogramas pueden visualizarse en diferentes pantallas, por ejemplo, el fotograma de la izquierda puede mostrarse durante un primer conjunto de instancias de tiempo, y el fotograma de la derecha durante un segundo conjunto entrelazado de instancias de tiempo, estando los ojos izquierdo y derecho del espectador bloqueados de manera sincronizada con la visualización por unas gafas de obturación. Un proyector con unos medios de polarización es otro ejemplo de una pantalla capaz de generar una impresión tridimensional de una escena, al menos de reproducir parte de la información tridimensional de la escena, es decir, lo que aproximadamente se ve como en una dirección determinada (es decir, en estéreo).

30 Diferentes calidades de aproximación de la escena pueden emplearse, por ejemplo, la escena 3 D puede representarse como un conjunto de capas planas una detrás de la otra. Sin embargo, estas diferentes calidades pueden codificarse por los formatos existentes.

35 Otra pantalla popular es la pantalla auto estereoscópica. Esta pantalla se forma, por ejemplo, colocando una pantalla LCD detrás de un conjunto de lentes, de tal manera que un grupo de píxeles se proyecta hacia una región en el espacio por una lente respectiva. De esta manera se genera un número de conos en el espacio que de dos en dos contienen imágenes izquierda y derecha para un ojo izquierdo y derecho, de manera que sin necesidad de gafas un usuario puede colocarse el mismo en un número de regiones en el espacio, y percibir 3D. Sin embargo, los datos de estos grupos de píxeles tienen que generarse a partir de las imágenes izquierda y derecha. Otra opción es que un usuario puede ver un objeto desde una serie de direcciones intermedias entre la vista izquierda y derecha de la codificación estéreo, vistas intermedias que pueden generarse calculando un campo de disparidad entre el fotograma de la izquierda y el de la derecha, y, posteriormente, interpolarlas. El documento WO 02/097733 describe una representación de este tipo de imágenes en 3D de múltiples ángulos mediante una imagen normal, una imagen de profundidad, y las imágenes adicionales que corresponden a diferentes puntos de vista.

40 Es una desventaja de la codificación a izquierda/derecha de la técnica anterior que se requieran muchos datos para obtener las vistas intermedias, y que aun así se obtengan resultados algo decepcionantes. Es difícil calcular un campo de disparidad precisamente coincidente, que dé lugar a artefactos en las interpolaciones, tales como unas partes de un fondo que se pega a un objeto en primer plano. Un deseo que llevó a las siguientes realizaciones tecnológicas presentadas en el presente documento era tener una forma de codificación que pueda llevar a resultados relativamente precisos al convertir a diferentes formatos, tal como a un conjunto de vistas con vistas intermedias y que sin embargo no comprenda una cantidad excesiva de datos.

45 Estos requisitos se cumplen al menos parcialmente mediante un método de codificación de información de imágenes de múltiples vistas en una señal de imagen (200) que comprende:

50 - añadir a la señal de imagen (200) una primera imagen (220) de los valores de píxel que representa uno o más objetos (110, 112) capturados por una primera cámara (101);

- añadir a la señal de imagen (200) un mapa (222) que comprende, para los conjuntos de píxeles respectivos de la primera imagen (220) unos valores respectivos que representan una posición tridimensional en el espacio de una región del uno o más objetos (110, 112) representados por el conjunto de píxeles respectivo;
- proporcionar una segunda imagen (224) de los valores de píxel que representan dicho uno o más objetos (110, 112) capturados por una segunda cámara (102) en una localización diferente de la primera cámara;
- determinar qué regiones están presentes en la segunda imagen y no en la primera imagen (220) en respuesta a las disparidades entre la primera imagen (220) y la segunda imagen (224); y
- añadir a la señal de imagen (200) una representación parcial (223) de la segunda imagen (224), comprendiendo la representación parcial (223) al menos la mayoría de los píxeles de las regiones determinadas, y que especifica las regiones de la segunda imagen (224) que no necesitan codificarse, y una señal obtenida por el método o un aparato que permite la realización del método.

Los inventores se han dado cuenta de que si uno entiende que por razones de calidad es mejor añadir a las imágenes a izquierda y derecha un mapa que contiene la información sobre la estructura tridimensional de la escena, lo que representa al menos esta parte de la información de escena tridimensional que se requiere para permitir la aplicación específica (con la calidad deseada), puede concebirse un formato de codificación interesante. Para la interpolación de vistas, el mapa puede ser, por ejemplo, un mapa de disparidad segmentado con precisión, los vectores de disparidad que llevarán a una buena interpolación de las vistas intermedias. Es importante observar que este mapa puede ajustarse de manera óptima en el lado de la creación/transmisión de acuerdo con su uso en el lado receptor, es decir, por ejemplo, de acuerdo con cómo se simule el entorno tridimensional en la pantalla, lo que significa que tendrá normalmente propiedades diferentes que cuando se usa para predecir de manera óptima las regiones de píxeles en la vista a izquierda y derecha.

El mapa puede, por ejemplo, ajustarse, o incluso crearse, por un operador humano, que puede obtener una vista previa a su lado como se comportaría una serie de pantallas previstas al recibir la señal. Hoy en día, y en el futuro incluso más aún, una parte del contenido ya se genera por ordenador, tal como por ejemplo un modelo tridimensional de un dinosaurio, o unas gráficas superpuestas, lo que significa que no es demasiado problemático crearlos al menos para las regiones que contienen este tipo de mapas de disparidad precisos de píxel de objetos hechos por el hombre, o mapas de profundidad, o mapas similares.

