

19



OFICINA ESPAÑOLA DE  
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 886 344**

51 Int. Cl.:

**H04N 7/00**

(2011.01)

12

## TRADUCCIÓN DE PATENTE EUROPEA

T3

86 Fecha de presentación y número de la solicitud internacional: **03.07.2013 PCT/US2013/049317**

87 Fecha y número de publicación internacional: **16.01.2014 WO14011480**

96 Fecha de presentación y número de la solicitud europea: **03.07.2013 E 13737774 (3)**

97 Fecha y número de publicación de la concesión europea: **18.08.2021 EP 2873233**

54 Título: **Predicción residual generalizada para codificación de vídeo escalable y codificación de vídeo 3D**

30 Prioridad:

**10.07.2012 US 201261670075 P  
27.09.2012 US 201261706692 P  
02.07.2013 US 201313933588**

45 Fecha de publicación y mención en BOPI de la traducción de la patente:  
**17.12.2021**

73 Titular/es:

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

72 Inventor/es:

**LI, XIANG;  
RAPAKA, KRISHNAKANTH;  
GUO, LIWEI;  
CHEN, JIANLE y  
KARCZEWICZ, MARTA**

74 Agente/Representante:

**ISERN JARA, Jorge**

ES 2 886 344 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

Predicción residual generalizada para codificación de vídeo escalable y codificación de vídeo 3D

### 5 CAMPO TÉCNICO

La presente divulgación se refiere a la codificación de vídeo.

### 10 ANTECEDENTES

Las capacidades de vídeo digital se pueden incorporar a una amplia gama de dispositivos, que incluyen televisores digitales, sistemas de radiodifusión digital directa, sistemas de radiodifusión inalámbrica, asistentes personales digitales (PDA), ordenadores portátiles o de escritorio, tabletas electrónicas, lectores de libros electrónicos, cámaras digitales, dispositivos de grabación digital, reproductores de medios digitales, dispositivos de videojuegos, consolas de videojuegos, teléfonos de radio celulares o por satélite, los denominados "teléfonos inteligentes", dispositivos de videoconferencia, dispositivos de transmisión continua de vídeo y similares. Los dispositivos de vídeo digital implementan técnicas de codificación 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 Eficacia (HEVC), actualmente en desarrollo, y las ampliaciones de dichas normas. Los dispositivos de vídeo pueden transmitir, recibir, codificar, decodificar y/o almacenar información de vídeo digital más eficazmente implementando dichas técnicas de codificación de vídeo.

Las técnicas de codificación de vídeo incluyen la predicción espacial (intraimagen) y/o la predicción temporal (interimagen) para reducir o eliminar la redundancia intrínseca en las secuencias de vídeo. En la codificación de vídeo basada en bloques, un fragmento de vídeo (por ejemplo, una trama de vídeo o una parte de una trama de vídeo) se puede dividir en bloques de vídeo, que también se pueden denominar bloques de árbol, unidades de codificación (CU) y/o nodos de codificación. Los bloques de vídeo en un fragmento intracodificado (I) de una imagen se codifican usando predicción espacial con respecto a unas muestras de referencia de bloques vecinos de la misma imagen. Los bloques de vídeo de un fragmento intercodificado (P o B) de una imagen pueden usar predicción espacial con respecto a unas muestras de referencia de bloques vecinos de la misma imagen o predicción temporal con respecto a unas muestras de referencia de otras imágenes de referencia. Las imágenes se pueden denominar tramas, y las imágenes de referencia se pueden denominar tramas de referencia.

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 de acuerdo con 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 se pueden transformar desde el dominio de píxel a un dominio de transformada, dando como resultado unos coeficientes de transformada residuales que, a continuación, se pueden cuantificar. Se puede hacer un barrido de los coeficientes de transformada cuantificados, dispuestos inicialmente en una matriz bidimensional, para producir un vector unidimensional de coeficientes de transformada, y se puede aplicar codificación por entropía para lograr aún más compresión.

Un documento titulado "Test model under consideration for HEVC based 3D video coding v 3.0" 30/4/12 IS)/IEC JTC1/SC29/WG11 describe técnicas para la predicción residual entre vistas. Un documento titulado "Multi-layer weighted prediction", convención del JVT, 24/07/2005, Poznan ISO/IEC JCT1/SC29/WG11 describe técnicas para determinar los parámetros de ponderación que se aplicarán a la predicción residual en capas sucesivas de una imagen multicapa.

### 50 SUMARIO

La invención se define en las reivindicaciones independientes, a las que se hace referencia con la característica preferida establecida en las reivindicaciones dependientes.

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

### 60 BREVE DESCRIPCIÓN DE LOS DIBUJOS

La figura 1 es un diagrama de bloques que ilustra un sistema de codificación y decodificación de vídeo de ejemplo que puede utilizar técnicas de acuerdo con los aspectos descritos en esta divulgación.

La figura 2 es un diagrama de bloques que ilustra un ejemplo de un codificador de vídeo que puede implementar técnicas de acuerdo con aspectos descritos en esta divulgación.

La figura 3 es un diagrama de bloques que ilustra un ejemplo de un decodificador de vídeo que puede implementar técnicas de acuerdo con aspectos descritos en esta divulgación.

La figura 4 es un diagrama de flujo que ilustra un procedimiento de ejemplo para una predicción residual generalizada de acuerdo con aspectos de esta divulgación.

5 La figura 4A es un diagrama de flujo que ilustra otro procedimiento de ejemplo para una predicción residual generalizada de acuerdo con aspectos de esta divulgación.

La figura 5 es un diagrama de flujo que ilustra un procedimiento de ejemplo para una predicción residual generalizada usando decodificación de bucle único de acuerdo con aspectos de esta divulgación.

10 La figura 6 es un diagrama de flujo que ilustra un procedimiento de ejemplo para una predicción residual generalizada usando decodificación multibucle de acuerdo con aspectos de esta divulgación.

La figura 7 es un diagrama de flujo que ilustra un procedimiento de ejemplo para señalar parámetros de predicción residual generalizada de acuerdo con aspectos de esta divulgación.

La figura 8 es un diagrama de flujo que ilustra un procedimiento de ejemplo para obtener parámetros de predicción residual generalizada de acuerdo con aspectos de esta divulgación.

15 La figura 9 es un diagrama de flujo que ilustra un procedimiento de ejemplo para muestrear de manera ascendente o descendente una capa en la predicción residual generalizada de acuerdo con aspectos de esta divulgación.

La figura 10 es un diagrama de flujo que ilustra un procedimiento de ejemplo para remapear, muestrear de manera ascendente o muestrear de manera descendente información de movimiento en la predicción residual generalizada de acuerdo con aspectos de esta divulgación.

20 La figura 11 es un diagrama de flujo que ilustra un procedimiento de ejemplo para determinar un factor de ponderación para codificar en la predicción residual generalizada de acuerdo con aspectos de esta divulgación.

#### DESCRIPCIÓN DETALLADA

25 Las técnicas descritas en esta divulgación se refieren, en general, a la codificación de vídeo escalable (SVC) y la codificación de vídeo 3D. Por ejemplo, las técnicas pueden referirse a y usarse con o en una ampliación de la codificación de vídeo escalable (SVC) de Codificación de Vídeo de Alta Eficacia (HEVC). En una ampliación de SVC, podría haber múltiples capas de información de vídeo. La capa en el nivel más bajo puede servir como capa base (BL), y la capa en la parte superior puede servir como capa mejorada (EL). La "capa mejorada" a veces se denomina "capa de mejora", y estos términos se pueden usar de manera intercambiable. Todas las capas centrales pueden servir como EL y/o BL. Por ejemplo, una capa central puede ser una EL para las capas por debajo de la misma, tal como la capa base o cualquier capa de mejora intermedia, y al mismo tiempo servir como RL para las capas de mejora dispuestas sobre la misma.

35 Para fines de ilustración solamente, las técnicas descritas en la divulgación se describen con ejemplos que incluyen solo dos capas (por ejemplo, una capa de nivel inferior, tal como la capa base, y una capa de nivel superior, tal como la capa mejorada). Debe entenderse que los ejemplos descritos en esta divulgación pueden ampliarse también a ejemplos con múltiples capas base y capas de mejora.

40 Las normas de codificación de vídeo 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), incluidas sus ampliaciones de Codificación de Vídeo Escalable (SVC) y de Codificación de Vídeo Multivista (MVC). Además, una nueva norma de codificación de vídeo, concretamente la Codificación de Vídeo de Alta Eficacia (HEVC), 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 reciente de HEVC está disponible en [http://wg11.sc29.org/jct/doc\\_end\\_user/current\\_document.php?id=5885/JCTVC-I1003-v2](http://wg11.sc29.org/jct/doc_end_user/current_document.php?id=5885/JCTVC-I1003-v2), a fecha de 7 de junio de 2012. Otro borrador reciente de la norma HEVC, denominado "HEVC Working Draft 7", se puede descargar de [http://phenix.it-sudparis.eu/ict/doc\\_end\\_user/documents/9\\_Geneva/wg11/JCTVC-11003-v3.zip](http://phenix.it-sudparis.eu/ict/doc_end_user/documents/9_Geneva/wg11/JCTVC-11003-v3.zip), a fecha de 7 de junio de 2012. La cita completa del borrador de trabajo 7 de HEVC es el documento HCTVC-11003, de Bross et al., titulado "High Efficiency Video Coding (HEVC) Text Specification Draft 7", Equipo de Colaboración Conjunta en Codificación de Vídeo (JCT-VC) de ITU-T SG16 WP3 e ISO/IEC JTC1/SC29/WG11, 9ª convención: Ginebra, Suiza, del 27 de abril de 2012 al 7 de mayo de 2012.

55 La codificación de vídeo escalable (SVC) se puede usar para proporcionar escalabilidad de calidad (también denominada relación de señal a ruido (SNR)), escalabilidad espacial y/o escalabilidad temporal. Una capa mejorada puede tener una resolución espacial diferente a una capa base. Por ejemplo, la relación de aspecto espacial entre la EL y la BL puede ser 1,0, 1,5, 2,0 u otras relaciones diferentes. En otras palabras, el aspecto espacial de la EL puede ser igual a 1,0, 1,5 o 2,0 veces el aspecto espacial de la BL. En algunos ejemplos, el factor de escalamiento de la EL puede ser mayor que la de la BL. Por ejemplo, un tamaño de imágenes en la EL puede ser mayor que un tamaño de imágenes en la BL. De esta manera, puede ser posible, aunque no es una limitación, que la resolución espacial de la EL sea mayor que la resolución espacial de la BL.

65 En la ampliación de SVC para H.264, la predicción de un bloque actual se puede realizar usando las diferentes capas que se proporcionan para SVC. Dicha predicción se puede denominar predicción entre capas. Los procedimientos de predicción entre capas se pueden utilizar en SVC para reducir la redundancia entre capas. Algunos ejemplos de predicción entre capas pueden incluir intrapredicción entre capas, predicción de movimiento entre capas y predicción

residual entre capas. La intrapredicción entre capas usa la reconstrucción de bloques ubicados conjuntamente en la capa base para predecir el bloque actual en la capa de mejora. La predicción de movimiento entre capas usa movimiento de la capa base para predecir el movimiento en la capa de mejora. La predicción residual entre capas usa el residuo de la capa base para predecir el residuo de la capa de mejora.

En la predicción residual entre capas, el residuo de la capa base se puede usar para predecir el bloque actual de la capa de mejora. El residuo puede definirse como la diferencia entre la predicción temporal para una unidad de vídeo y la unidad de vídeo de origen. En la predicción residual, el residuo de la capa base también se considera al predecir el bloque actual. Por ejemplo, el bloque actual puede reconstruirse usando el residuo de la capa de mejora, la predicción temporal a partir de la capa de mejora y el residuo de la capa base. El bloque actual se puede reconstruir de acuerdo con la siguiente ecuación:

$$\hat{I}_e = r_e + P_e + r_b \quad (1)$$

donde  $\hat{I}_e$  denota la reconstrucción del bloque actual,  $r_e$  denota el residuo de la capa de mejora,  $P_e$  denota la predicción temporal a partir de la capa de mejora y  $r_b$  denota la predicción residual a partir de la capa base.

Para usar la predicción residual entre capas para un macrobloque (MB) en la capa de mejora, el macrobloque ubicado conjuntamente en la capa base debe ser un inter-MB, y el residuo del macrobloque de capa base ubicado conjuntamente puede muestrearse de manera ascendente de acuerdo con la relación de resolución espacial de la capa de mejora (por ejemplo, porque las capas en SVC pueden tener diferentes resoluciones espaciales). En la predicción residual entre capas, la diferencia entre el residuo de la capa de mejora y el residuo de la capa base muestreada de manera ascendente puede codificarse en el flujo de bits. El residuo de la capa base puede normalizarse en base a la relación entre las etapas de cuantificación de las capas base y de mejora.

La ampliación de SVC a H.264 requiere decodificación de bucle único para la compensación de movimiento a fin de mantener una baja complejidad para el decodificador. En general, la compensación de movimiento se realiza añadiendo la predicción temporal y el residuo para el bloque actual como sigue:

$$\hat{I} = r + P \quad (2)$$

donde  $\hat{I}$  denota la trama actual,  $r$  denota el residuo y  $P$  denota la predicción temporal. En la decodificación de bucle único, cada capa admitida en SVC se puede decodificar con un solo bucle de compensación de movimiento. Para lograr esto, todas las capas que se usan para intraprededir entre capas las capas superiores se codifican usando intrapredicción restringida. En la intrapredicción restringida, los MB de intramodo se intracodifican sin hacer referencia a ninguna muestra de MB intercodificados vecinos. Por otro lado, HEVC permite la decodificación multibucle para SVC, en la que una capa de SVC se puede decodificar usando múltiples bucles de compensación de movimiento. Por ejemplo, primero se decodifica completamente la capa base y, a continuación, se decodifica la capa de mejora.

La predicción residual formulada en la ecuación (1) puede ser una técnica eficiente en la ampliación H.264 SVC. Sin embargo, su rendimiento se puede mejorar aún más en la ampliación HEVC SVC, especialmente cuando se usa decodificación multibucle en la ampliación HEVC SVC.

En el caso de la decodificación multibucle, se puede usar compensación de movimiento en el dominio de diferencia en lugar de la predicción residual. En SVC, una capa de mejora puede codificarse usando codificación de dominio de píxel o codificación de dominio de diferencia. En la codificación de dominio de píxel, los píxeles de entrada para los píxeles de una capa de mejora pueden codificarse, como para una capa HEVC que no es SVC. Por otro lado, en la codificación de dominio de diferencia, pueden codificarse valores de diferencia para una capa de mejora. Los valores de diferencia pueden ser la diferencia entre los píxeles de entrada para la capa de mejora y los píxeles reconstruidos de capa base escalada correspondiente. Dichos valores de diferencia se pueden usar en la compensación de movimiento para la compensación de movimiento en el dominio de diferencia.

En la intercodificación usando el dominio de diferencia, el bloque predicho actual se determina en base a los valores de diferencia entre las muestras de bloque predichas correspondientes en la imagen de referencia de capa de mejora y las muestras de bloque predichas correspondientes en la imagen de referencia de capa base escalada. Los valores de diferencia pueden denominarse bloque de predicción de diferencia. Las muestras reconstruidas de capa base ubicadas conjuntamente se añaden al bloque de predicción de diferencia para obtener muestras reconstruidas de capa de mejora.

Sin embargo, el uso de la compensación de movimiento de dominio de diferencia en la predicción entre capas introduce dos conjuntos de estimación de movimiento y compensación de movimiento, ya que la estimación de movimiento y la compensación de movimiento se utilizan a menudo tanto para el dominio de píxel como para el dominio de diferencia. La introducción de dos conjuntos de estimación de movimiento y compensación de movimiento puede dar lugar a mayor coste computacional y de búfer, lo que puede no ser práctico para un codificador o un decodificador. Además, la codificación de dos conjuntos de vectores de movimiento puede reducir la eficiencia de la codificación, ya que el

campo de movimiento puede volverse irregular cuando los dos conjuntos de vectores de movimiento tienen propiedades diferentes y están intercalados a nivel de unidad de codificación (CU). Además, la estimación de movimiento en el dominio de diferencia requiere que la capa base y la capa de mejora compartan el mismo movimiento. Además, la compensación de movimiento en el dominio de diferencia no funciona con la decodificación de bucle único, ya que la obtención de imágenes diferenciales entre dos capas se basa en imágenes completamente reconstruidas de cada capa. Por consiguiente, sería ventajoso evitar la redundancia al tener dos conjuntos de estimación de movimiento y compensación de movimiento cuando se usa compensación de movimiento en el dominio de diferencia. Además, sería ventajoso ampliar la compensación de movimiento en el dominio de diferencia en la decodificación de bucle único.

Las técnicas descritas en esta divulgación pueden abordar cuestiones relacionadas con la predicción residual entre capas y la compensación de movimiento en el dominio de diferencia en SVC. Las técnicas pueden proporcionar un marco de predicción residual generalizada (GRP). Como se explicó anteriormente, la predicción residual entre capas usa el residuo de la capa de referencia para predecir la unidad de vídeo actual, por ejemplo, un bloque o una trama. En la predicción residual generalizada, la predicción residual entre capas de la unidad de vídeo actual puede basarse en el residuo de la capa actual, la predicción temporal de la capa actual y el residuo de la capa de referencia. El residuo de la capa de referencia puede ajustarse mediante un factor de ponderación. El factor de ponderación puede basarse en e incluir diversos tipos de información. Ejemplos de dicha información pueden incluir un número de candidatos de ponderación, una etapa de ponderación, un índice de ponderación y una tabla de ponderación. El marco GRP también se puede aplicar a la intrapredicción. Cuando se usa la intrapredicción, la predicción espacial a partir de la capa actual se usa en el marco de GRP, en lugar de la predicción temporal a partir de la capa actual. Por ejemplo, la predicción residual entre capas de la unidad de vídeo actual puede basarse en el residuo de la capa actual, la predicción espacial de la capa actual y el residuo de la capa de referencia.

El marco de GRP de acuerdo con los aspectos de esta divulgación puede permitir diversos tipos de predicción residual mediante la incorporación de un factor de ponderación. El ajuste apropiado del factor de ponderación puede dar lugar a ganancias de codificación significativas para la predicción residual. Además, en el marco de GRP, la predicción residual se puede realizar usando una capa de referencia que no es necesariamente la capa base en la predicción residual tradicional. Por ejemplo, la capa de referencia puede obtenerse de la capa de mejora actual. La GRP también puede adaptarse a la predicción residual tradicional cuando el factor de ponderación se establece en 1. El marco de GRP se puede utilizar con decodificación de bucle único y decodificación multibucle. Además, en el marco de GRP, la estimación de movimiento en el dominio de diferencia puede no ser necesaria y, por lo tanto, la capa actual y la capa de mejora no tienen que compartir el mismo movimiento para la estimación de movimiento. El marco de GRP se puede aplicar a muchos tipos diferentes de predicción residual, y la predicción residual tradicional como se define en la ecuación (1) y la compensación de movimiento en el dominio de diferencia son dos escenarios específicos del uso del marco de GRP. Las técnicas pueden mejorar el rendimiento de la compensación de movimiento en la ampliación escalable de HEVC y también pueden aplicarse a la ampliación de codificación de vídeo 3D de HEVC.

