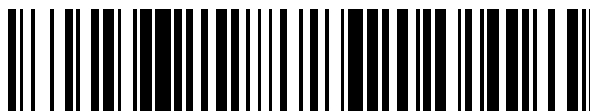


19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 703 062**

51 Int. Cl.:

H04N 19/58	(2014.01)
H04N 13/00	(2008.01)
H04N 19/176	(2014.01)
H04N 19/577	(2014.01)
H04N 19/597	(2014.01)
H04N 19/70	(2014.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

86 Fecha de presentación y número de la solicitud internacional: **10.01.2014 PCT/US2014/011119**

87 Fecha y número de publicación internacional: **17.07.2014 WO14110426**

96 Fecha de presentación y número de la solicitud europea: **10.01.2014 E 14701884 (0)**

97 Fecha y número de publicación de la concesión europea: **26.09.2018 EP 2944086**

54 Título: **Síntesis de visualización en vídeo 3D**

30 Prioridad:

10.01.2013 US 201361751211 P
09.01.2014 US 201414151586

45 Fecha de publicación y mención en BOPI de la traducción de la patente:
06.03.2019

73 Titular/es:

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

72 Inventor/es:

CHEN, YING;
WANG, YE-KUI y
ZHANG, LI

74 Agente/Representante:

FORTEA LAGUNA, Juan José

ES 2 703 062 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

Síntesis de visualización en vídeo 3D

- 5 **[0001]** Esta solicitud reivindica la prioridad de la solicitud provisional de Estados Unidos n.º 61/751,211 presentada el 10 de enero de 2013.

CAMPO TÉCNICO

- 10 **[0002]** Esta divulgación se refiere a la codificación de vídeo.

ANTECEDENTES

- 15 **[0003]** Las capacidades de vídeo digital pueden incorporarse en una amplia gama de dispositivos, incluyendo televisores digitales, sistemas de difusión directa digital, sistemas de difusión inalámbrica, asistentes digitales personales (PDA), ordenadores portátiles o de escritorio, ordenadores de tableta, lectores de libros electrónicos, cámaras digitales, dispositivos de grabación digital, reproductores de medios digitales, dispositivos de videojuegos, consolas de videojuegos, teléfonos celulares o de radio por satélite, los denominados "teléfonos inteligentes", dispositivos de videoconferencia, dispositivos de transmisión continua de vídeo y similares. Los dispositivos de vídeo digitales implementan técnicas de compresión de vídeo, tales como las descritas en las normas definidas por MPEG-2, MPEG-4, ITU-T H.263, ITU-T H.264/MPEG-4, parte 10, codificación de vídeo avanzada (AVC), la norma de codificación de vídeo de alta eficiencia (HEVC) actualmente en desarrollo y las extensiones de dichas normas. Los dispositivos de vídeo pueden transmitir, recibir, codificar, descodificar y/o almacenar información de vídeo digital con más eficiencia implementando dichas técnicas de compresión de vídeo.

- 25 **[0004]** Las técnicas de compresión de vídeo realizan la predicción espacial (intra-imagen) y/o la predicción temporal (entre imágenes) para reducir o eliminar la redundancia intrínseca en las secuencias de vídeo. Para la codificación de vídeo basada en bloques, un fragmento de vídeo (por ejemplo, una imagen o una parte de una imagen) puede ser dividida en bloques de vídeo, que también pueden ser denominados bloques arbolados, unidades de codificación (CU) y/o nodos de codificación. Los bloques de vídeo en un fragmento intracodificado (I) de una imagen se codifican utilizando predicción espacial con respecto a muestras de referencia en bloques contiguos en la misma imagen. Los bloques de vídeo en un fragmento intercodificado (P o B) de una imagen pueden usar la predicción espacial con respecto a muestras de referencia en bloques vecinos en la misma imagen o la predicción temporal con respecto a muestras de referencia en otras imágenes de referencia.

- 35 **[0005]** La predicción espacial o temporal da como resultado un bloque predictivo para un bloque que se va a codificar. Los datos residuales representan diferencias de píxeles entre el bloque original que se va a codificar y el bloque predictivo. Un bloque intercodificado se codifica de acuerdo con un vector de movimiento que apunta a un bloque de muestras de referencia que forman el bloque predictivo, y los datos residuales que indican la diferencia entre el bloque codificado y el bloque predictivo. Un bloque intracodificado se codifica de acuerdo con un modo de intracodificación y los datos residuales. Para una mayor compresión, los datos residuales pueden transformarse desde el dominio espacial a un dominio de transformada, dando como resultado unos coeficientes de transformada residuales, que posteriormente se pueden cuantificar. Los coeficientes de transformada cuantificados, dispuestos inicialmente en una matriz bidimensional, pueden examinarse con el fin de generar un vector unidimensional de coeficientes de transformada, y puede aplicarse codificación por entropía para lograr aún más compresión.

[0006] Los documentos relevantes de la técnica anterior son:

- 50 - ZHAO X ET AL: "3D-CE1.a: Modo de predicción de síntesis de visualización generalizada (GVSP)", 1. REUNIÓN DE JCT-3V; 101. REUNIÓN DE MPEG; 16-7-2012 - 20-7-2012; ESTOCOLMO; (EQUIPO DE COLABORACIÓN CONJUNTA EN DESARROLLO DE EXTENSIONES DE CODIFICACIÓN DE VÍDEO EN 3D DE ISO/IEC JTC1/SC29/WG11 E ITU-T SG.16); URL: HTTP://PHENIX.INT-EVRY.FR/JCT2/, n.º JCT3V-A0103, 11 de julio de 2012 (2012-07-11), y,
- 55 - US 2012/314027 A1 (TIAN DONG [US] ET AL) 13 de diciembre de 2012(2012-12-13).

SUMARIO

- 60 **[0007]** La presente invención se expone en las reivindicaciones.

En general, esta divulgación describe técnicas para codificar información relacionada con la predicción de síntesis de visualización. La predicción de síntesis de visualización en general se refiere a una técnica de codificación de vídeo para la codificación de vídeo de múltiples visualizaciones en la cual un componente de visualización sintetizada se usa como imagen de referencia cuando se codifica una imagen de una visualización usando predicción entre visualizaciones. Las técnicas de esta divulgación incluyen determinar si aplicar la predicción de

síntesis de visualización para un bloque particular de datos de vídeo basándose en una indicación proporcionada en el flujo de bits para imágenes de referencia entre visualizaciones.

5 [0008] En un ejemplo, un procedimiento para descodificar datos de vídeo incluye determinar si un índice de referencia para un bloque actual corresponde a una imagen de referencia entre visualizaciones, y cuando el índice de referencia para el bloque actual corresponde a la imagen de referencia entre visualizaciones, obtener, a partir de un flujo de bits codificado, datos que indican un modo de predicción de síntesis de visualización (VSP) del bloque actual, en el que el modo de VSP para el índice de referencia indica si el bloque actual se predice con la predicción de síntesis de visualización a partir de la imagen de referencia entre visualizaciones.

10 [0009] En otro ejemplo, un aparato para descodificar datos de vídeo incluye una memoria que almacena datos de vídeo que incluyen un bloque actual y uno o más procesadores configurados para determinar si un índice de referencia para el bloque actual corresponde a una imagen de referencia entre visualizaciones, y cuando el índice de referencia para el bloque actual corresponde a la imagen de referencia entre visualizaciones, obtener, a partir de un flujo de bits codificado, datos que indiquen un modo de predicción de síntesis de visualización (VSP) del bloque actual, en el que el modo VSP para el índice de referencia indica si se predice el bloque actual con visualización de la predicción de la síntesis a partir de la imagen de referencia entre visualizaciones.

15 [0010] En otro ejemplo, un procedimiento para codificar datos de vídeo incluye determinar si un índice de referencia para un bloque actual corresponde a una imagen de referencia entre visualizaciones, cuando el índice de referencia para el bloque actual corresponde a la imagen de referencia entre visualizaciones, generar datos que indican un modo de predicción de síntesis de visualización (VSP) del bloque actual, en el que el modo VSP para el índice de referencia indica si el bloque actual se predice con la predicción de síntesis de visualización a partir de la imagen de referencia entre visualizaciones, y codificar los datos que indican el modo VSP en un flujo de bits.

20 [0011] En otro ejemplo, un aparato para codificar datos de vídeo incluye una memoria que almacena datos de vídeo que incluyen un bloque actual, y uno o más procesadores configurados para determinar si un índice de referencia para un bloque actual corresponde a una imagen de referencia entre visualizaciones, cuando el índice de referencia para el bloque actual corresponde a la imagen de referencia entre visualizaciones, generar datos que indiquen un modo de predicción de síntesis de visualización (VSP) del bloque actual, en el que el modo de VSP para el índice de referencia indica si el bloque actual se predice con la predicción de síntesis de visualización a partir de la imagen de referencia entre visualizaciones, y codificar los datos que indican el modo VSP en un flujo de bits.

25 [0012] En otro ejemplo, un medio legible por ordenador no transitorio tiene instrucciones almacenadas en el mismo que, cuando se ejecutan, hacen que uno o más procesadores determinen si un índice de referencia para un bloque actual corresponde a una imagen de referencia entre visualizaciones, y cuando el índice de referencia para el bloque actual corresponde a la imagen de referencia entre visualizaciones, datos de codificación que indican un modo de predicción de síntesis de visualización (VSP) del bloque actual, en el que el modo de VSP para el índice de referencia indica si el bloque actual se predice con predicción de síntesis de visualización a partir de la imagen de referencia entre visualizaciones.

30 [0013] En otro ejemplo, un aparato para codificar datos de vídeo incluye medios para determinar si un índice de referencia para un bloque actual corresponde a una imagen de referencia entre visualizaciones, y cuando el índice de referencia para el bloque actual corresponde a la imagen de referencia entre visualizaciones, medios para codificar datos que indiquen un modo de predicción de síntesis de visualización (VSP) del bloque actual, en el que el modo de VSP para el índice de referencia indica si el bloque actual se predice con la predicción de síntesis de visualización a partir de la imagen de referencia entre visualizaciones.

35 [0014] Los detalles de uno o más ejemplos se exponen en los dibujos adjuntos y en la siguiente descripción. Otras características, objetivos y ventajas resultarán evidentes a partir de la descripción y de los dibujos, y a partir de las reivindicaciones.

40 **BREVE DESCRIPCIÓN DE LOS DIBUJOS**

45 [0015]

50 La FIG. 1 es un diagrama de bloques que ilustra un sistema de codificación y descodificación de vídeo de ejemplo que puede utilizar las técnicas descritas en esta divulgación.

55 La FIG. 2 es un diagrama de bloques que ilustra un codificador de vídeo de ejemplo que puede implementar las técnicas descritas en esta divulgación.

60 La FIG. 3 es un diagrama de bloques que ilustra un descodificador de vídeo de ejemplo que puede implementar las técnicas descritas en esta divulgación.

La FIG. 4 es un diagrama conceptual que ilustra la codificación de una secuencia de codificación de vídeo de múltiples visualizaciones (MVC).

La FIG. 5 es un diagrama conceptual que ilustra un patrón de predicción de MVC de ejemplo.

La FIG. 6 es un diagrama de flujo que ilustra un procedimiento de ejemplo para codificar información relacionada con la predicción de síntesis de visualización.

La FIG. 7 es un diagrama de flujo que ilustra otro procedimiento de ejemplo para codificar información relacionada con la predicción de síntesis de visualización.

DESCRIPCIÓN DETALLADA

[0016] En general, esta divulgación describe técnicas para codificar y procesar datos de vídeo de múltiples visualizaciones, por ejemplo, datos de vídeo utilizados para producir un efecto tridimensional (3D). Por lo tanto, los aspectos de esta divulgación pueden estar relacionados con la codificación de vídeo 3D basada en códecs avanzados, incluida la codificación de dos o más visualizaciones de una imagen con mapas de profundidad. En particular, los aspectos de esta divulgación pueden estar relacionados con la señalización de soporte de predicción de síntesis de visualización en un proceso de codificación de vídeo 3D.

[0017] Las técnicas de esta divulgación se pueden aplicar a una variedad de normas de codificación de vídeo diferentes. Por ejemplo, entre otras normas de codificación de vídeo se incluyen ITU-T H.261, ISO/IEC MPEG-1 Visual, ITU-T H.262 o ISO/IEC MPEG-2 Visual, ITU-T H.263, ISO/IEC MPEG-4 Visual e ITU-T H.264 (también conocida como ISO/IEC MPEG-4 AVC, o H.264/AVC), incluyendo sus extensiones de codificación de vídeo escalable (SVC) y de codificación de vídeo de múltiples visualizaciones (MVC). Un reciente borrador conjunto de la MVC se describe en "Advanced vídeo coding for generic audiovisual services" ["Codificación de vídeo avanzada para servicios audiovisuales genéricos"], Recomendación H.264 de la ITU-T, marzo de 2010.

[0018] Además, existe una nueva norma de codificación de vídeo, concretamente la HEVC que está siendo desarrollada por el Equipo de Colaboración Conjunta en Codificación de Vídeo (JCT-VC) del Grupo de Expertos en Codificación de Vídeo (VCEG) de ITU-T y el Grupo de Expertos en Imágenes en Movimiento (MPEG) de ISO/IEC. Un borrador de trabajo (WD) reciente de HEVC y denominado HEVC WD4 en adelante en el presente documento, está disponible en http://phenix.int-evry.fr/jct/doc_end_user/documents/6_Torino/wg11/JCTVC-F803-v8.zip. Otro WD de HEVC, y denominado HEVC WD9 en adelante en el presente documento, está disponible en http://phenix.int-evry.fr/jct/doc_end_user/documents/11_Shanghai/wg11/JCTVC-K1003-v10.zip.

[0019] Las actividades de estandarización recientes en MPEG incluyeron una Convocatoria de propuestas (CfP) para tecnologías de vídeo 3D. Entre las respuestas, las técnicas se seleccionan para ser incluidas en el software de referencia 3D actual. Un posible estándar es la codificación de vídeo 3D basada en H.264/AVC, que puede codificar la textura y la profundidad de cada visualización. Las principales herramientas de codificación para la codificación de vídeo 3D basada en AVC (3D-AVC) (denominadas en el presente documento 3DV WD4) se describen en el documento JCT3V-B1002 ("Proyecto de texto 3D-AVC borrador 4," Equipo de colaboración conjunta en el desarrollo de la extensión de codificación de vídeo 3D de ITU-T SG 16 WP 3 e ISO/IEC JTC 1/SC 29/WG 11, 2.^a reunión: Shanghái, China, 13-19 de octubre de 2012, disponible públicamente en http://phenix.it-sudparis.eu/jct2/doc_end_user/current_document.php?id=456). La última descripción del software de referencia es "Modelo de prueba para codificación de vídeo 3D basada en AVC", Miska M. Hannuksela, ISO/IEC JTC1/SC29/WG11 MPEG2011/N12558, San José, EE. UU., febrero de 2012. El software de referencia más reciente se puede consultar en el presente documento como ATM o 3D-ATM, y está disponible públicamente en <http://mpeg3dv.research.nokia.com/svn/mpeg3dv/trunk/>.

[0020] En algunos ejemplos, las técnicas de esta divulgación pueden aplicarse a una norma de codificación de vídeo 3D (3DV) que se basa en H.264/AVC. En consecuencia, con fines de ilustración, las técnicas de esta divulgación se describen principalmente con respecto a la extensión 3DV de ITU-T H.264/AVC. Sin embargo, debe entenderse que las técnicas pueden aplicarse a otras normas para la codificación de datos de vídeo utilizadas para producir un efecto tridimensional, como una extensión de vídeo 3D (3DV) de la norma HEVC u otra norma de codificación.

[0021] En cualquier caso, de acuerdo con ciertos sistemas de codificación de vídeo, la estimación de movimiento y la compensación de movimiento se pueden utilizar para reducir la redundancia temporal en una secuencia de vídeo, con el fin de lograr la compresión de datos. En este caso, puede generarse un vector de movimiento que identifica un bloque predictivo de datos de vídeo, por ejemplo, un bloque de otra imagen o fragmento de vídeo, que se puede utilizar para predecir los valores del bloque de vídeo actual que se está codificando. Los valores del bloque de vídeo predictivo se restan de los valores del bloque de vídeo actual para producir un bloque de datos residuales. La información de movimiento (por ejemplo, un vector de movimiento, índices de vectores de movimiento, direcciones de predicción u otra información) se comunica desde un codificador de vídeo a un descodificador de vídeo, junto con los datos residuales. El descodificador puede localizar el mismo bloque

predictivo (basándose en el vector de movimiento) y reconstruir el bloque de vídeo codificado mediante la combinación de los datos residuales con los datos del bloque predictivo.

5 **[0022]** Con respecto a la codificación de vídeo 3D, para producir un efecto tridimensional en vídeo, se pueden mostrar dos visualizaciones de una escena, por ejemplo, una visualización del ojo izquierdo y una visualización del ojo derecho, simultáneamente o casi simultáneamente. Dos imágenes de la misma escena, que corresponden a la visualización del ojo izquierdo y la visualización del ojo derecho de la escena, pueden capturarse (o generarse, por ejemplo, como gráficos generados por ordenador) desde posiciones horizontales ligeramente diferentes, lo cual representa la disparidad horizontal entre un los ojos izquierdo y derecho de un espectador. Visualizando estas dos imágenes de forma simultánea o casi simultánea, de manera que la imagen de la visualización del ojo izquierdo es percibida por el ojo izquierdo del espectador y la imagen de la visualización del ojo derecho es percibida por el ojo derecho del espectador, el espectador puede experimentar un efecto de vídeo tridimensional.

15 **[0023]** Las técnicas de codificación de vídeo relacionadas con la codificación de múltiples visualizaciones (o la codificación de visualización estéreo) en general pueden incluir predicción entre visualizaciones y/o información de profundidad de codificación. La predicción entre visualizaciones en general incluye la codificación de imágenes de una visualización base utilizando predicción inter o predicción intra estándar, y a continuación la predicción de imágenes de otras visualizaciones en relación con la visualización base u otras visualizaciones. De esta manera, algunas imágenes de una visualización no base se pueden predecir en relación con las imágenes de la visualización base u otra visualización no base.

25 **[0024]** Típicamente, cuando una imagen se codifica en relación con una imagen de referencia, los dispositivos de codificación de vídeo señalan la imagen de referencia como miembro de una lista de imágenes de referencia. Por lo tanto, para la codificación entre visualizaciones, las imágenes de referencia agregadas a la lista de imágenes de referencia pueden incluir imágenes de otras visualizaciones, además de otras imágenes de la visualización actual, por ejemplo, para la predicción temporal. El proceso de construcción de la lista de imágenes de referencia puede organizar de forma flexible las referencias de predicción temporal y de visualización. Esto puede proporcionar no solo posible ganancia de eficiencia de codificación sino también capacidad de recuperación de errores, ya que la selección de imágenes de referencia y los mecanismos de imagen redundantes se pueden a continuación extender a la dimensión de visualización.

35 **[0025]** En el contexto actual de 3DV, los codificadores de vídeo pueden codificar la información de textura (por ejemplo, luminancia/brillo y crominancia/color) y la información de profundidad para cada visualización de un flujo de bits de múltiples visualizaciones. Por lo tanto, los codificadores de vídeo 3DV pueden configurarse para codificar mapas de profundidad. Los mapas de profundidad son imágenes cuyos valores de píxel representan las profundidades tridimensionales de los objetos que se muestran en las correspondientes imágenes de textura. Un mapa de profundidad puede codificarse como otra visualización de una unidad de acceso. Es decir, además de imágenes de textura para diferentes visualizaciones, una o más visualizaciones representadas en una unidad de acceso puede incluir un mapa de profundidad respectivo. Un mapa de profundidad para una visualización también se puede denominar un "componente de visualización de profundidad" de la unidad de acceso. El término "componente de visualización" puede usarse para referirse al componente de visualización de textura y al componente de visualización de profundidad de una visualización particular de una unidad de acceso.

45 **[0026]** Por lo tanto, un flujo de bits 3DV puede incluir una representación codificada de múltiples visualizaciones, cada una de las cuales puede contener información de textura (componentes de visualización de textura) e información de profundidad (componentes de visualización de profundidad). Aunque 3DV también contiene la textura de múltiples visualizaciones, en algunos casos, un flujo de bits 3DV puede ser compatible con un flujo de bits de múltiples visualizaciones.

50 **[0027]** Al proporcionar información de profundidad, un dispositivo puede renderizar imágenes para varias visualizaciones distintas de las proporcionadas en el flujo de bits. Por ejemplo, si se proporcionan dos visualizaciones con información de profundidad en un flujo de bits, un dispositivo cliente puede usar la información de profundidad para generar una imagen para una tercera visualización, por ejemplo, entre las dos visualizaciones codificadas del flujo de bits. La generación de dichas imágenes se denomina renderización basada en imágenes en profundidad (DIBR). DIBR se puede usar para generar una visualización basada en la textura disponible y las imágenes de profundidad. Tal proceso de generación se llama síntesis de visualización.

60 **[0028]** La síntesis de visualización se puede utilizar para el post-procesamiento después de la decodificación de las visualizaciones (también conocida como síntesis de visualización "posterior al bucle"). De forma adicional o alternativa, la síntesis de visualización puede usarse para generar imágenes de referencia utilizadas para la predicción entre visualizaciones (también denominada síntesis de visualización "en bucle"). Debe entenderse que el DIBR se puede usar para renderizar una imagen de textura, una imagen de textura, una imagen de profundidad (en general denominada un mapa de profundidad), o ambas. En general, los mapas de profundidad se representan de manera similar a los datos de luminancia (luma) de las imágenes de textura, sin incluir información de crominancia, excepto que los valores de píxeles del mapa de profundidad representan la profundidad de la información de textura correspondiente, en lugar de la información de brillo.

[0029] Con respecto a la síntesis de visualización en bucle, una imagen de referencia generada utilizando la síntesis de visualización, a veces denominada imagen de referencia de síntesis de visualización (VSRP) (o, de forma alternativa, imagen de predicción de síntesis de visualización), puede utilizarse como una imagen de referencia similar a una imagen de referencia de predicción inter temporal o una imagen de referencia entre visualizaciones. Un VSRP se puede incluir en una lista de imágenes de referencia. En algunos diseños específicos de técnicas de codificación convencionales, se pueden usar múltiples visualizaciones (con profundidad) para generar solo un VSRP a través de DIBR. Por lo tanto, para una imagen que se está codificando actualmente, solo un VSRP puede estar disponible en estas técnicas de codificación convencionales.

[0030] La predicción de síntesis de la visualización en el bucle (VSP) se soporta en la norma de codificación de vídeo 3D basada en H.264/AVC actual para la codificación de textura mejorada (WD4, anotada anteriormente). Para permitir que la VSP codifique la visualización actual, los componentes de visualización de profundidad y de textura previamente codificados de la misma unidad de acceso pueden usarse para la síntesis de visualización. Una imagen sintetizada resultante de la VSP se incluye en las listas de imágenes de referencia iniciales (por ejemplo, RefPicList0 y/o RefPicList1) después de las tramas de referencia temporales y entre visualizaciones.

[0031] Documento "3D-CE1.a: modo de predicción de síntesis de visualización generada (GVSP)", Equipo de Colaboración Conjunta en el Desarrollo de la Extensión de Codificación de Vídeo 3D de ITU-T SG 16 WP 3 e ISO/IEC JTC 1/SC 29/WG 11, 2.^a reunión: Shanghái, China, 13-19 de octubre de 2012, JCT3V-B0054 propuso un proceso de síntesis de visualización denominado predicción de síntesis de visualización generalizada (GVSP). En GVSP, se introduce un indicador en varios niveles, macrobloque o inferior, para indicar si un bloque se codifica como modo de salto y se predice a partir de la imagen sintetizada de la visualización.

[0032] La señalización de predicción de síntesis de visualización se puede basar en un índice de referencia tradicional, que es transparente al diseño H.264/AVC en términos de señalización de nivel de MB. Por ejemplo, para cada MB, un codificador de vídeo puede determinar si usar la síntesis de visualización basándose en un índice de referencia que identifica la imagen de síntesis de visualización en una lista de imágenes de referencia, y el índice de referencia de síntesis de visualización se puede señalar explícitamente en una cabecera de fragmento. Este diseño puede soportar la predicción bidireccional, donde una dirección se predice con una imagen de síntesis de visualización y la otra dirección se predice a partir de una imagen temporal normal o una imagen de referencia entre visualizaciones. Este diseño también soporta múltiples imágenes de síntesis de visualización que se adaptan en el nivel de fragmento, así como en el nivel de macrobloque.

