

19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 796 535**

51 Int. Cl.:

H04N 21/235 (2011.01)
H04N 21/2381 (2011.01)
H04N 21/435 (2011.01)
H04N 21/436 (2011.01)
H04N 21/4363 (2011.01)
H04N 21/438 (2011.01)
H04N 21/643 (2011.01)
H04N 21/84 (2011.01)
H04N 21/845 (2011.01)
H04N 21/854 (2011.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

- 86 Fecha de presentación y número de la solicitud internacional: **20.07.2011 PCT/US2011/044742**
- 87 Fecha y número de publicación internacional: **26.01.2012 WO12012574**
- 96 Fecha de presentación y número de la solicitud europea: **20.07.2011 E 11738371 (1)**
- 97 Fecha y número de publicación de la concesión europea: **11.03.2020 EP 2596632**

54 Título: **Proporcionar conjuntos de datos de secuencia para la transmisión continua de datos de vídeo**

30 Prioridad:

20.07.2010 US 840020

45 Fecha de publicación y mención en BOPI de la traducción de la patente:
27.11.2020

73 Titular/es:

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

72 Inventor/es:

**CHEN, YING y
KARCZEWICZ, MARTA**

74 Agente/Representante:

FORTEA LAGUNA, Juan José

ES 2 796 535 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

Proporcionar conjuntos de datos de secuencia para la transmisión continua de datos de vídeo

5 CAMPO TÉCNICO

[0001] Esta divulgación se refiere al transporte de datos de vídeo codificados.

10 ANTECEDENTES

10 **[0002]** Las capacidades de vídeo digital se pueden incorporar a una amplia gama de dispositivos, incluyendo
 15 televisores digitales, sistemas de radiodifusión directa digital, sistemas de radiodifusión inalámbrica, asistentes
 personales digitales (PDA), ordenadores portátiles o de escritorio, cámaras digitales, dispositivos de grabación
 digitales, reproductores de medios digitales, dispositivos de videojuegos, consolas de videojuegos, teléfonos de radio
 20 celulares o por satélite, dispositivos de videoconferencia y similares. Los dispositivos de vídeo digitales implementan
 técnicas de compresión de vídeo, tales como las descritas en las normas definidas por MPEG-2, MPEG-4, ITU-T H.263
 o ITU-T H.264/MPEG-4, parte 10, Codificación de Vídeo Avanzada (AVC) y ampliaciones de dichas normas, para
 transmitir y recibir información de vídeo digital de manera más eficaz.

20 **[0003]** Las técnicas de compresión de vídeo realizan predicción espacial y/o predicción temporal para reducir o
 eliminar la redundancia inherente a las secuencias de vídeo. Para la codificación de vídeo basada en bloques, una
 trama o un fragmento de vídeo se pueden dividir en macrobloques. Cada macrobloque se puede dividir aún más. Los
 25 macrobloques de una trama o un sector intracodificados (I) se codifican usando predicción espacial con respecto a
 unos macrobloques vecinos. Los macrobloques de una trama o un fragmento intercodificados (P o B) pueden usar
 predicción espacial con respecto a unos macrobloques vecinos de la misma trama o fragmento, o predicción temporal
 con respecto a otras tramas de referencia.

30 **[0004]** Después de que se hayan codificado los datos de vídeo, los datos de vídeo se pueden agrupar en paquetes
 para su transmisión o almacenamiento. Los datos de vídeo se pueden ensamblar en un archivo de vídeo que se ajusta
 a cualquiera de una variedad de normas, tales como el formato de archivo de medios de base de la Organización
 Internacional de Normalización (ISO) y ampliaciones del mismo, tales como la AVC.

35 **[0005]** Se han dedicado esfuerzos para elaborar nuevas normas de codificación de vídeo basándose en H.264/AVC.
 Una de estas normas es la norma de codificación de vídeo ajustable a escala (SVC), que es la extensión ajustable a
 escala para la H.264/AVC. Otra norma es la codificación de vídeo multivista (MVC), que se ha convertido en la
 extensión multivista para H.264/AVC. Una versión de la norma AVC se describe en JVT-AD007, "Editors' draft revision
 to ITU-T Rec. H.264 | ISO/IEC 14496-10 Avance Video Coding - in preparation for ITU-T SG 16 AAP Consent (in
 40 integrated form)", 30ª reunión de JVT, Ginebra, Suiza, febrero de 2009", disponible en http://wftp3.itu.int/av-arch/jvt-site/2009_01_Geneva/JVT-AD007.zip. Este documento integra la SVC y la MVC en la memoria descriptiva de la AVC.

45 **[0006]** El documento US 2004/0006575 describe un procedimiento y un aparato para admitir formatos de codificación
 avanzados en archivos de medios. Los siguientes textos de estudio y especificaciones técnicas también son
 pertinentes para este campo: "Study Text of ISO/IEC 14496-10:2005/FPDAM3 Scalable Video Coding", MPEG or
 ISO/IEC JTC1/SC29/WG11, n.º N8455, 10 de noviembre de 2006; "Digital Cellular Telecommunications Systems
 (Phase 2+); Universal Mobile Telecommunications System (UMTS); LTE; Transparent end-to-end packet switched
 50 streaming service (PSS)" ETSI vol 3GPP SA 4 n.º 9.2.0, 1 de junio de 2010; Universal Mobile Telecommunications
 System (UMTS) LTE Transparent end-to-end Packet Switched Streaming Service Protocols and Codecs" ETSI 3GPP
 SA 4 V9.3.0.

50 BREVE EXPLICACIÓN

55 **[0007]** La invención se define en las reivindicaciones a las que está dirigida ahora la referencia. En general, esta
 divulgación describe técnicas para proporcionar conjuntos de datos de secuencia que describen imágenes de vídeo
 codificadas de un archivo de vídeo, por ejemplo, para admitir la transmisión continua del archivo de vídeo de acuerdo
 con la transmisión continua del HTTP. Los conjuntos de datos de secuencia pueden incluir, por ejemplo, conjuntos de
 60 parámetros de secuencia, conjuntos de parámetros de imagen y algunos mensajes de información de mejora
 complementaria (SEI), tales como, por ejemplo, mensajes de SEI a nivel de secuencia. Estas técnicas pueden incluir
 desacoplar los conjuntos de datos de secuencia de las imágenes de vídeo codificadas descritas por los conjuntos de
 datos de secuencia. Es decir, en lugar de incluir mensajes de SEI a nivel de secuencia con imágenes de vídeo
 65 codificadas descritas por los mensajes de SEI a nivel de secuencia, las técnicas de esta divulgación pueden incluir
 desacoplar los mensajes de SEI a nivel de secuencia de las imágenes de vídeo codificadas. Por consiguiente, los
 mensajes de SEI a nivel de secuencia pueden transferirse desde un dispositivo de origen a un dispositivo de destino
 por separado de las imágenes de vídeo codificadas. En algunos ejemplos, un conjunto de datos de secuencia que
 incluye los mensajes de SEI a nivel de secuencia puede transferirse desde el dispositivo de origen a un dispositivo de
 destino antes de transferir las imágenes de vídeo codificadas a las que pertenecen los conjuntos de datos de
 secuencia.

[0008] Los detalles de uno o más ejemplos se exponen en los dibujos adjuntos y en la descripción siguiente. Otras características, objetos y ventajas resultarán evidentes a partir de la descripción y de los dibujos, y a partir de las reivindicaciones.

5 BREVE DESCRIPCIÓN DE LOS DIBUJOS

[0009]

10 La FIG. 1 es un diagrama de bloques que ilustra un sistema de ejemplo en el que un dispositivo de origen de audio/vídeo (A/V) transfiere datos de audio y vídeo a un dispositivo de destino de A/V.

La FIG. 2 es un diagrama de bloques que ilustra componentes de una unidad de encapsulación de ejemplo.

15 La FIG. 3 es un diagrama de bloques que ilustra elementos de un archivo de vídeo de ejemplo que puede incluir mensajes de SEI a nivel de secuencia en diversas ubicaciones.

La FIG. 4 es un diagrama de bloques que ilustra elementos de otro archivo de vídeo de ejemplo, que incluye una pista de conjunto de parámetros que contiene mensajes de SEI a nivel de secuencia.

20 La FIG. 5 es un diagrama de bloques que ilustra un conjunto de presentaciones de ejemplo que incluye archivos del Proyecto de asociación de tercera generación (3GPP) que incluyen conjuntos de datos de secuencia respectivos para las presentaciones respectivas.

25 La FIG. 6 es un diagrama de flujo que ilustra un procedimiento de ejemplo para encapsular un conjunto de datos de secuencia dentro de un archivo de vídeo y proporcionar el archivo de vídeo desde un dispositivo de origen a un dispositivo de destino.

30 La FIG. 7 es un diagrama de bloques que ilustra otro ejemplo de un archivo de vídeo que incluye múltiples conjuntos de datos de secuencia.

DESCRIPCIÓN DETALLADA

35 **[0010]** En general, esta divulgación describe técnicas para admitir la entrega de datos de vídeo utilizando un protocolo de red de transmisión continua, como la transmisión continua del protocolo de transferencia de hipertexto (HTTP). En particular, las técnicas de esta divulgación están dirigidas a desacoplar ciertos mensajes de información de mejora complementaria (SEI), tales como mensajes de SEI a nivel de secuencia, de muestras de vídeo particulares al transmitir datos de vídeo, por ejemplo, en forma de un archivo de vídeo. Asimismo, las técnicas pueden incluir la interpretación de mensajes de SEI a nivel de secuencia en un dispositivo de destino para admitir la decodificación y representación de imágenes de vídeo codificadas.

45 **[0011]** La divulgación también describe el desacoplamiento de mensajes de SEI de las muestras de vídeo en el nivel de códec, donde los mensajes de SEI no necesariamente tienen que estar asociados con imágenes codificadas específicas, por ejemplo, cuando los mensajes de SEI están asociados con unidades de acceso. En caso de que los mensajes de SEI estén desacoplados de las imágenes codificadas, la información de temporización que especifica a qué duración es aplicable un mensaje de SEI puede señalizarse como parte del mensaje de SEI.

50 **[0012]** Los mensajes de SEI pueden contener información utilizada para ayudar a un decodificador y/o representador a decodificar y/o visualizar ciertas imágenes de vídeo codificadas. Los mensajes de SEI también se pueden usar para proporcionar resistencia a los errores. Algunos mensajes de SEI son a nivel de secuencia (por ejemplo, se refieren a un grupo codificado de imágenes (GOP) o a otra secuencia de imágenes de vídeo codificadas), mientras que otros pueden referirse a una imagen codificada particular. Los mensajes de SEI son típicamente transmitidos con una imagen codificada particular. Es decir, para recuperar un mensaje de SEI, un decodificador normalmente necesita recuperar la imagen codificada que incluye el mensaje de SEI. Cuando el decodificador no necesita los datos para la imagen codificada, en general conlleva un ancho de banda desaprovechado. Además, en algunos casos, dos mensajes de SEI para imágenes codificadas diferentes o dos GOP pueden contener fundamentalmente la misma información, de manera que la transmisión repetida de los mensajes de SEI también puede provocar una pérdida de ancho de banda.

60 **[0013]** Las técnicas de esta divulgación incluyen la transmisión de mensajes de SEI por separado (por ejemplo, de forma independiente) de datos de vídeo codificados, por ejemplo, donde los mensajes de SEI describen datos de secuencia para una secuencia de imágenes de vídeo codificadas. Los datos de secuencia pueden corresponder a datos que son comunes a la secuencia de imágenes de vídeo codificadas. En algunos ejemplos relacionados con H.264/AVC (codificación de vídeo avanzada), los mensajes de SEI pueden transmitirse junto con un flujo elemental de conjunto de parámetros cuando los mensajes de SEI incluyen información relacionada con rectángulos de barrido panorámico e información de vídeo estéreo. El flujo elemental de conjunto de parámetros puede incluir un conjunto de

parámetros de secuencia (SPS) y/o mensajes de conjunto de parámetros de imagen (PPS). Es decir, el flujo elemental de conjunto de parámetros puede incluir mensajes PPS y/o SPS además de uno o más mensajes de SEI, de acuerdo con estas técnicas. Un ejemplo de dicho mensaje de SEI en H.264/AVC es una disposición de empaquetado de tramas SEI, que especifica la información sobre cómo dos vistas de un par de imágenes se intercalan juntas en una trama.