A continuación en el presente documento se describen de forma más detallada diversos aspectos de los sistemas, aparatos y procedimientos novedosos, con referencia a los dibujos adjuntos. Sin embargo, esta divulgación se puede realizar de muchas formas diferentes y no se debería interpretar que está limitada a alguna estructura o función específica presentada a lo largo de esta divulgación. Más bien, estos aspectos se proporcionan de modo que esta divulgación sea exhaustiva y completa, y transmita de forma detallada el alcance de la divulgación a los expertos en la técnica. En base a las enseñanzas del presente documento, un experto en la técnica debe apreciar que el alcance de la divulgación está concebido para abarcar cualquier aspecto de los sistemas, aparatos y procedimientos novedosos divulgados en el presente documento, ya se implementen de forma independiente de, o en combinación con, cualquier otro aspecto de la invención. Por ejemplo, un aparato se puede implementar o un procedimiento se puede llevar a la práctica usando un número cualquiera de los aspectos expuestos en el presente documento. Además, el alcance de la invención pretende abarcar un aparato o procedimiento de este tipo que se lleve a la práctica usando otra estructura, funcionalidad, o estructura y funcionalidad, además de o aparte de los diversos aspectos de la invención expuestos en el presente documento. Se debe entender que cualquier aspecto divulgado en el presente documento se puede realizar mediante uno o más elementos de una reivindicación.

Aunque en el presente documento se describen aspectos particulares, muchas variantes y permutaciones de estos aspectos se hallan dentro del alcance de la divulgación. Aunque se mencionan algunos beneficios y ventajas de los aspectos preferentes, no se pretende limitar el alcance de la divulgación a beneficios, usos u objetivos particulares. En cambio, los aspectos de la divulgación pretenden ser ampliamente aplicables a diferentes tecnologías inalámbricas, configuraciones de sistema, redes y protocolos de transmisión, de los que algunos se ilustran a modo de ejemplo en las figuras y en la siguiente descripción de los aspectos preferentes. La descripción detallada y los dibujos son meramente ilustrativos de la divulgación, en lugar de limitantes, estando definido el alcance de la divulgación por las reivindicaciones adjuntas y equivalentes de las mismas.

La figura 1 es un diagrama de bloques que ilustra un sistema de codificación y decodificación de vídeo de ejemplo que puede utilizar técnicas de acuerdo con los aspectos descritos en esta divulgación. Como se muestra en la figura 1, el sistema 10 incluye un dispositivo de origen 12 que proporciona datos de vídeo codificados que se van a decodificar 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 por medio de un medio legible por ordenador 16. El dispositivo de origen 12 y el dispositivo de destino 14 pueden comprender cualquiera de entre una amplia gama de dispositivos, incluidos ordenadores de escritorio, ordenadores plegables (por ejemplo, portátiles), tabletas electrónicas, decodificadores, equipos telefónicos de mano tales como los denominados teléfonos "inteligentes", los denominados paneles "inteligentes", televisores, cámaras, dispositivos de visualización, reproductores de medios digitales, consolas de videojuegos, dispositivos de transmisión continua de vídeo o similares. En algunos casos, el dispositivo de origen 12 y el dispositivo de destino 14 pueden estar equipados para la comunicación inalámbrica.

El dispositivo de destino 14 puede recibir los datos de vídeo codificados que se van a decodificar por medio del medio legible por ordenador 16. El medio legible por ordenador 16 puede comprender cualquier tipo de medio o dispositivo que puede transportar 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 para permitir que el dispositivo de origen 12 transmita datos de vídeo codificados directamente al dispositivo de destino 14 en tiempo real. Los datos de vídeo codificados se pueden modular de acuerdo con una norma de comunicación, tal como un protocolo de comunicación inalámbrica, y transmitir al dispositivo de destino 14. 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 físicas de transmisión. 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 encaminadores, conmutadores, estaciones base o cualquier otro equipo que pueda ser útil para facilitar la comunicación desde el dispositivo de origen 12 al dispositivo de destino 14.

En algunos ejemplos, se pueden emitir datos codificados desde la interfaz de salida 22 hasta un dispositivo de almacenamiento. De forma similar, se puede acceder a los datos codificados desde el dispositivo de almacenamiento mediante una interfaz de entrada. El dispositivo de almacenamiento puede incluir cualquiera de una variedad de medios de almacenamiento de datos, distribuidos o de acceso local, tales como un disco duro, discos Blu-ray, DVD, CD-ROM, memoria flash, memoria volátil o no volátil, o cualquier otro medio de almacenamiento digital adecuado para almacenar datos de vídeo codificados. En un ejemplo adicional, el dispositivo de almacenamiento puede corresponder a un servidor de archivos o a otro dispositivo de almacenamiento intermedio que pueda almacenar el vídeo codificado generado por el dispositivo de origen 12. El dispositivo de destino 14 puede acceder a datos de vídeo almacenados desde el dispositivo de almacenamiento por medio de transmisión continua o descarga. El servidor de archivos puede ser cualquier tipo de servidor que pueda almacenar datos de vídeo codificados y transmitir esos datos de vídeo codificados al dispositivo de destino 14. Servidores de archivos de ejemplo incluyen un servidor web (por ejemplo, para un sitio web), un servidor FTP, dispositivos de almacenamiento conectado a red (NAS) o una unidad de disco local. El dispositivo de destino 14 puede acceder a los datos de vídeo codificados a través de cualquier conexión de datos estándar, incluida una conexión a Internet. Esto puede incluir un canal inalámbrico (por ejemplo, una conexión Wi-Fi), 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 las mismas.

Las técnicas de esta divulgación no están limitadas necesariamente a aplicaciones o configuraciones inalámbricas. Las técnicas se pueden aplicar a la codificación de vídeo como soporte de cualquiera de una variedad de aplicaciones multimedia, tales como radiodifusiones de televisión por el aire, transmisiones de televisión por cable, transmisiones de televisión por satélite, transmisiones continuas de vídeo por Internet, tales como la transmisión continua adaptativa dinámica por HTTP (DASH), vídeo digital que se codifica en un medio de almacenamiento de datos, decodificación de vídeo digital almacenado en un medio de almacenamiento de datos u otras aplicaciones. En algunos ejemplos, el sistema 10 puede estar configurado para admitir una transmisión de vídeo unidireccional o bidireccional para admitir aplicaciones tales como transmisión continua de vídeo, reproducción de vídeo, radiodifusión de vídeo y/o videotelefonía.

En el ejemplo de la figura 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 decodificador 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 para codificar un flujo de bits que incluye datos de vídeo según múltiples normas o ampliaciones de normas. En otros ejemplos, un dispositivo de origen y un dispositivo de destino pueden incluir otros componentes o disposiciones. Por ejemplo, el dispositivo de origen 12 puede recibir datos de vídeo desde una fuente de vídeo externa 18, tal como una cámara externa. Del mismo modo, el dispositivo de destino 14 puede interactuar con un dispositivo de visualización externo, en lugar de incluir un dispositivo de visualización integrado.

El sistema 10 ilustrado de la figura 1 es simplemente un ejemplo. Cualquier dispositivo de codificación y/o decodificación de vídeo digital puede realizar técnicas para determinar candidatos para una lista de candidatos para predictores de vectores de movimiento para un bloque actual. Aunque, en general, las técnicas de esta divulgación se realizan mediante un dispositivo de codificación de vídeo, las técnicas también se pueden realizar mediante un codificador/decodificador de vídeo, denominado típicamente "CÓDEC". Además, las técnicas de esta divulgación también se pueden realizar mediante un preprocesador de vídeo. El dispositivo de origen 12 y el dispositivo de destino

14 son simplemente ejemplos de dichos dispositivos de codificación, donde el dispositivo de origen 12 genera datos de vídeo codificados para su transmisión al dispositivo de destino 14. En algunos ejemplos, los dispositivos 12, 14 pueden funcionar de manera sustancialmente simétrica, de modo que cada uno de los dispositivos 12, 14 incluye componentes de codificación y decodificación de vídeo. Por consiguiente, el sistema 10 pueda admitir una transmisión de vídeo unidireccional o bidireccional entre los dispositivos de vídeo 12, 14, por ejemplo, para transmisión continua de vídeo, reproducción de vídeo, radiodifusión de vídeo o videotelefonía.

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 suministro de vídeo para recibir vídeo desde un proveedor de contenidos de vídeo. Como otra alternativa, la fuente de vídeo 18 puede generar datos basados en gráficos de ordenador como fuente de vídeo, o una combinación de vídeo en directo, vídeo archivado y vídeo generado por ordenador. En algunos casos, si la fuente de vídeo 18 es una cámara de vídeo, el dispositivo de origen 12 y el dispositivo de destino 14 pueden formar los denominados teléfonos con cámara o videotelefonos. Sin embargo, como se menciona anteriormente, las técnicas descritas en esta divulgación pueden ser aplicables a la codificación de vídeo en general, y se pueden aplicar a aplicaciones inalámbricas y/o alámbricas. En cada caso, el codificador de vídeo 20 puede codificar el vídeo capturado, precapturado o generado por ordenador. Después, la información de vídeo codificada se puede emitir por la interfaz de salida 22 a un medio legible por ordenador 16.

El medio legible por ordenador 16 puede incluir medios transitorios, tales como una radiodifusión inalámbrica o una transmisión de red alámbrica, o medios de almacenamiento (es decir, medios de almacenamiento no transitorios), 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, por medio de una transmisión en red, comunicación alámbrica directa, etc. De manera similar, un dispositivo informático de una instalación de producción de medios, tal como una instalación de estampado 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, se puede entender que el medio legible por ordenador 16 incluye uno o más medios legibles por ordenador de diversas formas, en diversos ejemplos.

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 de sintaxis definida por el codificador de vídeo 20, que también es usada por el decodificador de vídeo 30, que incluye elementos de sintaxis que describen características y/o procesamiento de bloques y otras unidades codificadas, por ejemplo, los GOP. El dispositivo de visualización 32 muestra los datos de vídeo decodificados 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.

El codificador de vídeo 20 y el decodificador de vídeo 30 pueden funcionar de acuerdo con una norma de codificación de vídeo, tal como la norma de Codificación de Vídeo de Alta Eficacia (HEVC), actualmente en fase de desarrollo, y pueden ajustarse al Modelo de Prueba de HEVC (HM). De forma alternativa, el codificador de vídeo 20 y el decodificador de vídeo 30 pueden funcionar de acuerdo con otras normas de propiedad o industriales, tales como la norma ITU-T H.264, denominada de forma alternativa MPEG-4, Parte 10, Codificación de Vídeo Avanzada (AVC), o ampliaciones de dichas normas. Sin embargo, las técnicas de esta divulgación no se limitan a ninguna norma de codificación particular, incluidas, pero sin limitarse a, cualquiera de las normas enumeradas anteriormente. Otros ejemplos de normas de codificación de vídeo incluyen MPEG-2 e ITU-T H.263. Aunque no se muestra en la figura 1, en algunos aspectos, el codificador de vídeo 20 y el decodificador de vídeo 30 se pueden integrar cada uno con un codificador y un decodificador de audio, y pueden incluir unidades MUX-DEMUX apropiadas, u otro hardware y software, para gestionar la codificación tanto de audio como de vídeo en un flujo de datos común o en flujos de datos separados. Cuando proceda, las unidades MUX-DEMUX pueden ajustarse al protocolo de multiplexor ITU H.223 u otros protocolos, tales como el protocolo de datagramas de usuario (UDP).

Tanto el codificador de vídeo 20 como el decodificador de vídeo 30 se pueden implementar como cualquiera de una variedad de circuitos de codificador adecuados, tales como uno o más microprocesadores, procesadores de señales digitales (DSP), circuitos integrados específicos de la aplicación (ASIC), matrices de puertas programables *in situ* (FPGA), lógica discreta, software, hardware, firmware o cualquier combinación de los mismos. Cuando las técnicas se implementan parcialmente en software, un dispositivo puede almacenar instrucciones para el software en un medio adecuado no transitorio legible por ordenador y ejecutar las instrucciones en hardware usando uno o más procesadores para realizar las técnicas de esta divulgación. Tanto el codificador de vídeo 20 como el decodificador de vídeo 30 pueden estar incluidos en uno o más codificadores o decodificadores, cualquiera de los cuales puede estar integrado como parte de un codificador/decodificador (CÓDEC) combinado en un dispositivo respectivo. Un dispositivo que incluye el codificador de vídeo 20 y/o el decodificador de vídeo 30 puede comprender un circuito integrado, un microprocesador y/o un dispositivo de comunicación inalámbrica, tal como un teléfono celular.

El JCT-VC está trabajando en la elaboración de la norma HEVC. Los trabajos de normalización de HEVC se basan en un modelo en evolución de un dispositivo de codificación de vídeo denominado Modelo de Prueba de HEVC (HM). El HM supone varias capacidades adicionales de los dispositivos de codificación de vídeo en relación con los dispositivos existentes de acuerdo con, por ejemplo, ITU-T H.264/AVC. Por ejemplo, mientras que H.264 proporciona nueve modos de codificación por intrapredicción, el HM puede proporcionar hasta treinta y tres modos de codificación por intrapredicción.

En general, el modelo de trabajo del HM describe que una trama o imagen de vídeo se puede dividir en una secuencia de bloques de árbol o unidades de codificación más grandes (LCU), que incluyen tanto muestras de luma como de croma. Los datos de sintaxis dentro de un flujo de bits pueden definir un tamaño para la LCU, que es la unidad de codificación más grande en cuanto al número de píxeles. Un fragmento incluye un número de bloques de árbol consecutivos en orden de codificación. Una trama o imagen de vídeo se puede dividir en uno o más fragmentos. Cada bloque de árbol se puede dividir en unidades de codificación (CU) de acuerdo con un árbol cuaternario. En general, una estructura de datos de árbol cuaternario incluye un nodo por CU, con un nodo raíz correspondiente al bloque de árbol. Si una CU se divide en cuatro sub-CU, el nodo correspondiente a la CU incluye cuatro nodos hoja, de los que cada uno corresponde a una de las sub-CU.

Cada nodo de la estructura de datos de árbol cuaternario puede proporcionar datos de sintaxis para la CU correspondiente. Por ejemplo, un nodo en el árbol cuaternario puede incluir una bandera de división, que indica si la CU correspondiente al nodo está dividida en diversas sub-CU. Los elementos de sintaxis para una CU se pueden definir de forma recursiva y pueden depender de si la CU está dividida en diversas sub-CU. Si una CU no se divide adicionalmente, se denomina CU hoja. En esta divulgación, cuatro sub-CU de una CU hoja también se denominarán CU hoja, incluso si no existe ninguna división explícita de la CU hoja original. Por ejemplo, si una CU de tamaño 16x16 no se divide adicionalmente, las cuatro sub-CU de 8x8 también se denominarán CU hoja aunque la CU de 16x16 nunca se dividiera.

Una CU tiene un propósito similar a un macrobloque de la norma H.264, excepto en que una CU no tiene una distinción de tamaño. Por ejemplo, un bloque de árbol se puede dividir en cuatro nodos hijo (también denominados sub-CU) y cada nodo hijo puede ser, a su vez, un nodo padre y dividirse en otros cuatro nodos hijo. Un nodo hijo final, no dividido, denominado nodo hoja del árbol cuaternario, comprende un nodo de codificación, también denominado CU hoja. Los datos de sintaxis asociados a un flujo de bits codificado pueden definir el número máximo de veces en que se puede dividir un bloque de árbol, lo que se denomina profundidad máxima de CU, y también pueden definir un tamaño mínimo de los nodos de codificación. En consecuencia, un flujo de bits también puede definir la unidad de codificación más pequeña (SCU). Esta divulgación usa el término "bloque" para referirse a cualquiera de una CU, PU o TU, en el contexto de la HEVC, o a estructuras de datos similares en el contexto de otras normas (por ejemplo, macrobloques y subbloques de las mismas en H.264/AVC).

Una CU incluye un nodo de codificación y unidades de predicción (PU) y unidades de transformada (TU) asociadas al nodo de codificación. El tamaño de la CU corresponde al tamaño del nodo de codificación y debe tener conformación cuadrada. El tamaño de la CU puede variar desde 8x8 píxeles hasta el tamaño del bloque de árbol, con un máximo de 64x64 píxeles o más. Cada CU puede contener una o más PU y una o más TU. Los datos de sintaxis asociados a una CU pueden describir, por ejemplo, la división de la CU en una o más PU. Los modos de división pueden diferir entre si la CU está codificada en modo directo o por omisión, codificada en modo de intrapredicción o codificada en modo de interpredicción. Las PU se pueden dividir para tener una conformación no cuadrada. Los datos de sintaxis asociados a una CU también pueden describir, por ejemplo, la división de la CU en una o más TU de acuerdo con un árbol cuaternario. Una TU puede tener una conformación cuadrada o no cuadrada (por ejemplo, rectangular).

La norma HEVC permite transformaciones de acuerdo con las TU, que pueden ser diferentes para CU diferentes. El tamaño de las TU se basa típicamente en el tamaño de las PU dentro de una CU dada definida para una LCU dividida, aunque puede que este no sea siempre el caso. Las TU son típicamente del mismo tamaño o más pequeñas que las PU. En algunos ejemplos, las muestras residuales correspondientes a una CU se pueden subdividir en unidades más pequeñas usando una estructura de árbol cuaternario conocida como "árbol cuaternario residual" (RQT). Los nodos hoja del RQT se pueden denominar unidades de transformada (TU). Los valores de diferencias de píxeles asociados a las TU se pueden transformar para producir coeficientes de transformada, que se pueden cuantificar.

Una CU hoja puede incluir una o más unidades de predicción (PU). En general, una PU representa un área espacial correspondiente a la totalidad, o a una parte, de la CU correspondiente, y puede incluir datos para recuperar una muestra de referencia para la PU. Además, una PU incluye datos relacionados con la predicción. Por ejemplo, cuando la PU está codificada mediante intramodo se pueden incluir datos para la PU en un árbol cuaternario residual (RQT), que puede incluir datos que describen un modo de intrapredicción para una TU correspondiente a la PU. Como otro ejemplo, cuando la PU está codificada mediante intermodo, la PU puede incluir datos que definen uno o más vectores de movimiento para la PU. Los datos que definen el vector de movimiento para una PU pueden describir, por ejemplo, una componente horizontal del vector de movimiento, una componente vertical del vector de movimiento, una resolución para el vector de movimiento (por ejemplo, precisión de un cuarto de píxel o precisión de un octavo de píxel), una imagen de referencia a la que apunta el vector de movimiento y/o una lista de imágenes de referencia (por ejemplo, lista 0, lista 1 o lista C) para el vector de movimiento.



