

19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 883 750**

21 Número de solicitud: 202190052

51 Int. Cl.:

H04N 13/117 (2008.01)

H04N 13/194 (2008.01)

H04N 13/204 (2008.01)

G06T 15/00 (2011.01)

H04N 13/366 (2008.01)

12

PATENTE DE INVENCION CON EXAMEN

B2

22 Fecha de presentación:

14.02.2020

30 Prioridad:

19.03.2019 EP 19163678

43 Fecha de publicación de la solicitud:

09.12.2021

88 Fecha de publicación diferida del informe sobre el estado de la técnica:

21.03.2023

Fecha de modificación de las reivindicaciones:

21.05.2024

Fecha de concesión:

05.06.2025

45 Fecha de publicación de la concesión:

12.06.2025

73 Titular/es:

**KONINKLIJKE PHILIPS N.V. (100.00%)
High Tech Campus 52
5656 AG Eindhoven NL**

72 Inventor/es:

**VAN GEEST, Bartholomeus, Wilhelmus,
Damianus y
KROON, Bart**

74 Agente/Representante:

GONZÁLEZ PECES, Gustavo Adolfo

54 Título: **SEÑAL DE IMAGEN QUE REPRESENTA UNA ESCENA**

ES 2 883 750 B2

Aviso: Se puede realizar consulta prevista por el art. 41 LP 24/2015.
Dentro de los seis meses siguientes a la publicación de la concesión en el Boletín Oficial de la Propiedad Industrial cualquier persona podrá oponerse a la concesión. La oposición deberá dirigirse a la OEPM en escrito motivado y previo pago de la tasa correspondiente (art. 43 LP 24/2015).

19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



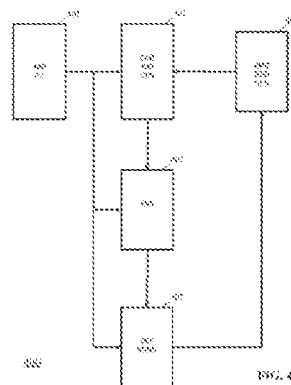
11 Número de publicación: **2 883 750**

21 Número de solicitud: 202190052

57 Resumen:

Señal de imagen que representa una escena.

La generación de una señal de imagen comprende un receptor (401) que recibe imágenes fuente que representan una escena. Un generador de imágenes combinadas (403) genera imágenes combinadas a partir de las imágenes fuente. Cada imagen combinada se deriva de solo partes de al menos dos imágenes de las imágenes fuente. Un evaluador (405) determina medidas de calidad de predicción para elementos de las imágenes fuente, donde la medida de calidad de predicción para un elemento de una primera imagen fuente es indicativa de una diferencia entre los valores de los píxeles en la primera imagen fuente y los valores de píxeles predichos para los píxeles en el elemento. Los valores de píxeles predichos son valores de píxeles resultantes de la predicción de píxeles de las imágenes combinadas. Un determinador (407) determina segmentos de las imágenes fuente que comprenden elementos para los que la medida de calidad de la predicción es indicativa de una diferencia por encima de un umbral. Un generador de señales de imagen (409) genera una señal de imagen que comprende datos de imagen que representan las imágenes combinadas y los segmentos de las imágenes fuente.



ES 2 883 750 B2

Aviso: Se puede realizar consulta prevista por el art. 41 LP 24/2015.
Dentro de los seis meses siguientes a la publicación de la concesión en el Boletín Oficial de la Propiedad Industrial cualquier persona podrá oponerse a la concesión. La oposición deberá dirigirse a la OEPM en escrito motivado y previo pago de la tasa correspondiente (art. 43 LP 24/2015).

DESCRIPCIÓN

Señal de imagen que representa una escena

Campo de la invención

5 La invención se refiere a una señal de imagen que representa una escena y en particular, pero no exclusivamente, a la generación de una señal de imagen que representa una escena y a la representación de imágenes a partir de esta señal de imagen como parte de una aplicación de realidad virtual.

Antecedentes de la invención

10

La variedad y el alcance de las aplicaciones de imagen y vídeo han aumentado considerablemente en los últimos años, con el continuo desarrollo e introducción de nuevos servicios y modos de usar y consumir vídeo.

15

Por ejemplo, un servicio cada vez más popular es el suministro de secuencias de imágenes de tal manera que el espectador pueda interactuar activa y dinámicamente con el sistema para modificar los parámetros de la representación. Una característica muy atractiva en muchas aplicaciones es la posibilidad de cambiar la posición efectiva de visualización y la dirección de visualización del espectador, tal como, por ejemplo, permitir que el espectador se mueva y "mire a su alrededor" de la escena que se está presentando.

20

Esta característica puede permitir específicamente proporcionar a un usuario una experiencia de realidad virtual. Esto puede permitir al usuario, por ejemplo, moverse de manera (relativamente) libre en un entorno virtual y cambiar dinámicamente su posición y hacia dónde mira. Normalmente, estas aplicaciones de realidad virtual se basan en un modelo tridimensional de la escena y el modelo se evalúa dinámicamente para proporcionar la vista específica solicitada. Este enfoque es bien conocido, por ejemplo, en las aplicaciones de juegos para ordenadores y consolas, como son las acciones en primera persona.

25

30

También es deseable, en particular para las aplicaciones de realidad virtual, que la imagen presentada sea una imagen tridimensional. En efecto, para optimizar la inmersión del espectador, se prefiere que el usuario experimente la escena presentada como una escena tridimensional. De hecho, una experiencia de realidad virtual debería permitir preferentemente

al usuario seleccionar su propia posición, el punto de visión de la cámara y el momento en el tiempo en relación con el mundo virtual.

5 Normalmente, las aplicaciones de realidad virtual están intrínsecamente limitadas en el sentido de que se basan en un modelo predeterminado de la escena, y normalmente en un modelo artificial de un mundo virtual. A menudo es deseable que se proporcione una experiencia de realidad virtual basándose en la captura del mundo real. Sin embargo, en muchos casos este enfoque está limitado o tiende a requerir que se construya un modelo virtual del mundo real a partir de las capturas del mundo real. La experiencia de realidad virtual se genera entonces evaluando este modelo.

10 Sin embargo, los enfoques actuales tienden a ser subóptimos y suelen requerir muchos recursos informáticos o de comunicación y/o proporcionan una experiencia de usuario subóptima con, por ejemplo, una calidad reducida o una libertad limitada.

15 En muchos sistemas, como cuando se basan de manera específica en una escena del mundo real, se proporciona una representación en imagen de la escena donde la representación en imagen incluye imágenes y profundidad para uno o más puntos de captura/puntos de visión en la escena. La representación en imagen más profundidad proporciona una caracterización muy eficiente, en particular, de una escena del mundo real donde la caracterización no solo es relativamente fácil de generar por la captura de la escena del mundo real, sino que también es muy adecuada para un renderizador que sintetiza vistas para los puntos de visión distintos a los capturados. Por ejemplo, un renderizador puede estar dispuesto a generar dinámicamente vistas que coincidan con una postura local actual del espectador. Por ejemplo, se puede determinar dinámicamente la postura del espectador y generar dinámicamente vistas que coincidan con esta postura del espectador basándose en las imágenes y, por ejemplo, en los mapas de profundidad proporcionados.

20 Sin embargo, estas representaciones en imágenes tienden a dar lugar a una velocidad de datos muy elevada para una calidad de imagen determinada. A fin de proporcionar una buena captura de la escena y, en concreto, para hacer frente a los fenómenos de oclusión, se desea que la escena sea capturada desde posiciones de captura cercanas y que cubran una amplia escala de posiciones. En consecuencia, se desea un número relativamente alto de imágenes. Además, las vistas de captura de las cámaras suelen solaparse y, por tanto, el conjunto de imágenes tiende a incluir una gran cantidad de información redundante. Estos problemas

tienden a ser independientes de la configuración de captura específica y específicamente de si se usan configuraciones de captura lineales o, por ejemplo, circulares.

5 Así, mientras que muchas de las representaciones y de los formatos de la imagen convencionales pueden ofrecer un buen rendimiento en muchas aplicaciones y servicios, tienden a ser subóptimas al menos en algunas circunstancias.

10 Por lo tanto, sería ventajoso un enfoque mejorado para procesar y generar una señal de imagen que comprenda una representación en imagen de una escena. En particular, sería ventajoso un sistema y/o un enfoque que permitieran un funcionamiento mejorado, una mayor flexibilidad, una experiencia de realidad virtual mejorada, velocidades de datos reducidas, una mayor eficiencia, una distribución facilitada, una complejidad reducida, una implementación facilitada, unos requisitos de almacenamiento reducidos, una calidad de imagen mejorada, una representación mejorada, una experiencia de usuario mejorada, una compensación
15 mejorada entre la calidad de la imagen y la velocidad de datos, y/o un rendimiento y/o un funcionamiento mejorados.

Sumario de la invención

20 En consecuencia, la Invención busca preferentemente mitigar, aliviar o eliminar de manera individual o en cualquier combinación una o más de las desventajas antes mencionadas.

Según un aspecto de la invención, se proporciona un aparato para generar una señal de imagen, comprendiendo el aparato: un receptor para recibir una pluralidad de imágenes fuente
25 que representan una escena desde diferentes posturas de visión; un generador de imágenes combinadas para generar una pluralidad de imágenes combinadas a partir de las imágenes fuente, siendo derivada cada imagen combinada de un conjunto de al menos dos imágenes fuente de la pluralidad de imágenes fuente, representando cada píxel de una imagen combinada la escena para una postura de rayo e incluyendo las posturas de rayo para cada
30 imagen combinada al menos dos posiciones diferentes, una postura de rayo para un píxel representando una postura para un rayo en una dirección de visión para el píxel y desde una posición de visión para el píxel; un evaluador para determinar las medidas de calidad de la predicción para los elementos de la pluralidad de imágenes fuente, una medida de calidad de la predicción para un elemento de una primera imagen fuente es indicativa de una diferencia
35 entre los valores de los píxeles en la primera imagen fuente para los píxeles en el elemento y los valores de los píxeles predichos para los píxeles en el elemento, los valores de los píxeles

5 predichos son valores de los píxeles resultantes de la predicción de los píxeles en el elemento de la pluralidad de imágenes combinadas; un determinador para determinar segmentos de las imágenes fuente que comprenden elementos para los que la medida de calidad de la predicción es indicativa de una diferencia por encima de un umbral; y un generador de señales de imagen para generar una señal de imagen que comprende datos de imagen que representan las imágenes combinadas y datos de imagen que representan los segmentos de las imágenes fuente.

10 La invención puede proporcionar una representación mejorada de una escena y puede proporcionar en muchas realizaciones y escenarios una calidad de imagen mejorada de las imágenes renderizadas frente a la velocidad de datos de la señal de imagen. En muchas realizaciones, se puede proporcionar una representación más eficiente de una escena, por ejemplo, permitiendo que se logre una calidad determinada con una velocidad de datos reducida. El enfoque puede proporcionar un enfoque más flexible y eficiente para la
15 representación de imágenes de una escena y puede permitir una mejor adaptación, por ejemplo, a las propiedades de la escena.

El enfoque puede, en muchas realizaciones, emplear una representación en imagen de una escena adecuada para una aplicación flexible, eficiente y de alto rendimiento de Realidad
20 Virtual (RV). En muchas realizaciones, puede permitir o posibilitar una aplicación de RV con una compensación sustancialmente mejorada entre la calidad de la imagen y la velocidad de datos. En muchas realizaciones, puede permitir una calidad mejorada de la imagen percibida y/o una velocidad de datos reducida.

25 El enfoque puede ser adecuado, por ejemplo, para los servicios de difusión de vídeo que admiten la adaptación al movimiento y la rotación de la cabeza en el extremo receptor.

Las imágenes fuente pueden ser específicamente imágenes de intensidad de luz con información de profundidad asociada, tales como mapas de profundidad.

30 El enfoque puede permitir, en particular, que las imágenes combinadas se optimicen para la información de primer plano y de fondo, respectivamente, con los segmentos que proporcionan datos adicionales cuando resulten apropiados de manera específica.

35 El generador de señales de imagen puede estar dispuesto para que use una codificación de las imágenes combinadas más eficiente que la de los segmentos. Sin embargo, los segmentos

pueden constituir normalmente una proporción relativamente pequeña de los datos de las imágenes combinadas.

5 Según una característica opcional de la invención, el generador de imágenes combinadas está dispuesto para generar al menos una primera imagen combinada de la pluralidad de imágenes combinadas mediante la síntesis de visión de los píxeles de la primera imagen combinada de la pluralidad de imágenes fuente, donde cada píxel de la primera imagen combinada representa la escena para una postura de rayo y las posturas de rayo para la primera imagen comprende al menos dos posiciones diferentes.

10

Esto puede proporcionar un funcionamiento particularmente ventajoso en muchas realizaciones, y puede, por ejemplo, permitir que las imágenes combinadas se generen para las posturas de visión donde pueden (normalmente en combinación) proporcionar una representación particularmente ventajosa de la escena.

15

Según una característica opcional de la invención, un producto de puntos entre un vector vertical y los vectores de producto cruzado de píxeles es no negativo para al menos el 90 % de los píxeles de la primera imagen combinada, siendo un vector de producto cruzado de píxeles para un píxel un producto cruzado entre una dirección de rayo para un píxel y un vector desde un punto central para las diferentes posturas de visión a una posición de rayo para el píxel.

20

Esto puede proporcionar en muchas realizaciones una generación particularmente eficiente y ventajosa de imágenes combinadas. En particular, puede proporcionar un enfoque de baja complejidad para determinar una imagen combinada que proporcione una representación ventajosa de los datos de fondo al tender a proporcionar una vista sesgada hacia una vista lateral.

25

Según una característica opcional de la invención, el generador de imágenes combinadas está dispuesto para generar una segunda imagen combinada de la pluralidad de imágenes combinadas mediante la síntesis de visión de los píxeles de la segunda imagen combinada a partir de la pluralidad de imágenes fuente, donde cada píxel de la segunda imagen combinada representa la escena para una postura de rayo y las posturas de rayo para la segunda imagen comprenden al menos dos posiciones diferentes; y donde un producto de puntos entre los vectores vertical y de producto cruzado de píxeles es no positivo para al menos el 90 % de los píxeles de la segunda imagen combinada.

35

5 Esto puede proporcionar en muchas realizaciones una generación particularmente eficiente y ventajosa de imágenes combinadas. En particular, puede proporcionar un enfoque de baja complejidad para determinar una imagen combinada que proporcione una representación ventajosa de los datos de fondo al tender a proporcionar una vista sesgada hacia diferentes vistas laterales.

10 Según una característica opcional de la invención, las posturas de rayo de la primera imagen combinada se seleccionan para que estén próximas a un borde de una región que comprende las diferentes posturas de visión de la pluralidad de imágenes fuente.

15 Esto puede proporcionar un funcionamiento ventajoso en muchas realizaciones y puede, por ejemplo, proporcionar una información de fondo mejorada por la señal de imagen, facilitando así y/o mejorando la síntesis de la visión basada en la señal de imagen.

20 Según una característica opcional de la invención, se determina que cada una de las posturas de rayo de la primera imagen combinada está a menos de una primera distancia de un borde de una región que comprende las diferentes posturas de visión de la pluralidad de imágenes fuente, siendo la primera distancia no superior al 50 % de una distancia interior máxima entre puntos del borde.

25 Esto puede proporcionar un funcionamiento ventajoso en muchas realizaciones y puede, por ejemplo, proporcionar una información de fondo mejorada por la señal de imagen, facilitando así y/o mejorando la síntesis de la visión basada en la señal de imagen. En algunas realizaciones, la primera distancia no es superior al 25 % o al 10 % de la distancia interior máxima.

30 En algunas realizaciones, se determina que al menos una postura de visión de las imágenes combinadas está a menos de una primera distancia de un borde de una región que comprende las diferentes posturas de visión de la pluralidad de imágenes fuente, siendo la primera distancia no superior al 20 %, al 10 % o incluso al 5 % de una distancia máxima entre dos posturas de visión de las diferentes posturas de visión.

35 En algunas realizaciones, se determina que al menos una postura de visión de las imágenes combinadas esté al menos a una distancia mínima de un punto central de las diferentes posturas de visión, siendo la distancia mínima al menos el 50 %, el 75 % o incluso el 90 % de

una distancia desde el punto central hasta un borde de una región que comprende las diferentes posturas de visión de la pluralidad de imágenes fuente a lo largo de una línea que pasa por el punto central y la al menos una postura de visión.

5 Según una característica opcional de la invención, para cada píxel de una primera imagen combinada de la pluralidad de imágenes combinadas, el generador de imágenes combinadas está dispuesto para determinar un píxel correspondiente en cada una de las imágenes fuente de visión para las que está presente un píxel correspondiente, siendo el píxel correspondiente uno que representa una misma dirección de rayo que el píxel de la primera imagen combinada;
10 seleccionar un valor de píxel para el píxel de la primera imagen combinada como un valor de píxel del píxel correspondiente en la imagen fuente de visión para la que el píxel correspondiente representa un rayo que tiene una mayor distancia desde un punto central para las diferentes posturas de visión, siendo la mayor distancia en una primera dirección a lo largo de un primer eje perpendicular a una dirección de rayo para el píxel correspondiente.