Esto es realmente cierto para las aplicaciones de juegos, en las que, por ejemplo, un usuario puede moverse ligeramente en comparación con la escena, y puede ver la escena de manera diferente, pero en un futuro próximo la invención también puede llegar a ser importante para la televisión 3D, capturada con dos cámaras, o incluso generada sobre la base de, por ejemplo, el paralaje de movimiento. Ya un número creciente de estudios (por ejemplo, la BBC) están usando, por ejemplo, los entornos virtuales para las noticias.

Este mapa puede codificarse con poca sobrecarga de datos, por ejemplo, como una imagen de valores de gris, comprimidos de acuerdo con la norma MPEG-2, y añadidos a la imagen a izquierda/derecha (o imágenes durante diversos instantes de tiempo de vídeo en movimiento) ya en la señal.

Teniendo este mapa, sin embargo, los inventores se dieron cuenta, que permite una reducción adicional de la cantidad de datos, debido a que una parte de la escena se imagina por ambas cámaras. Aunque la información de píxel puede ser útil para la interpolación bi-direccional (por ejemplo, pueden mitigarse las reflexiones especulares hacia una de las cámaras), de hecho no tanta información importante estará presente en las partes doblemente codificadas. Por lo tanto, al tener disponible el mapa, puede determinarse qué partes de la segunda imagen (por ejemplo, la imagen de la derecha) necesitan codificarse (y transmitirse), y qué partes son menos relevantes para la aplicación específica. Y en el lado del receptor puede realizarse una reconstrucción de buena calidad de los datos que faltan.

Por ejemplo, en una simple aproximación de escena (captura), con un objeto con una cara esencialmente plana hacia las cámaras (que puede colocarse en paralelo o en un pequeño ángulo hacia la escena), y no demasiado cerca, la parte que falta en la primera imagen (a la izquierda), que se captura en la segunda imagen (a la derecha) se compone de los píxeles de un objeto de fondo (por ejemplo, los elementos de la escena en el infinito).

Una realización interesante implica la codificación de un segundo mapa de disparidad o de profundidad parcial, o similar. Por ejemplo, este mapa de profundidad parcial contendrá sustancialmente los valores de profundidad de la región que no podía imaginarse por la primera cámara. A partir de estos datos de profundidad, a continuación puede deducirse en el lado receptor qué parte no cubierta pertenece a un objeto en primer plano que tiene una primera profundidad (indicada por 130 en la figura 1), y qué parte pertenece al fondo (132). Esto puede permitir mejores estrategias de interpolación, por ejemplo, la cantidad de estiramiento y relleno de los huecos puede ajustarse de manera fina, una representación pseudo-perspectiva de una oreja puede representarse en la imagen intermedia en lugar de solo en los píxeles de fondo, etc. Otro ejemplo es que la distorsión trapezoidal de las cámaras anguladas puede codificarse en este segundo mapa para la compensación del lado del receptor.

En el caso de una deformación trapezoidal procedente de una captura con cámaras convergentes (por lo general ligeramente), habrá, en general, una disparidad vertical en adición a una horizontal. Esta componente vertical puede

codificarse vectorialmente, o en un segundo mapa, como ya se ha previsto, por ejemplo, en las propuestas "auxiliary data representation" del subgrupo MPEG-4 Video-3DAV (por ejemplo, ISO/IEC JTC1/SC29/WG11 documentos. MPEG2005/12603, 12602, 12600, 12595). Los componentes de la disparidad pueden mapearse a las señales de luminancia y/o crominancia de un fotograma auxiliar, por ejemplo, la disparidad horizontal puede mapearse con una alta resolución a la luminancia, y las disparidades verticales pueden mapearse con un esquema a uno o dos componentes de crominancia (de tal manera que algunos de los datos están en la U y por una fracción de matemática como gran parte de los datos adicionales en la V).

Las ventajas de un formato a izquierda + derecha + "profundidad" parcial sobre, por ejemplo, una primera codificación a una vista central + "profundidad" + datos de oclusión son las siguientes. Al transformar los datos de oclusión para la vista central, en lugar de almacenarlos en una vista de captura de cámara original, se lleva al procesamiento a inexactitudes (en particular si el mapa(s) de profundidad se obtiene de manera automática y de menor calidad/consistencia, teniendo imperfecciones temporales y espaciales), y por lo tanto una ineficacia de codificación. Además, en el cálculo de una visión intermedia otras inexactitudes vendrán en la parte superior de la misma.

Estos y otros aspectos del método y el aparato de acuerdo con la invención serán evidentes a partir de y se aclararán con referencia a las implementaciones y las realizaciones descritas a continuación en el presente documento, y con referencia a los dibujos adjuntos, que sirven únicamente como ilustraciones específicas no limitativas que ejemplifican el concepto más general, y en las que los guiones se usan para indicar que un componente es opcional, no siendo necesariamente esenciales los componentes sin guiones.

En los dibujos:

- la figura 1 ilustra esquemáticamente la captura de una escena con al menos dos cámaras;
- la figura 2 ilustra esquemáticamente varias opciones de la codificación de los datos requeridos en la señal de imagen;
- la figura 3 ilustra esquemáticamente un aparato a modo de ejemplo para generar la señal de imagen; y
- la figura 4 ilustra esquemáticamente un aparato de recepción a modo de ejemplo capaz de usar la señal.

La figura 1 muestra una primera cámara 101 que captura una primera imagen de una escena que comprende un objeto cercano 110 y un objeto lejano 112. Su campo de visión está delimitado por las líneas 103 y 104. Su vista del fondo está ocluida por el objeto más cercano, es decir, la región 132 en el lado izquierdo de la tangente 120 no es visible. Sin embargo, una segunda cámara 102 es capaz de capturar una parte de esta región 132, en una segunda imagen, que puede, por razones de simplicidad, considerarse y llamarse la imagen de la derecha (pero esto no debería interpretarse como más estrecho que el que se ha capturado algo más a la derecha del otro fotograma). La segunda cámara también es capaz de capturar una parte adicional 130 del objeto más cercano 110.

