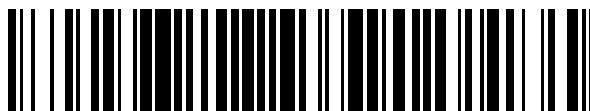


19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 389 401**

51 Int. Cl.:
H04N 7/15 (2006.01)
G06T 7/00 (2006.01)
H04N 13/02 (2006.01)
H04N 13/00 (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

- 96 Número de solicitud europea: **09765408 .1**
96 Fecha de presentación: **17.06.2009**
97 Número de publicación de la solicitud: **2299726**
97 Fecha de publicación de la solicitud: **23.03.2011**

54 Título: **Método, aparato y sistema de comunicación a través de vídeo**

30 Prioridad:
17.06.2008 CN 200810127007
29.08.2008 CN 200810210178

45 Fecha de publicación de la mención BOPI:
25.10.2012

45 Fecha de la publicación del folleto de la patente:
25.10.2012

73 Titular/es:
HUAWEI DEVICE CO., LTD. (100.0%)
Building B2 Huawei Industrial Base Bantian
Longgang District Shenzhen
Guangdong 518129, CN

72 Inventor/es:
FANG, PING;
LIU, CHEN y
LIU, YUAN

74 Agente/Representante:
LEHMANN NOVO, Isabel

ES 2 389 401 T3

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

DESCRIPCIÓN

Método, aparato y sistema de comunicación a través de vídeo

5 CAMPO DE LA TECNOLOGÍA

La presente invención se refiere al campo de la comunicación a través de vídeo y más en particular, a un método, un dispositivo y un sistema de comunicación a través de vídeo que utiliza el método de comunicación a través de vídeo.

10 ANTECEDENTES DE LA INVENCION

15 Con el desarrollo en curso de la tecnología de las comunicaciones, la tecnología de las comunicaciones a través de vídeo está ya ampliamente utilizada. Por ejemplo, la tecnología de comunicaciones a través de vídeo se utiliza en telefonía de vídeo, videoconferencia y funciones similares. Actualmente, una imagen en 2D convencional o vídeo se utiliza principalmente en varias aplicaciones de comunicaciones a través de vídeo.

20 Actualmente, para la extracción de objetivo de un contenido de imagen, se adopta principalmente un método de inserción croma adoptado para extraer un objeto en primer plano en un vídeo a través de la segmentación de colores. En la comunicación a través de vídeo, el objeto del primer plano extraído se sintetiza con otros vídeos distantes para mejorar el sentido de realidad. Por ejemplo, una persona (un objeto de primer plano) en el vídeo se sintetiza con notas de diapositivas a distancia. Sin embargo, el método de inserción croma tiene los inconvenientes siguientes.

- 25 1. El método de inserción croma requiere que un fondo del vídeo a segmentar sea de color azul, verde u otros colores simples para realizar la segmentación del objeto de primer plano desde el fondo y requiere que el color en el fondo no aparezca en el primer plano. Los requisitos de colores para el fondo y el primer plano en el método son estrictos, lo que causa una incomodidad operativa en el uso.
- 30 2. El método de inserción croma sólo puede diferenciar el objeto del primer plano respecto al fondo y el contenido de la escena no se puede dividir en más capas, por lo que no se puede conseguir la sustitución de una parte de objetos en el primer plano. Por ejemplo, en una escena de conferencia, podría existir un escritorio enfrente de una persona. Si un escritorio de otra parte se sustituye por un escritorio local, se puede mejorar el sentido de realidad. Sin embargo, el método de inserción croma no puede sustituir el escritorio de la otra parte con un escritorio local, por lo que no se puede mejorar todavía más el sentido de realidad.
- 35 3. La tecnología solamente realiza la sustitución de un contenido de vídeo en 2D y un usuario es incapaz de experimentar la profundidad de la escena, por lo que esta tecnología carece del sentido de realidad.

40 La imagen en 2D o vídeo sólo puede presentar un contenido de escenario, pero no puede reflejar la información de profundidad tal como las distancias y posiciones en el escenario.

45 Se utilizan seres humanos para observar el mundo exterior con dos ojos. Debido al paralaje de dos ojos, el escenario observado tiene una buena percepción de la distancia y de la posición y la sensación de estereo del escenario se puede presentar. En la tecnología de vídeo en estereo, el contenido de la escena que se muestra a un ojo izquierdo y a un ojo derecho de una persona son ligeramente distintas basándose en un principio de paralaje de dos ojos, con el fin de permitir a la persona obtener una sensación de profundidad y capas de la escena.

50 Con el fin de mejorar el sentido de realidad de la comunicación a través de vídeo, en la técnica anterior, se adopta una tecnología de vídeo en estereo y se realiza una decoración especial para escenas de ambas partes de la comunicación, de modo que los usuarios sientan que las dos partes de la comunicación están en la misma escena, con el fin de mejorar el sentido de realidad. Por ejemplo, el entorno de interiores de las dos partes de la comunicación está dispuesto en la misma escena. Por lo tanto, durante la comunicación, le parece al usuario que la otra parte en el vídeo está en la misma escena. Sin embargo, un alcance de aplicación del método está limitado por la disposición del entorno de las dos partes.

55 En la técnica anterior, se proporciona, además, un sistema de presentación interactivo. El sistema incluye principalmente una cámara de infrarrojos móvil y una unidad de reconocimiento de órdenes y una unidad de síntesis conectada a la cámara de infrarrojos. El sistema de presentación de la técnica anterior fotografía primero una imagen de un objeto de vídeo utilizando la cámara de infrarrojos móvil, con el fin de adquirir información del objeto de vídeo. Posteriormente, la unidad de reconocimiento de órdenes convierte la información en una señal de orden de salida y envía la señal de orden de salida a la unidad de síntesis. La información desde el objeto de vídeo puede incluir una imagen de infrarrojos de reconocimiento de un gesto del objeto de vídeo fotografiado por la cámara de infrarrojos o voz recibida del objeto de vídeo. Por último, la unidad de síntesis sintetiza la imagen del objeto de vídeo fotografiada por la cámara de infrarrojos y una imagen de un material de presentación, controla las posiciones del objeto de vídeo y controla la visualización en pantalla del material de presentación en función de la señal de orden de salida del sistema de reconocimiento de órdenes.

65

Sin embargo, el sistema de presentación dado a conocer en la técnica anterior soporta solamente un modo de visualización de vídeo en 2D, pero no soporta el modo de visualización de vídeo en 3D.

5 El documento WO02/0269272 A2 da a conocer un sistema para insertar imágenes de vídeo de seres humanos, animales u otros seres vivos o formas de vida y cualquier ropa u objetos que lleven, en un entorno virtual. El modelo en 3D se extrae en función de las imágenes registradas por más de una cámara. Los puntos de visualización se registran por múltiples cámaras, en donde cualquier punto de visualización se registra por una cámara.

10 El documento IZQUIERDO E ET AL: "Presentación basada en imágenes y creación de modelos en 3D: un marco de trabajo completo", SIGNAL PROCESSING IMAGE COMMUNICATION, ELSEVIER SCIENCE PUBLISHERS, AMSTERDAM, NODO LOCUTOR, vol. 15, nº 10, 1 agosto 2000 (01-08-2000), páginas 817-858, da a conocer una síntesis de múltiples puntos de visualización. Múltiples imágenes se adquieren por múltiples cámaras. El modelo en 3D se moviliza mediante análisis de múltiples puntos de visualización. De este modo, cualquier punto de visualización se adquiere de forma independiente.

15 El documento ZITNICK C L ET AL: "Interpolación de visualizaciones de vídeo de alta calidad utilizando una representación en capas", ACM TRANSACTIONS ON GRAPHICS: TOG, ACM, US, vol. 23, nº 3, 8 agosto 2004 (08-08-2004), páginas 600-608, se refiere a la interpolación de vistas de vídeo. Las tramas desde dos cámaras diferentes se interpolan para una sola imagen y se registran múltiples puntos de visualización.

20 El documento EP 1 796 047 A1 se refiere a un sistema para crear flujos de datos en 2D o en 3D desde un vídeo con objetos móviles.

25 El documento CN 1 275 871 A se refiere a un sistema de comunicación de imágenes-vídeo tomados con múltiples cámaras de objeto de vídeo.

SUMARIO DE LA INVENCION

30 La presente invención se refiere a un método, aparato y sistema de comunicación a través de vídeo, que no están limitados por la disposición del entorno de dos partes de comunicación, con el fin de mejorar el sentido de realidad para las dos partes de la comunicación durante la comunicación y para visualizar un contenido de escena en un modo de vídeo en 3D.

35 La presente invención da a conocer un método de procesamiento de vídeo, que incluye las etapas siguientes:

Imágenes en color/escala de grises de al menos dos puntos de visualización de un objeto de vídeo a presentar y una imagen en color/escala de grises de al menos un punto de visualización de al menos un material de presentación se adquieren, en donde el material de presentación es un contenido a presentar que se genera por un ordenador.

40 Las imágenes en color/escala de grises de los al menos dos puntos de visualización del objeto de vídeo y la imagen en color/escala de grises de al menos un punto de visualización del al menos un material de presentación se sintetizan para adquirir una imagen de presentación,

45 en donde la adquisición de las imágenes en color/escala de grises de los al menos dos puntos de visualización del objeto de vídeo a presentar comprende:

50 la adquisición, utilizando una sola cámara que adquiere información de profundidad e imagen en color/escala de grises al mismo tiempo, de información de profundidad en un punto de visualización y al menos una imagen en color/escala de grises del mismo punto de visualización de una escena en la que se encuentra el objeto de vídeo;

la realización de segmentación de vídeo en la imagen en color/escala de grises del mismo punto de visualización en función de la información de profundidad del punto de visualización, con el fin de adquirir el objeto de vídeo;

55 la generación de una imagen en color/escala de grises de al menos otro punto de visualización del objeto de vídeo en función de la información de profundidad del punto de visualización y de la imagen en color/escala de grises del mismo punto de visualización.

60 La presente invención da a conocer un dispositivo de procesamiento de vídeo, que incluye una unidad de adquisición y procesamiento de imagen.

65 La unidad de adquisición y procesamiento de imagen está configurada para adquirir imágenes en color/escala de grises de al menos dos puntos de visualización de un objeto de vídeo a presentar y una imagen en color/escala de grises de al menos un punto de visualización de al menos un material de presentación y para sintetizar las imágenes en color/escala de grises de los al menos dos puntos de visualización del objeto de vídeo y la imagen en color/escala de grises del al menos un punto de visualización del al menos un material de presentación para adquirir una imagen de presentación;

en donde la unidad de adquisición y procesamiento de imagen comprende:

5 un módulo de adquisición de imagen, configurado para adquirir, utilizando una sola cámara que puede adquirir información de profundidad y la imagen en color/escala de grises al mismo tiempo, la información de profundidad en un punto de visualización y al menos una imagen en color/escala de grises del mismo punto de visualización de una escena en la que se encuentra el objeto de vídeo;

10 un módulo de segmentación de vídeo, configurado para realizar una segmentación de vídeo en la imagen en color/escala de grises del mismo punto de visualización en función de la información de profundidad del punto de visualización para adquirir el objeto de vídeo;

15 un módulo de generación de imagen de múltiples puntos de visualización, configurado para generar una imagen en color/escala de grises de al menos otro punto de visualización del objeto de vídeo en función de la información de profundidad del punto de visualización y de la imagen en color/escala de grises del mismo punto de visualización y

un módulo de síntesis, configurado para sintetizar imágenes en color/escala de grises de los al menos dos puntos de visualización del objeto de vídeo y la imagen en color/escala de grises del al menos un punto de visualización del al menos un material de presentación para adquirir la imagen de presentación.

20 En las formas de realización de la presente invención, según se describe en las soluciones técnicas, durante la comunicación a través de vídeo, una imagen visualizada, al nivel local, necesita obtenerse sintetizando un contenido de fondo local y un contenido de objeto distante, de modo que un usuario pueda ver un fondo en la imagen exactamente el mismo que una escena en la que está el usuario y parece que dos partes de la comunicación están en el mismo entorno, con el fin de mejorar el sentido de realidad de los usuarios durante la comunicación. Además, en las formas de
 25 realización, las escenas de las dos partes de la comunicación no necesitan una disposición especial, a los entornos de las dos partes de la comunicación se les permite ser diferentes y el fondo tampoco necesita cambiarse a un color único, por lo que la puesta en práctica de la forma de realización de la presente invención no está limitada por los entornos de las dos partes de comunicación y se puede mejorar el sentido de realidad durante la comunicación. En las formas de
 30 realización de la presente invención, se genera una imagen de presentación adquiriendo imágenes en color/escala de grises de múltiples puntos de visualización de un objeto de vídeo y una imagen en color/escala de grises de un material de presentación y la imagen de presentación soporta un modo de visualización en 3D, de modo que visualizando la imagen de presentación se está en el modo en 3D y se resuelve así el problema de que solamente la presentación de vídeo en 2D esté soportada en la técnica anterior, con lo que se consigue la presentación de vídeo en 3D.

35 BREVE DESCRIPCIÓN DE LOS DIBUJOS

La Figura 1 es un diagrama de flujo de un método de preprocesamiento de vídeo según una primera forma de realización de la presente invención;

40 La Figura 2 es un diagrama de bloques de un dispositivo de preprocesamiento de vídeo según la primera forma de realización de la presente invención;

La Figura 3 es un diagrama de flujo de un método de recepción de vídeo según la primera forma de realización de la presente invención;

45 La Figura 4 es un diagrama de bloques de un dispositivo de recepción de vídeo según la primera forma de realización de la presente invención;

50 La Figura 5 es un diagrama de principio de un aparato de comunicación a través de vídeo según la primera forma de realización de la presente invención;

La Figura 6 es un diagrama de principio de un aparato de comunicación a través de vídeo según una segunda forma de realización de la presente invención;

55 La Figura 7 es un diagrama de principio de una cámara estéreo según la segunda forma de realización de la presente invención;

La Figura 8 es una vista esquemática de un proceso para sintetizar un contenido de escena según la segunda forma de realización de la presente invención;

60 Las Figuras 9A y 9B son un diagrama de flujo de comunicación a través de vídeo según una tercera forma de realización de la presente invención;

65 La Figura 10 es una vista estructural de un sistema de comunicación a través de vídeo según la tercera forma de realización de la presente invención;

La Figura 11 es un diagrama de flujo de un método de preprocesamiento de vídeo según una cuarta forma de realización del método de la presente invención;

5 La Figura 12 es un diagrama de flujo de un método para enviar una presentación de vídeo según una quinta forma de realización del método de la presente invención;

La Figura 13 es una vista de principio esquemática de una cámara para adquirir información de profundidad según la quinta forma de realización del método de la presente invención;

10 La Figura 14 es una vista estructural esquemática de una cámara para adquirir información de profundidad y una imagen en color/escala de grises según la quinta forma de realización del método de la presente invención;

La Figura 15 es una vista esquemática cuando se produce un bloqueo según la quinta forma de realización del método de la presente invención;

15 La Figura 16 es una vista esquemática de cuando se produce un agujero según una quinta forma de realización de la presente invención;

20 La Figura 17 es un diagrama de flujo de un método para enviar una presentación de vídeo según una sexta forma de realización del método de la presente invención;

La Figura 18 es un diagrama de flujo de un método para la recepción de una presentación de vídeo según la sexta forma de realización del método de la presente invención;

25 La Figura 19 es una vista esquemática de un dispositivo de presentación de vídeo según una cuarta forma de realización del dispositivo de la presente invención,

La Figura 20 es una vista esquemática de un dispositivo para enviar una presentación de vídeo según una quinta forma de realización del dispositivo de la presente invención;

30 La Figura 21 es una vista esquemática de un dispositivo para la recepción de presentación de vídeo según la quinta forma de realización del dispositivo de la presente invención;

35 La Figura 22 es una vista esquemática de un dispositivo para enviar una presentación de vídeo según una sexta forma de realización del dispositivo de la presente invención y

La Figura 23 es una vista esquemática de un sistema de comunicación según una forma de realización del sistema de la presente invención.

40 DESCRIPCIÓN DETALLADA DE LAS FORMAS DE REALIZACIÓN

En las formas de realización de la presente invención, un contenido del fondo local y un contenido del objeto distante están sintetizados en una imagen a visualizar, de modo que dos partes de la comunicación no necesiten realizar ninguna disposición especial para una escena con el fin de hacer que la escena en la imagen sea la misma que en la que se encuentra una de las dos partes, de forma que se mejore el sentido de realidad en un proceso de comunicación y para visualizar el contenido de la escena en un modo de vídeo en 3D. Un método, dispositivo y aparato de comunicación a través de vídeo, según las formas de realización de la presente invención, se ilustran en detalle a continuación haciendo referencia a los dibujos adjuntos.

50 Forma de realización uno

En esta forma de realización, la presente invención da a conocer un método de presentación de vídeo. Según se ilustra en la Figura 1, el método de preprocesamiento de vídeo incluye las etapas siguientes.

55 En la etapa 101, los contenidos de escenas locales y sus valores de profundidad se adquieren utilizando una cámara de profundidad o una cámara en estéreo.

60 En la etapa 102, los contenidos de escenas locales se pueden dividir en más de dos capas por intermedio de los valores de profundidad de los contenidos de escenas locales, de modo que una capa en donde existe un contenido de objeto local se segmenta, es decir, el contenido del objeto local se segmenta desde los contenidos de escenas locales.

En la etapa 103, el contenido de objeto local segmentado un valor de profundidad, correspondiente al contenido de objeto local, se envían a un extremo distante, que suele ser un extremo opuesto de la comunicación.

65 En esta forma de realización, el preprocesamiento de la imagen se realiza principalmente por intermedio de las etapas 101 y 102. En la etapa 103, se envía el contenido preprocesado, de modo que se pueda omitir la etapa 103.

En correspondencia con el método de preprocesamiento de vídeo, en esta forma de realización, la presente invención da a conocer, además, un dispositivo de preprocesamiento de vídeo. Según se ilustra en la Figura 2, el dispositivo de preprocesamiento de vídeo incluye un módulo de adquisición de información 21, un módulo de segmentación 22 y un módulo de envío 23.

5 El módulo de adquisición de información 21 está configurado para adquirir contenidos de escenas locales y sus valores de profundidad. El módulo de adquisición de información se puede poner en práctica como una cámara de profundidad o una cámara de estéreo. La cámara de profundidad adquiere una profundidad de la imagen por intermedio de la tecnología de infrarrojos y la cámara en estéreo adquiere la profundidad de la imagen utilizando cámaras duales.
10 Después de obtener los valores de profundidad de los contenidos de escenas locales, dichos contenidos de escenas locales se pueden dividir en más de dos capas. El módulo de segmentación 22 está configurado para segmentar un contenido de objeto local a partir de los contenidos de escenas locales en función de los valores de profundidad de los contenidos de escenas locales. El módulo de envío 23 está configurado para enviar el contenido objeto local y su valor de profundidad a un extremo distante.

15 El dispositivo de preprocesamiento de vídeo completa el preprocesamiento de vídeo principalmente utilizando el módulo de adquisición de información 21 y el módulo de segmentación 22. El módulo de envío 23 se puede omitir.

20 Con el fin de realizar la comunicación a través de vídeo, en esta forma de realización, la presente invención da a conocer, además, un método de recepción de vídeo correspondiente al método de preprocesamiento de vídeo. Según se ilustra en la Figura 3, el método de recepción de vídeo incluye las etapas siguientes.

En la etapa 301 se recibe un contenido objeto y su valor de profundidad enviados por un extremo distante.