15 Esto puede proporcionar en muchas realizaciones una generación particularmente eficiente y ventajosa de imágenes combinadas. En particular, puede proporcionar un enfoque de baja complejidad para determinar una imagen combinada que proporcione una representación ventajosa de los datos de fondo al tender a proporcionar una vista sesgada hacia una vista lateral.
20

Según una característica opcional de la invención, los píxeles correspondientes comprenden el remuestreo de cada imagen fuente para dar una representación en imagen que representa al menos una parte de una superficie de una esfera de visión que rodea las posturas de visión y la determinación de los píxeles correspondientes como píxeles que tienen una misma posición en la representación en imagen.
25

Esto puede proporcionar una determinación particularmente eficiente y precisa de los píxeles correspondientes.
30

La superficie de la esfera de visión se puede representar, por ejemplo, mediante un mapa equirectangular o cúbico. Cada píxel de la esfera de visión puede tener una dirección de rayo y el remuestreo de una imagen fuente puede incluir el establecimiento de un valor de píxel de la esfera de visión al valor de píxel de la imagen fuente para el que la dirección del rayo es la misma.
35

De acuerdo con una característica opcional de la invención, para cada píxel de una segunda imagen combinada el generador de imágenes combinadas está dispuesto para que seleccione un valor de píxel para el píxel en la segunda imagen combinada como un valor de píxel del píxel correspondiente en la imagen fuente de la vision para la cual el píxel correspondiente
5 representa un rayo que tiene una distancia mayor desde el punto central en una dirección opuesta a la primera dirección.

Esto puede proporcionar en muchas realizaciones una generación particularmente eficiente y ventajosa de imágenes combinadas. En particular, puede proporcionar un enfoque de baja
10 complejidad para determinar una imagen combinada que proporcione una representación ventajosa de los datos de fondo al tender a proporcionar una vista sesgada hacia una vista lateral. Además, la segunda imagen combinada puede complementar la primera imagen combinada proporcionando una vista lateral desde una dirección opuesta, combinándose así con la primera imagen combinada para proporcionar una representación particularmente
15 ventajosa de la escena y específicamente de la información de fondo.

Según una característica opcional de la invención, para cada píxel de una tercera imagen combinada el generador de imágenes combinadas está dispuesto para seleccionar un valor de píxel para el píxel en la tercera imagen combinada como un valor de píxel del píxel correspondiente en la imagen fuente de la vision para la cual el píxel correspondiente
20 representa un rayo que tiene una distancia más pequeña desde el punto central.

Esto puede proporcionar en muchas realizaciones una generación particularmente eficiente y ventajosa de imágenes combinadas. La tercera imagen combinada puede complementar la(s)
25 primera(s) imagen(es) combinada(s) proporcionando una vista más frontal de la escena que puede proporcionar una representación mejorada de los objetos de primer plano en la escena.

Según una característica opcional de la invención, para cada píxel en una cuarta imagen combinada el generador de imágenes combinadas está dispuesto para seleccionar un valor de píxel para el píxel en la cuarta imagen combinada como un valor de píxel del píxel correspondiente en la imagen fuente de la vision para la que el píxel correspondiente
30 representa un rayo que tiene una distancia mayor desde el punto central en una segunda dirección a lo largo de un segundo eje perpendicular a una dirección de rayo para el píxel correspondiente, teniendo el primer eje y el segundo eje direcciones diferentes.

35

Esto puede proporcionar en muchas realizaciones una generación particularmente eficiente y ventajosa de imágenes combinadas, y puede proporcionar una representación mejorada de la escena.

5 Según una característica opcional de la invención, el generador de imágenes combinadas está dispuesto para generar datos de origen para la primera imagen combinada, siendo los datos de origen indicativos de cuál de las imágenes fuente es un origen para cada píxel de la primera imagen combinada; y el generador de señales de imagen está dispuesto para incluir los datos de origen en la señal de imagen.

10

Esto puede proporcionar en muchas realizaciones un funcionamiento particularmente ventajoso.

15 Según una característica opcional de la invención, el generador de señales de imagen está dispuesto para incluir datos de postura de la visión fuente en la señal de imagen, siendo los datos de postura de la visión fuente indicativos de las diferentes posturas de la visión para las imágenes fuente.

20 Esto puede proporcionar en muchas realizaciones un funcionamiento particularmente ventajoso.

Según un aspecto de la invención, se proporciona un aparato para recibir una señal de imagen, comprendiendo el aparato: un receptor para recibir una señal de imagen, comprendiendo la señal de imagen una pluralidad de imágenes combinadas, representando cada imagen combinada datos de imagen derivados de un conjunto de al menos dos imágenes fuente de una pluralidad de imágenes fuente que representan una escena desde diferentes posturas de visión, representando cada píxel de una imagen combinada la escena para una postura de rayo e incluyendo las posturas de rayo para cada imagen combinada al menos dos posiciones diferentes, una postura de rayo para un píxel que representa una postura para un rayo en una dirección de visión para el píxel y desde una posición de visión para el píxel; datos de imagen para un conjunto de segmentos de la pluralidad de imágenes fuente, un segmento para una primera imagen fuente que comprende al menos un píxel de la primera imagen fuente para el que una medida de calidad de predicción para una predicción del segmento de la pluralidad de imágenes combinadas está por debajo de un umbral; y un procesador para procesar la señal de imagen.

25

30

35

Según un aspecto de la invención, se proporciona un procedimiento para generar una señal de imagen, comprendiendo el procedimiento: recibir una pluralidad de imágenes fuente que representan una escena desde diferentes posturas de visión; generar una pluralidad de imágenes combinadas a partir de las imágenes fuente, estando cada imagen combinada derivada de un conjunto de al menos dos imágenes fuente de la pluralidad de imágenes fuente, representando cada píxel de una imagen combinada la escena para una postura de rayo e incluyendo las posturas de rayo para cada imagen combinada al menos dos posiciones diferentes, representando una postura de rayo para un píxel una postura para un rayo en una dirección de visión para el píxel y desde una posición de visión para el píxel; determinar las medidas de calidad de la predicción para los elementos de la pluralidad de imágenes fuente, una medida de calidad de la predicción para un elemento de una primera imagen fuente que indica una diferencia entre los valores de los píxeles en la primera imagen fuente para los píxeles en el elemento y los valores de los píxeles predichos para los píxeles en el elemento, los valores de los píxeles predichos son valores de los píxeles resultantes de la predicción de los píxeles en el elemento de la pluralidad de imágenes combinadas; determinar los segmentos de las imágenes fuente que comprenden los elementos para los que la medida de calidad de la predicción es indicativa de una diferencia por encima de un umbral; y generar una señal de imagen que comprende los datos de imagen que representan las imágenes combinadas y los datos de imagen que representan los segmentos de las imágenes fuente.

Según un aspecto de la invención, se proporciona un procedimiento de procesamiento de una señal de imagen, comprendiendo el procedimiento: recibir una señal de imagen, comprendiendo la señal de imagen: una pluralidad de imágenes combinadas, representando cada imagen combinada datos de imagen derivados de un conjunto de al menos dos imágenes fuente de una pluralidad de imágenes fuente que representan una escena desde diferentes posturas de visión, representando cada píxel de una imagen combinada la escena para una postura de rayo e incluyendo las posturas de rayo para cada imagen combinada al menos dos posiciones diferentes, representando una postura de rayo para un píxel una postura para un rayo en una dirección de visión para el píxel y desde una posición de visión para el píxel; datos de imagen para un conjunto de segmentos de la pluralidad de imágenes fuente, un segmento para una primera imagen fuente que comprende al menos un píxel de la primera imagen fuente para el que una medida de calidad de predicción para una predicción del segmento de la pluralidad de imágenes combinadas está por debajo de un umbral; y procesamiento de la señal de imagen.

35

Según un aspecto de la invención, se proporciona una señal de imagen que comprende una pluralidad de imágenes combinadas, representando cada imagen combinada datos de imagen derivados de un conjunto de al menos dos imágenes fuente de una pluralidad de imágenes fuente que representan una escena desde diferentes posturas de visión, representando cada 5 píxel de una imagen combinada que que incluye al menos dos posiciones diferentes, representando una postura de rayo para un píxel una postura para un rayo en una dirección de visión para el píxel y desde una posición de visión para el píxel; datos de imagen para un conjunto de segmentos de la pluralidad de imágenes fuente, comprendiendo un segmento para una primera imagen fuente al menos un píxel de la primera imagen fuente para el que 10 una medida de calidad de predicción para una predicción del segmento de la pluralidad de imágenes combinadas está por debajo de un umbral.

Estos y otros aspectos, características y ventajas de la invención serán evidentes y se explicarán con referencia a la(s) realización(es) descrita(s) a continuación.

15

Breve descripción de los dibujos

Las realizaciones de la invención se describirán, solo a modo de ejemplo, con referencia a los dibujos, en los que

20

La Fig. 1 ilustra un ejemplo de una disposición para proporcionar una experiencia de realidad virtual;

La Fig. 2 ilustra un ejemplo de disposición de captura de una escena;

25

La Fig. 3 ilustra un ejemplo de disposición de captura de una escena;

La Fig. 4 ilustra un ejemplo de elementos de un aparato de acuerdo con algunas realizaciones de la invención;

30

La Fig. 5 ilustra un ejemplo de elementos de un aparato de acuerdo con algunas realizaciones de la invención;

La Fig. 6 ilustra un ejemplo de selección de píxeles de acuerdo con algunas 35 realizaciones de la invención; y

La Fig. 7 ilustra un ejemplo de selección de píxeles de acuerdo con algunas realizaciones de la invención.

La Fig. 8 ilustra un ejemplo de elementos de una disposición de postura de rayo para una imagen combinada generada de acuerdo con algunas realizaciones de la invención;

5

La Fig. 9 ilustra un ejemplo de elementos de una disposición de rayo para una imagen combinada generada de acuerdo con algunas realizaciones de la invención;

La Fig. 10 ilustra un ejemplo de elementos de una disposición de postura de rayo para una imagen combinada generada de acuerdo con algunas realizaciones de la invención;

10

La Fig. 11 ilustra un ejemplo de elementos de una disposición de postura de rayo para una imagen combinada generada de acuerdo con algunas realizaciones de la invención;

15

La Fig. 12 ilustra un ejemplo de elementos de una disposición de postura de rayo para una imagen combinada generada de acuerdo con algunas realizaciones de la invención; y

20

La Fig. 13 ilustra un ejemplo de elementos de una disposición de postura de rayo para una imagen combinada generada de acuerdo con algunas realizaciones de la invención.

25 **Descripción detallada de algunas realizaciones de la invención**

Las experiencias virtuales que permiten al usuario moverse en un mundo virtual son cada vez más populares y se están desarrollando servicios para satisfacer dicha demanda. Sin embargo, la prestación de servicios de realidad virtual eficientes es muy difícil, sobre todo si la experiencia debe basarse en una captura de un entorno del mundo real y no en un mundo artificial totalmente generado virtualmente.

30

En muchas aplicaciones de realidad virtual, se determina una entrada de postura del espectador que refleja la postura de un espectador virtual en la escena. A continuación, el aparato/sistema/aplicación de realidad virtual genera una o varias imágenes correspondientes

35

a las vistas y los puntos de visión de la escena para un espectador que corresponde a la postura del espectador.

5 Normalmente, la aplicación de realidad virtual genera una salida tridimensional en forma de imágenes separadas para los ojos izquierdo y derecho. Se pueden presentar al usuario por medios adecuados, tales como las pantallas individuales de los ojos izquierdo y derecho de un casco de RV. En otras realizaciones, la imagen puede presentarse, por ejemplo, en una pantalla autoestereoscópica (en cuyo caso se puede generar un mayor número de imágenes de visión para la postura del espectador), o incluso en algunas realizaciones puede generarse 10 una sola imagen bidimensional (por ejemplo, usando una pantalla bidimensional convencional).

La entrada de la postura del espectador puede determinarse de diferentes maneras en las distintas aplicaciones. En muchas realizaciones, el movimiento físico de un usuario puede ser 15 rastreado directamente. Por ejemplo, una cámara que inspeccione el área del usuario puede detectar y seguir la cabeza del usuario (o incluso los ojos). En muchas realizaciones, el usuario puede llevar un casco de RV que puede ser rastreado por medios externos y/o internos. Por ejemplo, los auriculares pueden incluir acelerómetros y giroscopios que proporcionan información sobre el movimiento y la rotación de los auriculares y, por tanto, de la cabeza. En 20 algunos ejemplos, el casco de RV puede transmitir señales o incluir identificadores (por ejemplo, visuales) que permitan a un sensor externo determinar el movimiento del casco de RV.

En algunos sistemas, la postura del espectador se puede ser proporcionada por medios 25 manuales, por ejemplo, controlando el usuario manualmente un joystick o una entrada manual similar. Por ejemplo, el usuario puede mover manualmente al espectador virtual en la escena controlando un primer joystick analógico con una mano y controlando manualmente con la otra mano la dirección hacia la que mira el espectador virtual moviendo manualmente un segundo joystick analógico.

30 En algunas aplicaciones se puede usar una combinación de enfoques manuales y automatizados para generar la postura del espectador de entrada. Por ejemplo, unos auriculares pueden rastrear la orientación de la cabeza y el movimiento/posición del espectador en la escena puede ser controlado por el usuario mediante un joystick.

35

La generación de imágenes se basa en una representación adecuada del mundo/entorno/escena virtual. En algunas aplicaciones, se puede proporcionar un modelo tridimensional completo para la escena y evaluando este modelo se pueden determinar las vistas de la escena desde una postura específica del espectador .

5

En muchos sistemas prácticos, se puede representar la escena mediante una representación en imagen que comprende datos de imagen. Los datos de la imagen pueden comprender normalmente imágenes asociadas a una o más posturas de captura o de anclaje, y, de manera específica, se pueden incluir las imágenes para uno o más puertos de visión correspondiendo cada uno de los puertos de visión a una postura específica. Puede usarse una representación de la imagen que comprenda una o más imágenes en las que cada imagen represente la vista de un puerto de visión determinado para una postura de visión determinada. Dichas posturas o posiciones de visión para las que se proporcionan datos de imagen se denominan a menudo posturas o posiciones de anclaje o posturas o posiciones de captura (ya que los datos de imagen pueden corresponder normalmente a imágenes que son o serían capturadas por cámaras situadas en la escena con la posición y la orientación correspondientes a la postura de captura).

10

15

Sobre la base de dicha representación en imagen, muchas aplicaciones típicas de RV pueden proporcionar imágenes de visión correspondientes a los puertos de visión de la escena para la postura actual del espectador, actualizándose dinámicamente las imágenes para reflejar los cambios en la postura del espectador y generándose las imágenes sobre la base de los datos de imagen que representan la escena/entorno/mundo (posiblemente) virtual. La aplicación puede hacer esto mediante la realización de algoritmos de síntesis de visión y de cambio de visión, como es conocido para el experto.

20

25

En este campo, los términos colocación y postura se usan como un término común para la posición y/o la dirección/orientación. La combinación de la posición y la dirección/orientación de, por ejemplo, un objeto, una cámara, una cabeza o una vista pueden denominarse postura o colocación. Por lo tanto, una indicación de colocación o de postura puede comprender seis valores/componentes/grados de libertad, y cada valor/componente suele describir una propiedad individual de la posición/ubicación o de la orientación/dirección del objeto correspondiente. Por supuesto, en muchas situaciones, una colocación o una postura puede considerarse o representarse con menos componentes, por ejemplo si uno o más componentes se consideran fijos o irrelevantes (por ejemplo, si se considera que todos los objetos están a la misma altura y tienen una orientación horizontal, cuatro componentes

30

35

pueden proporcionar una representación completa de la postura de un objeto). En lo sucesivo, el término postura se usa para referirse a una posición y/o una orientación que se pueden representar con uno a seis valores (correspondientes al máximo de grados de libertad posibles).

5

Muchas aplicaciones de RV se basan en una postura que tiene el máximo de grados de libertad, es decir, tres grados de libertad de cada una de las posiciones y la orientación, lo que da como resultado un total de seis grados de libertad. Por lo tanto, una postura puede ser representada por un conjunto o un vector de seis valores que representan los seis grados de libertad y, por lo tanto, un vector de postura puede proporcionar una posición tridimensional y/o una indicación de dirección tridimensional. Sin embargo, se apreciará que en otras realizaciones, la postura puede ser representada por menos valores.

10

Una postura puede ser al menos una de una orientación y una posición. Un valor de postura puede ser indicativo de al menos uno de los valores de orientación y de posición.

15

Un sistema o una entidad basados en proporcionar el máximo grado de libertad al espectador suele denominarse de 6 grados de libertad (6 GL). Muchos sistemas y entidades solo proporcionan una orientación o posición, y estos se conocen normalmente como teniendo 3 Grados de Libertad (3 GL).