La figura 2 muestra simbólicamente que estas imágenes capturadas se verán como un conjunto de píxeles. La señal de imagen 200 puede tener, por ejemplo, un formato de codificación prescrito de JPEG y contener un fotograma codificado de la escena, o puede ser una grabación de película codificada MPEG-4. En este último caso los datos 3D 210 comprenden la información necesaria para la reconstrucción de la escena en un instante temporal específico.

La imagen 220 es la imagen de la izquierda capturada por la primera cámara, que comprende un objeto más cercano 110 y un fondo 112.

El mapa 222 es un mapa que comprende toda la información relativa a cómo los objetos se colocan en su espacio tridimensional, que comprende, al menos, la información necesaria para la reproducción de un número de vistas necesarias (estática o dinámicamente, por ejemplo, en una interacción con un usuario en movimiento en un juego) en una pantalla. Varias de estas representaciones son posibles, por ejemplo, pueden ser un mapa de profundidad, que comprende por ejemplo una distancia ortogonal aproximada (por ejemplo, el promedio sobre todas las regiones de objeto) al centro de la cámara del objeto en el fondo, en sus posiciones de dos dimensiones como se percibe por la primera cámara, o pueden ser una disparidad o un paralaje, o solo un componente horizontal de la disparidad.

La profundidad y el paralaje etc. pueden relacionarse entre sí matemáticamente.

Este mapa de profundidad puede ser, por ejemplo, un píxel preciso o puede tener un solo valor para cada bloque de 8x8 píxeles, y puede codificarse, por ejemplo, como una imagen.

Una información adicional puede añadirse al mapa de profundidad (que puede comprender unos escalares o tuplas por conjunto de píxeles, incluyendo posiblemente un conjunto de solo un único píxel), tales como por ejemplo los datos de exactitud (en cómo de fiable es una cierta parte del mapa de profundidad) determinados sobre la base del algoritmo de coincidencia para obtenerlo.

- La estructura de datos parcial 223 (una parte de la imagen de la derecha 224) comprende la información de los píxeles (por ejemplo, la luminancia solamente, o el color, o cualquier otra representación usual, tal como por ejemplo, un modelo de textura, capaz de generar píxeles en una región) del fondo que pueden verse solamente por la segunda cámara (adyacente al objeto 225 más cercano cambiado de paralaje). Este región parcial codificada o al menos los datos necesarios para obtener los valores de píxel en una parte de una gran región codificada formada de acuerdo con un algoritmo de generación de parches de imagen puede ser algo menor que la actual región deocluída capturada en la imagen de la derecha, en el caso de que el aplicación del lado del receptor pueda tolerar algunos píxeles perdidos, por ejemplo, generándoles con una simple extrapolación, estiramiento, etc.
- La región codificada también puede ser mayor (por ejemplo, hasta el doble de la anchura y un tamaño de búfer similar añadido en la dirección vertical). Esto puede ser interesante por ejemplo, en el caso de duda acerca de la exactitud de la forma cuando se obtiene automáticamente, o en el caso de que por alguna razón pueda desearse la interpolación bi-direccional.
- También puede ser por razones de codificación. Puede ser más barato codificar bloques enteros, y uno puede beneficiarse de los píxeles extra-codificados, mientras que la codificación de forma compleja puede ser costosa. Al respecto en el lado de transmisión, un análisis (semi-) automático o manual puede realizarse en los datos de la imagen de la derecha, lo que se propone como una salida de una etapa de obtención anterior para ser útil además a los datos de la imagen de la izquierda. Por ejemplo, uno puede mirar las propiedades de píxel para identificar un reflejo especular, y decidir codificar una región de píxeles que componen la reflexión en ambas imágenes.
- También puede analizarse la forma de las diferentes regiones mediante un análisis morfológico, en particular, puede determinarse el tamaño o la anchura de la región. Las regiones pequeñas pueden implicar una sobrecarga de codificación considerable, pero a menudo pueden aproximarse en el lado receptor con poca o ninguna información. Por lo tanto, las regiones pequeñas pueden omitirse de la segunda imagen parcial. Esto puede estar bajo el control de un operador humano, que comprueba el efecto de cada eliminación.
- La forma (de abarque o exacta) de la región puede, por ejemplo, codificarse con aproximación poligonal o una caja delimitadora, y los valores de píxel interior (textura) pueden codificarse directamente, o por los coeficientes de una representación de transformación lineal sobre la forma, u otro modelo matemático. Además, pueden indicarse las partes que no necesitan codificarse/transmitirse.
- La representación parcial puede mapearse (por ejemplo, un simple desplazamiento en las líneas de corte, una transformación, o un corte en los sub-bloques que se vuelve a apilar de acuerdo con un orden predeterminado) sobre los datos de imagen o de usuario (por ejemplo, un modelo de regeneración) no usado para la primera imagen.
- Si la primera imagen con el acompañamiento del mapa de profundidad es una imagen central, puede haber unas segundas imágenes parciales para cada lado, es decir, a una cierta distancia angular (línea de base) entre las que pueden interpolarse.
- La primera cámara puede representar imágenes de un fondo y la segunda cámara puede representar imágenes del fondo con, por ejemplo, un lector de noticias que cubre parte de la misma, por ejemplo, desde el mismo punto vista a una hora diferente, es decir, las cámaras no necesitan ser cámaras reales presentes simultáneamente en un momento determinado, sino más bien, por ejemplo, una de las vistas puede descargarse, por ejemplo, desde un almacenamiento de fotogramas.
- Opcionalmente, al menos para la parte alrededor de las regiones de objetos no cubiertas representadas por imágenes en la segunda imagen puede añadirse un segundo mapa de profundidad 239 (una parte del mapa de profundidad total 240), o una representación similar a la señal. Este mapa de profundidad puede comprender un límite entre un objeto cercano y lejano. Con esta información, el lado de recepción puede añadir durante la interpolación los diferentes píxeles a las capas de objetos/profundidad correctas.
- También, pueden añadirse unos datos adicionales 230 a la señal, por ejemplo, en campos propietarios tales como información sobre la separación o en general de la composición tridimensional de los objetos en la escena. La indicación puede ser tan simple como una línea que siga un límite de objeto de escena representada por imágenes (si, por ejemplo, el mapa de profundidad no es suficiente o lo suficientemente preciso por sí mismo para demarcar los objetos), o incluso algo tan complicado como una malla metálica (por ejemplo, de la estructura de profundidad local en las partes deocluídas) o una información obtenida de la misma.
- También pueden incluirse la información de posición de cámara y la información de intervalo de escena, permitiendo que el lado de recepción haga reconstrucciones más avanzadas de las múltiples vistas (al menos dos).
- La figura 3 muestra un aparato 310 para generar la señal de imagen. Por lo general será un CI o una parte de un CI, o un procesador con software apropiado. El aparato puede estar comprendido en un aparato más grande tal como un aparato de autoría dedicado en un estudio, y puede conectarse a un ordenador, o puede estar comprendido en un ordenador. En la realización a modo de ejemplo, una primera cámara 301 y una segunda cámara 302 están