Una CU hoja que tiene una o más PU también puede incluir una o más unidades de transformada (TU). Las unidades de transformada se pueden especificar usando un RQT (también denominado estructura de árbol cuaternario de TU), como se analiza anteriormente. Por ejemplo, una bandera de división puede indicar si una CU hoja está dividida en cuatro unidades de transformada. A continuación, cada unidad de transformada se puede dividir adicionalmente en sub-TU adicionales. Cuando una TU no se divide adicionalmente, se puede denominar TU hoja. En general, en la intracodificación, todas las TU hoja que pertenecen a una CU hoja comparten el mismo modo de intrapredicción. Es decir, el mismo modo de intrapredicción se aplica, en general, para calcular valores predichos para todas las TU de una CU hoja. En la intracodificación, un codificador de vídeo puede calcular un valor residual para cada TU hoja usando el modo de intrapredicción, como una diferencia entre la parte de la CU correspondiente a la TU y el bloque original. Una TU no se limita necesariamente al tamaño de una PU. Por tanto, las TU pueden ser más grandes o más pequeñas que una PU. En la intracodificación, una PU puede estar ubicada junto con una TU hoja correspondiente para la misma CU. En algunos ejemplos, el tamaño máximo de una TU hoja puede corresponder al tamaño de la CU hoja correspondiente.

Además, las TU de las CU hoja también se pueden asociar a estructuras de datos de árbol cuaternario respectivas, denominadas árboles cuaternarios residuales (RQT). Es decir, una CU hoja puede incluir un árbol cuaternario que indica cómo se divide la CU hoja en las TU. El nodo raíz de un árbol cuaternario de TU corresponde, en general, a una CU hoja, mientras que el nodo raíz de un árbol cuaternario de CU corresponde, en general, a un bloque de árbol (o LCU). Las TU del RQT que no están divididas se denominan TU hoja. En general, esta divulgación usa los términos CU y TU para hacer referencia a una CU hoja y a una TU hoja, respectivamente, a menos que se indique de otro modo.

Una secuencia de vídeo incluye típicamente una serie de tramas o imágenes de vídeo. Un grupo de imágenes (GOP) comprende, en general, una serie de una o más de las imágenes de vídeo. Un GOP puede incluir datos de sintaxis en una cabecera del GOP, una cabecera de una o más de las imágenes, u otras ubicaciones, que describen un número de imágenes incluidas en el GOP. Cada fragmento de una imagen puede incluir datos de sintaxis de fragmento que describen un modo de codificación para el fragmento respectivo. El codificador de vídeo 20 funciona típicamente en bloques de vídeo dentro de fragmentos de vídeo individuales para codificar los datos de vídeo. Un bloque de vídeo puede corresponder a un nodo de codificación dentro de una CU. Los bloques de vídeo pueden tener tamaños fijos o variables y pueden diferir en tamaño de acuerdo con una norma de codificación especificada.

En un ejemplo, el HM admite la predicción en diversos tamaños de PU. Suponiendo que el tamaño de una CU particular es  $2N \times 2N$ , el HM admite la intrapredicción en tamaños de PU de  $2N \times 2N$  o  $N \times N$ , y la interpredicción en tamaños de PU simétricos de  $2N \times 2N$ ,  $2N \times N$ ,  $N \times 2N$  o  $N \times N$ . El HM también admite la división asimétrica para la interpredicción en tamaños de PU de  $2N \times nU$ ,  $2N \times nD$ ,  $nL \times 2N$  y  $nR \times 2N$ . En la división asimétrica, una dirección de una CU no está dividida, mientras que la otra dirección está dividida en un 25 % y un 75 %. La parte de la CU correspondiente a la partición del 25 % está indicada por una "n" seguida de una indicación de "arriba", "abajo", "izquierda" o "derecha". Por tanto, por ejemplo, " $2N \times nU$ " se refiere a una CU de  $2N \times 2N$  que está dividida horizontalmente, con una PU de  $2N \times 0,5N$  encima y una PU de  $2N \times 1,5N$  debajo.

En esta divulgación, " $N \times N$ " y "N por N" se puede usar de manera intercambiable para hacer referencia a las dimensiones de píxel de un bloque de vídeo en lo que respecta a dimensiones verticales y horizontales, por ejemplo,  $16 \times 16$  píxeles o 16 por 16 píxeles. En general, un bloque de  $16 \times 16$  tendrá 16 píxeles en una dirección vertical ( $y = 16$ ) y 16 píxeles en una dirección horizontal ( $x = 16$ ). Asimismo, un bloque de  $N \times N$  tiene, en general, N píxeles en una dirección vertical y N píxeles en una dirección horizontal, donde N representa un valor entero no negativo. Los píxeles de un bloque se pueden disponer en filas y columnas. Además, no es necesario que los bloques tengan necesariamente el mismo número de píxeles en la dirección horizontal que en la dirección vertical. Por ejemplo, los bloques pueden comprender  $N \times M$  píxeles, donde M no es necesariamente igual a N.

Después de la codificación de intrapredicción o interpredicción usando las PU de una CU, el codificador de vídeo 20 puede calcular datos residuales para las TU de la CU. Las PU pueden comprender datos de sintaxis que describen un procedimiento o modo de generación de datos de píxel predictivos en el dominio espacial (también denominado dominio de píxel) y las TU pueden comprender coeficientes en el dominio de transformada, después de la aplicación de una transformada, por ejemplo, una transformada de coseno discreta (DCT), una transformada de enteros, una transformada de ondícula o una transformada conceptualmente similar, a los datos de vídeo residuales. Los datos residuales pueden corresponder a diferencias de píxeles entre píxeles de la imagen no codificada y valores de predicción correspondientes a las PU. El codificador de vídeo 20 puede formar las TU, incluyendo los datos residuales para la CU, y, a continuación, transformar las TU para producir coeficientes de transformada para la CU.

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

Después de la cuantificación, el codificador de vídeo puede explorar los coeficientes de transformada, produciendo un vector unidimensional a partir de la matriz bidimensional que incluye los coeficientes de transformada cuantificados. La exploración se puede diseñar para colocar los coeficientes de mayor energía (y, por lo tanto, de menor frecuencia) en la parte delantera de la matriz y para colocar los coeficientes de menor energía (y, por lo tanto, de mayor frecuencia) en la parte trasera de la matriz. En algunos ejemplos, el codificador de vídeo 20 puede utilizar un orden de exploración predefinido para explorar los coeficientes de transformada cuantificados para producir un vector serializado que se puede codificar por entropía. En otros ejemplos, el codificador de vídeo 20 puede realizar una exploración adaptativa. Después de explorar los coeficientes de transformada cuantificados para formar un vector unidimensional, el codificador de vídeo 20 puede codificar por entropía el vector unidimensional, por ejemplo, de acuerdo con la codificación de longitud variable adaptativa al contexto (CAVLC), la codificación aritmética binaria adaptativa al contexto (CABAC), la codificación aritmética binaria adaptativa al contexto basada en sintaxis (SBAC), la codificación por entropía por división en intervalos de probabilidad (PIPE) o con otra metodología de codificación por entropía. El codificador de vídeo 20 también puede codificar por entropía los elementos de sintaxis asociados a los datos de vídeo codificados para su uso por el decodificador de vídeo 30 en la decodificación de los datos de vídeo.

Para realizar la CABAC, el codificador de vídeo 20 puede asignar un contexto, dentro de un modelo de contexto, a un símbolo que se va a transmitir. El contexto se puede referir, por ejemplo, a si los valores vecinos del símbolo son distintos de cero o no. Para realizar la CAVLC, el codificador de vídeo 20 puede seleccionar un código de longitud variable para un símbolo que se va a transmitir. Las palabras de código en la VLC se pueden construir de modo que los códigos relativamente más cortos correspondan a símbolos más probables, mientras que los códigos más largos correspondan a símbolos menos probables. De esta manera, el uso de la VLC puede conseguir un ahorro en bits con respecto, por ejemplo, al uso de palabras de código de igual longitud para cada símbolo que se va a transmitir. La determinación de la probabilidad se puede basar en un contexto asignado al símbolo.

Además, el codificador de vídeo 20 puede enviar además datos de sintaxis, tales como datos de sintaxis basados en bloques, datos de sintaxis basados en tramas y datos de sintaxis basados en GOP, al decodificador de vídeo 30, por ejemplo, en una cabecera de trama, una cabecera de bloque, una cabecera de fragmento o una cabecera de GOP. Los datos de sintaxis de GOP pueden describir un número de tramas en el GOP respectivo, y los datos de sintaxis de trama pueden indicar un modo de codificación/predicción usado para codificar la trama correspondiente.

La figura 2 es un diagrama de bloques que ilustra un ejemplo de un codificador de vídeo que puede implementar técnicas de acuerdo con aspectos descritos en esta divulgación. El codificador de vídeo 20 puede configurarse para realizar algunas de, o todas, las técnicas de esta divulgación. Como ejemplo, la unidad de selección de modo 40 se puede configurar para realizar cualquiera o todas las técnicas descritas en esta divulgación. Sin embargo, los aspectos de esta divulgación no se limitan a esto. En algunos ejemplos, las técnicas descritas en esta divulgación se pueden compartir entre los diversos componentes del codificador de vídeo 20. En algunos ejemplos, de manera adicional o alternativa, un procesador (no mostrado) puede configurarse para realizar cualquiera de, o todas, las técnicas descritas en esta divulgación.

En algunos modos de realización, la unidad de selección de modo 40, la unidad de estimación de movimiento 42, la unidad de compensación de movimiento 44, la unidad de intrapredicción 46 (u otro componente de la unidad de selección de modo 40, mostrada o no mostrada), u otro componente del codificador 20 (mostrado o no mostrado) pueden realizar las técnicas de esta divulgación. Por ejemplo, la unidad de selección de modo 40 puede recibir datos de vídeo que codificar, que pueden codificarse en una capa base y una o más capas de mejora correspondientes. La unidad de selección de modo 40, la unidad de estimación de movimiento 42, la unidad de compensación de movimiento 44, la unidad de intrapredicción 46 u otra unidad apropiada del codificador 20 pueden determinar un valor de una unidad de vídeo basándose, al menos en parte, en un valor de predicción y un valor de predicción residual ajustado asociado a una capa de referencia. El valor de predicción residual ajustado puede ser igual a una predicción residual a partir de la capa de referencia multiplicada por un factor de ponderación diferente de 1. El codificador 20 puede codificar la unidad de vídeo y señalar el factor de ponderación o la información de ponderación en un flujo de bits.

El codificador de vídeo 20 puede realizar la intracodificación y la 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 de un vídeo dentro de una trama o imagen de vídeo dada. La intercodificación se basa en la predicción temporal para reducir o eliminar la redundancia temporal de un vídeo dentro de tramas o imágenes contiguas de una secuencia de vídeo. Intramodo (modo I) se puede referir a cualquiera de varios modos de codificación de base espacial. Los intermodos, tales como la predicción unidireccional (modo P) o la predicción bidireccional (modo B), se pueden referir a cualquiera de varios modos de codificación de base temporal.

Como se muestra en la figura 2, el codificador de vídeo 20 recibe un bloque de vídeo actual dentro de una trama de vídeo que se va a codificar. En el ejemplo de la figura 1, el codificador de vídeo 20 incluye una unidad de selección de modo 40, una memoria de tramas 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 unidad de selección de modo 40 incluye una unidad de compensación de movimiento 44, una unidad de estimación de movimiento 42, una unidad de intrapredicción 46 y una unidad de división 48. Para la reconstrucción de bloques de vídeo, el codificador

de vídeo 20 incluye también una unidad de cuantificación inversa 58, una unidad de transformada inversa 60 y un sumador 62. También se puede incluir un filtro de eliminación de pixelado (no mostrado en la figura 2) para filtrar los límites de bloque y eliminar distorsiones de efecto pixelado del vídeo reconstruido. Si se desea, el filtro de eliminación de pixelado filtrará típicamente la salida del sumador 62. También se pueden usar filtros adicionales (en bucle o tras un bucle), además del filtro de eliminación de pixelado. Dichos filtros no se muestran por razones de brevedad, pero si se desea, pueden filtrar la salida del sumador 50 (como un filtro en bucle).

Durante el proceso de codificación, el codificador de vídeo 20 recibe una trama o un fragmento de vídeo que se va a codificar. La trama o el fragmento se pueden dividir en múltiples bloques de vídeo. La unidad de estimación de movimiento 42 y la unidad de compensación de movimiento 44 realizan la codificación interpredictiva del bloque de vídeo recibido en relación con uno o más bloques en una o más tramas de referencia para proporcionar predicción temporal. La unidad de intrapredicción 46, de forma alternativa, puede llevar a cabo la codificación intrapredictiva del bloque de vídeo recibido con respecto a uno o más bloques vecinos en la misma trama o fragmento que el bloque que se va a codificar para proporcionar predicción espacial. El codificador de vídeo 20 puede realizar múltiples pasadas de codificación, por ejemplo, para seleccionar un modo de codificación adecuado para cada bloque de datos de vídeo.

Además, la unidad de división 48 puede dividir bloques de datos de vídeo en subbloques, basándose en la evaluación de esquemas de división previos en las pasadas de codificación previas. Por ejemplo, la unidad de división 48 puede dividir inicialmente una trama o un fragmento en varias LCU, y dividir cada una de las LCU en varias sub-CU, basándose en un análisis de velocidad-distorsión (por ejemplo, una optimización de velocidad-distorsión). La unidad de selección de modo 40 puede producir además una estructura de datos de árbol cuaternario, indicativa de la división de una LCU en varias sub-CU. Las CU de nodos hoja del árbol cuaternario pueden incluir una o más PU y una o más TU.

La unidad de selección de modo 40 puede seleccionar uno de los modos de codificación, intra o inter, por ejemplo, basándose en los resultados de errores, y proporciona el bloque intracodificado o intercodificado resultante al sumador 50 para generar datos de bloques residuales, y al sumador 62 para reconstruir el bloque codificado para su uso como una trama de referencia. La unidad de selección de modo 40 también proporciona elementos de sintaxis, tales como vectores de movimiento, indicadores de intramodo, información de división y otra información de sintaxis de este tipo a la unidad de codificación por entropía 56.

La unidad de estimación de movimiento 42 y la unidad de compensación de movimiento 44 pueden estar altamente integradas, pero se ilustran por separado con propósitos 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 para los bloques de vídeo. Un vector de movimiento, por ejemplo, puede indicar el desplazamiento de una PU 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). Un bloque predictivo es un bloque que se considera estrechamente coincidente con el bloque que se va a codificar, en lo que respecta a la diferencia de píxeles, lo cual se puede determinar mediante una suma de diferencias absolutas (SAD), una suma de diferencias al cuadrado (SSD) u otras métricas de diferencia. En algunos ejemplos, el codificador de vídeo 20 puede calcular valores para posiciones fraccionarias de píxeles de imágenes de referencia almacenadas en la memoria de tramas de referencia 64. Por ejemplo, el codificador de vídeo 20 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 relativa a las posiciones de píxel completo y a las posiciones de píxel fraccionario, y proporcionar un vector de movimiento con una precisión de píxel fraccionario.

La unidad de estimación de movimiento 42 calcula un vector de movimiento para una PU de un bloque de vídeo en un fragmento intercodificado comparando la posición de la PU con la posición de un bloque predictivo de una imagen de referencia. La imagen de referencia puede seleccionarse a partir 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 de tramas de referencia 64. 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.

La compensación de movimiento, realizada por la unidad de compensación de movimiento 44, puede implicar buscar o generar el bloque predictivo en base al vector de movimiento determinado por la unidad de estimación de movimiento 42. De nuevo, la unidad de estimación de movimiento 42 y la unidad de compensación de movimiento 44 se pueden integrar funcionalmente, en algunos ejemplos. Tras recibir el vector de movimiento para la PU del 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 valores de píxel del bloque predictivo a los valores de píxel del bloque de vídeo actual que se está codificando, formando valores de diferencias de píxel, como se analiza a continuación. En general, la unidad de estimación de movimiento 42 realiza la estimación de movimiento en relación con las componentes de luma, y la unidad de compensación de movimiento 44 usa vectores de movimiento calculados en base a las componentes de luma tanto para componentes de croma como para componentes de luma. La unidad de selección de modo 40 también puede

generar elementos de sintaxis asociados a los bloques de vídeo y al fragmento de vídeo para su uso por el decodificador de vídeo 30 en la decodificación de los bloques de vídeo del fragmento de vídeo.

La unidad de intrapredicción 46 puede intrapredicir un bloque actual, como una alternativa a la interpredicción realizada por la unidad de estimación de movimiento 42 y la unidad de compensación de movimiento 44, como se describe anteriormente. En particular, la unidad de intrapredicción 46 puede determinar un modo de intrapredicción a usar para codificar un bloque actual. En algunos ejemplos, la unidad de intrapredicción 46 puede codificar un bloque actual usando diversos modos de intrapredicción, por ejemplo, durante pasadas de codificación separadas, y la unidad de intrapredicción 46 (o unidad de selección de modo 40, en algunos ejemplos) puede seleccionar un modo de intrapredicción apropiado a usar a partir de los modos sometidos a prueba.

Por ejemplo, la unidad de intrapredicción 46 puede calcular valores de velocidad-distorsión usando un análisis de velocidad-distorsión para los diversos modos de intrapredicción sometidos a prueba, y seleccionar el modo de intrapredicción que tenga las mejores características de velocidad-distorsión entre los modos sometidos a prueba. El análisis de velocidad-distorsión determina, en general, una cantidad de distorsión (o de error) entre un bloque codificado y un bloque original, no codificado, que se codificó para producir el bloque codificado, así como una velocidad de transferencia de bits (es decir, un número de bits) usada para producir el bloque codificado. La unidad de intrapredicción 46 puede calcular relaciones a partir de las distorsiones y velocidades para los diversos bloques codificados con el fin de determinar qué modo de intrapredicción presenta el mejor valor de velocidad-distorsión para el bloque.

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

El codificador de vídeo 20 forma un bloque de vídeo residual restando los datos de predicción de la unidad de selección de modo 40 al 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 transformada 52 aplica una transformada, tal como una transformada de coseno discreta (DCT) o una transformada conceptualmente similar, al bloque residual, produciendo un bloque de vídeo que comprende valores de coeficientes de transformada residuales. La unidad de procesamiento de transformada 52 puede realizar otras transformadas que son conceptualmente similares a la DCT. También se podrían usar transformadas de ondículas, transformadas de enteros, transformadas de subbanda u otros tipos de transformadas. En cualquier caso, la unidad de procesamiento de transformada 52 aplica la transformada al bloque residual, produciendo un bloque de coeficientes de transformada residuales. La transformada puede convertir la información residual desde un dominio de valores de píxel a un dominio de transformada, tal como un dominio de frecuencia. La unidad de procesamiento de transformada 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 además la velocidad de transferencia de bits. El proceso de cuantificación puede reducir la profundidad de bits asociada a algunos, o la totalidad, de los coeficientes. El grado de cuantificación se puede modificar ajustando un parámetro de cuantificación. En algunos ejemplos, la unidad de cuantificación 54 puede realizar, a continuación, una exploración de la matriz que incluye los coeficientes de transformada cuantificados. De forma alternativa, la unidad de codificación por entropía 56 puede realizar la exploración.