25 En la etapa 302, se adquiere un contenido de fondo local y un valor de profundidad del contenido de fondo.

En la etapa 303, en función de una diferencia entre los valores de profundidad, se determina una relación de bloqueo entre el contenido de fondo local y el contenido de objeto distante. En condiciones normales, un pixel que tiene un valor de profundidad más pequeño bloquea un pixel que tenga un valor de profundidad mayor. Por lo tanto, el contenido de objeto distante y el contenido de fondo local se pueden sintetizar en un contenido de escena en función de la relación de valores de profundidad.
30

En correspondencia con el método de recepción de vídeo, en esta forma de realización, la presente invención da a conocer, además, un dispositivo de recepción de vídeo. Según se ilustra en la Figura 4, el dispositivo de recepción de vídeo incluye un módulo de interfaz de transmisión 41, un módulo de extracción 42 y un módulo de síntesis 43.
35

El módulo de interfaz de transmisión 41 está configurado para recibir un contenido de objeto y su valor de profundidad enviado por un extremo distante. El módulo de extracción 42 está configurado para adquirir un contenido de fondo local y su valor de profundidad. El módulo de síntesis 43 está configurado para sintetizar el contenido de objeto distante y el contenido de fondo local en un contenido de escena en función de una relación de valores de profundidad. En condiciones normales, un pixel que tenga un valor de profundidad más pequeño bloquea un pixel que tenga un valor de profundidad mayor. Por último, el contenido de escena sintetizado se visualiza en un aparato tal como un monitor de presentación visual.
40

45 Según se ilustra en la Figura 5, en una forma de realización, la presente invención da a conocer, además, un aparato de comunicación a través de vídeo, que incluye concretamente un módulo de adquisición de información 21, un módulo de segmentación 22, un módulo de interfaz de transmisión 41, un módulo de extracción 42 y un módulo de síntesis 43.

50 El módulo de adquisición de información 21 está configurado para adquirir contenidos de escenas locales y sus valores de profundidad. El módulo de adquisición de información 21 se puede poner en práctica como una cámara de profundidad o una cámara en estéreo. La cámara de profundidad adquiere una profundidad de la imagen por intermedio de la tecnología de infrarrojos y la cámara en estéreo adquiere la profundidad de la imagen utilizando cámaras duales. El módulo de segmentación 22 está configurado para segmentar un contenido de objeto local a partir de los contenidos de escenas locales en función de los valores de profundidad de los contenidos de escenas locales. El módulo de interfaz de transmisión 41 está configurado para enviar el contenido de objeto local y su valor de profundidad a un extremo distante.
55

60 El módulo de interfaz de transmisión 41 está configurado, además, para recibir un contenido de objeto y su valor de profundidad enviados por el extremo distante. El módulo de extracción 42 está configurado para adquirir un contenido de fondo local y su valor de profundidad. El módulo de síntesis 43 está configurado para sintetizar el contenido de objeto distante y el contenido de fondo local en un contenido de escena en función de una relación de valores de profundidad. En condiciones normales, un pixel que tenga un valor de profundidad más pequeño bloquea un pixel que tenga un valor de profundidad mayor. Por último, el contenido de la escena sintetizado se visualiza en un módulo de presentación visual.

El contenido de fondo local puede ser el contenido residual después de que el módulo de segmentación 22 realice la segmentación del contenido de objeto local o un contenido de fondo local y su valor de profundidad opuesto al objeto local se pueden adquirir utilizando otra cámara.

5 En esta forma de realización, si se realiza la comunicación entre el dispositivo de preprocesamiento de vídeo y el dispositivo de recepción de vídeo, por ejemplo, ambos dispositivos acceden a la misma red para formar un sistema de comunicación a través de vídeo, un extremo emisor del sistema incluye el dispositivo de preprocesamiento de vídeo representado en la Figura 2 y un extremo receptor incluye el dispositivo de recepción de vídeo representado en la Figura 4.

10 Forma de realización dos

15 En esta forma de realización, la presente invención da a conocer un aparato de comunicación a través de vídeo. El aparato envía un contenido de objeto local en los contenidos de escenas locales y un valor de profundidad correspondiente al contenido de objeto local a un aparato opuesto. Después de recibir el contenido de objeto local, el aparato opuesto sintetiza el contenido de objeto local y el contenido de fondo opuesto en una escena y visualiza la escena a un usuario en el extremo opuesto. Por lo tanto, está garantizado que la escena vista por el usuario opuesto es exactamente la misma que la escena en la que se encuentra el usuario opuesto, con el fin de proporcionar el sentido de presencia y el sentido de realidad. Después de recibir el contenido de objeto distante, el aparato de comunicación a través de vídeo local sintetiza el contenido de objeto distante y el contenido de fondo local en una escena y visualiza la escena para el usuario local, con el fin de mejorar el sentido de presencia y el sentido de realidad del usuario local en el proceso de comunicación.

25 Según se ilustra en la Figura 6, el aparato de comunicación a través de vídeo incluye principalmente un módulo de adquisición de información 61, un módulo de segmentación 62, un módulo de codificación 63, un módulo de interfaz de transmisión 64, un módulo de decodificación 65, un módulo de síntesis 66 y un módulo de presentación visual 67.

30 El módulo de adquisición de información 61 está configurado para fotografiar los contenidos de escenas locales y para calcular los valores de profundidad correspondientes a los contenidos de escenas locales o para adquirir directamente los valores de profundidad correspondientes a los contenidos de escenas locales. El módulo de segmentación 62 está configurado para segmentar un contenido de objeto local a partir de los contenidos de escenas locales en función del valor de profundidad. El módulo de codificación 63 está configurado para codificar el contenido de objeto local segmentado y un valor de profundidad correspondiente al contenido de objeto local. El módulo de interfaz de transmisión 64 está configurado para enviar el contenido de objeto local y su valor de profundidad o para recibir el contenido de objeto y su valor de profundidad enviado por el extremo distante. El módulo de decodificación 65 está configurado para decodificar el contenido de objeto distante recibido y su valor de profundidad. El módulo de síntesis 66 está configurado para sintetizar el contenido de objeto distante decodificado y un contenido de fondo local y para generar una vista en estéreo en función del valor de profundidad correspondiente, en donde el contenido de fondo local puede ser un contenido residual en los contenidos de escenas locales después de que se segmente el contenido de objeto local o un contenido de escena opuesto al objeto local fotografiado utilizando otro grupo de cámaras. El modo de presentación visual 67 está configurado para visualizar la imagen sintetizada, que puede ser un aparato de presentación visual en estéreo o un aparato de presentación visual en 2D normal. Si el módulo de visualización 67 es el aparato de presentación visual en estéreo, necesita reconstruirse una imagen en 2D de otro punto de visualización.

45 A continuación, cada módulo en el aparato de comunicación a través de vídeo, según esta forma de realización, se ilustra en detalle, respectivamente.

50 El módulo de adquisición de información 61 se puede poner en práctica en dos modos. En el modo 1, los contenidos de escenas locales y sus valores de profundidad se obtienen al mismo tiempo utilizando una cámara de profundidad. En el modo 2, los contenidos de escenas locales se fotografían utilizando más de dos cámaras y se obtiene un valor de profundidad correspondiente por intermedio de un método de coincidencia de imágenes en estéreo.

55 La cámara de profundidad es un nuevo tipo de cámara. La cámara de profundidad puede adquirir un valor de profundidad correspondiente a cada pixel en una imagen de colores rojo-verde-azul (RGB) al mismo tiempo cuando se fotografía la imagen en color. Actualmente, la cámara de profundidad adquiere principalmente un valor de profundidad de un objeto en una escena en un modo de infrarrojos.

60 En el método de obtención del valor de profundidad correspondiente a través de la coincidencia de imágenes en estéreo, se necesitan más de dos cámaras para fotografiar una escena durante la adquisición de imágenes, con el fin de obtener más de dos imágenes de la escena en ángulos diferentes. Mediante la coincidencia de las imágenes, se puede obtener un paralaje de la escena en imágenes diferentes. Según los parámetros internos y externos del paralaje y de las cámaras, el valor de profundidad correspondiente a cada pixel en la imagen se puede calcular en este momento. A continuación, se toma como ejemplo dos cámaras para ilustrar la adquisición del valor de profundidad en un modo de coincidencia de imágenes.

65

La Figura 7 es una vista esquemática de la creación de imágenes por dos cámaras paralelas situadas horizontalmente en donde O_1 y O_2 son centros ópticos de dos cámaras, respectivamente, una distancia entre ellas es B , una distancia entre un punto A y un punto vertical de la cámara es Z (es decir, una profundidad del punto A) y A_1 y A_2 son puntos de creación de imágenes del punto A en la dos cámaras, respectivamente.

5 Por intermedio de la similitud entre un triángulo $A_1 O_1 O_1'$ y un triángulo $A O_1 C$, se puede obtener que

$$\frac{A_1 O_1'}{O_1} = \frac{f}{Z}$$

$$\frac{A_2 O_2'}{O_2} = \frac{f}{Z}$$

Gracias a la similitud entre un triángulo $A_2 O_2 O_2'$ y un triángulo $A O_2 C$, se puede obtener que

10 Por lo tanto, se puede deducir que un paralaje de los dos puntos de creación de imágenes es $d = A_1 O_1' - A_2 O_2' = f * (O_1 - O_2) / Z = f * B / Z$.

Por lo tanto, se puede obtener que un valor de profundidad del punto A es $Z = f * B / d$.

15 Puesto que la longitud focal f es conocida y B se puede medir, se puede calcular d por intermedio del método de coincidencia de imágenes. Por lo tanto, el valor de profundidad correspondiente a cada punto de escena se puede adquirir utilizando dos cámaras.

20 La adquisición de la información de profundidad utilizando la cámara de estéreo incluye las etapas siguientes. Los puntos de creación de imágenes correspondientes a un determinado punto en la escena se encuentran en más de dos imágenes. Posteriormente, en función de las coordenadas del punto en más de dos imágenes, se calcula su valor de profundidad. El proceso para encontrar los puntos de creación de imágenes correspondientes al punto en la escena en imágenes diferentes se realiza por intermedio de la coincidencia de imágenes. La tecnología de coincidencia de imágenes actual incluye principalmente coincidencias basadas en ventanas, coincidencias basadas en características y programación dinámica.

25 Un algoritmo de coincidencias basado en la escala de grises se adopta en la coincidencia basada en ventanas y en la programación dinámica. El algoritmo basado en la escala de grises es para segmentar una de las imágenes en más de dos subzonas pequeñas, cuyos valores de escala de grises se utilizan como modelos para encontrar subzonas que tengan una distribución de valores de escala de grises más similar a las subzonas en otras imágenes. Si dos subzonas satisfacen los requisitos para similaridad de distribución de valores de escala de grises, se puede considerar que los puntos en las subzonas son coincidentes, es decir, los puntos de creación de imágenes de las dos subzonas son imágenes del mismo punto en la escena. En un proceso de coincidencias, la similaridad de dos zonas se suele medir por intermedio de una función de correlación.

30 En la coincidencia basada en características, las escalas de grises de la imagen no se utilizan directamente sino que la coincidencia se realiza utilizando las características derivadas de la información de escala de grises de las imágenes, de modo que la coincidencia basada en las características es más estable en comparación con la coincidencia utilizando información de cambio simple de brillo y escalas de grises. Las características de la coincidencia pueden ser características potenciales importantes que pueden describir la estructura en 3D de la escena, por ejemplo, un punto de intersección (un punto de esquina) entre bordes. En la coincidencia basada en características, un diagrama de información de profundidad dispersa se suele obtener primero y posteriormente, un diagrama de información de profundidad denso de la imagen se obtiene por intermedio de un método que, por ejemplo, utiliza un valor de interpolación.

35 El módulo de segmentación 62 segmenta la imagen en función de los contenidos de escenas locales y los valores de profundidad correspondientes a los contenidos de escenas locales para obtener el contenido de objeto local en la escena local. El módulo de segmentación 62 se puede poner en práctica con una unidad de búsqueda 621 y una unidad de segmentación 622. La unidad de búsqueda 621 está configurada para buscar zonas en donde el contenido de objeto local aparezca en el contenido de escena local. La unidad de segmentación 622 está configurada para realizar una extracción de contorno de borde exacta en las zonas del contenido de objeto local en los contenidos de escenas locales, con el fin de segmentar en el contenido de objeto local y otro contenido de fondo local.

40 En general, para las zonas en donde el contenido de objeto local aparece en la escena local, un usuario local establece un margen de valores de profundidad del contenido de objeto local después de estimar una posición del objeto local en relación con la cámara. En un procesamiento de vídeo posterior, la unidad de búsqueda busca zonas en donde el contenido de objeto aparece en el margen de valores de profundidad.

- Si el contenido de objeto local a buscar es una imagen de persona, una unidad de reconocimiento de caras 623 puede reconocer automáticamente una posición en la que aparece una imagen de cara a partir de los contenidos de escenas locales por intermedio de una tecnología de reconocimiento de caras existente. Posteriormente, la unidad de búsqueda 621 busca un valor de profundidad correspondiente a la posición de la imagen de cara en los valores de profundidad de los contenidos de escenas locales. A continuación, un margen del valor de profundidad del contenido de objeto local se determina en función del valor de profundidad encontrado y se determina una zona del contenido de objeto local en el contenido de escena en función del margen del valor de profundidad. De este modo, un margen de profundidad de un objeto de persona que aparece en la escena, se determina de esta forma.
- 5
- 10 Cuando el valor de profundidad corresponde a la imagen de color, la zona de persona segmentada en función del valor de profundidad corresponde a la zona de persona en la imagen en color. El contenido de objeto local obtenido y su valor de profundidad de la imagen en color se envían al módulo de codificación 63. El módulo de codificación 63 codifica el contenido de objeto local y su valor de profundidad y los envía a un extremo distante por intermedio del módulo de interfaz de transmisión 64.
- 15
- 20 Cuando los contenidos de objetos locales, extraídos de la imagen en color, tienen tamaños diferentes, los contenidos de objetos locales necesitan ajustarse al mismo tamaño. Los contenidos de objetos locales se suelen ajustar a un tamaño que sea el mismo que el de los contenidos de escenas locales, con el fin de obtener una imagen a codificar que tenga el mismo tamaño desde cada trama, con lo que se facilita la operación de codificación. En el ajuste, el contenido de objeto local no está en escala sino solamente se cambia un tamaño de un 'lienzo' (soporte de imágenes) utilizado para el contenido de objeto local. Las zonas en blanco, que aparecen después del ajuste de tamaño, se pueden rellenar con un valor cero.
- 25
- 30 En esta forma de realización, el módulo de codificación 63 codifica el contenido de objeto local segmentado y su valor de profundidad. En comparación con un vídeo en 2D de canal único, el vídeo en estéreo tiene una cantidad de datos mucho mayor que el vídeo en estéreo del tipo de 'dos ojos' con dos canales de datos. El aumento de los datos de vídeo causa dificultades en la memorización y transmisión de datos. La codificación de vídeo en estéreo actual incluye también principalmente dos tipos, a saber, codificación basada en bloques y codificación basada en objetos. En la codificación de la imagen en estéreo, además de la eliminación de la redundancia de datos en dominios espaciales y dominios del tiempo por intermedio de la predicción intratrama y la predicción intertramas, ha de eliminarse también la redundancia de datos de dominios espaciales entre imágenes de múltiples canales. La estimación del paralaje y su compensación son tecnologías inserciones en la codificación de vídeo en estéreo, con el fin de eliminar la redundancia de dominios espaciales entre imágenes de múltiples canales. El elemento esencial de la estimación y compensación del paralaje es encontrar la correlación entre dos (o más de tres) imágenes. El contenido de la codificación de vídeo en estéreo incluye, en este caso, una imagen en color y un valor de profundidad correspondiente a la imagen de color y se puede adoptar la codificación en capas, es decir, la imagen en color se codifica de forma híbrida y se coloca en una capa base y un valor de profundidad se codifica de forma híbrida y se coloca en una capa de ampliación.
- 35
- 40 En esta forma de realización, el módulo de interfaz de transmisión 64 está configurado para enviar el contenido de objeto local codificado y su valor de profundidad, para recibir un contenido de objeto distante codificado y su valor de profundidad transmitido por un extremo distante y para enviar el contenido de objeto distante y el valor de profundidad al módulo de decodificación para su decodificación. En esta forma de realización, el módulo de interfaz de transmisión 64 puede ser varias interfaces inalámbricas o cableadas que pueden realizar la transmisión, por ejemplo, de una interfaz de banda ancha, una interfaz de Bluetooth, una interfaz de infrarrojos o una tecnología de acceso con la utilización de una red de comunicaciones móviles de teléfonos móviles. En esta forma de realización, el módulo de interfaz de transmisión sólo necesita transmitir el objeto local y su valor de profundidad en los contenidos de escenas locales, una cantidad de datos que es reducida en comparación con el contenido de escena local original, con lo que se reduce una tasa de ocupación de ancho de banda durante la transmisión de datos.
- 45
- 50 En esta forma de realización, pues de recibirse por el módulo de interfaz de transmisión 64 del aparato de comunicación a través de vídeo, el contenido del objeto distante y su valor de profundidad sólo se pueden visualizar después de su procesamiento.
- 55
- El módulo de decodificación 65 está configurado para decodificar los datos distantes recibidos, con el fin de obtener el contenido del objeto distante y el valor de profundidad correspondiente al contenido del objeto distante.
- 60
- 65 El módulo de síntesis 66 está configurado para sintetizar el contenido de objeto distante y el contenido de fondo local en función de los valores de profundidad, con el fin de obtener una imagen en color después de que se sinteticen el contenido del objeto distante y el fondo local y un valor de profundidad correspondiente a la imagen en color. El contenido de fondo local se extrae por el módulo de extracción 69. En el proceso de síntesis, se determina primero una relación de bloqueo en función del valor de profundidad del contenido de objeto distante y del valor de profundidad del contenido de fondo local y posteriormente, los contenidos de la imagen en color correspondiente se sintetizan en función de la relación de bloqueo. Cuando el módulo de presentación visual 67 es un aparato de visualización de estéreo en 3D, una imagen virtual en otro punto de visualización necesita reconstruirse todavía más en función de los contenidos sintetizados de la imagen en color y del valor de profundidad correspondiente a la imagen en color, de modo que, en esta forma de realización, la presente invención comprende, además, un módulo de reconstrucción de visualización 68, que

está configurado para realizar la reconstrucción de la visualización sobre el contenido de imagen sintetizado, con el fin de generar una imagen de punto de visualización virtual. La imagen de punto de visualización virtual y la imagen en color sintetizada constituyen una visualización en estéreo, que se envía al aparato de presentación visual de estéreo en 3D para realizar la presentación visual en estéreo.

5 Según se indica en la Figura 8, se proporciona un contenido de objeto distante (una persona), se ilustra una profundidad del contenido del objeto distante, se proporciona un contenido de fondo local (un árbol y un escritorio) adquirido localmente utilizando una cámara de profundidad y se ilustra una profundidad del contenido de fondo local. Posteriormente, se realiza la síntesis en función de una relación de profundidad con el fin de obtener una escena sintetizada. Cuando se obtienen las distancias del contenido de objeto distante y del contenido de fondo local en relación con la cámara, la persona distante se puede insertar entre el escritorio y el árbol locales.

Con el fin de hacer más natural la imagen sintetizada, necesita resolverse los problemas siguientes.

15 (1) Problema de escalamiento de un contenido del objeto distante: con el fin de sintetizar perfectamente una persona distante y el contenido del fondo local, una unidad de escalamiento 661 se necesita para ajustar una posición del contenido del objeto distante en relación con la cámara. En este momento, necesita ponerse en escala un tamaño del contenido del objeto distante. Cuando el contenido del objeto distante necesita llevarse a una distancia más próxima, es decir, cuando se disminuye el valor de profundidad, necesita ponerse en escala el contenido del objeto distante. Cuando el contenido del objeto distante está dispuesto a una distancia mayor, es decir, cuando se aumenta el valor de profundidad, el contenido del objeto distante necesita establecerse en escala descendente. Cuando el contenido del objeto distante es un objeto único, un margen del cambio de profundidad está limitado, de modo cuando se pone a escala la imagen, una relación de escalamiento en perspectiva se puede simplificar en un escalamiento lineal compatible con una profundidad de la imagen.