conectadas a la entrada del aparato 310. Cada cámara tiene un telémetro (308 resp. 309), que puede usar por ejemplo un haz láser o una rejilla proyectada, etc.

5 En el aparato existe una unidad de estimación de disparidad 312 que está dispuesta para determinar las disparidades entre al menos dos imágenes, al menos teniendo en cuenta la geometría del objeto (usando la información del mapa de profundidad). Se conocen diferentes técnicas de estimación de disparidad a partir de la técnica anterior, por ejemplo, por medio de una suma de las diferencias absolutas de los valores de píxeles en los bloques relacionados.

10 Está dispuesto para determinar al menos qué regiones están presentes en solo una de las imágenes y cuales están presentes en ambas, pero, además, puede tener unidades que sean capaces de aplicar los criterios de correspondencia a las regiones de los píxeles.

15 También puede haber una unidad de mapa de profundidad 314 capaz de generar y/o analizar y/o del refinar los mapas de profundidad (o una representación similar como un mapa de disparidad) o determinados por la unidad de estimación de disparidad 312, o extraídos de la señal de cámara introducida que contiene unos datos de intervalo de ejemplo. Puede comprender opcionalmente una unidad de representación 316, que puede generar, por ejemplo, vistas intermedias de tal manera que un artista de estudio puede comprobar el impacto de cualquier modificación y/o codificación más eficaz. Esto se realiza mediante la unidad de interfaz de usuario 318, lo que puede permitir, por ejemplo, que el usuario cambie los valores en la representación parcial 223, o cambie su forma (por ejemplo, haciendo que sea más grande o más pequeña). El usuario también puede modificar el mapa 222. Al respecto, puede conectarse una pantalla 335 y un medio de entrada de usuario. El aparato es capaz de transmitir la señal de imagen finalmente compuesta a una red 330 a través de los medios de transmisión y composición de señales 339, que un experto en la materia puede encontrar para la red apropiada (por ejemplo, la conversión a una señal de televisión implica una conversión ascendente a una frecuencia de transmisión, la transmisión a Internet implica una paquetización, puede haber más unidades de protección de errores, etc.).

30 La red presentada no debería interpretarse como limitativa, y está destinada también a comprender, por ejemplo, la transmisión a una unidad de memoria o medio de almacenamiento a través de una red interna de aparato tal como un bus.

La figura 4 muestra un receptor a modo de ejemplo 400, que puede ser de nuevo, por ejemplo, (una parte de) un CI, y que comprende medios para extraer la información relevante de la señal de imagen que puede recibirse de la red 330, al menos:

- 35 - medios (402) dispuestos para extraer la primera imagen (220) de los valores de píxel que representan uno o más objetos (110, 112) capturados por una primera cámara (101);
- medios (404) dispuestos para extraer de la señal de imagen (200) el mapa, por ejemplo, un mapa de profundidad que corresponde a las posiciones de objeto de la primera imagen; y
- 40 - medios (406) dispuestos para extraer la representación parcial (223) de la segunda imagen (224) de los valores de píxel que representan el uno o más objetos (110, 112) capturados por una segunda cámara (102).

45 Por supuesto, pueden estar presentes medios adicionales, ya que el receptor (y el método de extracción) puede duplicar cualquiera de las realizaciones posibles para la generación, por lo que pueden ser, por ejemplo, medios para extraer los datos adicionales, tales como la indicación de la frontera entre dos objetos.

50 Esta información extraída se transmite a un regenerador de imágenes, que puede generar por ejemplo, una imagen completa a izquierda y derecha. Una unidad de representación de imágenes 412 puede generar, por ejemplo, una vista intermedia (por ejemplo, mediante una interpolación mono- o bi-direccional, o cualquier otro algoritmo conocido), o las señales necesarias para dos vistas (estéreo) en una pantalla autoestereoscópica. En función del tipo de visualización en 3D y cómo se representa en realidad el 3D, estas dos unidades pueden realizarse en diferentes combinaciones.

55 El receptor puede normalmente conectarse a o estar comprendido en una pantalla 3D 415, que puede reproducir al menos dos vistas, o la señal(s) regenerada puede almacenarse en un dispositivo de memoria 420, por ejemplo, un escritor de disco 422, o en una memoria de estado sólido, etc.

60 Los componentes algorítmicos desvelados en este texto pueden realizarse en la práctica (por completo o en parte) como hardware (por ejemplo, las partes de un CI de aplicación específica) o como software que se ejecuta en un procesador de señal digital especial o un procesador genérico, etc.