Después de 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 adaptativa al contexto (CAVLC), la codificación aritmética binaria adaptativa al contexto (CABAC), la codificación aritmética binaria adaptativa al contexto basada en sintaxis (SBAC), la codificación por entropía por división en intervalos de probabilidad (PIPE) u otra técnica de codificación por entropía. En el caso de la codificación por entropía basada en contexto, el contexto se puede basar en bloques vecinos. Tras la codificación por entropía por parte de la unidad de codificación por entropía 56, el flujo de bits codificado se puede transmitir a otro dispositivo (por ejemplo, el decodificador de vídeo 30) o archivar para su posterior transmisión o recuperación.

La unidad de cuantificación inversa 58 y la unidad de transformada inversa 60 aplican la cuantificación inversa y la transformación inversa, respectivamente, para reconstruir el bloque residual en el dominio de 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 tramas de referencia 64. La unidad de compensación de movimiento 44 también puede aplicar uno o más filtros de interpolación al bloque residual reconstruido para calcular valores de píxeles fraccionarios, para su uso en la estimación de movimiento. 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 tramas de referencia 64. El bloque de vídeo reconstruido se puede usar por la unidad de estimación de movimiento 42 y la unidad de compensación de movimiento 44 como un bloque de referencia para intercodificar un bloque en una trama de vídeo subsiguiente.

La figura 3 es un diagrama de bloques que ilustra un ejemplo de un decodificador de vídeo que puede implementar técnicas de acuerdo con aspectos descritos en esta divulgación. El decodificador de vídeo 30 puede configurarse para realizar algunas de, o todas, las técnicas de esta divulgación. Como ejemplo, la unidad de compensación de movimiento 72 y/o la unidad de intrapredicción 74 pueden estar configuradas para realizar cualquiera o la totalidad de las técnicas descritas en esta divulgación. Sin embargo, los aspectos de esta divulgación no se limitan a esto. En algunos ejemplos, las técnicas descritas en esta divulgación se pueden compartir entre los diversos componentes del decodificador de vídeo 30. En algunos ejemplos, de manera adicional o alternativa, un procesador (no mostrado) puede configurarse para realizar cualquiera de, o todas, las técnicas descritas en esta divulgación.

En algunos modos de realización, la unidad de decodificación por entropía 70, la unidad de compensación de movimiento 72, la unidad de intrapredicción 74 u otro componente del decodificador 30 (mostrado o no mostrado) pueden realizar las técnicas de esta divulgación. Por ejemplo, la unidad de decodificación por entropía 70 puede recibir un flujo de bits de vídeo codificado, que puede codificar datos relacionados con una capa base y una o más capas de mejora correspondientes. La unidad de compensación de movimiento 72, la unidad de intrapredicción 74 u otra unidad apropiada del decodificador 30 pueden determinar un valor de una unidad de vídeo basándose, al menos en parte, en un valor de predicción y un valor de predicción residual ajustado asociado a una capa de referencia. El valor de predicción residual ajustado puede ser igual a una predicción residual a partir de la capa de referencia multiplicada por un factor de ponderación diferente de 1. El decodificador 30 puede decodificar la unidad de vídeo y recibir el factor de ponderación o la información de ponderación en un flujo de bits.

En el ejemplo de la figura 3, el decodificador de vídeo 30 incluye una unidad de decodificación por entropía 70, una unidad de compensación de movimiento 72, una unidad de intrapredicción 74, una unidad de cuantificación inversa 76, una unidad de transformación inversa 78, una memoria de tramas de referencia 82 y un sumador 80. En algunos ejemplos, el decodificador de vídeo 30 puede realizar una pasada de decodificación generalmente recíproca a la pasada de codificación descrita con respecto al codificador de vídeo 20 (figura 2). La unidad de compensación de movimiento 72 puede generar datos de predicción basándose en vectores de movimiento recibidos desde la unidad de decodificación por entropía 70, mientras que la unidad de intrapredicción 74 puede generar datos de predicción basándose en indicadores de modo de intrapredicción recibidos desde la unidad de decodificación por entropía 70.

Durante el proceso de decodificación, el decodificador de vídeo 30 recibe un flujo de bits de vídeo codificado que representa bloques de vídeo de un fragmento de vídeo codificado y elementos de sintaxis asociados desde el codificador de vídeo 20. La unidad de decodificación por entropía 70 del decodificador de vídeo 30 decodifica por entropía el flujo de bits para generar coeficientes cuantificados, vectores de movimiento o indicadores de modo de intrapredicción y otros elementos de sintaxis. La unidad de decodificación por entropía 70 reenvía los vectores de movimiento y otros elementos de sintaxis a la unidad de compensación de movimiento 72. El decodificador de vídeo 30 puede recibir los elementos de sintaxis a nivel de fragmento de vídeo y/o a nivel de bloque de vídeo.

Cuando el fragmento de vídeo se codifica como un fragmento intracodificado (I), la unidad de intrapredicción 74 puede generar datos de predicción para un bloque de vídeo del fragmento de vídeo actual basándose en un modo de intrapredicción señalizado y en datos de bloques previamente decodificados 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 de sintaxis recibidos desde la unidad de decodificación por entropía 70. Los bloques predictivos se pueden generar a partir de una de las imágenes de referencia dentro de una de las listas de imágenes de referencia. El decodificador de vídeo 30 puede construir las listas de tramas de referencia, la Lista 0 y la Lista 1, usando técnicas de construcción por defecto, basándose en imágenes de referencia almacenadas en la memoria de tramas de referencia 92. La unidad de compensación de movimiento 72 determina información de predicción para un bloque de vídeo del fragmento de vídeo actual analizando sintácticamente los vectores de movimiento y otros elementos de sintaxis, y usa la información de predicción para producir los bloques predictivos para el bloque de vídeo actual que se esté decodificando. Por ejemplo, la unidad de compensación de movimiento 72 usa algunos de los elementos de sintaxis recibidos para determinar un modo de predicción (por ejemplo, de intra- o interpredicción) usado para codificar los bloques de vídeo del fragmento de vídeo, un tipo de fragmento de interpredicción (por ejemplo, un fragmento B, un fragmento P o un fragmento GPB), información de construcción para una o más de las listas de imágenes de referencia para el fragmento, vectores de movimiento para cada bloque de vídeo intercodificado del fragmento, un estado de interpredicción para cada bloque de vídeo intercodificado del fragmento y otra información para decodificar los bloques de vídeo del fragmento de vídeo actual.

La unidad de compensación de movimiento 72 también puede realizar 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 usados por el codificador de vídeo 20 a partir de los elementos de sintaxis recibidos y usar los filtros de interpolación para producir bloques predictivos.

La unidad de cuantificación inversa 76 cuantifica de manera inversa, por ejemplo, descuantifica, los coeficientes de transformada cuantificados, proporcionados en el flujo de bits y decodificados por la unidad de decodificación por entropía 80. El proceso de cuantificación inversa puede incluir el uso de un parámetro de cuantificación QPy, calculado por el decodificador 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.

La unidad de transformada inversa 78 aplica una transformada inversa, por ejemplo, una DCT inversa, una transformada de enteros inversa, o un proceso de transformada inversa conceptualmente similar, a los coeficientes de transformada para producir bloques residuales en el dominio de píxel.

Después de que la unidad de compensación de movimiento 82 genere el bloque predictivo para el bloque de vídeo actual en base a los vectores de movimiento y otros elementos de sintaxis, el decodificador de vídeo 30 forma un bloque de vídeo decodificado sumando los bloques residuales de la unidad de transformada inversa 78 a los bloques predictivos correspondientes generados por la unidad de compensación de movimiento 72. El sumador 90 representa el componente o los componentes que realizan esta operación de suma. Si se desea, también se puede aplicar un filtro de eliminación de pixelado para filtrar los bloques decodificados para eliminar distorsiones de pixelado. También se pueden usar otros filtros de bucle (bien en el bucle de codificación o bien después del bucle de codificación) para suavizar las transiciones entre píxeles o mejorar de otro modo la calidad de vídeo. Los bloques de vídeo decodificados en una trama o imagen dada se almacenan a continuación en la memoria de imágenes de referencia 92, que almacena imágenes de referencia usadas para la compensación de movimiento posterior. La memoria de tramas de referencia 82 almacena también vídeo decodificado para su presentación posterior en un dispositivo de visualización, tal como el dispositivo de visualización 32 de la figura 1.

La figura 4 es un diagrama de flujo que ilustra un procedimiento de ejemplo para una predicción residual generalizada de acuerdo con aspectos de esta divulgación. Las técnicas descritas en esta divulgación pueden proporcionar un marco de predicción residual generalizada (GRP). Como se explicó anteriormente, la predicción residual entre capas usa el residuo de la capa de referencia para predecir la unidad de vídeo actual, por ejemplo, una trama. En la predicción residual generalizada, la predicción residual entre capas de la unidad de vídeo actual puede basarse en el residuo de la capa actual, la predicción temporal de la capa actual y el residuo de la capa de referencia. El residuo de la capa de referencia puede ajustarse mediante un factor de ponderación. El esquema GRP puede definirse como sigue:

$$\hat{I}_c = r_c + P_c + w \cdot r_r \quad (3)$$

donde  $\hat{I}_c$  denota la reconstrucción de la trama actual,  $r_c$  denota la predicción residual a partir de la capa actual,  $P_c$  denota la predicción temporal a partir de la misma capa,  $r_r$  denota la predicción residual a partir de la capa de referencia y  $w$  denota un factor de ponderación. El marco GRP también puede aplicarse a las intra-CU. En tal caso,  $P_e$  indica la predicción espacial a partir de la misma capa.

El factor de ponderación puede basarse en e incluir diversos tipos de información. Ejemplos de dicha información pueden incluir un número de candidatos de ponderación, una etapa de ponderación, un índice de ponderación y una tabla de ponderación. El número de candidatos de ponderación puede indicar el número de diferentes factores de ponderación que están disponibles para aplicarse al residuo de la capa de referencia. La etapa de ponderación puede indicar el tamaño del incremento o la unidad entre los factores de ponderación disponibles. El índice de ponderación puede indicar un factor de ponderación particular entre los factores de ponderación disponibles. La tabla de ponderación puede incluir información acerca del factor de ponderación y se puede acceder a ella mediante el índice de ponderación, de manera similar a una tabla de consulta. En un ejemplo específico, pueden estar disponibles tres candidatos a factor de ponderación: 0,0, 0,5 y 1,0. En este ejemplo, el número de candidatos de ponderación es 3 porque hay disponibles tres candidatos a factor de ponderación. La etapa de ponderación entre los 3 candidatos de ponderación es 0,5. Cada candidato de ponderación puede identificarse mediante un índice de ponderación. El factor de ponderación 0 se identifica con el índice 0, el factor de ponderación 0,5 con el índice 1 y el factor de ponderación 1,0 con el índice 2. La etapa de ponderación y el índice se pueden usar para obtener el factor de ponderación ya que las fracciones de señalización pueden ser costosas.

El marco de GRP de acuerdo con los aspectos de esta divulgación puede permitir diversos tipos de predicción residual mediante la incorporación de un factor de ponderación. El ajuste apropiado del factor de ponderación puede dar lugar a ganancias de codificación significativas para la predicción residual. La GRP puede mejorar el rendimiento de codificación al tiempo que reduce la cantidad de memoria y el coste computacional al incorporar información de ponderación para la capa de referencia en la predicción residual. Por ejemplo, la GRP puede mejorar el rendimiento de codificación ya que la predicción residual ponderada es más precisa. Además, la cantidad de memoria y el coste computacional se pueden reducir, por ejemplo, ya que dos conjuntos de bucles de compensación de movimiento no se usan típicamente como compensación de movimiento en el dominio de diferencia. Además, en el marco de GRP, la predicción residual se puede realizar usando una capa de referencia que no es necesariamente la capa base en la

predicción residual tradicional. Por ejemplo, la capa de referencia puede obtenerse a partir de la capa de mejora de la capa actual. La GRP también puede adaptarse a la predicción residual tradicional cuando el factor de ponderación se establece en 1. El marco de GRP se puede utilizar con decodificación de bucle único y decodificación multibucle.

Con respecto a la compensación de movimiento en el dominio de diferencia, el marco GRP puede aplicarse en la decodificación de bucle único. Como se explicó anteriormente, en H.264, la compensación de movimiento en el dominio de diferencia no puede emplearse en un escenario de decodificación de bucle único, ya que las imágenes diferenciales entre capas han de calcularse en base a la imagen completamente reconstruida de cada capa. Para obtener la imagen de diferencia en la compensación de movimiento en el dominio de diferencia, a menudo se usa la reconstrucción completa de cada capa, y para cada capa se puede usar un bucle de compensación de movimiento para la reconstrucción completa. Por ejemplo, a menudo se usan dos bucles de compensación de movimiento para tener una reconstrucción completa de dos capas. Por consiguiente, la compensación de movimiento en el dominio de diferencia no puede emplearse en la decodificación de bucle único. Por el contrario, la GRP puede admitir tanto decodificación de bucle único como decodificación multibucle. Además, en el marco de GRP, la estimación de movimiento en el dominio de diferencia puede no ser necesaria. Por lo tanto, es posible que la capa actual y la capa de mejora no tengan que compartir el mismo movimiento para la estimación de movimiento. El marco de GRP se puede aplicar a muchos tipos diferentes de predicción residual, y la predicción residual tradicional como se define en la ecuación (1) y la compensación de movimiento en el dominio de diferencia son dos escenarios específicos del uso del marco de GRP.

El procedimiento de ejemplo para la predicción residual generalizada de acuerdo con aspectos de esta divulgación se explicará a continuación con referencia a la figura 4. El proceso 400 puede realizarse mediante un codificador (por ejemplo, el codificador mostrado en la figura 2, etc.), un decodificador (por ejemplo, el decodificador mostrado en la figura 3, etc.) o cualquier otro componente. Los bloques del proceso 400 se describen con respecto al decodificador 30 en la figura 3, pero el proceso 400 puede ser realizado por otros componentes, tal como un codificador, mencionados anteriormente.

En el bloque 401, el decodificador 30 aplica un factor de ponderación a la predicción residual a partir de la capa de referencia. Como se explicó anteriormente, la predicción residual generalizada (GRP) puede aplicar un factor de ponderación al residuo de la capa de referencia. Se puede determinar que el factor de ponderación es óptimo para un escenario particular, tal como la decodificación de bucle único. El factor de ponderación puede incluir información tal como el número de candidatos de ponderación, la etapa de ponderación, el índice de ponderación y la tabla de ponderación.

En el bloque 402, el decodificador 30 obtiene la predicción residual a partir de la capa de mejora. En el bloque 403, el decodificador 30 obtiene la predicción temporal a partir de la capa de mejora.

En el bloque 404, el decodificador 30 determina la unidad de vídeo actual basándose en la predicción residual a partir de la capa de referencia ajustada por el factor de ponderación, la predicción residual a partir de la capa de mejora y la predicción temporal a partir de la capa de mejora. Como se explicó anteriormente, en GRP, la unidad de vídeo actual se puede predecir de acuerdo con la ecuación (3).

El procedimiento de ejemplo para la predicción residual generalizada de acuerdo con aspectos de esta divulgación descritos con respecto a la figura 4 puede implementarse en varios niveles de codificación, tales como una secuencia, una imagen, un grupo de tramas, una trama, un grupo de fragmentos, un fragmento, un grupo de unidades de codificación (CU), una unidad de codificación (CU), un grupo de unidades de predicción (PU), una unidad de predicción (PU), bloques, una región de píxeles o un píxel. Además, el procedimiento de ejemplo para la predicción residual generalizada de acuerdo con los aspectos de esta divulgación descritos con respecto a la figura 4 también puede aplicarse a componentes de señal, tales como componentes de luma y componentes de croma. Por ejemplo, el procedimiento de ejemplo puede aplicarse solo a componentes luma, solo a componentes de croma, o una combinación de componentes de luma y componentes de croma (por ejemplo, cualquier combinación de una componente de luma y dos componentes de croma). Además, todos los modos de realización descritos con respecto a la figura 4 pueden implementarse por separado o en combinación entre sí.

La figura 4A es un diagrama de flujo que ilustra otro procedimiento de ejemplo para una predicción residual generalizada de acuerdo con aspectos de esta divulgación. El proceso 400A puede realizarse mediante un codificador (por ejemplo, el codificador mostrado en la figura 2, etc.), un decodificador (por ejemplo, el decodificador mostrado en la figura 3, etc.) o cualquier otro componente. Los bloques del proceso 400A se describen con respecto al decodificador 30 en la figura 3, pero el proceso 400A puede ser realizado por otros componentes, tal como un codificador, mencionados anteriormente. El proceso 400A puede aplicarse tanto a SVC como a la codificación de vídeo 3D. Todos los modos de realización descritos con respecto a la figura 4A pueden implementarse por separado o en combinación entre sí.

En el bloque 401A, el decodificador 30 obtiene la predicción residual a partir de la capa de referencia. La capa de referencia puede ser una capa base para la capa actual. La capa de referencia también puede ser una capa de mejora que corresponda a la capa de mejora actual. La predicción residual a partir de la capa de referencia se puede

determinar de varias formas. Por ejemplo, el residuo de una unidad de vídeo ubicada conjuntamente en la capa de referencia se puede muestrear de manera ascendente para que coincida con la resolución de la capa actual.

En el bloque 402A, el decodificador 30 ajusta la predicción residual a partir de la capa de referencia mediante un factor de ponderación. El factor de ponderación puede ser diferente de 1. Si el factor de ponderación es 1, el decodificador 30 puede realizar la predicción residual tradicional. El factor de ponderación se puede determinar en base a información de ponderación, tal como una etapa de ponderación, una tabla de ponderación, un número de candidatos a factor de ponderación y un índice de ponderación. En algunos modos de realización, el factor de ponderación puede recibirse en un flujo de bits. En otros modos de realización, el factor de ponderación puede obtenerse, al menos parcialmente, de la información recibida en un flujo de bits o de información decodificada previamente. El factor de ponderación puede codificarse a cualquier nivel de codificación, incluidos, pero sin limitarse a, los siguientes niveles de sintaxis: una secuencia, un grupo de tramas, una trama, un grupo de fragmentos, un fragmento, un grupo de unidades de codificación (CU), una unidad de codificación (CU), un grupo de unidades de predicción (PU), una unidad de predicción (PU), un bloque, una región de píxeles, un píxel y cualquier combinación de una componente de luma y dos componentes de croma.