25 (2) Problema de bloqueo mutuo entre el contenido del objeto distante y el contenido del fondo local: cuando el contenido del objeto distante y el contenido del fondo local se sintetizan, el problema de bloqueo mutuo necesita considerarse con la unidad de síntesis 662. En función del valor de profundidad en escala del contenido del objeto distante y del valor de profundidad del contenido de fondo local, se determina una relación de bloqueo entre el contenido del objeto distante en escala y el contenido del fondo local. Cuando se solapan las posiciones horizontal y vertical de puntos de pixels, un punto de pixel que tenga un valor de profundidad más pequeño bloquea un punto que tenga un valor de profundidad mayor (una visualización cercana bloquea una visualización distante). En función de la relación de bloqueo entre el contenido del objeto distante escalado y el contenido del fondo local, el contenido del objeto distante escalado y el contenido del fondo local son sintetizados en un contenido de escena.

35 (3) Problema de relleno de agujeros: un agujero podría existir en el contenido de fondo local obtenido después de que se elimine el contenido del objeto local y el agujero podría existir todavía después de que se sinteticen el contenido del fondo local y el contenido del objeto distante. Se proporcionan dos soluciones:

40 En una primera solución, se utilizan otro grupo de cámaras para adquirir y fotografiar un contenido de una escena o puesta al objeto local, que suele ser un contenido de escena que ve una persona. Durante la síntesis, el contenido de la escena es directamente sintetizado con el contenido del objeto distante. Este modo tiene un efecto adecuado, es decir, un fondo que ve la persona se sintetiza con una persona distante. Cuando se utiliza directamente la escena opuesta, no existe el problema de relleno de agujeros. Sin embargo, un grupo de cámaras debe añadirse en cada extremo de la comunicación a través de vídeo.

En otra solución, se utiliza un contenido de fondo local residual después de que se elimine el contenido del objeto local y los posibles agujeros se rellenan mediante un método de relleno de pixels de borde.

50 En esta forma de realización, cuando el aparato de comunicación a través de vídeo adopta un aparato de visualización de estéreo en 3D y el aparato de presentación visual solamente soporta la visualización de imágenes izquierdas y derechas de entrada, necesita reconstruirse otra imagen, con el fin de realizar una presentación visual en estéreo. Algunas visualizaciones en estéreo automáticas soportan la visualización de estéreo en 3D de una imagen de color en 2D con un valor de profundidad correspondiente a la imagen en color, de modo que otra imagen no necesite reconstruirse. En cambio, la visualización en estéreo automática realiza la reconstrucción de la otra imagen y el relleno de agujeros correspondiente se realiza en el proceso de reconstrucción, por ejemplo, un monitor de presentación visual en estéreo de Philips.

60 La reconstrucción de visualización se refiere también como una síntesis de imagen de puntos de visualización virtual, lo que suele significar la reconstrucción de una imagen en otro ángulo de vista desde un modelo o una imagen en un ángulo diferente. En esta forma de realización, el módulo de reconstrucción de visualización 68 realiza la reconstrucción. Cuando se conoce una profundidad de una imagen, un paralaje entre un punto de visualización virtual y una vista conocida se puede calcular aplicando la fórmula siguiente:

65
$$d=A1O1'-A2O2'=f*(CO1-CO2)/Z=f*B/Z.$$

en donde d es el paralaje entre la vista del punto de visualización virtual y la vista conocida, f es una longitud focal de la cámara, B es una distancia entre el punto de vista virtual y un punto de fotografía original y Z es una profundidad de la imagen.

- 5 Cuando se reconstruye una imagen derecha en función de la imagen sintetizada y su profundidad, un color de un pixel en una determinada línea de barrido x_r en la imagen derecha se determina por un color de un pixel en una línea de barrido x_l correspondiente en una imagen izquierda (una imagen sintetizada). Una coordenada de x_l se determina en función de la ecuación siguiente:

$$x_l = x_r + d = x_r + \frac{fB}{Z}$$

- 10 Cuando el contenido de la vista sintetizada se determina en función de la fórmula anterior, puntos correspondientes a algunos puntos en la imagen derecha no se pueden encontrar en la imagen izquierda debido al bloqueo, es decir, ocurre el bloque de los agujeros. De forma similar, un punto de pixel en un borde del agujero se utiliza para rellenar el agujero y el relleno se puede realizar por intermedio de un modo de interpolación bilineal.

- 15 En esta forma de realización, el módulo de presentación visual está configurado para visualizar la imagen sintetizada. El módulo de presentación visual 67 puede ser un aparato de presentación visual en estéreo tal como un aparato de presentación visual en estéreo automático, un par de gafas de estéreo y un aparato de visualización holográfica capaz de
 20 presentación visual de estéreo en 3D, con el fin de obtener una visualización en estéreo de una imagen de estéreo, de modo que un usuario experimente la profundidad de una escena y sienta el efecto de estéreo. Cuando se necesita la presentación visual en estéreo, la reconstrucción de la vista y el relleno de agujeros suele necesitar realizarse. En esta forma de realización, el módulo de presentación visual puede ser también un aparato de presentación visual en 2D normal, que solamente visualiza una imagen sintetizada en 2D. Cuando solamente necesita visualizarse una imagen en
 25 2D, la reconstrucción de la vista no se necesita y se visualiza directamente la imagen en 2D sintetizada.

Forma de realización tres

- 30 En esta forma de realización, se proporciona un ejemplo de un proceso de comunicación en el sistema de comunicación a través de vídeo, que es concretamente un proceso completo en el que dos usuarios (A y B) se comunican a través del aparato de comunicación a través de vídeo en una segunda forma de realización y el usuario A envía datos de vídeo al usuario B y el usuario B recibe los datos de vídeo del usuario A en el proceso de comunicación. En la Figura 10 se ilustra una estructura del sistema de comunicación a través de vídeo que incluye un extremo emisor y un extremo receptor. El extremo emisor y el extremo receptor están conectados a través de una red.

- 35 El extremo emisor está configurado para adquirir un contenido de escena y su valor de profundidad del extremo emisor, efectúa la segmentación de un contenido de objeto del extremo emisor a partir del contenido de escena del extremo emisor en función de un valor de profundidad del contenido de escena del extremo emisor y para enviar el contenido del objeto del extremo emisor y su valor de profundidad al extremo receptor. El extremo emisor incluye un módulo de adquisición de información 1001, un módulo de segmentación 1002, un módulo de codificación 1003 y un módulo de interfaz de transmisión 1004. El módulo de adquisición de información 1001 está configurado para fotografiar los contenidos de escenas locales y para calcular los valores de profundidad correspondientes a los contenidos de escenas locales o adquirir directamente los valores de profundidad correspondientes a los contenidos de escenas locales. El
 40 módulo de segmentación 1002 está configurado para segmentar el contenido del objeto local a partir de los contenidos de escenas locales en función del valor de profundidad. El módulo de codificación 1003 está configurado para codificar el contenido del objeto local segmentado y un valor de profundidad correspondiente al contenido del objeto local. El módulo de interfaz de transmisión 1004 está configurado para enviar el contenido del objeto local y su valor de profundidad al extremo receptor.

- 50 El extremo receptor está configurado para recibir un contenido de objeto y su valor de profundidad enviado por el extremo emisor, para adquirir un contenido de fondo y su valor de profundidad del extremo receptor y para sintetizar el contenido de objeto del extremo emisor y el contenido de fondo del extremo receptor en un contenido de escena en función de los valores de profundidad. El extremo receptor incluye un módulo de interfaz de transmisión 1005, un módulo de decodificación 1006, un módulo de síntesis 1007, un módulo de reconstrucción de vista 1008, un módulo de presentación visual 1009 y un módulo de extracción 1010. El módulo de interfaz de transmisión 1005 está configurado
 55 para recibir un contenido de objeto y su valor de profundidad enviado por un extremo distante. El módulo de decodificación 1006 está configurado para decodificar el contenido del objeto distante recibido y su valor de profundidad. El módulo de síntesis 1007 está configurado para sintetizar el contenido del objeto distante decodificado y el contenido de fondo local y para generar una vista en estéreo en función de los valores de profundidad correspondientes, en donde el contenido del fondo local puede ser un contenido residual después de que sea segmentado el contenido del objeto local en los contenidos de escenas locales. El módulo de extracción 1010 está configurado para extraer el contenido residual; el contenido de fondo local puede ser también un contenido de escena opuesto al objeto local fotografiado utilizando otro grupo de cámaras. El módulo de presentación visual 1009 está configurado para visualizar la imagen

sintetizada, que puede ser un aparato de presentación visual en estéreo o un aparato de presentación visual en 2D normal; si el módulo de presentación visual 1009 es un aparato de visualización en estéreo, necesita reconstruirse una imagen en 2D en otro punto de visualización. La reconstrucción de la imagen en 2D en otro punto de visualización se puede realizar utilizando el módulo de reconstrucción de visualización 1008.

5 El proceso de comunicación es según se ilustra en la Figura 9, que incluye concretamente las etapas siguientes.

10 En la etapa 901, un módulo de adquisición de información de un aparato de comunicación a través de vídeo de un usuario A adquiere los contenidos de escenas locales y sus valores de profundidad. Los contenidos de escenas locales y los valores de profundidad de los contenidos de escenas se pueden adquirir utilizando una cámara de profundidad o una cámara de estéreo. La cámara de profundidad puede adquirir una profundidad directamente a través de rayos infrarrojos. La cámara de estéreo suele adquirir el contenido de escena utilizando dos cámaras en paralelo y posteriormente, se calcula un valor de profundidad de cada pixel en el contenido de escena. La fórmula de cálculo es $Z=fB/\Delta x$, en donde f es una longitud focal, B es una distancia entre las dos cámaras y Δx es una diferencia de posiciones de cada pixel en dos cámaras.

15 En la etapa 902, el módulo de segmentación del aparato de comunicación a través de vídeo del usuario A efectúa la segmentación del contenido del objeto local a partir de los contenidos de escenas locales, que incluyen concretamente las etapas siguientes. Una unidad de reconocimiento de caras, en el módulo de segmentación, realiza el reconocimiento de caras en el contenido de escena local fotografiado para obtener una posición de una imagen de cara y posteriormente, una unidad de búsqueda en el módulo de segmentación busca un valor de profundidad correspondiente a la posición de la imagen de cara en los valores de profundidad de los contenidos de escenas locales y determina un margen del valor de profundidad de una persona en la imagen fotografiada según el valor de profundidad encontrado. De este modo, se puede determinar una zona del contenido del objeto local en el contenido de escena. Por último, la unidad de segmentación, en el módulo de segmentación, realiza la segmentación de un objeto de persona a partir de los contenidos de escenas locales en función de la zona determinada.

20 En la etapa 903, después de que se segmente un objeto de persona local, un contenido residual, después del objeto de persona local, se segmenta a partir de los contenidos de escenas locales y se puede memorizar un valor de profundidad del contenido residual. Además, se puede utilizar otra cámara para adquirir y memorizar un contenido de fondo opuesto al objeto de persona y su valor de profundidad al mismo tiempo.

25 En la etapa 904, con el fin de obtener un tamaño de una persona de objeto local uniforme, necesita ponerse a escala un objeto de persona local hasta un tamaño de una imagen de adquisición original o personalizada en una imagen en otros tamaños. Las zonas de agujeros resultantes de la personalización se rellenan con un valor 0.

30 En la etapa 905, se codifica el objeto de persona local y su valor de profundidad obtenido en la etapa 904, respectivamente. La codificación en capas se prefiere cuando la cantidad de datos que necesita transmitirse en la codificación en capas es pequeña.

35 En la etapa 906, el objeto de persona local codificado y su valor de profundidad se envían al aparato de comunicación a través de vídeo del usuario B por intermedio del módulo de interfaz de transmisión.

40 A través de las etapas anteriores, las operaciones de envío del usuario A son concluidas. En las etapas siguientes, un usuario B recibe datos y se procesan los datos.

45 En la etapa 907, un aparato de comunicación a través de vídeo del usuario B recibe un objeto de persona y su valor de profundidad enviado por el usuario A por intermedio de un módulo de interfaz de transmisión.

50 En la etapa 908, el aparato de comunicación a través de vídeo del usuario B decodifica los datos recibidos utilizando un módulo de decodificación, con el fin de obtener un objeto de persona y su valor de profundidad del usuario A. Al mismo tiempo, el aparato de comunicación a través de vídeo del usuario B necesita, además, adquirir un contenido de fondo y un valor de profundidad del contenido de fondo. En términos generales, un contenido residual después de que se elimine el objeto local en los contenidos de escenas locales se puede utilizar como el contenido de fondo. Si otra cámara adquiere un contenido de fondo opuesto al usuario B y su valor de profundidad, el usuario B puede ver más imágenes naturales y el problema de los agujeros no ocurre durante la síntesis de imagen.

55 En la etapa 909, una unidad de escalamiento, en el módulo de síntesis, pone a escala el objeto de persona y su valor de profundidad enviado por el usuario A, con el fin de obtener un objeto de persona que tenga un tamaño óptimo. Cuando el contenido de objeto distante necesita llevarse a una distancia más corta, es decir, cuando se disminuye un valor de profundidad, el contenido del objeto distante necesita ponerse a escala ascendente. Cuando el contenido del objeto distante está dispuesto a una mayor distancia, es decir, cuando se aumenta el valor de profundidad, necesita ponerse a escala descendente el contenido del objeto distante.

Posteriormente, se determina una relación de bloqueo entre un objeto de persona distante y un contenido de fondo local en función de un valor de profundidad de un objeto de persona después del escalamiento y del valor de profundidad del contenido de fondo del usuario A. Un principio de bloqueo es como sigue. Cuando se solapan las posiciones horizontal y vertical del punto de pixel, un punto de pixel que tenga un pequeño valor de profundidad bloquea un punto que tenga un valor de profundidad mayor (una vista cercana bloquea una vista distante).

Una unidad de síntesis en el módulo de síntesis sinteriza entonces el objeto de persona y el contenido de fondo en un contenido de escena, en función de la relación de bloqueo determinada.

Si el contenido de fondo es un contenido residual después de que se elimine el contenido de objeto, necesita realizarse un relleno de pixels en los agujeros en el contenido de escena sintetizado. Si el contenido de fondo es una escena opuesta al usuario B de adquisición directa, no se necesita el relleno de pixels.

En la etapa 910, el módulo de reconstrucción de visualización realiza una síntesis de imagen de puntos de vista virtual en el contenido de escena sintetizado, que es concretamente para calcular un paralaje entre una vista de visualización virtual y una vista conocida en función de la fórmula siguiente.

$$d=A1O1'-A2O2'=f*(CO1-CO2)/Z=f*B/Z.$$

en donde d es el paralaje entre una vista virtual y la vista conocida, f es una longitud focal de la cámara, B es una distancia entre el punto de vista virtual y un punto de fotografía original y Z es una profundidad de la imagen.

Cuando una imagen derecha de la imagen sintetizada se reconstruye basada en la imagen sintetizada y su profundidad, un color de un pixel en una determinada línea de barrido x_r , en la imagen derecha, se determina por un color de un pixel en una línea de barrido correspondiente en una imagen izquierda (una imagen sintetizada) x_l . Una coordenada de x_l se determina aplicando la fórmula siguiente.

$$x_l = x_r + d = x_r + \frac{fB}{Z}$$

Una vez reconstruida la vista, necesita realizarse un relleno de pixel en agujeros en el contenido de escena después de la síntesis de la imagen del punto de visualización virtual.

En la etapa 911, el contenido de escena sintetizado se visualiza utilizando un módulo de presentación visual tal como un aparato de visualización en estéreo automático, un par de gafas de estéreo o un aparato de visualización holográfica capaz de proporcionar una presentación visual de estéreo en 3D, con el fin de obtener una visualización en estéreo de una imagen de estéreo o solamente se visualiza una imagen sintetizada en 2D utilizando un aparato de presentación visual en 2D normal.

En esta forma de realización, en el sistema de comunicación a través de vídeo, el aparato del usuario A puede incluir, además, un dispositivo de recepción de vídeo y el aparato del usuario B puede incluir, además, un dispositivo de preprocesamiento de vídeo, con el fin de garantizar que el usuario B pueda enviar datos de vídeo al usuario A. Si el usuario B necesita enviar los datos de vídeo al usuario A, el proceso es el mismo que el ilustrado en la Figura 9 y solamente se intercambian las posiciones del emisor y del receptor. Esta forma de realización de la presente invención se utiliza principalmente en la comunicación a través de vídeo, tal como una conversación en vídeo normal, telefonía de vídeo de negocios y videoconferencia.

Forma de realización cuatro

En una forma de realización, la presente invención da a conocer un método de procesamiento de vídeo. Según se ilustra en la Figura 11, el método comprende las etapas siguientes.

En la etapa 111, las imágenes en color/escala de grises de al menos dos puntos de visualización de un objeto de vídeo a presentar y una imagen en color/escala de grises de al menos un punto de visualización de al menos un material de presentación son objeto de adquisición.

En la etapa 112, las imágenes en color/escala de grises de los al menos dos puntos de visualización del objeto de vídeo y la imagen en color/escala de grises del al menos un punto de visualización del al menos un material de presentación se sintetizan para adquirir una imagen de presentación.

En la etapa 113, la imagen de presentación se visualiza en un modo en 3D.

En una forma de realización, la presente invención da a conocer una solución técnica. Adquiriendo las imágenes en color/escala de grises de múltiples puntos de visualización del objeto de vídeo y la imagen en color/escala de grises del

material de presentación, se genera la imagen de presentación. La imagen de presentación soporta un modo de presentación visual en 3D y a continuación, se visualiza la imagen de presentación en el modo de 3D, con el fin de resolver el problema de que solamente la presentación de vídeo en 2D se soporta en la técnica anterior, consiguiendo de este modo, la presentación de vídeo en 3D.

5 Con el fin de ilustrar la solución técnica de las formas de realización de la presente invención con más claridad, a continuación se ilustra brevemente la tecnología de vídeo en 3D/vídeo en estéreo en la forma de realización de la presente invención.

10 La tecnología de vídeo convencional solamente proporciona información en 2D. Sin embargo, en la forma de realización de la presente invención, utilizando una tecnología de vídeo en 3D, un usuario, como un observador, no solamente aprende información relacionada con el contenido del escenario sino que también obtiene información de profundidad tal como distancias y posiciones del escenario.

15 La tecnología de vídeo en 3D puede proporcionar una imagen que está conforme con un principio de visión en estéreo y tiene información de profundidad, con el fin de reproducir escenas del mundo subjetivo de forma resaltada y para presentar una sensación de profundidad, una sensación de capas y naturalidad de la escena, lo que es una orientación importante de desarrollo de la tecnología de vídeo actual.

20 El principio básico de la tecnología de vídeo en 3D es simular los principios de creación de imágenes de ojos humanos. Una imagen en el ojo izquierdo y una imagen en el ojo derecho se obtienen utilizando cámaras duales. Durante la presentación, una persona es habilitada para ver las imágenes del ojo izquierdo y del ojo derecho con los ojos izquierdo y derecho, respectivamente, con el fin de sintetizar una imagen que tenga finalmente una sensación de estéreo, de modo que un observador pueda sentir una profundidad de la escena. Por lo tanto, el vídeo en estéreo de dos ojos se puede
25 considerar como información de profundidad añadida al vídeo en 2D existente.