65 Bajo un producto de programa informático debería entenderse cualquier realización física de un conjunto de comandos que permiten a un procesador genérico o de propósito especial, después de una serie de etapas de carga (que pueden incluir etapas de conversión intermedias, como la traducción a un lenguaje intermedio, y a un lenguaje de procesador final) obtener los comandos en el procesador, para ejecutar cualquiera de las funciones características de una invención. En particular, el producto de programa informático puede realizarse como datos en

un portador tal como, por ejemplo, un disco o una cinta, datos presentes en una memoria, datos que viajan a través de una conexión de red cableada o inalámbrica, o un código de programa en papel. Aparte del código de programa, los datos de características requeridos para el programa también pueden realizarse como un producto de programa informático.

- 5 Algunas de las etapas necesarias para el funcionamiento del método pueden estar ya presentes en la funcionalidad del procesador en lugar de descritas en el producto de programa informático, tales como las etapas de entrada y de salida de datos.
- 10 Debería observarse que las realizaciones mencionadas anteriormente ilustran en lugar de limitar la invención. Además de las combinaciones de los elementos de la invención como se combinan en las reivindicaciones, son posibles otras combinaciones de los elementos. Cualquier combinación de los elementos puede realizarse en un único elemento dedicado.
- 15 Cualquier signo de referencia entre paréntesis en la reivindicación no está destinado a limitar la reivindicación. La palabra "comprende" no excluye la presencia de elementos o aspectos no mencionados en una reivindicación. La palabra "un" o "una" precediendo un elemento no excluye la presencia de una pluralidad de tales elementos.

REIVINDICACIONES

1. Un método de codificación de información de imágenes de múltiples vistas en una señal de imagen (200) que comprende:
- 5
- añadir a la señal de imagen (200) una primera imagen (220) de los valores de píxel que representan uno o más objetos (110, 112) capturados por una primera cámara (101);
 - añadir a la señal de imagen (200) un mapa (222) que comprende para los conjuntos de píxeles respectivos de la primera imagen (220) unos valores respectivos, que representan una posición tridimensional en el espacio de

10

 - una región del uno o más objetos (110, 112) representados por el conjunto de píxeles respectivo; y
 - proporcionar una segunda imagen (224) de los valores de píxel que representan dicho uno o más objetos (110, 112) capturados por una segunda cámara (102) en una localización diferente de la primera cámara;
 - determinar qué regiones están presentes en la segunda imagen y no en la primera imagen (220) en respuesta a las disparidades entre la primera imagen (220) y la segunda imagen (224); y estando el método caracterizado por

15

 - que comprende:
 - añadir a la señal de imagen (200) una representación parcial (223) de la segunda imagen (224), comprendiendo la representación parcial (223) al menos una información de la mayoría de los píxeles de las regiones determinadas, y que especifica las regiones de la segunda imagen (224) que no necesitan codificarse.
- 20
2. Un método de acuerdo con la reivindicación 1, que comprende además añadir a la señal de imagen (200) un segundo mapa (240) que comprende al menos una representación parcial (239) de, para los conjuntos de píxeles respectivos de la segunda imagen (224), los valores respectivos que representan una posición tridimensional en el espacio de una región del uno o más objetos (110, 112) representados por el conjunto de píxeles respectivo.
- 25
3. Un método de acuerdo con la reivindicación 1 o 2, que comprende obtener de manera automática por medio de un algoritmo de cálculo de disparidad para los conjuntos de píxeles de la primera imagen (220), los vectores de disparidad que apuntan a los conjuntos de píxeles correspondientes en la segunda imagen (224), y la etapa de añadir a la señal de imagen (200) un mapa (222), que consiste en añadir el mapa (222) que comprende los vectores de disparidad.
- 30
4. Un método de acuerdo con la reivindicación 1 o 2, que comprende obtener, a partir de un dispositivo de determinación de intervalo (308), los intervalos respectivos para los conjuntos de píxeles de la primera imagen (220), y la etapa de añadir a la señal de imagen (200) un mapa (222), que consiste en añadir el mapa (222) que comprende los intervalos.
- 35
5. Un método de acuerdo con una de las reivindicaciones anteriores, en el que los valores del mapa (222) se ajustan de manera fina por un humano antes de la adición a la señal de imagen (200).
- 40
6. Un método de acuerdo con una de las reivindicaciones anteriores, en el que la representación parcial (223) la ajusta de manera fina un humano antes de la adición a la señal de imagen (200).
- 45
7. Un método de acuerdo con la reivindicación 2, que comprende además añadir a la señal de imagen (200) una indicación (250), tal como una curva que comprende las posiciones en el mapa (222), que indica un límite entre un primer objeto (110) y un segundo objeto (112).
- 50
8. Un método de acuerdo con una de las reivindicaciones anteriores, en el que la etapa de añadir a la señal de imagen (200) una representación parcial (223) de una segunda imagen (224), comprende especificar y añadir a la señal de imagen una representación de una forma envolvente que abarca la región de píxeles en la representación parcial (223) de una segunda imagen (224).
- 55
9. Un método de acuerdo con una de las reivindicaciones anteriores, en el que se realiza un análisis de imágenes tal como, por ejemplo, un análisis morfológico en las regiones comprendidas en la representación parcial (223) de una segunda imagen (224), y se realiza una modificación en la representación parcial (223) antes de añadir la representación parcial (223) a la señal de imagen (200), comprendiendo el análisis morfológico, por ejemplo, la determinación de la mayor anchura de las regiones respectivas.
- 60
10. Un aparato (310) para generar una codificación en una señal de imagen (200) de información de imágenes de múltiples vistas que comprende:
- medios (340) dispuestos para añadir a la señal de imagen (200) una primera imagen (220) de los valores de píxel que representan uno o más objetos (110, 112) capturados por una primera cámara (101);
 - medios (341) dispuestos para añadir a la señal de imagen (200) un mapa (222) que comprende, para los conjuntos de píxeles respectivos de la primera imagen (220), los valores respectivos que representan una posición tridimensional en el espacio de una región del uno o más objetos (110, 112) representados por el