En algunos modos de realización, la información de ponderación también puede recibirse en un flujo de bits, o puede obtenerse al menos parcialmente de la información recibida en un flujo de bits o información decodificada previamente. La información de ponderación también puede codificarse a cualquier nivel de codificación, incluidos, pero sin limitarse a, los siguientes niveles de sintaxis: una secuencia, un grupo de tramas, una trama, un grupo de fragmentos, un fragmento, un grupo de unidades de codificación (CU), una unidad de codificación (CU), un grupo de unidades de predicción (PU), una unidad de predicción (PU), un bloque, una región de píxeles, un píxel y cualquier combinación de una componente de luma y dos componentes de croma.

En el bloque 403A, el decodificador 30 determina un valor de predicción. El valor de predicción puede ser el valor de predicción para la unidad de vídeo actual. En algunos modos de realización, el valor de predicción es una unidad de predicción para la unidad de vídeo actual. El valor de predicción puede ser un valor de predicción temporal o un valor de intrapredicción espacial. El proceso 400A se puede usar tanto con interpredicción como con intrapredicción.

En el bloque 404A, el decodificador 30 determina el valor de una unidad de vídeo basándose, al menos en parte, en el valor de predicción y la predicción residual ajustada asociada a la capa de referencia. Una unidad de vídeo puede ser cualquier unidad de datos de vídeo y puede incluir, pero sin limitarse a: una trama, un fragmento, una unidad de codificación más grande (LCU), una unidad de codificación (CU), un bloque, un píxel y un subpíxel. El valor de la unidad de vídeo se puede determinar generando una unidad de predicción (PU) para la unidad de vídeo.

En algunos modos de realización, el decodificador 30 puede determinar el valor de la unidad de vídeo actual en una capa basándose en la predicción residual para la unidad de vídeo actual, el valor de predicción para la unidad de vídeo actual y la predicción residual ajustada asociada a la capa de referencia. La predicción residual y el valor de predicción para la unidad de vídeo actual pueden ser de la capa actual. El valor de la unidad de vídeo actual puede determinarse sumando el valor de predicción para la unidad de vídeo actual, la predicción residual para la unidad de vídeo actual y la predicción residual ajustada asociada a la capa de referencia.

La figura 5 es un diagrama de flujo que ilustra un procedimiento de ejemplo para una predicción residual generalizada usando decodificación de bucle único de acuerdo con aspectos de esta divulgación. Como se explicó anteriormente, en la decodificación de bucle único, se usa un bucle para la compensación de movimiento de la capa de mejora. En el escenario de la decodificación de bucle único, no se dispone de una reconstrucción completa de la capa base. Por consiguiente, el residuo normalizado de la capa base puede emplearse directamente como predictor residual base. Para la capa de mejora, la reconstrucción  $\hat{I}_e$  puede determinarse como sigue:

$$\hat{I}_e = r_e + P_e + w \cdot r_b = r_e + P_e + w \cdot r_b' \cdot (Q_e / Q_b) \quad (4)$$

donde  $r_e$  y  $P_e$  denotan una predicción residual y temporal descuantificada de la capa de mejora,  $r_b$  denota el predictor residual de capa base normalizado (muestreado de manera ascendente en el caso escalable espacial),  $r_b'$  denota el residuo de capa base y  $Q_e$  y  $Q_b$  denotan la etapa de cuantificación de la capa de mejora y la capa base, respectivamente.

El procedimiento de ejemplo para la predicción residual generalizada que usa decodificación de bucle único de acuerdo con aspectos de esta divulgación se explicará a continuación con referencia a la figura 5. El proceso 500 puede realizarse mediante un codificador (por ejemplo, el codificador mostrado en la figura 2, etc.), un decodificador (por ejemplo, el decodificador mostrado en la figura 3, etc.) o cualquier otro componente. Los bloques del proceso 500 se describen con respecto al decodificador 30 en la figura 3, pero el proceso 500 puede ser realizado por otros componentes, tal como un codificador, mencionados anteriormente. En el bloque 501, el decodificador 30 determina un factor de ponderación para la predicción residual a partir de la capa de referencia en la decodificación de bucle único para el marco de GRP. En el bloque 502, el decodificador 30 determina la unidad de vídeo actual basándose en la predicción residual a partir de la RL ajustada por el factor de ponderación, la predicción residual a partir de la EL y



la predicción temporal a partir de la EL. Por ejemplo, como se explicó anteriormente con respecto a la ecuación (4), el residuo de capa base normalizado se puede usar para la predicción residual de RL. El procedimiento de ejemplo para la predicción residual generalizada que usa decodificación de bucle único de acuerdo con aspectos de esta divulgación descritos con respecto a la figura 5 puede implementarse en varios niveles de codificación, tales como una secuencia, una imagen, un grupo de tramas, una trama, un grupo de fragmentos, un fragmento, un grupo de unidades de codificación (CU), una unidad de codificación (CU), un grupo de unidades de predicción (PU), una unidad de predicción (PU), bloques, una región de píxeles o un píxel. Además, el procedimiento de ejemplo para la predicción residual generalizada que usa decodificación de bucle único de acuerdo con los aspectos de esta divulgación descritos con respecto a la figura 5 también puede aplicarse a componentes de señal, tales como componentes de luma y componentes de croma. Por ejemplo, el procedimiento de ejemplo puede aplicarse solo a componentes luma, solo a componentes de croma, o una combinación de componentes de luma y componentes de croma (por ejemplo, cualquier combinación de una componente de luma y dos componentes de croma). Además, todos los modos de realización descritos con respecto a la figura 5 pueden implementarse por separado o en combinación entre sí.

La figura 6 es un diagrama de flujo que ilustra un procedimiento de ejemplo para una predicción residual generalizada usando decodificación multibucle de acuerdo con aspectos de esta divulgación. Como se explicó anteriormente, en la decodificación multibucle, se usan múltiples bucles para la compensación de movimiento de la capa de mejora. En el escenario de la decodificación multibucle, la reconstrucción completa de la capa base está disponible cuando se codifica/decodifica la capa de mejora. En consecuencia, el valor diferencial entre la reconstrucción de la capa de mejora previamente codificada y la capa base (muestreada de manera ascendente si es necesario) puede emplearse como predictor residual. Para la capa de mejora, la reconstrucción  $\hat{I}_e$  puede determinarse como sigue:

$$\hat{I}_e = r_e + P_e + w \cdot (\hat{I}_b - P_b) \quad (5)$$

donde  $r_e$  indica el residuo descuantificado de la unidad de vídeo actual en la capa de mejora,  $P_e$  y  $P_b$  indican la predicción temporal para la unidad de vídeo actual en la capa de mejora y la capa base, respectivamente, y  $\hat{I}_b$  indica la reconstrucción completa de la unidad de vídeo actual en la capa base. Debido a que la capa de mejora y la capa base pueden tener un objetivo de calidad diferente, el movimiento de las predicciones temporales  $P_e$  y  $P_b$  puede ser diferente.

Si la capa base y la capa de mejora tienen el mismo movimiento, el movimiento de las predicciones temporales  $P_e$  y  $P_b$  son iguales, y la ecuación (5) puede emplearse directamente. Cuando se decodifica una interunidad de vídeo de la capa de mejora, su capa de mejora y las predicciones temporales de capa base  $P_e$  y  $P_b$  están disponibles. La reconstrucción de capa base  $\hat{I}_b$  también está disponible. En consecuencia, la reconstrucción  $\hat{I}_e$  puede obtenerse del residuo descuantificado  $r_e$  y  $w$ , que puede señalizarse u obtenerse como se explica en más detalle con respecto a las figuras 7 y 8. De forma similar, cuando se aplica el marco de GRP a las intra-CU,  $P_e$  y  $P_b$  indican la intrapredicción espacial para la unidad de vídeo actual en la capa de mejora y la capa base, respectivamente. Además,  $P_e$  y  $P_b$  pueden obtenerse aplicando la misma dirección de intrapredicción, que es similar al caso de interpredicción en que se puede suponer el mismo movimiento de  $P_e$  y  $P_b$ .

Si la capa base y la capa de mejora tienen un movimiento diferente, el movimiento de la capa de mejora y las predicciones temporales de capa base  $P_e$  y  $P_b$  es diferente, y el residuo de la capa base y el residuo de la capa de mejora pueden no estar correlacionados. En tal caso, la predicción residual puede no dar lugar a buenos resultados. Para mejorar el rendimiento de la predicción residual, se puede suponer que la capa de mejora y las predicciones temporales de capa base comparten el mismo movimiento. Además de o en lugar de suponer que las predicciones temporales de EL y BL comparten el mismo movimiento, puede requerirse que el movimiento de la capa base o el movimiento de la capa de mejora se aplique a otra capa para generar el predictor residual. Por ejemplo, el movimiento de la predicción temporal de capa de mejora  $P_e$  puede aplicarse a la capa base para obtener  $P_b$ . En tal caso, a menudo se usan dos compensaciones de movimiento para decodificar la capa de mejora, ya que tanto  $P_e$  como  $P_b$  pueden generarse con el movimiento de  $P_e$ .

El procedimiento de ejemplo para la predicción residual generalizada que usa decodificación multibucle de acuerdo con aspectos de esta divulgación se explicará a continuación con referencia a la figura 6. El proceso 600 puede realizarse mediante un codificador (por ejemplo, el codificador mostrado en la figura 2, etc.), un decodificador (por ejemplo, el decodificador mostrado en la figura 3, etc.) o cualquier otro componente. Los bloques del proceso 600 se describen con respecto al decodificador 30 en la figura 3, pero el proceso 600 puede ser realizado por otros componentes, tal como un codificador, mencionados anteriormente. En el bloque 601, el decodificador 30 determina un factor de ponderación para la predicción residual a partir de la capa de referencia en la decodificación multibucle para el marco de GRP. En el bloque 602, el decodificador 30 determina la unidad de vídeo actual basándose en la predicción residual a partir de la RL ajustada por el factor de ponderación, la predicción residual a partir de la EL y la predicción temporal a partir de la EL. Por ejemplo, como se explicó anteriormente con respecto a la ecuación (5),  $\hat{I}_b - P_b$  puede usarse para la predicción residual de RL. El procedimiento de ejemplo para la predicción residual generalizada que usa decodificación multibucle de acuerdo con aspectos de esta divulgación descritos con respecto a la figura 6 puede implementarse en varios niveles de codificación, tales como una secuencia, una imagen, un grupo de tramas, una trama, un grupo de fragmentos, un fragmento, un grupo de unidades de codificación (CU), una unidad de

codificación (CU), un grupo de unidades de predicción (PU), una unidad de predicción (PU), bloques, una región de píxeles o un píxel. Además, el procedimiento de ejemplo para la predicción residual generalizada que usa decodificación multibucle de acuerdo con los aspectos de esta divulgación descritos con respecto a la figura 6 también puede aplicarse a componentes de señal, tales como componentes de luma y componentes de croma. Por ejemplo, el procedimiento de ejemplo puede aplicarse solo a componentes luma, solo a componentes de croma, o una combinación de componentes de luma y componentes de croma (por ejemplo, cualquier combinación de una componente de luma y dos componentes de croma). Además, todos los modos de realización descritos con respecto a la figura 6 pueden implementarse por separado o en combinación entre sí.

La figura 7 es un diagrama de flujo que ilustra un procedimiento de ejemplo para señalar parámetros de predicción residual generalizada de acuerdo con aspectos de esta divulgación. Como se explicó anteriormente, la información de ponderación puede incluir el número de candidatos de ponderación, la etapa de ponderación (o la tabla de ponderación) y el índice de ponderación. El factor de ponderación  $w$  puede determinarse en base a dicha información de ponderación. El número de candidatos de ponderación puede denotarse como  $N_w$ . La etapa de ponderación puede denotarse como  $S_w$ , y la tabla de ponderación como  $W_T$ . El índice de ponderación puede denotarse como  $i_w$ . En un modo de realización, el factor de ponderación  $w$  se obtiene en base a la etapa de ponderación  $S_w$  y el índice de ponderación  $i_w$  como sigue:

$$w = S_w \cdot i_w \quad (6)$$

En otro modo de realización,  $w$  puede obtenerse de una tabla de consulta  $W_T$  de acuerdo con el índice  $i_w$ .

La información del factor de ponderación, que puede incluir, pero no se limita a,  $N_w$ ,  $S_w$ ,  $W_T$  e  $i_w$ , puede señalizarse de diversas formas. En algunos modos de realización, la etapa de ponderación  $S_w$  o la tabla de ponderación  $W_T$  puede estar codificada de forma fija o señalizada.  $S_w$  o  $W_T$  pueden señalizarse a nivel de secuencia o a nivel de imagen. El índice de ponderación  $i_w$  puede señalizarse en un nivel inferior, tal como CU y PU.

En un modo de realización, la etapa de ponderación  $S_w$  está representada con una cuantificación de 3 bits ( $S_w$  puede ser  $1/8$ ,  $2/8$ , ...,  $8/8$ ) y está codificada mediante Exp-Golomb de enteros sin signo en el Conjunto de Parámetros de Secuencia (SPS). Considerando que  $N_w \geq 1$ ,  $(N_w - 1)$  también está codificado mediante Exp-Golomb de enteros sin signo en SPS. En primer lugar, el índice de ponderación  $i_w$  se binariza con código unario truncado (donde  $N_w$  es el número máximo) y a continuación se codifica mediante CABAC. En la codificación CABAC, el primer *bin* se codifica con un contexto y el resto de los *bins* se codifica con otro contexto. Para codificar el índice de ponderación  $i_w$ , el contexto puede depender de parámetros codificados previamente. Por ejemplo,  $i_w$  de CU espacialmente vecinas, tales como las CU izquierda y superior, se puede utilizar como contexto para el índice de ponderación  $i_w$  de la CU actual. Además, el tipo de la CU actual, tal como si la CU actual está codificada por omisión o fusión, o el tamaño de la CU actual, se puede utilizar como contexto.

En otros modos de realización, diferentes modos de CU pueden tener diferentes procedimientos de señalización de factores de ponderación. Por ejemplo, para los modos de omisión y fusión, se pueden seleccionar y señalar tres factores de ponderación (tales como  $w = 0$ ,  $w = 0,5$  y  $w = 1$ ). Para intermodos distintos de los modos de omisión y fusión, solo se pueden seleccionar y señalar dos factores de ponderación (tales como  $w = 0$  y  $w = 1$ ). De forma alternativa, para intermodos distintos de los modos de omisión y fusión, solo se puede aplicar un factor de ponderación fijo. En tal caso, no se puede utilizar ninguna señalización adicional para el factor de ponderación.

El procedimiento de ejemplo para señalar parámetros de predicción residual generalizada de acuerdo con aspectos de esta divulgación se explicará a continuación con referencia a la figura 7. El proceso 700 puede realizarse mediante un codificador (por ejemplo, el codificador mostrado en la figura 2, etc.), un decodificador (por ejemplo, el decodificador mostrado en la figura 3, etc.) o cualquier otro componente. Los bloques del proceso 700 se describen con respecto al codificador 20 en la figura 2, pero el proceso 700 puede ser realizado por otros componentes, tal como un decodificador, mencionados anteriormente. En el bloque 701, el codificador 20 señala la etapa de ponderación o la tabla de ponderación. En el bloque 702, el codificador 20 señala el número de candidatos de ponderación. En el bloque 703, el codificador 20 señala el índice de ponderación. Los bloques del proceso 700 pueden realizarse en un orden diferente. Por ejemplo, el número de candidatos de ponderación puede señalizarse antes de la etapa de ponderación (o la tabla de ponderación). El procedimiento de ejemplo para señalar parámetros de predicción residual generalizada de acuerdo con aspectos de esta divulgación descritos con respecto a la figura 7 puede implementarse en varios niveles de codificación, tales como una secuencia, una imagen, un grupo de tramas, una trama, un grupo de fragmentos, un fragmento, un grupo de unidades de codificación (CU), una unidad de codificación (CU), un grupo de unidades de predicción (PU), una unidad de predicción (PU), bloques, una región de píxeles o un píxel. Además, el procedimiento de ejemplo para señalar parámetros de predicción residual generalizada de acuerdo con los aspectos de esta divulgación descritos con respecto a la figura 7 también puede aplicarse a componentes de señal, tales como componentes de luma y componentes de croma. Por ejemplo, el procedimiento de ejemplo puede aplicarse solo a componentes luma, solo a componentes de croma, o una combinación de componentes de luma y componentes de croma (por ejemplo, cualquier combinación de una componente de luma y dos componentes de croma). Además,

todos los modos de realización descritos con respecto a la figura 7 pueden implementarse por separado o en combinación entre sí.

La figura 8 es un diagrama de flujo que ilustra un procedimiento de ejemplo para obtener parámetros de predicción residual generalizada de acuerdo con aspectos de esta divulgación. Los parámetros de GRP se pueden señalar como se explica con respecto a la figura 7. Los parámetros de GRP también pueden obtenerse de la información incluida en el flujo de bits. Los parámetros de GRP pueden obtenerse total o parcialmente de la información del flujo de bits. En un modo de realización, la etapa de ponderación  $S_w$  se obtiene a nivel de CU de acuerdo con el tamaño de CU relacionado. En la siguiente tabla se muestra un ejemplo de mapeo entre la etapa de ponderación  $S_w$  y el tamaño de CU.

Tabla 1 - Ejemplo de mapeo entre la etapa de ponderación y el tamaño de CU

Tamaño de CU	$S_w$
64x64	1/8
32x32	1/4
16x16	1/2
8x8	1/2

En otro modo de realización, el número máximo de candidatos de ponderación se ajusta a nivel de CU en base a información previamente codificada, tal como el modo de CU, el tamaño de CU y la cuantificación. Por ejemplo, para CU menores que 16x16, solo se pueden permitir dos candidatos de ponderación, por ejemplo, para ahorrar costes de señalización.

El procedimiento de ejemplo para obtener parámetros de predicción residual generalizada de acuerdo con aspectos de esta divulgación se explicará a continuación con referencia a la figura 8. El proceso 800 puede realizarse mediante un codificador (por ejemplo, el codificador mostrado en la figura 2, etc.), un decodificador (por ejemplo, el decodificador mostrado en la figura 3, etc.) o cualquier otro componente. Los bloques del proceso 800 se describen con respecto al decodificador 30 en la figura 3, pero el proceso 800 puede ser realizado por otros componentes, tal como un codificador, mencionados anteriormente.