5 Para mencionar otro ejemplo, con respecto a la codificación de vídeo ajustable a escala (SVC), los mensajes de SEI pueden enviarse en un flujo elemental de conjunto de parámetros cuando los mensajes de SEI proporcionan información de la ajustabilidad a escala. Como otro ejemplo más, en la codificación de vídeo multivista (MVC), los mensajes de SEI pueden incluirse en el flujo elemental de conjunto de parámetros cuando los mensajes de SEI incluyen información de ajustabilidad a escala de la vista, información de adquisición multivista y/o información de escena multivista.

[0014] Los mensajes de SEI a nivel de secuencia convencional se han asociado con ciertas unidades de acceso, tales como una unidad de acceso que comprende una imagen de actualización instantánea del descodificador (IDR). Sin embargo, las técnicas de esta divulgación están dirigidas a separar los mensajes de SEI a nivel de secuencia de las unidades de acceso. Por consiguiente, en lugar de recuperar una unidad de acceso que incluye un mensaje de SEI a nivel de secuencia, las técnicas de esta divulgación pueden permitir que un dispositivo cliente recupere los mensajes de SEI a nivel de secuencia por separado de las imágenes de vídeo codificadas descritas por los mensajes de SEI a nivel de secuencia. Por lo tanto, si el dispositivo cliente está configurado para descodificar y visualizar solo una porción de una secuencia de imágenes de vídeo codificadas, el dispositivo cliente puede recuperar solo la porción de la secuencia, sin necesidad de recuperar otras imágenes, como la imagen IDR con la cual, en caso contrario, los mensajes de SEI se incluirían. El dispositivo cliente puede recuperar los mensajes de SEI por separado de la porción deseada de las imágenes de vídeo codificadas, de acuerdo con las técnicas de esta divulgación.

[0015] En algunos ejemplos, en lugar de incluir mensajes de SEI a nivel de secuencia (y/u otros datos de secuencia) en una pista de conjunto de parámetros, los mensajes de SEI a nivel de secuencia pueden incluirse de forma continua en un archivo, de manera que una operación HTTP-Get pueda ser utilizada por un cliente para recuperar información que puede usarse para descodificar y visualizar un archivo de vídeo en cualquier momento. Es decir, un usuario puede buscar cualquier punto en un archivo de vídeo, y el descodificador puede recuperar los mensajes de SEI sin recuperar datos de vídeo codificados adicionales e innecesarios, por ejemplo, datos de vídeo codificados que preceden a una ubicación temporal dentro de un archivo de vídeo correspondiente a la petición de búsqueda y que no se utilizaría para descodificar datos de vídeo en y más allá de la ubicación temporal de la petición de búsqueda.

[0016] Una representación de vídeo o audio puede incluir uno o más segmentos. Cada representación puede incluir un segmento de inicialización, o cada segmento de medios en una representación puede ser autoinicializador. Un segmento de inicialización puede contener información de inicialización para acceder a datos de la representación correspondiente, por ejemplo, segmentos de medios descritos por el segmento de inicialización. Los segmentos de inicialización en general no contienen datos de medios, sino que pueden incluir solo datos de inicialización. Cada representación también puede incluir uno o más componentes de medios, donde cada componente de medios puede ser una versión codificada de un tipo individual de medios, tal como audio, vídeo o texto temporizado. Los componentes de medios pueden tener continuidad temporal entre fronteras de segmentos de medios consecutivos dentro de una representación.

[0017] Para mencionar aún otro ejemplo, las técnicas de esta divulgación pueden incluir el envío de datos de secuencia (por ejemplo, PPS, SPS y mensajes de SEI a nivel de secuencia) y, en particular, segmentos de archivo que no contienen datos de vídeo codificados, concretamente, muestras de vídeo. Un dispositivo de origen y un dispositivo de destino pueden configurarse con ubicaciones de las muestras de vídeo, incluidos los datos de secuencia, y/o con procedimientos para determinar las ubicaciones de estas muestras de vídeo. Los datos de secuencia pueden, en algunos ejemplos, enviarse desde un servidor a un cliente antes de enviar los datos de vídeo codificados. Es decir, un dispositivo de origen puede proporcionar datos de secuencia, como PPS, SPS y mensajes de SEI a nivel de secuencia, a un dispositivo cliente antes de enviar datos de vídeo codificados a los que hacen referencia los datos de secuencia. En la transmisión continua del HTTP basada en segmentos de archivo, los mensajes de SEI, por ejemplo, mensajes de SEI a nivel de secuencia, junto con los conjuntos de parámetros, como conjuntos de parámetros de secuencia y conjuntos de parámetros de imagen, se pueden colocar en un segmento de inicialización, que no contiene ninguna muestra codificada de audio o vídeo.

[0018] Las técnicas de esta divulgación se pueden aplicar a los archivos de vídeo que cumplen con cualquier formato de archivo de medios de base de la ISO, el formato de archivo de codificación de vídeo ajustable a escala (SVC), el formato de archivo de codificación de vídeo avanzada (AVC), el formato de archivo del Proyecto de asociación de tercera generación (3GPP) y/o el formato de archivo de codificación de vídeo multivista (MVC). El formato de archivo de medios de base de la ISO está diseñado para contener la información temporizada de medios para una presentación en un formato flexible y extensible que facilite el intercambio, la gestión, la edición y la presentación de los medios. El formato de archivo de medios de base de la ISO (ISO/IEC 14496-12:2004) se especifica en la norma MPEG-4 Parte-12, que define una estructura general para los archivos de medios basados en el tiempo. Se usa como la base para otros formatos de archivo en la familia, tales como el soporte definido del formato de archivo AVC (ISO/IEC 14496-15) para la compresión de vídeo de la norma AVC H.264/MPEG-4, el formato de archivo 3GPP, el formato de archivo SVC y el formato de archivo MVC. El formato de archivo 3GPP y el formato de archivo MVC son extensiones

del formato de archivo AVC. El formato de archivo de medios de base de la ISO contiene la temporización, la estructura y la información de medios para las secuencias temporizadas de datos de medios, tales como las presentaciones audiovisuales. La estructura del archivo está orientada a objetos. Un archivo puede descomponerse en objetos básicos muy sencillamente y la estructura de los objetos está implícita a partir de su tipo.

[0019] Los archivos que cumplen con el formato de archivo de medios de base de la ISO (y sus ampliaciones) pueden formarse como una serie de objetos, llamados "cuadros". Los datos en el formato de archivo de medios de base de la ISO pueden estar contenidos en cuadros y no es necesario que haya datos fuera de los cuadros dentro del archivo. Esto incluye cualquier rúbrica inicial requerida por el formato de archivo específico. Un "cuadro" puede ser un bloque de construcción orientado a objetos, definido por un identificador y una longitud de tipo único. Típicamente, una presentación está contenida en un archivo y la presentación de medios está autocontenida. El contenedor de la película (cuadro de película) contiene los metadatos de los medios, y las tramas de vídeo y audio pueden estar contenidas en el contenedor de datos de medios y podrían estar en otros archivos.

[0020] Una presentación (secuencia de movimiento) puede estar contenida en varios archivos. La información de temporización y entramado (posición y tamaño) se encuentra en general en el archivo de medios de base de la ISO y los archivos auxiliares pueden usar fundamentalmente cualquier formato. Esta presentación puede ser "local" al sistema que contenga la presentación, o puede proporcionarse a través de una red u otro mecanismo de suministro de transmisión continua.

[0021] Los archivos pueden tener una estructura lógica, una estructura de tiempo y una estructura física, y no es necesario que estas estructuras estén acopladas. La estructura lógica del archivo puede ser de una película o videoclip (que potencialmente incluye datos de vídeo y audio) que a su vez contiene un conjunto de pistas paralelas en el tiempo. La estructura temporal del archivo puede ser que las pistas contengan secuencias de muestras en el tiempo, y esas secuencias se asignen a la línea de tiempo de toda la película mediante listas de edición opcionales. La estructura física del archivo puede separar los datos necesarios para la descomposición lógica, temporal y estructural de las propias muestras de datos de medios. Esta información estructural puede concentrarse en un cuadro de película, posiblemente extendida en el tiempo por cuadros de fragmentos de película. El cuadro de película puede documentar las relaciones lógicas y de temporización de las muestras, y también puede contener punteros a donde están ubicados. Esos punteros pueden estar en el mismo archivo o en otro, por ejemplo, referenciado por una URL.

[0022] Cada flujo de medios puede estar contenido en una pista especializada para ese tipo de medios (audio, vídeo, etc.), y además puede estar parametrizado por una entrada de muestra. Cada una de las entradas de muestras puede contener un nombre de un tipo exacto de medios (por ejemplo, un tipo de decodificador necesario para decodificar el flujo) y cualquier parametrización necesaria para ese decodificador. El nombre también puede adoptar la forma de un código de cuatro caracteres, por ejemplo, "moov" o "trak". Existen formatos de entrada de muestra definidos no solo para los medios MPEG-4, sino también para los tipos de medios utilizados por otras organizaciones que usan esta familia de formatos de archivo.

[0023] El soporte para metadatos en general toma dos formas. Primero, los metadatos temporizados pueden almacenarse en una pista apropiada y sincronizarse según se desee con los datos de medios que están describiendo. En segundo lugar, puede haber soporte general para metadatos no temporizados adjuntos a la película o a una pista individual. El soporte estructural es general y permite el almacenamiento de recursos de metadatos en otro lugar del archivo o en otro archivo, de manera similar al almacenamiento de los datos medios, es decir, las imágenes de vídeo codificadas. Así mismo, estos recursos pueden nombrarse y pueden protegerse.

[0024] En el formato de archivo de medios de base de la ISO, una agrupación de muestras es una asignación de cada una de las muestras en una pista para ser un miembro de un grupo de muestras. No se requiere que las muestras en un grupo de muestras sean contiguas. Por ejemplo, cuando se presenta H.264/AVC en formato de archivo AVC, las muestras de vídeo en un nivel temporal se pueden muestrear en un grupo de muestra. Los grupos de muestras pueden estar representados por dos estructuras de datos: un cuadro MuestrearEnGrupo y un cuadro DescripciónGrupoMuestra. El cuadro MuestrearEnGrupo representa la asignación de muestras a grupos de muestras. Puede haber una instancia del cuadro DescripciónGrupoMuestra para cada entrada de grupo de muestra, para describir las propiedades del grupo correspondiente.

[0025] Se puede usar una pista de metadatos opcional para etiquetar cada pista con la "característica interesante" que tiene, para la cual su valor puede diferir de otros miembros del grupo (por ejemplo, su tasa de bits, tamaño de pantalla o idioma). Algunas muestras dentro de una pista pueden tener características especiales o pueden identificarse individualmente. Un ejemplo de la característica es el punto de sincronización (a menudo una trama-I de vídeo). Estos puntos pueden ser identificados por una tabla especial en cada pista. De manera más general, la naturaleza de las dependencias entre muestras de pista también se puede documentar utilizando metadatos. Los metadatos se pueden estructurar como una secuencia de muestras de formato de archivo, al igual que una pista de vídeo. Dicha pista puede denominarse una pista de metadatos. Cada muestra de metadatos se puede estructurar como una declaración de metadatos. Hay varios tipos de declaraciones, correspondientes a las diversas preguntas que se pueden formular acerca de la correspondiente muestra de formato de archivo o sus muestras constituyentes.