65

 - conjunto de píxeles respectivo; y
 - medios dispuestos para proporcionar una segunda imagen (224) de los valores de píxel que representan uno o

más objetos (110, 112) capturados por una segunda cámara (102) en una localización diferente de la primera cámara;

- medios dispuestos para determinar al menos qué regiones están presentes en la segunda imagen y no en la primera imagen (220) en respuesta a las disparidades entre la primera imagen (220) y la segunda imagen (224); y estando el aparato caracterizado por que comprende:

- medios (342) dispuestos para añadir a la señal de imagen (200) una representación parcial (223) de la segunda imagen (224), comprendiendo la representación parcial (223) al menos una información de la mayoría de los píxeles de la regiones,

y que especifica las regiones de la segunda imagen (224) que no necesitan codificarse.

11. Un aparato de acuerdo con la reivindicación 10, que comprende además una unidad de estimación de disparidad automática dispuesta para estimar los vectores de disparidad entre el conjunto de píxeles de la primera imagen (220) y los conjuntos de píxeles correspondientes de la segunda imagen (224).

12. Un aparato de acuerdo con la reivindicación 10 u 11, que comprende además una unidad de interfaz de usuario (318) dispuesta para proporcionar un control de operador humano sobre los contenidos del mapa (222) y/o la representación parcial (223).

13. Un receptor de señal de imagen (400), que comprende:

- medios (402) dispuestos para extraer de una señal de imagen (200) una primera imagen (220) de los valores de píxel que representan uno o más objetos (110, 112) capturados por una primera cámara (101);

- medios (404) dispuestos para extraer de la señal de imagen (200) un mapa (222) que comprende, para los conjuntos de píxeles respectivos de la primera imagen (220), los valores respectivos que representan una posición tridimensional en el espacio de una región del uno o más objetos (110, 112) representados por el conjunto de píxeles respectivo;

- medios (406) dispuestos para extraer de la señal de imagen (200) una representación parcial (223) de una segunda imagen (224) de los valores de píxel que representan uno o más objetos (110, 112) capturados por una segunda cámara (102) en una localización diferente de la primera cámara, y caracterizado por que comprende además: la representación parcial (223) que comprende al menos una información de la mayoría de los píxeles que representan las regiones del uno o más objetos (110, 112) presentes en la segunda imagen y no en la primera imagen (220), y la señal de imagen (200) comprende para la segunda imagen (224) solo una parte (223) de la segunda imagen.

14. Una pantalla (415) capaz de generar al menos dos vistas de una imagen que comprende:

- un receptor de señal de imagen (400) de acuerdo con la reivindicación 13;

- un regenerador de imagen (410) dispuesto para generar dos imágenes a partir de los datos de señal de imagen recibidos por el receptor de señal de imagen (400); y

- una unidad de reproducción de imágenes (412), dispuesta para generar a partir de las dos imágenes unas imágenes adicionales de un formato apropiado para la pantalla.

15. Un método de extracción de información de imágenes de múltiples vistas a partir de una señal de imagen (200) que comprende:

- extraer a partir de la señal de imagen (200) una primera imagen (220) de los valores de píxel que representan uno o más objetos (110, 112) capturados por una primera cámara (101);

- extraer a partir de la señal de imagen (200) un mapa (222) que comprende, para los conjuntos de píxeles respectivos de la primera imagen (220) los valores respectivos que representan una posición tridimensional en el espacio de una región del uno o más objetos (110, 112) representados por el conjunto de píxeles respectivo;

- extraer a partir de la señal de imagen (200) una representación parcial (223) de una segunda imagen (224) de los valores de píxel que representan uno o más objetos (110, 112) capturados por una segunda cámara (102) en una localización diferente de la primera cámara, y estando el método caracterizado por que comprende además: la representación parcial (223) que comprende al menos una información de la mayoría de los píxeles que representan las regiones del uno o más objetos (110, 112) presentes en la segunda imagen y no en la primera imagen (220), y la señal de imagen (200) comprende para la segunda imagen (224) solo una parte (223) de la segunda imagen.