En una parte de la adquisición de las imágenes en color/escala de grises de los al menos dos puntos de visualización del objeto de vídeo a presentar, se adoptan tecnologías de vídeo tales como vídeo de múltiples puntos de visualización (MVV)/vídeo de punto de visualización libre (FVV). Una idea básica de la tecnología de MVV es fotografiar escenas al
30 mismo tiempo utilizando más de dos cámaras. Por ejemplo, para una escena de deportes o teatro, diferentes cámaras tienen diferentes ángulos de toma de fotografías, de modo que se generen más de dos flujos de vídeo. Los flujos de vídeo que tienen diferentes puntos de vista se envían a un terminal de usuario. El usuario puede seleccionar un punto de vista aleatorio y la dirección para observar la escena. El punto de vista puede ser un punto de vista de fotografía de una cámara fija o un punto de visualización virtual y una imagen en el punto de vista virtual se obtiene sintetizando imágenes
35 fotografiadas por cámaras reales a su alrededor.

Forma de realización cinco

40 En una forma de realización, la presente invención da a conocer un método de envío de señales de vídeo. Según se ilustra en la Figura 12, el método comprende las etapas siguientes.

En la etapa 121, se adquieren imágenes en color/escala de grises de al menos dos puntos de visualización de un objeto de vídeo a presentar y una imagen en color/escala de grises de al menos un punto de visualización de al menos un material de presentación.

45 El objeto de vídeo a presentar incluye un objeto tal como una persona y un objeto que sirve como un primer plano después de que una escena de naturaleza se segmente o también puede ser un objeto de fondo. El material de presentación puede ser un contenido tal como un documento, una fotografía, un vídeo o un gráfico generado por un ordenador ha de presentarse.

50 En una forma de realización, la presente invención da a conocer el método para adquirir las imágenes en color/escala de grises de los al menos dos puntos de visualización del objeto de vídeo que incluye las etapas siguientes.

55 En la etapa 121, se adquiere información de profundidad y al menos una imagen en color/escala de grises en el mismo punto de visualización de la escena en la que se encuentra el objeto de vídeo.

Más de dos cámaras de color normales se pueden utilizar para adquirir una imagen de profundidad aplicando un método basado en la coincidencia de imágenes. En consideración de la complejidad y del rendimiento en tiempo real deficiente de los algoritmos en el método, en una forma de realización preferida de la presente invención, la información de
60 profundidad se adquiere utilizando una cámara que puede adquirir información de profundidad de una escena y se adquiere una imagen en color/escala de grises utilizando una cámara que puede adquirir una imagen en color/escala de grises de una escena. A continuación, se ilustran brevemente los principios de funcionamiento de las dos cámaras.

65 La Figura 13 es una vista de principio esquemática de adquisición de una imagen de profundidad utilizando una cámara provista de un dispositivo de carga acoplada (CCD) de un obturador de ultra-alta velocidad y un iluminador modulado en intensidad. Un objeto a y un objeto b existen en la escena y el objeto a es un objeto cuadrado y el objeto b es un objeto

triangular. En la Figura 13, I_1 e I_2 son intensidades de luz de los objetos a y b adquiridas por la cámara cuando se abre el obturador (según se ilustra por c en la Figura 13), respectivamente, d_1 y d_2 representan una luz modulada con intensidad creciente y una luz modulada con intensidad decreciente, respectivamente.

5 La intensidad de luz instantánea I_1 de una luz reflejada transmitida desde el objeto más próximo a la cámara se detecta por el obturador de ultra-alta velocidad de un dispositivo de detección en la cámara, se obtiene la distribución cuadrática en la imagen A y se obtiene la distribución triangular en la imagen A para una luz reflejada del objeto b. Cuando el objeto a está más próximo a la cámara, la intensidad de luz instantánea I_1 detectada por la cámara es más alta que la intensidad de luz instantánea I_2 y el brillo de la imagen cuadrada es mayor que el de la imagen triangular. Por lo tanto, la profundidad del objeto se puede detectar por intermedio de una diferencia de brillo de la imagen capturada a. Sin embargo, el brillo de la luz reflejada del objeto es incluido por parámetros tales como la reflectividad del objeto, una distancia entre el objeto y la cámara, un índice de modulación de una fuente de luz y la falta de uniformidad espacial de la iluminación. En este momento, se puede obtener una imagen B mediante decremento lineal de la distribución espacial de la intensidad luminosa. La imagen A y la imagen B se combinan y se pueden eliminar las influencias adversas mediante un algoritmo de procesamiento de señales, con el fin de obtener un mapa de profundidad preciso. En el mapa de profundidad, una profundidad del objeto B (el objeto triangular en la Figura 13) es mayor que una profundidad del objeto A (el objeto cuadrado en la Figura 13), es decir, visualmente, el objeto A está más próximo a la cámara y el objeto B está más alejado de la cámara.

20 De forma opcional, en la forma de realización de la presente invención, la información de la profundidad se puede adquirir utilizando la cámara y una cámara para adquirir la imagen en color/escala de grises. Como alternativa, una cámara puede adquirir directamente la información de la profundidad y la imagen en color/escala de grises se adopta al mismo tiempo.

25 La Figura 14 es una vista esquemática de una estructura básica de una cámara Axi-visión de alta definición (HD). La cámara puede adquirir la información de profundidad y la imagen en color/escala de grises al mismo tiempo. El sistema de cámara Axi-visión de alta definición HD incluye una unidad de procesamiento de imágenes de profundidad y una unidad de procesamiento de imágenes en color. Un conjunto matricial de diodos LED casi -infrarrojos se utiliza para un iluminador modulado en intensidad, que es capaz de una modulación directa rápida. Una longitud de onda de una luz emitida por el conjunto matricial de diodos LED de casi-infrarrojos es de 850 nm, que está más allá del margen de la luz visible, por lo que no se interfiere la luz visible. Cuatro conjuntos matriciales de diodos LED se pueden disponer alrededor de una lente de la cámara, con el fin de iluminar uniformemente la escena fotografiada. Al mismo tiempo, se puede proporcionar, además, una fuente de luz visible tal como una fuente de luz fluorescente, que está configurada para irradiar el objeto que se va a fotografiar. La fuente de luz tiene un espectro que excede una zona de la luz de casi-infrarrojos.

30 Cuando la luz refleja del objeto pasa a través de un prisma dicroico de la lente de la cámara, se separan la luz visible y la luz de casi-infrarrojos. La luz visible entra en la unidad de procesamiento de imágenes en color y se procesa por la unidad de procesamiento de imágenes en color, con el fin de obtener una imagen en color del objeto, es decir, una imagen en 2D. La unidad de procesamiento de imágenes en color puede ser una cámara HD de color. La luz de casi-infrarrojos pasa a través de la unidad de procesamiento de imágenes profundas y se procesa con el fin de obtener una imagen profunda del objeto. En la unidad de procesamiento de imágenes profundas, la luz de casi-infrarrojos, separada en el prisma dicroico, se enfoca en un fotocátodo al mismo tiempo. Una polarización de pulsos cortos se aplica entre el fotocátodo y una placa de microcanales (MCP), con el fin de obtener una velocidad del obturador de una billonésima de segundo. Una frecuencia de apertura del obturador es la misma que una frecuencia de la modulación de rayos de luz, con el fin de obtener una mejor relación de señal a ruido (SNR). Al abrir el obturador, se obtiene una imagen óptica del objeto en un fósforo. La imagen óptica pasa, entonces, a través de una lente de relé y se enfoca en una cámara CCD progresiva de alta resolución y se convierte en una imagen fotoelectrónica y a la larga, la imagen fotoelectrónica forma el mapa de profundidad del objeto por intermedio de un procesador de señales.

50 En la descripción anterior, se dan a conocer dos métodos preferibles para adquirir información de profundidad y una imagen en color/escala de grises. Sin embargo, las formas de realización de la presente invención no están así limitadas. Todos los demás métodos similares y pertinentes para adquirir la información de profundidad y las imágenes en color/escala de grises habrán de caer dentro del alcance de protección de las formas de realización de la presente invención.

60 En la etapa 121, en función de la información de profundidad adquirida en un punto de visualización, se realiza la segmentación de vídeo en una imagen en color/escala de grises en el mismo punto de visualización para adquirir el objeto de vídeo.

65 La segmentación de vídeo se puede realizar mediante más de un método. Una imagen de vídeo está dividida en un primer plano y en un fondo, por ejemplo, mediante una tecnología de segmentación de inserción croma, una tecnología de segmentación de inserciones de profundidad o una tecnología de segmentación que detecta una diferencia entre una imagen actual y una imagen de fondo fotografiada con anticipación. Las primera y tercera tecnologías tienen más condiciones de límite para la escena. En una forma de realización preferida de la presente invención, la segmentación de

vídeo se realiza mediante la tecnología de segmentación de inserciones-profundidad, que incluye principalmente los puntos inserciones técnicos como sigue.

5 Una máscara de profundidad binarizada se genera mediante un valor de umbral en función del mapa de profundidad y el objeto de vídeo se extrae en función de la máscara. Por ejemplo, un pixel que tenga un valor de máscara de 1 es un pixel de objeto de primer plano y un pixel que tenga un valor de máscara de 0 es un pixel de objeto de fondo, de modo que el objeto de vídeo se puede extraer o eliminar en función de los valores de máscara.

10 Una descripción de código- cadena se construye para un contorno de la máscara de profundidad. Un contorno del objeto de primer plano se recupera en función de la descripción de código-cadena. Una zona de procesamiento se define en función del contorno del objeto como una zona de síntesis Alfa.

15 En la etapa 121, al menos se genera una imagen en color/escala de grises en otro punto de visualización del objeto de vídeo en función de la información de profundidad del punto de visualización y la imagen en color/escala de grises del mismo punto de visualización.

A continuación, se describe brevemente la tecnología para generar una imagen de múltiples puntos de visualización utilizada en las formas de realización de la presente invención.

20 Un determinado punto de pixel $P=[X, Y, Z]$ en el espacio se proyecta a un punto $[x, y]$ en un plano de imágenes en 2D de la cámara, que satisface las ecuaciones siguientes:

$$x = F \frac{X}{Z}, \quad y = F \frac{Y}{Z},$$

, en donde F es una longitud focal de la cámara.

25 Se supone que ambos puntos de pixel $P_1=[X_1, Y_1, Z_1]$ y $P_2=[X_2, Y_2, Z_2]$, en el espacio, se proyectan en un par de cámaras. Una cámara está situada en $[0,0,0]$ (en una cámara izquierda) y otra cámara está situada en $[B,0,0]$ (una cámara derecha). Las longitudes focales de las dos cámaras son iguales y ambas son F, que están en posiciones paralelas y satisfacen las ecuaciones siguientes:

$$30 \quad x_{L,1} = F \frac{X_1}{Z_1}; \quad x_{L,2} = F \frac{X_2}{Z_2}; \quad x_{R,1} = F \frac{X_1 - B}{Z_1}; \quad x_{R,2} = F \frac{X_2 - B}{Z_2}$$

en donde $x_{L,1}$ y $x_{L,2}$ son posiciones de los puntos P_1 y P_2 obtenidas por la cámara izquierda respectivamente y $x_{R,1}$ y $x_{R,2}$ son posiciones de los puntos P_1 y P_2 obtenidas por la cámara derecha respectivamente, de modo que las profundidades de los puntos P_1 y P_2 (es decir, un paralaje), satisfaga la ecuación siguiente:

$$35 \quad d_i = x_{L,i} - x_{R,i} = F_i \cdot \left[\frac{X_i}{Z_i} - \frac{X_i - B}{Z_i} \right] = \frac{F \cdot B}{Z_i} \Rightarrow x_{R,i} = x_{L,i} - d_i$$

, en donde i es 1 o 2.

Se puede deducir de lo anterior que en tanto que se conozcan los valores de x_L y de d, se puede calcular el valor de x_R .

40 En la forma de realización de la presente invención, x_L se obtiene mediante una imagen en color/escala de grises de la cámara de profundidad y d se puede calcular a través del mapa de profundidad, de modo que se pueda generar una imagen en color/escala de grises del objeto de vídeo en otro punto de visualización.

45 Puesto que la profundidad no suele ser un número entero, en la forma de realización de la presente invención, un nivel de subpixel se adopta para las posiciones de pixel calculadas. En este momento, los valores del brillo y de croma de un nuevo pixel correspondiente se puede determinar por intermedio de un método de media ponderada en función de los valores de brillo y de croma de un pixel adyacente a un determinado pixel en la imagen original.

50 Mediante este método, se pueden generar más de dos imágenes en color/escala de grises en diferentes puntos de visualización del objeto de vídeo.

55 Con el fin de garantizar la adquisición de imágenes en color/escala de grises de alta calidad, cuando se genera la imagen de múltiples puntos de visualización, necesitan resolverse los problemas de "bloqueo" y "agujero" causados por los cambios de puntos de visualización en la imagen. A continuación se ilustra brevemente el método para resolver los problemas de los "blocajes" y "agujeros" en la forma de realización de la presente invención.

La Figura 15 representa una situación en que ocurre un “bloqueo” cuando se genera una imagen de múltiples puntos de visualización. La Figura 15 (1), a la izquierda representa una situación de observación de la escena en un punto de visualización de imagen original O_1 y la Figura 15 (2), a la derecha, representa una situación de observación de la escena en un nuevo punto de visualización O_2 en donde la imagen necesita reconstruirse. Cuando se observa la escena en el punto O_1 a la izquierda, un objeto frontal (según se representa por A en la Figura 15) bloquea una pequeña parte de un objeto posterior (según se representa por B en la Figura 15) y C en la Figura 15 representa una zona de bloqueo. Cuando se observa la escena en el punto O_2 a la derecha, la zona de bloqueo del objeto posterior por el objeto frontal es ampliada, de modo que una parte en la imagen obtenida en el punto O_1 no se puede visualizar en la imagen en el punto O_2 . Por lo tanto, en el proceso de mapeo de pixels, necesita determinarse si un pixel está, o no, bloqueado. Si el pixel no está bloqueado, se realiza el mapeo de pixels. Si el pixel está bloqueado, se salta el procesamiento.

La Figura 16 representa una situación en la que ocurre un “agujero” cuando se genera una imagen de múltiples puntos de visualización. La Figura 16 (1), a la izquierda, representa una situación de observación de una escena en el punto de visualización O_1 de la imagen original y la Figura 16 (2), a la derecha, representa una situación de observación de una escena en el punto de visualización O_2 en donde necesita reconstruirse la imagen. En la posición O_1 a la izquierda, debido al bloqueo del objeto posterior (según se representa por B en la Figura 16) por el objeto frontal (según se representa por A en la Figura 16), no se puede observar una parte izquierda del objeto posterior, de modo que la imagen generada en el punto O_1 no tiene pixel en esta parte. Cuando se observa la escena en la nueva posición O_2 a la derecha, una parte izquierda del objeto posterior no está bloqueada; sin embargo, puesto que la imagen generada en el punto O_1 no tiene pixel en esta parte, la imagen generada en el punto O_2 carece de zona de agujeros correspondientes a los pixel de imagen en el punto O_1 (según se ilustra por C en la Figura 16).

Con el fin de procesar los agujeros, todos los pixels en una nueva imagen se establecen a valores cromáticos especiales antes de que se genere la nueva imagen. Una vez concluido el proceso de mapeado, las zonas que tienen valores cromáticos especiales, todavía mantenidas en la imagen, son zonas de agujeros. Para una zona de agujeros pequeña, la información correspondiente de pixel en la zona de agujeros se puede determinar en función de la información de profundidad, brillo y croma de los pixels alrededor del agujero y se repara la zona del agujero, por ejemplo, por intermedio de un método de interpolación lineal o no lineal. Para un agujero grande, se puede adoptar un método de compensación de movimiento, en el que la información de pixel correspondiente a la zona de agujero se busca a partir de una secuencia de tramas de vídeo antes de la trama actualmente reconstruida y la zona de agujeros se repara en consecuencia.

El método de adquisición de las imágenes en color/escala de grises de los al menos dos puntos de visualización del objeto de vídeo a presentar es también adecuado para adquirir la imagen en color/escala de grises del al menos un punto de visualización de al menos un material de presentación.

En la etapa 122, las imágenes en color/escala de grises de los al menos dos puntos de visualización del objeto de vídeo y la imagen en color/escala de grises del al menos un punto de visualización del al menos un material de presentación se sintetizan para adquirir una imagen de presentación.

La síntesis de imágenes es para integrar, de forma continua, el objeto de primer plano segmentado en el fondo. De forma opcional, el fondo, en este caso, puede ser una imagen en color/escala de grises generada a partir del material de presentación por intermedio de la imagen de múltiples puntos de visualización o una imagen en color/escala de grises de un punto de visualización del material de presentación. En el método en la forma de realización de la presente invención, se realiza la síntesis de valores alfa en bordes del objeto de primer plano y el fondo basado en la imagen de profundidad. El método incluye principalmente los puntos inserciones técnicos como sigue.

En primer lugar, se define una zona a procesar. La zona de procesamiento está situada principalmente cerca de un contorno del objeto de primer plano, que es una zona rectangular con el punto de código-cadena como centro.

Un tamaño de la zona de procesamiento está relacionado con la agudeza de los bordes del objeto de primer plano. La agudeza de los bordes del objeto de primer plano se obtiene calculando una derivada en una dirección vertical de los bordes del objeto. Una zona de transición desde un primer plano que tenga bordes agudos para el fondo es pequeña, de modo que la zona a procesar es pequeña. Una zona borrosa es gradualmente transicional desde un primer plano a un fondo, de modo que una zona a procesar es grande.

En un contorno de la zona de procesamiento definida, se supone que todos los puntos de pixel están en un objeto de primer plano (un valor alfa opaco y un color de primer plano puro) o fuera del objeto de primer plano (un valor de alfa transparente y un color de fondo puro). En la zona de procesamiento, el punto de pixel es híbrido del primer plano y del fondo, que tiene un valor alfa semitransparente y un valor de color $I(i,j)$ del punto de pixel es el híbrido del color de primer plano $F(i,j)$ y del color de fondo $B(i,j)$:

$$I(i,j) = \alpha * F(i,j) + (1 - \alpha) * B(i,j).$$

Un valor de color $I(i, j)$ de un determinado punto del pixel en la zona de procesamiento se puede calcular mediante la estimación del valor alfa α .

5 Aunque se describió anteriormente la tecnología clave de la síntesis de imágenes, adoptada en la forma de realización de la presente invención, la tecnología no está así limitada y se pueden adoptar otros métodos para síntesis de imágenes.

10 De forma opcional, en la forma de realización de la presente invención, la información de posición del objeto de vídeo y la información de órdenes de control de objeto de vídeo se pueden adquirir mediante la información de la profundidad. Las imágenes en color/escala de grises de los al menos dos puntos de visualización del objeto de vídeo y la imagen en color/escala de grises del al menos un punto de visualización del al menos un material de presentación se sintetizan para adquirir la imagen de presentación.

15 La información de posición se adquiere a través de la imagen de profundidad y se configura para controlar la posición del objeto de vídeo en el material de presentación. La información de orden de control se configura para controlar el contenido del material de presentación y la posición del objeto de vídeo en el material de presentación.

20 El método para adquirir la información de órdenes de control, dado a conocer en la forma de realización de la presente invención, comprende las etapas siguientes.

25 En función de la imagen de profundidad, se reconoce el gesto del objeto de vídeo y la información de reconocimiento del gesto se convierte en la información de órdenes de control.

30 Analizando la imagen de profundidad, se puede obtener la información de coordenadas en 3D de un punto característico de gesto del objeto de vídeo en la escena. El gesto del objeto de vídeo se reconoce a través de la información de coordenadas en 3D de los puntos característicos. Puesto que el mapa de profundidad puede expandir el reconocimiento del gesto del objeto de vídeo en un espacio en 3D, se pueden reconocer también, con precisión, los movimientos ascendentes y descendentes del gesto del objeto de vídeo. La información de reconocimiento se convierte en la información de órdenes de control.