[0026] Cuando los medios se suministran mediante un protocolo de transmisión continua, los medios pueden necesitar transformarse a partir de la forma en que se representan en el archivo. Un ejemplo de esto es cuando los medios de comunicación se transmiten mediante el protocolo de transporte en tiempo real (RTP). En el archivo, por ejemplo, cada trama de vídeo se almacena de forma contigua como una muestra de formato de archivo. En el RTP, se deben obedecer las reglas de paquetización específicas del códec usado para colocar estas tramas en paquetes del RTP. Un servidor de transmisión continua puede configurarse para calcular dicha paquetización en tiempo de ejecución. Sin embargo, hay soporte para la asistencia de los servidores de transmisión continua. Se pueden colocar pistas especiales llamadas pistas de indicación en los archivos.

[0027] Las pistas de indicación contienen instrucciones generales para los servidores de transmisión continua respecto a cómo formar flujos de paquetes a partir de pistas de medios para un protocolo específico. Puesto que la forma de estas instrucciones es independiente de los medios, es posible que no sea necesario revisar los servidores cuando se introducen nuevos códecs. Así mismo, el software de codificación y edición puede no tener conocimiento de los servidores de transmisión continua. Una vez que finaliza la edición de un archivo, se puede usar un software llamado indicador que añade pistas de indicaciones en el archivo, antes de colocarlo en un servidor de transmisión continua. Como ejemplo, hay un formato de pista de indicación definido para los flujos de RTP en la memoria descriptiva del formato de archivo MPEG-4.

[0028] El término "descarga progresiva" se usa para describir la transferencia de archivos de medios digitales de un servidor a un cliente, típicamente usando el protocolo HTTP. Cuando se inicia desde un ordenador, el ordenador puede comenzar la reproducción de los medios antes de que finalice la descarga. Una diferencia entre la transmisión continua de los medios y la descarga progresiva radica en cómo el dispositivo de usuario final que accede a los medios digitales recibe y almacena los datos de los medios digitales. Un reproductor de medios que es capaz de reproducir la descarga progresivamente se basa en que los metadatos ubicados en la cabecera del archivo estén intactos y en un búfer local del archivo de medios digitales a medida que se descarga de un servidor web. En el momento en que una cantidad especificada de datos almacenados en el búfer está disponible para el dispositivo de reproducción local, el dispositivo puede comenzar a reproducir los medios. El productor del contenido puede incorporar esta cantidad especificada de datos almacenados en búfer en el archivo en los ajustes del codificador y puede reforzarse con ajustes adicionales del búfer impuestos por el reproductor de medios del ordenador cliente.

[0029] AVC y 3GPP son extensiones del formato de archivo de medios de base de la ISO, mientras que SVC y MVC son extensiones del formato de archivo AVC. Por consiguiente, las técnicas de esta divulgación pueden aplicarse con respecto a los archivos de vídeo que se ajustan al formato de archivo de medios de base de la ISO, el formato de archivo AVC y extensiones de los mismos, por ejemplo, SVC y MVC, y/o formato de archivo del Proyecto de asociación de tercera generación (3GPP). Las técnicas pueden aplicarse además a estas y otras extensiones de estos formatos, y pueden aplicarse además para extender otros formatos de archivo para proporcionar mensajes de SEI por separado de los datos de vídeo que describen los mensajes de SEI. Por tanto, aunque en general se describe con respecto a AVC con fines de ejemplo, debe entenderse que las técnicas de esta divulgación pueden aplicarse en general a cualquier otro formato de archivo.

[0030] El Grupo de expertos en imágenes en movimiento (MPEG) ha desarrollado un formato de archivo AVC, como una extensión del formato de archivo de medios de base de la ISO. En el formato de archivo AVC, los conjuntos de parámetros, incluidos el conjunto de parámetros de secuencia y los conjuntos de parámetros de imagen, pueden asociarse con el flujo de vídeo elemental, que se encuentra en la pista de vídeo. También es posible tener los conjuntos de parámetros en otra pista, llamada pista de conjunto de parámetros, que incluye el flujo elemental de conjunto de parámetros que contiene las muestras que son las unidades de capa de abstracción de red (NAL) de SPS o PPS. Sin embargo, la ubicación de las unidades NAL de mensajes de SEI no está restringida en el formato de archivo AVC, siempre y cuando se presente cualquier mensaje de SEI en una muestra con un tiempo de descodificación antes de cualquier imagen que use este mensaje de SEI.

[0031] El formato de archivo AVC tiene las extensiones SVC y MVC para admitir el almacenamiento de SVC y MVC, respectivamente. En SVC y MVC, se pueden contener múltiples capas/vistas en una pista de vídeo o en múltiples pistas de vídeo. Existen herramientas definidas en las memorias descriptivas de SVC y MVC para construir puntos de funcionamiento cuando las capas/vistas se presentan en diversas organizaciones.

[0032] Algunas funcionalidades de SVC se heredan de H.264/AVC. En comparación con las normas anteriores ajustables a escala, la norma SVC puede proporcionar diversas ventajas, como, por ejemplo, ajustabilidad a escala temporal jerárquica, predicción entre capas, descodificación de un solo bucle e interfaz de transporte flexible. H.264/AVC proporciona una estructura de codificación de imagen B jerárquica y flexible que permite una ajustabilidad a escala temporal avanzada. Con esta característica heredada de H.264/AVC, SVC admite la ajustabilidad a escala temporal para capas con diferentes resoluciones. En SVC, un grupo de imágenes (GOP) incluye una denominada imagen clave, y todas las imágenes que están situadas en el orden de salida/visualización entre esta imagen clave y la imagen clave anterior. Una imagen clave puede estar codificada en intervalos regulares o irregulares, y puede ser bien intracodificada o bien intercodificada, usando la imagen clave anterior como una referencia para la predicción compensada por movimiento. Las imágenes no claves pueden predecirse jerárquicamente a partir de imágenes con niveles temporales inferiores, y la imagen clave puede tener el nivel temporal más bajo. El nivel temporal de una

imagen puede indicarse mediante el elemento de sintaxis temporal_id en la extensión SVC de la cabecera de la unidad NAL.

5 **[0033]** SVC introduce la predicción entre capas para las ajustabilidades a escala espaciales y de relación señal-ruido (SNR) basadas en textura, residuos y movimiento. La ajustabilidad espacial a escala en SVC puede generalizarse para cualquier relación de resolución entre dos capas. La ajustabilidad a escala de la SNR puede ser realizada por la ajustabilidad a escala de granularidad gruesa (CGS), la ajustabilidad a escala de granularidad media (MGS) o la ajustabilidad a escala de grano fino (FGS). En SVC, dos capas espaciales o de CGS pertenecen a distintas capas de dependencia (indicadas mediante id_dependencia en la cabecera de la unidad NAL), mientras que dos capas de MGS pueden estar en la misma capa de dependencia. Una capa de dependencia incluye capas de calidad con id_calidad desde 0 hasta valores más altos, correspondientes a capas de mejora de la calidad. SVC proporciona procedimientos de predicción entre capas que pueden utilizarse para reducir la redundancia entre capas. Se resumen como predicción de textura entre capas, predicción residual entre capas y predicción de movimiento entre capas.

15 **[0034]** SVC también proporciona sistemas flexibles y diseños de interfaz de transporte que pueden permitir la integración sin fisuras del códec a los sistemas de aplicación multimedia ajustables a escala. Además de la disposición de compresión y ajustabilidad a escala, los sistemas y la interfaz de transporte se centran en las funcionalidades del códec, como, por ejemplo, para el códec de vídeo en general, la interoperabilidad y la conformidad, la extensibilidad, el acceso aleatorio, la temporización, la gestión del búfer, así como resistencia a los errores y para la codificación ajustable a escala en particular, la compatibilidad con versiones anteriores, disposición de información de ajustabilidad a escala y adaptación de la ajustabilidad a escala.

25 **[0035]** SVC además proporciona un esquema de descodificación de un solo bucle. En el esquema de descodificación de un solo bucle, solamente la capa objetivo necesita ser compensada por movimiento y completamente descodificada. Por lo tanto, en comparación con otros esquemas de descodificación de bucle múltiple, donde la compensación de movimiento y la descodificación completa se realizan típicamente para cada capa espacial o de SNR ajustable a escala, el esquema de descodificación de un solo bucle de SVC puede reducir en gran medida la complejidad de la descodificación, así como el tamaño de DPB.

30 **[0036]** SVC y MVC tienen diseños de sintaxis de alto nivel similares, por ejemplo, en la cabecera de la unidad NAL, el conjunto de parámetros de secuencia y los mensajes de SEI. Tanto en SVC como en MVC, la extracción de todo el flujo de bits se admite hasta cierto punto para la adaptación del ancho de banda o la capacidad del dispositivo. Los subflujos de bits extraídos corresponden a diferentes puntos de funcionamiento.

35 **[0037]** En algunos ejemplos, las técnicas de esta divulgación pueden aplicarse a códecs H.264/AVC o códecs basados en AVC, como SVC, MVC u otras extensiones de H.264/AVC. Dichos códecs pueden configurarse para reconocer mensajes de SEI cuando los mensajes de SEI están asociados dentro de una imagen codificada, por ejemplo, una unidad de acceso, donde el mensaje de SEI puede encapsularse dentro de la unidad de acceso en un formato de archivo de medios de base de la ISO o flujo de bits de sistemas MPEG-2. Las técnicas también se ocupan de futuras normas de codificación, por ejemplo, H.265/HEVC, para permitir que los mensajes de SEI se transmitan por separado de cualquier unidad de acceso descrita por los mensajes de SEI.

45 **[0038]** Con respecto a 3GPP para mencionar otro ejemplo, el transporte HTTP/TCP/IP se admite en archivos 3GP para la descarga y descarga progresiva. Además, el uso de HTTP para la transmisión continua de vídeo puede proporcionar algunas ventajas, y los servicios de transmisión continua de vídeo basados en HTTP se están volviendo populares. La transmisión continua del HTTP puede proporcionar ciertas ventajas, incluido el uso de los componentes y protocolos de Internet existentes, de manera que no se necesitan nuevos esfuerzos para desarrollar nuevas técnicas para transportar datos de vídeo a través de una red. Otros protocolos de transporte, por ejemplo, el formato de carga útil RTP, requieren dispositivos de red intermedios, por ejemplo, cuadros intermedios, para conocer el formato de los medios y el contexto de señalización. Además, la transmisión continua del HTTP puede ser dirigida por el cliente, lo que puede evitar problemas de control.

55 **[0039]** Por ejemplo, para aprovechar las características para obtener un rendimiento óptimo, el servidor puede realizar un seguimiento del tamaño y el contenido de los paquetes que aún no se han confirmado. El servidor también puede analizar la estructura del archivo y reconstruir el estado del búfer del cliente para tomar decisiones óptimas de conmutación/baja resolución de RD. Así mismo, las restricciones en las variaciones de flujo de bits pueden cumplirse con el fin de seguir cumpliendo con los perfiles negociados. HTTP no requiere necesariamente nuevas implementaciones de hardware o software en un servidor web que tenga implementado HTTP 1.1. La transmisión continua del HTTP también proporciona facilidad de manejo del TCP y cruce de cortafuegos.