[0033] Sin embargo, en el diseño actual (como se establece en el software de referencia 3D AVC actual, denominado 3D-ATM, como se señaló anteriormente), la predicción de síntesis basada en bloques se realiza sobre la marcha. La ejecución de la predicción de síntesis de visualización basada en bloques "sobre la marcha" significa que un bloque de referencia de síntesis de visualización no se genera hasta que se necesita el bloque de síntesis de visualización. En consecuencia, una imagen de predicción de síntesis de visualización (como imagen de referencia) no está presente; por ejemplo, se genera previamente y se incluye en una lista de imágenes de referencia.

[0034] Sin embargo, ciertos equipos para realizar la compensación de movimiento pueden tener un módulo de compensación de movimiento heredado (por ejemplo, maduro) que utiliza una trama de referencia y un vector de movimiento como entrada. Por consiguiente, al menos algunos módulos de compensación de movimiento H.264/AVC actualmente en circulación no se pueden usar para la predicción de síntesis de visualización sobre la marcha, porque no hay ninguna imagen de síntesis de visualización generada previamente como entrada inicial.

[0035] Para compensar el hardware heredado, en lugar de realizar una síntesis de visualización sobre la marcha como se describe anteriormente, un codificador de vídeo siempre puede generar una imagen de síntesis de visualización. Es decir, el codificador de vídeo puede generar una imagen de síntesis de visualización para una secuencia de imágenes y asignar a la imagen de síntesis de visualización un valor de índice de imagen de referencia en una lista de imágenes de referencia. Sin embargo, si siempre se genera una imagen de predicción de síntesis de visualización, no es posible aprovechar la predicción de síntesis de visualización basada en bloques ("sobre la marcha"), que puede afectar la eficiencia de la memoria y de computación. En casos típicos, tal proceso de señalización puede conducir a una degradación de la eficiencia de codificación. Además, el mecanismo basado en la señalización de índice de referencia puede requerir un comando explícito de modificación de la lista de imágenes de referencia para colocar una imagen de síntesis de visualización en una posición deseable de una lista de imágenes de referencia, lo cual puede afectar aún más la eficiencia computacional.

[0036] Esta divulgación reconoce ciertos problemas que pueden surgir con las técnicas de codificación convencionales, por ejemplo, de una extensión 3DV de H.264/AVC. Por ejemplo, no es posible desactivar el mecanismo de señalización basado en el índice de referencia descrito anteriormente para ver la predicción de la síntesis. Por ejemplo, puede que no sea posible desactivar el mecanismo de señalización basado en el índice de referencia actual de 3D-ATM sin interrumpir los objetivos de compatibilidad retroactivos de la extensión 3D-ATM.

[0037] En consecuencia, para los codificadores de vídeo puede ser necesario un nuevo módulo de hardware que soporta la predicción de síntesis de visualización basada en bloques. Además, cuando una secuencia de imágenes tiene una imagen de síntesis de visualización asociada (seq_view_synthesis_flag es igual a 1), un codificador de vídeo debe generar valores de índice de referencia adicionales para las imágenes de síntesis de visualización, de modo que los valores de índice de referencia estén disponibles para asignarse en instancias en que se genera una imagen de síntesis de visualización y se utiliza con fines de predicción. En consecuencia, el codificador de vídeo genera valores de índice de referencia adicionales, incluso cuando una imagen de síntesis de visualización no se genera realmente durante la codificación, lo cual afecta la eficiencia computacional y/o de codificación.

[0038] De acuerdo con los aspectos de esta divulgación, un índice de referencia de una imagen de VSP no se indica específicamente en una cabecera de fragmento ni se agrega durante la construcción de la lista de imágenes de referencia. En su lugar, uno o más elementos sintácticos (p. ej., un indicador) se pueden señalar en el nivel de partición de MB o el MB para cada índice de referencia que corresponde a una imagen de referencia entre visualizaciones. Este indicador puede indicar si VSP se aplica para la partición de MB o el MB para ese índice de referencia. En otras palabras, por ejemplo, el indicador puede indicar si se debe usar una imagen de referencia entre visualizaciones con fines de predicción, o usar una imagen de VSP (por ejemplo, generada a partir de la imagen de referencia entre visualizaciones) con fines de predicción.

[0039] El indicador descrito anteriormente es diferente del tradicional vsp_flag de 3D-ATM. Por ejemplo, el indicador descrito anteriormente puede reemplazar al vsp_flag de 3D-ATM. De acuerdo con los aspectos de esta divulgación, al señalar un índice de referencia de la manera descrita anteriormente, se pueden usar diferentes índices de referencia para VSP. Es decir, por ejemplo, cualquier índice de referencia entre visualizaciones se puede usar para VSP, en lugar de que VSP esté vinculado a un índice de referencia de VSP específico señalado en una cabecera de fragmento.

[0040] En un ejemplo, de acuerdo con los aspectos de esta divulgación, un indicador de VSP solo está presente cuando un índice de referencia para el bloque que se está codificando actualmente corresponde a una imagen de referencia entre visualizaciones. En otro ejemplo, el indicador de VSP puede estar presente para cualquier modo con predicción inter. En otro ejemplo, de forma adicional o alternativa, el indicador solo puede estar presente para un índice de referencia correspondiente a RefPicListO. En otro ejemplo, de forma adicional o alternativa, el indicador de VSP puede estar presente para un índice de referencia correspondiente a RefPicListO y RefPicList1 (si está presente).

[0041] La FIG. 1 es un diagrama de bloques que ilustra un sistema de codificación y decodificación de vídeo de ejemplo 10 que puede utilizar las técnicas de esta divulgación para síntesis de visualización. Como se muestra en la FIG. 1, el sistema 10 incluye un dispositivo de origen 12 que proporciona datos de vídeo codificados, a ser descodificados en un momento posterior por un dispositivo de destino 14. En particular, el dispositivo de origen 12 proporciona los datos de vídeo al dispositivo de destino 14 mediante un medio legible por ordenador 16. El dispositivo 12 de origen y el dispositivo 14 de destino pueden comprender cualquiera de entre una amplia gama de dispositivos, incluidos ordenadores de escritorio, ordenadores plegables (es decir, portátiles), ordenadores de tableta, decodificadores, teléfonos tales como los denominados teléfonos «inteligentes», los denominados paneles «inteligentes», televisores, cámaras, dispositivos de visualización, reproductores de medios digitales, consolas de videojuegos, un dispositivo de transmisión en continuo de vídeo o similares. En algunos casos, el dispositivo 12 de origen y el dispositivo de 14 destino pueden estar equipados para la comunicación inalámbrica.

[0042] El dispositivo de destino 14 puede recibir los datos de vídeo codificados, a descodificar mediante el medio legible por ordenador 16. El medio legible por ordenador 16 puede comprender cualquier tipo de medio o dispositivo capaz de desplazar los datos de vídeo codificados desde el dispositivo de origen 12 al dispositivo de destino 14. En un ejemplo, el medio legible por ordenador 16 puede comprender un medio de comunicación que habilita el dispositivo de origen 12 para transmitir datos de vídeo codificados directamente al dispositivo de destino 14 en tiempo real.

[0043] Los datos de vídeo codificados pueden modularse de acuerdo con una norma de comunicación, tal como un protocolo de comunicación inalámbrica, y transmitirse al dispositivo 14 de destino. El medio de comunicación puede comprender cualquier medio de comunicación, inalámbrica o alámbrica, tal como un espectro de radiofrecuencia (RF) o una o más líneas de transmisión física. El medio de comunicación puede formar parte de una red basada en paquetes, tal como una red de área local, una red de área amplia o una red global tal como Internet. El medio de comunicación puede incluir routers, conmutadores, estaciones base o cualquier otro equipo que pueda ser útil para facilitar la comunicación desde el dispositivo 12 de origen hasta el dispositivo 14 de destino.

[0044] En algunos ejemplos, pueden emitirse datos codificados desde la interfaz de salida 22 hasta un dispositivo de almacenamiento. De forma similar, se puede acceder a los datos codificados del dispositivo de almacenamiento mediante una interfaz de entrada. El dispositivo de almacenamiento puede incluir cualquiera entre una diversidad de medios de almacenamiento de datos, de acceso distribuido o local, tales como una unidad de disco duro, discos Blu-ray, discos DVD, discos CD-ROM, memoria flash, memoria volátil o no volátil u otros medios adecuados cualesquiera de almacenamiento digital para almacenar datos de vídeo codificados. En un ejemplo adicional, el

dispositivo de almacenamiento puede corresponder a un servidor de ficheros o a otro dispositivo de almacenamiento intermedio que pueda almacenar los datos de vídeo codificados generados por el dispositivo de origen 12.

5 **[0045]** El dispositivo de destino 14 puede acceder a datos de vídeo almacenados del dispositivo de almacenamiento mediante transmisión continua o descarga. El servidor de archivos puede ser cualquier tipo de servidor capaz de almacenar datos de vídeo codificados y transmitir esos datos de vídeo codificados al dispositivo 14 de destino. Entre los ejemplos de servidores de archivos se incluyen un servidor web (por ejemplo, para un sitio web), un servidor FTP, dispositivos de almacenamiento conectado en red (NAS) o una unidad de disco local. El dispositivo 14 de destino puede acceder a los datos de vídeo codificados a través de cualquier conexión de datos estándar, incluyendo una conexión a Internet. Esto puede incluir un canal inalámbrico (por ejemplo, una conexión wifi), una conexión alámbrica (por ejemplo, DSL, módem de cable, etc.) o una combinación de ambos que sea adecuada para acceder a datos de vídeo codificados almacenados en un servidor de archivos. La transmisión de datos de vídeo codificados desde el dispositivo de almacenamiento puede ser una transmisión continua, una transmisión de descarga o una combinación de ambas.

10 **[0046]** Las técnicas de esta divulgación no están limitadas necesariamente a aplicaciones o configuraciones inalámbricas. Las técnicas pueden aplicarse a la codificación de vídeo como soporte de cualquiera entre una diversidad de aplicaciones de multimedios, tales como difusiones de televisión por el aire, transmisiones de televisión por cable, transmisiones de televisión por satélite, transmisiones continuas de vídeo por Internet, tales como el flujo de transmisión adaptativo dinámico sobre HTTP (DASH), el vídeo digital que se codifica en un medio de almacenamiento de datos, la descodificación de vídeo digital almacenado en un medio de almacenamiento de datos, u otras aplicaciones. En algunos ejemplos, el sistema 10 puede estar configurado para soportar transmisión de vídeo unidireccional o bidireccional para soportar aplicaciones tales como la transmisión en continuo de vídeo, la reproducción de vídeo, la radiodifusión de vídeo y/o la videotelefonía.

15 **[0047]** En el ejemplo de la FIG. 1, el dispositivo de origen 12 incluye una fuente de vídeo 18, un codificador de vídeo 20 y una interfaz de salida 22. El dispositivo de destino 14 incluye una interfaz de entrada 28, un descodificador de vídeo 30 y un dispositivo de visualización 32. De acuerdo con esta divulgación, el codificador de vídeo 20 del dispositivo de origen 12 puede estar configurado para aplicar las técnicas de predicción de vectores de movimiento en la codificación de múltiples visualizaciones. En otros ejemplos, un dispositivo de origen y un dispositivo de destino pueden incluir otros componentes o disposiciones. Por ejemplo, el dispositivo 12 de origen puede recibir datos de vídeo desde una fuente 18 de vídeo externa, tal como una cámara externa. Del mismo modo, el dispositivo 14 de destino puede interactuar con un dispositivo de visualización externo, en lugar de incluir un dispositivo de visualización integrado.

20 **[0048]** El sistema ilustrado 10 de la FIG. 1 es simplemente un ejemplo. Las técnicas para la síntesis de visualización pueden realizarse mediante cualquier dispositivo de codificación y/o descodificación de vídeo digital. Aunque, en general, las técnicas de esta divulgación se realizan mediante un dispositivo de codificación de vídeo, las técnicas también pueden realizarse mediante un codificador/descodificador de vídeo, denominado típicamente «CÓDEC». Además, las técnicas de esta divulgación también pueden realizarse mediante un preprocesador de vídeo. El dispositivo 12 de origen y el dispositivo 14 de destino son simplemente ejemplos de dichos dispositivos de codificación en los que el dispositivo 12 de origen genera datos de vídeo codificados para su transmisión al dispositivo 14 de destino. En algunos ejemplos, los dispositivos 12, 14 pueden funcionar de manera esencialmente simétrica, de modo que cada uno de los dispositivos 12, 14 incluye componentes de codificación y de descodificación de vídeo. Por lo tanto, el sistema 10 puede soportar transmisión de vídeo unidireccional o bidireccional entre los dispositivos 12, 14 de vídeo, por ejemplo para transmisión de vídeo en continuo, reproducción de vídeo, radiodifusión de vídeo o videotelefonía.

25 **[0049]** La fuente de vídeo 18 del dispositivo de origen 12 puede incluir un dispositivo de captura de vídeo, tal como una cámara de vídeo, un archivo de vídeo que contiene vídeo capturado previamente y/o una interfaz de alimentación de vídeo para recibir vídeo desde un proveedor de contenidos de vídeo. De forma alternativa adicional, la fuente 18 de vídeo puede generar datos basados en gráficos por ordenador como el vídeo de origen, o una combinación de vídeo en directo, vídeo archivado y vídeo generado por ordenador. En algunos casos, si la fuente 18 de vídeo es una videocámara, el dispositivo 12 de origen y el dispositivo 14 de destino pueden formar los denominados teléfonos con cámara o videoteléfonos. Sin embargo, como se ha mencionado anteriormente, las técnicas descritas en esta divulgación pueden ser aplicables a la codificación de vídeo en general, y pueden aplicarse a aplicaciones inalámbricas y/o alámbricas. En cada caso, el codificador 20 de vídeo puede codificar el vídeo captado, precaptado o generado por ordenador. La información de vídeo codificada puede entonces emitirse mediante la interfaz de salida 22 hacia un medio legible por ordenador 16.

30 **[0050]** El medio legible por ordenador 16 puede incluir medios transitorios, tales como una difusión inalámbrica o una transmisión de red alámbrica, o medios de almacenamiento (es decir, medios de almacenamiento no transitorio), tales como un disco duro, una unidad de memoria flash, un disco compacto, un disco de vídeo digital, un disco Blu-ray u otros medios legibles por ordenador. En algunos ejemplos, un servidor de red (no mostrado) puede recibir datos de vídeo codificados desde el dispositivo de origen 12 y proporcionar los datos de vídeo

codificados al dispositivo de destino 14, por ejemplo, mediante transmisión por red. De manera similar, un dispositivo informático de una utilidad de producción de un medio, tal como una utilidad de grabación de discos, puede recibir datos de vídeo codificados desde el dispositivo de origen 12 y producir un disco que contiene los datos de vídeo codificados. Por lo tanto, puede entenderse que el medio legible por ordenador 16 incluye uno o más medios legibles por ordenador de diversas formas, en varios ejemplos.

[0051] La interfaz de entrada 28 del dispositivo de destino 14 recibe información desde el medio legible por ordenador 16. La información del medio legible por ordenador 16 puede incluir información sintáctica definida por el codificador de vídeo 20, que también usa el descodificador de vídeo 30, que incluye elementos sintácticos que describen características y/o procesamiento de bloques y otras unidades codificadas, por ejemplo, grupos de imágenes (GOP). El dispositivo 32 de visualización muestra los datos de vídeo descodificados a un usuario, y puede comprender cualquiera de una variedad de dispositivos de visualización, tales como un tubo de rayos catódicos (CRT), una pantalla de cristal líquido (LCD), una pantalla de plasma, una pantalla de diodos orgánicos emisores de luz (OLED) u otro tipo de dispositivo de visualización.

[0052] Tanto el codificador 20 de vídeo como el descodificador 30 de vídeo pueden implementarse como cualquiera de entre una variedad de circuitos codificadores o descodificadores adecuados, según corresponda, tales como uno o más de entre microprocesadores, procesadores de señales digitales (DSP), circuitos integrados específicos de la aplicación (ASIC), matrices de puertas programables *in situ* (FPGA), circuitos de lógica discreta, software, hardware, firmware o cualquier combinación de los mismos. Tanto el codificador 20 de vídeo como el descodificador 30 de vídeo pueden estar incluidos en uno o más codificadores o descodificadores, cada uno de los cuales puede estar integrado como parte de un codificador/descodificador (códec) de vídeo combinado. Un dispositivo que incluye el codificador de vídeo 20 y/o el descodificador de vídeo 30 puede comprender un circuito integrado, un microprocesador y/o un dispositivo de comunicación inalámbrica, tal como un teléfono celular.

[0053] Aunque no se muestra en la FIG. 1, en algunos aspectos, tanto el codificador 20 de vídeo como el descodificador 30 de vídeo pueden estar integrados en un codificador y descodificador de audio, y pueden incluir unidades MUX-DEMUX adecuadas, u otro tipo de hardware y software, para ocuparse de la codificación tanto de audio como de vídeo en un flujo de datos común o en flujos de datos separados. Si procede, las unidades MUX-DEMUX pueden ajustarse al protocolo de multiplexador ITU H.223 o a otros protocolos, tales como el protocolo de datagramas de usuario (UDP).

[0054] Esta divulgación puede referirse en general al codificador de vídeo 20 que "señala" cierta información a otro dispositivo, tal como el descodificador de vídeo 30. Sin embargo, debería entenderse que el codificador de vídeo 20 puede señalar información asociando ciertos elementos sintácticos con varias partes codificadas de datos de vídeo. Es decir, el codificador de vídeo 20 puede "señalar" datos almacenando ciertos elementos sintácticos en cabeceras de varias partes codificadas de datos de vídeo. En algunos casos, dichos elementos sintácticos pueden ser codificados y almacenados (por ejemplo, almacenados en un dispositivo de almacenamiento 24) antes de ser recibidos y descodificados por el descodificador de vídeo 30. Por lo tanto, el término "señalización" puede referirse en general a la comunicación de datos sintácticos u otros, usados para descodificar datos de vídeo comprimidos, ya sea que dicha comunicación se produzca en tiempo real o casi real o en un intervalo de tiempo, tal como podría ocurrir cuando se almacenan elementos sintácticos en un medio en el momento de la codificación, que entonces pueden ser recuperados por un dispositivo de descodificación en cualquier momento después de ser almacenados en este medio.

[0055] El codificador de vídeo 20 y el descodificador de vídeo 30 pueden funcionar de acuerdo con una variedad de normas de codificación de vídeo. Por ejemplo, el codificador de vídeo 20 y el descodificador de vídeo 30 pueden funcionar de acuerdo con la norma ITU-T H.264/MPEG-4 (AVC), que fue formulada por el Grupo de Expertos en Codificación de Vídeo de ITU-T (VCEG), junto con el Grupo de Expertos en Imágenes en Movimiento (MPEG) de ISO/IEC, como el producto de una asociación colectiva conocida como Equipo Mixto de Vídeo (JVT, por sus siglas en inglés). En algunos aspectos, las técnicas descritas en esta divulgación pueden aplicarse a dispositivos que se ajustan en general a la norma H.264. La norma H.264 se describe en la recomendación ITU-T H.264, codificación avanzada de vídeo para servicios audiovisuales genéricos, por el Grupo de estudio de la ITU-T, con fecha de marzo de 2005, que puede denominarse en el presente documento norma H.264 o especificación H.264, o norma o especificación H.264/AVC.

[0056] El Equipo mixto de vídeo (JVT) continúa trabajando en extensiones de la norma H.264/MPEG-4 AVC. En particular, una posible norma es una norma de codificación de vídeo 3D basado en H.264/AVC que proporciona codificación de textura y profundidad para cada visualización de datos de vídeo. Las principales herramientas de codificación para 3DV WD4 se pueden describir en el documento JCT3V-B1002, como se señaló anteriormente. Además, el software de referencia más reciente se puede denominar aquí ATM o 3D-ATM, como se indicó anteriormente.

[0057] Sin embargo, las técnicas de esta divulgación no están limitadas a ninguna norma de codificación particular. Entre otros ejemplos de normas de codificación de vídeo se incluyen MPEG-2 y ITU-T H.263, y la norma de codificación de vídeo de alta eficiencia (HEVC). Por ejemplo, en algunos casos, el codificador de vídeo 20 y el

descodificador de vídeo 3D pueden realizar codificación de vídeo de múltiples visualizaciones (MVC), y pueden implementar una extensión de múltiples visualizaciones de HEVC, llamada MV-HEVC, o un códec 3DV completo basado en HEVC mejorado en profundidad, 3D-HEVC.

5 **[0058]** En H.264/AVC, cada macrobloque (MB) con predicción inter se puede dividir en una de cuatro posibles formas diferentes, incluyendo una partición de MB 16x16, dos particiones de MB 16x8, dos particiones de MB 8x16 o cuatro particiones de MB 8x8. Las diferentes particiones de MB de un MB pueden tener diferentes valores de índice de referencia para cada dirección (RefPicList0 o RefPicList1). Cuando un MB no está dividido en cuatro
10 particiones de MB 8x8, el MB puede tener solo un vector de movimiento para toda la partición de MB en cada dirección. Cuando un MB está dividido en cuatro particiones de MB 8x8, cada partición de MB 8x8 puede dividirse además en sub-bloques, cada uno de los cuales puede tener un vector de movimiento diferente en cada dirección. Puede haber cuatro formas diferentes de dividir una partición de MB 8x8 en sub-bloques, incluyendo un sub-bloque 8x8, dos sub-bloques 8x4, dos sub-bloques 4x8 y cuatro sub-bloques 4x4. Cada uno de los sub-bloques puede tener un vector de movimiento diferente en cada dirección.

15 **[0059]** En general, el bucle de compensación de movimiento de H.264/AVC y HEVC es el mismo. Por ejemplo, la reconstrucción de una trama actual \hat{I} en el bucle de compensación de movimiento puede ser igual a los coeficientes des-cuantificados r más la predicción temporal P :

$$\hat{I} = r + P.$$

20 **[0060]** En la fórmula anterior, P indica una predicción inter uni-predictiva para tramas P o una predicción inter predictiva bi-predictiva para tramas B .

[0061] Sin embargo, la unidad de compensación de movimiento en HEVC es diferente de la de las normas de codificación de vídeo anteriores. Por ejemplo, el concepto de un macrobloque en normas de codificación de vídeo anteriores no existe en HEVC. En lugar de eso, los macrobloques se reemplazan por una estructura jerárquica flexible basada en un esquema de árbol cuaternario genérico. Dentro de este sistema, se definen tres tipos de bloques, es decir, unidades de codificación (CU), unidades de predicción (PU) y unidades de transformada (TU). Una CU es una unidad básica de división de zonas. El concepto de CU es análogo al concepto de macrobloque, pero una CU no está restringida a un tamaño máximo y una CU permite la división recursiva en cuatro CU de igual tamaño para mejorar el contenido adaptativamente. Una PU es una unidad básica de predicción inter/intra. En algunos ejemplos, una PU puede contener varias particiones de forma arbitraria en una sola PU para codificar eficazmente patrones de imagen irregulares. Una TU es una unidad básica de transformada. Las TU de una CU pueden definirse independientemente de las PU de la CU. Sin embargo, el tamaño de una TU está limitado a la CU a la que pertenece la TU. Esta separación de la estructura de bloques en tres conceptos diferentes puede permitir optimizar cada uno de ellos de acuerdo con su función, lo cual puede dar como resultado una mejora de la eficacia de codificación.

35 **[0062]** En cualquier caso, en especificaciones de codificación de vídeo, una secuencia de vídeo incluye típicamente una serie de imágenes. Las imágenes también pueden denominarse "tramas". Una imagen puede incluir tres matrices de muestra, indicadas como S_L , S_{Cb} y S_{Cr} . S_L es una matriz bidimensional (es decir, un bloque) de muestras de luminancia. S_{Cb} es una matriz bidimensional de muestras de crominancia C_b . S_{Cr} es una matriz bidimensional de muestras de crominancia C_r . Las muestras de crominancia también se pueden denominar en el presente documento muestras de "croma". En otros casos, una imagen puede ser monocromática y puede solo incluir una matriz de muestras de luminancia.