16. Una señal de imagen (200) que comprende:

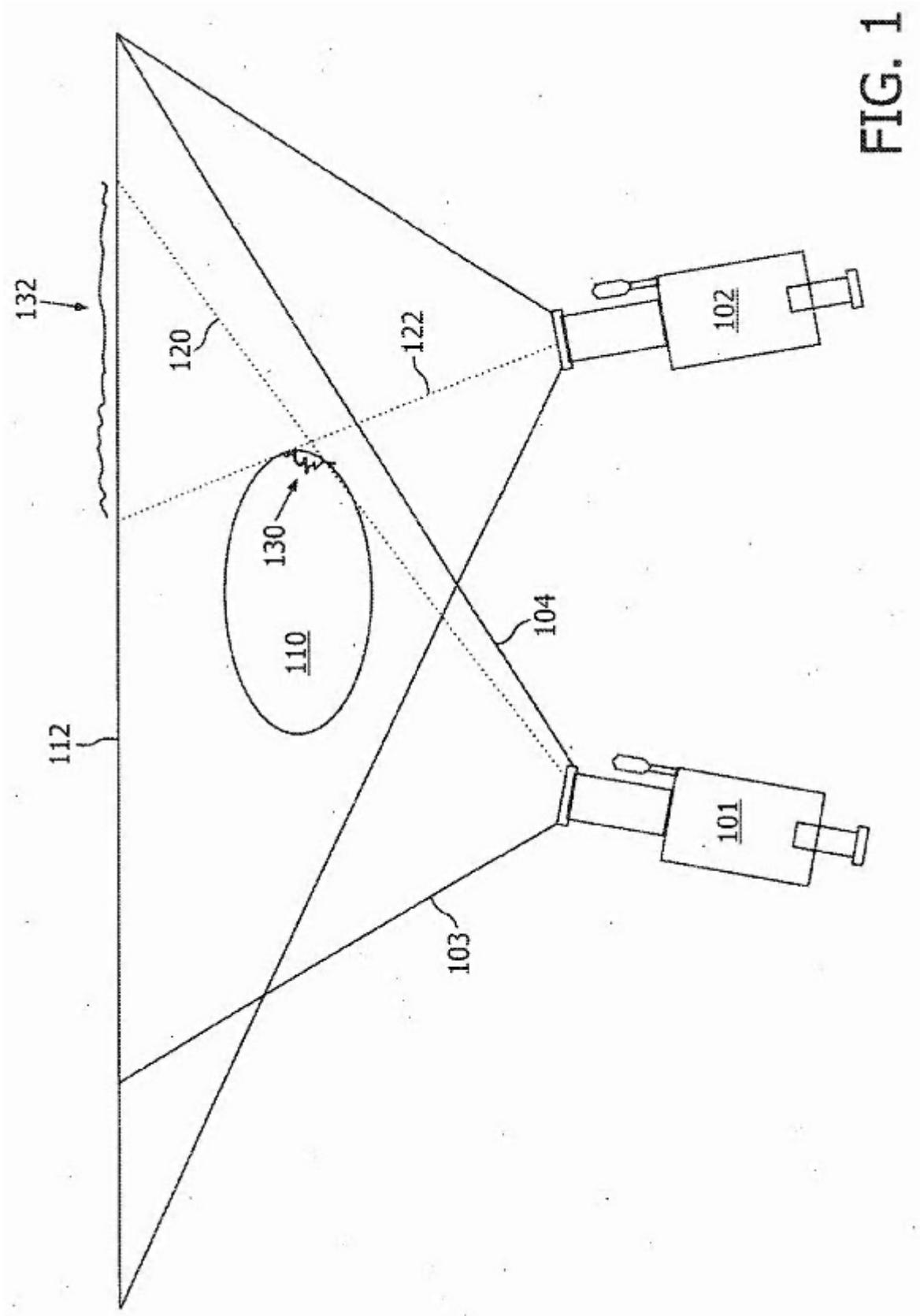
- una primera imagen (220) de los valores de píxel que representan uno o más objetos (110, 112) capturados por una primera cámara (101);

- un mapa (222) que comprende, para los conjuntos de píxeles respectivos de la primera imagen (220), los valores respectivos que representan una posición tridimensional en el espacio de una región del uno o más objetos (110, 112) representados por el conjunto de píxeles respectivo; y caracterizado por que comprende:

- una representación parcial (223) de solo una parte de una segunda imagen (224) de los valores de píxel que

representan uno o más objetos (110, 112) capturados por una segunda cámara (102) en una localización diferente de la primera cámara, comprendiendo la representación parcial (223) al menos una información de la mayoría de los píxeles que representan las regiones del uno o más objetos (110, 112) presentes en la segunda imagen y no en la primera imagen (220).

- 5
17. Una señal de imagen (200) de acuerdo con la reivindicación 16, en la que el mapa (222) se codifica en un valor de gris o un fotograma a color, codificándose al menos un escalar para los conjuntos de píxeles respectivos en una representación de luminancia o de crominancia del fotograma.
- 10
18. Una señal de imagen (200) de acuerdo con la reivindicación 16 o 17, en la que la representación parcial se mapea en unos datos de imagen o de usuario en relación con, pero no usados para, la primera imagen.