35 Además, en las formas de realización de la presente invención, la información de cambio de profundidad del objeto de vídeo se puede obtener, además, detectando el mismo punto característico en diferentes espacios de dominios del tiempo y la información del cambio se convierte en la información de órdenes de control.

40 Con el fin de soportar la presentación a distancia del vídeo en una forma de realización, la presente invención comprende, además, las etapas siguientes.

En la etapa 123, se codifica y envía la imagen de presentación.

45 La imagen de presentación adquirida es objeto de compresión. Por ejemplo, el protocolo H.264 o el protocolo de Grupo - 4 de Expertos en Imágenes en Movimiento (MPEG-4) se adopta para codificar la imagen de presentación para adaptar el ancho de banda de red limitado y la imagen de presentación se transmite a un extremo distante a través de la red para su presentación.

50 En el extremo receptor de la red, en correspondencia, se decodifica la imagen comprimida. Por ejemplo, en el extremo receptor, se decodifica la imagen codificada en conformidad con el protocolo H.264 o el protocolo MPEG-4, con el fin de adquirir la imagen de presentación y visualizar la imagen de presentación en un modo en 3D. La imagen de presentación se puede visualizar en el modo en 3D mediante aparatos tales como un par de gafas de estereó, un monitor de estereó automático o un proyector.

55 En las formas de realización de la presente invención, la visualización de la imagen de presentación en un modo en 2D está incluida, de modo que los modos de presentación de imágenes en 2D y en 3D se hacen compatibles. En este caso, no se necesita la generación de imágenes de múltiples puntos de visualización del objeto de vídeo y las imágenes en color/escala de grises del objeto de vídeo directamente adquirido y el material de presentación se sintetizan y visualizan en un modo en 2D.

Forma de realización seis

60 En la forma de realización, la presente invención da a conocer, además, un método de envío de señales de vídeo. Según se ilustra en la Figura 17, el método consiste en garantizar la presentación a distancia de un vídeo en 3D, que comprende las etapas siguientes.

65 En la etapa 171, se adquieren imágenes en color/escala de grises de al menos dos puntos de visualización de un objeto de vídeo a presentar y una imagen en color/escala de grises de al menos un punto de visualización de al menos un material de presentación.

En la etapa 172, las imágenes en color/escala de grises de los al menos dos puntos de visualización del objeto de vídeo a presentar y la imagen en color/escala de grises del al menos un punto de visualización del al menos un material de presentación se codifican y envían.

5 En correspondencia, en la forma de realización de la presente invención, se da a conocer, además, un método de recepción de señales de vídeo, con el fin de poner en práctica la presentación a distancia del vídeo en 3D. Según se ilustra en la Figura 18, el método comprende las etapas siguientes.

10 En la etapa 181, se adquieren imágenes codificadas y las imágenes codificadas adquiridas se decodifican para obtener imágenes en color/escala de grises de los al menos dos puntos de visualización de un objeto de vídeo a presentar y una imagen en color/escala de grises del al menos un punto de visualización del al menos un material de presentación.

15 En la etapa 182, las imágenes en color/escala de grises de los al menos dos puntos de visualización del objeto de vídeo y la imagen en color/escala de grises del al menos un punto de visualización del al menos un material de presentación se sintetizan para adquirir una imagen de presentación.

En la etapa 183, la imagen de presentación se visualiza en un modo en 3D.

20 En la forma de realización de la presente invención, está incluida la visualización de la imagen de presentación en un modo en 2D, de modo que los modos de presentación de imágenes en 2D y en 3D se hacen compatibles. En este caso, no se necesita la generación de imágenes con múltiples puntos de visualización del objeto de vídeo y las imágenes en color/escala de grises del objeto de vídeo directamente adquirido y el material de presentación se sintetizan para su visualización en un modo en 2D.

25 En esta forma de realización de la presente invención, los métodos específicos de las etapas 171 y 172 y las etapas 181 a 183 son según se describe en la quinta forma de realización y las diferencias principales son como sigue.

30 Con el fin de poner en práctica la presentación a distancia, en consideración de factores tales como el ancho de banda de la red, en la quinta forma de realización de la presente invención, el extremo emisor procesa principalmente la imagen de vídeo adquirida, tal como una segmentación de vídeo y generación de múltiples puntos de visualización y el extremo emisor codifica la imagen de presentación sintetizada para transmisiones de la red. La imagen de presentación incluye el objeto de vídeo y el material de presentación, con el fin de poner en práctica la presentación en 3D de la imagen decodificada (es decir, la imagen de presentación). Sin embargo, en la sexta forma de realización de la presente invención, el extremo emisor procesa principalmente la imagen de vídeo adquirida en correspondencia y el extremo emisor solamente codifica la información de profundidad y las imágenes en color/escala de grises del objeto de vídeo y el material de presentación. El extremo receptor decodifica primero las imágenes codificadas, a continuación se sintetizan las imágenes decodificadas con el material de presentación para generar una imagen de presentación y la imagen de presentación se visualiza en el modo en 3D.

40 En las soluciones técnicas de la forma de realización de la presente invención, la imagen de presentación se genera adquiriendo imágenes en color/escala de grises de múltiples puntos de visualización del objeto de vídeo y la imagen en color/escala de grises del material de presentación, la imagen de presentación soporta el modo de presentación visual en 3D y la imagen de presentación se visualiza luego en el modo en 3D, con el fin de resolver el problema de que solamente se soporta la presentación de vídeo en 2D en la técnica anterior, con lo que se consigue la presentación de vídeo en 3D.

45 Forma de realización siete

50 En una forma de realización, la presente invención proporciona, además, un dispositivo de procesamiento de vídeo. Según se representa en la Figura 19, el dispositivo de procesamiento de vídeo incluye una unidad de adquisición y procesamiento de imágenes 191 y una unidad de presentación visual 192.

55 La unidad de adquisición y procesamiento de imágenes 191 está configurada para adquirir imágenes en color/escala de grises de al menos dos puntos de visualización de un objeto de vídeo a presentar y una imagen en color/escala de grises de al menos un punto de visualización de al menos un material de presentación y para sintetizar las imágenes en color/escala de grises de los al menos dos puntos de visualización del objeto de vídeo y la imagen en color/escala de grises del al menos un punto de visualización del al menos un material de presentación para adquirir una imagen de presentación.

60 La unidad de presentación visual 192 está configurada para visualizar la imagen de presentación en un modo en 3D.

La unidad de adquisición y procesamiento de imágenes 191 incluye un módulo de adquisición de imágenes, un módulo de segmentación de vídeo, un módulo de generación de imágenes de múltiples puntos de visualización y un módulo de síntesis.

65 El módulo de adquisición de imágenes está configurado para adquirir información de profundidad y al menos una imagen en color/escala de grises en el mismo punto de visualización en la escena en donde se encuentra el objeto de vídeo.

El módulo de segmentación de vídeo está configurado para realizar una segmentación de vídeo en la imagen en color/escala de grises en el mismo punto de visualización en función de la información de profundidad adquirida de un punto de visualización, con el fin de adquirir el objeto de vídeo.

5 El módulo de generación imágenes de múltiples puntos de visualización está configurado para generar una imagen en color/escala de grises de al menos otro punto de visualización del objeto de vídeo en función de la información de profundidad adquirida de un punto de visualización y la imagen en color/escala de grises del mismo punto de visualización.

10 El módulo de síntesis está configurado para sinterizar las imágenes en color/escala de grises de los al menos dos puntos de visualización del objeto de vídeo y la imagen en color/escala de grises del al menos un punto de visualización del al menos un material de presentación, con el fin de adquirir una imagen de presentación.

15 En la solución técnica dada a conocer en la forma de realización de la presente invención, la imagen de presentación se genera a través de las imágenes en color/escala de grises de múltiples puntos de visualización adquiridas del objeto de vídeo y la imagen en color/escala de grises del material de presentación, la imagen de presentación soporta el modo de presentación en 3D y la imagen de presentación se visualiza luego en el modo en 3D, con el fin de resolver el problema de que solamente se soporta la presentación de vídeo en 2D en la técnica anterior, con lo que se consigue la presentación de vídeo en 3D.

20 Forma de realización ocho

25 Con el fin de soportar la presentación de vídeo distante, según se representa en la Figura 20, en la forma de realización, la presente invención proporciona un dispositivo de envío de señales de vídeo, que incluye una unidad de adquisición de imágenes 201, una unidad de codificación 202 y una unidad de envío 203.

30 La unidad de adquisición de imágenes 201 está configurada para adquirir imágenes en color/escala de grises de los al menos dos puntos de visualización de un objeto de vídeo a presentar y una imagen en color/escala de grises del al menos un punto de visualización del al menos un material de presentación.

La unidad de codificación 202 está configurada para codificar las imágenes en color/escala de grises de los al menos dos puntos de visualización del objeto de vídeo a presentar y la imagen en color/escala de grises del al menos un punto de visualización del al menos un material de presentación para obtener imágenes codificadas.

35 La unidad de envío 203 está configurada para enviar las imágenes codificadas.

La unidad de adquisición de imágenes 201 incluye un módulo de adquisición de imágenes, un módulo de segmentación de vídeo y un módulo de generación de imágenes de múltiples puntos de visualización.

40 El módulo de adquisición de imágenes está configurado para adquirir información de profundidad y al menos una imagen en color/escala de grises del mismo punto de visualización en una escena en la que se encuentra el objeto de vídeo.

45 El módulo de segmentación de vídeo está configurado para realizar la segmentación de vídeo en la imagen en color/escala de grises en el mismo punto de visualización en función de la información de profundidad adquirida del punto de visualización, con el fin de adquirir un objeto de vídeo.

50 El módulo de generación de imágenes de múltiples puntos de visualización está configurado para generar una imagen en color/escala de grises de al menos otro punto de visualización del objeto de vídeo en función de la información de profundidad adquirida de un punto de visualización y la imagen en color/escala de grises del mismo punto de visualización.

55 Con el fin de conseguir, además, la presentación en 3D del vídeo a distancia, en correspondencia, según se representa en la Figura 21, en la forma de realización, la presente invención da a conocer, además, un dispositivo de recepción de vídeo, que incluye una unidad de recepción 210, una unidad de decodificación 211, una unidad de síntesis 212 y una unidad de presentación visual 213.

La unidad de recepción 210 está configurada para recibir imágenes codificadas.

60 La unidad de decodificación 211 está configurada para decodificar las imágenes codificadas recibidas para adquirir imágenes en color/escala de grises de al menos dos puntos de visualización de un objeto de vídeo a presentar y una imagen en color/escala de grises del al menos un punto de visualización de un al menos material de presentación.

65 La unidad de síntesis 212 está configurada para sintetizar imágenes en color/escala de grises de los al menos dos puntos de visualización del objeto de vídeo y la imagen en color/escala de grises del al menos un punto de visualización del al menos un material de presentación para adquirir una imagen de presentación.

La unidad de presentación visual 213 está configurada para visualizar la imagen de presentación en un modo en 3D.

Los modos de trabajo específicos de los módulos funcionales, en la forma de realización de la presente invención, son según se describen en la sexta forma de realización.

5 En la solución técnica de la forma de realización de la presente invención, la imagen de presentación se genera adquiriendo las imágenes en color/escala de grises de múltiples puntos de visualización del objeto de vídeo y la imagen en color/escala de grises del material de presentación, la imagen de presentación soporta el modo de presentación visual en 3D y la imagen de presentación se visualiza luego en el modo en 3D, con el fin de resolver el problema de que
10 solamente la presentación de vídeo en 2D se soporta en la técnica anterior, con lo que se consigue la presentación de vídeo en 3D.

Forma de realización nueve

15 En la forma de realización, la presente invención da a conocer un dispositivo de envío de señales de vídeo. Según se ilustra en la Figura 22, el dispositivo de envío de señales de vídeo incluye una unidad de adquisición y de procesamiento de imágenes 221, una unidad de codificación 222 y una unidad de envío 223.

20 La unidad de adquisición y procesamiento de imágenes 221 está configurada para adquirir imágenes en color/escala de grises de los al menos dos puntos de visualización de un objeto de vídeo a presentar y para sintetizar las imágenes en color/escala de grises de los al menos dos puntos de visualización del objeto de vídeo y la imagen en color/escala de grises del al menos un punto de visualización del al menos un material de presentación para adquirir una imagen de presentación.

25 La unidad de codificación 222 está configurada para codificar la imagen de presentación para obtener una imagen codificada.

La unidad de envío 223 está configurada para enviar la imagen codificada.

30 Modos de trabajo específicos de los modos funcionales, en la forma de realización de la presente invención, son según se describe en la quinta forma de realización.

35 En las soluciones técnicas dadas a conocer en la forma de realización de la presente invención, la imagen de presentación se genera adquiriendo las imágenes en color/escala de grises de múltiples puntos de visualización del objeto de vídeo y la imagen en color/escala de grises del material de presentación, la imagen de presentación soporta el modo de presentación visual en 3D y la imagen de presentación se visualiza luego en el modo de 3D, con el fin de resolver el problema de que solamente la presentación de vídeo en 2D se soporta en la técnica anterior, con lo que se consigue la presentación de vídeo en 3D.

40 Forma de realización diez

En esta forma de realización, la presente invención da a conocer un sistema de comunicación a través de vídeo. Según se representa en la Figura 23, el sistema de comunicación a través de vídeo incluye un dispositivo de envío de señales de vídeo 231 y un dispositivo de recepción de señales de vídeo 232.

45 El dispositivo de envío de señales de vídeo 231 incluye una unidad de adquisición y procesamiento de imágenes 2311, una unidad de codificación 2312 y una unidad de envío 2313.

50 La unidad de adquisición y procesamiento de imágenes 2311 está configurada para adquirir imágenes en color/escala de grises de los al menos dos puntos de visualización de un objeto de vídeo a presentar y una imagen en color/escala de grises del al menos un punto de visualización del al menos un material de presentación y para sintetizar las imágenes en color/escala de grises de los al menos dos puntos de visualización del objeto de vídeo y la imagen en color/escala de grises del al menos un punto de visualización del al menos un material de presentación para adquirir una imagen de presentación.

55 La unidad de codificación 2312 está configurada para codificar la imagen de presentación para obtener una imagen codificada.

La unidad de envío 2313 está configurada para enviar la imagen codificada.

60 El dispositivo de recepción de señales de vídeo 232 incluye una unidad de recepción 2321, una unidad de decodificación 2322 y una unidad de presentación visual 2323.

La unidad de recepción 2321 está configurada para recibir la imagen transmitida.

65

La unidad de decodificación 2322 está configurada para decodificar la imagen codificada recibida para adquirir la imagen de presentación.

La unidad de presentación visual 2323 está configurada para visualizar la imagen de presentación en un modo de 3D.

Además, cuando se realiza la presentación a distancia, el sistema de comunicación a través de vídeo puede realizar, además, la presentación visual en 3D de la imagen de vídeo del extremo receptor de red en el extremo emisor. El extremo emisor y el extremo receptor pueden tener las mismas funciones de procesamiento y de visualización de imágenes de vídeo y, en este momento, el dispositivo emisor de vídeo 231 incluye, además, una segunda unidad de decodificación y una segunda unidad de presentación visual.

La segunda unidad de decodificación está configurada para decodificar la imagen codificada recibida y para obtener la imagen de presentación codificada.

La segunda unidad de presentación visual está configurada para realizar la presentación visual en 3D de la imagen de presentación.

El dispositivo de recepción de señales de vídeo 232 incluye, además, una segunda unidad de adquisición y procesamiento de imágenes y una segunda unidad de codificación.

La segunda unidad de adquisición y procesamiento de imágenes está configurada para adquirir imágenes en color/escala de grises de los al menos dos puntos de visualización de un objeto de vídeo a presentar y una imagen en color/escala de grises del al menos un punto de visualización del al menos un material de presentación y para sintetizar las imágenes en color/escala de grises de los al menos dos puntos de visualización del objeto de vídeo y la imagen en color/escala de grises del al menos un punto de visualización del al menos un material de presentación para adquirir una imagen de presentación.

La segunda unidad de codificación está configurada para codificar la imagen de presentación para obtener una imagen codificada y enviar la imagen codificada.

En la solución técnica dada a conocer en una forma de realización del sistema de la presente invención, la imagen de presentación se genera adquiriendo las imágenes en color/escala de grises de múltiples puntos de visualización del objeto de vídeo y la imagen en color/escala de grises del material de presentación, la imagen de presentación soporta el modo de presentación visual en 3D y la imagen de presentación se visualiza luego en el modo en 3D, con lo que se resuelve el problema de que solamente la presentación de vídeo en 2D se soporta en la técnica anterior, con lo que se consigue la presentación de vídeo en 3D.

En la forma de realización del sistema de la presente invención, el extremo emisor de la red puede codificar la imagen de presentación que incluye el objeto de vídeo y el material de presentación para la transmisión a través de la red. La unidad de decodificación en el extremo receptor de la red, decodifica los datos recibidos y posteriormente la unidad de presentación visual presenta directamente la imagen decodificada. Sin embargo, la presente invención no está así limitada. De forma opcional, el proceso incluye también: en el extremo emisor solamente la información de profundidad y las imágenes en color/escala de grises del objeto de vídeo y el material de presentación se codifican para la transmisión a través de la red; en el extremo receptor de la red, la imagen decodificada y la imagen en estéreo del material de presentación se sintetizan para adquirir una imagen de presentación, que se presenta luego por la unidad de presentación visual.

Durante la síntesis de las imágenes de múltiples puntos de visualización y la adquisición de la imagen de presentación, la información de posición del objeto de vídeo se puede calcular, además, en función de la información de profundidad. La información de reconocimiento se genera para el gesto del objeto de vídeo y la información de reconocimiento se convierte en información de órdenes de control, con el fin de controlar la unidad de procesamiento de señales de vídeo.

En el sistema de comunicación a través de vídeo, dado a conocer en la forma de realización del sistema de la presente invención, la presentación en 2D y la presentación en 3D son compatibles, con el fin de visualizar la imagen decodificada en un modo de 2D o de 3D.

En la solución técnica dada a conocer en la forma de realización del sistema de la presente invención, la imagen de presentación se genera adquiriendo las imágenes en color/escala de grises de múltiples puntos de visualización del objeto de vídeo y la imagen en color/escala de grises del material de presentación, la imagen de presentación soporta el modo de presentación visual en 3D y la imagen de presentación se visualiza en el modo en 3D, con el fin de resolver el problema de que solamente la presentación de vídeo en 2D se soporta en la técnica anterior, con lo que se consigue la presentación de vídeo en 3D.

Los expertos en esta materia pueden entender que la totalidad o parte de las etapas del método según la forma de realización de la presente invención se pueden poner en práctica por un programa que proporcione instrucciones al hardware pertinente. El programa informático se puede memorizar en un medio de memorización legible por ordenador.

Las anteriores descripciones son meramente formas de realización de la presente invención, pero no están previstas para limitar el alcance de protección de la presente invención. Diversas variaciones o sustituciones realizadas por expertos en esta técnica, sin desviarse por ello del alcance técnico de la presente invención, caen dentro del alcance de protección de la presente invención según se define por las reivindicaciones adjuntas. Por lo tanto, el alcance de protección de la presente invención está sujeto a las reivindicaciones adjuntas.