45 **[0063]** En algunos casos, las imágenes se pueden dividir en varios fragmentos, y cada fragmento incluye varios macrobloques o unidades de árbol de codificación (CTU, por sus siglas en inglés) con respecto a HEVC. Un fragmento puede incluir un número entero de macrobloques o CTU ordenadas consecutivamente en un orden de escaneado cuadrado. Un fragmento codificado puede comprender una cabecera de fragmento y datos de fragmento. La cabecera de fragmento de un fragmento puede ser una estructura sintáctica que incluye elementos sintácticos que proporcionan información sobre el fragmento.

50 **[0064]** Esta divulgación puede usar el término "unidad de vídeo", "bloque de vídeo" o "bloque" para referirse a uno o más bloques de muestras y estructuras sintácticas usadas para codificar muestras del uno o más bloques de muestras. Entre los tipos de bloques o unidades de vídeo de ejemplo pueden incluirse macrobloques, particiones de macrobloques, CTU, CU, PU, unidades de transformada (TU), etc. En general, se puede hacer referencia a que los bloques de datos de vídeo se encuentran en el "nivel de bloque", en comparación con los niveles más altos, como el nivel de división, el nivel de imagen o el nivel de secuencia.

60 **[0065]** El codificador de vídeo 2D puede utilizar predicción intra o predicción inter para generar un bloque predictivo. Si el codificador de vídeo 2D utiliza predicción intra para generar el bloque predictivo, el codificador de vídeo 2D puede generar el bloque predictivo basándose en muestras de la imagen asociada con el bloque codificado. En esta divulgación, la frase "basado en" puede indicar "basado al menos en parte en".

5 **[0066]** Si el codificador de vídeo 20 utiliza predicción inter para generar los bloques predictivos, el codificador de vídeo 20 puede generar los bloques predictivos basándose en muestras descodificadas de una o más imágenes distintas a la imagen actualmente codificada. Cuando se utiliza predicción inter para generar los bloques predictivos de un bloque, esta divulgación puede referirse al bloque como "con intercodificación" o "con predicción inter". La predicción inter puede ser uni-predictiva (es decir, uni-predicción) o bi-predictiva (es decir, bi-predicción). Para realizar uni-predicción o bi-predicción, el codificador de vídeo 20 puede generar una primera lista de imágenes de referencia (RefPicList0) y una segunda lista de imágenes de referencia (RefPicList1) para una imagen actual. Cada una de las listas de imágenes de referencia puede incluir una o más imágenes de referencia. Después de que se construya una lista de imágenes de referencia (concretamente RefPicList0 y RefPicList1, si están disponibles), se puede usar un índice de referencia a una lista de imágenes de referencia para identificar cualquier imagen de referencia incluida en la lista de imágenes de referencia.

15 **[0067]** Cuando se utiliza uni-predicción, el codificador de vídeo 20 puede buscar las imágenes de referencia en RefPicList0, RefPicList1 o en ambas para determinar una ubicación de referencia dentro de una imagen de referencia. Además, cuando se usa uni-predicción, el codificador de vídeo 20 puede generar, basándose al menos en parte en muestras correspondientes a la ubicación de referencia, los bloques predictivos. Además, cuando se usa uni-predicción, el codificador de vídeo 20 puede generar un solo vector de movimiento que indica un desplazamiento espacial entre un bloque de predicción y la ubicación de referencia. El vector de movimiento puede incluir un componente horizontal que especifica un desplazamiento horizontal entre el bloque de predicción y la ubicación de referencia y puede incluir un componente vertical que especifica un desplazamiento vertical entre el bloque de predicción y la ubicación de referencia.

25 **[0068]** Cuando se utiliza la predicción doble para codificar un bloque, el codificador de vídeo 20 puede determinar una primera ubicación de referencia en una imagen de referencia en RefPicList0 y una segunda ubicación de referencia en una imagen de referencia en RefPicList1. El codificador de vídeo 20 puede generar, basándose al menos en parte en muestras correspondientes a las ubicaciones de referencia primera y segunda, los bloques predictivos. Además, cuando se usa la bi-predicción, el codificador de vídeo 20 puede generar un primer vector de movimiento que indica un desplazamiento espacial entre un bloque de predicción y la primera ubicación de referencia y un segundo vector de movimiento que indica un desplazamiento espacial entre el bloque de predicción y la segunda ubicación de referencia.

35 **[0069]** Tras la codificación intrapredictiva o interpredictiva, el codificador de vídeo 20 puede calcular datos residuales para el bloque actualmente codificado. Los datos residuales pueden corresponder a diferencias de píxel entre píxeles de la imagen no codificada y los bloques predictivos. El codificador de vídeo 20 puede a continuación transformar los datos residuales. Por ejemplo, el codificador de vídeo 20 puede transformar los datos de vídeo del dominio espacial al dominio de transformada aplicando, por ejemplo, una transformada de coseno discreta (DCT), una transformada entera, una transformada de ondícula o una transformada conceptualmente similar.

40 **[0070]** Tras cualquier transformada para generar coeficientes de transformada, el codificador 20 de vídeo puede realizar la cuantificación de los coeficientes de transformada. La cuantificación se refiere, en general, a un proceso en el que los coeficientes de transformada se cuantifican para reducir posiblemente la cantidad de datos usados para representar los coeficientes, proporcionando compresión adicional. El proceso de cuantificación puede reducir la profundidad de bits asociada a algunos, o a la totalidad, de los coeficientes. Por ejemplo, un valor de n bits puede redondearse a la baja hasta un valor de m bits durante la cuantificación, donde n es mayor que m.

45 **[0071]** Después de la cuantificación, el codificador de vídeo 20 puede recorrer los coeficientes de transformada, produciendo un vector unidimensional a partir de la matriz bidimensional que incluye los coeficientes de transformada cuantificados. El recorrido puede estar diseñado para colocar los coeficientes de energía más alta (y por lo tanto de menor frecuencia) en la parte frontal de la matriz y para colocar los coeficientes de energía más baja (y por lo tanto de mayor frecuencia) en la parte posterior de la matriz. En algunos ejemplos, el codificador 20 de vídeo puede utilizar un orden de examen predefinido para examinar los coeficientes de transformada cuantificados con el fin de generar un vector en serie que se pueda someter a codificación por entropía. En otros ejemplos, el codificador 20 de vídeo puede realizar un examen adaptativo.

55 **[0072]** Después de recorrer los coeficientes de transformada cuantificados para formar un vector unidimensional, el codificador de vídeo 20 puede realizar la codificación por entropía del vector unidimensional, por ejemplo, de acuerdo con la codificación de longitud variable adaptable al contexto (CAVLC), la codificación aritmética binaria adaptable al contexto (CABAC), la codificación aritmética binaria adaptable al contexto basada en la sintaxis (SBAC), la codificación por entropía por división de intervalos de probabilidad (PIPE) o con otra metodología de codificación por entropía. El codificador 20 de vídeo también puede realizar la codificación por entropía de los elementos sintácticos asociados a los datos de vídeo codificados, para su uso por el descodificador 30 de vídeo en la descodificación de los datos de vídeo.

65 **[0073]** Además, el codificador de vídeo 20 puede enviar datos sintácticos, tales como datos sintácticos basados en bloques, datos sintácticos basados en imágenes y datos sintácticos basados en GOP, al descodificador de vídeo 30, por ejemplo, en una cabecera de imagen, una cabecera de bloque, una cabecera de fragmento o una

cabecera de GOP. Los datos sintácticos del GOP pueden describir un cierto número de imágenes en el GOP respectivo, y los datos sintácticos de imagen pueden indicar un modo de codificación/predicción usada para codificar la imagen correspondiente.

5 **[0074]** El codificador de vídeo 20 puede emitir un flujo de bits que incluye una secuencia de bits que forman una representación de imágenes datos de vídeo (es decir, imágenes codificadas y datos asociados). El flujo de bits puede comprender una secuencia de unidades de capa de abstracción de red (NAL). Una unidad NAL es una estructura sintáctica que contiene una indicación del tipo de datos en la unidad NAL y los bytes que contienen dichos datos en forma de una carga útil de secuencia de bytes sin procesar (RBSP) intercalados según sea necesario con bits de prevención de emulación. Cada una de las unidades NAL incluye una cabecera de la unidad NAL y encapsula una RBSP. La cabecera de la unidad NAL puede incluir un elemento sintáctico que indique un código de tipo de unidad NAL. El código de tipo de unidad NAL especificado por la cabecera de unidad NAL de una unidad NAL indica el tipo de la unidad NAL. Una RBSP puede ser una estructura de sintaxis que contenga un número entero de bytes que se encapsule dentro de una unidad NAL. En algunos casos, una RBSP incluye cero bits.

10 **[0075]** Diferentes tipos de unidades NAL pueden encapsular diferentes tipos de las RBSP. Por ejemplo, diferentes tipos de unidades NAL pueden encapsular diferentes RBSP para conjuntos de parámetros de secuencia (SPS), conjuntos de parámetros de imagen (PPS), fragmentos codificados, mensajes de información de mejora suplementaria (SEI), etc. Las unidades NAL que encapsulan las RBSP para datos de codificación de vídeo (a diferencia de las RBSP para conjuntos de parámetros y mensajes SEI) pueden denominarse unidades NAL de la capa de codificación de vídeo (VCL).

15 **[0076]** El descodificador de vídeo 30 puede recibir un flujo de bits generado por el codificador de vídeo 20. Además, el descodificador de vídeo 30 puede analizar el flujo de bits para obtener elementos sintácticos del flujo de bits. El descodificador de vídeo 30 puede reconstruir las imágenes de los datos de vídeo basándose, al menos en parte, en los elementos sintácticos obtenidos a partir del flujo de bits. El proceso para reconstruir los datos de vídeo puede ser, en general, recíproco al proceso realizado por el codificador de vídeo 20.

20 **[0077]** Por ejemplo, el descodificador de vídeo 30 puede usar vectores de movimiento para determinar bloques predictivos para un bloque que se está descodificando actualmente (un bloque actual). Además, el descodificador de vídeo 30 puede cuantificar inversamente coeficientes de transformada del bloque actual. El descodificador de vídeo 30 puede aplicar una transformada inversa en los coeficientes que determinan los datos residuales para el bloque actual. El descodificador de vídeo 30 puede reconstruir el bloque actual agregando las muestras del bloque predictivo a las muestras residuales correspondientes. Reconstruyendo cada uno de los bloques para una imagen, el descodificador de vídeo 30 puede reconstruir la imagen.

25 **[0078]** En general, para realizar predicción de síntesis de visualización, un codificador de vídeo, como el codificador de vídeo 20 o el descodificador de vídeo 30, puede usar la información de textura y profundidad de una o más visualizaciones para generar un componente de visualización sintetizado. Por ejemplo, VSP puede referirse a generar una trama de referencia sintetizada adicional con el fin de predecir datos de vídeo. Una versión virtual de una trama puede sintetizarse para codificarse a partir de tramas previamente codificadas de otras cámaras y puede usar la trama virtual como referencia de predicción.

30 **[0079]** Este proceso es sustancialmente similar al proceso para generar un componente de visualización sintetizado utilizando información de textura y profundidad con fines de visualización. Sin embargo, en la predicción de síntesis de visualización, se puede agregar un componente de visualización sintetizada (es decir, una imagen) a una lista de imágenes de referencia para usarla como imagen de referencia cuando se codifica una visualización diferente. La predicción de síntesis de visualización se puede usar para generar imágenes de referencia para codificar información de textura o información de profundidad. Por supuesto, las imágenes de síntesis de visualización generadas también pueden mostrarse en un dispositivo cliente como parte de la reproducción. De forma adicional o alternativa, el descodificador de vídeo 30, u otra unidad del dispositivo de destino 14, como una unidad de posprocesamiento (no mostrada), puede realizar una síntesis de visualización con el fin de generar imágenes sintetizadas para la reproducción. Debe entenderse que, aunque las imágenes de síntesis de visualización utilizadas para la predicción pueden mostrarse, en algunos ejemplos, puede generarse una imagen separada utilizando DIBR con fines de visualización. Las imágenes de síntesis de visualización con fines de predicción pueden generarse durante la codificación o "en bucle", mientras que las imágenes de síntesis de visualización con fines de visualización pueden generarse "en bucle" o "posterior al bucle", es decir, siguiendo la codificación.

35 **[0080]** Para generar una imagen de referencia adicional, se puede usar un proceso de interpolación de visualización o distorsión 3D. Por ejemplo, un proceso de distorsión 3D puede incluir realizar una proyección utilizando información de profundidad y parámetros de la cámara. Un proceso de interpolación de visualización puede incluir explotar información de disparidad entre imágenes de puntos de visualización adyacentes. En general, para VSP, existen dos tipos de técnicas de distorsión de imagen, a saber, distorsión hacia delante y hacia atrás, dependiendo de la disponibilidad del mapa de profundidad de la visualización actual. La distorsión hacia

adelante genera la visualización sintética cuando el mapa de profundidad desde el punto de visualización de referencia está disponible. En este ejemplo, el mapa de profundidad desde el punto de visualización de referencia se codifica/descodifica antes de codificar/descodificar el componente de textura de la visualización actual. Se puede obtener un resultado similar para la distorsión hacia atrás utilizando la profundidad de la visualización que se va a sintetizar.

[0081] VSP en bucle se soporta en la norma de codificación de vídeo 3D basado en H.264/AVC actual para la codificación de textura mejorada (WD4, anotado anteriormente). Para permitir que la VSP codifique la visualización actual, los componentes de visualización de profundidad y de textura previamente codificados de la misma unidad de acceso pueden usarse para la síntesis de visualización. Una imagen sintetizada resultante de la VSP se incluye en las listas de imágenes de referencia iniciales (por ejemplo, RefPicList0 y/o RefPicList1) después de las tramas de referencia temporales y entre visualizaciones.

[0082] Como se señaló anteriormente, el documento "3D-CE1.a: modo de predicción de síntesis de visualización generada (GVSP)", Equipo de Colaboración Conjunta en el Desarrollo de la Extensión de Codificación de Vídeo 3D de ITU-T SG 16 WP 3 e ISO/IEC JTC 1/SC 29/WG 11, 2.^a reunión: Shanghái, China, 13-19 de octubre de 2012, JCT3V-B0054 propuso introducir un indicador en varios niveles, macrobloque o inferior, para indicar si un bloque se codifica como modo de salto y se predice desde la imagen sintetizada de la visualización.

[0083] Un ejemplo del conjunto de elementos sintácticos se presenta en la Tabla 1 - Tabla 3 a continuación. Por conveniencia editorial, los indicadores pueden denominarse vsp_flag. Un ejemplo de la sintaxis de la capa de macrobloque se muestra en la Tabla 1 a continuación:

TABLA 1 - Sintaxis de capa de macrobloque

macroblock_layer() {	C	Descriptor
if (slice_vsp_flag && VspRefExist)		
vsp_mb_flag	2	ae(v)
if (!vsp_mb_flag)		
mb_type	2	ue(v) ae(v)
if(mb_type == I_PCM) {		
while(!byte_aligned())		
pcm_alignment_zero_bit	3	f(1)
for(i = 0; i < 256; i++)		
pcm_sample_luma[i]	3	u(v)
for(i = 0; i < 2 * MbWidthC * MbHeightC; i++)		
pcm_sample_chroma[i]	3	u(v)
} else {		
noSubMbPartSizeLessThan8x8Flag = 1		
if(mb_type != I_NxN && MbPartPredMode(mb_type, 0) != Intra_16x16 && NumMbPart(mb_type) == 4) {		
if(!vsp_mb_flag)		
sub_mb_pred(mb_type)	2	
for(mbPartIdx = 0; mbPartIdx < 4; mbPartIdx++)		
if(sub_mb_type[mbPartIdx] != B_Direct_8x8) {		
if(NumSubMbPart(sub_mb_type[mbPartIdx]) > 1)		
noSubMbPartSizeLessThan8x8Flag = 0		
} else if(!direct_8x8_inference_flag)		
noSubMbPartSizeLessThan8x8Flag = 0		
} else {		
if(transform_8x8_mode_flag && mb_type == I_NxN)		
transform_size_8x8_flag	2	u(1) ae(v)
if(!vsp_mb_flag)		

macroblock_layer() {	C	Descriptor
mb_pred(mb_type)	2	
}		
if(MbPartPredMode(mb_type, 0) != Intra_16x16) {		
coded_block_pattern	2	me(v) ae(v)
if(CodedBlockPatternLuma > 0 && transform_8x8_mode_flag && mb_type != I_NxN && noSubMbPartSizeLessThan8x8Flag && (vsp_mb_flag (lvsp mb_flag && (mb_type != B_Direct_16x16 direct_8x8_inference_flag))))		
transform_size_8x8_flag	2	u(1) ae(v)
}		
if(CodedBlockPatternLuma > 0 CodedBlockPatternChroma > 0 MbPartPredMode(mb_type, 0) == Intra_16x16) {		
mb_qp_delta	2	se(v) ae(v)
residual(0, 15)	3 4	
}		
}		
}		

5 **[0084]** En el ejemplo de la Tabla 1 anterior, vsp_mb_flag igual a 1 indica que todo el MB se predice a partir de una imagen VSP. Este indicador igual a 0 indica que todo el MB puede predecirse por otros modos. Cuando este indicador es igual a 1, el mb_type no se señala. Cuando no está presente, se deduce que es igual a 0.

[0085] Un ejemplo de la sintaxis de predicción de macrobloques se muestra en la Tabla 2 a continuación:

10

TABLA 2 - Sintaxis de predicción de macrobloque

mb_pred(mb_type) {	C	Descriptor
if(MbPartPredMode(mb_type, 0) == Intra_4x4 MbPartPredMode(mb_type, 0) == Intra_16x16) {		
if(MbPartPredMode(mb_type, 0) == Intra_4x4)		
for(luma4x4BlkIdx=0; luma4x4BlkIdx<16; luma4x4BlkIdx++) {		
prev_intra4x4_pred_mode_flag[luma4x4BlkIdx]	2	u(1) ae(v)
if(!prev_intra4x4_pred_mode_flag[luma4x4BlkIdx])		
rem_intra4x4_pred_mode[luma4x4BlkIdx]	2	u(3) ae(v)
}		
intra_chroma_pred_mode	2	ue(v) ae(v)
} else if(MbPartPredMode(mb_type, 0) != Direct) {		
for(mbPartIdx = 0; mbPartIdx < NumMbPart(mb_type); mbPartIdx++)		

mb_pred(mb_type) {	C	Descriptor
if (slice_vsp_flag && NumMbPart(mb_type) != 1 && VspRefExist) //vsp_mb_flag is not 1		
mb_part_vsp_flag[mbPartIdx]	2	ae(v)
if(!mb_part_vsp_flag[mbPartIdx] &&(num_ref_idx_l0_active_minus1 > 0 mb_field_decoding_flag) && MbPartPredMode(mb_type, mbPartIdx) != Pred_L1)		
ref_idx_l0[mbPartIdx]	2	te(v) ae(v)
for (mbPartIdx = 0; mbPartIdx < NumMbPart(mb_type) && !mb_part_vsp_flag[mbPartIdx]; mbPartIdx++)		
if((num_ref_idx_l1_active_minus1 > 0 mb_typefield_decoding_flag) && MbPartPredMode(mb_type, mbPartIdx) != Pred_L0)		
ref_idx_l1[mbPartIdx]	2	te(v) ae(v)
for(mbPartIdx = 0; mbPartIdx < NumMbPart(mb_type); mbPartIdx++)		
if(!mb_part_vsp_flag[mbPartIdx] && MbPartPredMode (mb_type, mbPartIdx) != Pred L1)		
for(compldx = 0; compldx < 2; compldx++)		
mvd_l0[mbPartIdx][0][compldx]	2	se(v) ae(v)
for(mbPartIdx = 0; mbPartIdx < NumMbPart(mb_type); mbPartIdx++)		
if(!mb_part_vsp_flag[mbPartIdx] && MbPartPredMode(mb_type, mbPartIdx) != Pred L0)		
for(compldx = 0; compldx < 2; compldx++)		
mvd_l1[mbPartIdx][0][compldx]	2	se(v) ae(v)
}		
}		

5 **[0086]** En el ejemplo de la Tabla 2 anterior, mb_part_vsp_flag [mbPartIdx] igual a 1 indica que la partición de MB actual se predice a partir de una imagen VSP. Este indicador igual a 0 indica que toda la partición de MB no se predice a partir de una imagen VSP. Cuando este indicador no está presente, se deduce que este indicador es igual a 0.

10 **[0087]** En algunos ejemplos, si el número de particiones de MB es menor que 4, noSubMbPartSizeLessThan8x8Flag es verdadero (igual que en H.264/AVC). Si el número de particiones de MB es igual a 4, se puede obtener un elemento sintáctico noSubMbPartSizeLessThan8x8Flag comprobando las particiones de 8x8 MB con mb_part_vsp_flag igual a 0, de la forma siguiente:

- 15 1. Si cualquier partición de MB tiene una partición de subMB menor que 8x8, noSubMbPartSizeLessThan8x8Flag es falso, retorno
2. noSubMbPartSizeLessThan8x8Flag es verdadero.

20 **[0088]** Un ejemplo de la sintaxis de predicción de sub-macrobloques se muestra en la Tabla 3 a continuación:

TABLA 3 - Sintaxis de predicción de sub-macrobloque

sub_mb_pred(mb_type) {	C	Descriptor
for(mbPartIdx = 0; mbPartIdx < 4; mbPartIdx++) {		
if (slice_vsp_flag && VspRefExist) // vsp_mb_flag is not 1		
sub_mb_vsp_flag[mbPartIdx]		
if (!sub_mb_vsp_flag[mbPartIdx])		
sub_mb_type[mbPartIdx]	2	ue(v) ae(v)

sub_mb_pred(mb_type) {	C	Descriptor
}		
for(mbPartIdx = 0; mbPartIdx < 4; mbPartIdx++)		
if(!sub_mb_vsp_flag[mbPartIdx] &&(num_ref_idx_l0_active_minus1 > 0 mb_field_decoding_flag) && mb_type != P_8x8ref0 && sub_mb_type[mbPartIdx] != B_Direct_8x8 && SubMbPredMode(sub_mb_type[mbPartIdx]) != Pred_L1)		
ref_idx_l0[mbPartIdx]	2	te(v) ae(v)
for(mbPartIdx = 0; mbPartIdx < 4; mbPartIdx++)		
if(!sub_mb_vsp_flag[mbPartIdx] &&(num_ref_idx_l1_active_minus1 > 0 mb_field_decoding_flag) && sub_mb_type[mbPartIdx] != B_Direct_8x8 && SubMbPredMode(sub_mb_type[mbPartIdx]) != Pred_L0)		
ref_idx_l1[mbPartIdx]	2	te(v) ae(v)
for(mbPartIdx = 0; mbPartIdx < 4; mbPartIdx++)		
if(!sub_mb_vsp_flag[mbPartIdx] &&(sub_mb_type[mbPartIdx] != B_Direct_8x8 && SubMbPredMode(sub_mb_type[mbPartIdx]) != Pred_L1)		
for(subMbPartIdx = 0; subMbPartIdx < NumSubMbPart(sub_mb_type[mbPartIdx]); subMbPartIdx++)		
for(compldx = 0; compldx < 2; compldx++)		
mvd_l0[mbPartIdx][subMbPartIdx][compldx]	2	se(v) ae(v)
for(mbPartIdx = 0; mbPartIdx < 4; mbPartIdx++)		
if(!sub_mb_vsp_flag[mbPartIdx] &&sub_mb_type[mbPartIdx] != B_Direct_8x8 && SubMbPredMode(sub_mb_type[mbPartIdx]) != Pred_L0)		
for(subMbPartIdx = 0; subMbPartIdx < NumSubMbPart(sub_mb_type[mbPartIdx]); subMbPartIdx++)		
for(compldx = 0; compldx < 2; compldx++)		
mvd_l1[mbPartIdx][subMbPartIdx][compldx]	2	se(v) ae(v)
}		

[0089] En el ejemplo de la Tabla 3 anterior, sub_mb_vsp_flag[mbPartIdx] igual a 1 indica que la partición de MB actual (8x8) se predice a partir de una imagen VSP. Este indicador igual a 0 indica que toda la partición de MB no se predice a partir de una imagen VSP. Cuando no está presente, se deduce que este indicador es igual a 0.