En el bloque 801, el decodificador 30 obtiene información a partir del flujo de bits u obtiene información codificada previamente para determinar la información de ponderación. Por ejemplo, como se explicó anteriormente, los parámetros de GRP pueden basarse en el tamaño de CU. De forma alternativa, los parámetros de GRP pueden basarse en información codificada previamente, tal como el modo de CU, el tamaño de CU y la cuantificación. En el bloque 802, el decodificador 30 determina uno o más parámetros para la predicción residual generalizada basándose en la información obtenida en el bloque 801. Por ejemplo, el decodificador 30 puede determinar la etapa de ponderación  $S_w$  basándose en el tamaño de CU. El decodificador 30 también puede determinar el número de candidatos de ponderación  $N_w$  basándose en el tamaño de CU. El decodificador 30 también puede ajustar la información de ponderación basándose en información codificada previamente, tal como el modo de CU, el tamaño de CU y la cuantificación. El procedimiento de ejemplo para obtener parámetros de predicción residual generalizada de acuerdo con aspectos de esta divulgación descritos con respecto a la figura 8 puede implementarse en varios niveles de codificación, tales como una secuencia, una imagen, un grupo de tramas, una trama, un grupo de fragmentos, un fragmento, un grupo de unidades de codificación (CU), una unidad de codificación (CU), un grupo de unidades de predicción (PU), una unidad de predicción (PU), bloques, una región de píxeles o un píxel. Además, el procedimiento de ejemplo para obtener parámetros de predicción residual generalizados de acuerdo con aspectos de esta divulgación descritos con respecto a la figura 8 también puede aplicarse a componentes de señal, como componentes de luma y componentes de croma. Por ejemplo, el procedimiento de ejemplo puede aplicarse solo a componentes luma, solo a componentes de croma, o una combinación de componentes de luma y componentes de croma (por ejemplo, cualquier combinación de una componente de luma y dos componentes de croma). Además, todos los modos de realización descritos con respecto a la figura 8 pueden implementarse por separado o en combinación entre sí.

La figura 9 es un diagrama de flujo que ilustra un procedimiento de ejemplo para muestrear de manera ascendente o descendente una capa en la predicción residual generalizada de acuerdo con aspectos de esta divulgación. En el proceso de predicción entre capas, se aplica un proceso de filtrado de muestreo ascendente o descendente a la imagen de la capa base para que coincida con la relación de aspecto espacial de la capa de mejora. También se puede aplicar un proceso de filtrado, tal como un filtro de suavizado, incluso cuando el tamaño de imagen de la capa base y de la capa de mejora son idénticos. En general, se usa y se codifica de forma fija un conjunto fijo de filtros de muestreo ascendente, muestreo descendente y suavizado. El filtro puede seleccionarse del conjunto de acuerdo con el desplazamiento de píxeles fraccionarios (a veces denominado fase), que se calcula en base a la relación de aspecto espacial entre la capa base y las imágenes de capa de mejora.

En el marco de GRP, se pueden aplicar conjuntos de filtrado variantes para mejorar el rendimiento de predicción entre capas. Los conjuntos de filtrado pueden estar codificados de forma fija o señalizados a nivel de secuencia o de imagen. El índice de conjunto de filtros puede señalizarse u obtenerse a un nivel inferior, tal como CU y PU. El índice de conjunto de filtros puede obtenerse en base al valor del factor de ponderación  $w$ , o puede obtenerse en base al índice de ponderación  $i_w$ . La tabla de mapeo de derivación entre el índice del conjunto de filtrado y el factor de ponderación  $w$ , o entre el conjunto de filtrado y el índice de ponderación  $i_w$  puede estar codificada de forma fija o señalizada a nivel de secuencia o imagen.

El procedimiento de ejemplo para muestrear de manera ascendente o descendente una capa en la predicción residual generalizada de acuerdo con los aspectos de esta divulgación se explicará a continuación con referencia a la figura 9. El proceso 900 puede realizarse mediante un codificador (por ejemplo, el codificador mostrado en la figura 2, etc.), un decodificador (por ejemplo, el decodificador mostrado en la figura 3, etc.) o cualquier otro componente. Los bloques del proceso 900 se describen con respecto al decodificador 30 en la figura 3, pero el proceso 900 puede ser realizado por otros componentes, tal como un codificador, mencionados anteriormente.

En el bloque 901, el decodificador 30 determina si muestrear de manera ascendente la capa de referencia o muestrear de manera descendente la capa de mejora. En la escalabilidad espacial, dicho muestreo ascendente y dicho muestreo descendente se realiza de modo que la predicción entre capas se pueda realizar con la misma resolución. Si se determina que la capa de referencia será muestreada de manera ascendente en el bloque 902, el decodificador 30 muestrea de manera ascendente la capa de referencia hasta la resolución de la capa de mejora en el bloque 903. Por otro lado, si se determina que la capa de mejora será muestreada de manera descendente en el bloque 902, el decodificador 30 muestrea de manera descendente la capa de mejora hasta la resolución de la capa de referencia en el bloque 904. En el bloque 905, el decodificador 30 aplica un filtro de suavizado a las imágenes muestreadas de manera ascendente o descendente. El filtro de suavizado se puede aplicar incluso si la capa de mejora y las imágenes de capa de referencia son las mismas. El filtro de suavizado se puede seleccionar de forma apropiada. En el bloque 906, el decodificador 30 determina la unidad de vídeo actual usando GRP basándose en las imágenes muestreadas de manera ascendente o descendente.

El procedimiento de ejemplo para muestrear de manera ascendente o descendente una capa en la predicción residual generalizada de acuerdo con aspectos de esta divulgación descritos con respecto a la figura 9 puede implementarse en varios niveles de codificación, tales como una secuencia, una imagen, un grupo de tramas, una trama, un grupo de fragmentos, un fragmento, un grupo de unidades de codificación (CU), una unidad de codificación (CU), un grupo de unidades de predicción (PU), una unidad de predicción (PU), bloques, una región de píxeles o un píxel. Además, el procedimiento de ejemplo para muestrear de manera ascendente o descendente una capa en la predicción residual generalizada de acuerdo con los aspectos de esta divulgación descritos con respecto a la figura 9 también puede aplicarse a componentes de señal, tales como componentes de luma y componentes de croma. Por ejemplo, el procedimiento de ejemplo puede aplicarse solo a componentes de luma, solo a componentes de croma, o una combinación de componentes de luma y componentes de croma (por ejemplo, cualquier combinación de una componente de luma y dos componentes de croma). Además, todos los modos de realización descritos con respecto a la figura 9 pueden implementarse por separado o en combinación entre sí.

La figura 10 es un diagrama de flujo que ilustra un procedimiento de ejemplo para remapear, muestrear de manera ascendente o muestrear de manera descendente información de movimiento en la predicción residual generalizada de acuerdo con aspectos de esta divulgación. En algunos casos, cuando se aplica información de movimiento de una capa a otra capa para generar el predictor residual, la referencia disponible en una capa puede no estar disponible en otra capa. En tal caso, se necesita un remapeo del movimiento. En un modo de realización, si una referencia solo está disponible en una capa, se marca como no disponible de modo que esta referencia no se usará para generar el predictor residual en el marco de GRP propuesto. En otro modo de realización, la referencia no disponible se reemplaza por la referencia al comienzo de la lista de referencias relacionadas, y el movimiento se establece en movimiento cero.

En la codificación de vídeo 3D, los datos de vídeo SVC también incluyen datos de vídeo para diferentes vistas. Debido a que las vistas pueden referirse a diferentes ángulos, puede haber disparidad entre las diferentes vistas. Si el movimiento se remapea en el contexto de la codificación de vídeo 3D, el vector de disparidad se puede considerar a la hora de remapear el movimiento.

En casos de escalabilidad espacial, el vector de movimiento se puede muestrear de manera ascendente o descendente debido a las diferentes resoluciones entre la capa de mejora y la capa base. En un modo de realización, el escalado del vector de movimiento se basa directamente en la relación de resolución. En otro modo de realización, se puede aplicar un desplazamiento de fase adicional (+1 o -1) después del escalado directo. El desplazamiento de fase adicional se puede señalar en el flujo de bits u obtener en base a información codificada previamente, tal como el tamaño de PU, un vector de movimiento, la profundidad de CU, etc.

El procedimiento de ejemplo para remapear, muestrear de manera ascendente o muestrear de manera descendente información de movimiento de acuerdo con los aspectos de esta divulgación se explicará a continuación con referencia a la figura 10. El proceso 1000 puede realizarse mediante un codificador (por ejemplo, el codificador mostrado en la

figura 2, etc.), un decodificador (por ejemplo, el decodificador mostrado en la figura 3, etc.) o cualquier otro componente. Los bloques del proceso 1000 se describen con respecto al decodificador 30 en la figura 3, pero el proceso 1000 puede ser realizado por otros componentes, tal como un codificador, mencionados anteriormente. En el bloque 1001, si una referencia para la información de movimiento no está disponible en una de las capas, el decodificador 30 remapea la información de movimiento en el bloque 1002. Por ejemplo, el decodificador 30 puede marcar una referencia como no disponible si la referencia correspondiente en otra capa no está disponible. De forma alternativa, el decodificador 30 puede remapear la referencia a una referencia en una lista de referencias relacionada. Si hay disponible una referencia para la información de movimiento en las capas usadas para la interpolación en el bloque 1001, el decodificador 30 puede no realizar un procesamiento adicional como se muestra en el bloque 1003. En el bloque 1004, si se usa SVC espacial, el decodificador 30 determina si se muestrear de manera ascendente la información de movimiento de capa de referencia o muestrear de manera descendente la información de movimiento de capa de mejora en el bloque 1005. Si no se usa la escalabilidad espacial, el decodificador 30 puede no realizar ningún procesamiento adicional como se muestra en el bloque 1006. En el bloque 1007, si se determina que la información de movimiento de capa de referencia será muestreada de manera ascendente, el decodificador 30 muestrea de manera ascendente la información de movimiento de capa de referencia hasta la resolución de la capa de mejora en el bloque 1008. Por otro lado, si se determina que la información de movimiento de capa de mejora será muestreada de manera descendente en el bloque 1007, el decodificador 30 muestrea de manera descendente la información de movimiento de capa de mejora hasta la resolución de la capa de referencia en el bloque 1009. En el bloque 1010, el decodificador 30 determina la unidad de vídeo actual usando GRP que usa las imágenes muestreadas de manera ascendente o descendente.

El procedimiento de ejemplo para remapear, muestrear de manera ascendente o muestrear de manera descendente información de movimiento de acuerdo con aspectos de esta divulgación descritos con respecto a la figura 10 puede implementarse en varios niveles de codificación, tales como una secuencia, una imagen, un grupo de tramas, una trama, un grupo de fragmentos, un fragmento, un grupo de unidades de codificación (CU), una unidad de codificación (CU), un grupo de unidades de predicción (PU), una unidad de predicción (PU), bloques, una región de píxeles o un píxel. Además, el procedimiento de ejemplo para remapear, muestrear de manera ascendente o muestrear de manera descendente información de movimiento de acuerdo con los aspectos de esta divulgación descritos con respecto a la figura 10 también puede aplicarse a componentes de señal, tales como componentes de luma y componentes de croma. Por ejemplo, el procedimiento de ejemplo puede aplicarse solo a componentes luma, solo a componentes de croma, o una combinación de componentes de luma y componentes de croma (por ejemplo, cualquier combinación de una componente de luma y dos componentes de croma). Además, todos los modos de realización descritos con respecto a la figura 10 pueden implementarse por separado o en combinación entre sí.

La figura 11 es un diagrama de flujo que ilustra un procedimiento de ejemplo para determinar un factor de ponderación para codificar en la predicción residual generalizada de acuerdo con aspectos de esta divulgación. El procedimiento de ejemplo puede aplicarse a la optimización en el lado de codificador. En un modo de realización, el mejor factor de ponderación  $w$  para cada CU se determina verificando el coste de velocidad-distorsión de CU con cada candidato a factor de ponderación. El factor de ponderación con coste mínimo se selecciona como factor de ponderación  $w$  para la CU. En otro modo de realización, el predictor residual se obtiene aplicando el movimiento de la predicción temporal de capa de mejora  $P_e$  a la predicción temporal de capa base  $P_b$ . El factor de ponderación  $w$  se puede determinar como sigue:

$$w = \frac{\sum_{x,y} \{(I - P_e) \cdot (\hat{I}_b - P_b)\}}{\sum_{x,y} \{(\hat{I}_b - P_b)^2\}} \quad (7)$$

donde  $I$  indica la imagen de origen para la capa de mejora,  $\sum_{x,y} \{(I - P_e) \cdot (\hat{I}_b - P_b)\}$  indica la suma del producto escalar del bloque diferencial  $(I - P_e)$  e  $(\hat{I}_b - P_b)$ .

El procedimiento de ejemplo para determinar un factor de ponderación para codificar en la predicción residual generalizada de acuerdo con aspectos de esta divulgación se explicará a continuación con referencia a la figura 11. El proceso 1100 puede realizarse mediante un codificador (por ejemplo, el codificador mostrado en la figura 2, etc.), un decodificador (por ejemplo, el decodificador mostrado en la figura 3, etc.) o cualquier otro componente. Los bloques del proceso 1100 se describen con respecto al codificador 20 en la figura 2, pero el proceso 1100 puede ser realizado por otros componentes, tal como un decodificador, mencionados anteriormente. En el bloque 1101, el codificador 20 obtiene la predicción residual de la EL aplicando el movimiento de la predicción temporal de EL a la predicción temporal de BL. En el bloque 1102, el decodificador 30 obtiene el factor de ponderación basándose en la predicción residual obtenida. El procedimiento de ejemplo para determinar un factor de ponderación para la codificación en la predicción residual generalizada de acuerdo con aspectos de esta divulgación descritos con respecto a la figura 11 puede implementarse en varios niveles de codificación, tales como una secuencia, una imagen, un grupo de tramas, una trama, un grupo de fragmentos, un fragmento, un grupo de unidades de codificación (CU), una unidad de codificación (CU), un grupo de unidades de predicción (PU), una unidad de predicción (PU), bloques, una región de píxeles o un píxel. Además, el procedimiento de ejemplo para determinar un factor de ponderación para la codificación en la

predicción residual generalizada de acuerdo con los aspectos de esta divulgación descritos con respecto a la figura 11 también puede aplicarse a componentes de señal, tales como componentes de luma y componentes de croma. Por ejemplo, el procedimiento de ejemplo puede aplicarse solo a componentes luma, solo a componentes de croma, o una combinación de componentes de luma y componentes de croma (por ejemplo, cualquier combinación de una componente de luma y dos componentes de croma). Además, todos los modos de realización descritos con respecto a la figura 11 pueden implementarse por separado o en combinación entre sí.

Se debe reconocer que, dependiendo del ejemplo, determinadas acciones o eventos de cualquiera de las técnicas descritas en el presente documento se pueden realizar en una secuencia distinta, se pueden añadir, fusionar o excluir por completo (por ejemplo, no todas las acciones o eventos descritos son necesarios para la puesta en práctica de las técnicas). Además, en determinados ejemplos, las acciones o eventos se pueden realizar simultáneamente, por ejemplo, a través de procesamientos de múltiples hilos, procesamientos por interrupciones o mediante múltiples procesadores, en lugar de secuencialmente.

En uno o más ejemplos, las funciones descritas se pueden implementar en hardware, software, firmware o en cualquier combinación de los mismos. Si se implementan en software, las funciones pueden almacenarse en, o 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 corresponden a un medio tangible, tales como unos medios de almacenamiento de datos, o unos medios de comunicación que incluyen cualquier medio que facilita la transferencia de un programa informático de un lugar a otro, por ejemplo, de acuerdo con un protocolo de comunicación. De esta manera, los medios legibles por ordenador pueden corresponder, en general, a (1) unos medios de almacenamiento tangibles legibles por ordenador que no son transitorios o a (2) un medio de comunicación tal como una señal u onda portadora. Los medios de almacenamiento de datos pueden ser unos medios cualesquiera disponibles a los que se puede acceder desde uno o más ordenadores o uno o más procesadores para recuperar instrucciones, un código y/o unas estructuras de datos para la implementación de las técnicas descritas en la presente divulgación. Un producto de programa informático puede incluir un medio legible por ordenador.

A modo de ejemplo, y no de limitación, dichos medios de almacenamiento legibles por ordenador pueden comprender RAM, ROM, EEPROM, CD-ROM u otro almacenamiento en disco óptico, almacenamiento en disco magnético u otros dispositivos de almacenamiento magnético, memoria *flash* o cualquier otro medio que se puede usar para almacenar un código de programa deseado en forma de instrucciones o estructuras de datos y al que se puede acceder mediante un ordenador. Además, cualquier conexión recibe apropiadamente la denominación de medio legible por ordenador. Por ejemplo, si las instrucciones se transmiten desde un sitio web, un servidor u otra fuente remota usando un cable coaxial, un cable de fibra óptica, un par trenzado, una línea de abonado digital (DSL) o unas 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 están incluidos en la definición de medio. Sin embargo, se deberá entender 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, están dirigidos a medios de almacenamiento tangibles no transitorios. El término disco, como se usa en el presente documento, incluye disco compacto (CD), disco láser, disco óptico, disco versátil digital (DVD), disco flexible y disco Blu-ray, donde los discos flexibles normalmente reproducen datos magnéticamente, mientras que los demás discos reproducen datos ópticamente con láseres. Combinaciones de lo anterior también se deben incluir dentro del alcance de los medios legibles por ordenador.

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 lógicas programables *in situ* (FPGA) u otros circuitos lógicos integrados o discretos equivalentes pueden ejecutar las instrucciones. En consecuencia, el término "procesador", como se usa en el presente documento, se puede referir 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 se puede proporcionar dentro de módulos de hardware y/o de software dedicados, configurados para la codificación y la decodificación, o incorporar en un códec combinado. Asimismo, las técnicas se podrían implementar por completo en uno o más circuitos o elementos lógicos.

Las técnicas de esta divulgación se pueden implementar en una amplia variedad de dispositivos o aparatos, incluidos un teléfono inalámbrico, un circuito integrado (IC) o un conjunto de IC (por ejemplo, un conjunto de chips). En esta divulgación se describen diversos componentes, módulos o unidades para destacar aspectos funcionales de dispositivos configurados para realizar las técnicas divulgadas, pero no se requiere necesariamente su realización mediante diferentes unidades de hardware. En cambio, como se describe anteriormente, diversas unidades se pueden combinar en una unidad de hardware de códec o proporcionar mediante un grupo de unidades de hardware interoperativas, que incluyen uno o más procesadores descritos anteriormente, junto con software y/o firmware adecuados.

Se han descrito diversos ejemplos. Estos y otros ejemplos están dentro del alcance de las siguientes reivindicaciones.

## REIVINDICACIONES

1. Un aparato para codificar información de vídeo, que comprende:

5 una unidad de memoria configurada para almacenar información de vídeo asociada a una capa de referencia; y un procesador en comunicación con la unidad de memoria, estando configurado el procesador para:

determinar una reconstrucción de una unidad de vídeo en una capa de mejora, en el que la reconstrucción de la unidad de vídeo en la capa de mejora es una suma de:

10 (i) una predicción residual de la unidad de vídeo en la capa de mejora,  
(ii) una predicción espacial o temporal de la unidad de vídeo en la capa de mejora, y  
(iii) una predicción residual ajustada asociada a la capa de referencia,

15 en el que la predicción residual ajustada es una predicción residual asociada a la capa de referencia igual a una diferencia entre una predicción espacial o temporal de la unidad de vídeo en la capa de referencia y una reconstrucción de la unidad de vídeo en la capa de referencia,  
multiplicada por un factor de ponderación que es diferente de 1, en el que el factor de ponderación se selecciona de una pluralidad de candidatos a factor de ponderación, donde cada uno de la pluralidad de candidatos a factor de  
20 ponderación se indica mediante un índice respectivo.