20

En algunos sistemas, se puede proporcionar localmente la aplicación de RV a un espectador, por ejemplo, por medio de un dispositivo autónomo que no usa, ni siquiera tiene acceso, a ningún dato o procesamiento de RV remoto. Por ejemplo, un dispositivo tal como una consola de juegos puede comprender un almacén para guardar los datos de la escena, una entrada para recibir/generar la postura del espectador y un procesador para generar las imágenes correspondientes a partir de los datos de la escena.

25

En otros sistemas, la aplicación de RV puede implementarse y ejecutarse a distancia del espectador. Por ejemplo, un dispositivo local del usuario puede detectar/recibir datos de movimiento/postura que se transmiten a un dispositivo remoto que procesa los datos para generar la postura del espectador. El dispositivo remoto puede entonces generar imágenes de visión adecuadas para la postura del espectador basándose en los datos de la escena que la describen. Las imágenes de la vision se transmiten entonces al dispositivo local del espectador donde se presentan. Por ejemplo, el dispositivo remoto puede generar directamente un flujo de vídeo (normalmente un flujo de vídeo estéreo/3D) que es presentado

35

directamente por el dispositivo local. Así, en este ejemplo, el dispositivo local puede no realizar ningún procesamiento de RV, excepto la transmisión de datos de movimiento y la presentación de los datos de vídeo recibidos.

5 En muchos sistemas, la funcionalidad puede estar distribuida entre un dispositivo local y un dispositivo remoto. Por ejemplo, el dispositivo local puede procesar los datos de entrada y de los sensores recibidos para generar las posturas del espectador que se transmiten continuamente al dispositivo de RV remoto. El dispositivo de RV remoto puede entonces generar las correspondientes imágenes de la visión y transmitir las al dispositivo local para su
10 presentación. En otros sistemas, el dispositivo de RV remoto puede no generar directamente las imágenes de la visión, pero puede seleccionar los datos relevantes de la escena y transmitirlos al dispositivo local, que puede entonces generar las imágenes de la visión que se presentan. Por ejemplo, el dispositivo de RV remoto puede identificar el punto de captura más cercano y extraer los datos de la escena correspondientes (por ejemplo, imagen esférica
15 y datos de profundidad del punto de captura) y transmitirlos al dispositivo local. El dispositivo local puede entonces procesar los datos de la escena recibida para generar las imágenes para la postura de visión específica y actual. La postura de la visión corresponderá normalmente a la postura de la cabeza, y las referencias a la postura de la visión pueden considerarse equivalentes a las referencias a la postura de la cabeza.

20 En muchas aplicaciones, especialmente para los servicios de difusión, una fuente puede transmitir datos de la escena en forma de una representación en imagen (incluido el vídeo) de la escena que es independiente de la postura del espectador. Por ejemplo, una representación en imagen para una única esfera de visión para una única posición de captura puede ser
25 transmitida a una pluralidad de clientes. Los clientes individuales pueden entonces sintetizar localmente las imágenes de la visión correspondientes a la postura actual del espectador.

Una aplicación que despierta especial interés es aquella en la que se admite una cantidad limitada de movimiento, de manera que las vistas presentadas se actualizan para seguir
30 pequeños movimientos y rotaciones correspondientes a un espectador sustancialmente estático que solo realiza pequeños movimientos y rotaciones de la cabeza. Por ejemplo, un espectador sentado puede girar su cabeza y moverla ligeramente y las vistas/imágenes presentadas se adaptan para seguir estos cambios de postura. Este enfoque puede proporcionar una experiencia de vídeo altamente inmersiva. Por ejemplo, un espectador que
35 vea un evento deportivo puede sentir que está presente en un punto concreto del estadio.

Las aplicaciones de libertad limitada de este tipo tienen la ventaja de proporcionar una experiencia mejorada al tiempo que no requieren una representación precisa de una escena desde muchas posiciones diferentes, lo que reduce sustancialmente los requisitos de captura. Del mismo modo, la cantidad de datos que hay que proporcionar a un renderizador puede reducirse sustancialmente. De hecho, en muchos escenarios, solo es necesario proporcionar datos de imagen y, normalmente, de profundidad para un único punto de visión, y a partir de ellos el renderizador local puede generar las vistas deseadas.

El enfoque puede ser específicamente muy adecuado para aplicaciones en las que los datos necesitan ser comunicados desde una fuente a un destino a través de un canal de comunicación de banda limitada, como por ejemplo para una aplicación de difusión o de cliente-servidor.

La Fig. 1 ilustra un ejemplo de sistema de RV en el que un dispositivo cliente de RV 101 remoto se comunica con un servidor de RV 103, por ejemplo, a través de una red 105, tal como Internet. El servidor 103 puede estar preparado para soportar simultáneamente un número potencialmente grande de dispositivos cliente 101.

El servidor de RV 103 puede, por ejemplo, apoyar una experiencia de transmisión transmitiendo una señal de imagen que comprende una representación en imagen en forma de datos de imagen que pueden ser usados por los dispositivos cliente para sintetizar localmente imágenes de visión correspondientes a las posturas apropiadas.

En muchas aplicaciones, como la de la Fig.1, puede ser deseable capturar una escena y generar una representación en imagen eficiente que pueda incluirse de forma eficiente en una señal de imagen. La señal de imagen puede entonces ser transmitida a varios dispositivos que pueden sintetizar localmente vistas para otras posturas de visión diferentes a las de la captura. Para ello, la representación de la imagen puede incluir normalmente información de profundidad y, por ejemplo, se pueden proporcionar imágenes con profundidad asociada. Por ejemplo, pueden obtenerse mapas de profundidad usando la captura estereoscópica en combinación con la estimación de la disparidad o usando sensores de alcance, y estos mapas de profundidad pueden proporcionarse con las imágenes de intensidad luminosa.

Sin embargo, un problema particular de estos enfoques es que el cambio de la postura de la visión puede cambiar las características de oclusión, lo que hace que los segmentos del fondo

que no son visibles en una imagen capturada determinada se vuelvan visibles para una postura de vision diferente.

Para ello, se suele usar un número relativamente grande de cámaras para capturar una
 5 escena. La Fig. 2 muestra un ejemplo de captura mediante un equipo de cámara circular de 8 vistas. En el ejemplo, las cámaras están orientadas hacia el exterior. Como puede verse, diferentes cámaras, y por tanto diferentes imágenes de captura/fuente, pueden tener visibilidad de diferentes partes de la escena. Por ejemplo, la región de fondo 1 solo es visible desde la cámara 2. Sin embargo, como también puede observarse, gran parte de la escena
 10 es visible desde varias cámaras, por lo que se crea una cantidad significativa de información redundante.

La Fig. 3 muestra un ejemplo de un conjunto lineal de cámaras. De nuevo, las cámaras proporcionan información de diferentes partes de la escena, por ejemplo, c1 es la única
 15 cámara que captura la región 2, c3 es la única cámara que captura la región 4, y c4 es la única cámara que captura la región 3. Al mismo tiempo, algunas partes de la escena son capturadas por más de una de las cámaras. Por ejemplo, todas las cámaras capturan la parte delantera de los objetos en primer plano fg1 y fg2, pero algunas cámaras proporcionan una mejor captura que otras. La Fig. 3 muestra un ejemplo A para cuatro cámaras y un ejemplo B para
 20 dos cámaras. Como se puede ver, la configuración de cuatro cámaras proporciona una mejor captura, que incluye la captura de parte de la escena (región 4 del fondo bg) pero, por supuesto, también genera una mayor cantidad de datos, incluyendo más datos redundantes.

Una desventaja de la captura de varias vistas con respecto a una sola vista central es,
 25 obviamente, la mayor cantidad de datos de imagen. Otra desventaja es la gran cantidad de píxeles generados, es decir, la tasa de píxeles que hay que procesar y que el decodificador debe producir. Esto también requiere una mayor complejidad y uso de recursos para la síntesis de visión durante la reproducción.

30 A continuación, se describirá un enfoque específico que usa una representación en imagen más eficiente y menos redundante de las vistas capturadas. Pretende preservar cierta coherencia espacial y temporal de los datos de la imagen, permitiendo que los codificadores de vídeo sean más eficientes. Reduce la tasa de bits, la tasa de píxeles y la complejidad de la síntesis de visiones en el lugar de reproducción.

35

Esta representación comprende una pluralidad de imágenes combinadas, cada una de las cuales se genera a partir de dos o más de las imágenes fuente (que específicamente pueden ser imágenes 3D capturadas, por ejemplo, representadas como imagen más mapa de profundidad) y normalmente solo se considera una parte de cada una de las imágenes fuente.

5 Las imágenes combinadas pueden proporcionar una referencia para la síntesis de la visión y proporcionar información sustancial de la escena. Las imágenes combinadas pueden ser generadas para estar sesgadas hacia visiones más externas de la escena, y específicamente hacia los bordes de la región de captura. En algunas realizaciones, también se pueden proporcionar una o más imágenes combinadas centrales.

10

En muchas realizaciones, cada una de las imágenes combinadas representa vistas desde diferentes posiciones de visión, es decir, cada imagen puede comprender al menos píxeles que corresponden a diferentes posturas de visión/captura/anclaje. Específicamente, cada píxel de una imagen combinada puede representar una postura de rayo correspondiente a un origen/posición y una dirección/orientación para un rayo desde ese origen/posición dirigido en esa dirección/orientación y que termina en el punto de la escena/objeto que está representado por el valor del píxel para ese píxel. Al menos dos píxeles de una imagen combinada pueden tener diferentes orígenes/posiciones del rayo. Por ejemplo, en algunas realizaciones, los píxeles de una imagen combinada pueden estar divididos en N grupos en los que todos los píxeles de un grupo tienen el mismo origen/posición del rayo, pero éste es diferente para los grupos individuales. N puede ser dos o más. En algunas realizaciones, N puede ser igual al número máximo de píxeles horizontales en una fila (y/o al número de columnas en la imagen combinada), y de hecho en algunas realizaciones, N puede ser igual al número de píxeles, es decir, todos los píxeles pueden tener un origen/posición de rayo único.

25

Una postura de rayo para un píxel puede representar un origen/posición, y/o una orientación/dirección para un rayo entre el origen/posición y el punto de la escena representado por el píxel. El origen/posición puede ser específicamente una posición de visión para el píxel y la orientación/dirección puede ser la dirección de visión para el píxel. Puede representar efectivamente el rayo de luz que sería capturado en la posición del rayo desde la dirección del rayo para el píxel, y por lo tanto refleja el rayo de la luz que está representado por el valor del píxel.

30

Así, cada píxel puede representar la escena vista desde una posición de visión en una dirección de visión. La posición de la visión y la dirección de la visión definen en consecuencia un rayo. Cada píxel puede tener un rayo de visión asociado desde la posición de la visión para

35

el píxel y en la dirección de la vision para el píxel. Cada píxel representa la escena para una postura de rayo (de visión) que es la postura de un rayo desde un punto de visión/posición para el píxel y en una dirección de visión. El píxel puede representar específicamente el punto de la escena (punto en la escena) en el que el rayo de visión se cruza con un objeto de la escena (incluyendo el fondo). Un píxel puede representar rayos de la luz desde un punto de la escena hasta la posición de la vision y en la dirección de la vision. El rayo de la vision puede ser un rayo desde la posición de la vision en la dirección que interseca el punto de la escena.

Además, las imágenes combinadas se complementan con segmentos o fragmentos de las vistas capturadas que han sido identificados como no previstos lo suficientemente bien por las imágenes combinadas. Por lo tanto, se definen e incluyen un número, normalmente elevado, de segmentos, normalmente pequeños, para representar específicamente partes individuales de las imágenes capturadas que pueden proporcionar información sobre elementos de la escena no suficientemente bien representados por las imágenes combinadas.

Una ventaja de esta representación es que se pueden proporcionar diferentes codificaciones a las distintas partes de los datos de la imagen que se van a transmitir. Por ejemplo, se puede aplicar una codificación y una compresión eficientes y complejas a las imágenes combinadas, ya que éstas tienden a constituir la mayor parte de la señal de imagen, mientras que a los segmentos se les puede aplicar una codificación menos eficiente. Además, las imágenes combinadas pueden generarse de forma que sean muy adecuadas para una codificación eficiente, por ejemplo, generándolas de forma que sean similares a las imágenes convencionales, lo que permite usar enfoques de codificación de imágenes eficientes. Por el contrario, las propiedades de los segmentos pueden variar mucho más en función de las características específicas de las imágenes y, por lo tanto, pueden resultar más difíciles de codificar de forma tan eficiente. Sin embargo, esto no es un problema, ya que los segmentos tienden a proporcionar muchos menos datos de imagen.

La Fig. 4 ilustra un ejemplo de un aparato para generar una señal de imagen que incluye una representación de una pluralidad de imágenes fuente de la escena a partir de diferentes posturas de visión de origen (posturas de anclaje) tal como se ha descrito anteriormente. El aparato también se denominará transmisor de señales de imagen 400. El transmisor de señales de imagen 400 puede estar comprendido, por ejemplo, en el servidor de RV 103 de la Fig. 1.

35

La Fig. 5 ilustra un ejemplo de un aparato para la representación de imágenes de visión basado en una señal de imagen recibida que incluye una representación de una pluralidad de imágenes de la escena. El aparato puede recibir específicamente la señal de datos de imagen generada por el aparato de la Fig. 4 y proceder a procesarla para representar imágenes para posturas de visión específicas. El aparato de la Fig. 5 también se denominará receptor de señales de imagen 500. El receptor de señales de imagen 500 puede, por ejemplo, estar incluido en el dispositivo cliente 101 de la Fig. 1.

El transmisor de señales de imagen 400 comprende un receptor de fuente de imagen 401 que está dispuesto para recibir una pluralidad de imágenes fuente de la escena. Las imágenes fuente pueden representar vistas de la escena desde diferentes posiciones de visión. Las imágenes fuente pueden ser normalmente imágenes capturadas, por ejemplo, capturadas por las cámaras de un equipo de cámaras. Las imágenes fuente pueden comprender, por ejemplo, imágenes de una fila de cámaras de captura equidistantes o de un anillo de cámaras.

En muchas realizaciones, las imágenes fuente pueden ser imágenes 3D que comprenden imágenes 2D con información de profundidad asociada. Las imágenes 2D pueden ser específicamente imágenes de visión de la escena desde la correspondiente postura de captura, y la imagen 2D puede ir acompañada de una imagen o de un mapa de profundidad que comprende valores de profundidad para cada uno de los píxeles de la imagen 2D. La imagen 2D puede ser un mapa de textura. La imagen 2D puede ser una imagen de intensidad de luz.

Los valores de profundidad pueden ser, por ejemplo, valores de disparidad o valores de distancia, por ejemplo, indicados por una coordenada z. En algunas realizaciones, una imagen fuente puede ser una imagen 3D en forma de mapa de textura con una malla 3D asociada. En algunas realizaciones, los mapas de textura y las representaciones de malla de este tipo pueden ser convertidos en representaciones de imagen más profundidad por el receptor de la imagen fuente antes de su posterior procesamiento por el transmisor de la señal de imagen 400.

El receptor de imágenes fuente 401 recibe, por consiguiente, una pluralidad de imágenes fuente que caracterizan y representan la escena desde diferentes posturas de visión de origen. Este conjunto de imágenes fuente permitirá generar imágenes de visión para otras posturas usando algoritmos tal como el de desplazamiento de visión, como es conocido para el experto. En consecuencia, el transmisor de señales de imagen 400 está dispuesto para

generar una señal de imagen que comprende datos de imagen para las imágenes fuente y transmitir estos datos a un dispositivo remoto para su representación local. Sin embargo, la transmisión directa de todas las imágenes fuente requerirá una velocidad de datos inviable y comprenderá una gran cantidad de información redundante. El transmisor de señales de imagen 400 está preparado para reducir la velocidad de datos usando una representación en imagen como la descrita anteriormente.

En concreto, el receptor de la fuente de entrada 401 está acoplado a un generador de imágenes combinadas 403 que está dispuesto para generar una pluralidad de imágenes combinadas. Las imágenes combinadas comprenden información derivada de una pluralidad de imágenes fuente. El enfoque exacto para derivar las imágenes combinadas puede ser distinto entre las diferentes realizaciones, y los ejemplos específicos se describirán con más detalle más adelante. En algunas realizaciones, una imagen combinada puede generarse mediante la selección de píxeles de diferentes imágenes fuente. En otras realizaciones, las imágenes combinadas pueden generar alternativa o adicionalmente una o más de las imágenes combinadas mediante la síntesis de visión de las imágenes fuente.

Sin embargo, mientras que cada imagen combinada incluye una contribución de al menos dos, y a menudo más, de las imágenes fuente, normalmente solo se considera una parte de las imágenes fuente individuales para cada imagen combinada. De este modo, para cada imagen fuente usada para generar una determinada imagen combinada, hay algunos píxeles que se excluyen/descartan. Por consiguiente, los valores de los píxeles generados para la imagen combinada específica no dependen de los valores de estos píxeles.