5

[0090] Para simplificar la síntesis de visualización, en algunos ejemplos, la predicción de síntesis de visualización se puede unificar con la predicción de vectores de movimiento. En tales ejemplos, la predicción de síntesis de visualización puede simplificarse para realizar solo la distorsión hacia atrás y no realizar ninguna técnica de síntesis de visualización adicional, como el llenado de agujeros. Además, se aplica un valor de disparidad o valor de profundidad a un bloque 4x4 completo. Por consiguiente, es posible simplificar la predicción de síntesis de visualización para utilizar técnicas de compensación de movimiento tradicionales, de modo que el vector de movimiento se obtiene a partir de la profundidad o disparidad y se asocia con cada bloque 4x4 de un MB, partición de MB o partición de sub-MB. Este proceso de síntesis de visualización simplificada puede denominarse modo VSP unificado.

10

15

[0091] Como se señaló anteriormente, la señalización de la predicción de síntesis de visualización se puede basar en un índice de referencia tradicional, que es transparente al diseño H.264/AVC en términos de señalización de nivel de MB. Por ejemplo, para cada MB, un codificador de vídeo puede determinar si usar la síntesis de visualización basándose en un índice de referencia que identifica la imagen de síntesis de visualización en una lista de imágenes de referencia, y el índice de referencia de síntesis de visualización se puede señalar explícitamente en una cabecera de fragmento. Este diseño puede soportar la predicción bidireccional, donde una dirección se predice con una imagen de síntesis de visualización y la otra dirección se predice a partir de una

20

imagen temporal normal o una imagen de referencia entre visualizaciones. Este diseño también soporta múltiples imágenes de síntesis de visualización que se adaptan en el nivel de fragmento, así como en el nivel de macrobloque.

5 **[0092]** No es posible desactivar el mecanismo de señalización basado en el índice de referencia actual de 3D-ATM. Por ejemplo, puede que no sea posible desactivar el mecanismo de señalización basado en el índice de referencia actual de 3D-ATM sin interrumpir los objetivos de compatibilidad retroactivos de la extensión 3D-ATM. En consecuencia, para los codificadores de vídeo puede ser necesario un nuevo módulo de hardware que soporta la predicción de síntesis de visualización basada en bloques. Además, cuando una secuencia de imágenes tiene una imagen de síntesis de visualización asociada (seq_view_synthesis_flag es igual a 1), un codificador de vídeo debe generar valores de índice de referencia adicionales para las imágenes de síntesis de visualización, de modo que los valores de índice de referencia estén disponibles para asignarse en instancias en que se genera una imagen de síntesis de visualización y se utiliza con fines de predicción. En consecuencia, el codificador de vídeo genera valores de índice de referencia adicionales, incluso cuando una imagen de síntesis de visualización no se genera realmente durante la codificación, lo cual afecta la eficiencia computacional y/o de codificación.

10 **[0093]** Los aspectos de esta divulgación incluyen técnicas para indicar si aplicar la predicción de síntesis de visualización al codificar un bloque de datos de vídeo. En un ejemplo, un codificador de vídeo, como el codificador de vídeo 20 y/o el descodificador de vídeo 30, puede determinar si un índice de referencia para un bloque actual corresponde a una imagen de referencia entre visualizaciones, y cuando el índice de referencia para el bloque actual corresponde a la imagen de referencia entre visualizaciones, determinar los datos que indican un modo VSP del bloque actual, donde el modo VSP para el índice de referencia indica si el bloque actual se predice utilizando la predicción de síntesis de visualización a partir de la imagen entre visualizaciones (por ejemplo, a partir de al menos una parte de una imagen sintetizada a partir de la imagen entre visualizaciones). En los casos en que el codificador de vídeo está configurado como descodificador de vídeo 30, el descodificador de vídeo 30 puede determinar los datos que indican el modo VSP obteniendo los datos de un flujo de bits codificado.

20 **[0094]** En consecuencia, en el ejemplo anterior, un índice de referencia de una imagen de VSP no se indica específicamente en una cabecera de fragmento ni se agrega durante la construcción de la lista de imágenes de referencia. En su lugar, el codificador de vídeo 20 puede codificar (y el descodificador de vídeo 30 puede obtener y descodificar, a partir de un flujo de bits codificado) uno o más elementos sintácticos en el nivel de MB o partición de MB para cada índice de referencia que corresponde a una imagen de referencia con predicción interna. Como se describe en el presente documento, una imagen de referencia de predicción inter se asocia en general con un modo de predicción inter. Es decir, una imagen de predicción inter puede incluir una imagen de referencia intra-visualización o una imagen de referencia entre visualizaciones.

30 **[0095]** En algunos ejemplos, uno o más elementos sintácticos pueden ser un indicador de VSP. Por ejemplo, el indicador puede indicar si VSP se aplica al MB o la partición de MB para un índice de referencia particular. En otras palabras, por ejemplo, el indicador puede indicar si usar una imagen de referencia entre visualizaciones con fines de predicción, o usar una imagen VSP con fines de predicción.

40 **[0096]** En un ejemplo, de acuerdo con los aspectos de esta divulgación, el codificador de vídeo 20 puede codificar solo un indicador de VSP cuando un índice de referencia para codificar de manera predictiva un bloque corresponde a una imagen de referencia entre visualizaciones. En otro ejemplo, el codificador de vídeo 20 puede codificar el indicador de VSP para cualquier modo con predicción inter. En otro ejemplo, de forma adicional o alternativa, el codificador de vídeo 20 puede codificar el indicador solo para un índice de referencia correspondiente a RefPicListO. En otro ejemplo, el codificador de vídeo 20 puede codificar el indicador para un índice de referencia correspondiente a RefPicListO y RefPicList1 (si está presente).

50 **[0097]** De acuerdo con otros aspectos de esta divulgación, cuando se habilita la predicción de síntesis de visualización, el codificador de vídeo 20 puede codificar (y el descodificador de vídeo 30 puede obtener, a partir de un flujo de bits codificado) uno o más elementos sintácticos en un conjunto de parámetros, como una extensión SPS o un SPS indicando si la señalización basada en referencia se usa para VSP (por ejemplo, en la que se genera un índice de imagen de referencia para una imagen de referencia de síntesis de visualización y VSP se realiza basándose en el índice de referencia), por ejemplo, un primer modo de VSP, o si se usa GVSP para VSP (por ejemplo, en el que VSP se realiza basándose en la señalización de macrobloques descrita anteriormente, por ejemplo, con respecto a las Tablas 1-3 anteriores), por ejemplo, un segundo modo VSP. En algunos ejemplos, el uno o más elementos sintácticos pueden incluir un indicador en una extensión SPS.

60 **[0098]** De acuerdo con otros aspectos de esta divulgación, cuando no se utiliza la señalización basada en el índice de referencia, un proceso de modificación de la lista de imágenes de referencia (RPLM) no debe contener ningún comando RPLM que corresponda a una imagen VSP. Por ejemplo, la construcción de la lista de imágenes de referencia puede incluir los siguientes pasos: (1) se aplica el proceso de inicialización de la lista de imágenes de referencia para las imágenes de referencia temporales (intra-visualización) como se especifica en AVC (2) las imágenes de referencia sintetizadas y/o entre visualizaciones se incorporan al final de la lista, y (3) una proceso RPLM se aplica para reordenar las imágenes de referencia en la lista de imágenes de referencia. El codificador de

65

vídeo 20 puede emitir uno o más comandos RPLM que ordenan al descodificador de vídeo 30 aplicar el proceso RPLM. El descodificador de vídeo 30 puede aplicar el proceso RPLM en respuesta a los comandos RPLM.

5 **[0099]** De acuerdo con los aspectos de esta divulgación, cuando no se utiliza la señalización basada en el índice de referencia, el codificador de vídeo 20 y el descodificador de vídeo 30 pueden no emitir un comando RPLM que corresponda a una imagen VSP. Es decir, el codificador de vídeo 20 y el descodificador de vídeo 30 pueden abstenerse de usar un RPLM para cambiar la posición de una imagen VSP, porque no se ha generado ninguna imagen VSP.

10 **[0100]** De acuerdo con otros aspectos de esta divulgación, cuando se usa GVSP, el codificador de vídeo 20 y el descodificador de vídeo 30 pueden sintetizar una imagen de predicción de síntesis de visualización solo desde la primera imagen de referencia entre visualizaciones para RefPicList0. En este ejemplo, no es necesaria la señalización de la imagen de predicción de síntesis de visualización, y se determina (por ejemplo, se deduce) automáticamente que el modo GVSP genera un vector de movimiento de disparidad que apunta a la primera imagen de referencia de la visualización, por ejemplo, RefPicList 0. De forma alternativa, en otro ejemplo, el
15 el codificador de vídeo 20 y el descodificador de vídeo 30 pueden realizar la predicción de síntesis de visualización basándose en una imagen entre visualizaciones en RefPicList1. De forma alternativa, en otro ejemplo, el codificador de vídeo 20 y el descodificador de vídeo 30 pueden realizar la predicción de síntesis de visualización basándose en las imágenes entre visualizaciones, si están presentes, tanto en RefPicList0 como en RefPicList1.

20 **[0101]** La FIG. 2 es un diagrama de bloques que ilustra un ejemplo de codificador de vídeo 20 que puede implementar las técnicas descritas en esta divulgación para síntesis de visualización. El codificador 20 de vídeo puede realizar intracodificación e intercodificación de bloques de vídeo dentro de fragmentos de vídeo. La intracodificación se basa en la predicción espacial para reducir o eliminar la redundancia espacial en el vídeo dentro de una imagen dada. La intercodificación se basa en la predicción temporal para reducir o eliminar la redundancia temporal en el vídeo dentro de imágenes adyacentes o imágenes de una secuencia de vídeo. El modo intra (modo I) puede referirse a cualquiera de varios modos de compresión espacial. Los modos inter, tales como la predicción unidireccional (modo P) o la predicción bidireccional (modo B), pueden referirse a cualquiera de varios modos de compresión temporal.

25 **[0102]** Como se ha indicado anteriormente, el descodificador de vídeo 20 puede adaptarse para realizar una codificación de vídeo de múltiples visualizaciones. Además de codificar mapas de textura (es decir, valores de luma y croma) para cada visualización, el codificador de vídeo 20 puede codificar adicionalmente un mapa de profundidad para cada visualización.

30 **[0103]** Como se muestra en la FIG. 2, el codificador 20 de vídeo recibe un bloque de vídeo actual dentro de una trama de vídeo que se va a codificar. En el ejemplo de la FIG. 2, el codificador de vídeo 20 incluye una unidad de selección de modo 40, una memoria de imágenes de referencia 64, un sumador 50, una unidad de procesamiento de transformada 52, una unidad de cuantificación 54 y una unidad de codificación por entropía 56. A su vez, la
35 unidad de selección de modo 40 incluye una unidad de compensación de movimiento 44, una unidad de estimación de movimiento 42, una unidad de predicción intra 46 y una unidad de división 48. Para la reconstrucción de bloques de vídeo, el codificador 20 de vídeo incluye también la unidad 58 de cuantificación inversa, la unidad 60 de transformada inversa y el sumador 62.

40 **[0104]** Puede incluirse, asimismo, un filtro de eliminación de bloques (no mostrado en la FIG. 2) para filtrar fronteras de bloques, con el fin de eliminar distorsiones de efecto pixelado del vídeo reconstruido. Si se desea, el filtro de eliminación de bloques filtrará típicamente la salida del sumador 62. También pueden usarse filtros adicionales (en bucle o bucle posterior), además del filtro de eliminación de bloques. Dichos filtros no se muestran por razones de brevedad pero, si se desea, pueden filtrar la salida del sumador 50 (tal como un filtro en bucle).

45 **[0105]** Durante el proceso de codificación, el codificador 20 de vídeo recibe una trama o un fragmento de vídeo que se va a codificar. La trama o el fragmento pueden dividirse en múltiples bloques de vídeo. La unidad de estimación de movimiento 42 y la unidad de compensación de movimiento 44 llevan a cabo la codificación interpredictiva del bloque de vídeo recibido con respecto a uno o más bloques en una o más tramas de referencia, para proporcionar la predicción temporal. La unidad de predicción intra 46, de forma alternativa, puede llevar a
50 cabo la codificación intrapredictiva del bloque de vídeo recibido, con respecto a uno o más bloques contiguos en la misma trama o fragmento que el bloque a codificar, para proporcionar predicción espacial. El codificador de vídeo 20 puede llevar a cabo múltiples pases de codificación, por ejemplo, para seleccionar un modo de codificación adecuada para cada bloque de datos de vídeo.

55 **[0106]** Además, la unidad de división 48 puede dividir bloques de datos de vídeo en sub-bloques, basándose en la evaluación de los anteriores esquemas de división en los anteriores pases de codificación. Por ejemplo, la unidad de división 48 puede dividir inicialmente una trama o un fragmento en bloques (por ejemplo, macrobloques o LCU), y dividir cada uno de los sub- bloques de bloques basándose en un análisis de velocidad-distorsión (por ejemplo, optimización de velocidad-distorsión).

65

[0107] La unidad de selección de modo 40 puede seleccionar uno de los modos de codificación, como predicción intra, inter temporal, o entre visualizaciones, por ejemplo, basándose en los resultados de error, y proporciona un bloque predicho al sumador 50 para generar datos de bloque residual, y al sumador 62 para reconstruir el bloque codificado para su uso como una trama de referencia. La unidad de selección de modo 40 proporciona además elementos sintácticos, tales como vectores de movimiento, indicadores de intra-modo, información de división y otra información sintáctica de este tipo, a la unidad de codificación por entropía 56.

[0108] En algunos ejemplos, la unidad de selección de modo 40 puede configurarse para seleccionar la predicción entre visualizaciones, por ejemplo, en relación con una visualización codificada previamente o una visualización sintetizada para la predicción de síntesis de visualización. Como se analiza con mayor detalle a continuación, la unidad de síntesis de visualización 66 puede configurarse para sintetizar una visualización (es decir, sintetizar imágenes, incluidos los valores de textura y/o píxeles de profundidad) para la predicción de síntesis de visualización. Por ejemplo, la unidad de síntesis de visualización 66 puede configurarse para realizar técnicas sustancialmente similares a la renderización basada en imágenes en profundidad (DIBR).

[0109] Debe entenderse que cuando la predicción de síntesis de visualización está habilitada, la unidad de selección de modo 40 aún puede seleccionar entre los otros modos de codificación disponibles, por ejemplo, predicción intra, predicción temporal inter o predicción entre visualizaciones en relación con una imagen codificada previamente. Por lo tanto, el codificador de vídeo 20 puede proporcionar una indicación de qué modo de codificación se selecciona para un bloque de datos de vídeo, separado de la información que indica si la predicción de síntesis de visualización está habilitada.

[0110] La unidad de estimación de movimiento 42 y la unidad de compensación de movimiento 44 (así como una o más unidades del codificador de vídeo 20) pueden estar altamente integradas, pero se ilustran por separado con fines conceptuales. La estimación de movimiento, realizada por la unidad de estimación de movimiento 42, es el proceso de generación de vectores de movimiento, que estiman el movimiento de los bloques de vídeo. Un vector de movimiento, por ejemplo, puede indicar el desplazamiento de un bloque de vídeo dentro de una trama o imagen de vídeo actual, con respecto a un bloque predictivo dentro de una trama de referencia (u otra unidad codificada), con respecto al bloque actual que se está codificando dentro de la trama actual (u otra unidad codificada).

[0111] Un bloque predictivo es un bloque que se revela como estrechamente coincidente con el bloque que se va a codificar, en lo que respecta a la diferencia de píxeles, lo cual puede determinarse mediante la suma de una diferencia absoluta (SAD), suma de diferencia de cuadrados (SSD) u otras métricas de diferencia. Los vectores de movimiento pueden incluir vectores de movimiento temporal, que describen el movimiento de un bloque en relación con un bloque codificado previamente de la misma visualización en una imagen temporalmente distinta, y vectores de movimiento de disparidad, que describen la disparidad entre bloques similares en diferentes visualizaciones (con diferentes perspectivas de cámara horizontales) pero que pueden tener la misma posición temporal. En algunos casos, los vectores de movimiento pueden describir el movimiento con respecto a una imagen temporalmente distinta que también se encuentra en una visualización diferente.

[0112] En algunos ejemplos, el codificador 20 de vídeo puede calcular valores para posiciones de píxel de subentero de imágenes de referencia almacenadas en la memoria 64 de imágenes de referencia. Por ejemplo, el codificador 20 de vídeo puede interpolar valores de posiciones de un cuarto de píxel, posiciones de un octavo de píxel u otras posiciones de píxel fraccionario de la imagen de referencia. Por lo tanto, la unidad de estimación de movimiento 42 puede realizar una búsqueda de movimiento en relación con las posiciones de píxeles completas y las posiciones de píxeles fraccionarias, y generar un vector de movimiento con una precisión de píxel fraccionaria.

[0113] Al realizar la predicción entre visualizaciones, la unidad de estimación de movimiento 42 puede calcular vectores de movimiento de disparidad en relación con imágenes codificadas previamente de una visualización diferente o imágenes sintetizadas para la predicción de síntesis de visualización. Por lo tanto, la unidad de estimación de movimiento 42 también puede denominarse unidad de estimación de movimiento/disparidad.

[0114] En general, las imágenes sintetizadas pueden almacenarse en la memoria de imágenes de referencia 64 mediante la unidad de síntesis de visualización 66 y, por lo tanto, la unidad de estimación de movimiento 42 y la unidad de compensación de movimiento 44 no necesitan configurarse para determinar si una imagen de referencia es una imagen codificada previamente de una visualización diferente o una imagen sintetizada a partir de un proceso de síntesis de visualización. El proceso para buscar un vector de movimiento de disparidad se puede restringir a la búsqueda horizontal, en lugar de horizontal y vertical, porque las imágenes de diferentes visualizaciones en la misma posición temporal típicamente solo incluyen diferencias horizontales y no verticales debido a las imágenes de una escena capturadas o generadas desde perspectivas de cámara horizontales dispares.

[0115] La unidad de estimación de movimiento 42 calcula un vector de movimiento (por ejemplo, un vector de movimiento temporal o un vector de movimiento de disparidad) para un bloque de vídeo en un fragmento intercodificado, comparando la posición del bloque con la posición de un bloque predictivo de una imagen de referencia. La imagen de referencia puede seleccionarse de una primera lista de imágenes de referencia (lista 0)

o una segunda lista de imágenes de referencia (lista 1), cada una de las cuales identifica una o más imágenes de referencia almacenadas en la memoria 64 de imágenes de referencia. La unidad de estimación de movimiento 42 envía el vector de movimiento calculado a la unidad de codificación por entropía 56 y a la unidad de compensación de movimiento 44.

[0116] La compensación de movimiento, llevada a cabo por la unidad de compensación de movimiento 44, puede implicar capturar o generar el bloque predictivo basándose en el vector de movimiento determinado por la unidad de estimación de movimiento 42 que nuevamente puede ser un vector de movimiento temporal o un vector de movimiento de disparidad. De nuevo, la unidad 42 de estimación de movimiento y la unidad 44 de compensación de movimiento pueden integrarse funcionalmente, en algunos ejemplos. Tras recibir el vector de movimiento para el bloque de vídeo actual, la unidad de compensación de movimiento 44 puede localizar el bloque predictivo al que apunta el vector de movimiento en una de las listas de imágenes de referencia. El sumador 50 forma un bloque de vídeo residual restando los valores de píxel del bloque predictivo a los valores de píxel del bloque de vídeo actual que se está codificando, generando valores de diferencia de píxel, como se analiza posteriormente.

[0117] En general, la unidad de estimación de movimiento 42 lleva a cabo la estimación de movimiento con respecto a los componentes de luma, y la unidad de compensación de movimiento 44 utiliza los vectores de movimiento calculados basándose en los componentes de luma, tanto para los componentes de croma como para los componentes de luma. La unidad de estimación de movimiento 42 y la unidad de compensación de movimiento 44 pueden configurarse para reutilizar vectores de movimiento de componentes de luma para codificar mapas de profundidad, o para calcular independientemente vectores de movimiento para mapas de profundidad. Por lo tanto, en algunos casos, la estimación de movimiento 42 y la unidad de compensación de movimiento 44 pueden configurarse para predecir mapas de profundidad de una manera similar a los componentes de croma o de una manera similar a los componentes de luma. La unidad de selección de modo 40 también puede generar elementos sintácticos asociados a los bloques de vídeo y el fragmento de vídeo para su uso por el descodificador de vídeo 30 al descodificar los bloques de vídeo del fragmento de vídeo.

[0118] La unidad de predicción intra 46 puede realizar la predicción intra de un bloque actual, de forma alternativa a la predicción inter (por ejemplo, predicción entre visualizaciones y/o predicción inter temporal) realizada por la unidad de estimación de movimiento 42 y la unidad de compensación de movimiento 44, como se ha descrito anteriormente. En particular, la unidad 46 de predicción intra puede determinar un modo de predicción intra para usar con el fin de codificar un bloque actual. En algunos ejemplos, la unidad 46 de predicción intra puede codificar un bloque actual usando diversos modos de predicción intra, por ejemplo, durante diferentes pasadas de codificación, y la unidad 46 de predicción intra (o la unidad 40 de selección de modo, en algunos ejemplos) puede seleccionar un modo de predicción intra adecuado para usar a partir de los modos probados.

[0119] Por ejemplo, la unidad 46 de predicción intra puede calcular valores de velocidad-distorsión usando un análisis de velocidad-distorsión para los diversos modos de predicción intra probados, y seleccionar el modo de predicción intra que tenga las mejores características de velocidad-distorsión entre los modos probados. El análisis de velocidad-distorsión determina en general una magnitud de distorsión (o de errores) entre un bloque codificado y un bloque original, no codificado, que se codificó para generar el bloque codificado, así como una velocidad binaria (es decir, un número de bits) usada para generar el bloque codificado. La unidad 46 de predicción intra puede calcular razones a partir de las distorsiones y velocidades para los diversos bloques codificados, con el fin de determinar qué modo de predicción intra presenta el mejor valor de velocidad-distorsión para el bloque.

[0120] Después de seleccionar un modo de predicción intra para un bloque, la unidad de predicción intra 46 puede proporcionar información, indicativa del modo de predicción intra seleccionado para el bloque, a la unidad de codificación por entropía 56. La unidad de codificación por entropía 56 puede codificar la información indicando el modo de predicción intra seleccionado. El codificador de vídeo 20 puede incluir datos de configuración en el flujo de bits transmitido, que pueden incluir una pluralidad de tablas de índices de modos de predicción intra y una pluralidad de tablas de índices de modos de predicción intra modificadas (también denominadas tablas de correlación de palabras de código), definiciones de contextos de codificación para varios bloques e indicaciones del modo de predicción intra más probable, una tabla de índices de modos de predicción intra y una tabla de índices de modos de predicción intra modificadas a utilizar para cada uno de los contextos.

[0121] Además, como se señaló anteriormente, la unidad de selección de modo 40 puede determinar si utilizar la predicción de síntesis de visualización para codificar un bloque particular de una visualización no básica. La unidad de selección de modo 40 puede recibir datos de configuración que indican si la predicción de síntesis de visualización está habilitada para un flujo de bits o para un punto de funcionamiento particular de un flujo de bits de múltiples visualizaciones. Por ejemplo, un usuario puede haber proporcionado datos de configuración que indiquen que el flujo de bits debe ser compatible con dispositivos que no están equipados para realizar la predicción de síntesis de visualización, en cuyo caso, la unidad de selección de modo 40 puede inhabilitar la predicción de síntesis de visualización para todo el flujo de bits. De forma alternativa, los datos de configuración pueden indicar un subconjunto de visualizaciones, formando un punto de funcionamiento, para el cual la predicción de síntesis de visualización está deshabilitada, en cuyo caso la unidad de selección de modo 40 puede deshabilitar la predicción

de síntesis de visualización para cada una de las visualizaciones en el subconjunto, pero probar la síntesis de visualización de prueba predicción para otras visualizaciones del flujo de bits no incluidas en el subconjunto.

5 **[0122]** En general, suponiendo que la predicción de síntesis de visualización está permitida para una visualización no base particular, la unidad de selección de modo 40 puede probar la predicción de síntesis de visualización para bloques de imágenes de la visualización no base para determinar si la predicción de síntesis de visualización produce un mejor rendimiento que otros modos de codificación, por ejemplo, la predicción intra temporal, la predicción intra y/o la predicción entre visualizaciones en relación con una visualización previamente codificada (no sintetizada). La unidad de selección de modo 40 puede probar el rendimiento de varios modos de codificación usando optimización de velocidad-distorsión (RDO) y/o usando velocidades de Bjontegaard-Delta (velocidades de BD). En general, la optimización de la velocidad-distorsión es una medida del número de bits necesarios para codificar un conjunto de datos de vídeo usando un modo particular o una combinación de modos, en relación con la cantidad de distorsión introducida por el modo o combinación de modos. La unidad de selección de modo 40 puede seleccionar el modo o combinación de modos que ofrezcan el mejor rendimiento según lo indicado por estas medidas.