5

REIVINDICACIONES

1. Un método de procesamiento de vídeo que comprende:

5 la adquisición (121) de imágenes en color/escala de grises de al menos dos puntos de visualización de un objeto de vídeo a presentar y de una imagen en color/escala de grises de al menos un punto de visualización de al menos un material de presentación, en donde el material de presentación es un contenido a presentar que se genera por un ordenador y

10 la sintetización (122) de las imágenes en color/escala de grises de los al menos dos puntos de visualización del objeto de vídeo y de la imagen en color/escala de grises del al menos un punto de visualización del al menos un material de presentación para adquirir una imagen de presentación;

15 caracterizado porque,

en donde la adquisición de las imágenes en color/escala de grises de los al menos dos puntos de visualización del objeto de vídeo a presentar, comprende:

20 la adquisición, utilizando una cámara que adquiere información de profundidad e imagen en color/escala de grises al mismo tiempo, de la información de profundidad en un punto de visualización y al menos una imagen en color/escala de grises del mismo punto de visualización de una escena en la que se encuentra el objeto de vídeo;

25 la realización de una segmentación de vídeo en la imagen en color/escala de grises del mismo punto de visualización en función de la información de profundidad del punto de visualización, de modo que se adquiriera el objeto de vídeo y

la generación de una imagen en color/escala de grises del al menos otro punto de visualización del objeto de vídeo en función de la información de profundidad del punto de visualización y de la imagen en color/escala de grises del mismo punto de visualización.

30 2. El método de procesamiento de vídeo según la reivindicación 1 que comprende, además:

la visualización de la imagen de presentación mediante una presentación de vídeo en 3D o

35 la codificación de la imagen de presentación y el envío (123) de la imagen de presentación codificada.

3. El método de procesamiento de vídeo según la reivindicación 1 o 2 que comprende, además:

40 la codificación, por un extremo emisor, de las imágenes en color/escala de grises de los al menos dos puntos de visualización del objeto de vídeo a presentarse y la imagen en color/escala de grises del al menos un punto de visualización del al menos un material de presentación;

el envío, por el extremo emisor, de las imágenes codificadas a un extremo receptor y

45 en donde, el extremo receptor está configurado para: recibir las imágenes codificadas, decodificar las imágenes codificadas para obtener imágenes en color/escala de grises de al menos dos puntos de visualización de un objeto de vídeo a presentarse y de una imagen en color/escala de grises de al menos un punto de visualización del al menos un material de presentación, para sintetizar las imágenes en color/escala de grises de los al menos dos puntos de visualización del objeto de vídeo y de la imagen en color/escala de grises de al menos un punto de visualización del al menos un material de presentación para adquirir una imagen de presentación y para visualizar la imagen de presentación mediante una presentación de vídeo en 3D.

4. El método de procesamiento de vídeo según una de las reivindicaciones 1 o 2 que comprende, además:

55 la adquisición, por un extremo emisor, de información de posición e información de orden de control del objeto de vídeo, en donde la información de posición se adquiere por intermedio de una imagen de profundidad, y configurada para controlar la posición del objeto de vídeo en el material de presentación, estando la información de orden de control configurada para controlar el contenido del material de presentación y la posición del objeto de vídeo en el material de presentación, en función de la imagen de profundidad, siendo reconocido el gesto desde el objeto de vídeo y la información de reconocimiento de gesto convertida en la información de orden de control, en donde

60 la sintetización de las imágenes en color/escala de grises de los al menos dos puntos de visualización del objeto de vídeo y de la imagen en color/escala de grises del al menos un punto de visualización de al menos un material de presentación para adquirir la imagen de presentación comprende:

65 la sintetización, por el extremo emisor, de las imágenes en color/escala de grises de los al menos dos puntos de visualización del objeto de vídeo y de la imagen en color/escala de grises del al menos un punto de visualización del al

menos un material de presentación para adquirir la imagen de presentación en función de la información de posición y de la información de orden de control.

5. El método de procesamiento de vídeo según la reivindicación 3 que comprende, además:

la adquisición, por extremo receptor, de información de posición y de información de orden de control del objeto de vídeo, en donde la información de posición se adquiere por intermedio de una imagen de profundidad y se configura para controlar la posición del objeto de vídeo en el material de presentación, estando la información de orden de control configurada para controlar el contenido del material de presentación y la posición del objeto de vídeo en el material de presentación, en función de la imagen de profundidad, siendo reconocido el gesto del objeto de vídeo y la información de reconocimiento de gesto se convierte en la información de orden de control, en donde

la sintetización de las imágenes en color/escala de grises de los al menos dos puntos de visualización del objeto de vídeo y de la imagen en color/escala de grises del al menos un punto de visualización del al menos un material de presentación para adquirir la imagen de presentación comprende:

la sintetización, por el extremo receptor, de las imágenes en color/escala de grises de los al menos dos puntos de visualización del objeto de vídeo y de la imagen en color/escala de grises del al menos un punto de visualización del al menos un material de presentación para adquirir la imagen de presentación en función de la información de posición y de la información de orden de control.

6. Un dispositivo de procesamiento de vídeo que comprende:

una unidad de adquisición y de procesamiento de imágenes (191), configurada para adquirir imágenes en color/escala de grises de al menos dos puntos de visualización de un objeto de vídeo a presentar y de una imagen en color/escala de grises de al menos un punto de visualización del al menos un material de presentación y sintetizar las imágenes en color/escala de grises de los al menos dos puntos de visualización del objeto de vídeo y de la imagen en color/escala de grises del al menos un punto de visualización del al menos un material de presentación para adquirir una imagen de presentación;

caracterizado porque,

en donde la unidad de adquisición y procesamiento de imágenes comprende:

un módulo de adquisición de imágenes, configurado para adquirir, utilizando una sola cámara que puede adquirir información de profundidad e imagen en color/escala de grises al mismo tiempo, información de profundidad en un punto de visualización y al menos una imagen en color/escala de grises del mismo punto de visualización de una escena en la que se encuentra el objeto de vídeo;

un módulo de segmentación de vídeo, configurado para realizar la segmentación de vídeo en la imagen en color/escala de grises del mismo punto de visualización en función de la información de profundidad del punto de visualización para adquirir el objeto de vídeo;

un módulo de generación de imágenes con puntos de visualización múltiples, configurado para generar una imagen en color/escala de grises de al menos otro punto de visualización del objeto de vídeo en función de la información de profundidad del punto de visualización y de la imagen en color/escala de grises del mismo punto de visualización y

un módulo de síntesis, configurado para sintetizar imágenes en color/escala de grises de los al menos dos puntos de visualización del objeto de vídeo y la imagen en color/escala de grises del al menos un punto de visualización del al menos un material de presentación con el fin de adquirir la imagen de presentación.

7. El dispositivo de procesamiento de vídeo según la reivindicación 6 que comprende, además:

una unidad de presentación visual (192), configurada para visualizar la imagen de presentación mediante una presentación de vídeo en 3D.

8. El dispositivo de procesamiento de vídeo según la reivindicación 6 que comprende, además:

una unidad de codificación configurada para codificar la imagen de presentación para obtener una imagen codificada y una unidad de envío configurada para enviar la imagen codificada a un dispositivo de recepción de vídeo.

9. El dispositivo de procesamiento de vídeo según la reivindicación 7 que comprende, además:

una unidad de recepción, configurada para recibir una imagen codificada desde un dispositivo de envío de vídeo y

una unidad de decodificación, configurada para decodificar la imagen codificada para adquirir la imagen de presentación en donde la imagen de presentación se visualiza por una presentación de vídeo en 3D por la unidad de presentación visual.

5 **10.** Un sistema de comunicación a través de vídeo, que comprende el dispositivo de procesamiento de vídeo según una de las reivindicaciones 8 a 9 y un dispositivo de recepción de vídeo, en donde el dispositivo de recepción de vídeo comprende:

una unidad de recepción (210), configurada para recibir la imagen codificada;

10 una unidad de decodificación (211), configurada para decodificar la imagen codificada para adquirir la imagen de presentación y

15 una unidad de presentación visual (213), configurada para visualizar la imagen de presentación mediante una presentación de vídeo en 3D.

11. El sistema de comunicación a través de vídeo según la reivindicación 10, en donde el dispositivo de procesamiento de vídeo comprende, además:

20 una unidad de decodificación, configurada para decodificar la imagen codificada para obtener la imagen de presentación decodificada y

una unidad de presentación visual, configurada para realizar la visualización en 3D de la imagen de presentación.

25 **12.** El sistema de comunicación a través de vídeo según la reivindicación 10 o 11, en donde el dispositivo de recepción de vídeo comprende, además:

30 una unidad de adquisición y procesamiento de imágenes (221), configurada para adquirir las imágenes en color/escala de grises de los al menos dos puntos de visualización del objeto de vídeo a presentar y la imagen en color/escala de grises del al menos un punto de visualización del al menos un material de presentación y para sintetizar las imágenes en color/escala de grises de los al menos dos puntos de visualización del objeto de vídeo y la imagen en color/escala de grises del al menos un punto de visualización del al menos un material de presentación para adquirir la imagen de presentación y

35 una unidad de codificación (222), configurada para codificar la imagen de presentación para obtener la imagen codificada y para enviar la imagen codificada.