Las imágenes combinadas pueden generarse de manera que cada imagen no represente simplemente una vista/captura/posición de anclaje, sino que represente dos o más vistas/capturas/posiciones de anclaje. En concreto, el origen del rayo/la posición de al menos algunos píxeles en una imagen combinada será diferente y, por lo tanto, una imagen combinada puede representar una vista de la escena desde diferentes direcciones.

El generador de imágenes combinadas 403 puede, en consecuencia, estar dispuesto para generar una pluralidad de imágenes combinadas a partir de las imágenes fuente, donde cada imagen combinada se deriva de un conjunto de al menos dos imágenes fuente, y donde normalmente la derivación de una primera imagen combinada incluye solo una parte de cada una de estas al menos dos imágenes fuente. Además, cada píxel de una imagen combinada

dada representa la escena para una postura de rayo y las posturas de rayo para cada imagen combinada pueden comprender al menos dos posiciones diferentes.

5 El generador de imágenes combinadas 403 está acoplado a un evaluador 405 que recibe las imágenes combinadas y las imágenes fuente. El evaluador 405 está preparado para determinar las medidas de calidad de la predicción para los elementos de las imágenes fuente. Un elemento puede ser un píxel individual y el evaluador 405 puede estar dispuesto para determinar una medida de calidad de predicción para cada píxel de cada imagen fuente. En otras realizaciones, los elementos pueden comprender una pluralidad de píxeles y cada elemento puede ser un grupo de píxeles. Por ejemplo, una medida de la calidad de predicción de puede determinar para bloques, por ejemplo, de 4x4 o 16x16 bloques de píxeles. Esto puede reducir la granularidad de los segmentos o de los fragmentos que se determinan, pero puede reducir sustancialmente la complejidad de procesamiento y el uso de recursos.

15 La medida de calidad de la predicción para un elemento dado se genera para ser indicativa de una diferencia entre los valores de los píxeles en la primera imagen fuente para los píxeles en el elemento y los valores de los píxeles predichos para los píxeles en el elemento. De este modo, un elemento puede estar formado por uno o más píxeles y la medida de calidad de la predicción para el elemento puede ser indicativa de la diferencia entre los valores de los píxeles para esos píxeles en la imagen fuente original y los valores de los píxeles que resultarían de una predicción de las imágenes combinadas.

25 Se apreciará que se pueden usar diferentes enfoques para determinar las medidas de calidad de la predicción en diferentes realizaciones. Específicamente, en muchas realizaciones, el evaluador 405 puede proceder a realizar una predicción de cada una de las imágenes fuente a partir de las imágenes combinadas. A continuación, para cada imagen individual y cada píxel individual, se puede determinar la diferencia entre el valor del píxel original y el valor del píxel predicho. Se apreciará que se puede usar cualquier medida de diferencia adecuada, como por ejemplo una simple diferencia absoluta, una diferencia de raíz cuadrada de la suma aplicada a los componentes del valor de los píxeles de, por ejemplo, múltiples canales de color, etc.

35 Una predicción de este tipo puede emular la síntesis de predicción/vista que puede realizar el receptor de la señal de imagen 500 para generar vistas para las posturas de visión de las imágenes fuente. Las medidas de calidad de la predicción reflejan, por lo tanto, lo bien que un

receptor de las imágenes combinadas puede ser capaz de generar las imágenes fuente originales basándose únicamente en las imágenes combinadas.

5 Una imagen predicha para una imagen fuente de las imágenes combinadas puede ser una imagen para la postura de la visión de la imagen fuente generada por la síntesis de la visión de las imágenes combinadas. La síntesis de la visión incluye normalmente un cambio de postura de la visión, y normalmente un cambio de posición de la visión. La síntesis de la visión puede ser una síntesis de la imagen de desplazamiento de la visión.

10 Una predicción de una primera imagen a partir de una segunda imagen puede ser específicamente una síntesis de visión de una imagen en la postura de visión de la primera imagen basada en la segunda imagen (y en la postura de visión de ésta). Por lo tanto, una operación de predicción para predecir una primera imagen a partir de una segunda imagen puede ser un desplazamiento de la postura de visión de la segunda imagen desde la postura
15 de visión asociada a ésta a la postura de visión de la imagen del puño.

Se apreciará que se pueden usar diferentes procedimientos y algoritmos para la síntesis y la predicción de visión en diferentes realizaciones. En muchas realizaciones, se puede usar un algoritmo de síntesis y predicción de visión que toma como entrada una postura de visión de
20 síntesis para la que se va a generar la imagen sintetizada, y una pluralidad de imágenes de entrada, cada una de las cuales está asociada a una postura de visión diferente. El algoritmo de síntesis de la visión puede entonces generar la imagen sintetizada para esta postura de la visión basándose en las imágenes de entrada que pueden incluir normalmente tanto un mapa de textura como la profundidad.

25 Se conocen varios algoritmos de este tipo, y se puede usar cualquier algoritmo adecuado sin desviarse de la invención. Como ejemplo de este enfoque, se pueden generar primero imágenes intermedias de síntesis y predicción para cada imagen de entrada. Esto puede lograrse, por ejemplo, generando primero una malla para la imagen de entrada basada en el
30 mapa de profundidad de la imagen. A continuación, la malla se puede deformar/desplazar desde la postura de la visión de la imagen de entrada hasta la postura de la visión de síntesis basándose en cálculos geométricos. Los vértices de la malla resultante pueden proyectarse en la imagen intermedia de síntesis/ predicción y el mapa de textura puede superponerse a esta imagen. Este proceso puede implementarse, por ejemplo, usando el procesamiento de
35 vértices y los sombreadores de fragmentos conocidos, por ejemplo, en las canalizaciones gráficas estándar.

De este modo, para cada una de las imágenes de entrada se puede generar una imagen intermedia de síntesis/previsión (en adelante, solo imagen intermedia de predicción) para la postura de la visión de síntesis.

5

Las imágenes de predicción intermedias pueden entonces combinarse entre sí, por ejemplo, mediante una combinación/suma ponderada o mediante una combinación de selección. Por ejemplo, en algunas realizaciones, cada píxel de la imagen de síntesis/predicción para la postura de la visión de síntesis puede generarse seleccionando el píxel de la imagen de predicción intermedia que esté más adelantado, o el píxel puede generarse mediante una suma ponderada del valor del píxel correspondiente para todas las imágenes de predicción intermedias, donde el peso para una imagen de predicción intermedia dada depende de la profundidad determinada para ese píxel. La operación de combinación también se conoce como operación de mezcla.

10

15

En algunas realizaciones, las medidas de calidad de la predicción pueden llevarse a cabo sin realizar una predicción completa, sino que se puede usar una medida indirecta de la calidad de la predicción.

20

La medida de calidad de la predicción, por ejemplo, se puede determinar indirectamente evaluando un parámetro del proceso implicado en el cambio de visión. Por ejemplo, la cantidad de distorsión geométrica (estiramiento) que se produce en una primitiva (normalmente un triángulo) al realizar el cambio de postura de la visión. Cuanto mayor sea la distorsión geométrica, menor será la medida de calidad de la predicción para cualquier píxel representado por esta primitiva.

25

El evaluador 405 puede así determinar las medidas de calidad de la predicción para los elementos de la pluralidad de imágenes fuente, donde una medida de calidad de la predicción para un elemento de una primera imagen fuente es indicativa de una diferencia entre los valores de píxeles predichos para los píxeles en el elemento predicho desde la pluralidad de imágenes combinadas y los valores de píxeles en la primera imagen fuente para los píxeles en el elemento.

30

El evaluador 405 está acoplado a un determinador 407 que está dispuesto para determinar segmentos de las imágenes fuente que comprenden elementos para los que la medida de calidad de la predicción es indicativa de que la diferencia está por encima de un umbral/la

35

medida de calidad de la predicción es indicativa de que la calidad de la predicción está por debajo de un umbral.

5 Los segmentos pueden corresponder a elementos inducibles determinados por el evaluador 405 y para los cuales la medida de calidad de la predicción está por debajo de un umbral de calidad. Sin embargo, en muchas realizaciones, el determinador 407 puede estar dispuesto para generar segmentos agrupando dichos elementos, y de hecho la agrupación puede incluir también algunos elementos para los que la medida de calidad de la predicción está por encima del umbral.

10

Por ejemplo, en algunas realizaciones, el determinador 407 puede estar dispuesto para generar segmentos agrupando todos los elementos adyacentes que tienen una medida de calidad de predicción por debajo de un umbral de calidad (en adelante, medidas de calidad de predicción bajas y elementos de baja calidad, respectivamente).

15

En otras realizaciones, el determinador 407 puede estar dispuesto, por ejemplo, para ajustar segmentos de un tamaño y una forma determinados a las imágenes, de tal manera que incluyan el mayor número posible de elementos de baja calidad.

20

En consecuencia, el determinador 407 genera un conjunto de segmentos que incluyen los elementos de baja calidad y que, por lo tanto, no pueden predecirse con suficiente precisión a partir de las imágenes combinadas. Normalmente, los segmentos corresponderán a una proporción baja de las imágenes fuente y, por tanto, a una cantidad relativamente pequeña de datos de imagen y de píxeles.

25

El determinador 407 y el generador de imágenes combinadas 403 están acoplados a un generador de señales de imagen 409 que recibe las imágenes combinadas y los segmentos. El generador de señales de imagen 409 está dispuesto para generar una señal de imagen que comprende datos de imagen que representan las imágenes combinadas y datos de imagen que representan los segmentos.

30

El generador de señales de imagen 409 puede codificar específicamente las imágenes combinadas y los segmentos y puede hacerlo específicamente de forma diferente y usar diferentes algoritmos y estándares de codificación para las imágenes combinadas y para los segmentos.

35

Normalmente, las imágenes combinadas se codifican usando algoritmos y estándares de codificación de imágenes altamente eficientes, o algoritmos y estándares de codificación de vídeo altamente eficientes si las imágenes son fotogramas de una señal de vídeo.

5 La codificación de los segmentos puede ser normalmente menos eficiente. Por ejemplo, los segmentos pueden combinarse en imágenes de segmentos en las que cada imagen puede comprender segmentos de una pluralidad de imágenes fuente. Estas imágenes de segmentos combinadas pueden codificarse usando un algoritmo estándar de codificación de imágenes o de vídeo. Sin embargo, debido a la naturaleza mixta y parcial de tales imágenes de segmento
10 combinadas, la codificación es normalmente menos eficiente que para las imágenes completas normales.

Como otro ejemplo, debido a la naturaleza dispersa de los segmentos, éstos pueden no ser almacenados en cuadros/imágenes completos. En algunas realizaciones, los segmentos
15 pueden representarse, por ejemplo, como mallas en el espacio 3D usando VRML (Virtual Reality Modeling Language).

Los datos de la imagen de los segmentos pueden ir acompañados normalmente de metadatos indicativos del origen de los segmentos, como por ejemplo las coordenadas de la imagen
20 original y el origen de la cámara/imagen fuente.

La señal de imagen se transmite en este ejemplo al receptor de señal de imagen 500 que forma parte del dispositivo cliente de RV 101. El receptor de señal de imagen 500 comprende un receptor de señal de imagen 501 que recibe la señal de imagen del transmisor de señal de
25 imagen 400. El receptor de señal de imagen 501 está dispuesto para decodificar la señal de imagen recibida para recuperar las imágenes combinadas y los segmentos.

El receptor de la señal de imagen 501 está acoplado a un procesador de imágenes 503 que está dispuesto para procesar la señal de imagen, y específicamente las imágenes combinadas
30 y los segmentos.

En muchas realizaciones, el procesador de imágenes 503 puede estar dispuesto para sintetizar imágenes de visión para diferentes posturas de visión basadas en las imágenes y los segmentos combinados.

35

En algunas realizaciones, el procesador de imágenes 503 puede proceder a sintetizar primero las imágenes fuente. Las partes de los mensajes fuente sintetizados para los que se incluye un segmento en la señal de imagen pueden entonces sustituirse por los datos de imagen de los segmentos proporcionados. Las imágenes fuente resultantes pueden usarse entonces para la síntesis de imágenes convencional.

En otras realizaciones, las imágenes y los segmentos combinados pueden usarse directamente sin recuperar primero las imágenes fuente.

Se apreciará que el transmisor de la señal de imagen 400 y el receptor de la señal de imagen 500 comprenden la funcionalidad necesaria para comunicar la señal de imagen, incluyendo la funcionalidad para codificar, modular, transmitir, recibir, etc. la señal de imagen. Se apreciará que dicha funcionalidad dependerá de las preferencias y requisitos de la realización individual y que dichas técnicas serán conocidas por el experto en la materia y, por lo tanto, en aras de la claridad y de la brevedad, no se tratarán más en el presente documento.

Se pueden usar diferentes enfoques para generar las imágenes combinadas en diferentes realizaciones.

En algunas realizaciones, el generador de imágenes combinadas 403 puede estar dispuesto para generar las imágenes combinadas mediante la selección de píxeles de las imágenes fuente. Por ejemplo, para cada píxel de una imagen combinada, el generador de imágenes combinadas 403 puede seleccionar un píxel de una de las imágenes fuente.

Un mapa de imagen y/o de profundidad comprende píxeles que tienen valores que pueden considerarse que representan la correspondiente propiedad de la imagen (intensidad/intensidad de luz o profundidad) de la escena a lo largo de un rayo que tiene una dirección de rayo (orientación) desde un origen de rayo (posición). El origen del rayo es normalmente la postura de visión de la imagen, pero en algunas representaciones puede variar sobre una base de píxeles (como, por ejemplo, en el caso de la estereoscopía omnidireccional, donde se puede considerar que la imagen como tal tiene una postura de visión correspondiente al centro del círculo de estereoscopía omnidireccional, pero cada píxel tiene una postura de visión individual correspondiente a la posición en el círculo de estereoscopía omnidireccional). La dirección del rayo puede variar en función del píxel, especialmente en el caso de las imágenes en las que todos los píxeles tienen el mismo origen del rayo (es decir, hay una única postura de visión común de la imagen). El origen y/o la

dirección del rayo también suelen denominarse postura de rayo o postura de proyección de rayo.

De este modo, cada píxel está vinculado a una posición que es el origen de un rayo/línea
5 recta. Además, cada píxel está vinculado a una dirección que es la dirección del rayo/línea
recta desde el origen. Por consiguiente, cada píxel está vinculado a un rayo/línea recta que
está definido por una posición/origen y una dirección desde esta posición/origen. El valor del
píxel viene dado por la propiedad apropiada para la escena en la primera intersección del rayo
10 para el píxel y un objeto de la escena (incluyendo un fondo). Por lo tanto, el valor del píxel
representa una propiedad de la escena a lo largo de un rayo/línea recta que se origina en una
posición de origen del rayo y que tiene una dirección de rayo asociada al píxel. El valor del
píxel representa una propiedad de la escena a lo largo de un rayo que tiene la posición del
rayo del píxel.

15 Por lo tanto, para un primer píxel dado en la imagen combinada que se está generando, el
generador de imágenes combinadas 403 puede determinar los píxeles correspondientes en
las imágenes fuente como píxeles que representan la misma dirección del rayo. Los píxeles
correspondientes pueden ser píxeles que representan la misma dirección del rayo pero que
pueden tener posiciones diferentes ya que las imágenes fuente pueden corresponder a
20 posiciones diferentes.

Así, en principio, para un píxel dado de la imagen combinada, el generador de imágenes
combinadas 403 puede determinar una dirección de rayo y, a continuación, determinar todos
los píxeles de las imágenes fuente que tienen las mismas (dentro de un requisito de similitud
25 dado) direcciones de rayo y considerarlas como píxeles correspondientes. De este modo, los
píxeles correspondientes tendrán normalmente la misma dirección de rayo pero diferentes
posiciones/origen de rayo.

Las vistas de las diferentes imágenes de postura de la visión de origen pueden, por ejemplo,
30 remuestrearse de forma que las coordenadas de la imagen correspondientes tengan las
direcciones de rayo correspondientes. Por ejemplo, cuando las vistas de origen se
representan en un formato de proyección equirectangular parcial, se remuestran a una
versión completa de $360^\circ/180^\circ$. Por ejemplo, se puede definir una esfera de visión que rodee
toda la configuración de la visión fuente. Esta esfera de visión puede dividirse en píxeles y
35 cada píxel tiene una dirección de rayo. Para una imagen fuente determinada, cada píxel puede
ser remuestreado a la representación de la esfera de visión mediante el valor del píxel de la

esfera de visión para una dirección de rayo determinada que se establece en el valor del píxel de la visión de origen que tiene la misma dirección de rayo.

5 El remuestreo de las imágenes fuente en una representación de la superficie de la esfera de visión completa suele dar como resultado N imágenes parcialmente llenas, ya que las imágenes individuales suelen tener puertos de visión limitados y N es el número de imágenes fuente. Sin embargo, los puertos de visión tienden a superponerse y, en consecuencia, el conjunto de representaciones de la superficie de la esfera de visión tiende a proporcionar múltiples valores de píxeles para cualquier dirección dada.

10

El generador de imágenes combinadas 403 puede ahora proceder a generar al menos una, pero normalmente una pluralidad de imágenes combinadas seleccionando entre los píxeles correspondientes.