20 **[0123]** Además, la unidad de selección de modo 40 puede determinar un número de bloques en un fragmento, imagen, secuencia u otra unidad codificada (por ejemplo, un elemento o frente de onda) para los cuales la predicción de síntesis de visualización produce el mejor rendimiento y determinar si este número de bloques es suficientemente alto para justificar el gasto de procesamiento de usar la predicción de síntesis de visualización. Por ejemplo, sintetizar una visualización mediante un descodificador de vídeo puede requerir una cantidad bastante grande de recursos de procesamiento, y por lo tanto, la unidad de selección de modo 40 puede seleccionar un modo de codificación diferente para un bloque particular aparte de la predicción de síntesis de visualización si no hay un número suficientemente grande de bloques en la unidad codificada para la cual se utilizará la predicción de síntesis de visualización.

30 **[0124]** La unidad de síntesis de visualización 66 representa una unidad que sintetiza visualizaciones para la predicción entre visualizaciones. Cuando se habilita la predicción de síntesis de visualización, la unidad de síntesis de visualización 66 puede sintetizar una imagen para una visualización sintetizada utilizando información de textura y profundidad almacenada en la memoria de imágenes de referencia 64. La unidad de síntesis de visualización 66 puede usar una o más visualizaciones de referencia para sintetizar otra visualización.

35 **[0125]** En general, para sintetizar una imagen, la unidad de síntesis de visualización 66 puede usar información de textura y profundidad de una o más visualizaciones previamente codificadas. La unidad de síntesis de visualización 66 puede calcular la disparidad horizontal para la información de textura basándose en la información de profundidad correspondiente (por ejemplo, de ubicación conjunta) de una imagen de textura y un mapa de profundidad correspondiente. En general, los objetos que deben aparecer en la profundidad de la pantalla (por ejemplo, en el plano de convergencia) pueden tener una disparidad de cero. Es decir, los píxeles representativos de este objeto pueden tener sustancialmente la misma posición horizontal en la visualización sintetizada que en la visualización de referencia. Para los objetos que se muestran delante de la pantalla, se puede asignar una disparidad positiva, de manera que en una imagen del "ojo izquierdo", los píxeles del objeto se posicionan a la derecha de los píxeles correspondientes en la imagen del "ojo derecho". De forma alternativa, para que los objetos se muestren detrás de la pantalla, se puede asignar una disparidad negativa, de manera que en una imagen del "ojo izquierdo", los píxeles del objeto se posicionan a la izquierda de los píxeles correspondientes en la imagen del "ojo derecho". El desplazamiento horizontal relativo se puede determinar en función de la cantidad de profundidad a alcanzar, el desplazamiento horizontal relativo, la distancia al plano de convergencia, las distancias del mundo real, los parámetros de la cámara y similares.

50 **[0126]** De esta manera, la unidad de síntesis de visualización 66 puede sintetizar una imagen de una visualización para usarla como imagen de referencia. La unidad de síntesis de visualización 66 puede usar procesos similares para sintetizar una o ambas imágenes de textura y/o un mapa de profundidad. Por lo tanto, el codificador de vídeo 20 puede codificar una imagen de textura y/o un mapa de profundidad en relación con un componente de visualización sintetizado, por ejemplo, una imagen de textura sintetizada o un mapa de profundidad sintetizado. La unidad de síntesis de visualización 66 puede almacenar datos para el componente de visualización sintetizado en la memoria de imágenes de referencia 64. De esta manera, la unidad de selección de modo 40, la unidad de estimación de movimiento 42 y la unidad de compensación de movimiento 44 pueden tratar el componente de visualización sintetizado como si el componente de visualización sintetizado fuera una imagen de referencia convencional.

60 **[0127]** De acuerdo con los aspectos de esta divulgación, la unidad de selección de modo 40 puede generar uno o más elementos sintácticos que indican si aplicar VSP para codificar un bloque particular. En algunos ejemplos, la unidad de selección de modo 40 solo puede generar los elementos sintácticos para los bloques que tienen predicción inter, como la predicción entre visualizaciones. Por lo tanto, en este ejemplo, cuando un índice de imagen de referencia para el bloque actual corresponde a una imagen de referencia entre visualizaciones, la unidad de selección de modo 40 puede enviar un indicador a la unidad de codificación por entropía 56 para indicar si se usa VSP para codificar el bloque actual.

- 5 **[0128]** En el ejemplo anterior, un índice de referencia de una imagen VSP no se indica específicamente en una cabecera de fragmento o se agrega durante la construcción de la lista de imágenes de referencia. En su lugar, el codificador de vídeo 20 puede codificar un indicador en el nivel de partición de MB o el MB para cada índice de referencia que corresponde a una imagen de referencia entre visualizaciones. Dicha señalización puede proporcionar un enfoque más flexible a la señalización de VSP, ya que la imagen sintetizada no está vinculada a un índice de imagen de referencia particular.
- 10 **[0129]** En otro ejemplo, la unidad de selección de modo 40 puede generar el indicador de VSP para cualquier modo de predicción inter. En otro ejemplo, de forma adicional o alternativa, la unidad de selección de modo 40 puede generar el indicador solo para un índice de referencia correspondiente a RefPicList0. En otro ejemplo más, la unidad de selección de modo 40 puede generar el indicador para un índice de referencia correspondiente a RefPicList0 y RefPicList1 (si está presente).
- 15 **[0130]** De acuerdo con otros aspectos de esta divulgación, cuando se habilita la predicción de síntesis de visualización, la unidad de selección de modo 40 puede generar uno o más elementos sintácticos para un conjunto de parámetros, como una extensión SPS o un SPS que indica si la señalización basada en referencias se utiliza para VSP (por ejemplo, en el que se genera un índice de imagen de referencia para una imagen de referencia de síntesis de visualización y VSP se realiza basándose en el índice de referencia) o si se utiliza GVSP para VSP (por ejemplo, en el que se realiza VSP basándose en la señalización de macrobloques descrita anteriormente, por ejemplo, con respecto a las Tablas 1-3 anteriores). En algunos ejemplos, el uno o más elementos sintácticos pueden incluir un indicador en una extensión SPS.
- 20 **[0131]** De acuerdo con otros aspectos de esta divulgación, cuando no se utiliza la señalización basada en el índice de referencia, el codificador de vídeo 20 puede no emitir ningún comando RPLM que corresponda a una imagen VSP.
- 25 **[0132]** De acuerdo con otros aspectos de esta divulgación, cuando se usa GVSP, la unidad de síntesis de visualización 66 puede sintetizar una imagen de predicción de síntesis de visualización solo desde la primera imagen de referencia entre visualizaciones para RefPicList0. En este ejemplo, no es necesaria la señalización de la imagen de predicción de síntesis de visualización, y se determina (por ejemplo, se deduce) automáticamente que el modo GVSP genera un vector de movimiento de disparidad que apunta a la primera imagen de referencia de la visualización, por ejemplo, RefPicList 0. De forma alternativa, en otro ejemplo, la unidad de síntesis de visualización 66 puede realizar una predicción de síntesis de visualización basada en una imagen entre visualizaciones en RefPicList1. De forma alternativa, en otro ejemplo, la unidad de síntesis de visualización 66 puede realizar una predicción de síntesis de visualización basada en imágenes entre visualizaciones, si están presentes, tanto en RefPicList0 como en RefPicList1.
- 30 **[0133]** El codificador de vídeo 20 forma un bloque de vídeo residual restando los datos de predicción de la unidad de selección de modo 40 del bloque de vídeo original que se está codificando. El sumador 50 representa el componente o los componentes que realizan esta operación de resta. La unidad de procesamiento de transformadas 52 aplica una transformada, tal como una transformada discreta de coseno (DCT) o una transformada conceptualmente similar, al bloque residual, generando un bloque de vídeo que comprende valores residuales de coeficientes de transformada. La unidad de procesamiento de transformadas 52 puede llevar a cabo otras transformadas que son conceptualmente similares a la DCT. También podrían usarse transformadas de ondícula, transformadas de números enteros, transformadas de sub-bandas u otros tipos de transformadas.
- 35 **[0134]** En cualquier caso, la unidad de procesamiento de transformadas 52 aplica la transformada al bloque residual, generando un bloque de coeficientes de transformada residuales. La transformada puede convertir la información residual, desde un dominio de valor de píxel a un dominio de transformada, tal como un dominio de frecuencia. La unidad de procesamiento de transformadas 52 puede enviar los coeficientes de transformada resultantes a la unidad de cuantificación 54. La unidad de cuantificación 54 cuantifica los coeficientes de transformada para reducir adicionalmente la velocidad de transmisión de bits. El proceso de cuantificación puede reducir la profundidad de bits asociada a algunos, o a la totalidad, de los coeficientes. El grado de cuantificación puede modificarse ajustando un parámetro de cuantificación. En algunos ejemplos, la unidad de cuantificación 54 puede realizar, a continuación, un recorrido de la matriz que incluye los coeficientes de transformada cuantificados. De forma alternativa, la unidad de codificación por entropía 56 puede llevar a cabo el recorrido.
- 40 **[0135]** Tras la cuantificación, la unidad de codificación por entropía 56 codifica por entropía los coeficientes de transformada cuantificados. Por ejemplo, la unidad de codificación por entropía 56 puede llevar a cabo la codificación de longitud variable adaptable al contexto (CAVLC), la codificación aritmética binaria adaptable al contexto (CABAC), la codificación aritmética binaria adaptable al contexto y basada en sintaxis (SBAC), la codificación por entropía mediante la división en intervalos de probabilidades (PIPE) u otra técnica de codificación por entropía. En el caso de la codificación por entropía basada en el contexto, el contexto puede basarse en bloques contiguos. Tras la codificación por entropía realizada por la unidad de codificación por entropía 56, el flujo
- 45 **[0135]** Tras la cuantificación, la unidad de codificación por entropía 56 codifica por entropía los coeficientes de transformada cuantificados. Por ejemplo, la unidad de codificación por entropía 56 puede llevar a cabo la codificación de longitud variable adaptable al contexto (CAVLC), la codificación aritmética binaria adaptable al contexto (CABAC), la codificación aritmética binaria adaptable al contexto y basada en sintaxis (SBAC), la codificación por entropía mediante la división en intervalos de probabilidades (PIPE) u otra técnica de codificación por entropía. En el caso de la codificación por entropía basada en el contexto, el contexto puede basarse en bloques contiguos. Tras la codificación por entropía realizada por la unidad de codificación por entropía 56, el flujo
- 50 **[0135]** Tras la cuantificación, la unidad de codificación por entropía 56 codifica por entropía los coeficientes de transformada cuantificados. Por ejemplo, la unidad de codificación por entropía 56 puede llevar a cabo la codificación de longitud variable adaptable al contexto (CAVLC), la codificación aritmética binaria adaptable al contexto (CABAC), la codificación aritmética binaria adaptable al contexto y basada en sintaxis (SBAC), la codificación por entropía mediante la división en intervalos de probabilidades (PIPE) u otra técnica de codificación por entropía. En el caso de la codificación por entropía basada en el contexto, el contexto puede basarse en bloques contiguos. Tras la codificación por entropía realizada por la unidad de codificación por entropía 56, el flujo
- 55 **[0135]** Tras la cuantificación, la unidad de codificación por entropía 56 codifica por entropía los coeficientes de transformada cuantificados. Por ejemplo, la unidad de codificación por entropía 56 puede llevar a cabo la codificación de longitud variable adaptable al contexto (CAVLC), la codificación aritmética binaria adaptable al contexto (CABAC), la codificación aritmética binaria adaptable al contexto y basada en sintaxis (SBAC), la codificación por entropía mediante la división en intervalos de probabilidades (PIPE) u otra técnica de codificación por entropía. En el caso de la codificación por entropía basada en el contexto, el contexto puede basarse en bloques contiguos. Tras la codificación por entropía realizada por la unidad de codificación por entropía 56, el flujo
- 60 **[0135]** Tras la cuantificación, la unidad de codificación por entropía 56 codifica por entropía los coeficientes de transformada cuantificados. Por ejemplo, la unidad de codificación por entropía 56 puede llevar a cabo la codificación de longitud variable adaptable al contexto (CAVLC), la codificación aritmética binaria adaptable al contexto (CABAC), la codificación aritmética binaria adaptable al contexto y basada en sintaxis (SBAC), la codificación por entropía mediante la división en intervalos de probabilidades (PIPE) u otra técnica de codificación por entropía. En el caso de la codificación por entropía basada en el contexto, el contexto puede basarse en bloques contiguos. Tras la codificación por entropía realizada por la unidad de codificación por entropía 56, el flujo
- 65 **[0135]** Tras la cuantificación, la unidad de codificación por entropía 56 codifica por entropía los coeficientes de transformada cuantificados. Por ejemplo, la unidad de codificación por entropía 56 puede llevar a cabo la codificación de longitud variable adaptable al contexto (CAVLC), la codificación aritmética binaria adaptable al contexto (CABAC), la codificación aritmética binaria adaptable al contexto y basada en sintaxis (SBAC), la codificación por entropía mediante la división en intervalos de probabilidades (PIPE) u otra técnica de codificación por entropía. En el caso de la codificación por entropía basada en el contexto, el contexto puede basarse en bloques contiguos. Tras la codificación por entropía realizada por la unidad de codificación por entropía 56, el flujo

de bits codificado puede transmitirse a otro dispositivo (por ejemplo, el descodificador de vídeo 30) o archivar para su posterior transmisión o recuperación.

[0136] La unidad 58 de cuantificación inversa y la unidad 60 de transformada inversa aplican la cuantificación inversa y la transformada inversa, respectivamente, para reconstruir el bloque residual en el dominio del píxel, por ejemplo, para su uso posterior como bloque de referencia. La unidad de compensación de movimiento 44 puede calcular un bloque de referencia añadiendo el bloque residual a un bloque predictivo de una de las tramas de la memoria de imágenes de referencia 64. La unidad 44 de compensación de movimiento también puede aplicar uno o más filtros de interpolación al bloque residual reconstruido para calcular valores de píxeles de subentero para su uso en la estimación de movimiento.

[0137] El sumador 62 añade el bloque residual reconstruido al bloque de predicción compensado por movimiento, generado por la unidad de compensación de movimiento 44, para generar un bloque de vídeo reconstruido para su almacenamiento en la memoria de imágenes de referencia 64. El bloque de vídeo reconstruido puede ser utilizado por la unidad 42 de estimación de movimiento y la unidad 44 de compensación de movimiento, como bloque de referencia para intercodificar un bloque en una trama de vídeo subsiguiente.

[0138] De acuerdo con los aspectos de esta divulgación, el codificador de vídeo 20 puede realizar un procedimiento que incluya determinar si un índice de referencia para un bloque actual corresponde a una imagen de referencia entre visualizaciones, y cuando el índice de referencia para el bloque actual corresponde a la imagen de referencia entre visualizaciones, generar datos que indiquen un modo de predicción de síntesis de visualización (VSP) del bloque actual, en el que el modo de VSP para el índice de referencia indica si el bloque actual se predice con la predicción de síntesis de visualización a partir de la imagen de referencia entre visualizaciones. El codificador de vídeo 20 también está configurado para codificar los datos que indican el modo VSP en un flujo de bits.

[0139] La FIG. 3 es un diagrama de bloques que ilustra un ejemplo de descodificador de vídeo 30 que puede implementar técnicas para codificar información relacionada con la predicción de síntesis de visualización. En el ejemplo de la FIG. 3, el descodificador de vídeo 30 incluye una unidad de descodificación por entropía 70, unidad de compensación de movimiento 72, unidad de predicción intra 74, unidad de cuantificación inversa 76, unidad de transformada inversa 78, memoria de imágenes de referencia 82 y sumador 80. En algunos ejemplos, el descodificador 30 de vídeo puede realizar una pasada de descodificación en general recíproca a la pasada de codificación descrita con respecto al codificador 20 de vídeo (FIG. 2). La unidad de compensación de movimiento 72 puede generar datos de predicción basándose en vectores de movimiento recibidos desde la unidad de descodificación por entropía 70, mientras que la unidad de predicción intra 74 puede generar datos de predicción basándose en indicadores de modo de predicción intra recibidos de la unidad de descodificación por entropía 70.

[0140] Durante el proceso de descodificación, el descodificador 30 de vídeo recibe un flujo de bits de vídeo codificado que representa bloques de vídeo de un fragmento de vídeo codificado y elementos sintácticos asociados, desde el codificador 20 de vídeo. La unidad de descodificación por entropía 70 descodifica el flujo de bits para generar coeficientes cuantificados, vectores de movimiento o indicadores de modo de predicción intra y otros elementos sintácticos. La unidad de descodificación por entropía 70 remite los vectores de movimiento y otros elementos sintácticos a la unidad de compensación de movimiento 72. El descodificador 30 de vídeo puede recibir los elementos sintácticos en el nivel del fragmento de vídeo y/o el nivel del bloque de vídeo.

[0141] En algunos ejemplos, el descodificador de vídeo 30 puede recibir información de sintaxis que indica si la predicción de síntesis de visualización está habilitada para algunas o todas las secuencias de imágenes, una imagen individual, un fragmento, un elemento o un frente de onda. La unidad de descodificación por entropía 70 puede descodificar la información de sintaxis y enviar información que indica si la predicción de síntesis de visualización está habilitada para ver la unidad de síntesis 84.

[0142] La unidad de síntesis de visualización 84 puede configurarse para funcionar de una manera sustancialmente similar a la de la unidad de síntesis de visualización 66 (FIG. 2). Por ejemplo, cuando se habilita la predicción de síntesis de visualización, la unidad de síntesis de visualización 84 puede usar información de textura y profundidad de imágenes descodificadas previamente, almacenadas en la memoria de imágenes de referencia 82, para generar una imagen de referencia utilizando síntesis de visualización, es decir, una imagen de referencia de síntesis de visualización. La unidad de síntesis de visualización 84 puede almacenar la imagen sintetizada en la memoria de imágenes de referencia 82. Cuando la predicción de síntesis de visualización no está habilitada, la unidad de síntesis de visualización 84 no necesita generar una imagen, que puede conservar los recursos de procesamiento y/o la energía de la batería.

[0143] De acuerdo con los aspectos de esta divulgación, la unidad de descodificación por entropía 70 puede recibir y descodificar uno o más elementos sintácticos que indican si aplicar VSP para codificar un bloque particular. En algunos ejemplos, los elementos sintácticos solo pueden generarse para bloques que tienen predicción inter, como predicción entre visualizaciones. Por lo tanto, en este ejemplo, cuando un índice de imagen de referencia para el bloque actual corresponde a una imagen de referencia entre visualizaciones, la unidad de síntesis de visualización 84 puede recibir un indicador que indica si se usa VSP para codificar el bloque actual.

- 5 **[0144]** En el ejemplo anterior, un índice de referencia de una imagen VSP no se indica específicamente en una cabecera de fragmento o se agrega durante la construcción de la lista de imágenes de referencia. En su lugar, el descodificador de vídeo 30 descodifica un indicador en el nivel de MB o partición de MB para cada índice de referencia que corresponde a una imagen de referencia entre visualizaciones. Dicha señalización puede proporcionar un enfoque más flexible a la señalización de VSP, ya que la imagen sintetizada no está vinculada a un índice de imagen de referencia particular.
- 10 **[0145]** En otro ejemplo, la unidad de descodificación por entropía 70 puede recibir y descodificar el indicador de VSP para cualquier modo de predicción inter. En otro ejemplo, de forma adicional o alternativa, la unidad de descodificación por entropía 70 puede recibir y descodificar el indicador solo para un índice de referencia correspondiente a RefPicList0. En otro ejemplo más, la unidad de descodificación por entropía 70 puede recibir y descodificar el indicador para un índice de referencia correspondiente a RefPicList0 y RefPicList1 (si está presente).
- 15 **[0146]** De acuerdo con otros aspectos de esta divulgación, cuando se habilita la predicción de síntesis de visualización, el descodificador de vídeo 30 puede recibir y descodificar uno o más elementos sintácticos en un conjunto de parámetros, como una extensión SPS o un SPS que indica si se usa la señalización basada en referencias para VSP (por ejemplo, en el que se genera un índice de imagen de referencia para una imagen de referencia de síntesis de visualización y VSP se realiza basándose en el índice de referencia) o si se usa GVSP para VSP (por ejemplo, en el que VSP se realiza basándose en la señalización de macrobloques descrita anteriormente, por ejemplo, con respecto a las Tablas 1-3 anteriores). En algunos ejemplos, el uno o más elementos sintácticos pueden incluir un indicador en una extensión SPS.
- 20 **[0147]** De acuerdo con otros aspectos de esta divulgación, cuando no se usa la señalización basada en el índice de referencia, el descodificador de vídeo 30 no puede ejecutar ningún comando RPLM que corresponda a una imagen VSP.
- 25 **[0148]** De acuerdo con otros aspectos de esta divulgación, cuando se usa GVSP, la unidad de síntesis de visualización 84 puede sintetizar una imagen de predicción de síntesis de visualización solo desde la primera imagen de referencia entre visualizaciones para RefPicList0. En este ejemplo, no es necesaria la señalización de la imagen de predicción de síntesis de visualización, y se determina (por ejemplo, se deduce) automáticamente que el modo GVSP genera un vector de movimiento de disparidad que apunta a la primera imagen de referencia de la visualización, por ejemplo, RefPicList 0. De forma alternativa, en otro ejemplo, la unidad de síntesis de visualización 84 puede realizar una predicción de síntesis de visualización basada en una imagen entre visualizaciones en RefPicList1. De forma alternativa, en otro ejemplo, la unidad de síntesis de visualización 84 puede realizar una predicción de síntesis de visualización basada en imágenes entre visualizaciones, si están presentes, tanto en RefPicList0 como en RefPicList1.
- 30 **[0149]** Cuando el fragmento de vídeo se codifica como un fragmento intracodificado (I), la unidad de predicción intra 74 puede generar datos de predicción para un bloque de vídeo del fragmento de vídeo actual basándose en un modo de predicción intra señalado y datos de los bloques previamente descodificados de la trama o imagen actual. Cuando la trama de vídeo se codifica como un fragmento intercodificado (por ejemplo, B, P o GPB), la unidad de compensación de movimiento 72 genera bloques predictivos para un bloque de vídeo del fragmento de vídeo actual basándose en los vectores de movimiento y en otros elementos sintácticos recibidos de la unidad de descodificación por entropía 70. Los bloques predictivos pueden generarse a partir de una de las imágenes de referencia dentro de una de las listas de imágenes de referencia. El descodificador 30 de vídeo puede construir las listas de tramas de referencia, lista 0 y lista 1, mediante técnicas de construcción predeterminadas, basándose en imágenes de referencia almacenadas en la memoria 82 de imágenes de referencia.
- 35 **[0150]** La unidad de compensación de movimiento 72 determina la información de predicción para un bloque de vídeo del fragmento de vídeo actual, analizando los vectores de movimiento y otros elementos sintácticos, y usa la información de predicción para generar los bloques predictivos del bloque de vídeo actual que se está descodificando. Por ejemplo, la unidad de compensación de movimiento 72 usa algunos de los elementos sintácticos recibidos para determinar un modo de predicción (por ejemplo, predicción intra, predicción inter temporal o predicción entre visualizaciones) usada para codificar los bloques de vídeo del fragmento de vídeo, un tipo de fragmento de predicción inter (por ejemplo, fragmento B, fragmento P o fragmento GPB), información de construcción para una o más de las listas de imágenes de referencia del fragmento, vectores de movimiento (por ejemplo, vectores de movimiento de disparidad y/o vectores de movimiento temporales) para cada bloque de vídeo intercodificado del fragmento, el estado de predicción inter para cada bloque de vídeo intercodificado del fragmento y otra información para descodificar los bloques de vídeo en el fragmento de vídeo actual.
- 40 **[0151]** La unidad de compensación de movimiento 72 también puede realizar la interpolación basándose en filtros de interpolación. La unidad de compensación de movimiento 72 puede usar filtros de interpolación como los usados por el codificador de vídeo 20 durante la codificación de los bloques de vídeo, para calcular valores interpolados para píxeles fraccionarios de bloques de referencia. En este caso, la unidad de compensación de movimiento 72
- 45 **[0152]** La unidad de compensación de movimiento 72 también puede recibir y descodificar el indicador de VSP para cualquier modo de predicción inter. En otro ejemplo, de forma adicional o alternativa, la unidad de descodificación por entropía 70 puede recibir y descodificar el indicador solo para un índice de referencia correspondiente a RefPicList0. En otro ejemplo más, la unidad de descodificación por entropía 70 puede recibir y descodificar el indicador para un índice de referencia correspondiente a RefPicList0 y RefPicList1 (si está presente).
- 50 **[0153]** De acuerdo con otros aspectos de esta divulgación, cuando se habilita la predicción de síntesis de visualización, el descodificador de vídeo 30 puede recibir y descodificar uno o más elementos sintácticos en un conjunto de parámetros, como una extensión SPS o un SPS que indica si se usa la señalización basada en referencias para VSP (por ejemplo, en el que se genera un índice de imagen de referencia para una imagen de referencia de síntesis de visualización y VSP se realiza basándose en el índice de referencia) o si se usa GVSP para VSP (por ejemplo, en el que VSP se realiza basándose en la señalización de macrobloques descrita anteriormente, por ejemplo, con respecto a las Tablas 1-3 anteriores). En algunos ejemplos, el uno o más elementos sintácticos pueden incluir un indicador en una extensión SPS.
- 55 **[0154]** De acuerdo con otros aspectos de esta divulgación, cuando no se usa la señalización basada en el índice de referencia, el descodificador de vídeo 30 no puede ejecutar ningún comando RPLM que corresponda a una imagen VSP.
- 60 **[0155]** De acuerdo con otros aspectos de esta divulgación, cuando se usa GVSP, la unidad de síntesis de visualización 84 puede sintetizar una imagen de predicción de síntesis de visualización solo desde la primera imagen de referencia entre visualizaciones para RefPicList0. En este ejemplo, no es necesaria la señalización de la imagen de predicción de síntesis de visualización, y se determina (por ejemplo, se deduce) automáticamente que el modo GVSP genera un vector de movimiento de disparidad que apunta a la primera imagen de referencia de la visualización, por ejemplo, RefPicList 0. De forma alternativa, en otro ejemplo, la unidad de síntesis de visualización 84 puede realizar una predicción de síntesis de visualización basada en una imagen entre visualizaciones en RefPicList1. De forma alternativa, en otro ejemplo, la unidad de síntesis de visualización 84 puede realizar una predicción de síntesis de visualización basada en imágenes entre visualizaciones, si están presentes, tanto en RefPicList0 como en RefPicList1.
- 65 **[0156]** Cuando el fragmento de vídeo se codifica como un fragmento intracodificado (I), la unidad de predicción intra 74 puede generar datos de predicción para un bloque de vídeo del fragmento de vídeo actual basándose en un modo de predicción intra señalado y datos de los bloques previamente descodificados de la trama o imagen actual. Cuando la trama de vídeo se codifica como un fragmento intercodificado (por ejemplo, B, P o GPB), la unidad de compensación de movimiento 72 genera bloques predictivos para un bloque de vídeo del fragmento de vídeo actual basándose en los vectores de movimiento y en otros elementos sintácticos recibidos de la unidad de descodificación por entropía 70. Los bloques predictivos pueden generarse a partir de una de las imágenes de referencia dentro de una de las listas de imágenes de referencia. El descodificador 30 de vídeo puede construir las listas de tramas de referencia, lista 0 y lista 1, mediante técnicas de construcción predeterminadas, basándose en imágenes de referencia almacenadas en la memoria 82 de imágenes de referencia.
- [0157]** La unidad de compensación de movimiento 72 determina la información de predicción para un bloque de vídeo del fragmento de vídeo actual, analizando los vectores de movimiento y otros elementos sintácticos, y usa la información de predicción para generar los bloques predictivos del bloque de vídeo actual que se está descodificando. Por ejemplo, la unidad de compensación de movimiento 72 usa algunos de los elementos sintácticos recibidos para determinar un modo de predicción (por ejemplo, predicción intra, predicción inter temporal o predicción entre visualizaciones) usada para codificar los bloques de vídeo del fragmento de vídeo, un tipo de fragmento de predicción inter (por ejemplo, fragmento B, fragmento P o fragmento GPB), información de construcción para una o más de las listas de imágenes de referencia del fragmento, vectores de movimiento (por ejemplo, vectores de movimiento de disparidad y/o vectores de movimiento temporales) para cada bloque de vídeo intercodificado del fragmento, el estado de predicción inter para cada bloque de vídeo intercodificado del fragmento y otra información para descodificar los bloques de vídeo en el fragmento de vídeo actual.
- [0158]** La unidad de compensación de movimiento 72 también puede realizar la interpolación basándose en filtros de interpolación. La unidad de compensación de movimiento 72 puede usar filtros de interpolación como los usados por el codificador de vídeo 20 durante la codificación de los bloques de vídeo, para calcular valores interpolados para píxeles fraccionarios de bloques de referencia. En este caso, la unidad de compensación de movimiento 72