60 **[0040]** En la transmisión continua del HTTP, las operaciones usadas con frecuencia incluyen el GET y el GET parcial. La operación GET recupera un archivo completo asociado a un localizador uniforme de recursos (URL) o a un nombre de recurso uniforme (URN) dado. La operación GET parcial recibe un intervalo de bytes como parámetro de entrada y recupera un número continuo de bytes de un archivo correspondiente al intervalo de bytes recibido. Por tanto, se pueden proporcionar fragmentos de película para la transmisión continua del HTTP, porque una operación GET parcial puede obtener uno o más fragmentos de película individuales. Obsérvese que, en un fragmento de película, puede

haber varios fragmentos de pista de diferentes pistas. En la transmisión continua del HTTP, una presentación de medios puede ser una recopilación de datos estructurados que es accesible para el cliente. El cliente puede solicitar y descargar la información de datos de medios para presentar un servicio de transmisión continua a un usuario.

5 **[0041]** La FIG. 1 es un diagrama de bloques que ilustra un sistema de ejemplo 10 en el que un dispositivo de origen de audio/vídeo (A/V) 20 transporta datos de audio y vídeo a un dispositivo de destino de A/V 40. El sistema 10 de la FIG. 1 puede corresponder a un sistema de teleconferencia por vídeo, un sistema de cliente/servidor, un sistema de difusor/receptor o cualquier otro sistema en el que se envíen datos de vídeo desde un dispositivo de origen, tal como un dispositivo de origen de A/V 20, a un dispositivo de destino, tal como un dispositivo de destino de A/V 40. En algunos ejemplos, el dispositivo de origen de A/V 20 y el dispositivo de destino de A/V 40 pueden realizar el intercambio de información bidireccional. Es decir, el dispositivo de origen de A/V 20 y el dispositivo de destino de A/V 40 pueden ser capaces tanto de codificar como de decodificar (y transmitir y recibir) datos de audio y vídeo. En algunos ejemplos, el codificador de audio 26 puede comprender un codificador de voz, también denominado vocodificador.

15 **[0042]** El dispositivo de origen de A/V 20, en el ejemplo de la FIG. 1, comprende una fuente de audio 22 y una fuente de vídeo 24. La fuente de audio 22 puede comprender, por ejemplo, un micrófono que produce señales eléctricas representativas de datos de audio captados que el codificador de audio 26 va a codificar. De forma alternativa, la fuente de audio 22 puede comprender un medio de almacenamiento que almacena datos de audio previamente registrados, un generador de datos de audio tal como un sintetizador informatizado, o cualquier otra fuente de datos de audio. La fuente de vídeo 24 puede comprender una cámara de vídeo que produce datos de vídeo que han de ser codificados por el codificador de vídeo 28, un medio de almacenamiento codificado con datos de vídeo previamente grabados, una unidad de generación de datos de vídeo o cualquier otra fuente de datos de vídeo.

25 **[0043]** Los datos de audio y vídeo no procesados pueden comprender datos analógicos o digitales. Los datos analógicos se pueden digitalizar antes de que el codificador de audio 26 y/o el codificador de vídeo 28 los codifiquen. La fuente de audio 22 puede obtener datos de audio a partir de un participante que habla mientras el participante que habla está hablando, y la fuente de vídeo 24 puede obtener simultáneamente datos de vídeo del participante que habla. En otros ejemplos, la fuente de audio 22 puede comprender un medio de almacenamiento legible por ordenador que comprende datos de audio almacenados, y la fuente de vídeo 24 puede comprender un medio de almacenamiento legible por ordenador que comprende datos de vídeo almacenados. De esta manera, las técnicas descritas en esta divulgación se pueden aplicar a la transmisión continua en directo y en tiempo real de datos de audio y vídeo, o de datos de audio y vídeo archivados y preregistrados.

35 **[0044]** Las tramas de audio que corresponden a tramas de vídeo son, en general, tramas de audio que contienen datos de audio que fueron captados por la fuente de audio 22 junto con los datos de vídeo captados por la fuente de vídeo 24 que están contenidos dentro de las tramas de vídeo. Por ejemplo, mientras un participante que habla en general produce, al hablar, datos de audio, la fuente de audio 22 capta los datos de audio, y la fuente de vídeo 24 capta los datos de vídeo del participante que habla al mismo tiempo, es decir, mientras la fuente de audio 22 está captando los datos de audio. Así pues, una trama de audio puede corresponder temporalmente a una o más tramas de vídeo en particular. Por consiguiente, una trama de audio correspondiente a una trama de vídeo corresponde en general a una situación en la que se han captado datos de audio y datos de vídeo al mismo tiempo, y para la que una trama de audio y una trama de vídeo comprenden, respectivamente, los datos de audio y los datos de vídeo que se han captado al mismo tiempo.

45 **[0045]** En algunos ejemplos, el codificador de audio 26 puede codificar una marca de tiempo en cada trama de audio codificada, que representa un tiempo en el que se han registrado los datos de audio para la trama de audio codificada y, de forma similar, el codificador de vídeo 28 puede codificar una marca de tiempo en cada trama de vídeo codificada, que representa un tiempo en el que se han registrado los datos de vídeo para la trama de vídeo codificada. En dichos ejemplos, una trama de audio correspondiente a una trama de vídeo puede comprender una trama de audio que comprende una marca de tiempo y una trama de vídeo que comprende la misma marca de tiempo. El dispositivo de origen de A/V 20 puede incluir un reloj interno a partir del cual el codificador de audio 26 y/o el codificador de vídeo 28 pueden generar las marcas de tiempo, o que la fuente de audio 22 y la fuente de vídeo 24 pueden usar para asociar datos de audio y vídeo, respectivamente, con una marca de tiempo.

55 **[0046]** En algunos ejemplos, la fuente de audio 22 puede enviar datos al codificador de audio 26, correspondientes a un tiempo en el que se han registrado los datos de audio, y la fuente de vídeo 24 puede enviar datos al codificador de vídeo 28, correspondientes a un tiempo en el que se han registrado los datos de vídeo. En algunos ejemplos, el codificador de audio 26 puede codificar un identificador de secuencia en unos datos de audio codificados para indicar un orden temporal relativo de los datos de audio codificados, pero sin indicar necesariamente un tiempo absoluto en el que se han registrado los datos de audio y, de forma similar, el codificador de vídeo 28 también puede usar identificadores de secuencia para indicar un orden temporal relativo de los datos de vídeo codificados. De forma similar, en algunos ejemplos, un identificador de secuencia se puede asignar a, o en cualquier caso correlacionar con, una marca de tiempo.

65 **[0047]** Las técnicas de esta divulgación son en general dirigidas al transporte de datos de multimedia (por ejemplo, audio y vídeo) codificados, y a la recepción y la posterior interpretación y decodificación de los datos de multimedia

transportados. Como se muestra en el ejemplo de la FIG. 1, la fuente de vídeo 24 puede proporcionar una pluralidad de vistas de una escena al codificador de vídeo 28.

[0048] El dispositivo de origen de A/V 20 puede proporcionar un "servicio" al dispositivo de destino de A/V 40. Un servicio en general corresponde a un subconjunto de vistas disponibles de datos de MVC. Por ejemplo, los datos de MVC pueden estar disponibles para ocho vistas, ordenadas de cero a siete. Un servicio puede corresponder a vídeo estéreo que tiene dos vistas, mientras que otro servicio puede corresponder a cuatro vistas, y otro servicio más puede corresponder a las ocho vistas. En general, un servicio corresponde a cualquier combinación (es decir, a cualquier subconjunto) de las vistas disponibles. Un servicio también puede corresponder a una combinación de vistas disponibles, así como datos de audio. Un punto de operación puede corresponder a un servicio, de manera que el dispositivo de origen de A/V 20 pueda proporcionar además un descriptor de punto de funcionamiento para cada servicio proporcionado por el dispositivo de origen de A/V 20.

[0049] Cada flujo individual de datos (ya sea de audio o vídeo) se denomina un flujo elemental. Un flujo elemental es un componente único, codificado digitalmente (posiblemente comprimido) de un programa. Por ejemplo, la parte de vídeo o audio codificado del programa puede ser un flujo elemental. Un flujo elemental se puede convertir en un flujo elemental paquetizado (PES) antes de encapsularse dentro de un archivo de vídeo. Dentro del mismo programa, se utiliza un identificador de flujo para distinguir los paquetes PES pertenecientes a un flujo elemental del otro. La unidad básica de datos de un flujo elemental es un paquete de flujo elemental paquetizado (PES). Por tanto, cada vista de datos de vídeo de MVC corresponde a respectivos flujos elementales. De forma similar, los datos de audio corresponden a uno o más flujos elementales respectivos. Un conjunto de datos de secuencia, que puede incluir, por ejemplo, un conjunto de parámetros de secuencia, un conjunto de parámetros de imagen y mensajes de SEI a nivel de secuencia, puede incluirse en una pista de conjunto de parámetros o un flujo elemental de conjunto de parámetros, de acuerdo con las técnicas de esta divulgación. Las técnicas de esta divulgación se pueden aplicar a archivos de vídeo o transmisiones de vídeo que se ajusten a los sistemas MPEG-2, que incluyen flujo elemental paquetizado (PES), flujo de programa (PS) y flujo de transporte (TS), y sus extensiones AVC, SVC y MVC.

[0050] Una secuencia de vídeo de MVC codificada puede separarse en varios subflujos de bits, cada uno de los cuales es un flujo elemental. Cada subflujo de bits puede identificarse utilizando un subconjunto de los id_vista de la MVC. Basándose en el concepto de cada subconjunto de los id_vista de la MVC, se define un subflujo de bits de vídeo de la MVC. Un subflujo de bits de vídeo de MVC contiene las unidades de NAL de las vistas enumeradas en el subconjunto de los id_vista de la MVC. Un flujo de programa contiene en general solo las unidades de NAL que son de las de los flujos elementales. También se diseña que dos flujos elementales cualesquiera no puedan contener una vista idéntica.

[0051] En el ejemplo de la FIG. 1, la unidad de encapsulación 30 recibe flujos elementales que comprenden datos de vídeo desde el codificador de vídeo 28 y flujos elementales que comprenden datos de audio desde el codificador de audio 26. En algunos ejemplos, el codificador de vídeo 28 y el codificador de audio 26 pueden incluir, cada uno, paquetizadores para formar paquetes PES a partir de datos codificados. En otros ejemplos, el codificador de vídeo 28 y el codificador de audio 26 pueden interactuar, cada uno, con los paquetizadores respectivos para formar paquetes PES a partir de datos codificados. En otros ejemplos más, la unidad de encapsulación 30 puede incluir paquetizadores para formar paquetes PES a partir de datos de audio y de vídeo codificados.

[0052] Un "programa", como se utiliza en esta divulgación, puede comprender una combinación de datos de audio y datos de vídeo, por ejemplo, un flujo elemental de audio y un subconjunto de vistas disponibles suministradas por un servicio del dispositivo de origen de A/V 20. Cada paquete PES incluye un id_flujo que identifica el flujo elemental al que pertenece el paquete PES. La unidad de encapsulación 30 es responsable de ensamblar flujos elementales en un archivo de vídeo.

[0053] La unidad de encapsulación 30 recibe paquetes PES para flujos elementales de un programa desde el codificador de audio 26 y el codificador de vídeo 28 y forma las correspondientes unidades de capa de abstracción de red (NAL) a partir de los paquetes PES. En el ejemplo de la H.264/AVC (codificación de vídeo avanzada), los segmentos de vídeo codificados están organizados en unidades de NAL, que proporcionan una representación de vídeo "apta para redes" que aborda aplicaciones tales como la videotelefonía, el almacenamiento, la radiodifusión o la transmisión continua. Las unidades NAL se pueden clasificar en unidades NAL de capa de codificación de vídeo (VCL) y unidades NAL no VCL. Las unidades VCL pueden contener el motor de compresión central y pueden incluir datos a nivel de bloque, macrobloque y/o fragmento. Otras unidades NAL pueden ser unidades NAL no VCL. En algunos ejemplos, una imagen codificada en una instancia de tiempo, normalmente presentada como una imagen codificada primaria, puede estar contenida en una unidad de acceso, que puede incluir una o más unidades NAL.