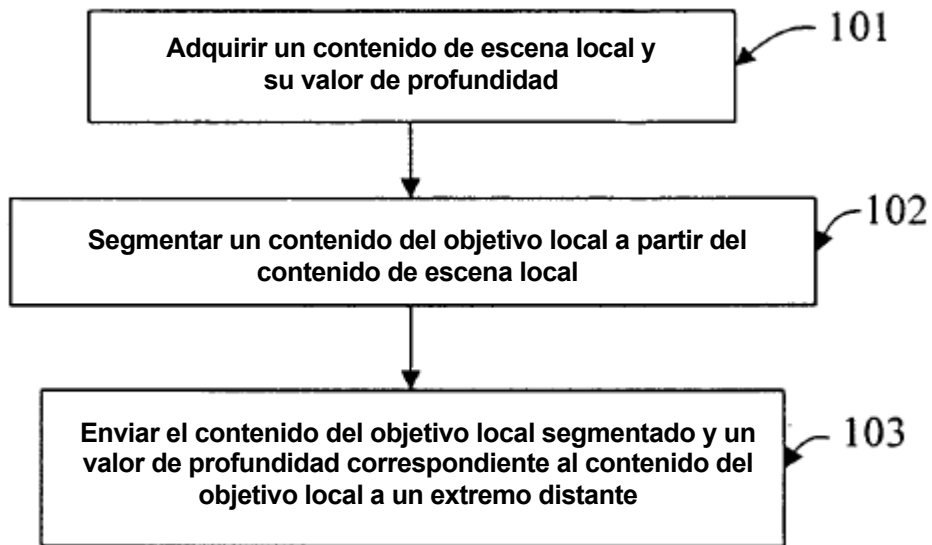


Figura 1

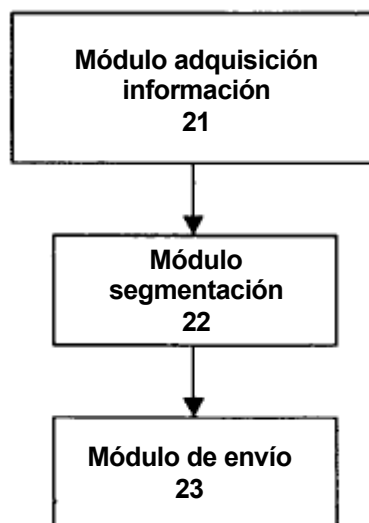


Figura 2

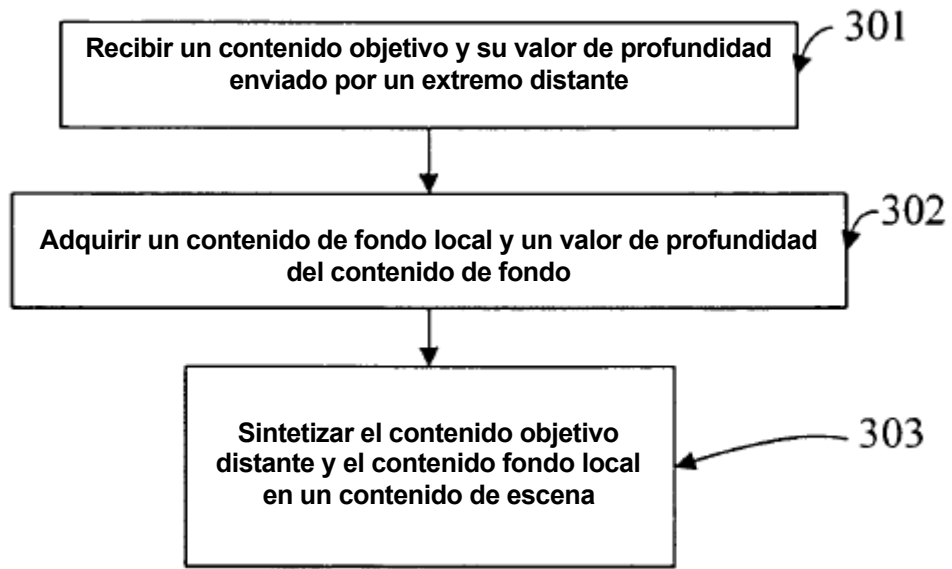


Figura 3

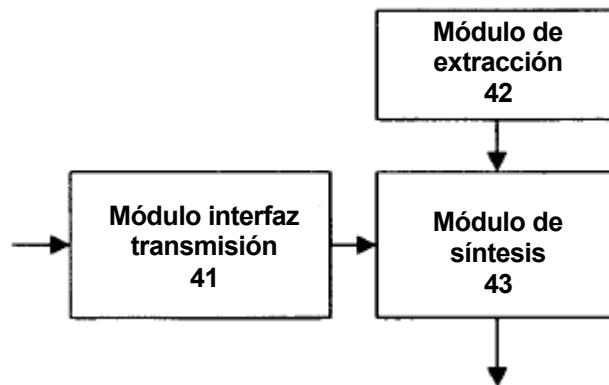


Figura 4

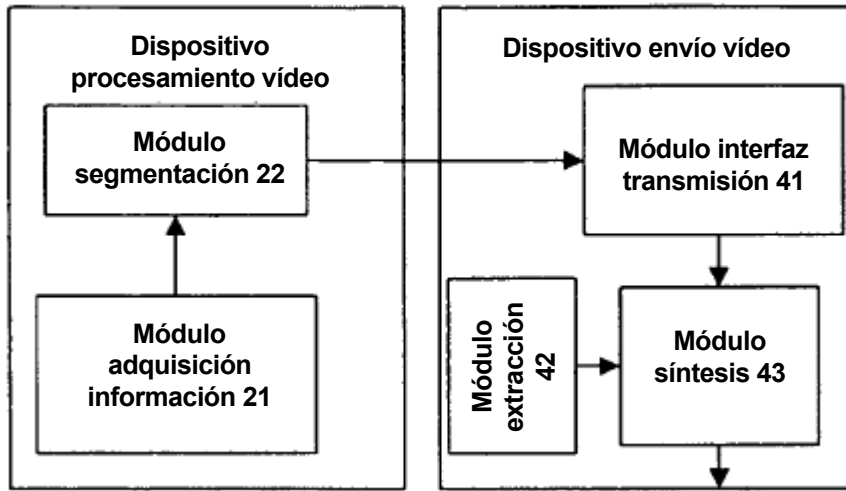


Figura 5

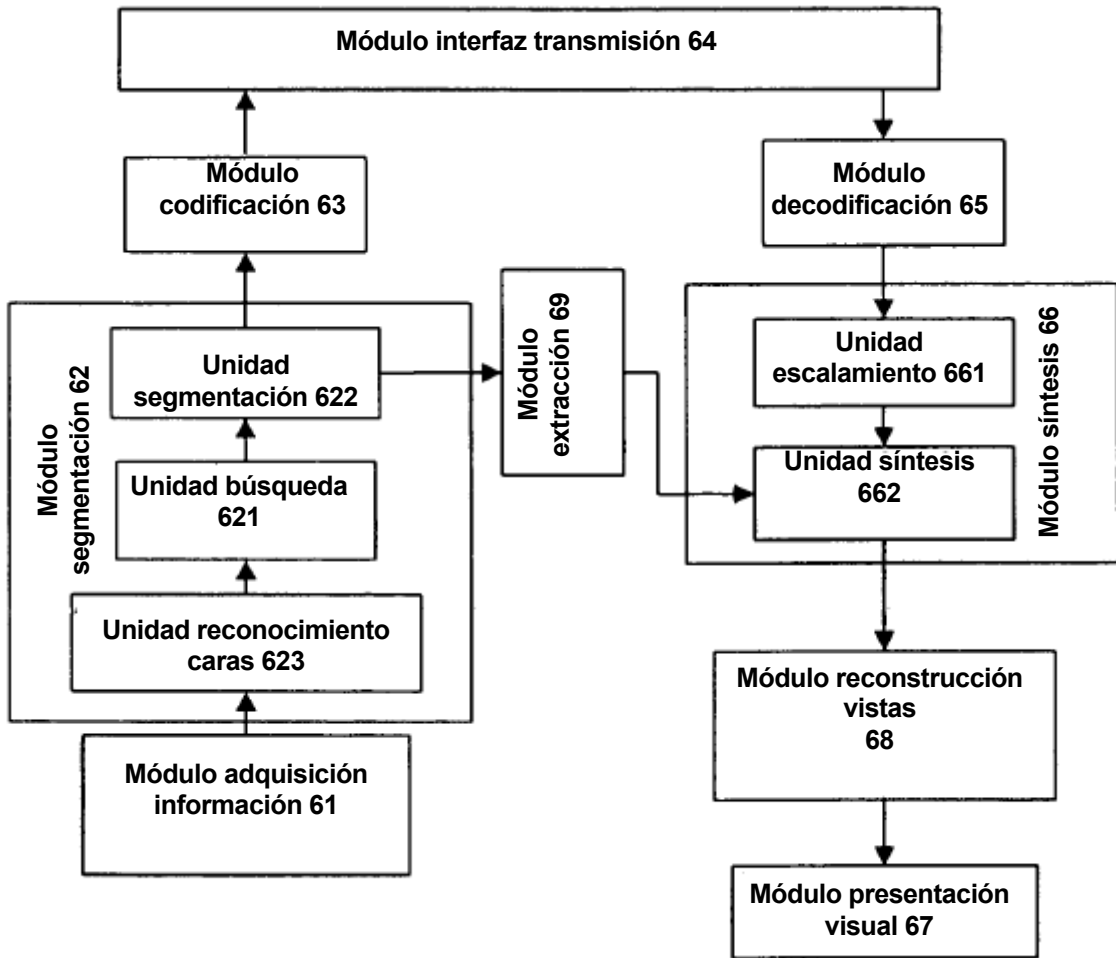


Figura 6

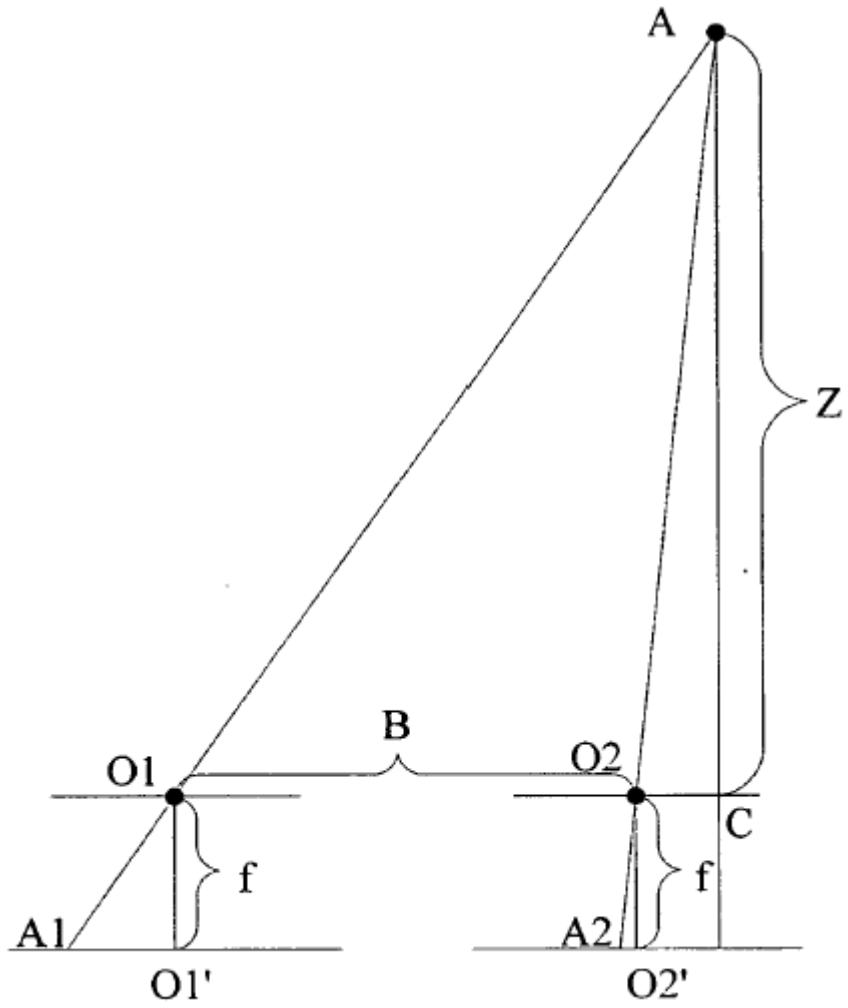


Figura 7

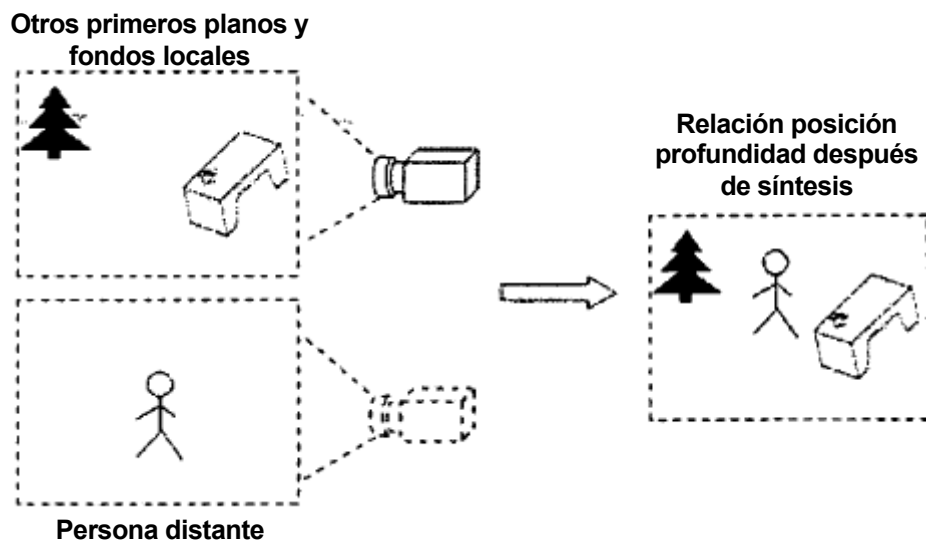
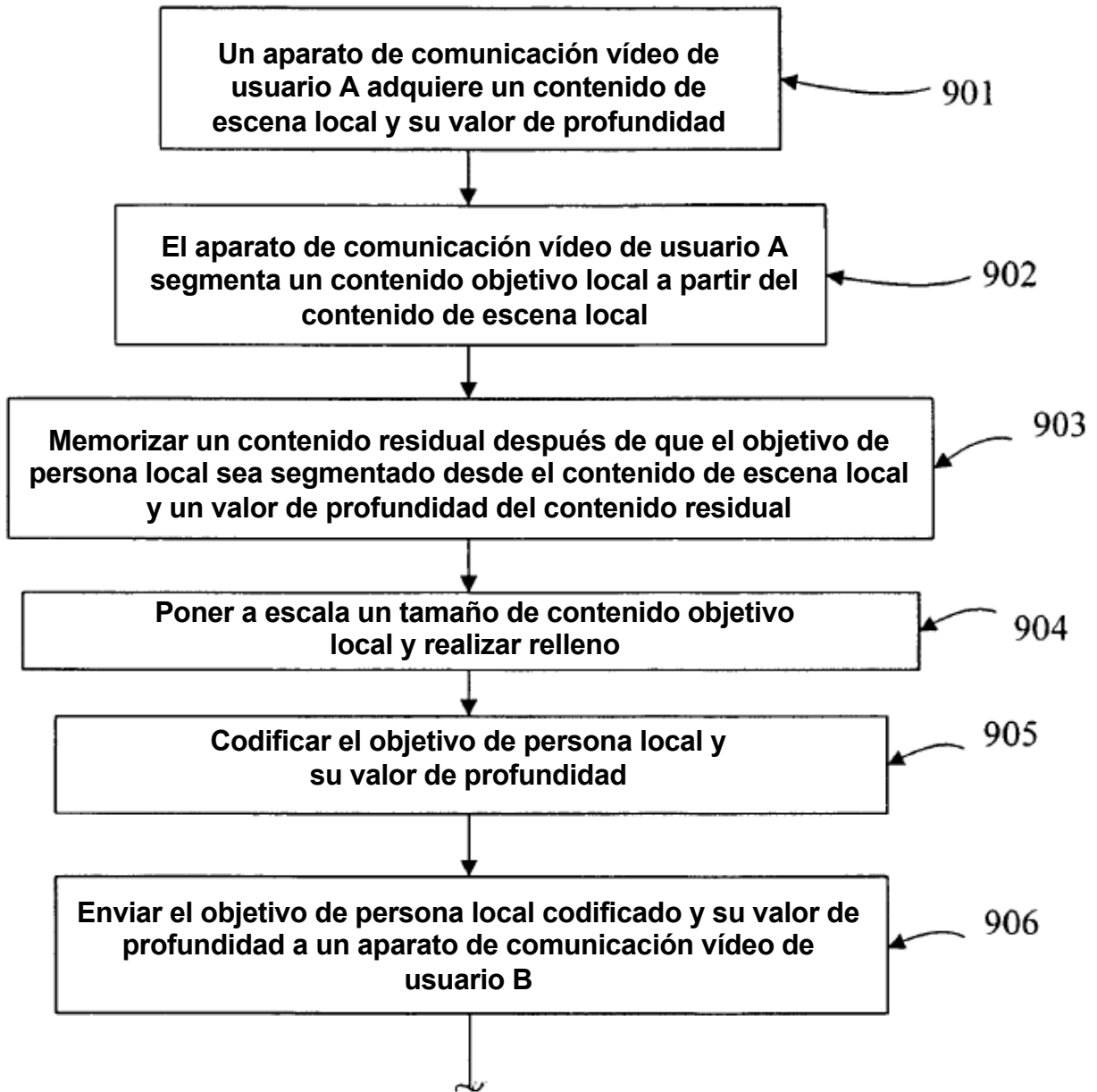