puede determinar los filtros de interpolación utilizados por el codificador de vídeo 20 a partir de los elementos sintácticos recibidos y utilizar los filtros de interpolación para generar bloques predictivos.

[0152] En algunos ejemplos, la unidad de compensación de movimiento 72 puede realizar una predicción entre visualizaciones utilizando vectores de movimiento de disparidad. Dicha predicción entre visualizaciones puede ser relativa a una imagen previamente descodificada de otra visualización o relativa a una imagen de referencia generada usando la síntesis de visualización, suponiendo que la predicción de síntesis de visualización está habilitada. Por consiguiente, la unidad de compensación de movimiento 72 puede denominarse unidad de compensación de movimiento/disparidad 72.

[0153] La unidad de cuantificación inversa 76 cuantifica de manera inversa, es decir, descuantifica, los coeficientes de transformada cuantificados, proporcionados en el flujo de bits y descodificados por la unidad de descodificación por entropía 70. El proceso de cuantificación inversa puede incluir el uso de un parámetro de cuantificación QPy, calculado por el descodificador de vídeo 30 para cada bloque de vídeo en el fragmento de vídeo, para determinar un grado de cuantificación y, asimismo, un grado de cuantificación inversa que debería aplicarse.

[0154] La unidad 78 de transformada inversa aplica una transformada inversa, por ejemplo, una DCT inversa, una transformada inversa de enteros, o un proceso de transformada inversa conceptualmente similar, a los coeficientes de transformada, con el fin de generar bloques residuales en el dominio del píxel.

[0155] Después de que la unidad de compensación de movimiento 72 o la unidad de predicción intra 74 genera el bloque predictivo para el bloque de vídeo actual basándose en los vectores de movimiento y otros elementos sintácticos, el descodificador de vídeo 30 forma un bloque de vídeo descodificado sumando los bloques residuales de la unidad de transformada inversa 78 y el bloque predictivo correspondiente. El sumador 80 representa el componente o los componentes que realizan esta operación de suma.

[0156] Si se desea, también puede aplicarse un filtro de eliminación de bloques para filtrar los bloques descodificados con el fin de eliminar distorsiones de efecto pixelado. También pueden utilizarse otros filtros de bucle (ya sea en el bucle de codificación o después del bucle de codificación) para allanar las transiciones de píxeles o mejorar de otro modo la calidad del vídeo. Los bloques de vídeo descodificados de una trama o imagen determinada se almacenan a continuación en la memoria 82 de imágenes de referencia, que almacena imágenes de referencia usadas para una subsiguiente compensación de movimiento. La memoria 82 de imágenes de referencia almacena también vídeo descodificado para su presentación posterior en un dispositivo de visualización, tal como el dispositivo 32 de visualización de la FIG. 1.

[0157] De esta manera, el descodificador de vídeo 30 de la FIG. 3 representa un ejemplo de un descodificador de vídeo configurado para determinar si un índice de referencia para el bloque actual corresponde a una imagen de referencia entre visualizaciones, y cuando el índice de referencia para el bloque actual corresponde a la imagen de referencia entre visualizaciones, obtener, a partir de un flujo de bits codificado, datos que indican un modo de predicción de síntesis de visualización (VSP) del bloque actual, donde el modo de VSP para el índice de referencia indica si el bloque actual se predice con la predicción de síntesis de visualización a partir de la imagen de referencia entre visualizaciones.

[0158] La FIG. 4 ilustra un orden de descodificación (es decir, orden de flujo de bits) de MVC típico. La disposición de orden de descodificación se denomina codificación de tiempo primero.. Cada unidad de acceso se define para contener las imágenes codificadas de todas las visualizaciones para un instante de tiempo de salida. El orden de descodificación de las unidades de acceso puede no ser idéntico al orden de salida o de visualización.

[0159] La FIG. 5 es un diagrama conceptual que ilustra un patrón de predicción de MVC de ejemplo. La codificación de vídeo multivisualización (MVC) es una extensión de la norma ITU-T H.264/AVC. Una técnica similar puede aplicarse a la HEVC. En el ejemplo de la FIG. 4, se ilustran ocho visualizaciones (que tienen ID de visualización "S0" a "S7"), y se ilustran doce ubicaciones temporales ("T0" a "T11") para cada visualización. Es decir, cada fila en la FIG. 4 corresponde a una visualización, mientras que cada columna indica una ubicación temporal.

[0160] Aunque la MVC tiene una denominada visualización de base que es descodificable mediante los descodificadores de la H.264/AVC y el par de visualizaciones estéreo podría tener soporte también de la MVC, una ventaja de la MVC es que podría dar soporte a un ejemplo que usa más de dos visualizaciones como una entrada de vídeo tridimensional y descodifica este vídeo tridimensional representado por las múltiples visualizaciones. Un renderizador de un cliente que tiene un descodificador de MVC puede esperar contenido de vídeo tridimensional con múltiples visualizaciones.

[0161] Las imágenes de la FIG. 5 se indican usando un bloque sombreado que incluye una letra, que designa si la imagen correspondiente está intracodificada (es decir, una trama I), o intercodificada en una dirección (es decir, como una trama P) o en múltiples direcciones (es decir, como una trama B). En general, las predicciones se indican

mediante flechas, donde la imagen a la que se apunta utiliza el objeto desde el que se apunta como referencia de predicción. Por ejemplo, la trama P de la visualización S2 en la ubicación temporal T0 se predice a partir de la trama I de la visualización S0 en la ubicación temporal T0.

5 **[0162]** Al igual que con la codificación de vídeo de visualización única, las imágenes de una secuencia de vídeo de la norma 3D-AVC pueden codificarse predictivamente con respecto a las imágenes en diferentes ubicaciones temporales. Por ejemplo, la trama B de la visualización S0 en la ubicación temporal T1 tiene una flecha apuntando a la misma desde la trama I de la visualización S0 en la ubicación temporal T0, lo cual indica que la trama B se predice a partir de la trama I. Adicionalmente, sin embargo, en el contexto de la codificación de vídeo de múltiples visualizaciones, las imágenes se pueden predecir entre visualizaciones. Es decir, un componente de visualización puede utilizar los componentes de visualización en otras visualizaciones como referencia. En la MVC, por ejemplo, la predicción entre visualizaciones se realiza como si el componente de visualización en otra visualización es una referencia de predicción inter. Las posibles referencias entre visualizaciones pueden señalarse en la extensión de MVC del conjunto de parámetros de secuencia (SPS) y pueden ser modificadas por el proceso de construcción de la lista de imágenes de referencia, lo cual habilita el ordenamiento flexible de las referencias de predicción inter o de predicción entre visualizaciones.

20 **[0163]** La FIG. 5 proporciona varios ejemplos de predicción entre visualizaciones. Las imágenes de visualización S1, en el ejemplo de la FIG. 5, se ilustran como predichas a partir de imágenes en diferentes ubicaciones temporales de la visualización S1, así como también como entre visualizaciones predichas a partir de imágenes de imágenes de las visualizaciones S0 y S2 en las mismas ubicaciones temporales. Por ejemplo, la trama-B de la visualización S1 en la ubicación temporal T1 se predice a partir de cada una de las tramas-B de la visualización S1 en las ubicaciones temporales T0 y T2, así como las tramas-B de las visualizaciones S0 y S2 en la ubicación temporal T1.

25 **[0164]** En el ejemplo de la FIG. 5, la mayúscula "B" y la minúscula "b" están destinadas a indicar diferentes relaciones jerárquicas entre imágenes, en lugar de diferentes metodologías de codificación. En general, las tramas con "B" mayúscula están relativamente más altas en la jerarquía de predicción que las tramas con "b" minúscula. La FIG. 5 también ilustra las variaciones en la jerarquía de predicción utilizando diferentes niveles de sombreado, donde las imágenes con una mayor cantidad de sombreado (es decir, relativamente más oscuras) están más altas en la jerarquía de predicción que aquellas imágenes que tienen menos sombreado (es decir, relativamente más claras). Por ejemplo, todas las tramas I en la FIG. 5 se ilustran con sombreado completo, mientras que las tramas P tienen un sombreado algo más claro, y las tramas B (y las tramas b minúscula) tienen varios niveles de sombreado entre sí, pero siempre más claro que el sombreado de las tramas P y las tramas I.

35 **[0165]** En general, la jerarquía de predicción se relaciona con índices del orden de visualizaciones, en cuanto a que las imágenes relativamente más altas en la jerarquía de predicción deberían ser descodificadas antes de la descodificación de imágenes que están relativamente más bajas en la jerarquía, de tal modo que esas imágenes relativamente más altas en la jerarquía se puedan utilizar como imágenes de referencia durante la descodificación de las imágenes relativamente más bajas en la jerarquía. Un índice de orden de visualización es un índice que indica el orden de descodificación de componentes de visualización en una unidad de acceso. Los índices de orden de visualizaciones pueden estar implícitos en un conjunto de parámetros, tal como un SPS.

40 **[0166]** De esta manera, las imágenes utilizadas como imágenes de referencia pueden ser descodificadas antes de la descodificación de las imágenes que se codifican con referencia a las imágenes de referencia. Un índice de orden de visualización es un índice que indica el orden de descodificación de componentes de visualización en una unidad de acceso. Para cada índice i de orden de visualización, se señala el ID de visualización correspondiente. La descodificación de los componentes de visualización sigue el orden ascendente de los índices de orden de visualización. Si se presentan todas las visualizaciones, entonces el conjunto de índices de orden de visualización comprende un conjunto consecutivamente ordenado de cero a uno menos que el número total de visualizaciones.

45 **[0167]** En la MVC, un subconjunto de todo un flujo de bits puede ser extraído para formar un sub-flujo de bits que aún se ajusta a la MVC. Hay muchos posibles sub-flujos de bits que las aplicaciones específicas puedan requerir, basándose, por ejemplo, en un servicio proporcionado por un servidor, la capacidad, el apoyo y las capacidades de los descodificadores de uno o más clientes y/o la preferencia de uno o más clientes. Por ejemplo, un cliente podría requerir solo tres visualizaciones, y podría haber dos escenarios. En un ejemplo, un cliente puede requerir una experiencia de exhibición sin fisuras y podría preferir visualizaciones con los valores de view_id S0, S1 y S2, mientras que otro cliente puede requerir ajustabilidad a escala de las visualizaciones y preferir visualizaciones con los valores de view_id S0, S2 y S4. Obsérvese que ambos sub-flujos de bits se pueden descodificar como flujos de bits independientes de MVC y pueden disponer de soporte de forma simultánea.

60 **[0168]** Por consiguiente, en la FIG. 5 se muestra una estructura típica de predicción de MVC (que incluye tanto la predicción entre imágenes dentro de cada visualización como la predicción entre visualizaciones) para la codificación de vídeo de múltiples visualizaciones, donde las predicciones se indican mediante flechas, y el objeto apuntado utiliza el objeto apuntador como referencia de predicción. En MVC, la predicción entre visualizaciones

se soporta mediante la compensación de movimiento de disparidad, que utiliza la sintaxis de la compensación de movimiento H.264/AVC, pero permite utilizar una imagen de una visualización diferente como imagen de referencia.

5 **[0169]** La codificación de dos visualizaciones también podría ser soportada por MVC, y una de las posibles ventajas de MVC es que un codificador MVC podría tomar más de dos visualizaciones como una entrada de vídeo 3D y un decodificador MVC puede decodificar una representación de múltiples visualizaciones de este tipo. En consecuencia, cualquier renderizador con un decodificador MVC puede esperar contenidos de vídeo 3D con más de dos visualizaciones.

10 **[0170]** Con respecto a la predicción entre visualizaciones, en MVC, se permite la predicción entre visualizaciones entre imágenes en la misma unidad de acceso (es decir, con la misma instancia de tiempo). Cuando se codifica una imagen en una de las visualizaciones no básicas, se puede añadir una imagen a una lista de imágenes de referencia si está en una visualización diferente pero con una misma instancia temporal. Una imagen de referencia de predicción entre visualizaciones puede disponerse en cualquier posición de una lista de imágenes de referencia, tal como cualquier imagen de referencia de predicción inter.

15 **[0171]** Las visualizaciones S0 a S7 de la FIG. 4 representan ejemplos de visualizaciones codificadas, es decir, visualizaciones para las que se proporciona información codificada en un flujo de bits. En algunos ejemplos, se pueden sintetizar visualizaciones adicionales entre las visualizaciones S0 a S7. Por ejemplo, una visualización puede sintetizarse entre las visualizaciones S0 y S1. Para sintetizar una imagen de dicha visualización, se puede utilizar la información de textura y/o profundidad de las imágenes en las visualizaciones S0 y S2. Por ejemplo, para sintetizar una imagen de referencia en el momento T1, se puede usar la información de textura y profundidad de las imágenes en el momento T1 a partir de las visualizaciones S0 y/o S2. Un codificador de vídeo, tal como el codificador de vídeo 20 o el decodificador de vídeo 30, puede interpolar datos de píxeles para que una imagen de este tipo de una visualización sintetizada se use como referencia para codificar otra imagen, por ejemplo, una imagen de la visualización S1 en el momento T1.

20 **[0172]** Los aspectos de esta divulgación incluyen técnicas para indicar si aplicar la predicción de síntesis de visualización al codificar un bloque de datos de vídeo. En un ejemplo, un codificador de vídeo, como el codificador de vídeo 20 y/o el decodificador de vídeo 30, puede determinar si un índice de referencia para un bloque actual corresponde a una imagen de referencia de predicción inter, y cuando el índice de referencia para el bloque actual corresponde a la imagen de referencia de predicción inter, determinar los datos que indican un modo VSP del bloque actual, donde el modo VSP para el índice de referencia indica si el bloque actual se predice utilizando la predicción de síntesis de visualización, por ejemplo, a partir de una imagen de referencia entre visualizaciones. En los casos en que el codificador de vídeo está configurado como decodificador de vídeo 30, el decodificador de vídeo 30 puede determinar los datos que indican el modo VSP obteniendo los datos de un flujo de bits codificado.

30 **[0173]** En algunos ejemplos, la imagen de referencia de predicción inter está en general asociada con un modo de predicción inter. En tales ejemplos, los datos que indican el modo VSP pueden determinarse para cualquier modo intermedio, por ejemplo, para imágenes de referencia intra-visualización e imágenes de referencia entre visualizaciones. En otros ejemplos, los datos que indican el modo VSP solo pueden determinarse para imágenes de referencia entre visualizaciones. Es decir, por ejemplo, el decodificador de vídeo 30 solo puede decodificar los datos que indican el modo VSP cuando un índice de imagen de referencia corresponde a una imagen de referencia entre visualizaciones.

40 **[0174]** En el ejemplo anterior, un índice de referencia de una imagen VSP no se indica específicamente en una cabecera de fragmento o se agrega durante la construcción de la lista de imágenes de referencia. En su lugar, el codificador de vídeo puede codificar uno o más elementos sintácticos en el nivel de MB o partición de MB para cada índice de referencia que corresponda a una imagen de referencia de predicción inter. En algunos ejemplos, uno o más elementos sintácticos pueden ser un indicador de VSP. Por ejemplo, el indicador puede indicar si VSP se aplica al MB o la partición de MB para un índice de referencia particular.

45 **[0175]** Con respecto a la semántica de la capa de macrobloque, un `vsp_flag` igual a 1 puede especificar que la partición de MB o el MB actual se predice unidireccionalmente a partir una imagen de referencia entre visualizaciones, identificada por `non_anchor_ref_10[VOIdx][0]` cuando `anchor_pic_flag` es 0 o `anchor_ref_10[VOIdx][0]` cuando `anchor_pic_flag` es 1, donde `VOIdx` es el índice de orden de visualización del componente de visualización actual. Los vectores de movimiento del MB o la partición de MB pueden obtenerse como se especifica en la subcláusula J.8.3.1.3 (por ejemplo, de WD 4). Además, un `vsp_flag` igual a 0 puede especificar que la predicción de síntesis de visualización no se usa para la partición de MB o el MB actual. De forma alternativa, un `vsp_flag` igual a 0 puede especificar que la predicción de síntesis de visualización bajo GVSP está deshabilitada, pero la predicción de síntesis de visualización usando el mecanismo basado en señalización de índice de referencia todavía se puede permitir.

60 **[0176]** En otro ejemplo, cuando `vsp_flag` es igual a 1, la partición de MB o el MB se puede predecir a partir de la primera referencia entre visualizaciones en `RefPicList0`. En otro ejemplo, cuando `vsp_flag` es igual a 1, la predicción de MB o el MB se puede predecir a partir de la primera entrada en un RPLM que tiene

modification_of_pic_nums_idc igual a 6, cuando no se usa vsp_pic_flag para restringir el presente de modification_of_pic_nums_id. En este caso, el comando con modified_of_pic_nums_idc igual a 6 en el RPLM no indica una nueva entrada en una lista de imágenes de referencia.

5 **[0177]** Con respecto a las restricciones de nivel, cuando una partición de MB o un MB está codificado con un indicador de VSP y tiene diferentes vectores de movimiento en al menos un bloque de 8x8 dentro de la partición de MB o el MB, un codificador de vídeo, como el codificador de vídeo 20 o el descodificador de vídeo 30, puede contar el número de sub-bloques más pequeños que 8x8 basados en la partición de MB o el MB. El número de sub-bloques no debe exceder a MaxSubMbRectSize. En este ejemplo, el codificador de vídeo puede determinar que un MB predicho por VSP tiene uno o más sub-bloques más pequeños que 8x8.

10 **[0178]** De forma alternativa, en otro ejemplo, después de predecir un MB usando VSP (con un vsp_flag), el codificador de vídeo puede determinar que el MB tiene sub-bloques más pequeños que 8x8. De forma alternativa, en otro ejemplo, cuando un indicador de VSP está presente para una partición de MB o un MB, para cada partición de MB, solo se obtiene un vector de movimiento entre visualizaciones, de manera que la partición de MB o el MB nunca tiene una partición de bloque más pequeña que 8x8. De forma alternativa, en otro ejemplo, cuando un indicador de VSP está presente para una partición de MB o un MB, para cada partición de MB, solo se obtiene un vector de movimiento entre visualizaciones para cada uno de los 4 bloques de 8x8, de modo que la partición de MB o el MB nunca tiene una partición de bloques más pequeña que 8x8.

15 **[0179]** De acuerdo con otros aspectos de esta divulgación, cuando se habilita la predicción de síntesis de visualización, un codificador de vídeo puede codificar uno o más elementos sintácticos en un conjunto de parámetros, como una extensión SPS o un SPS que indica si la señalización basada en referencias se usa para VSP (por ejemplo, en el cual se genera un índice de imagen de referencia para una imagen de referencia de síntesis de visualización y VSP se realiza basándose en el índice de referencia) o si se usa GVSP para VSP (por ejemplo, en el que VSP se realiza basándose en la señalización de macrobloques descrita anteriormente, por ejemplo, con respecto de las Tablas 1-3 anteriores). En algunos ejemplos, el uno o más elementos sintácticos pueden incluir un indicador en una extensión SPS.