15 En concreto, se puede generar una primera imagen combinada para cubrir una parte de la escena. Por ejemplo, se puede generar una imagen combinada que tenga un tamaño predeterminado para cubrir una determinada área de píxeles en las representaciones de la esfera de visión, describiendo así esta sección de la escena. En algunas realizaciones, cada una de las imágenes combinadas puede cubrir toda la escena e incluir toda la superficie de la
20 esfera de visión.

Para cada píxel en la primera imagen combinada, el generador de imágenes combinadas 403 puede ahora considerar los píxeles correspondientes en las representaciones de la esfera de visión y proceder a seleccionar uno de los píxeles. El generador de imágenes combinadas
25 403 puede generar específicamente la primera imagen combinada seleccionando el valor de píxel para la imagen combinada como el valor de píxel para el píxel correspondiente en la imagen fuente de visión para la cual el píxel correspondiente representa un rayo que tiene la mayor distancia desde el punto central en una primera dirección a lo largo de un primer eje perpendicular a una dirección de rayo para el píxel correspondiente.

30

La distancia desde el punto central a una dirección de rayo puede determinarse como la distancia entre los rayos del punto central y el píxel correspondiente para ese píxel de la imagen combinada.

35 La selección puede ejemplificarse en la Fig. 6, que se basa en el ejemplo de una configuración de postura de visión de origen circular que tiene un punto central C.

En este ejemplo, se considera la determinación de un píxel de una imagen combinada que tiene una dirección de rayo r_c . Las cámaras/vistas de origen 1-4 capturan esta dirección y, por lo tanto, existen cuatro píxeles correspondientes. Cada uno de estos píxeles correspondientes
5 representa una postura diferente y, en consecuencia, representan rayos que se originan en diferentes posiciones, tal como se muestra. En consecuencia, existe una distancia de desplazamiento p_1 - p_4 entre los rayos y el rayo de la imagen combinada r_c , correspondiente a la distancia entre el punto central C y los rayos cuando éstos se extienden hacia atrás (para cruzar el eje 601).

10

La Fig. 6 también muestra una dirección/eje 601 perpendicular al rayo r_c . Para una primera imagen combinada, el generador de imágenes combinadas 403 puede ahora seleccionar el píxel correspondiente para el cual la distancia del rayo en esta dirección es la mayor. Por lo tanto, en este caso, el valor del píxel de la imagen combinada se seleccionará como el valor
15 del píxel para la cámara/vista 1, ya que p_1 es la mayor distancia en esta dirección.

El generador de imágenes combinadas 403 puede proceder normalmente a determinar una segunda imagen combinada realizando la misma operación pero seleccionando los píxeles correspondientes que tengan la mayor distancia en la dirección opuesta (podría considerarse
20 que la generación de la primera y segunda imágenes combinadas puede ser mediante la selección de la mayor distancia positiva y negativa respectivamente con respecto a la primera dirección si la distancia se mide como positiva cuando está en la misma dirección que el eje y negativa cuando está en la otra dirección). Por lo tanto, en este caso, el generador de imágenes combinadas 403 seleccionará el valor de los píxeles de la imagen combinada como
25 el valor de los píxeles de la cámara/vista 4, ya que p_4 es la mayor distancia en esta dirección.

En muchas realizaciones, el generador de imagen combinada 403 puede proceder además a generar una tercera imagen combinada realizando la misma operación pero seleccionando los píxeles correspondientes que tengan la menor distancia en cualquier dirección (la menor
30 distancia absoluta). Por lo tanto, en este caso, el generador de imágenes combinadas 403 seleccionará el valor del píxel de la imagen combinada como el valor del píxel de la cámara/vista 3, ya que p_3 es la distancia más pequeña.

De este modo, el generador de imágenes combinadas 403 puede generar tres imágenes
35 combinadas para la misma parte de la escena (y posiblemente para toda la escena). Una de las imágenes corresponderá a una selección de píxeles que proporcionan la vista más lateral

de la escena desde una dirección, otra que representa la vista más lateral de la escena desde la dirección opuesta y otra que representa la vista más central de la escena. Esto puede ilustrarse con la Fig. 7, que muestra las direcciones de visión seleccionadas de cada vista/cámara para, respectivamente, la imagen combinada central y las dos imágenes combinadas laterales.

Las imágenes resultantes proporcionan así una representación muy eficiente de la escena, con una imagen combinada que suele proporcionar la mejor representación de los objetos en primer plano y las otras dos que se combinan para proporcionar datos centrados en el fondo.

En algunas realizaciones, el generador de imágenes combinadas 403 puede estar dispuesto para generar además una o más imágenes combinadas seleccionando los píxeles correspondientes según una dirección de eje que es perpendicular a la dirección de rayo, pero que es diferente de la dirección de eje usada anteriormente. Este enfoque puede ser adecuado para configuraciones de visión de origen no planas (es decir, configuraciones tridimensionales). Por ejemplo, para una configuración de postura de la vision fuente esférica, pueden considerarse más de dos planos. Por ejemplo, se puede considerar un plano a 0, 60 y 120 grados, o dos planos ortogonales (por ejemplo, planos izquierda-derecha y arriba-abajo).

En algunas realizaciones, las imágenes combinadas pueden generarse mediante la síntesis/previsión de visión a partir de las imágenes fuente. El generador de imágenes 403 puede generar específicamente imágenes combinadas que representen vistas de la escena desde diferentes posiciones de visión, y específicamente desde posiciones de visión distintas de las de las imágenes fuente. Además, a diferencia de la síntesis de imágenes convencional, una imagen combinada no se genera para representar la vista de la escena desde una única posición de visión/captura, sino que puede representar la escena desde diferentes posiciones de visión incluso dentro de la misma imagen combinada. Por lo tanto, una imagen combinada puede generarse generando valores de píxeles para los píxeles de la imagen combinada mediante la síntesis/previsión de la vision a partir de las imágenes fuente, pero con los valores de los píxeles representando diferentes posiciones de la vision.

Específicamente, para un píxel dado en la imagen combinada, se puede realizar la síntesis/predicción de la vision para determinar el valor del píxel correspondiente a la postura de rayo específica para ese píxel. Esto puede repetirse para todos los píxeles de la imagen

combinada, pero con al menos algunos de los píxeles que tienen posturas de rayo con diferentes posiciones.

5 Por ejemplo, una sola imagen combinada puede proporcionar una representación de 360° de la escena correspondiente, por ejemplo, a una superficie de una esfera de visión que rodea toda la configuración de la vision de origen. Sin embargo, las vistas de diferentes partes de la escena pueden representarse desde diferentes posiciones dentro de la misma imagen combinada. La Fig. 8 ilustra un ejemplo en el que la imagen combinada comprende píxeles que representan dos posiciones de rayo diferentes (y, por tanto, posiciones de visión de 10 píxeles), a saber, un primer origen de rayo 801 que se usa para los píxeles que representan un hemisferio y un segundo origen de rayo 803 que representa el otro hemisferio. Para cada una de estas posiciones/origen de rayo, los píxeles están provistos de diferentes direcciones de rayo como se muestra en las flechas. En el ejemplo específico, la configuración de la vision de origen comprende ocho vistas de origen (1-8) en una disposición circular. Cada vista de 15 cámara solo proporciona una vista parcial, por ejemplo una vista de 90°, pero con una superposición entre las vistas. Para un píxel dado en la imagen combinada, puede haber una postura de rayo asociada, y el valor del píxel para esta postura de rayo puede determinarse mediante la síntesis/previsión de la vision a partir de las vistas de origen.

20 En principio, cada píxel de la imagen combinada puede sintetizarse individualmente, pero en muchas realizaciones se realiza una síntesis combinada para una pluralidad de píxeles. Por ejemplo, puede sintetizarse una única imagen de 180° para la primera posición 801 a partir de las imágenes fuente de la vision (por ejemplo, usando las posiciones 2, 1, 8, 7, 6, 5, 4) y puede sintetizarse una única imagen de 180° para la segunda posición 803 a partir de las 25 imágenes fuente de la vision (por ejemplo, usando las posiciones 6, 5, 4, 3, 2, 1, 8). La imagen combinada puede generarse entonces combinando estas imágenes. Si las imágenes sintetizadas por separado se superponen, se puede usar la combinación o la mezcla para generar la imagen combinada. Alternativamente, las partes superpuestas de las imágenes combinadas pueden ser silenciadas, por ejemplo, asignando un color reservado o un valor de 30 profundidad. De este modo, se aumenta la eficacia de la codificación de vídeo.

En muchas realizaciones, una o más de las imágenes combinadas pueden generarse para representar la escena desde un punto de visión que proporcione una mirada más lateral a la escena. Por ejemplo, en la Fig. 8, el centro del círculo de la vision corresponde al punto central 35 de las posturas de la vision de origen y al centro de las posiciones de los orígenes de rayo para la imagen combinada. Sin embargo, las direcciones de rayo para un determinado origen

de rayo 801, 803 no están en una dirección predominantemente radial, sino que proporcionan una vista lateral de la escena. Específicamente, en el ejemplo, tanto el primer origen de rayo 801 como el segundo origen 803 proporcionan vistas en dirección a la izquierda, es decir, las direcciones de rayo para ambos están a la izquierda cuando se mira el origen de rayo 801, 803 desde el punto central.

El generador de imágenes 403 puede proceder a generar una segunda imagen combinada que represente una vista diferente de la escena, y específicamente puede a menudo generar ventajosamente una segunda vista de la escena que sea complementaria a la primera vista pero que mire en la dirección opuesta. Por ejemplo, el generador de imágenes 403 puede generar una segunda imagen combinada que utilice los mismos orígenes de rayo pero en la que las direcciones de rayo estén en la dirección opuesta. Por ejemplo, el generador de imágenes 403 puede generar una segunda imagen combinada correspondiente a la configuración de la Fig. 9.

Las dos imágenes pueden proporcionar una representación muy ventajosa y complementaria de la escena, y pueden proporcionar normalmente una representación mejorada de las partes del fondo de la escena.

En muchas realizaciones, la imagen combinada también puede incluir una o más imágenes que se generan para proporcionar una vista más frontal, como por ejemplo una correspondiente a la configuración de la Fig. 10. Este ejemplo puede proporcionar en muchas realizaciones una representación mejorada de la parte frontal de los objetos en primer plano.

Se apreciará que se pueden usar diferentes configuraciones de origen de rayo en diferentes realizaciones y que específicamente se pueden usar más orígenes. Por ejemplo. Las Figs. 11 y 12 muestran ejemplos de dos configuraciones complementarias para generar imágenes combinadas de aspecto lateral en las que los orígenes de rayo se distribuyen en una curva (específicamente un círculo) que en este caso rodea la configuración de la vision fuente (a menudo dicha curva se seleccionaría para ajustarse a la configuración de la postura de la vision fuente). Las figuras solo muestran los orígenes y las posturas para una parte del círculo/curva y se apreciará que en muchas realizaciones se generará una vista esférica completa o de 360°.

La Fig. 7 puede considerarse que ilustra otra configuración a modo de ejemplo en la que se generan tres imágenes combinadas basadas en ocho posiciones de rayo en un círculo

alrededor de un punto central. Para la primera imagen combinada, se seleccionan las direcciones alrededor de un círculo radial, para la segunda imagen se seleccionan las direcciones de rayo alrededor de un ángulo de 90° a la derecha, y para la tercera imagen se seleccionan las direcciones de rayo alrededor de un ángulo de 90° a la izquierda. Esta combinación de imágenes combinadas puede proporcionar una representación combinada altamente eficiente de una escena.

En algunas realizaciones, el generador de imágenes 403 puede estar dispuesto para generar valores de píxeles para las imágenes combinadas para posturas de rayo específicas mediante la síntesis de visión de las imágenes fuente. Las posturas de rayo pueden seleccionarse de forma diferente para las distintas imágenes combinadas.

Específicamente, en muchas realizaciones, las posturas de rayo para una imagen pueden ser seleccionadas para proporcionar una vista lateral de la escena desde el origen de rayo, y las posturas de rayo de otra imagen pueden ser seleccionadas para proporcionar una vista lateral complementaria.

Específicamente, para una primera imagen combinada, las posturas de rayo pueden ser tales que un producto de puntos entre un vector vertical y los vectores del producto cruzado de los píxeles sea no negativo para al menos el 90 % (a veces el 95 % o incluso todos) de los píxeles de la primera imagen combinada. El vector de producto cruzado de píxeles para un píxel se determina como un producto cruzado entre una dirección de rayo para un píxel y un vector desde un punto central para las diferentes posturas de visión fuente hasta una posición de rayo para el píxel.

El punto central para las posturas de la visión de origen puede generarse como una posición media o promedio para las posturas de la visión de origen. Por ejemplo, cada coordenada (por ejemplo, x, y, z) puede ser promediada individualmente y la coordenada promedio resultante puede ser el punto central. Hay que tener en cuenta que el punto central de una configuración no está (necesariamente) en el centro del círculo/esfera más pequeño que comprende las posturas de la visión de origen.

El vector desde el punto central hasta el origen del rayo para un píxel dado es, por tanto, un vector en el espacio de la escena que define una distancia y una dirección desde el punto central hasta la posición de la visión para ese píxel. La dirección del rayo puede ser representada por cualquier vector que tenga la misma dirección, es decir, puede ser un vector

desde el origen del rayo hacia el punto de la escena representado por el píxel (y por lo tanto también puede ser un vector en el espacio de la escena).

5 El producto cruzado entre estos dos vectores será perpendicular a ambos. Para un plano horizontal (en el sistema de coordenadas de la escena), una dirección de rayo hacia la izquierda (vista desde el punto central) dará como resultado un vector de producto cruzado que tiene un componente hacia arriba, es decir, que tiene un componente z positivo en un sistema de coordenadas de la escena x,y,z donde z indica la altura. El vector producto cruzado será hacia arriba para cualquier vista hacia la izquierda, independientemente del origen del rayo, por ejemplo, será hacia arriba para todos los píxeles/posiciones del rayo de la Fig. 8.

15 Por el contrario, para una vista hacia la derecha, el vector del producto cruzado será hacia abajo para todas las posturas de rayo, por ejemplo, para todos los píxeles/posiciones de rayo de la Fig. 9 resultará una coordenada z negativa.

20 El producto de puntos entre un vector vertical en el espacio de la escena y todos los vectores que tengan una coordenada z positiva será el mismo, concretamente será positivo para un vector vertical que apunte hacia arriba y negativo para un vector vertical que apunte hacia abajo. A la inversa, para una coordenada z negativa, el producto de puntos será negativo para un vector vertical que apunta hacia arriba y positivo para un vector vertical que apunta hacia abajo. En consecuencia, el producto de puntos tendrá el mismo signo para las posturas de rayo hacia la derecha y el signo opuesto para todas las posturas de rayo hacia la izquierda.

25 En algunos escenarios, puede resultar un vector nulo o un producto de puntos (por ejemplo, para puntos polares en un círculo de visión) y para tales posturas de rayo, el signo no será diferente de las vistas a la izquierda o a la derecha.

30 Se apreciará que las consideraciones anteriores también se aplican, mutatis mutandis, a una representación tridimensional, como, por ejemplo, cuando los orígenes de rayo se sitúan en una esfera.

35 De este modo, en algunas realizaciones, al menos el 90 %, y en algunas realizaciones al menos el 95 % o incluso todos los píxeles de una imagen combinada dan como resultado un producto de puntos que no tiene signos diferentes, es decir, al menos ese número de píxeles tendrá una vista lateral hacia el mismo lado.

En algunas realizaciones, se pueden generar las imágenes combinadas para que tengan bandas de guarda o, por ejemplo, algunos píxeles de borde específicos pueden tener circunstancias concretas para las que el producto de puntos puede no cumplir el requisito. Sin embargo, para la gran mayoría de los píxeles, el requisito se cumple, y los píxeles proporcionan las correspondientes vistas laterales.

Además, en muchas realizaciones, al menos dos imágenes combinadas cumplen estos requisitos pero con los signos de los productos de puntos opuestos. Así, para una imagen combinada, al menos el 90 % de los píxeles puede representar una vista a la derecha y para otra imagen combinada al menos el 90 % de los píxeles puede representar una vista a la izquierda.

Las imágenes combinadas pueden generarse para posturas que proporcionan una vista particularmente ventajosa de la escena. Los inventores se han dado cuenta de que, en muchos escenarios, puede ser particularmente ventajoso generar imágenes combinadas para posturas de visión que den como resultado una vista más lateral de la parte principal de la escena, y además que para una configuración dada de las vistas de origen, puede ser ventajoso generar al menos algunas vistas que estén cerca de las posiciones extremas de la configuración en lugar de cerca del centro de la configuración.

Por lo tanto, en muchas realizaciones, al menos una, y normalmente al menos dos, de las imágenes combinadas se generan para posturas de rayo que están próximas al borde de una región correspondiente a la configuración de la visión de origen.