[0054] Las unidades NAL no VCL pueden incluir unidades NAL de conjunto de parámetros y unidades NAL SEI, entre otras. Los conjuntos de parámetros contienen información de cabecera a nivel de secuencia (en conjuntos de parámetros de secuencia (SPS)) y la información de cabecera a nivel de imagen que cambia ocasionalmente (en conjuntos de parámetros de imagen (PPS)). Con los conjuntos de parámetros (por ejemplo, PPS y SPS), la información que cambia ocasionalmente no necesita ser repetida para cada secuencia o imagen, de ahí que pueda mejorarse la eficacia de la codificación. Además, el uso de conjuntos de parámetros puede permitir la transmisión fuera de banda

de la información de cabecera importante, evitando la necesidad de transmisiones redundantes para la resistencia a los errores. En los ejemplos de transmisión fuera de banda, las unidades NAL de conjunto de parámetros se pueden transmitir en un canal diferente al de otras unidades NAL, tales como las unidades NAL SEI.

5 **[0055]** La información de mejora complementaria (SEI) puede contener información que no es necesaria para descodificar las muestras de imágenes codificadas a partir de las unidades NAL VCL, pero puede ayudar en los procesos relacionados con la descodificación, visualización, resistencia a los errores y otros propósitos. Los mensajes de SEI pueden estar contenidos en unidades NAL no VCL. Los mensajes de SEI son la parte normativa de algunas memorias descriptivas habituales y, por tanto, no siempre son obligatorios para la implementación de descodificadores que cumplen las normas. Los mensajes de SEI pueden ser mensajes de SEI a nivel de secuencia o mensajes de SEI a nivel de imagen. Parte de la información a nivel de secuencia puede estar contenida en mensajes de SEI, tales como mensajes de SEI de información de ajustabilidad a escala en el ejemplo de SVC y mensajes de SEI de información de ajustabilidad a escala de la vista en MVC. Estos ejemplos de mensajes de SEI pueden transmitir información, por ejemplo, sobre extracción de puntos de funcionamiento y características de los puntos de funcionamiento.

15 **[0056]** De acuerdo con las técnicas de esta divulgación, la unidad de encapsulación 30 puede organizar los datos de un archivo de vídeo de manera que los mensajes de SEI a nivel de secuencia se desacoplan de las imágenes de vídeo codificadas descritas por los mensajes de SEI a nivel de secuencia. De esta manera, la unidad de encapsulación 30 puede desacoplar mensajes de SEI a nivel de secuencia de unidades de acceso específicas que contienen datos de vídeo codificados. Por ejemplo, la unidad de encapsulación 30 puede incluir mensajes de SEI a nivel de secuencia en un flujo elemental de conjunto de parámetros, junto con datos SPS y PPS, por ejemplo, para datos de vídeo codificados de acuerdo con AVC o extensiones de los mismos, tales como MVC y SVC. Así mismo, para ejemplos en los que los datos de vídeo se codifican de acuerdo con MVC y/o SVC, la unidad de encapsulación 30 puede incluir SPS de subconjunto en el flujo elemental de conjunto de parámetros.

25 **[0057]** En algunos ejemplos, la unidad de encapsulación 30 puede configurarse para colocar solo ciertos mensajes de SEI en un flujo elemental de conjunto de parámetros. Por ejemplo, para H.264/AVC, la unidad de encapsulación 30 puede incluir mensajes de SEI que describen el rectángulo de barrido panorámico y la información de vídeo estéreo en el flujo elemental de conjunto de parámetros. Para mencionar otro ejemplo, para SVC, la unidad de encapsulación 30 puede incluir mensajes de SEI que describen información de ajustabilidad a escala en el flujo elemental de conjunto de parámetros. Para mencionar aún otro ejemplo, en MVC, la unidad de encapsulación 30 puede incluir mensajes de SEI que describen información de ajustabilidad a escala de la vista, información de adquisición multivista e información de escena multivista.

35 **[0058]** En general, los SPS y PPS (y los subconjuntos de SPS, por ejemplo, para MVC y SVC), junto con los mensajes de SEI a nivel de secuencia, pueden denominarse colectivamente como un "conjunto de datos de secuencia". Como se analiza anteriormente, el conjunto de datos de secuencia puede incluirse en una pista separada, tal como el flujo elemental de conjunto de parámetros. En algunos ejemplos, en lugar de colocar el conjunto de datos de secuencia en una pista de parámetros separada, la unidad de encapsulación 30 puede presentar el conjunto de datos de secuencia para una secuencia de forma continua en un archivo, de manera que el dispositivo de destino 40 pueda recuperar todo el conjunto de datos de secuencia usando una petición GET HTTP parcial. Por ejemplo, el conjunto de datos de secuencia puede estar contenido en una muestra de vídeo especializada, que no contiene datos de vídeo codificados, por ejemplo, porciones de imágenes de vídeo codificadas o enteras.

45 **[0059]** Al separar el conjunto de datos de secuencia de los datos de vídeo descritos por el conjunto de datos de secuencia, el dispositivo de destino 40 puede recuperar el conjunto de datos de secuencia por separado de los datos de vídeo descritos por el conjunto de datos de secuencia. De esta manera, el dispositivo de origen 20 puede transmitir el conjunto de datos de secuencia al dispositivo de destino 40 antes de transmitir los datos de vídeo descritos por el conjunto de datos de secuencia al dispositivo de destino 40. Por tanto, el dispositivo de origen 20 puede enviar el conjunto de datos de secuencia al dispositivo de destino 40 antes de que un fragmento de película que contiene una imagen de vídeo codificada se envíe al dispositivo de destino 40.

50 **[0060]** En algunos ejemplos, por ejemplo, con respecto al uso de archivos de vídeo 3GPP, el conjunto de datos de secuencia puede incluirse en un archivo 3GPP separado, que puede enviarse al dispositivo de destino 40 después de que el dispositivo de destino 40 acceda a un descriptor de presentación de medios. Es decir, el dispositivo de destino 40 puede recuperar el descriptor de presentación de medios para una secuencia de datos de vídeo, luego recuperar un archivo 3GPP que incluye el conjunto de datos de secuencia para la secuencia de datos de vídeo, y luego recuperar la propia secuencia de datos de vídeo. En algunos ejemplos, el conjunto de datos de secuencia puede formar parte del propio descriptor de presentación de medios. En algunos ejemplos, el conjunto de datos de secuencia se puede incluir en parte de un cuadro de película que no contiene datos de vídeo codificados. En algunos ejemplos, el conjunto de datos de secuencia puede formar un fragmento de película que se coloca después del cuadro de película.

60 **[0061]** La unidad de encapsulación 30 puede formar unidades de NAL que comprenden una cabecera que identifica un programa al cual pertenece la NAL, así como una carga útil, por ejemplo, datos de audio, datos de vídeo o datos que describen el flujo de transporte o de programa al cual corresponde la unidad de NAL. Por ejemplo, en H.264/AVC, una unidad NAL incluye una cabecera de 1 byte y una carga útil de tamaño variable. En un ejemplo, una cabecera de

65

unidad de NAL comprende un elemento `id_prioridad`, un elemento `id_temporal`, un elemento `indicador_img_ancla`, un elemento `id_vista`, un elemento `indicador_no_idr` y un elemento `indicador_entre_vistas`. En la MVC convencional, se conserva la unidad NAL definida por la H.264, excepto para las unidades NAL de prefijo y las unidades NAL de fragmento codificado por MVC, que incluyen una cabecera de unidad NAL de MVC de 4 bytes y la carga útil de la unidad NAL.

[0062] El elemento `id_prioridad` de una cabecera de NAL se puede utilizar para un proceso sencillo de adaptación de flujo de bits de una ruta. El elemento `id_temporal` puede usarse para especificar el nivel temporal de la unidad de NAL correspondiente, donde diferentes niveles temporales corresponden a diferentes frecuencias de trama.

[0063] El elemento `indicador_img_ancla` puede indicar si una imagen es una imagen anclada o una imagen no anclada. Las imágenes ancladas y todas las imágenes que suceden a estas en el orden de salida (es decir, el orden de visualización) se pueden descodificar correctamente sin descodificar las imágenes anteriores en el orden de descodificación (es decir, el orden del flujo de bits) y, por tanto, pueden utilizarse como puntos de acceso aleatorio. Las imágenes ancladas y las imágenes no ancladas pueden tener dependencias diferentes, ambas señalizadas en el conjunto de parámetros de secuencia. Otros indicadores serán expuestos y usados en las siguientes secciones de este capítulo. Una imagen de anclaje de ese tipo también puede denominarse punto de acceso de GOP (Grupo de imágenes) abierto, mientras que también se admite un punto de acceso de GOP cerrado cuando el elemento `indicador_no_idr` es igual a cero. El elemento `indicador_no_idr` indica si una imagen es una imagen de actualización instantánea del descodificador (IDR) o de IDR de vista (V-IDR). En general, una imagen de IDR y todas las imágenes que la suceden en orden de salida u orden de flujo de bits, se pueden descodificar correctamente sin la descodificación de imágenes anteriores, ya sea en orden de descodificación o en orden de visualización.

[0064] El elemento `id_vista` puede comprender información sintáctica que se puede utilizar para identificar una vista, que puede ser utilizada para la interactividad de datos dentro de un descodificador de la MVC, por ejemplo, para la predicción entre vistas, y fuera de un descodificador, por ejemplo, para la representación. El elemento `indicador_entre_vistas` puede especificar si la unidad de NAL correspondiente es utilizada por otras vistas para la predicción entre vistas. Para transmitir la información de cabecera de la unidad de NAL de 4 bytes para una vista base, que puede ser compatible con la AVC, se define una unidad de prefijo de NAL en la MVC. En el contexto de la MVC, la unidad de acceso de vista base incluye las unidades de NAL de la VCL de la instancia de tiempo actual de la vista, así como su unidad de prefijo de NAL, que contiene solamente la cabecera de la unidad de NAL. Un descodificador de H.264/AVC puede ignorar la unidad de prefijo de NAL.

[0065] Una unidad NAL que incluye datos de vídeo en su carga útil puede comprender diversos niveles de granularidad de datos de vídeo. Por ejemplo, una unidad de NAL puede comprender un bloque de datos de vídeo, un macrobloque, una pluralidad de macrobloques, un fragmento de datos de vídeo o una trama completa de datos de vídeo. La unidad de encapsulación 30 puede recibir datos de vídeo codificados desde el codificador de vídeo 28 en forma de paquetes PES de flujos elementales. La unidad de encapsulación 30 puede asociar cada flujo elemental a un programa correspondiente.

[0066] La unidad de encapsulación 30 también puede ensamblar unidades de acceso desde una pluralidad de unidades NAL. En general, una unidad de acceso puede comprender una o más unidades NAL para representar una trama de datos de vídeo, así como datos de audio correspondientes a la trama cuando dichos datos de audio están disponibles. Una unidad de acceso incluye en general todas las unidades NAL para una instancia de tiempo de salida, por ejemplo, todos los datos de audio y vídeo para una instancia de tiempo. Por ejemplo, si cada vista tiene una frecuencia de trama de 20 tramas por segundo (fps), entonces cada instancia de tiempo puede corresponder a un intervalo de tiempo de 0,05 segundos. Durante este intervalo de tiempo, las tramas específicas para todas las vistas de la misma unidad de acceso (la misma instancia de tiempo) se pueden representar simultáneamente. En un ejemplo correspondiente a la H.264/AVC, una unidad de acceso puede comprender una imagen codificada en una instancia de tiempo, que puede presentarse como una imagen codificada primaria. Por consiguiente, una unidad de acceso puede comprender todas las tramas de audio y vídeo de una instancia temporal común, por ejemplo, todas las vistas correspondientes al tiempo X. Esta divulgación también se refiere a una imagen codificada de una vista particular como un "componente de vista". Es decir, un componente de vista puede comprender una imagen (o trama) codificada para una vista en particular en un tiempo en particular. Por consiguiente, se puede definir una unidad de acceso que comprende todos los componentes de vista de una instancia temporal común. El orden de descodificación de las unidades de acceso no necesita ser el mismo que el orden de salida o de visualización.