20 **[0180]** Por ejemplo, de acuerdo con los aspectos de esta divulgación, el codificador de vídeo 20 y/o el descodificador de vídeo 30 pueden usar la sintaxis de extensión del conjunto de parámetros de secuencia en relación con WD4 (anotado arriba), como se muestra en la Tabla 4 a continuación:

TABLA 4 - Sintaxis de extensión de 3DVC del conjunto de parámetros de secuencia

35

seq parameter set_3dvc_extension() {		Descriptor
for(i = 0; i <= num_level_values_signalled_minus_1; i++)		
for(j = 0; j <= num_applicable_ops_minus_1[i];j++)		
applicable_op_num_depth_views_minus1[i][j]		ue(v)
depth_info_present_flag		u(1)
if(depth_info_present_flag) {		
3dv_acquisition_idc		ue(v)
for(i = 0; i < num_views_minus_1; i++)		
view_id_3dv[i]		ue(v)
if(3dv_acquisition_idc) {		
depth_ranges(2, 0)		
if(profile_idc == 139 && num_views_minus1)		
vsp_param(2,0)		
}		
if (profile_idc == 139)		
for(i = 0; i <= num_views_minus_1; i++)		

seq parameter set_3dvc_extension() {		Descriptor
depth_preceding_texture_flag[i]		u(1)
}		
if(profile_idc == 139 && depth_info_present_flag) {		
reduced_resolution_flag		u(1)
slice_header_prediction_flag		u(1)
inside_view_mvpc_flag		u(1)
seq_view_synthesis_flag		u(1)
if(seq_view_synthesis_flag)		
vsp_pic_flag		u(1)
disp_flag		u(1)
psip_flag		u(1)
nonlinear_depth_representation_num		ue(v)
for i = 1; i <= nonlinear_depth_representation_num; i++)		
nonlinear_depth_representation_model[i]		ue(v)
}		
if(profile_idc == 139 && !depth_info_present_flag)		
alc_sps_enable_flag		u(1)
}		

5 **[0181]** En el ejemplo de la Tabla 4 anterior, vsp_pic_flag igual a 1 indica que la imagen de predicción de síntesis de visualización se puede generar y hacer referencia a la misma mediante un índice de referencia en una lista de imágenes de referencia. Además, vsp_pic_flag igual a 0 indica que la imagen de predicción de síntesis de visualización no se genera y la predicción de síntesis de visualización siempre se refiere a una imagen de referencia entre visualizaciones. Cuando no está presente, se puede deducir que vsp_pic_flag es igual a 0.

10 **[0182]** De acuerdo con otros aspectos de esta divulgación, cuando no se utiliza la señalización basada en el índice de referencia, un RPLM no debe contener ningún comando RPLM que corresponda a una imagen VSP. Por ejemplo, cuando no se utiliza la señalización basada en el índice de referencia, un codificador de vídeo (como el codificador de vídeo 20 y/o el decodificador de vídeo 30) puede no emitir un comando RPLM que corresponda a una imagen VSP. Es decir, el codificador de vídeo puede abstenerse de usar un RPLM para cambiar la posición de una imagen VSP, porque no se ha generado ninguna imagen VSP.

15 **[0183]** En un ejemplo, con respecto a las modificaciones de 3DVC de la lista de imágenes de referencia, la semántica especificada en la subcláusula 1.7.4.3.1.1 (por ejemplo, en WD 4, señalada anteriormente) se aplica con las siguientes adiciones:

20 La siguiente entrada de tabla adicional se inserta en la Tabla H-1:

modification_of_pic_nums_idc	Modificación especificada
6	vsp_ref_idx está presente y corresponde a un índice de referencia de VSP

donde modification_of_pic_nums_idc no es igual a 6, si seq_view_synthesis_flag o vsp_pic_flag es igual a 0.

25 **[0184]** De acuerdo con otros aspectos de esta divulgación, cuando se usa GVSP, el codificador de vídeo puede sintetizar una imagen de predicción de síntesis de visualización solo desde la primera imagen de referencia entre visualizaciones para RefPicList0. En este ejemplo, no es necesaria la señalización de la imagen de predicción de síntesis de visualización, y se determina (por ejemplo, se deduce) automáticamente que el modo GVSP genera un

vector de movimiento de disparidad que apunta a la primera imagen de referencia de la visualización, por ejemplo, RefPicList 0. De forma alternativa, en otro ejemplo, el codificador de vídeo puede realizar la predicción de síntesis de visualización basándose en una imagen entre visualizaciones en RefPicList1. De forma alternativa, en otro ejemplo, el codificador de vídeo puede realizar la predicción de síntesis de visualización basándose en imágenes entre visualizaciones si están presentes, tanto en RefPicList0 como en RefPicList1.

[0185] La FIG. 6 es un diagrama de flujo que ilustra un procedimiento de ejemplo para codificar información relacionada con la predicción de síntesis de visualización. El procedimiento de la FIG. 6 se explica con respecto al codificador de vídeo 20 (FIGs. 1 y 2). Sin embargo, debe entenderse que otros dispositivos de codificación de vídeo pueden configurarse para realizar un procedimiento similar. Además, ciertos pasos en el procedimiento pueden realizarse en un orden diferente o en paralelo. Del mismo modo, ciertos pasos pueden omitirse, y otros pasos pueden agregarse, en varios ejemplos.

[0186] En este ejemplo, el codificador de vídeo 20 predice inicialmente el bloque actual y el índice de referencia del bloque predictivo (120). En este ejemplo, se supone que el codificador de vídeo 20 realiza predicción inter del bloque actual. Por ejemplo, la unidad de estimación de movimiento 42 puede calcular un vector de movimiento para el bloque actual realizando una búsqueda de movimiento de imágenes codificadas previamente, por ejemplo, imágenes entre visualizaciones e imágenes temporales. Por lo tanto, la unidad de estimación de movimiento 42 puede producir un vector de movimiento temporal o un vector de movimiento de disparidad para predecir el bloque actual.

[0187] En algunos casos, como se señaló anteriormente, el codificador de vídeo 20 puede predecir el bloque actual a partir de un bloque sintetizado. Por ejemplo, el codificador de vídeo 20 puede realizar un proceso de síntesis de visualización para generar una imagen sintetizada, que se puede agregar a una lista de imágenes de referencia y tener un índice de imagen de referencia. El codificador de vídeo 20 puede realizar VSP para predecir el bloque actual en relación con un bloque de la imagen sintetizada.

[0188] En el ejemplo de la FIG. 6, el codificador de vídeo 20 puede determinar si el índice de referencia de la imagen de referencia que incluye el bloque predictivo es una imagen de referencia entre visualizaciones (122). En algunos ejemplos, el codificador de vídeo 20 puede determinar si la imagen de referencia se incluye en una visualización diferente a la del bloque que se codifica basándose en un identificador de visualización (view_id).

[0189] De acuerdo con los aspectos de esta divulgación, si el índice de referencia corresponde a una imagen de referencia entre visualizaciones (la rama Sí del paso 122), el codificador de vídeo 20 puede codificar datos que indiquen si aplicar VSP para el índice de referencia. Por ejemplo, en los casos en que el codificador de vídeo 20 realiza VSP para determinar el bloque predictivo, el codificador de vídeo 20 puede codificar datos que indiquen que se ha usado VSP (por ejemplo, establecer un vsp_flag igual a 1). En los casos en que el codificador de vídeo 20 simplemente realizó una predicción de visualización para determinar el bloque predictivo (sin VSP), el codificador de vídeo 20 puede codificar datos que indiquen que no se ha usado VSP (por ejemplo, establecer un vsp_flag igual a 0).

[0190] A continuación, el codificador de vídeo 20 puede calcular un bloque residual para el bloque actual (126). Para calcular el bloque residual, el codificador de vídeo 20 puede calcular una diferencia entre el bloque original, no codificado y el bloque predictivo, que puede ser un bloque en la misma visualización que el bloque codificado, un bloque en una visualización diferente que el bloque codificado, o un bloque sintetizado. Posteriormente, el codificador de vídeo 20 puede transformar y cuantificar los coeficientes del bloque residual (128). A continuación, el codificador de vídeo 20 puede escanear los coeficientes de transformada cuantificados del bloque residual (130). Durante o tras el escaneado, el codificador de vídeo 20 puede codificar por entropía los coeficientes (132). Por ejemplo, el codificador de vídeo 20 puede codificar los coeficientes usando CAVLC o CABAC. A continuación, el codificador de vídeo 20 puede emitir los datos codificados del bloque así como la indicación de VSP (134).

[0191] De esta manera, el procedimiento de la FIG. 6 representa un ejemplo de un procedimiento para codificar datos de vídeo, con el procedimiento que incluye determinar si un índice de referencia para un bloque actual corresponde a una imagen de referencia entre visualizaciones, cuando el índice de referencia para el bloque actual corresponde a la imagen de referencia entre visualizaciones, generar datos que indiquen un modo de predicción de síntesis de visualización (VSP) del bloque actual, donde el modo de VSP para el índice de referencia indica si el bloque actual se predice con la predicción de síntesis de visualización a partir de la imagen de referencia entre visualizaciones y codificar los datos que indican el modo VSP en un flujo de bits.

[0192] La FIG. 7 es un diagrama de flujo que ilustra un procedimiento de ejemplo para codificar información relacionada con la predicción de síntesis de visualización. El procedimiento de la FIG. 7 se explica con respecto al decodificador de vídeo 30 (FIGs. 1 y 3). Sin embargo, debe entenderse que otros dispositivos de codificación de vídeo pueden configurarse para realizar un procedimiento similar. Además, ciertos pasos en el procedimiento pueden realizarse en un orden diferente o en paralelo. Del mismo modo, ciertos pasos pueden omitirse, y otros pasos pueden agregarse, en varios ejemplos.

[0193] La unidad de descodificación por entropía 70 descodifica por entropía los datos para los coeficientes del bloque que se está descodificando actualmente, al menos un vector de movimiento o vector de disparidad, y al menos un índice de referencia correspondiente (160). El ejemplo de la FIG. 7 supone que el bloque actual está uni-predicho, pero debe entenderse que en otros ejemplos, el bloque actual puede ser bi-predicho, como se describe en el presente documento.

[0194] El vector de movimiento o vector de disparidad puede identificar el bloque predictivo en la imagen de referencia asociada con el índice de imagen de referencia descodificada. Por ejemplo, se puede realizar una predicción inter del bloque actual a partir un bloque en la misma visualización que el bloque actual, una predicción entre visualizaciones a partir de un bloque en una visualización diferente a la del bloque actual, o una predicción entre visualizaciones a partir un bloque sintetizado utilizando VSP. Independientemente del proceso de predicción en particular, el índice de referencia descodificado identifica una imagen almacenada en una memoria de imágenes de referencia.

[0195] En el ejemplo de la FIG. 7, el descodificador de vídeo 30 puede determinar si el índice de referencia de la imagen de referencia que incluye el bloque predictivo es una imagen de referencia entre visualizaciones (162). En algunos ejemplos, el descodificador de vídeo 30 puede determinar si la imagen de referencia se incluye en una visualización diferente a la del bloque que se codifica basándose en un identificador de visualización (view_id).

[0196] De acuerdo con los aspectos de esta divulgación, si el índice de referencia corresponde a una imagen de referencia entre visualizaciones (la rama SÍ del paso 162), el descodificador de vídeo 30 puede descodificar datos que indiquen si aplicar VSP para el índice de referencia. Por ejemplo, en los casos en que se va a usar VSP para determinar el bloque predictivo, el descodificador de vídeo 30 puede descodificar datos que indican que se va a usar VSP (por ejemplo, el descodificador de vídeo 30 puede obtener, a partir de un flujo de bits codificado, un vsp_flag igual a 1). En los casos en que el descodificador de vídeo 30 debe realizar una predicción entre visualizaciones sin VSP para determinar el bloque predictivo, el descodificador de vídeo 30 puede descodificar datos que indican que no se debe usar VSP (por ejemplo, el descodificador de vídeo 30 puede obtener, a partir de un flujo de bits codificado, un vsp_flag igual a 0). En algunos ejemplos, si el índice de referencia corresponde a una imagen de referencia entre visualizaciones pero no hay una indicación de VSP incluida en el flujo de bits (por ejemplo, no está presente vsp_flag), el descodificador de vídeo 30 puede determinar (inferir) automáticamente que se realiza una predicción entre visualizaciones del bloque actual sin VSP.

[0197] El descodificador de vídeo 30 puede predecir el bloque actual utilizando el vector de disparidad o movimiento descodificado y el índice de imagen de referencia (166). En los casos en que se usa VSP, el descodificador de vídeo 30 puede generar una imagen sintetizada para determinar el bloque predictivo. Posteriormente, el descodificador de vídeo 30 puede escanear de forma inversa los coeficientes reproducidos (168), para crear un bloque de coeficientes de transformada cuantificados. Posteriormente, el descodificador de vídeo 30 puede cuantificar de manera inversa y transformar de manera inversa los coeficientes para producir un bloque residual (170). El descodificador de vídeo 30 puede finalmente descodificar el bloque actual combinando el bloque predicho y el bloque residual (172).

[0198] De esta manera, el procedimiento de la FIG. 7 representa un ejemplo de un procedimiento que incluye determinar si un índice de referencia para un bloque actual corresponde a una imagen de referencia entre visualizaciones, y cuando el índice de referencia para el bloque actual corresponde a la imagen de referencia entre visualizaciones, obtener, a partir de un flujo de bits codificado, datos que indican un modo de predicción de síntesis de visualización (VSP) del bloque actual, donde el modo de VSP para el índice de referencia indica si el bloque actual se predice con la predicción de síntesis de visualización a partir de la imagen de referencia entre visualizaciones.

[0199] Las técnicas descritas anteriormente con respecto a la síntesis de visualización pueden realizarse mediante el codificador de vídeo 20 (FIGs. 1 y 2) y/o el descodificador de vídeo 30 (FIGs. 1 y 3), los cuales pueden denominarse en general codificadores de vídeo. Del mismo modo, "codificación" y "codificación de vídeo" pueden referirse a codificación de vídeo (por ejemplo, mediante un codificador de vídeo) o descodificación de vídeo (por ejemplo, mediante un descodificador de vídeo), según corresponda.

[0200] Debería entenderse que, dependiendo del ejemplo, ciertos actos o sucesos de cualquiera de los procedimientos descritos en el presente documento pueden realizarse en una secuencia distinta, pueden añadirse, fundirse u omitirse por completo (por ejemplo, no todos los actos o sucesos descritos son necesarios para la puesta en práctica del procedimiento). Además, en ciertos ejemplos, los actos o sucesos pueden realizarse simultáneamente, por ejemplo, mediante el procesamiento de múltiples hilos, el procesamiento de interrupciones o múltiples procesadores, en lugar de secuencialmente. Además, aunque ciertos aspectos de esta divulgación se describen como realizados por un único módulo o unidad, con fines de claridad, se debería entender que las técnicas de esta divulgación pueden ser realizadas por una combinación de unidades o módulos asociados a un codificador de vídeo.

5 **[0201]** Si bien se describen anteriormente combinaciones particulares de diversos aspectos de las técnicas, estas combinaciones se proporcionan meramente para ilustrar ejemplos de las técnicas descritas en esta divulgación. En consecuencia, las técnicas de esta divulgación no deberían limitarse a estas combinaciones a modo de ejemplo y pueden abarcar cualquier combinación concebible de los diversos aspectos de las técnicas descritas en esta divulgación.

10 **[0202]** En uno o más ejemplos, las funciones descritas pueden implementarse en hardware, software, firmware o cualquier combinación de los mismos. Si se implementan en software, las funciones pueden almacenarse en, y transmitirse por, un medio legible por ordenador como una o más instrucciones o código, y ejecutarse mediante una unidad de procesamiento basada en hardware. Los medios legibles por ordenador pueden incluir medios de almacenamiento legibles por ordenador, que correspondan a un medio tangible tal como medios de almacenamiento de datos o medios de comunicación que incluyan cualquier medio que facilite la transferencia de un programa informático desde un lugar a otro, por ejemplo, de acuerdo con un protocolo de comunicación.

15 **[0203]** De esta manera, los medios legibles por ordenador pueden corresponder en general a (1) medios de almacenamiento tangibles legibles por ordenador que sean no transitorios o (2) un medio de comunicación tal como una señal o una onda portadora. Los medios de almacenamiento de datos pueden ser cualquier medio disponible al que se pueda acceder desde uno o más ordenadores o uno o más procesadores para recuperar instrucciones, código y/o estructuras de datos para la implementación de las técnicas descritas en esta divulgación. Un producto de programa informático puede incluir un medio legible por ordenador.

20 **[0204]** A modo de ejemplo, y no de manera limitativa, tales medios de almacenamiento legibles por ordenador pueden comprender RAM, ROM, EEPROM, CD-ROM u otro almacenamiento de disco óptico, almacenamiento de disco magnético u otros dispositivos de almacenamiento magnético, memoria flash o cualquier otro medio que pueda usarse para almacenar código de programa deseado en forma de instrucciones o estructuras de datos y al que pueda accederse mediante un ordenador. Además, cualquier conexión recibe debidamente la denominación de medio legible por ordenador. Por ejemplo, si las instrucciones se transmiten desde una sede de la Red, un servidor u otro origen remoto usando un cable coaxial, un cable de fibra óptica, un par trenzado, una línea de abonado digital (DSL) o tecnologías inalámbricas tales como infrarrojos, radio y microondas, entonces el cable coaxial, el cable de fibra óptica, el par trenzado, la DSL o las tecnologías inalámbricas tales como infrarrojos, radio y microondas se incluyen en la definición de medio.

25 **[0205]** Sin embargo, debería entenderse que los medios de almacenamiento legibles por ordenador y los medios de almacenamiento de datos no incluyen conexiones, ondas portadoras, señales u otros medios transitorios, sino que, en cambio, se orientan a medios de almacenamiento tangibles no transitorios. Los discos, como se usan en el presente documento, incluyen el disco compacto (CD), el disco láser, el disco óptico, el disco versátil digital (DVD), el disco flexible y el disco Blu-ray, de los cuales el disco flexible habitualmente reproduce los datos magnéticamente, mientras que los demás discos reproducen los datos ópticamente con láseres. Las combinaciones de lo anterior también deberían incluirse dentro del alcance de los medios legibles por ordenador.

30 **[0206]** Las instrucciones pueden ser ejecutadas por uno o más procesadores, tales como uno o más procesadores de señales digitales (DSP), microprocesadores de propósito general, circuitos integrados específicos de la aplicación (ASIC), matrices de puertas programables *in situ* (FPGA) u otros circuitos lógicos integrados o discretos equivalentes. En consecuencia, el término «procesador», como se usa en el presente documento, puede referirse a cualquiera de las estructuras anteriores o a cualquier otra estructura adecuada para la implementación de las técnicas descritas en el presente documento. Además, en algunos aspectos, la funcionalidad descrita en el presente documento puede proporcionarse dentro de módulos de hardware y/o software dedicados, configurados para la codificación y la decodificación, o incorporados en un códec combinado. Asimismo, las técnicas podrían implementarse por completo en uno o más circuitos o elementos lógicos.

35 **[0207]** Las técnicas de esta divulgación se pueden implementar en una amplia variedad de dispositivos o aparatos, que incluyen un teléfono inalámbrico, un circuito integrado (CI) o un conjunto de CI (por ejemplo, un conjunto de chips). Diversos componentes, módulos o unidades se describen en esta divulgación para enfatizar aspectos funcionales de dispositivos configurados para realizar las técnicas divulgadas, pero no requieren necesariamente su realización mediante diferentes unidades de hardware. En cambio, como se ha descrito anteriormente, diversas unidades pueden combinarse en una unidad de hardware de códec o proporcionarse por medio de un grupo de unidades de hardware interoperativas, incluyendo uno o más procesadores, como se ha descrito anteriormente, conjuntamente con software y/o firmware adecuados.

REIVINDICACIONES

1. Un procedimiento de descodificación de datos de vídeo de múltiples visualizaciones, el procedimiento que comprende:

5 determinar si un índice de referencia de una imagen de referencia en una lista de imágenes de referencia para predecir un bloque actual corresponde a una imagen de referencia entre visualizaciones; y

10 basándose en el índice de referencia correspondiente a la imagen de referencia entre visualizaciones, obtener, a partir de un flujo de bits codificado, datos que indican si el bloque actual se predice con la predicción de síntesis de visualización (VSP) utilizando al menos una parte de una imagen sintetizada a partir de la imagen de referencia entre visualizaciones utilizando renderización basada en imágenes de profundidad;

15 en el que obtener los datos que indican si el bloque actual se predice con VSP comprende obtener un indicador de VSP a un nivel de bloque solo cuando el índice de referencia para el bloque actual corresponde a una imagen de referencia entre visualizaciones; y

20 descodificar el bloque actual utilizando la predicción de síntesis de visualización (VSP) si el indicador de VSP indica que se debe usar VSP.
2. El procedimiento según la reivindicación 1, que comprende además descodificar una cabecera de fragmento que no incluye una indicación de un índice de referencia de una imagen de referencia de síntesis de visualización para descodificar el bloque actual.
3. El procedimiento según la reivindicación 1, en el que los datos que indican si el bloque actual se predice con VSP comprenden datos que indican un primer modo VSP para el bloque actual, el procedimiento que comprende además obtener, a partir del flujo de bits codificado, datos que indican si se debe realizar uno del primer modo VSP y un segundo modo VSP, en el que realizar el segundo modo VSP comprende determinar una imagen de referencia de síntesis de visualización basándose en un índice de imagen de referencia de la imagen de referencia de síntesis de visualización.
4. El procedimiento según la reivindicación 1, que comprende además:

35 determinar una imagen de referencia de síntesis de visualización a partir de la imagen de referencia entre visualizaciones;

40 agregar la imagen de predicción de síntesis de visualización a la lista de imágenes de referencia, en el que un proceso de modificación de lista de imágenes de referencia no se utiliza para reordenar la imagen de predicción de síntesis de visualización; y

45 descodificar el bloque actual en relación con la imagen de referencia de síntesis de visualización.
5. Un procedimiento de codificación de datos de vídeo de múltiples visualizaciones, el procedimiento que comprende:

50 determinar si un índice de referencia de una imagen de referencia en una lista de imágenes de referencia para predecir un bloque actual corresponde a una imagen de referencia entre visualizaciones;

55 basándose en el índice de referencia correspondiente a la imagen de referencia entre visualizaciones, generar datos que indican si el bloque actual se predice con la predicción de síntesis de visualización (VSP) usando al menos una parte de una imagen sintetizada a partir de la imagen de referencia entre visualizaciones usando renderización basada en imágenes de profundidad; y

60 codificar los datos que indican el modo VSP en un flujo de bits;

65 en el que generar los datos que indican si el bloque actual se predice con VSP comprende generar un indicador de VSP a un nivel de bloque solo cuando el índice de referencia para el bloque actual corresponde a una imagen de referencia entre visualizaciones; y

60 codificar el bloque actual utilizando la predicción de síntesis de visualización (VSP) si el indicador de VSP indica que se debe usar VSP.
6. El procedimiento según la reivindicación 1 o 5, en el que el nivel de bloque es un nivel de macrobloque cuando el bloque actual es un macrobloque o un nivel de partición de macrobloque cuando el bloque actual es una partición de macrobloque.