2. Un aparato para decodificar información de vídeo, que comprende:

25 una unidad de memoria configurada para almacenar información de vídeo asociada a una capa de referencia; y un procesador en comunicación con la unidad de memoria, estando configurado el procesador para:

determinar una reconstrucción de una unidad de vídeo en una capa de mejora, en el que la reconstrucción de la unidad de vídeo en la capa de mejora es una suma de:

30 (i) una predicción residual de la unidad de vídeo en la capa de mejora,  
(ii) una predicción espacial o temporal de la unidad de vídeo en la capa de mejora, y  
(iii) una predicción residual ajustada asociada a la capa de referencia,

35 en el que la predicción residual ajustada es una predicción residual asociada a la capa de referencia igual a una diferencia entre una predicción espacial o temporal de la unidad de vídeo en la capa de referencia y una reconstrucción de la unidad de vídeo en la capa de referencia,  
multiplicada por un factor de ponderación que es diferente de 1, en el que el factor de ponderación se selecciona de una pluralidad de candidatos a factor de ponderación, donde cada uno de la pluralidad de candidatos a factor de  
40 ponderación se indica mediante un índice respectivo.

3. El aparato de la reivindicación 1 o 2, en el que el procesador está configurado además para aplicar el factor de ponderación a nivel de codificación seleccionado de un grupo que comprende: una secuencia, un grupo de tramas, una trama, un grupo de fragmentos, un fragmento, un grupo de unidades de codificación (CU), una unidad de codificación (CU), un grupo de unidades de predicción (PU), una unidad de predicción (PU), un bloque, una región de  
45 píxeles, un píxel y cualquier combinación de una componente de luma y dos componentes de croma.

4. El aparato de la reivindicación 1 o 2, en el que el índice del factor de ponderación se recibe en un flujo de bits.

5. El aparato de la reivindicación 1 o 2, en el que el factor de ponderación se determina en base a la información de ponderación,  
50 en el que la información de ponderación comprende uno o más de una etapa de ponderación, una tabla de ponderación, un número de candidatos a factor de ponderación y un índice de ponderación y, opcionalmente, la información de ponderación se recibe en un flujo de bits.

6. El aparato de la reivindicación 5, en el que la información de ponderación se obtiene, al menos parcialmente, en base a, al menos en parte, información codificada previamente,  
55 en el que la información codificada previamente se proporciona a nivel de codificación y comprende uno o más de: un parámetro de cuantificación, un tamaño de CU, un tamaño de PU y un modo de codificación de CU.

7. El aparato de la reivindicación 6, en el que el número de candidatos a factor de ponderación se obtiene en base a la información codificada previamente.

8. El aparato de la reivindicación 1 o 2, en el que el procesador está configurado además para realizar codificación de vídeo 3D, y en el que la capa de referencia comprende una pluralidad de capas de referencia o vistas de referencia.

65

9. El aparato de la reivindicación 1 o 2, en el que el procesador está configurado además para determinar el valor de predicción residual en la codificación de vídeo escalable espacial mediante muestreo ascendente o muestreo descendente de una unidad de vídeo de la capa de referencia.

- 5 10. El aparato de la reivindicación 1, en el que dicho procesador está configurado además para codificar información de vídeo no codificada y determinar el factor de ponderación ( $w$ ) de acuerdo con la relación:

$$w = \frac{\sum \{(I - P_e) \cdot (\hat{I}_b - P_b)\}}{\sum \{(\hat{I}_b - P_b)^2\}}$$

- 10 en la que  $I$  corresponde a una imagen de origen,  $P_e$  corresponde a una predicción temporal o una intrapredicción espacial de una capa de mejora,  $P_b$  corresponde a una predicción temporal o una intrapredicción espacial de una capa base, e  $\hat{I}_b$  corresponde a una reconstrucción de capa base, determinada a partir de la información de vídeo sin codificar.

- 15 11. Un procedimiento para codificar información de vídeo, que comprende:

- almacenar información de vídeo asociada a una capa de referencia; y  
- determinar una reconstrucción de una unidad de vídeo en una capa de mejora en el que la reconstrucción de la unidad de vídeo en la capa de mejora es una suma de:

- 20 (i) una predicción residual de la unidad de vídeo en la capa de mejora,  
(ii) una predicción espacial o temporal de la unidad de vídeo en la capa de mejora, y  
(iii) una predicción residual ajustada asociada a la capa de referencia, en el que la predicción residual ajustada es una predicción residual asociada a la capa de referencia igual a una diferencia entre una predicción espacial o  
25 temporal de la unidad de vídeo en la capa de referencia y una reconstrucción de la unidad de vídeo en la capa de referencia, multiplicada por un factor de ponderación que es diferente de 1, en el que el factor de ponderación se selecciona de una pluralidad de candidatos a factor de ponderación, donde cada uno de la pluralidad de candidatos a factor de ponderación se indica mediante un índice respectivo.

- 30 12. Un procedimiento para codificar información de vídeo, que comprende:

almacenar información de vídeo asociada a una capa de referencia; y  
determinar una reconstrucción de una unidad de vídeo en una capa de mejora en el que la reconstrucción de la unidad de vídeo en la capa de mejora es una suma de:

- 35 (i) una predicción residual de la unidad de vídeo en la capa de mejora,  
(ii) una predicción espacial o temporal de la unidad de vídeo en la capa de mejora, y  
(iii) una predicción residual ajustada asociada a la capa de referencia, en el que la predicción residual ajustada es una predicción residual asociada a la capa de referencia igual a una diferencia entre una predicción espacial o  
40 temporal de la unidad de vídeo en la capa de referencia y una reconstrucción de la unidad de vídeo en la capa de referencia, multiplicada por un factor de ponderación que es diferente de 1, en el que el factor de ponderación se selecciona de una pluralidad de candidatos a factor de ponderación, donde cada uno de la pluralidad de candidatos a factor de ponderación se indica mediante un índice respectivo.

- 45 13. Un medio de almacenamiento legible por ordenador que tiene instrucciones almacenadas en el mismo que, cuando se ejecutan, hacen que un aparato lleve a cabo el procedimiento de cualquiera de las reivindicaciones 11 y 12.



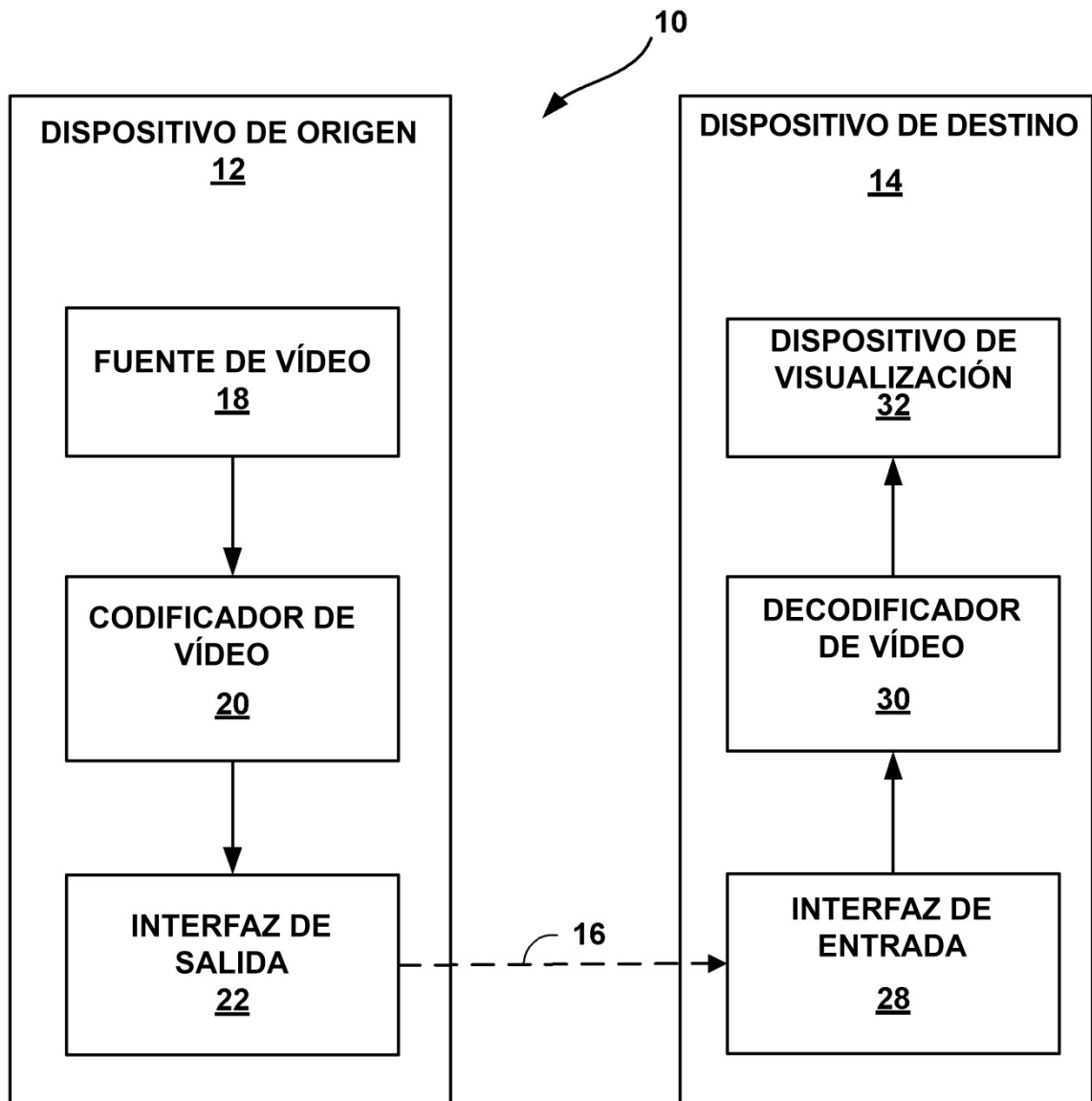


FIG. 1

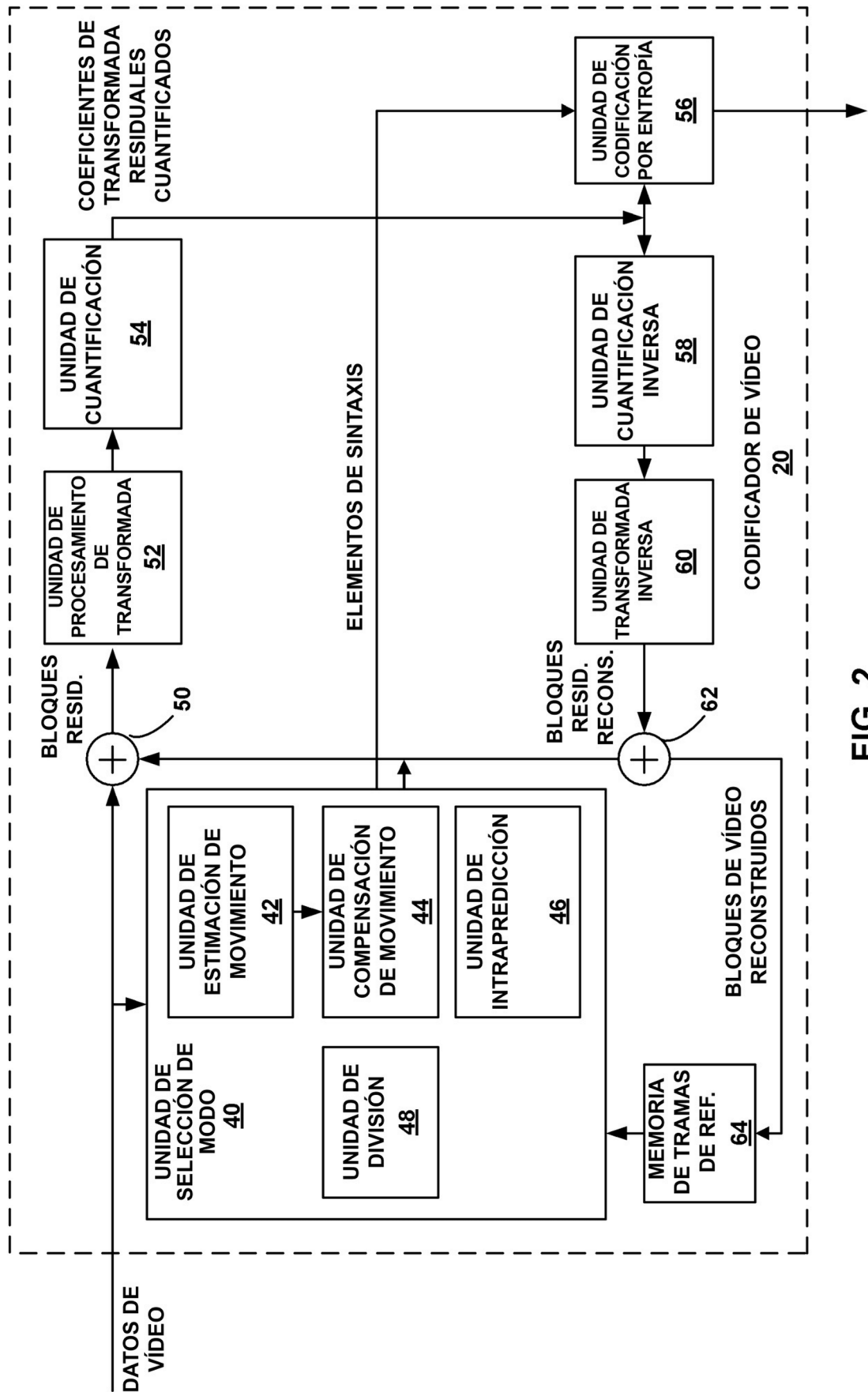


FIG. 2

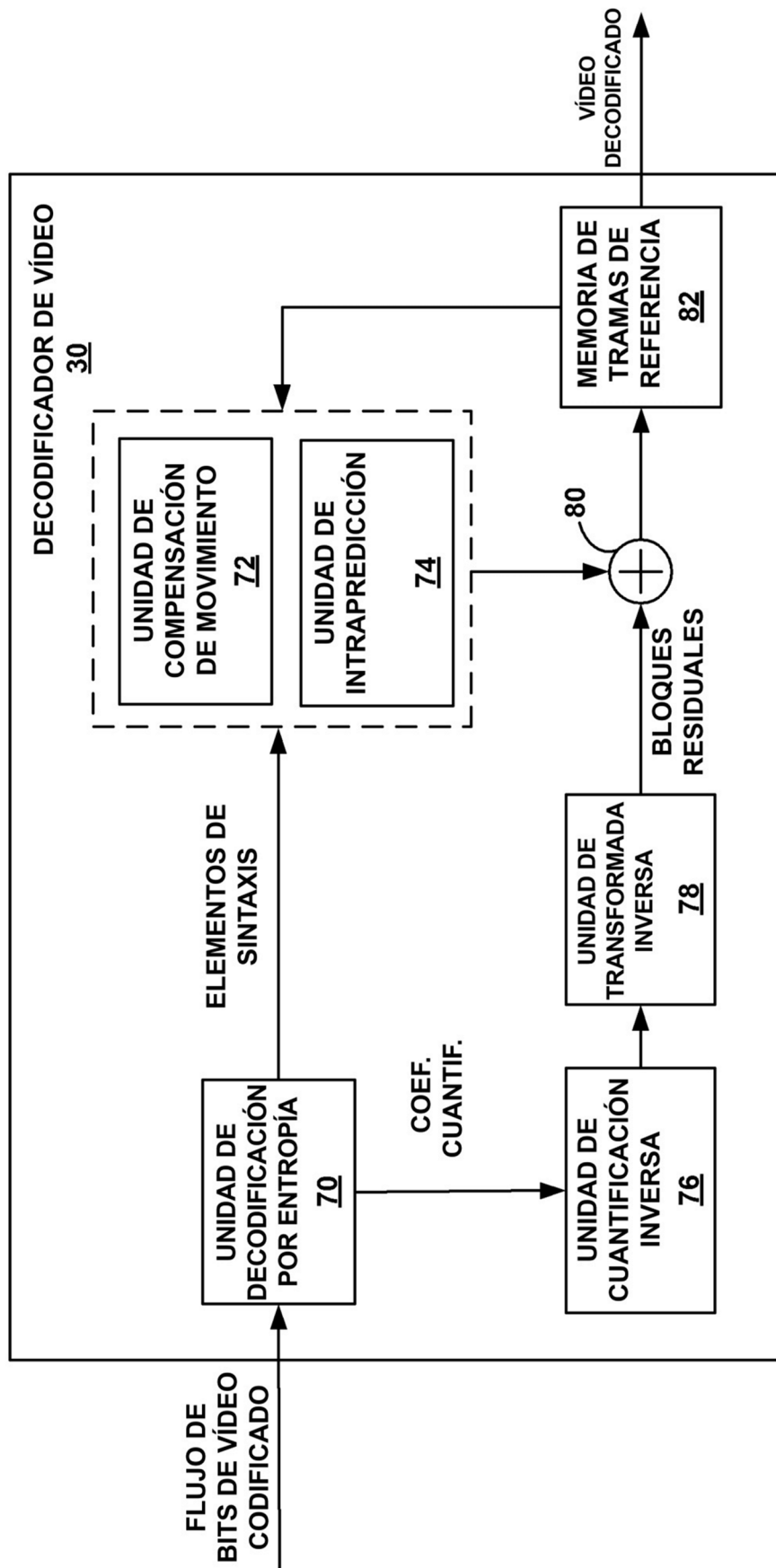
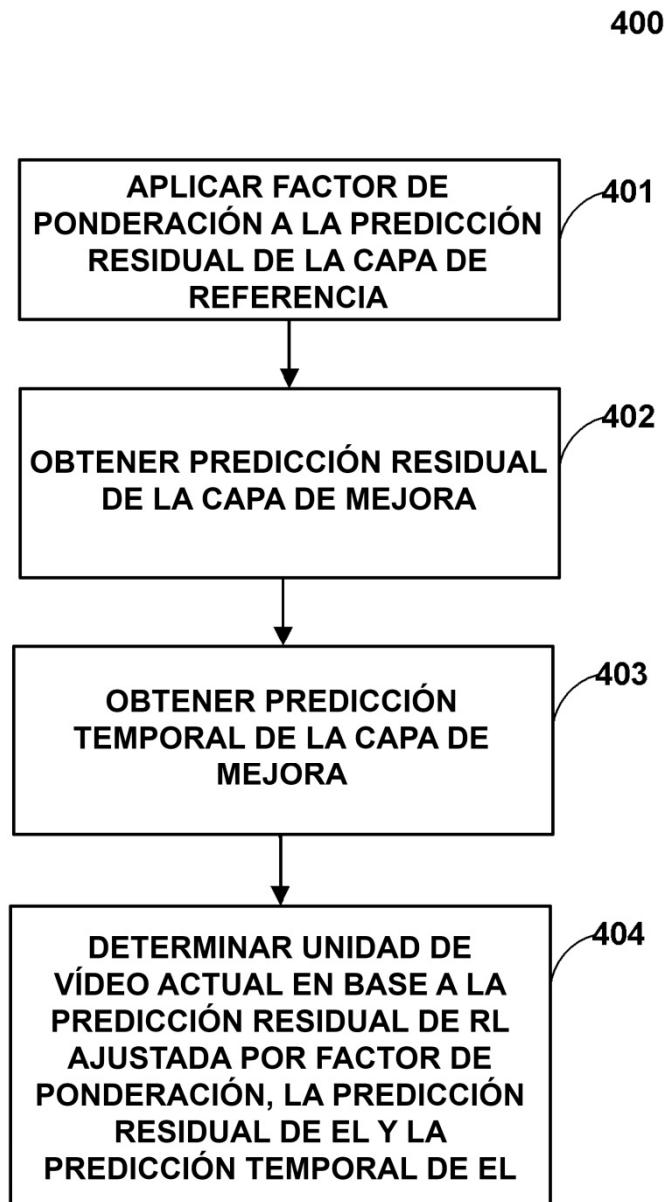