[0067] Como ocurre con la mayoría de normas de codificación de vídeo, la norma H.264/AVC define la sintaxis, la semántica y el proceso de descodificación para flujos de bits libres de errores, cualquiera de los cuales es conforme a un cierto perfil o nivel. La norma H.264/AVC no especifica el codificador, pero el codificador tiene la tarea de garantizar que los flujos de bits generados sean compatibles con la norma para un descodificador. En el contexto de la norma de codificación de vídeo, un "perfil" corresponde a un subconjunto de algoritmos, características o herramientas y restricciones que se les aplican. Como se define en la norma H.264, por ejemplo, un "perfil" es un subconjunto de toda la sintaxis del flujo de bits especificada por la norma H.264. Un "nivel" corresponde a las limitaciones del consumo de recursos del descodificador, tales como, por ejemplo, memoria de descodificador y cálculo, que están relacionadas con la resolución de las imágenes, tasa de bits y tasa de procesamiento de macrobloques (MB). Un perfil se puede

señalizar con un valor de perfil_idc (indicador de perfil), mientras que un nivel se puede señalar con un valor de nivel_idc (indicador de nivel).

[0068] La norma H.264, por ejemplo, reconoce que, dentro de los límites impuestos por la sintaxis de un perfil dado, todavía es posible requerir una gran variación del rendimiento de los codificadores y descodificadores, dependiendo de los valores adoptados por los elementos de sintaxis en el flujo de bits, tales como el tamaño especificado de las imágenes descodificadas. La norma H.264 reconoce además que, en muchas aplicaciones, no es ni práctico ni económico implementar un descodificador capaz de tratar todos los usos hipotéticos de la sintaxis dentro de un perfil en particular. Por consiguiente, la norma H.264 define un "nivel" como un conjunto especificado de restricciones impuestas a los valores de los elementos de sintaxis en el flujo de bits. Estas restricciones pueden ser simples limitaciones de valores. De forma alternativa, estas restricciones pueden adoptar la forma de restricciones sobre combinaciones aritméticas de valores (por ejemplo, la anchura de la imagen multiplicada por la altura de la imagen multiplicada por el número de imágenes descodificadas por segundo). La norma H.264 establece además que las implementaciones individuales pueden admitir un nivel diferente para cada perfil admitido.

[0069] Un descodificador que se ajusta a un perfil normalmente admite todas las características definidas en el perfil. Por ejemplo, como característica de codificación, la codificación de imágenes B no está admitida en el perfil de línea de base de H.264/AVC, pero está admitida en otros perfiles de H.264/AVC. Un descodificador que se ajusta a un nivel deberá ser capaz de descodificar cualquier flujo de bits que no requiere recursos fuera de las limitaciones definidas en el nivel. Las definiciones de perfiles y niveles pueden ser útiles para la interpretabilidad. Por ejemplo, durante la transmisión de vídeo, se pueden negociar y acordar un par de definiciones de perfil y nivel para una sesión de transmisión completa. Más específicamente, en la norma H.264/AVC, un nivel puede definir, por ejemplo, limitaciones en el número de macrobloques que necesitan ser procesados, el tamaño del búfer de imágenes descodificadas (DPB), el tamaño del búfer de imágenes codificadas (CPB), el intervalo vertical del vector de movimiento, el número máximo de vectores de movimiento por dos MB consecutivos y si un bloque B puede tener particiones de submacrobloque inferiores a 8x8 píxeles. De esta manera, un descodificador puede determinar si el descodificador es capaz de descodificar apropiadamente el flujo de bits.

[0070] Los conjuntos de parámetros contienen en general la información de cabecera de capa-secuencia en los conjuntos de parámetros de secuencia (SPS), y la información de cabecera de capa-imagen que cambia ocasionalmente en conjuntos de parámetros de imagen (PPS). Con los conjuntos de parámetros, esta información que cambia ocasionalmente no necesita repetirse para cada secuencia o imagen, de ahí que pueda mejorarse la eficacia de la codificación. Además, el uso de conjuntos de parámetros puede permitir la transmisión fuera de banda de la información de cabecera, evitando la necesidad de transmisiones redundantes para lograr la resistencia a los errores. En la transmisión fuera de banda, las unidades de NAL de conjuntos de parámetros se transmiten por un canal diferente al de las otras unidades de NAL.

[0071] Los ejemplos de SVC y MVC proporcionan extensiones de conjunto de parámetros de secuencia (SPS) para vídeo ajustable a escala o vídeo multivista. La información de extensión SPS puede estar contenida en unidades NAL sub-SPS, que pueden tener un tipo de unidad NAL diferente a las otras unidades NAL. La información de perfil y nivel para los puntos de funcionamiento admitidos puede incluirse en las unidades NAL, incluida la información de extensión SPS.

[0072] Una presentación de medios puede incluir una descripción de presentación de medios (MPD), que puede contener descripciones de diferentes representaciones alternativas (por ejemplo, servicios de vídeo con diferentes calidades) y la descripción puede incluir, por ejemplo, información de códec, un valor de perfil y un valor de nivel. El dispositivo de destino 40 puede recuperar la MPD de una presentación de medios para determinar cómo acceder a fragmentos de película de diversas presentaciones. Los fragmentos de película pueden estar situados en cuadros de fragmento de película (cuadros moof) de archivos de vídeo.

[0073] Las normas de compresión vídeo, tales como ITU-T H.261, H.262, H.263, MPEG-1, MPEG-2 y H.264/MPEG-4 parte 10, aprovechan la predicción temporal compensada por movimiento para reducir la redundancia temporal. El codificador utiliza una predicción compensada por movimiento de algunas imágenes previamente codificadas (también mencionadas aquí como tramas) para predecir las imágenes codificadas actuales de acuerdo con los vectores de movimiento. Existen tres tipos de imagen principales en la codificación de vídeo típica. Se trata de imágenes intracodificadas ("imágenes-I" o "tramas-I"), imágenes predictivas ("imágenes-P" o "tramas-P") e imágenes predictivas bidireccionales ("imágenes-B" o "tramas-B"). Las imágenes-P utilizan solamente la imagen de referencia antes de la imagen actual en orden temporal. En una imagen B, cada bloque de la imagen B puede predecirse a partir de una o dos imágenes de referencia. Estas imágenes de referencia podrían situarse antes o después de la imagen actual en el orden temporal.

[0074] De acuerdo con la norma de codificación H.264, como ejemplo, las imágenes-B usan dos listas de imágenes de referencia codificadas previamente, la lista 0 y la lista 1. Estas dos listas pueden cada una contener imágenes codificadas pasadas y/o futuras en orden temporal. Los bloques en una imagen-B se pueden predecir de una entre varias maneras: predicción compensada por movimiento a partir de una imagen de referencia de lista 0, predicción compensada por movimiento a partir de una imagen de referencia de lista 1 o predicción compensada por movimiento

a partir de la combinación de imágenes de referencia de ambas listas, lista 0 y lista 1. Para obtener la combinación de las imágenes de referencia, tanto de la lista 0 como de la lista 1, se obtienen dos áreas de referencia compensadas por movimiento de imágenes de referencia de la lista 0 y de la lista 1, respectivamente. Su combinación se usará para predecir el bloque actual.

[0075] La norma ITU-T H.264 admite la intrapredicción en diversos tamaños de bloque, tales como 16 por 16, 8 por 8 o 4 por 4 para componentes de luma, y 8x8 para componentes de croma, así como la interpredicción en diversos tamaños de bloque, tales como 16x16, 16x8, 8x16, 8x8, 8x4, 4x8 y 4x4 para componentes de luma y los tamaños ajustados a escala correspondientes para componentes de croma. En la presente divulgación, "x" y "por" pueden utilizarse indistintamente para hacer referencia a las dimensiones de píxeles del bloque en términos de las dimensiones vertical y horizontal, por ejemplo, 16x16 píxeles o 16 por 16 píxeles. En general, un bloque de 16x16 tendrá 16 píxeles en una dirección vertical ($y = 16$) y 16 píxeles en una dirección horizontal ($x = 16$). Asimismo, un bloque de tamaño $N \times N$ 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.

[0076] Los tamaños de bloque que son inferiores a 16 por 16 pueden denominarse particiones de un macrobloque de 16 por 16. Los bloques de vídeo pueden comprender bloques de datos de píxeles en el dominio de píxel, o bloques de coeficientes de transformada en el dominio de transformada, por ejemplo, tras la aplicación de una transformada, tal como una transformada discreta del coseno (DCT), una transformada de números enteros, una transformada de ondículas o una transformada conceptualmente similar a los datos de bloques de vídeo residuales que representan diferencias de píxeles entre los bloques de vídeo codificados y los bloques de vídeo predictivos. En algunos casos, un bloque de vídeo puede comprender bloques de coeficientes de transformada cuantificados en el dominio de transformada.

[0077] Los bloques de vídeo más pequeños pueden proporcionar una mejor resolución y se pueden usar para localizaciones de una trama de vídeo que incluyen altos niveles de detalle. En general, los macrobloques y las diversas particiones, denominadas en ocasiones subbloques, pueden considerarse bloques de vídeo. Así mismo, un fragmento puede considerarse una pluralidad de bloques de vídeo, tales como macrobloques y/o subbloques. Cada fragmento puede ser una unidad independientemente descodificable de una trama de vídeo. De forma alternativa, las propias tramas pueden ser unidades descodificables, o pueden definirse otras porciones de una trama como unidades descodificables. El término "unidad codificada" o "unidad de codificación" puede referirse a cualquier unidad independientemente descodificable de una trama de vídeo, tal como una trama completa, un fragmento de una trama, un grupo de imágenes (GOP), denominado también secuencia u otra unidad independientemente descodificable definida de acuerdo con técnicas de codificación aplicables.

[0078] El término macrobloque se refiere a una estructura de datos para la codificación de datos de imagen y/o vídeo de acuerdo con una matriz bidimensional de píxeles que comprende 16x16 píxeles. Cada píxel comprende una componente de crominancia y una componente de luminancia. Por consiguiente, el macrobloque puede definir cuatro bloques de luminancia, cada uno de los cuales comprende una formación bidimensional de 8x8 píxeles, dos bloques de crominancia, cada uno de los cuales comprende una formación bidimensional de 16x16 píxeles, y una cabecera que comprende información sintáctica, tal como un patrón de bloque codificado (CBP), un modo de codificación (por ejemplo, modos de intracodificación (I) o de intercodificación (P o B)), un tamaño de partición para particiones de un bloque intracodificado (por ejemplo, 16x16, 16x8, 8x16, 8x8, 8x4, 4x8 o 4x4), o uno o más vectores de movimiento para un macrobloque intercodificado.

[0079] El codificador de vídeo 28, el descodificador de vídeo 48, el codificador de audio 26, el descodificador de audio 46, la unidad de encapsulación 30 y la unidad de desencapsulación 38 pueden, cada uno, implementarse como cualquiera de una diversidad de circuitos de procesamiento adecuados, según corresponda, 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), circuitos lógicos discretos, software, hardware, firmware o cualquier combinación de los mismos. Tanto el codificador de vídeo 28 como el descodificador de vídeo 48 pueden estar incluidos en uno o más codificadores o descodificadores, ambos de los cuales pueden estar integrados como parte de un codificador/descodificador (CÓDEC) de vídeo combinado. Asimismo, tanto el codificador de audio 26 como el descodificador de audio 46 pueden estar incluidos en uno o más codificadores o descodificadores, ambos de los cuales puede estar integrado como parte de un CÓDEC combinado. Un aparato que incluye un codificador de vídeo 28, un descodificador de vídeo 48, un codificador de audio 26, un descodificador de audio 46, una unidad de encapsulación 30 y/o una unidad de desencapsulación 38 puede comprender un circuito integrado, un microprocesador y/o un dispositivo de comunicación inalámbrica, tal como un teléfono celular.