- 5
7. El procedimiento según la reivindicación 5, que comprende además codificar una cabecera de fragmento que no incluye una indicación de un índice de referencia de una imagen de referencia de síntesis de visualización para codificar el bloque actual.
8. El procedimiento según la reivindicación 1 o 5, en el que la lista de imágenes de referencia comprende solo la lista de imágenes de referencia cero.
- 10
9. El procedimiento según la reivindicación 1 o 5, en el que la lista de imágenes de referencia comprende una de una primera lista de imágenes de referencia y una segunda lista de imágenes de referencia.
- 15
10. El procedimiento según la reivindicación 5, en el que los datos que indican si el bloque actual se predice con VSP comprenden datos que indican un primer modo VSP para el bloque actual, y que comprenden además datos de codificación que indican si se debe realizar uno del modo VSP y un segundo modo VSP, en el que realizar el segundo modo VSP comprende determinar una imagen de referencia de síntesis de visualización basándose en un índice de imagen de referencia de la imagen de referencia de síntesis de visualización.
- 20
11. El procedimiento según la reivindicación 5, que comprende además:
- determinar una imagen de referencia de síntesis de visualización a partir de la imagen de referencia entre visualizaciones;
- 25
- agregar la imagen de predicción de síntesis de visualización a la lista de imágenes de referencia, en el que un proceso de modificación de lista de imágenes de referencia no se utiliza para reordenar la imagen de predicción de síntesis de visualización; y
- codificar del bloque actual en relación con la imagen de referencia de síntesis de visualización.
- 30
12. El procedimiento según la reivindicación 1 o 5, que comprende además sintetizar una imagen de referencia de síntesis de visualización a partir de la imagen de referencia entre visualizaciones solo cuando la imagen de referencia entre visualizaciones se incluye en la lista de imágenes de referencia cero.
- 35
13. El procedimiento según la reivindicación 1 o 5, que comprende además sintetizar una imagen de referencia de síntesis de visualización a partir de la imagen de referencia entre visualizaciones solo cuando la imagen de referencia entre visualizaciones se incluye en la lista de imágenes de referencia uno.
- 40
14. Un aparato para codificar datos de vídeo de múltiples visualizaciones, el aparato que comprende:
- medios para determinar si un índice de referencia de una imagen de referencia en una lista de imágenes de referencia para predecir un bloque actual corresponde a una imagen de referencia entre visualizaciones; y
- 45
- basándose en el índice de referencia correspondiente a la imagen de referencia entre visualizaciones, medios para codificar datos que indican si el bloque actual se predice con la predicción de síntesis de visualización (VSP) usando al menos una parte de una imagen sintetizada a partir de la imagen de referencia entre visualizaciones usando renderización basada en imágenes de profundidad;
- 50
- en el que los medios para codificar los datos que indican si el bloque actual se predice con VSP comprenden medios para codificar un indicador de VSP a nivel de bloque solo cuando el índice de referencia para el bloque actual corresponde a una imagen de referencia entre visualizaciones; y
- medios para codificar el bloque actual utilizando la predicción de síntesis de visualización (VSP) si el indicador de VSP indica que se debe usar VSP.
- 55
15. Un medio no transitorio legible por ordenador que tiene instrucciones almacenadas en el mismo que, al ejecutarse, hacen que uno o más procesadores realicen el procedimiento de cualquiera de las reivindicaciones 1 a 13.

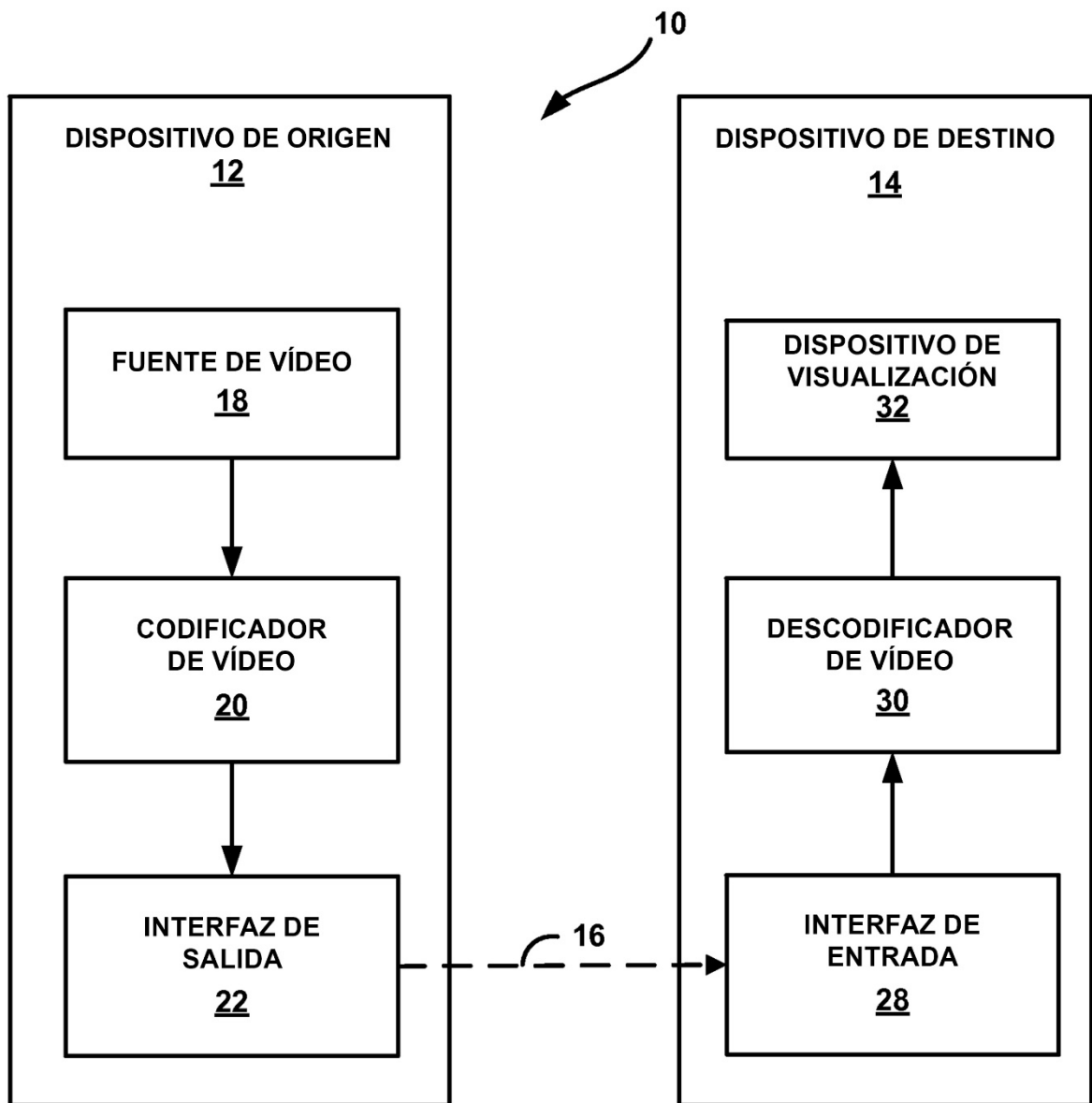


FIG. 1

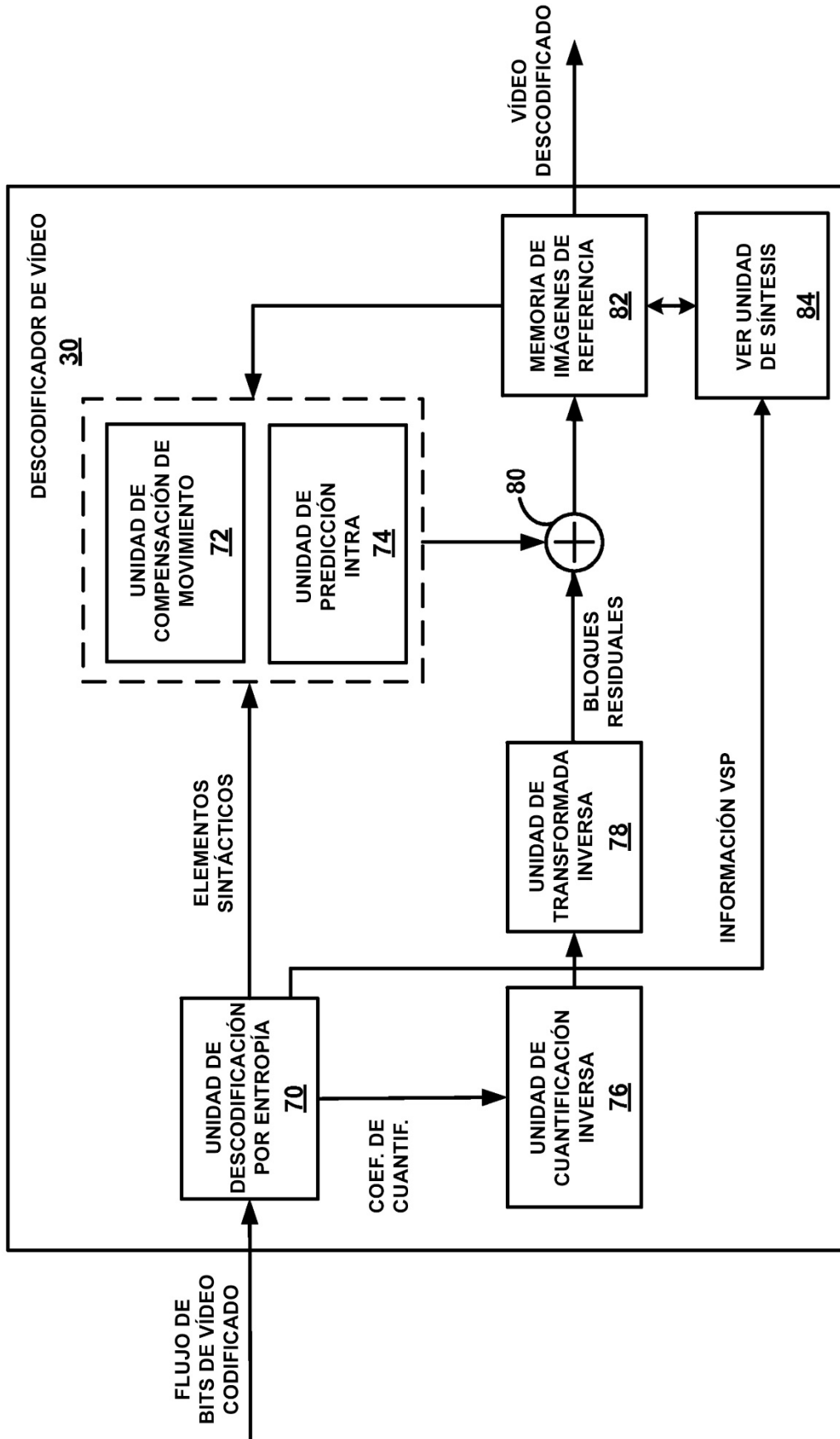


FIG. 3

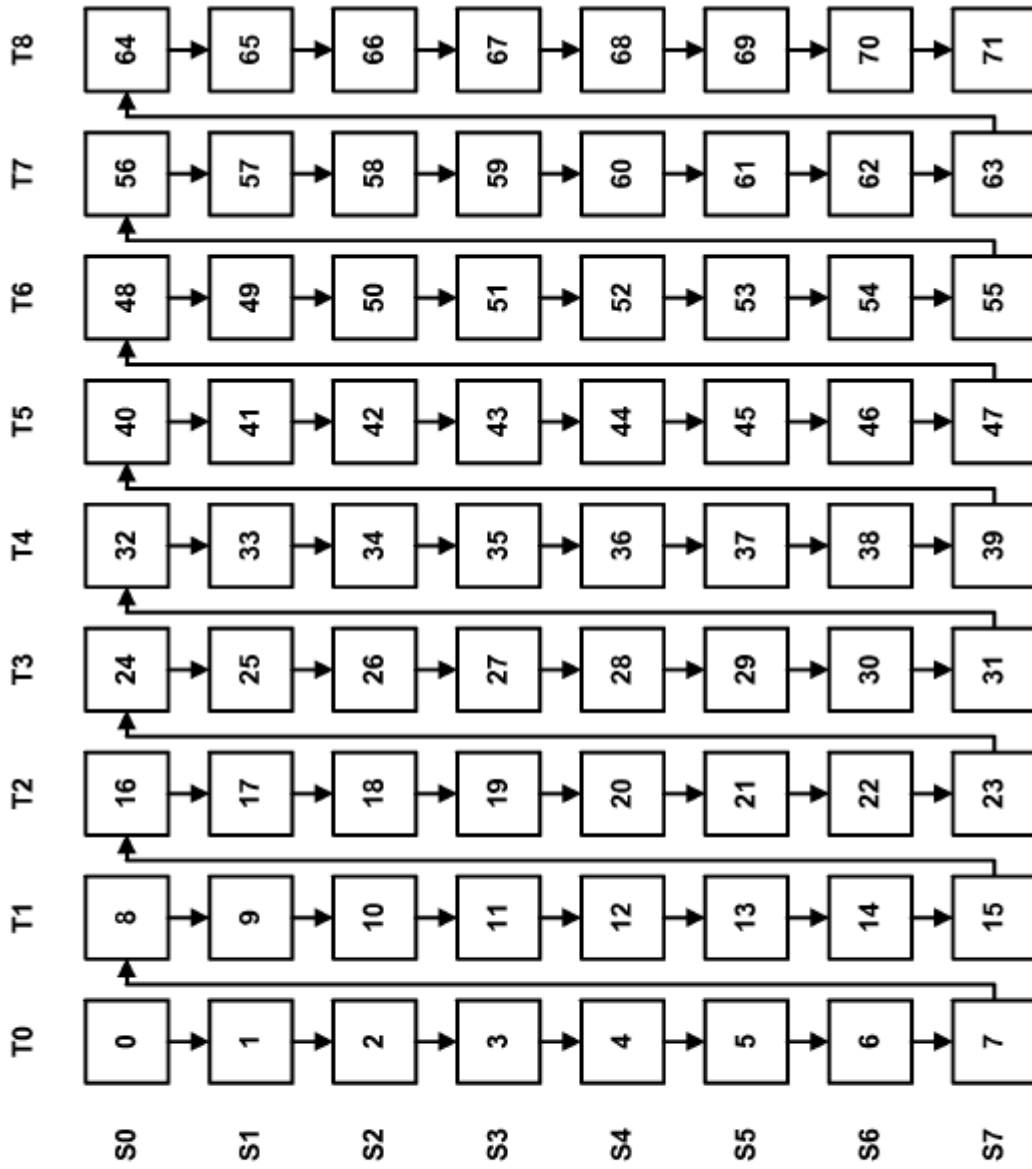


FIG. 4

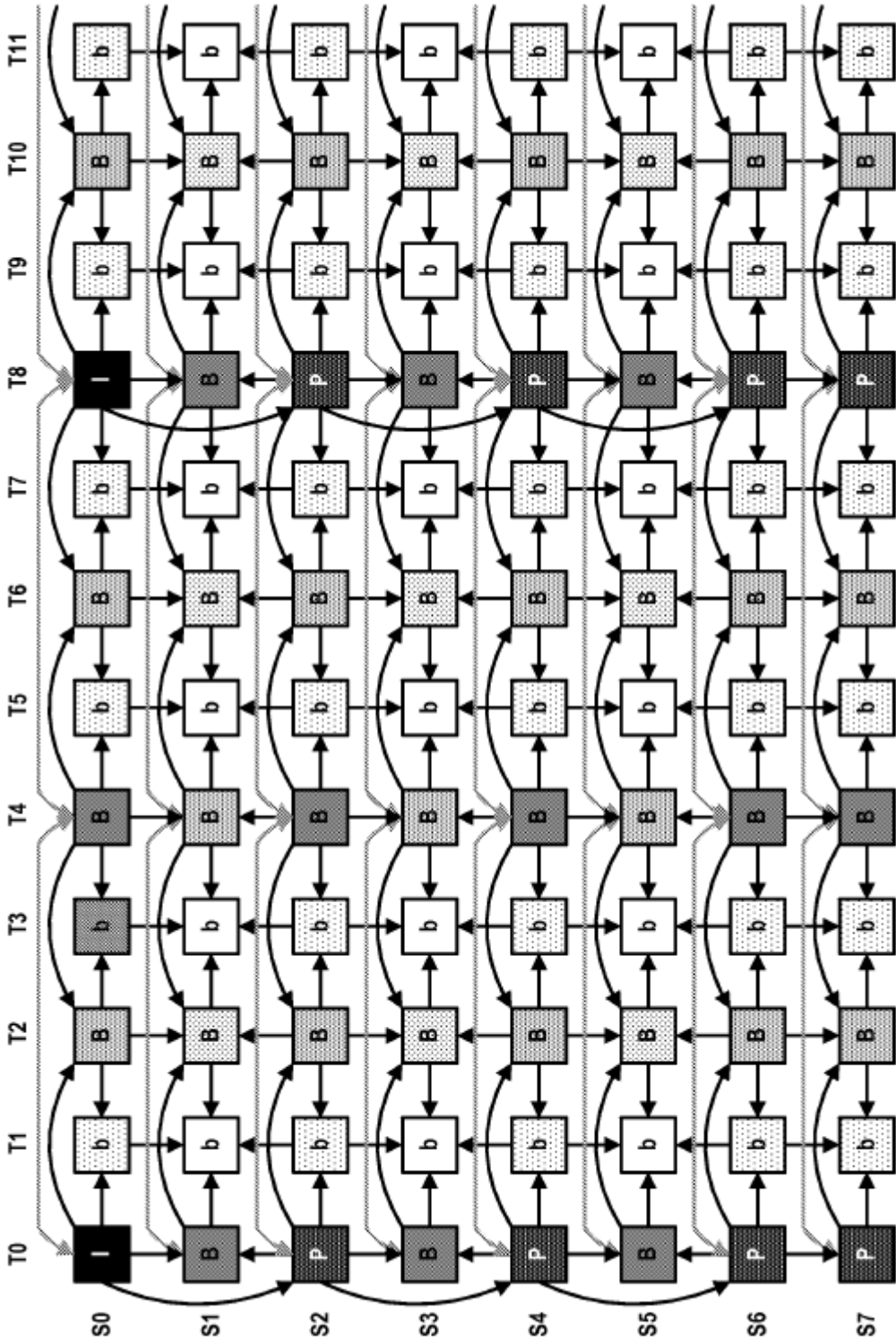


FIG. 5

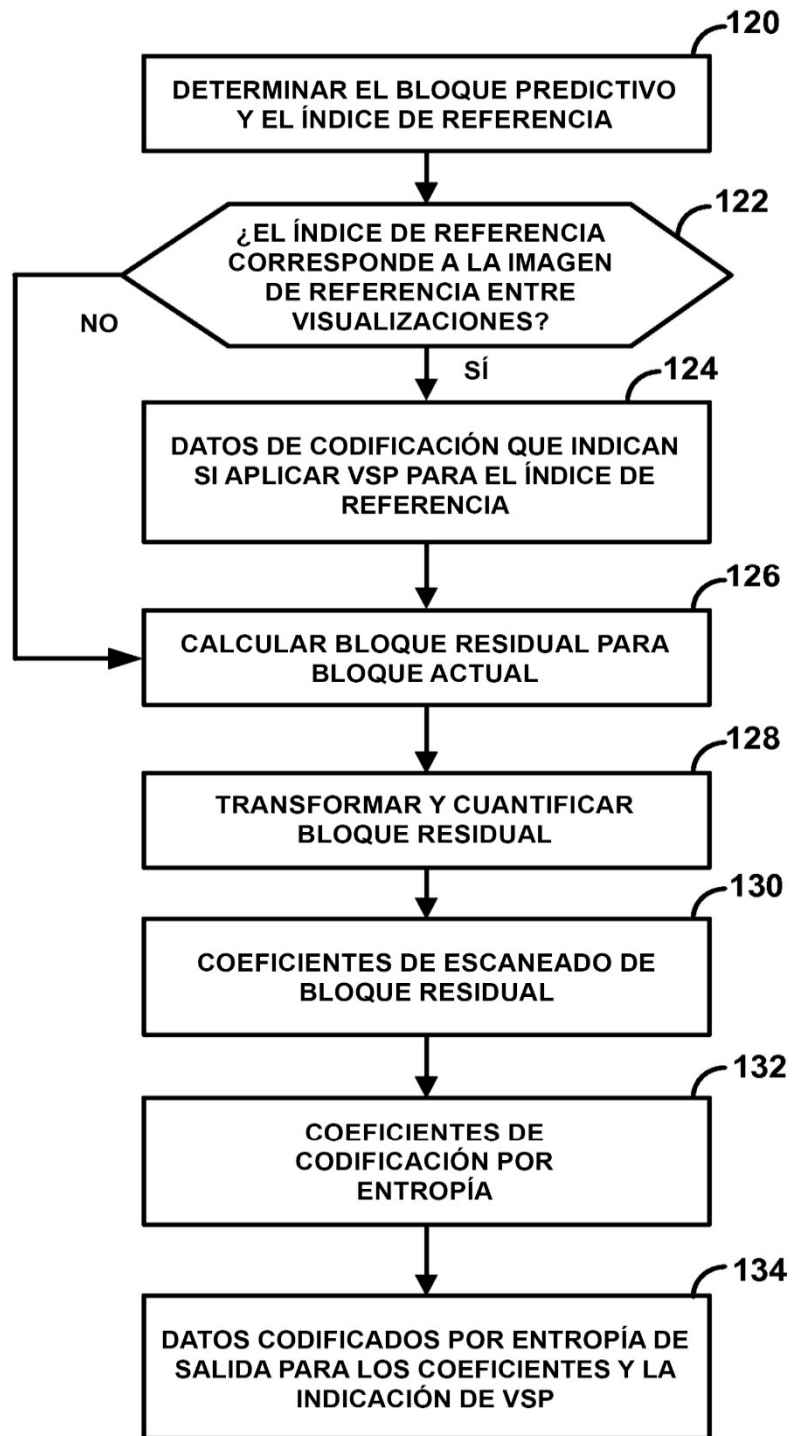


FIG. 6

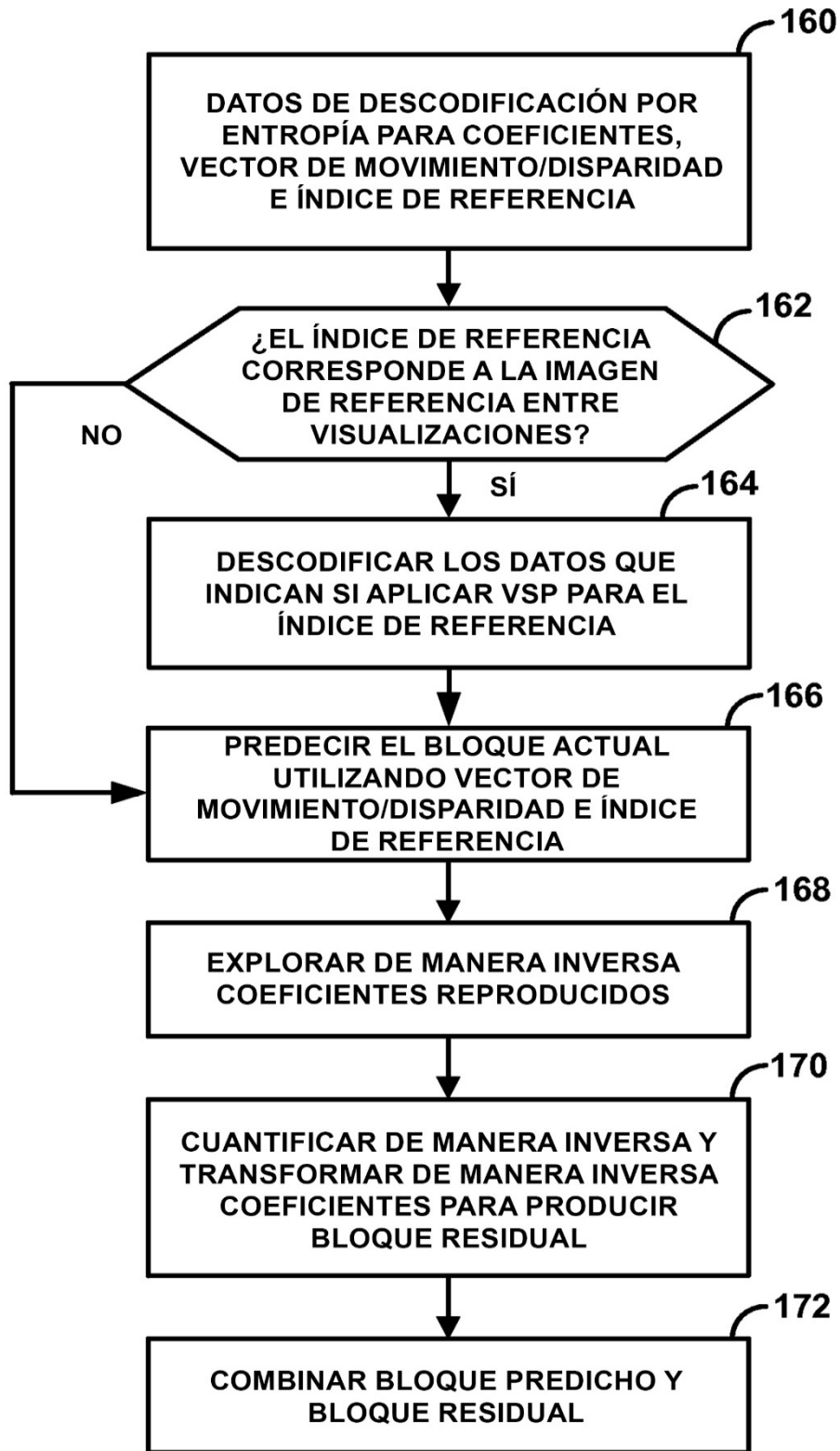


FIG. 7