FIG. 3



**FIG. 4**

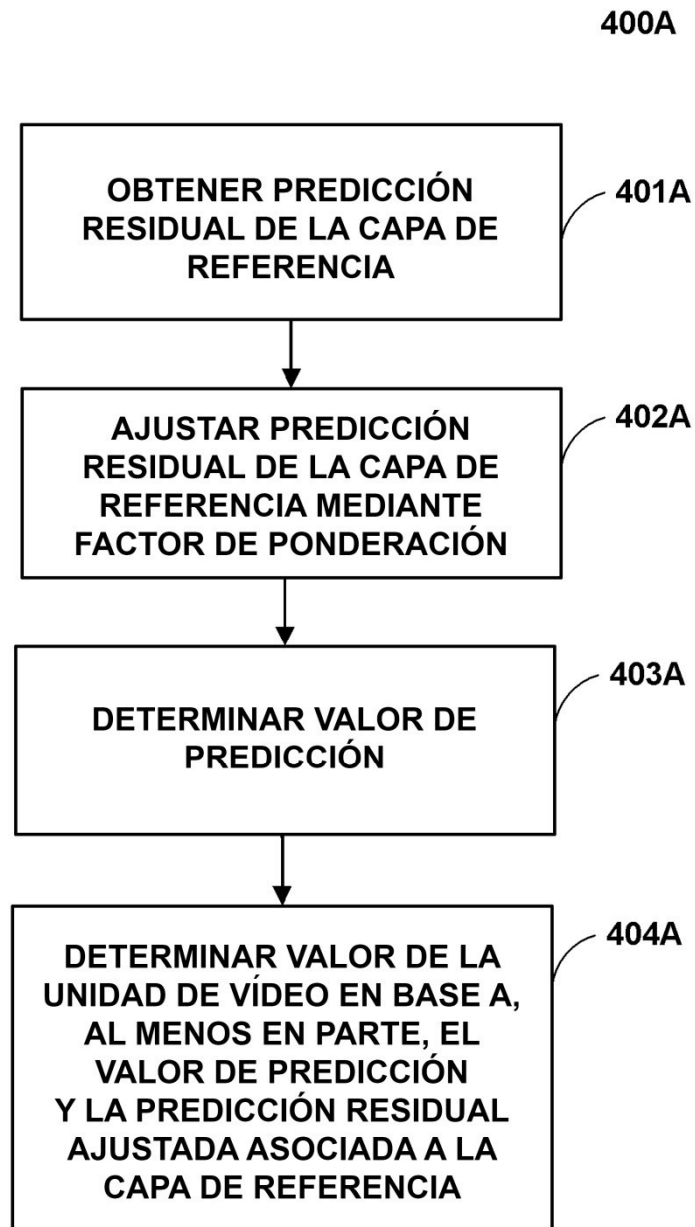


FIG. 4A

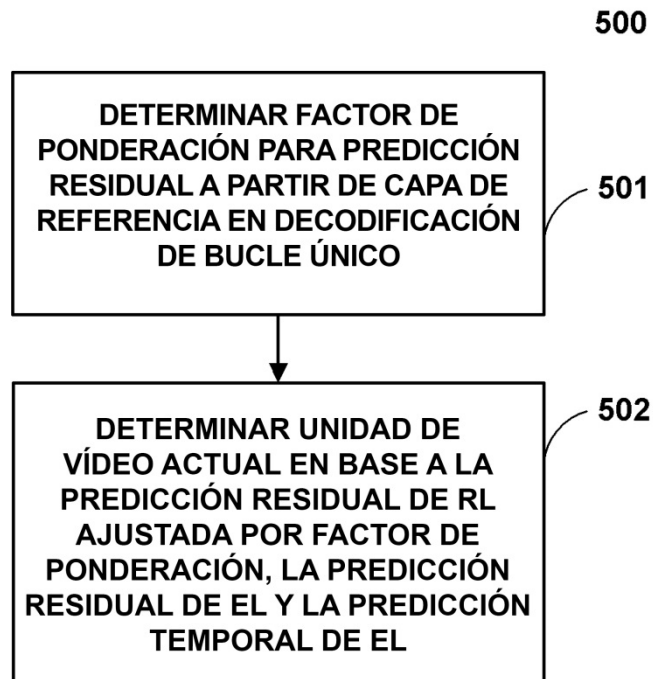


FIG. 5

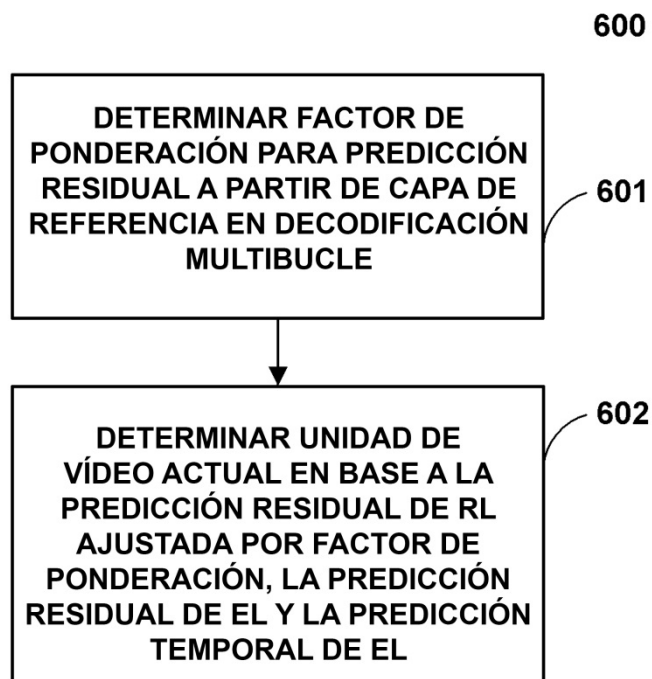
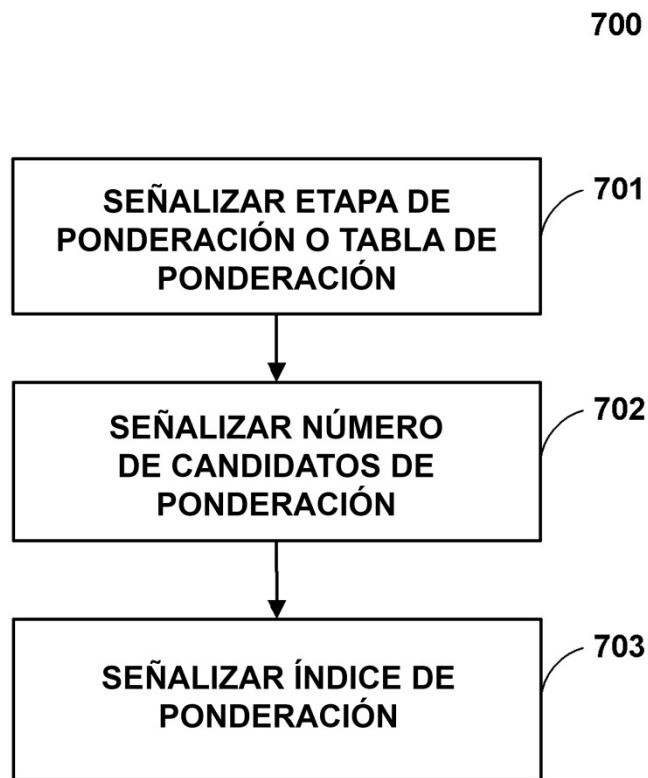
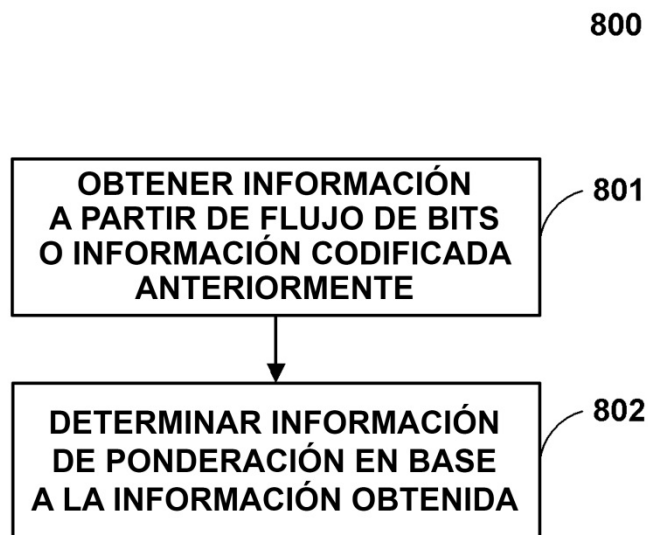


FIG. 6



**FIG. 7**



**FIG. 8**

900

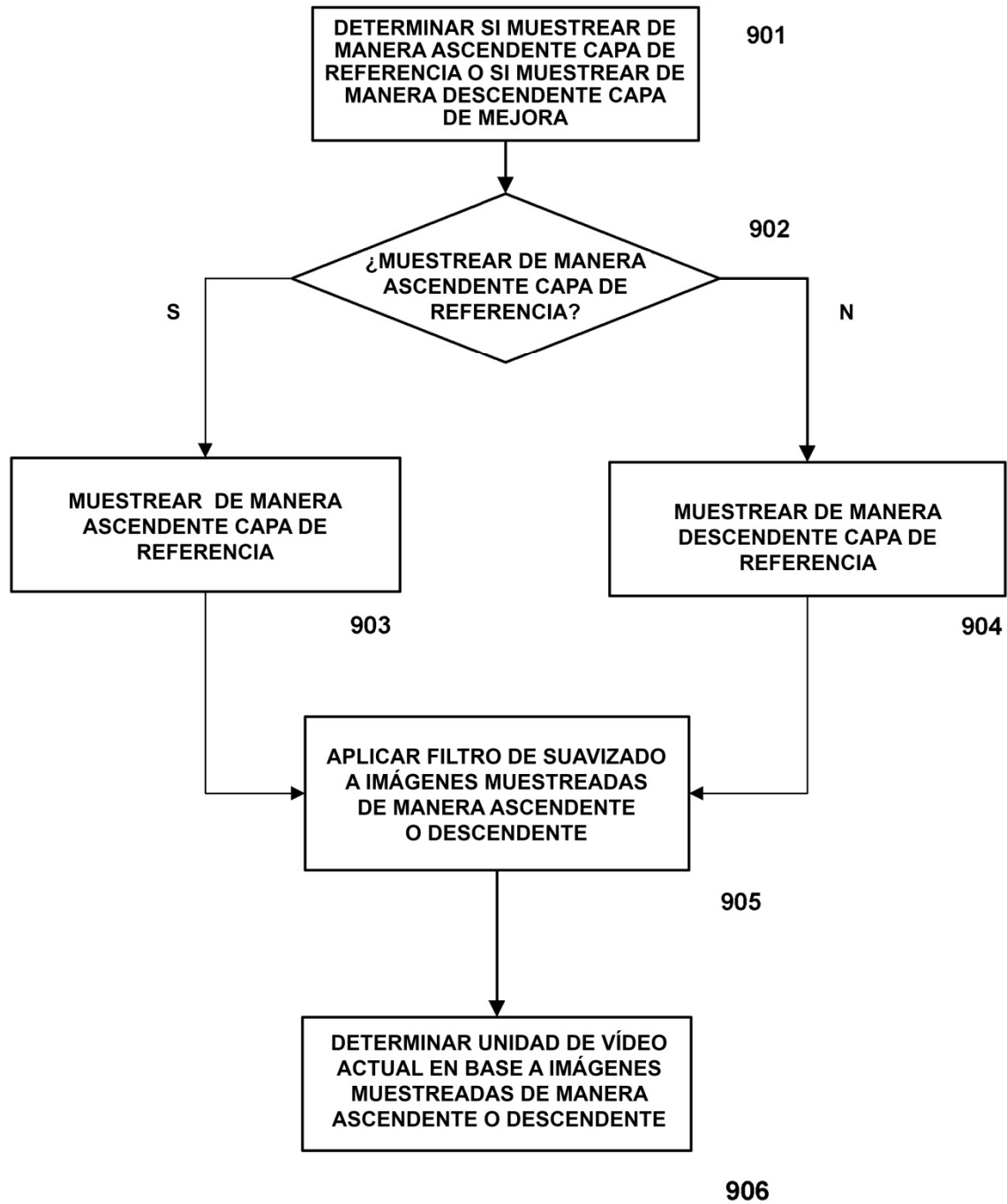


FIG. 9



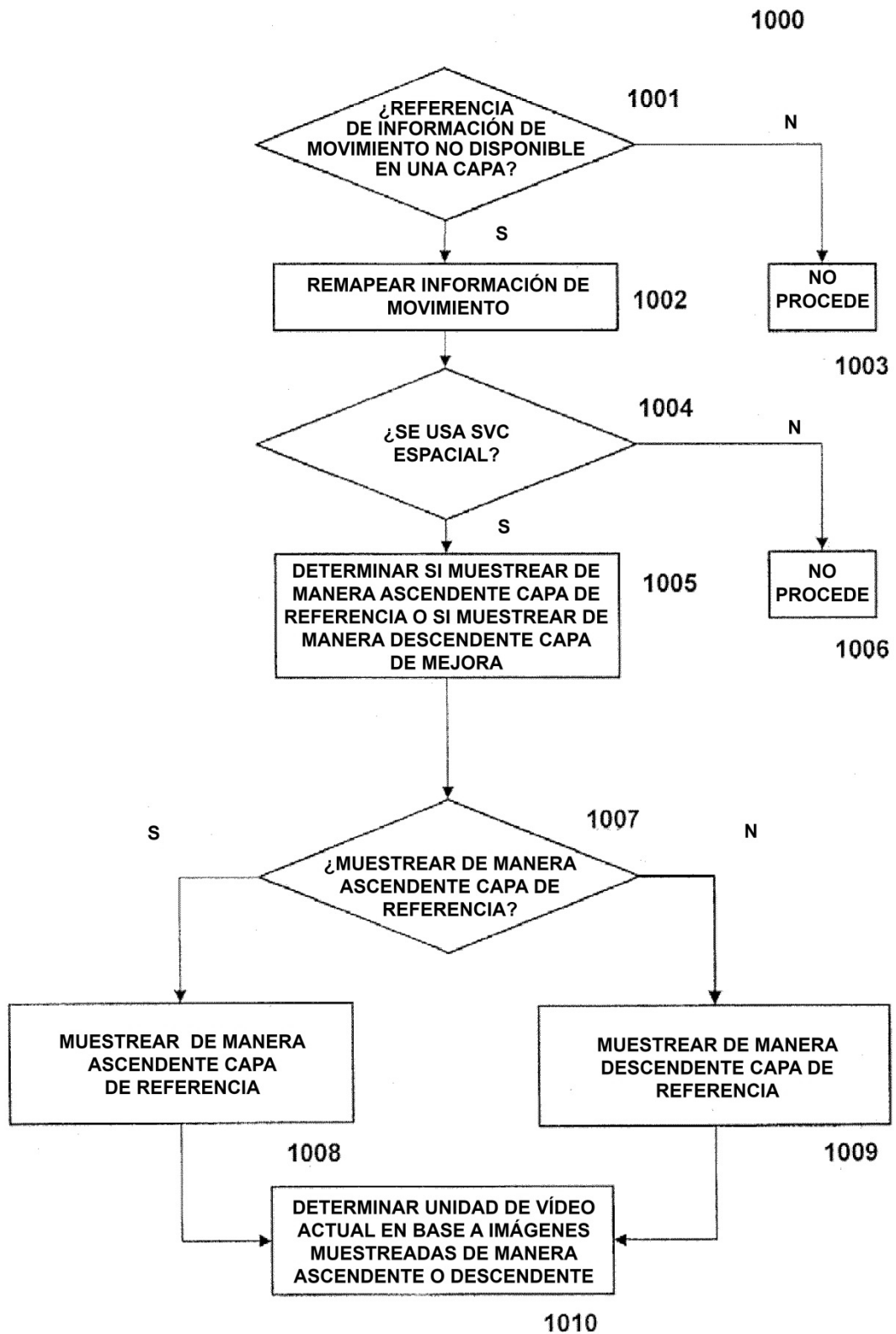
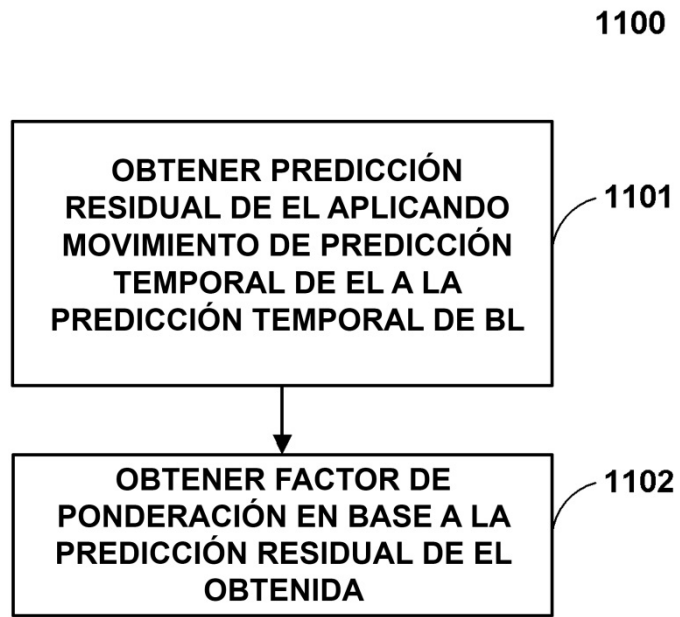


FIG. 10



**FIG. 11**