[0080] Después de que la unidad de encapsulación 30 haya ensamblado las unidades de NAL y/o las unidades de acceso en un archivo de vídeo, en base a los datos recibidos, la unidad de encapsulación 30 pasa el archivo de vídeo a la interfaz de salida 32 para su salida. En algunos ejemplos, la unidad de encapsulación 30 puede almacenar el archivo de vídeo localmente o enviar el archivo de vídeo a un servidor remoto por medio de la interfaz de salida 32, en lugar de enviar el archivo de vídeo directamente al dispositivo de destino 40. La interfaz de salida 32 puede comprender, por ejemplo, un transmisor, un transceptor, un dispositivo para escribir datos en un medio legible por ordenador tal como, por ejemplo, una unidad óptica, una unidad de medios magnéticos (por ejemplo, una unidad de

disquetes), un puerto de bus serie universal (USB), una interfaz de red u otra interfaz de salida. La interfaz de salida 32 envía el archivo de vídeo a un medio legible por ordenador 34, tal como, por ejemplo, una señal de transmisión, un medio magnético, un medio óptico, una memoria, una unidad de memoria flash u otro medio legible por ordenador.

5 **[0081]** En última instancia, la interfaz de entrada 36 recupera los datos del medio legible por ordenador 34. La interfaz de entrada 36 puede comprender, por ejemplo, una unidad óptica, una unidad de medios magnéticos, un puerto de USB, un receptor, un transceptor u otra interfaz de medio legible por ordenador. La interfaz de entrada 36 puede proporcionar la unidad de NAL o la unidad de acceso a la unidad de desencapsulación 38. La unidad de desencapsulación 38 puede desencapsular un elemento de un archivo de vídeo en flujos PES constituyentes, desempaquetar los flujos PES para recuperar los datos codificados y enviar los datos codificados al descodificador de audio 46 o al descodificador de vídeo 48, dependiendo de si los datos codificados forman parte de un flujo de audio o vídeo, por ejemplo, como se indica en las cabeceras de paquetes PES del flujo. El descodificador de audio 46 descodifica datos de audio codificados y envía los datos de audio descodificados a la salida de audio 42, mientras que el descodificador de vídeo 48 descodifica datos de vídeo codificados y envía los datos de vídeo descodificados, que pueden incluir una pluralidad de vistas de un flujo, a la salida de vídeo 44.

10 **[0082]** Así mismo, la unidad de desencapsulación 38 puede enviar un conjunto de datos de secuencia descodificados, que puede incluir datos PPS, datos SPS y/o mensajes de SEI a nivel de secuencia, a la salida de vídeo 44. La salida de vídeo 44 puede comprender una pantalla que utiliza una única vista de una escena o una pluralidad de vistas de una escena, por ejemplo, una visualización estereoscópica o autoestereoscópica que presenta cada vista de una escena simultáneamente o en una sucesión rápida alternante. La salida de vídeo 44 puede usar el conjunto de datos de secuencia para visualizar las imágenes de vídeo descodificadas. Los mensajes de SEI a nivel de secuencia también se pueden usar para ayudar en el proceso de descodificación, aunque los mensajes de SEI no son necesarios para descodificar correctamente las muestras de una secuencia de vídeo. Por consiguiente, el descodificador de vídeo 48 también puede usar los mensajes de SEI a nivel de secuencia al descodificar muestras de vídeo del archivo de vídeo.

15 **[0083]** Como se ha indicado anteriormente, los mensajes de SEI pueden usarse para una variedad de propósitos. Por ejemplo, los mensajes de SEI a nivel de secuencia pueden describir información de panoramización y exploración para ajustar los datos de vídeo y satisfacer diversos tamaños de pantalla cuando se visualizan. Por consiguiente, la salida de vídeo 44 (o una unidad de control de la misma) puede usar los mensajes de SEI a nivel de secuencia para ajustar los datos de vídeo que se visualizarán por la salida de vídeo 44.

20 **[0084]** Para mencionar otro ejemplo, los mensajes de SEI a nivel de secuencia pueden describir información de vídeo estéreo. Es decir, los mensajes de SEI a nivel de secuencia pueden describir qué imágenes son imágenes "izquierda" o "derecha" para aplicaciones de vídeo tridimensional. Por consiguiente, la salida de vídeo 44 puede usar los mensajes de SEI a nivel de secuencia para determinar cómo presentar datos de vídeo tridimensionales. Por ejemplo, la salida de vídeo 44 puede visualizar imágenes izquierda y derecha usando diferentes polarizaciones de luz o con diferentes filtros de color (por ejemplo, rojo y verde), o puede alternar rápidamente entre imágenes izquierda y derecha, de manera que las gafas activas asociadas con la salida de vídeo 44 pueden alternar entre qué lente deja pasar la luz.

25 **[0085]** Para mencionar aún otro ejemplo, los mensajes de SEI a nivel de secuencia pueden proporcionar información de ajustabilidad a escala, por ejemplo, para ejemplos en los que el archivo de vídeo se construye de acuerdo con SVC o MVC. Por consiguiente, el descodificador de vídeo 48 puede usar los mensajes de SEI a nivel de secuencia para determinar si se ha recibido información de la capa de mejora, y por consiguiente si va a usar los datos de la capa de mejora recibidos o va a interpolar valores para sobremuestrear datos de vídeo de la capa base.

30 **[0086]** Para mencionar otro ejemplo, los mensajes de SEI a nivel de secuencia pueden proporcionar información de adquisición multivista o información de escena multivista, por ejemplo, para ejemplos en los que el archivo de vídeo se construye de acuerdo con MVC. En general, los mensajes de SEI de información de adquisición multivista especifican parámetros de un entorno de adquisición, como, por ejemplo, parámetros de cámara intrínsecos y extrínsecos, un número de vistas, una distancia focal, un punto principal, un factor de inclinación, parámetros de rotación y/o parámetros de traslación. Los mensajes de SEI de información de escena multivista pueden especificar la disparidad máxima entre múltiples vistas, cuando se captan las vistas. Esta información puede ser utilizada por el descodificador de vídeo 48 y/o la salida de vídeo 44 al descodificar y/o representar los datos de vídeo tridimensionales, por ejemplo, al construir una imagen adicional a partir de una imagen existente. La información de escena multivista también puede describir características de una o más cámaras o condiciones durante la adquisición de imágenes, como, por ejemplo, cuál es la disparidad máxima causada por los ajustes de la cámara de, por ejemplo, hasta 1024 vistas. La información de escena multivista también puede proporcionar detalles pertinentes para la calibración y rectificación de la cámara.

35 **[0087]** La FIG. 2 es un diagrama de bloques que ilustra componentes de una unidad de encapsulación de ejemplo 30. En el ejemplo de la FIG. 2, la unidad de encapsulación 30 incluye una interfaz de entrada de vídeo 80, una interfaz de entrada de audio 82, una unidad de creación de archivos de vídeo 60 y una interfaz de salida de archivos de vídeo 84. La unidad de creación de archivos de vídeo 60, en este ejemplo, incluye el constructor de unidad de capa de

abstracción de red (NAL) 62, la unidad de extracción de mensajes de SEI 64 y la unidad de creación del conjunto de datos de secuencia 66.

[0088] La interfaz de entrada de vídeo 80 y la interfaz de entrada de audio 82 reciben datos de vídeo y audio codificados, respectivamente. La interfaz de entrada de vídeo 80 y la interfaz de entrada de audio 82 pueden recibir datos de vídeo y audio codificados a medida que se codifican los datos, o puede recuperar datos de vídeo y audio codificados a partir de un medio legible por ordenador. Tras recibir datos de vídeo y audio codificados, la interfaz de entrada de vídeo 80 y la interfaz de entrada de audio 82 pasan los datos de vídeo y audio codificados a la unidad de creación de archivos de vídeo 60 para ensamblarlos en un archivo de vídeo.

[0089] La unidad de creación de archivos de vídeo 60 puede corresponder a una unidad de control que incluya hardware, software y/o firmware configurado para realizar las funciones y procedimientos atribuidos a los mismos. La unidad de control puede llevar a cabo además las funciones atribuidas a la unidad de encapsulación 30 en general. Para ejemplos en los que la unidad de creación de archivos de vídeo 60 está incorporada en software y/o firmware, la unidad de encapsulación 30 puede incluir un medio legible por ordenador que comprende instrucciones para la unidad de creación de archivos de vídeo 60 y una unidad de procesamiento para ejecutar las instrucciones. Cada una de las subunidades de la unidad de creación de archivos de vídeo 60 (constructor de unidad de NAL 62, unidad de extracción de mensajes de SEI 64 y unidad de creación del conjunto de datos de secuencia 66, en este ejemplo) se puede implementar como unidades de hardware y/o módulos de software individuales, y se puede integrar funcionalmente o separarse además en subunidades adicionales. La unidad de creación de archivos de vídeo 60 puede corresponder a cualquier unidad de procesamiento o circuitos de procesamiento adecuados, tales como, por ejemplo, uno o más microprocesadores, circuitos integrados específicos de la aplicación (ASIC), matrices de puertas programables *in situ* (FPGA), procesadores de señales digitales (DSP) o cualquier combinación de los mismos. La unidad de creación de archivos de vídeo 60 puede incluir además un medio no transitorio y legible por ordenador que almacena instrucciones para cualquiera o la totalidad del constructor de unidad de NAL 62, la unidad de extracción de mensajes de SEI 64 y la unidad de creación del conjunto de datos de secuencia 66, así como un procesador para ejecutar las instrucciones.

[0090] En general, la unidad de creación de archivos de vídeo 60 puede crear un archivo de vídeo incluyendo los datos de audio y vídeo recibidos. El constructor de unidad de NAL 62 puede formar unidades de NAL que incluyen muestras codificadas de vídeo y audio. La unidad de creación de archivos de vídeo 60 puede configurarse además para ensamblar unidades de acceso que incluyan todas las unidades de NAL para una instancia de tiempo en particular. Además, la unidad de creación de archivos de vídeo 60 puede configurarse para desacoplar mensajes de SEI a nivel de secuencia a partir de imágenes de vídeo codificadas descritas por los mensajes de SEI a nivel de secuencia, y almacenar los mensajes de SEI a nivel de secuencia en el archivo de vídeo por separado a partir de las imágenes de vídeo codificadas descritas por los mensajes de SEI a nivel de secuencia.

[0091] El codificador de vídeo 28 (FIG. 1) puede incluir mensajes de SEI con muestras de datos de vídeo. Los mensajes de SEI recibidos desde el codificador de vídeo 28 para un número de muestras individuales pueden ser los mismos para cada una de las muestras, en algunos casos. Por consiguiente, en lugar de transmitir los mismos mensajes de SEI para cada una de las muestras, tras determinar que los mensajes de SEI para las muestras individuales son iguales (o muy similares) para cada una de las muestras, la unidad de extracción de mensajes de SEI 64 puede separar los mensajes de SEI de las muestras descritas por los mensajes de SEI. Dichos mensajes de SEI que pueden estar relacionados con más de una imagen de vídeo codificada pueden denominarse mensajes de SEI a nivel de secuencia.