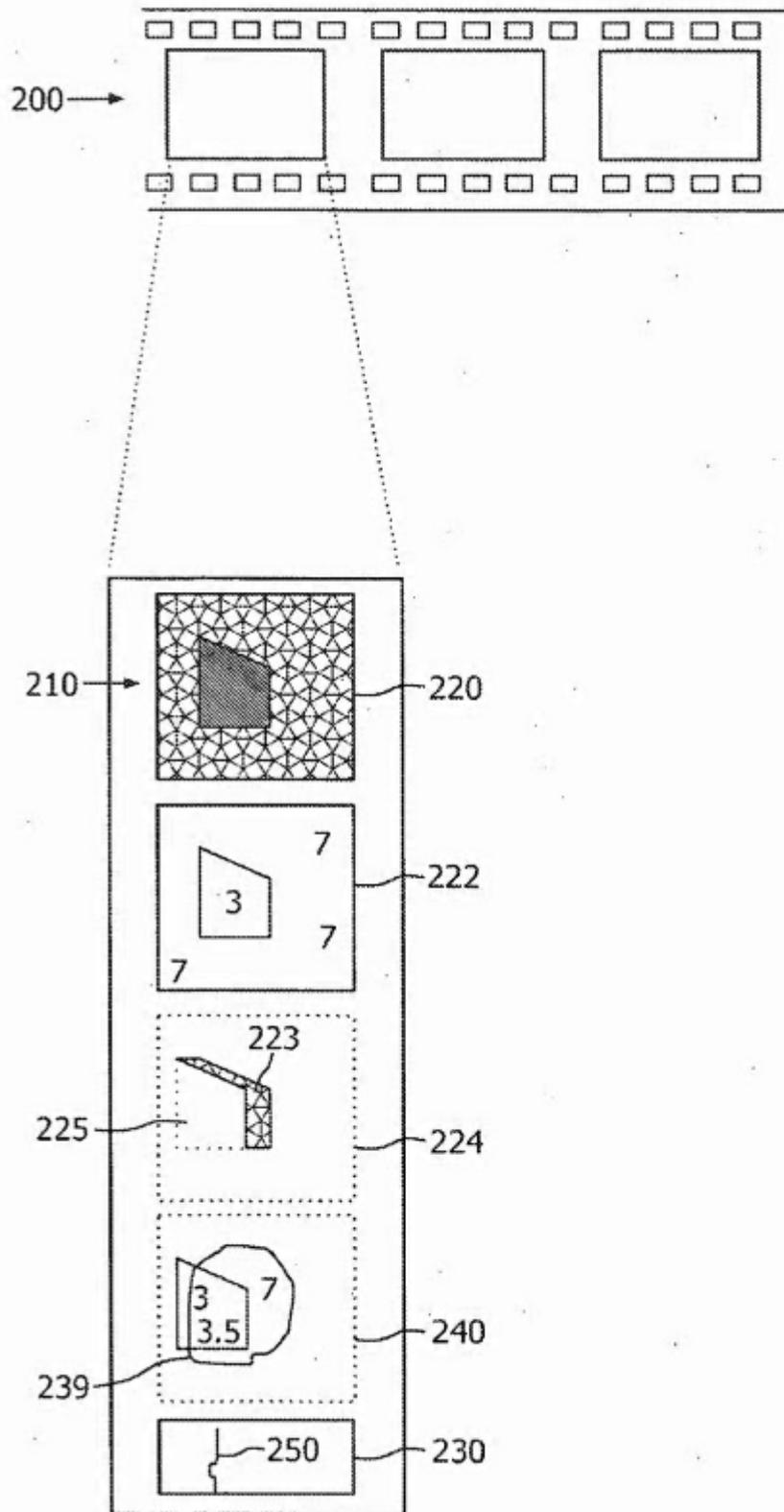


FIG. 2

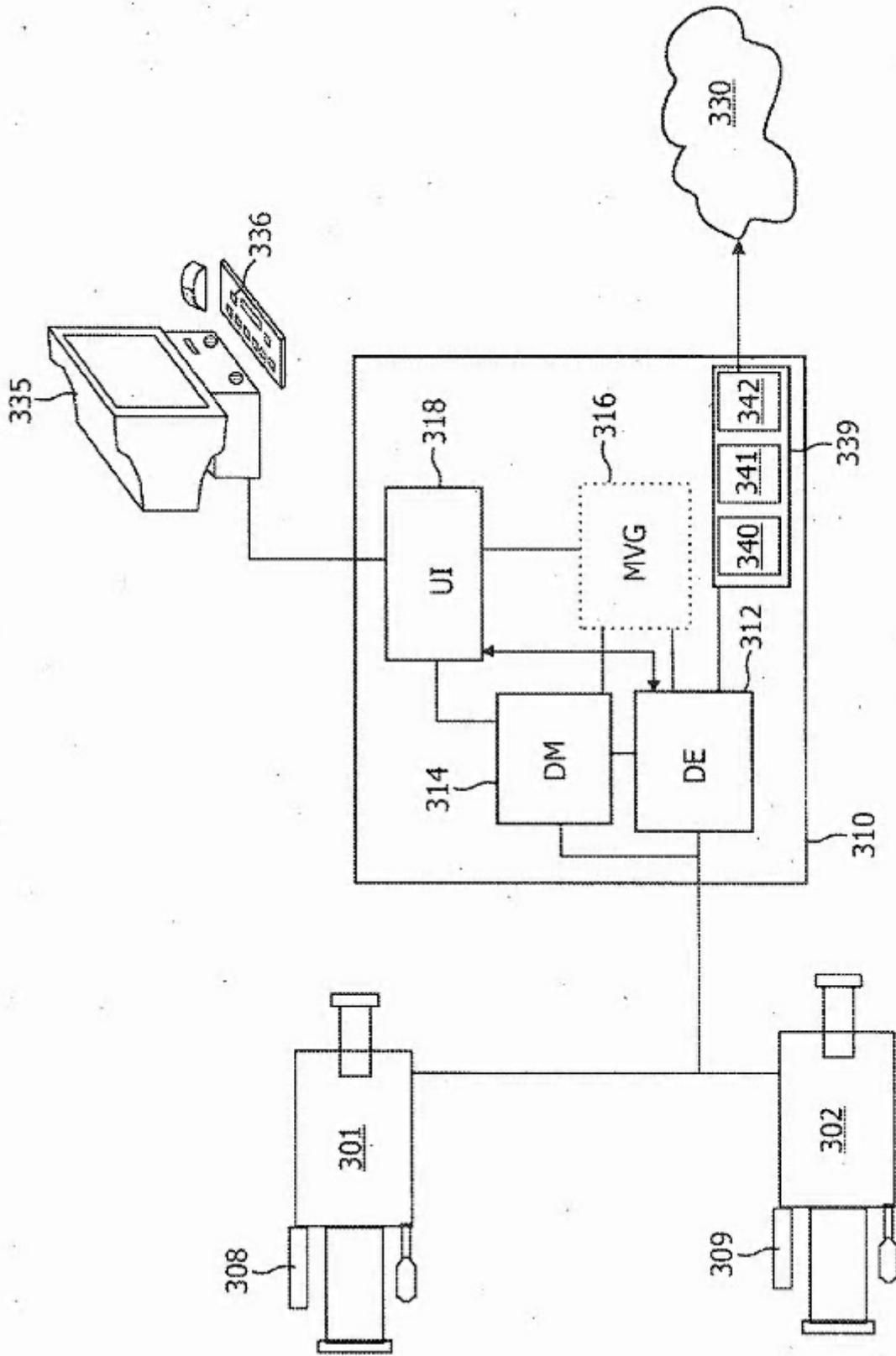


FIG. 3