La región puede ser específicamente una región del espacio (una colección o un conjunto de puntos en el espacio), que está limitada por un polígono más grande que se puede formar usando al menos algunas de las posiciones de la visión como vértices para las líneas rectas del polígono. El polígono puede ser una figura plana delimitada por una cadena finita de segmentos de líneas rectas que se cierran en un bucle para formar una cadena o un circuito cerrado, y esto puede incluir una configuración unidimensional como la de la Fig. 2A (también conocida como polígono degenerado). Para una configuración tridimensional, la región puede corresponder a un poliedro más grande posible formado por al menos algunas de las posiciones de la visión de origen. Por lo tanto, la región puede ser el mayor polígono o poliedro que se puede formar usando al menos algunas de las posiciones de la visión de origen como vértices para las líneas del polígono o poliedro.

Como alternativa, una región que comprenda las diferentes posiciones de visión de la pluralidad de imágenes fuente puede ser una línea, un círculo o una esfera más pequeños que incluya todas las posiciones de visión. La región puede ser específicamente una esfera más pequeña que incluya todas las posiciones de visión fuente.

5

De este modo, en muchas realizaciones, las posturas de rayo de al menos una de las imágenes combinadas se seleccionan para que estén cerca del borde de la región que comprende la configuración de posturas de la visión de origen.

10 En muchas realizaciones, se determina que al menos una posición de rayo de las imágenes combinadas es inferior a una primera distancia del borde de la región, donde esta primera distancia no es superior al 50 % o, en muchos casos, al 25 % o al 10 % de la distancia máxima (interior) entre puntos del borde de la región. Así, desde la posición de la visión, una distancia mínima al borde puede ser no más del 50 %, 25 % o 10 % de una distancia máxima al borde.

15

Esto puede ilustrarse en la Fig. 13, que muestra un ejemplo de puntos de visión fuente indicados por puntos negros. La Fig. 13 ilustra además una región correspondiente a la esfera más pequeña que incluye las posturas de visión. En el ejemplo, la configuración de visión es una configuración plana, bidimensional y la consideración de una esfera se reduce a una consideración de un círculo 1301. La Fig. 13 muestra además una postura de rayo 1303 para una imagen combinada que está próxima al borde de la esfera/ círculo/ región. Específicamente, la distancia mínima d_{min} al borde de la región es mucho menor (alrededor del 10 %) de la distancia máxima d_{max} al borde/ de la región.

20

25 En algunas realizaciones, las posturas de rayo de una imagen combinada pueden determinarse para que sean inferiores a una primera distancia desde el borde de la región, donde la primera distancia no es superior al 20 %, o a menudo incluso al 10 % o al 5 % de la distancia máxima entre dos posturas de visión de origen. En el ejemplo en el que la región se determina como la esfera/círculo más pequeño que incluye todas las posturas de la visión de origen, la distancia máxima entre dos posturas de la visión es igual al diámetro de la esfera/círculo y, por lo tanto, la postura de la visión de la imagen combinada puede seleccionarse de forma que la distancia mínima d_{min} cumpla este requisito.

30

En algunas realizaciones, las posturas de rayo de una imagen combinada pueden determinarse para estar al menos a una distancia mínima de un punto central de las diferentes posturas de visión, donde la distancia mínima es al menos el 50 %, y a menudo incluso el 75

35

% o el 90 %, de la distancia desde el punto central hasta el borde a lo largo de una línea que pasa por el punto central y la postura de rayo.

5 En algunas realizaciones, se seleccionan dos posturas de visión para las imágenes combinadas de tal manera que una distancia entre éstas es al menos el 80 %, y a veces incluso el 90 % o el 95 %, de la distancia máxima entre dos puntos de un borde que cruza una línea a través de las dos posturas de visión. Por ejemplo, si se traza una línea a través de las dos posturas, la distancia entre las dos posturas es al menos el 80 %, 90 % o 95 % de la distancia entre los puntos en los que la línea cruza el círculo.

10

En algunas realizaciones, una distancia máxima entre dos de las posturas de rayo de la primera imagen combinada es al menos el 80 % de una distancia máxima entre puntos del borde de una región que comprende las diferentes posturas de visión de la pluralidad de imágenes fuente.

15

Los inventores han tenido la idea de que el enfoque de generar imágenes combinadas para posiciones cercanas al borde de la región que comprende las posturas de la visión de origen puede ser particularmente ventajoso, ya que tiende a proporcionar una mayor información de los objetos de fondo en la escena. La mayoría de los datos de fondo son capturados normalmente por las cámaras o áreas de imagen que tienen la mayor distancia lateral con respecto a un punto de visión central. Esto puede ser ventajoso combinado con una imagen combinada más central ya que esto tiende a proporcionar una mejor información de la imagen para los objetos de primer plano.

20

25 En muchas realizaciones, el generador de señales de imagen 409 puede estar dispuesto para incluir además metadatos para los datos de imagen generados. Específicamente, el generador de imágenes combinadas 403 puede generar datos de origen para las imágenes combinadas, donde los datos de origen indican cuál de las imágenes fuente es el origen para los píxeles individuales en las imágenes combinadas. El generador de señales de imagen 409 puede entonces incluir estos datos en la señal de imagen generada.

30

En muchas realizaciones, el generador de señales de imagen 409 puede incluir datos de postura de la visión fuente indicativos de las posturas de la visión para las imágenes fuente. Los datos pueden incluir específicamente datos que definen la posición y la dirección de cada imagen/vista fuente.

35

La señal de imagen puede incluir metadatos que indiquen, posiblemente de forma individual para cada píxel, la posición y la dirección para las que se proporcionan los valores de los píxeles, es decir, una indicación de la posición del rayo. En consecuencia, el receptor de la señal de imagen 500 puede estar dispuesto a procesar estos datos para realizar, por ejemplo, la síntesis de la visión.

Por ejemplo, para cada píxel de las tres vistas generadas por la selección de los píxeles correspondientes, se pueden incluir metadatos que indiquen la identidad de la visión de origen. Esto puede dar lugar a tres mapas de etiquetas, uno para la vista central y dos para las vistas laterales. Las etiquetas pueden estar vinculadas a datos específicos de la visión, incluyendo, por ejemplo, la óptica de la cámara y la geometría del equipo.

Se apreciará que la descripción anterior, para mayor claridad, ha descrito realizaciones de la invención con referencia a diferentes circuitos funcionales, unidades y procesadores. Sin embargo, será evidente que cualquier distribución adecuada de la funcionalidad entre diferentes circuitos funcionales, unidades o procesadores puede ser usada sin apartarse de la invención. Por ejemplo, la funcionalidad ilustrada para ser realizada por procesadores o controladores separados puede ser realizada por el mismo procesador o controladores. Por lo tanto, las referencias a unidades o circuitos funcionales específicos deben considerarse únicamente como referencias a los medios adecuados para proporcionar la funcionalidad descrita y no como una indicación de una estructura u organización lógica o física estricta.

REIVINDICACIONES

1. Un aparato para generar una señal de imagen, comprendiendo el aparato:

un receptor (401) que comprende medios para recibir una pluralidad de imágenes fuente que representan una escena desde diferentes posturas de visión, donde una postura de visión comprende una posición de visión y una dirección de visión;

un generador de imágenes combinadas (403) que comprende medios para generar una pluralidad de imágenes combinadas a partir de las imágenes fuente, derivándose cada imagen combinada de un conjunto de al menos dos imágenes fuente de la pluralidad de imágenes fuente, representando cada píxel respectivo de una imagen combinada una propiedad de la escena en un punto de intersección de un rayo respectivo con la escena, el rayo teniendo una posición de origen y una dirección que juntas definen una postura de rayo, y donde cada imagen combinada incluye al menos dos píxeles representando posturas de rayo de posiciones de origen de postura de rayo respectivas diferentes;

un evaluador (405) que comprende medios para determinar las medidas de calidad de la predicción para los elementos de la pluralidad de imágenes fuente, siendo una medida de calidad de la predicción para un elemento de una primera imagen fuente indicativa de una diferencia entre los valores de píxel en el elemento y los valores de píxel predichos en el elemento, siendo los valores de píxel predichos los valores de píxel resultantes de la predicción de valores de píxel en el elemento basada en la pluralidad de imágenes combinadas;

un determinador (407) que comprende medios para determinar segmentos de las imágenes fuente que comprenden elementos para los que la medida de calidad de la predicción es indicativa de una diferencia por encima de un umbral; y

un generador de señales de imagen (409) que comprende medios para generar una señal de imagen que comprende datos de imagen que representan las imágenes combinadas y datos de imagen que representan los segmentos de las imágenes fuente.

2. El aparato de la reivindicación 1, en el que el generador de imágenes combinadas (403) está dispuesto para generar al menos una primera imagen combinada de la pluralidad de imágenes combinadas mediante la síntesis de visión de los píxeles de la primera imagen combinada de la pluralidad de imágenes fuente, donde cada píxel de la primera imagen combinada representa la escena para una postura de rayo y las posturas de rayo para la primera imagen comprenden posturas de rayo de al menos dos posiciones de origen de postura de rayo diferentes.

3. El aparato de la reivindicación 2, en el que un producto de puntos entre un vector vertical y vectores de producto cruzado de píxeles es no negativo para al menos el 90 % de los píxeles de la primera imagen combinada, siendo un vector de producto cruzado de píxeles para un píxel un producto cruzado entre una dirección de postura de rayo para un píxel y un vector desde un punto central para las diferentes posturas de visión hasta una posición de origen de postura de rayo para el píxel.
- 5
4. El aparato de la reivindicación 3, en el que el generador de imágenes combinadas (403) está dispuesto para generar una segunda imagen combinada de la pluralidad de imágenes combinadas mediante la síntesis de visión de los píxeles de la segunda imagen combinada a partir de la pluralidad de imágenes fuente, en donde cada píxel de la segunda imagen combinada representa la escena para una postura de rayo y las posturas de rayo para la segunda imagen comprenden posturas de rayo de al menos dos posiciones de origen de postura de rayo diferentes; y
- 10
- en donde un producto de puntos entre el vector vertical y los vectores del producto cruzado de píxeles es no positivo para al menos el 90 % de los píxeles de la segunda imagen combinada.
- 15
5. El aparato de la reivindicación 2, en el que las posturas de rayo de la primera imagen combinada se seleccionan para que estén próximas a un borde de una región que comprende las diferentes posturas de visión de la pluralidad de imágenes fuente.
- 20
6. El aparato de las reivindicaciones 2 o 3, en el que se determina que cada una de las posturas de rayo de la primera imagen combinada está a menos de una primera distancia de un borde de una región que comprende las diferentes posturas de visión de la pluralidad de imágenes fuente, siendo la primera distancia no superior al 50 % de una distancia interior máxima entre puntos del borde.
- 25
7. El aparato de cualquier reivindicación anterior, en el que el generador de imágenes combinadas (403), para cada píxel de una primera imagen combinada de la pluralidad de imágenes combinadas, está dispuesto para:
- 30
- determinar un píxel correspondiente en cada una de las imágenes fuente de la visión para las que está presente un píxel correspondiente, siendo el píxel correspondiente uno que representa una misma dirección de rayo que el píxel de la primera imagen combinada;
- 35
- seleccionar un valor de píxel para el píxel de la primera imagen combinada como un valor de píxel del píxel correspondiente en la imagen fuente de la visión para la que el píxel

correspondiente representa un rayo que tiene una mayor distancia desde un punto central para las diferentes posturas de visión, siendo la mayor distancia en una primera dirección a lo largo de un primer eje perpendicular a una dirección de rayo para el píxel correspondiente.

5 8. El aparato de la reivindicación 7, en el que la determinación de los píxeles correspondientes comprende el remuestreo de cada imagen fuente a una representación en imagen que representa al menos una parte de una superficie de una esfera de visión que rodea las posturas de visión y la determinación de los píxeles correspondientes como píxeles que tienen una misma posición en la representación en imagen.

10

9. El aparato de las reivindicaciones 7 u 8, en el que el generador de imágenes combinadas (403), para cada píxel de una segunda imagen combinada, está dispuesto para:

seleccionar un valor de píxel para el píxel en la segunda imagen combinada como un valor de píxel del píxel correspondiente en la imagen fuente de la visión para la cual el píxel correspondiente representa un rayo que tiene una distancia mayor desde el punto central en una dirección opuesta a la primera dirección.

15

10. El aparato de cualquiera de las reivindicaciones 7-9, en el que el generador de imágenes combinadas (403), para cada píxel de una tercera imagen combinada, está dispuesto para:

seleccionar un valor de píxel para el píxel en la tercera imagen combinada como un valor de píxel del píxel correspondiente en la imagen fuente de la visión para la cual el píxel correspondiente representa un rayo que tiene una distancia más pequeña desde el punto central.

20

11. El aparato de cualquiera de las reivindicaciones 7-10 en el que el generador de imágenes combinado (403), para cada píxel de una cuarta imagen combinada, está dispuesto para:

seleccionar un valor de píxel para el píxel en la cuarta imagen combinada como un valor de píxel del píxel correspondiente en la imagen fuente de la visión para la que el píxel correspondiente representa un rayo que tiene una distancia máxima desde el punto central en una segunda dirección a lo largo de un segundo eje perpendicular a una dirección de rayo para el píxel correspondiente, teniendo el primer eje y el segundo eje direcciones diferentes.

30

12. El aparato de cualquiera de las reivindicaciones 7-11, en el que el generador de imágenes combinadas (403) está dispuesto para generar datos de origen para la primera imagen combinada, siendo los datos de origen indicativos de cuál de las imágenes fuente es un origen

35

para cada píxel de la primera imagen combinada; y el generador de señales de imagen (409) está dispuesto para incluir los datos de origen en la señal de imagen.

5 13. El aparato de cualquiera de las reivindicaciones anteriores, en el que el generador de señales de imagen (403) está dispuesto para incluir datos de postura de la vision fuente en la señal de imagen, siendo los datos de postura de la vision fuente indicativos de las diferentes posturas de la vision para las imágenes fuente.

14. Un aparato para recibir una señal de imagen, comprendiendo el aparato:
10 un receptor (501) para recibir una señal de imagen, comprendiendo la señal de imagen una pluralidad de imágenes combinadas, representando cada imagen combinada datos de imagen derivados de un conjunto de al menos dos imágenes fuente de una pluralidad de imágenes fuente que representan una escena desde diferentes posturas de visión, donde una postura de visión comprende una posición de visión y una dirección de visión, representando
15 cada píxel respectivo de una imagen combinada una propiedad de la escena en un punto de intersección de un rayo respectivo con la escena, el rayo teniendo una posición de origen y una dirección que juntas definen una postura de rayo, y donde cada imagen combinada incluye al menos dos píxeles representando posturas de rayo de posiciones de origen de postura de rayo respectivas diferentes;
20 datos de imagen para un conjunto de segmentos de la pluralidad de imágenes fuente, un segmento para una primera imagen fuente comprende al menos un píxel de la primera imagen fuente para el que una medida de calidad de predicción relativa a una predicción de valores de pixel en el segmento basada en la pluralidad de imágenes combinadas, está por debajo de un umbral; y
25 un procesador (503) para procesar la señal de imagen.

15. Un procedimiento para generar una señal de imagen, comprendiendo el procedimiento:
recibir una pluralidad de imágenes fuente que representen una escena desde diferentes posturas de visión, donde una postura de visión comprende una posición de visión y una
30 dirección de visión;
generar una pluralidad de imágenes combinadas a partir de las imágenes fuente, cada imagen combinada derivada de un conjunto de al menos dos imágenes fuente de la pluralidad de imágenes fuente, representando cada píxel respectivo de una imagen combinada una propiedad de la escena en un punto de intersección de un rayo respectivo con la escena, el
35 rayo teniendo una posición de origen y una dirección que juntas definen una postura de rayo,

y donde cada imagen combinada incluye al menos dos píxeles representando posturas de rayo de posiciones de origen de postura de rayo respectivas diferentes ;

determinar medidas de calidad de predicción para elementos de la pluralidad de imágenes fuente, siendo una medida de calidad de predicción para un elemento de una primera imagen fuente indicativa de una diferencia entre los valores de píxel en el elemento y los valores de píxel predichos en el elemento, siendo los valores de píxel predichos los valores de píxel resultantes de la predicción de valores de píxel en el elemento basada en la pluralidad de imágenes combinadas;

5 10 determinar segmentos de las imágenes fuente que comprenden elementos para los que la medida de calidad de la predicción es indicativa de una diferencia superior a un umbral; y

generar una señal de imagen que comprenda datos de imagen que representen las imágenes combinadas y datos de imagen que representen los segmentos de las imágenes fuente.

15 16. Un procedimiento de procesamiento de una señal de imagen, comprendiendo el procedimiento:

recibir una señal de imagen, comprendiendo la señal de imagen:

una pluralidad de imágenes combinadas, representando cada imagen combinada datos de imagen derivados de un conjunto de al menos dos imágenes fuente de una pluralidad de imágenes fuente que representan una escena desde diferentes posturas de visión, donde una postura de visión comprende una posición de visión y una dirección de visión, representando cada píxel respectivo de una imagen combinada una propiedad de la escena en un punto de intersección de un rayo respectivo con la escena, el rayo teniendo una posición de origen y una dirección que juntas definen una postura de rayo, y donde los píxeles de cada imagen combinada incluyen al menos dos píxeles representando posturas de rayo de posiciones de origen de postura de rayo respectivas diferentes; datos de imagen para un conjunto de segmentos de la pluralidad de imágenes fuente, un segmento para una primera imagen fuente comprende al menos un píxel de la primera imagen fuente para el que una medida de calidad de predicción relativa a una predicción de valores de píxel en el segmento basada en la pluralidad de imágenes combinadas está por debajo de un umbral; y

procesar la señal de imagen.