Figura 8



A
FIG. 9B

Figura 9A

CONT. DESDE
FIG. 9A

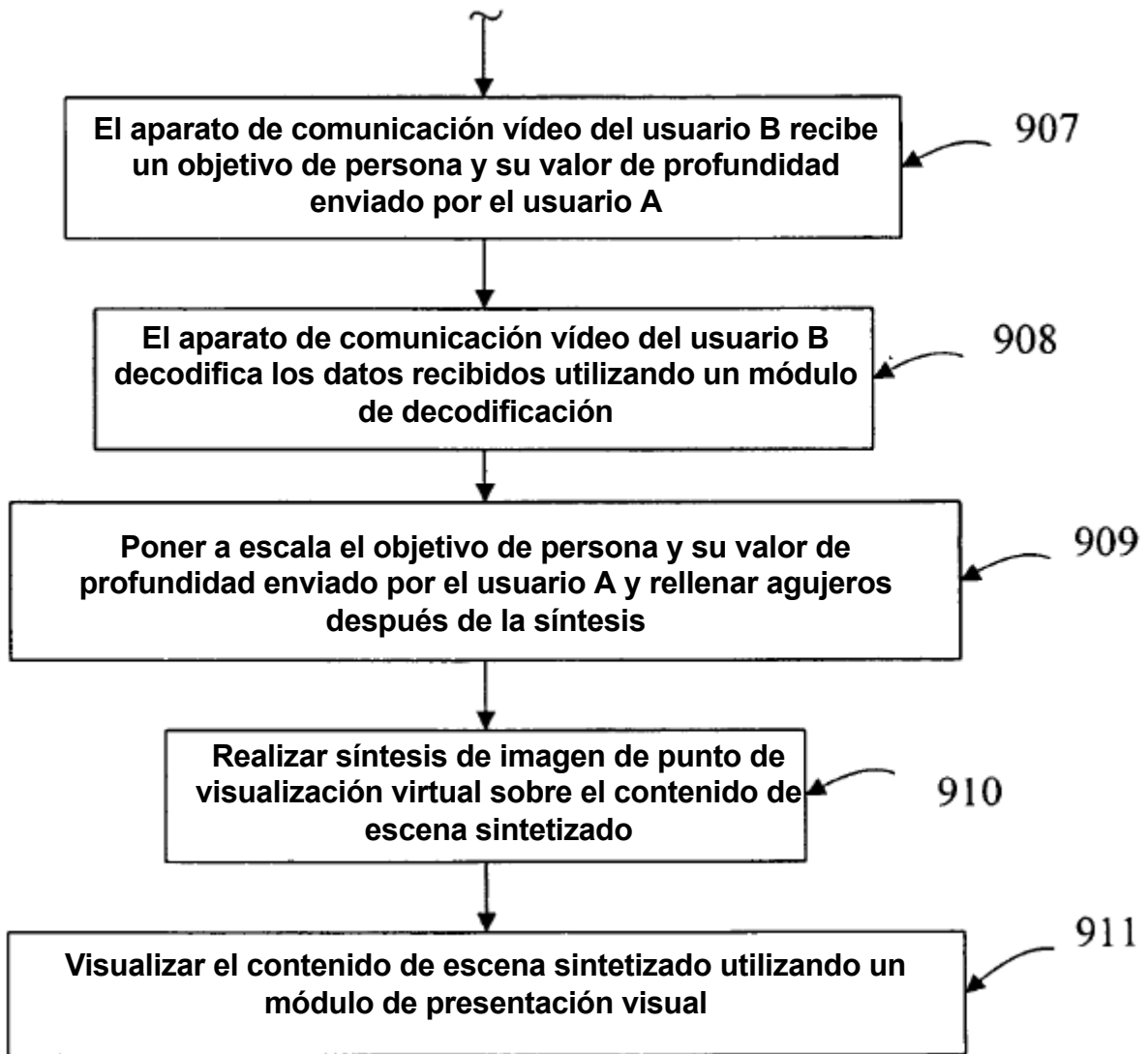


Figura 9B

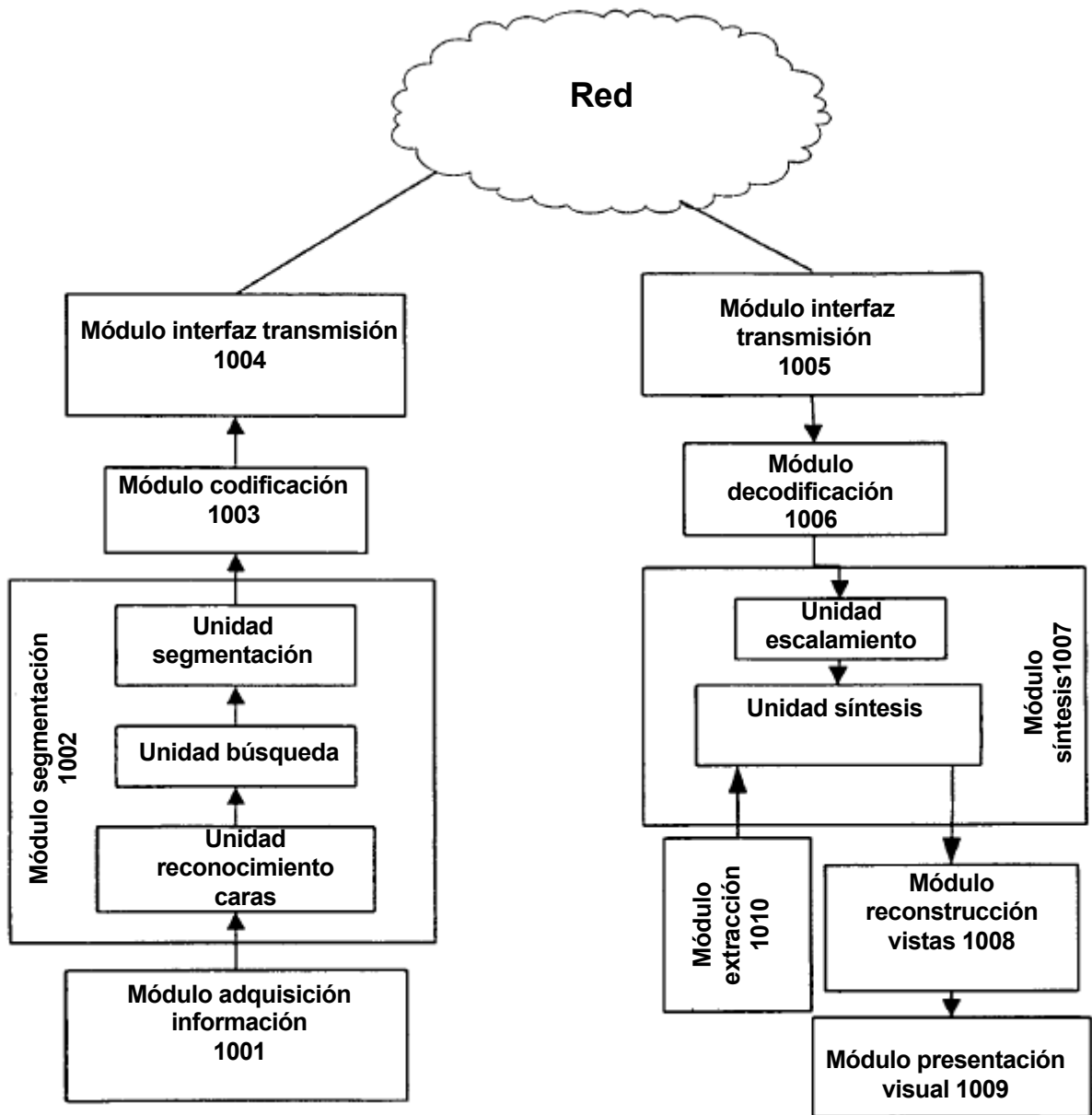


Figura 10

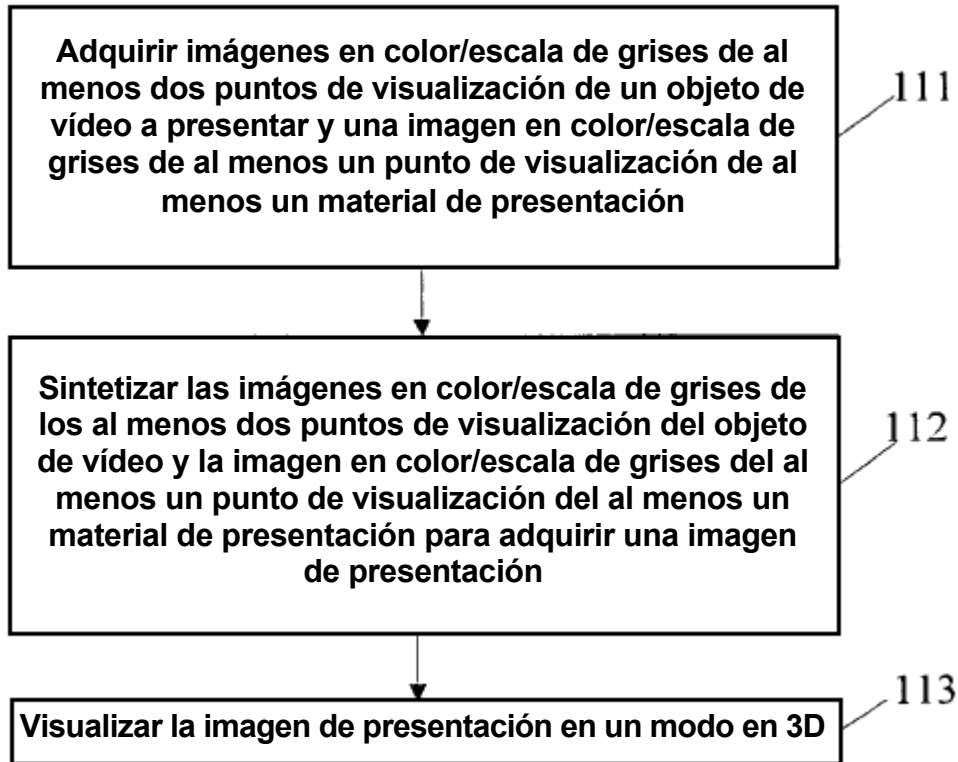


Figura 11

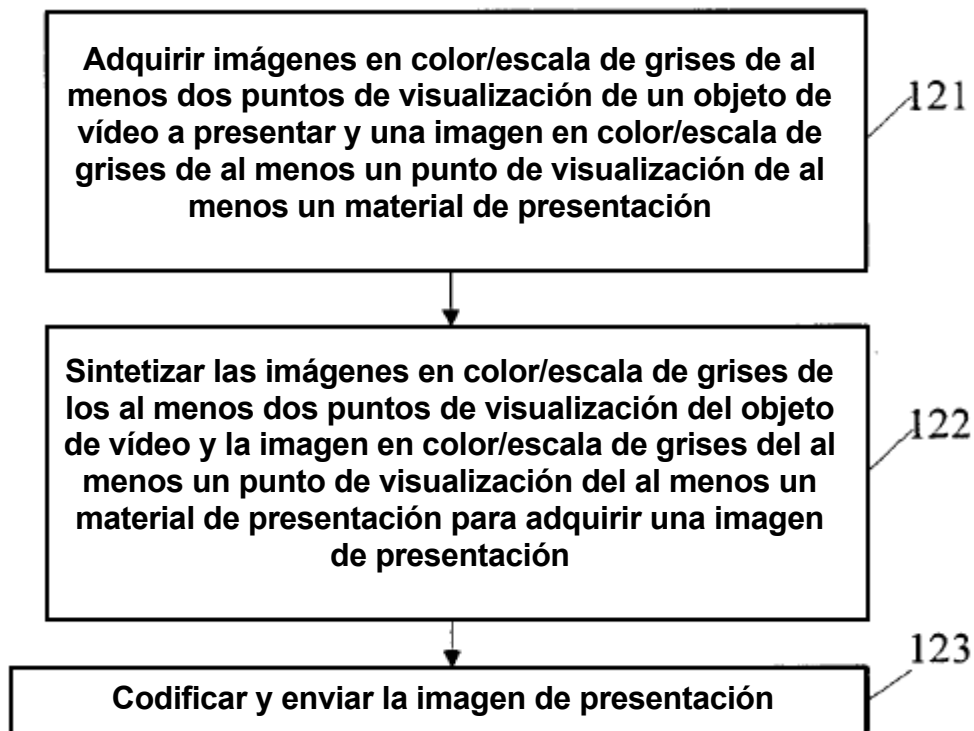


Figura 12

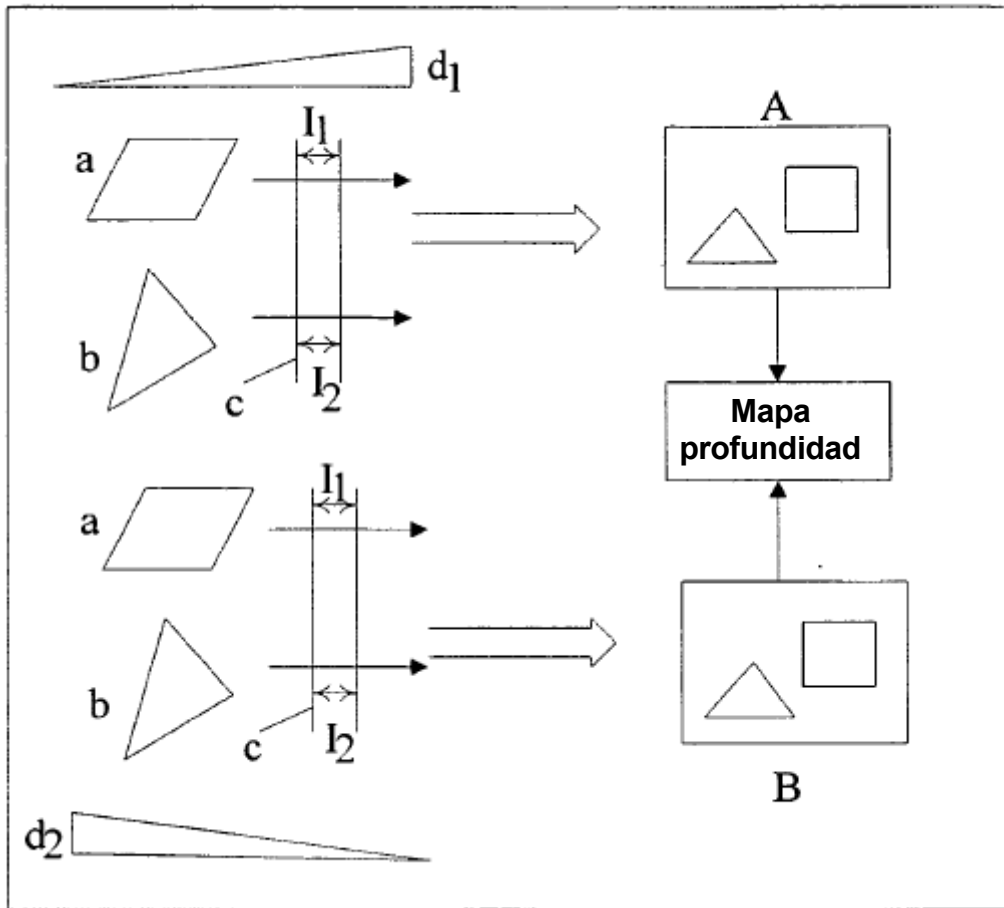


Figura 13

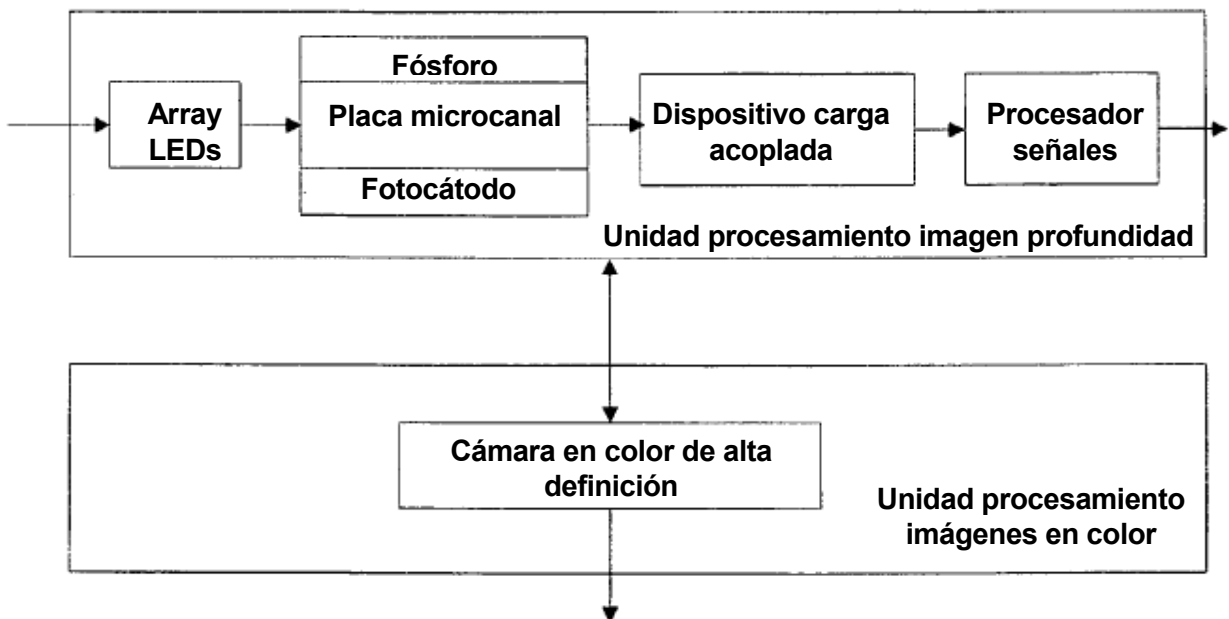


Figura 14

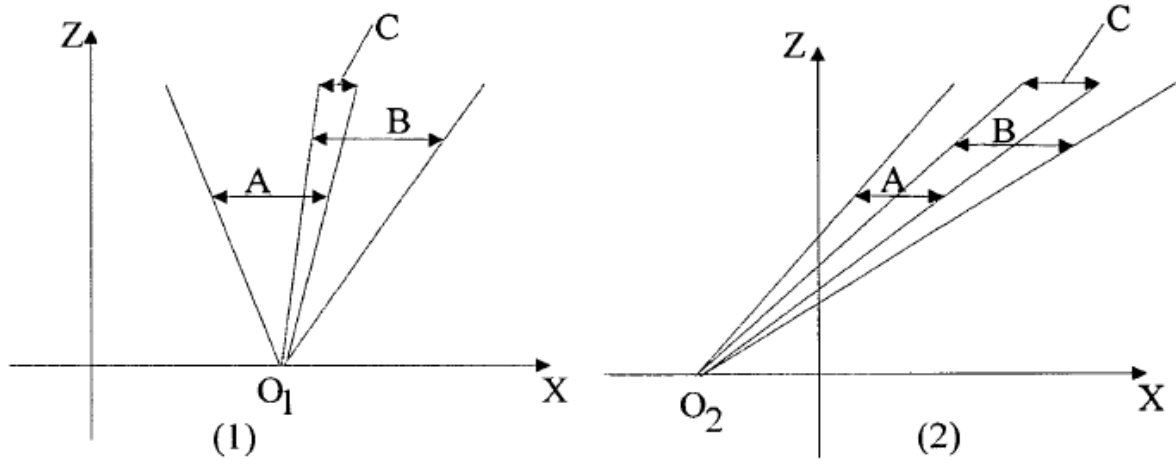


Figura 15

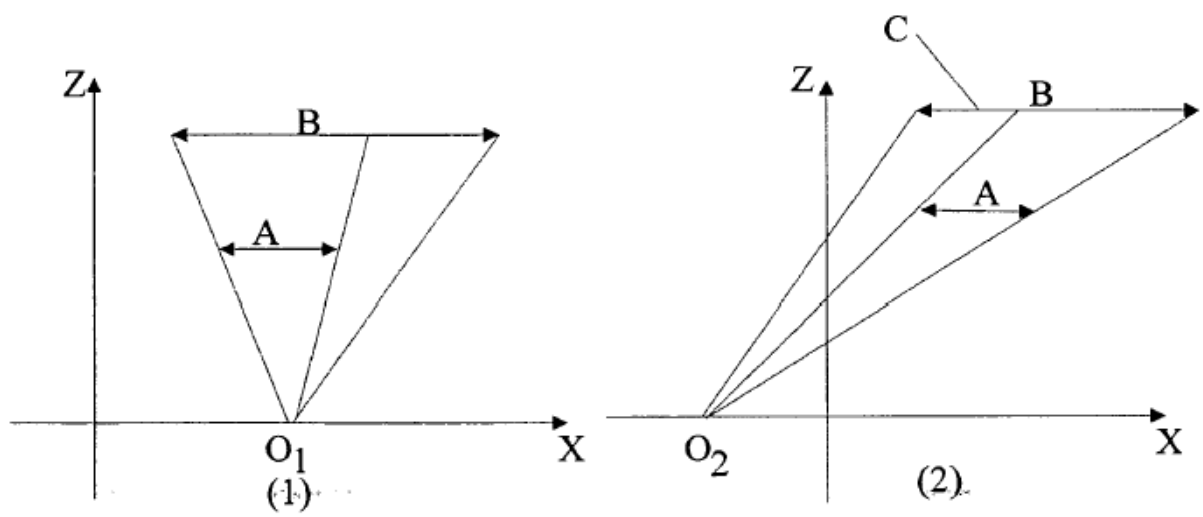


Figura 16

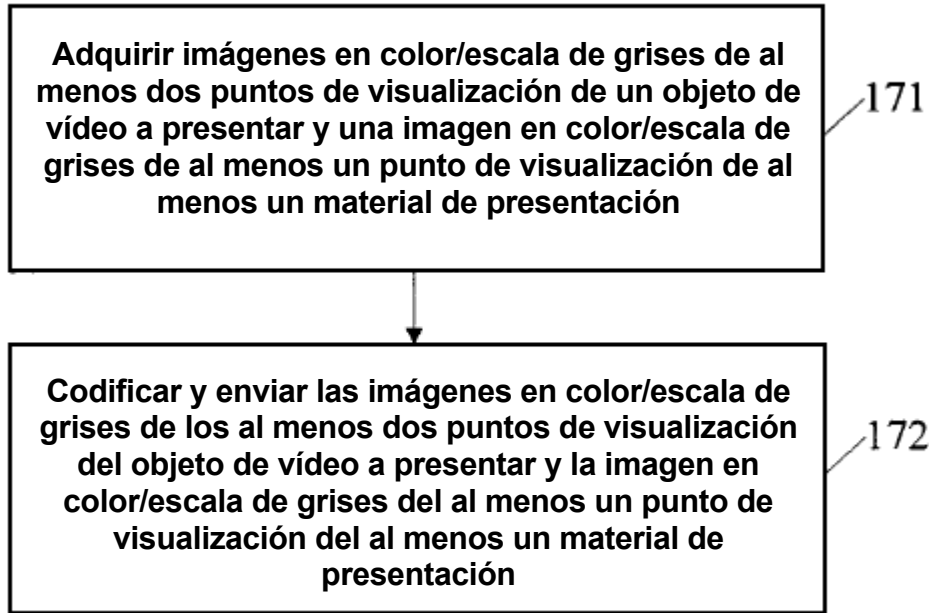


Figura 17

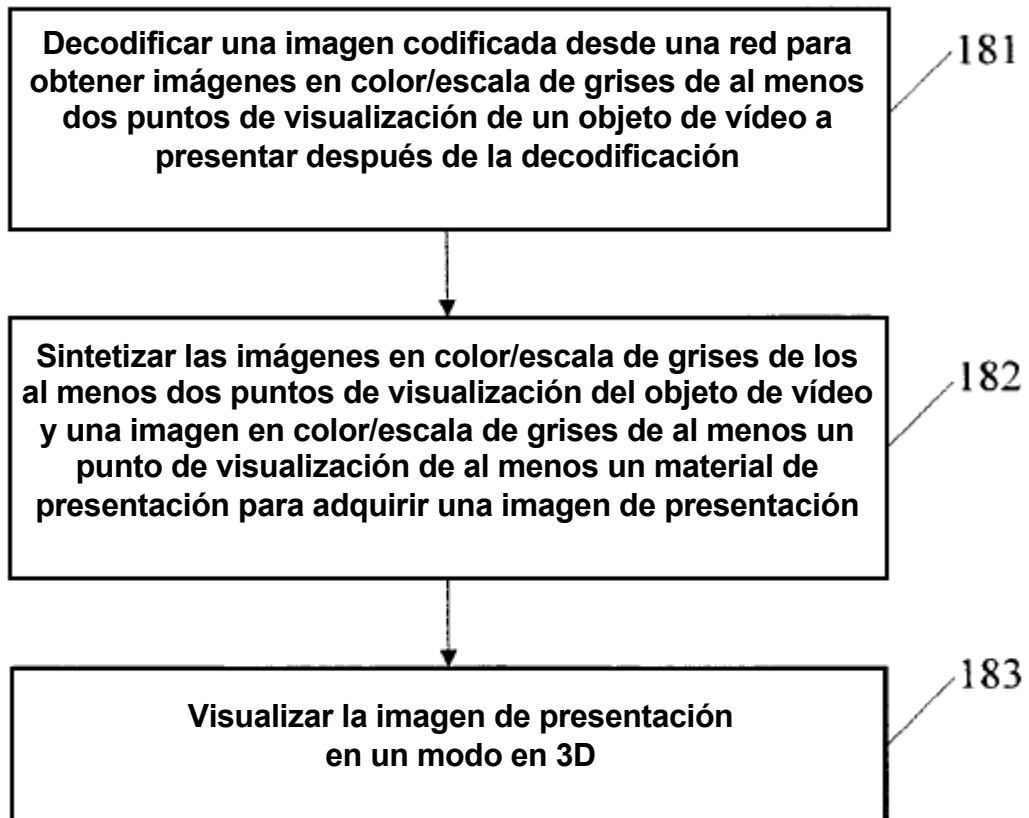


Figura 18

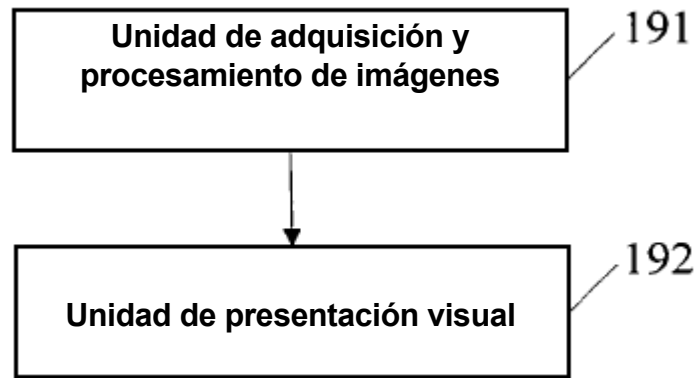


Figura 19

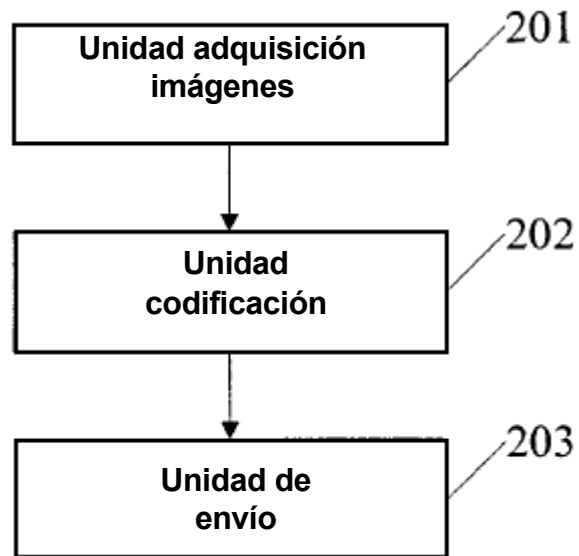


Figura 20

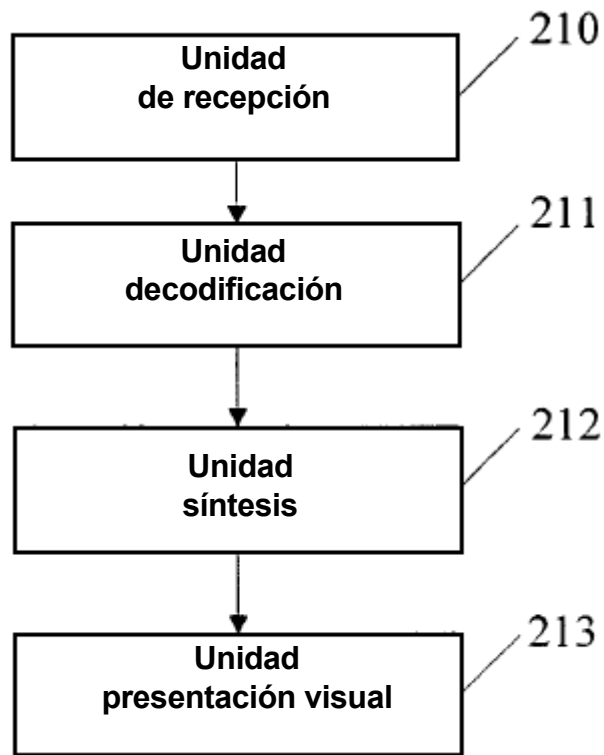


Figura 21

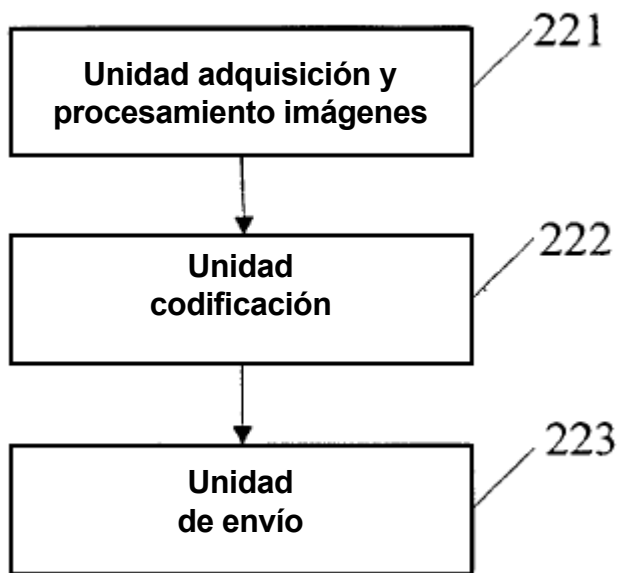


Figura 22

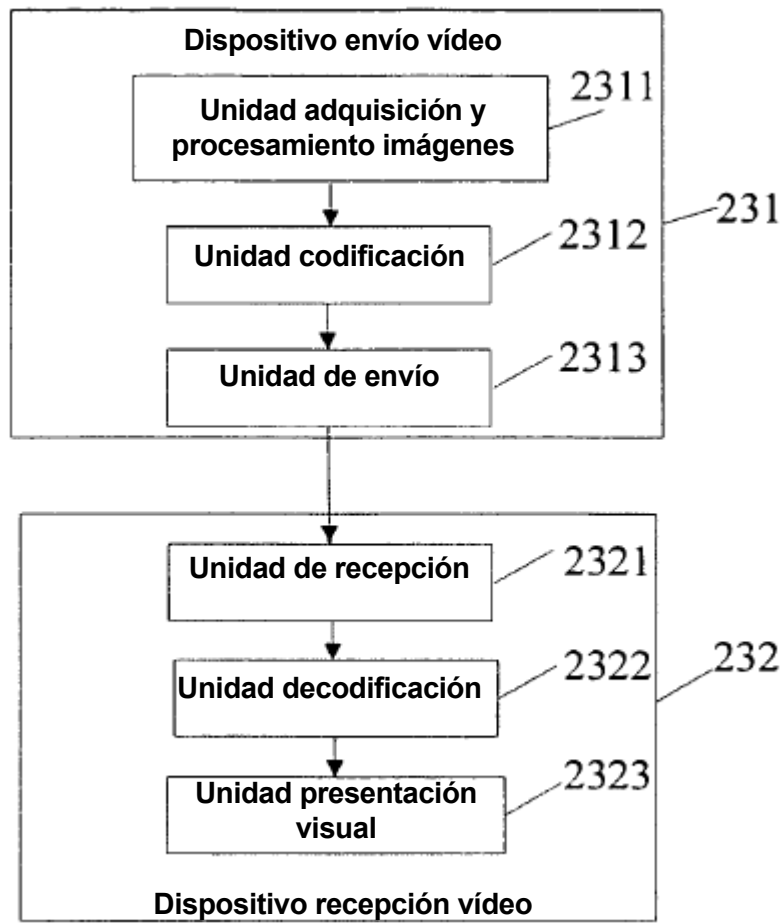


Figura 23