17. Una señal de imagen que comprende

35 una pluralidad de imágenes combinadas, representando cada imagen combinada datos de imagen derivados de un conjunto de al menos dos imágenes fuente de una pluralidad de

imágenes fuente que representan una escena desde diferentes posturas de visión, donde una postura de visión comprende una posición de visión y una dirección de visión, representando cada píxel respectivo de una imagen combinada una propiedad de la escena en un punto de intersección de un rayo respectivo con la escena, el rayo teniendo una posición de origen y una dirección que juntas definen una postura de rayo, y donde los píxeles de cada imagen combinada incluyen al menos dos píxeles representando posturas de rayo de posiciones de origen de postura de rayo respectivas diferentes; datos de imagen para un conjunto de segmentos de la pluralidad de imágenes fuente, comprendiendo un segmento para una primera imagen fuente al menos un píxel de la primera imagen fuente para el que una medida de calidad de predicción para una predicción del segmento basada en la pluralidad de imágenes combinadas está por debajo de un umbral.

18. Un producto de programa informático que comprende medios de código de programa informático adaptados para hacer que un ordenador lleve a cabo todos los pasos de las reivindicaciones 15 o 16 cuando se ejecuta dicho programa en el ordenador.

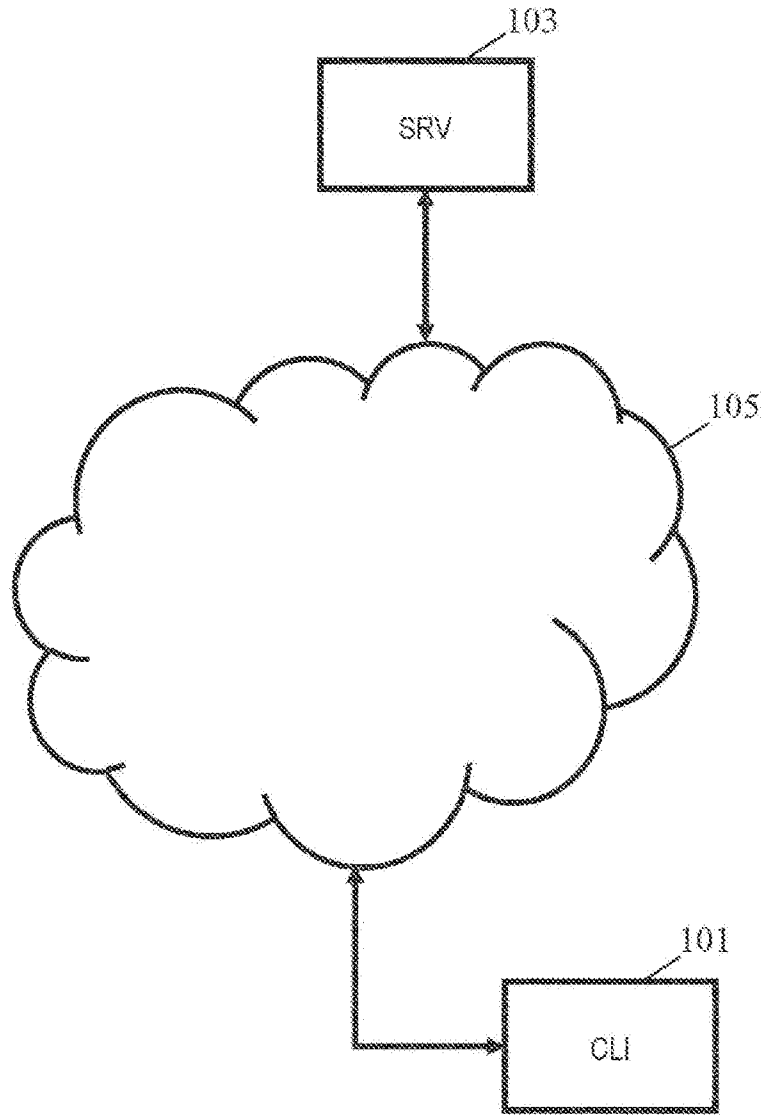


FIG. 1

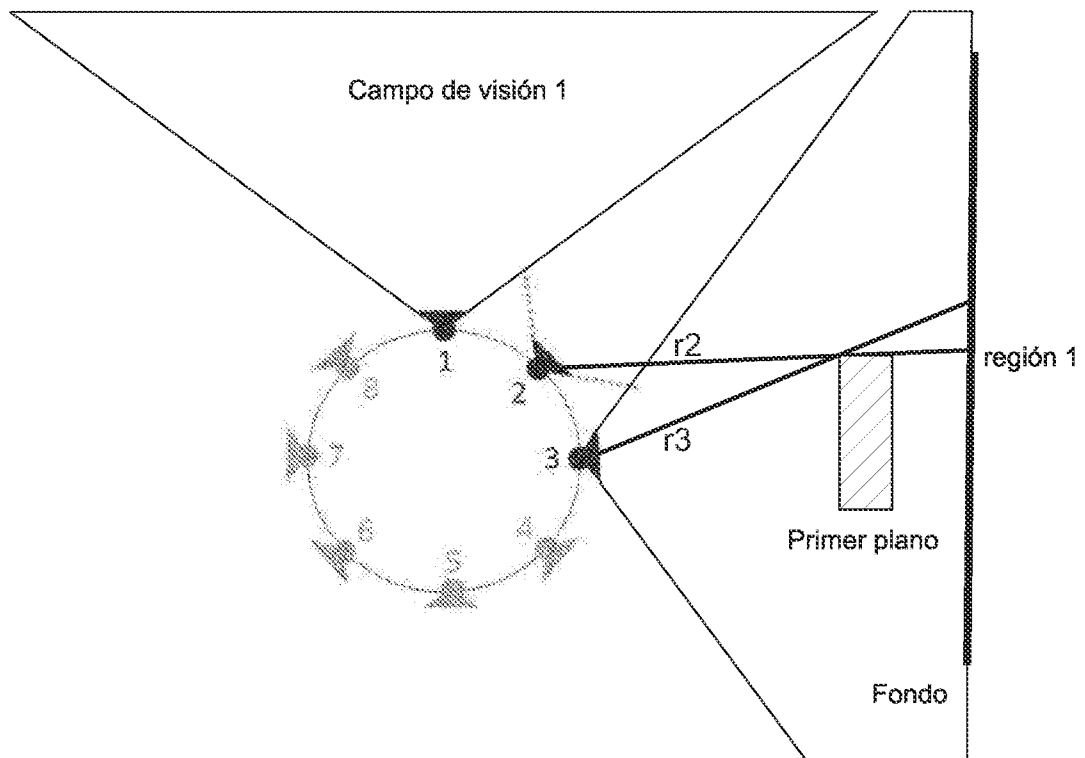


FIG. 2

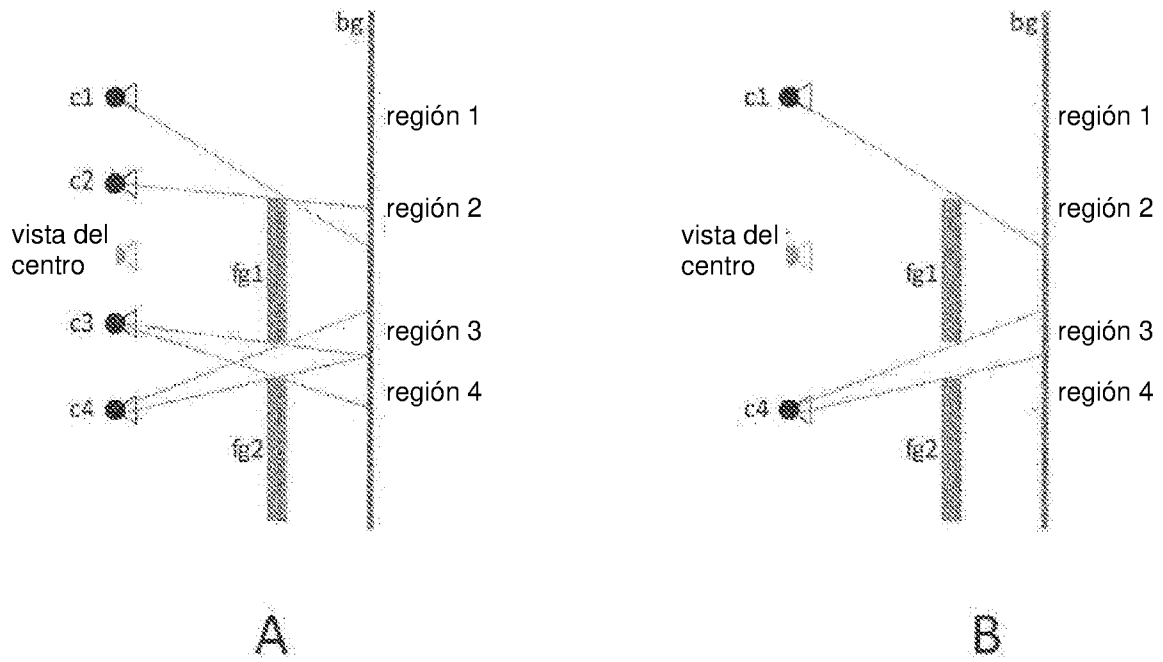
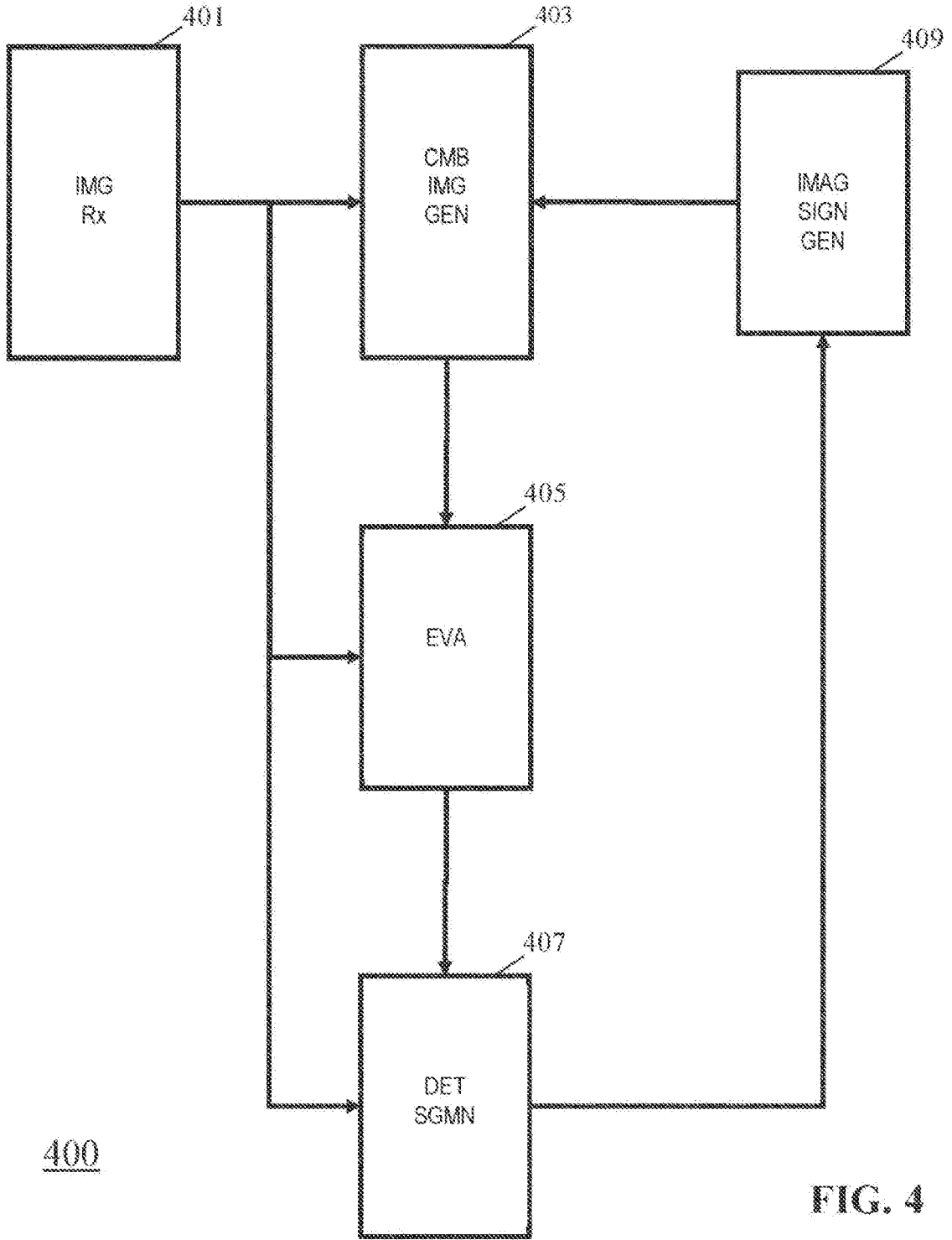
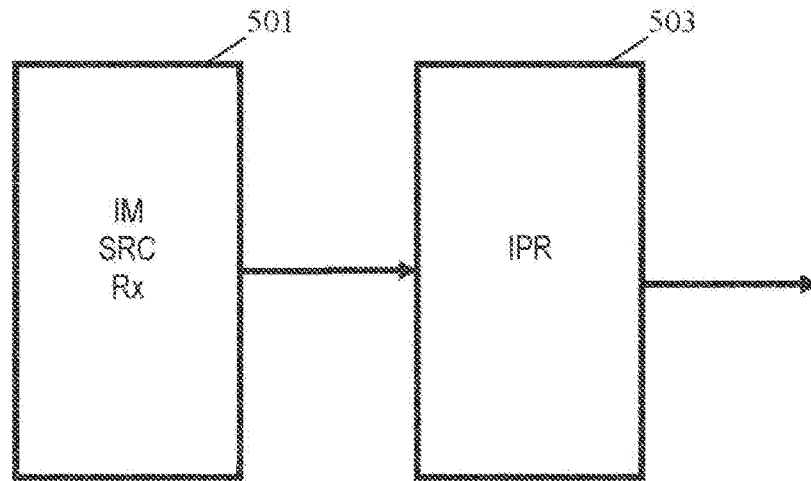


FIG. 3





500

FIG. 5

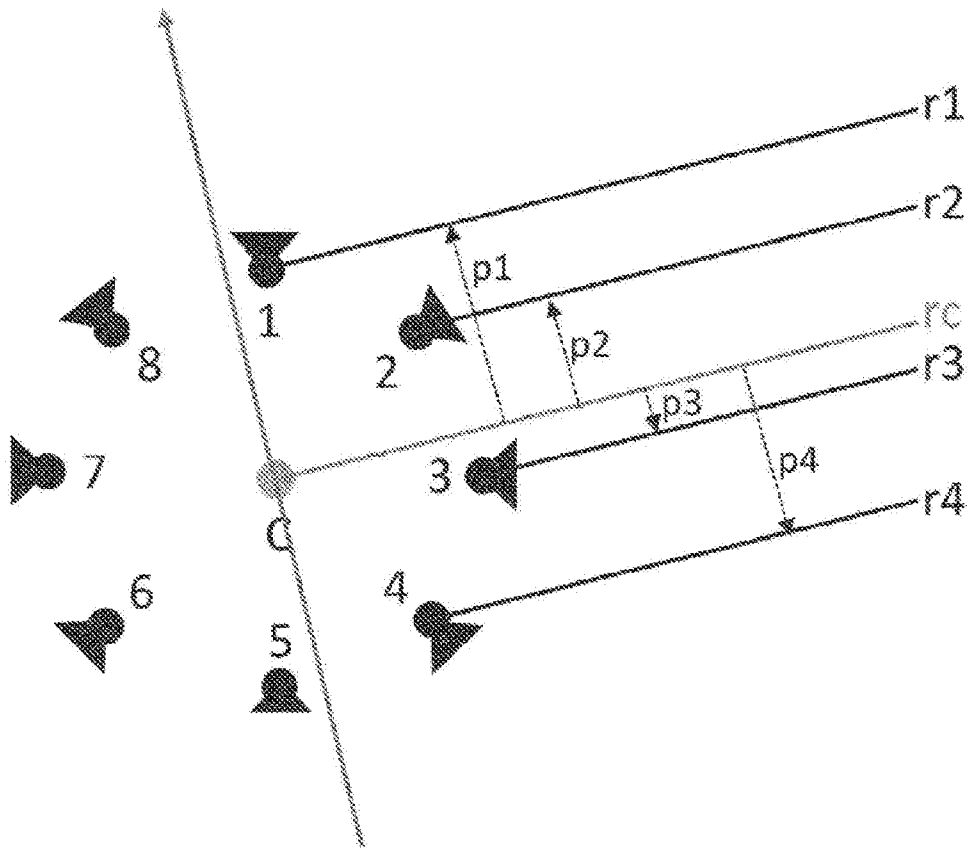


FIG. 6

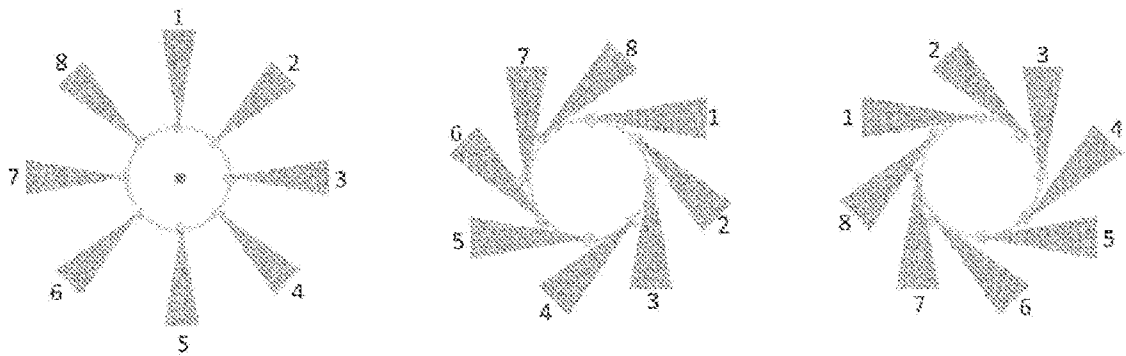


FIG. 7

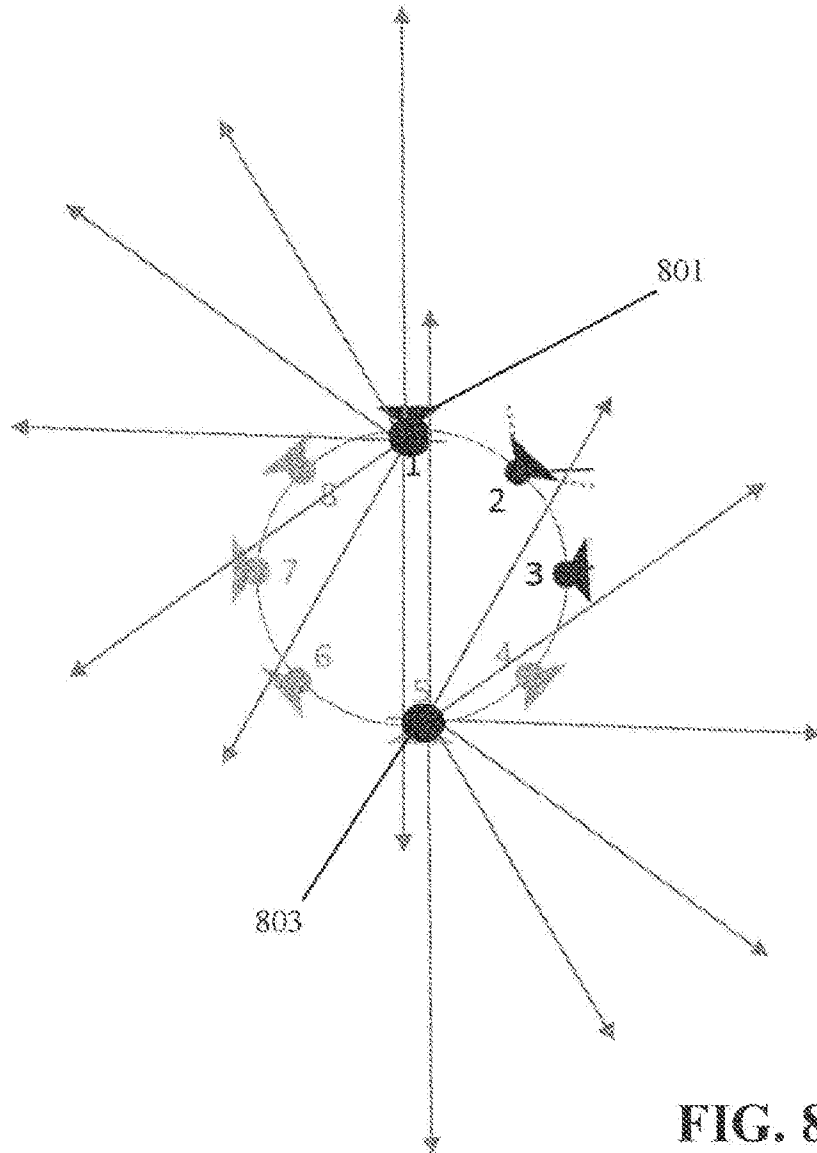


FIG. 8

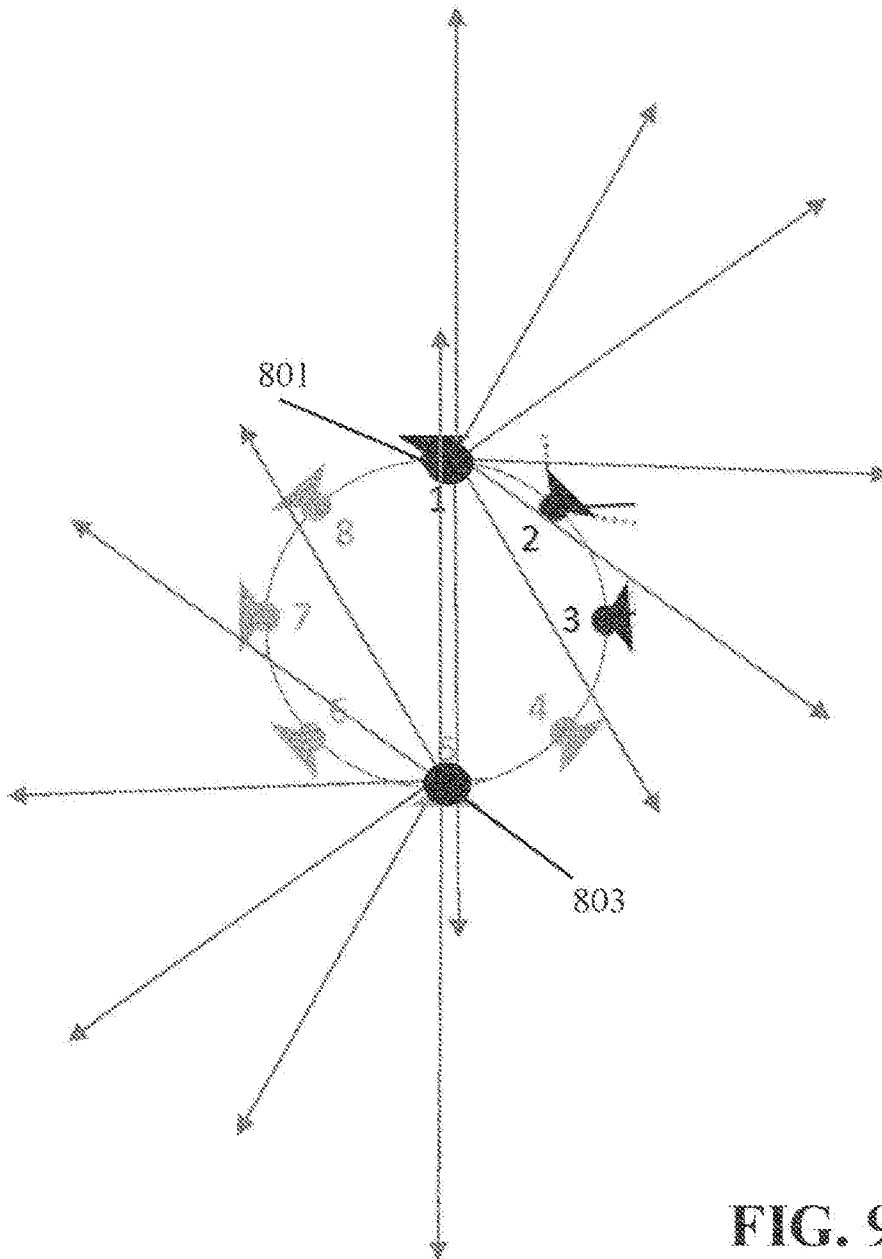


FIG. 9

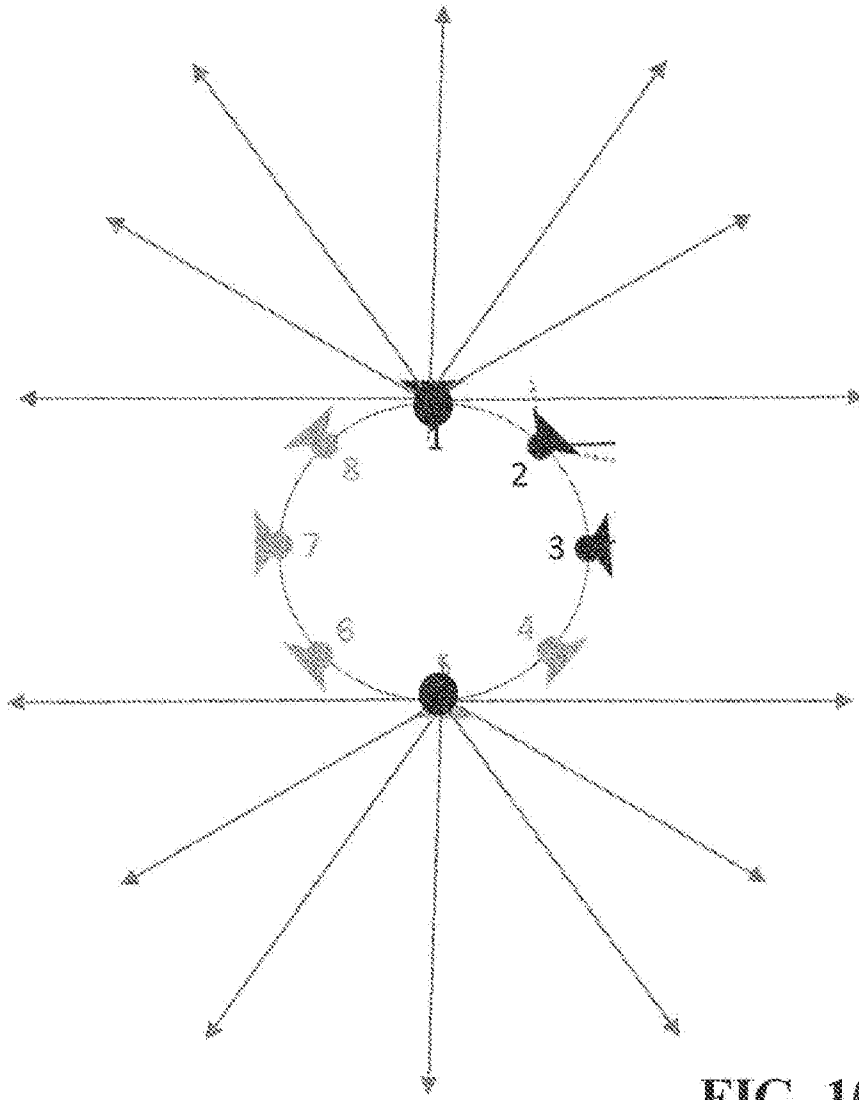


FIG. 10

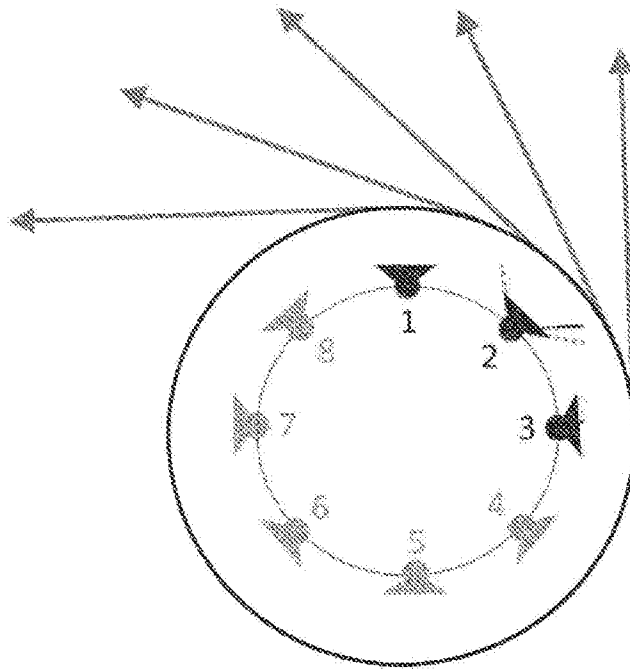


FIG. 11

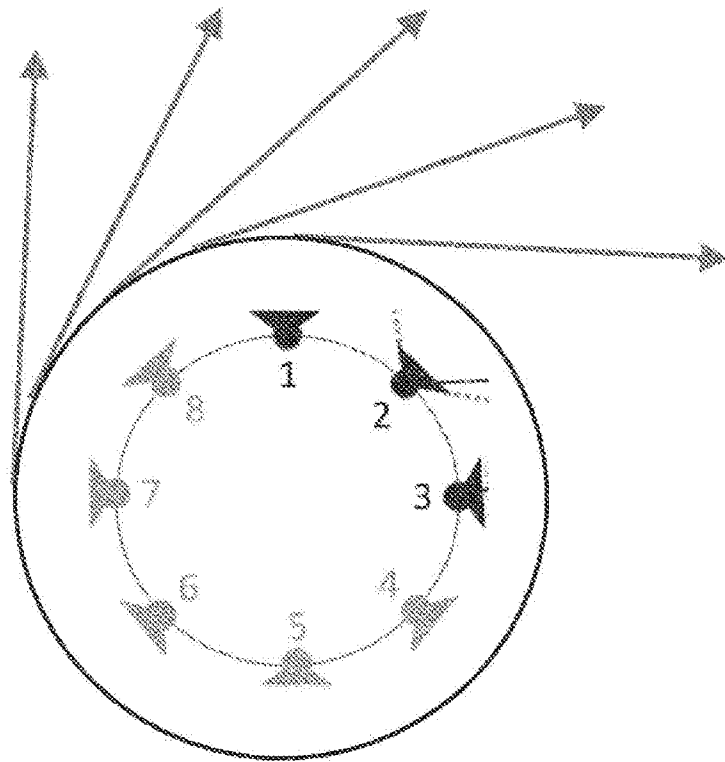


FIG. 12

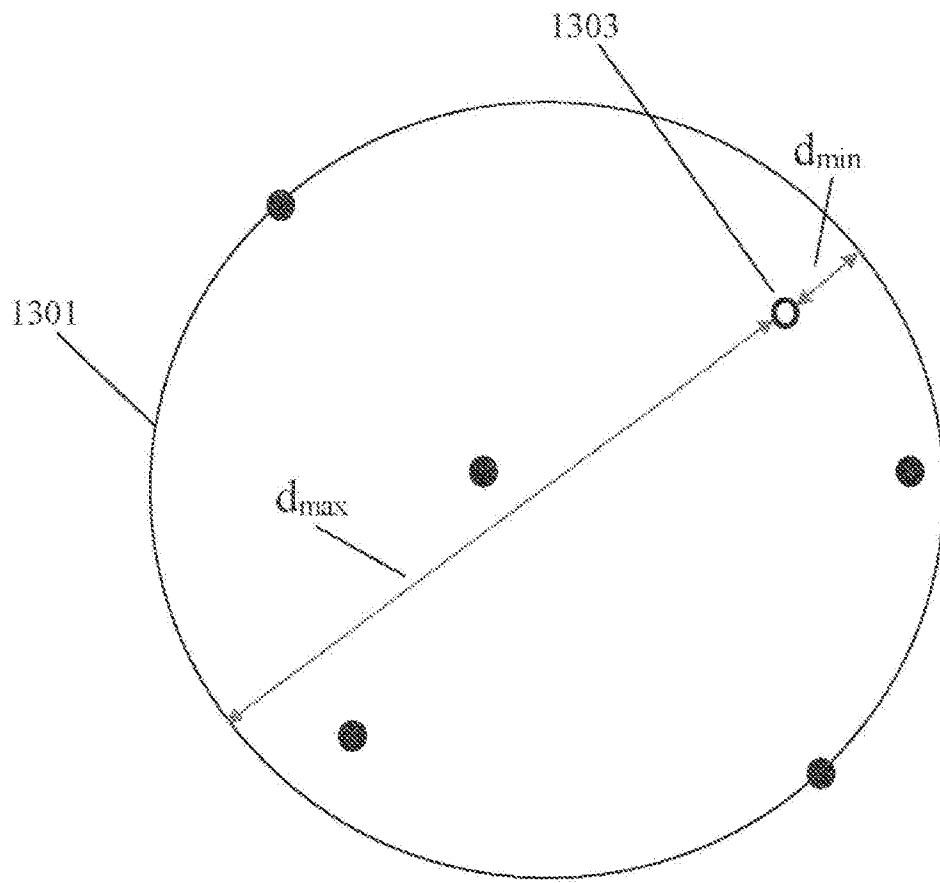


FIG. 13