[0092] De forma adicional o alternativa, la unidad de creación de archivos de vídeo 60 puede configurarse para generar mensajes de SEI a nivel de secuencia para una secuencia de imágenes de vídeo codificadas. De esta manera, la unidad de creación de archivos de vídeo 60 puede generar mensajes de SEI a nivel de secuencia o separar mensajes de SEI existentes de las muestras de vídeo para formar mensajes de SEI a nivel de secuencia que describen una o más de las muestras de vídeo.

[0093] La unidad de creación del conjunto de datos de secuencia 66 puede crear un conjunto de datos de secuencia para un archivo de vídeo producido por la unidad de creación de archivos de vídeo 60. La unidad de creación del conjunto de datos de secuencia 66 puede agrupar mensajes de SEI a nivel de secuencia con el conjunto de parámetros de secuencia y/o los datos de conjunto de parámetros de imagen. Por ejemplo, la unidad de creación del conjunto de datos de secuencia 66 puede incluir los mensajes de SEI a nivel de secuencia en un flujo elemental de conjunto de parámetros que incluye los SPS y los PPS, y en algunos ejemplos (por ejemplo, con respecto a la SVC y MVC), los SPS del subconjunto. Para mencionar otro ejemplo, la unidad de creación del conjunto de datos de secuencia 66 puede incluir los mensajes de SEI a nivel de secuencia para una secuencia de imágenes de vídeo codificadas en una sección continua del archivo de vídeo creado. Es decir, la unidad de creación del conjunto de datos de secuencia 66 puede asegurar que los datos para mensajes de SEI a nivel de secuencia para una secuencia particular de imágenes de vídeo codificadas se agrupen en una serie continua de uno o más bytes. Por ejemplo, la unidad de creación del conjunto de datos de secuencia 66 puede ensamblar los mensajes de SEI a nivel de secuencia y otros datos para el conjunto de datos de secuencia en una muestra de vídeo particular que no contiene ninguna parte de una imagen de vídeo codificada.

[0094] Para mencionar otro ejemplo, por ejemplo, con respecto a 3GPP, la unidad de creación del conjunto de datos de secuencia 66 puede ensamblar el conjunto de datos de secuencia para una secuencia de imágenes de vídeo codificadas en un archivo 3GPP relativamente pequeño que precede a uno o más archivos 3GPP incluyendo la secuencia de imágenes de vídeo codificadas descrita por el conjunto de datos de secuencia. De forma alternativa, la unidad de creación del conjunto de datos de secuencia 66 puede incluir el conjunto de datos de secuencia en un archivo de descripción de presentación de medios (MPD) para una o más presentaciones relacionadas que incluyen uno o más archivos 3GPP.

[0095] Para mencionar aún otro ejemplo, la unidad de creación del conjunto de datos de secuencia 66 puede incluir el conjunto de datos de secuencia dentro de un cuadro de película (cuadro de moov) de un archivo de vídeo cuando el cuadro de película no incluye imágenes de vídeo codificadas. Para mencionar otro ejemplo, la unidad de creación del conjunto de datos de secuencia 66 puede incluir el conjunto de datos de secuencia dentro de un cuadro de fragmento de película (cuadro de moof) que se produce después del cuadro de película (cuadro de moov) de un archivo de vídeo, pero antes de cualquier cuadro de fragmento de película que incluya imágenes de vídeo codificadas.

[0096] Al agrupar SPS, PPS, mensajes de SEI a nivel de secuencia y/o subconjunto de SPS en un conjunto de datos de secuencia, la unidad de creación del conjunto de datos de secuencia 66 puede simplificar la tarea de proporcionar datos de secuencia a un dispositivo cliente, como el dispositivo de destino 40. De esta manera, los mensajes de SEI a nivel de secuencia pueden transmitirse al dispositivo de destino 40 antes de que las imágenes de vídeo codificadas descritas por los mensajes de SEI a nivel de secuencia se transmitan al dispositivo de destino 40.

[0097] En general, se espera que los mensajes SPS, PPS y SEI a nivel de secuencia de un conjunto de datos de secuencia permanezcan constantes para una secuencia de muestras de vídeo descritas por el conjunto de datos de secuencia. Sin embargo, en algunos casos, ciertas porciones de los SPS, PPS y/o los mensajes de SEI a nivel de secuencia pueden variar ligeramente para imágenes de vídeo codificadas individuales o para conjuntos o subsecuencias de imágenes de vídeo codificadas. Por ejemplo, en un archivo de vídeo que encapsula un número de imágenes de vídeo codificadas, puede haber una o más secuencias que no necesariamente comparten la misma información de un SPS, un PPS o mensajes de SEI a nivel de secuencia.

[0098] Cuando esto se produce, la unidad de creación del conjunto de datos de secuencia 66 puede crear múltiples pistas de conjuntos de parámetros, donde cada pista puede incluir un conjunto de mensajes SPS, PPS y/o de SEI, así como indicaciones de la información de temporización que indica cuándo se aplica el conjunto de datos de secuencia de la pista. Por ejemplo, la información de temporización puede indicar un tiempo de inicio para la secuencia de vídeo que indica un tiempo en el que la pista comienza a aplicarse a la secuencia de vídeo y/o un tiempo de finalización para la secuencia de vídeo en el que ya no se aplica la pista que incluye la información de temporización. De forma alternativa, la información de temporización para cada pista puede indicar solo un tiempo de finalización, de manera que al alcanzar el tiempo de finalización para una pista de conjunto de parámetros particular, un dispositivo cliente puede saltar a la siguiente pista de conjunto de parámetros. Dicha información de temporización puede corresponder a un tiempo universal coordinado (UTC) o una duración de una última muestra de vídeo (para el tiempo final) en el que se aplica la pista de conjunto de parámetros en una escala de tiempo que se utiliza para todo el vídeo en el formato de encapsulación de archivos.

[0099] Una de las múltiples pistas de conjuntos de parámetros puede aplicarse a todas las muestras del archivo de vídeo. La unidad de creación del conjunto de datos de secuencia 66 puede indicar una pista de conjunto de parámetros que se aplica a todas las muestras de un archivo de vídeo estableciendo la información de temporización para la pista de conjunto de parámetros a un tiempo UTC posterior a cualquier muestra de película o fragmento de película. De forma alternativa, la unidad de creación del conjunto de datos de secuencia 66 puede indicar una duración de cero para la información de temporización para una pista de conjunto de parámetros que se aplica a todas las muestras de un archivo de vídeo.

[0100] Para los ejemplos en los que la unidad de creación del conjunto de datos de secuencia 66 crea múltiples pistas de conjuntos de parámetros, la unidad de creación del conjunto de datos de secuencia 66 puede asignar identificadores únicos respectivos a cada una de las pistas de conjuntos de parámetros. Cuando un dispositivo cliente, como el dispositivo de destino 40, solicita buscar una ubicación temporal particular, por ejemplo, mediante la reproducción local o durante la transmisión continua del HTTP, el dispositivo de destino 40 puede determinar cuál de un número de pistas de conjuntos de parámetros se aplica a la búsqueda de ubicación temporal. El dispositivo cliente puede solicitar identificadores para las pistas de conjuntos de parámetros que tienen información de temporización que indica que las pistas respectivas de conjuntos de parámetros se aplican a muestras de vídeo más tarde que la búsqueda de ubicación temporal. Un servidor (por ejemplo, el dispositivo de origen 20) puede mantener una estructura de datos para marcar dichas pistas pertinentes de conjuntos de parámetros para asegurar que las pistas pertinentes de conjuntos de parámetros se suministren al cliente, por ejemplo, en el contexto de la transmisión continua del HTTP. Durante la reproducción local, el cliente puede mantener una estructura de datos similar para asegurar que las pistas pertinentes de conjuntos de parámetros sean analizadas por el dispositivo cliente.

[0101] Además, en lugar de incluir un número de pistas de conjuntos de parámetros, la unidad de creación del conjunto de datos de secuencia 66 puede incluir un número de conjuntos de datos de secuencia en un archivo de

vídeo, cada uno de los cuales puede almacenarse de forma continua dentro del archivo. Es decir, cada uno de los conjuntos de datos de secuencia puede almacenarse de forma continua de manera que todos los datos para un conjunto particular de datos de secuencia sean continuos, pero los conjuntos de datos de secuencia no necesitan almacenarse de forma continua entre sí. Por ejemplo, la unidad de creación del conjunto de datos de secuencia 66 puede almacenar un conjunto de datos de secuencia en un fragmento de vídeo que precede a una serie de fragmentos para los cuales el conjunto de datos de secuencia es pertinente, y luego almacenar un nuevo conjunto de datos de secuencia después de la serie de fragmentos. El nuevo conjunto de datos de secuencia puede ser pertinente para otros fragmentos de vídeo, que van detrás del fragmento que almacena el nuevo conjunto de datos de secuencia. Al almacenar múltiples conjuntos de datos de secuencia en un archivo, la unidad de creación del conjunto de datos de secuencia 66 puede proporcionar información de temporización para cada uno de los conjuntos de datos de secuencia, así como un identificador único para cada uno de los conjuntos de datos de secuencia.

[0102] Después de que la unidad de creación de archivos de vídeo 60 produce un archivo de vídeo que incluye mensajes de SEI a nivel de secuencia que están separados de las imágenes de vídeo codificadas descritas por los mensajes de SEI a nivel de secuencia, la unidad de creación de archivos de vídeo 60 puede pasar el archivo de vídeo a la interfaz de salida de archivos de vídeo 84. La interfaz de salida de archivos de vídeo 84 puede enviar el archivo de vídeo, por ejemplo, a la interfaz de salida 32 del dispositivo de origen 20. En algunos ejemplos, la interfaz de salida de archivos de vídeo 84 puede enviar el archivo de vídeo a un medio de almacenamiento del dispositivo de origen 20 (no se muestra). El archivo de vídeo puede almacenarse localmente dentro del dispositivo de origen 20, almacenarse en un medio de almacenamiento portátil como un DVD, disco Blu-ray, unidad de memoria flash, disco flexible u otro medio de almacenamiento portátil, salida a través de una red, por ejemplo, de acuerdo con un protocolo de transmisión continua, como la transmisión continua del HTTP, o de otro modo emitido de tal manera que el archivo de vídeo pueda ser recibido por un dispositivo cliente, como el dispositivo de destino 40. De esta manera, la emisión de datos de vídeo, por ejemplo, mensajes de SEI a nivel de secuencia e imágenes de vídeo codificadas descritas por los mensajes de SEI a nivel de secuencia, puede incluir almacenar, transferir, transmitir y radiodifundir los datos de vídeo.

[0103] La FIG. 3 es un diagrama de bloques que ilustra elementos de un archivo de vídeo de ejemplo 100 que puede incluir mensajes de SEI a nivel de secuencia en diversas ubicaciones. Como se ha descrito anteriormente, los archivos de vídeo, de acuerdo con el formato de archivo de medios de base de la ISO, y las ampliaciones del mismo, almacenan los datos en una serie de objetos, denominados "cuadros". En el ejemplo de la FIG. 3, el archivo de vídeo 100 incluye el cuadro del tipo de archivo (FTYP) 102, el cuadro de película (MOOV) 104, los cuadros de fragmentos de película (MOOF) 112 y el cuadro de acceso aleatorio de fragmento de película (MFRA) 114.

[0104] El cuadro de tipo de archivo (FTYP) 102 describe en general un tipo de archivo para el archivo de vídeo 100. El cuadro de tipo de archivo 102 puede incluir datos que identifican una especificación que describe un mejor uso para el archivo de vídeo 100. El cuadro de tipo de archivo 102 se puede colocar antes del cuadro de MOOV 104, de los cuadros de fragmentos de película 112 y del cuadro de MFRA 114.

[0105] El cuadro de MOOV 104, en el ejemplo de la FIG. 3, incluye el cuadro de cabecera de película (MVHD) 106, el cuadro de pista (TRAK) 108 y uno o más cuadros de ampliación de película (MVEX) 110. En general, el cuadro de MVHD 106 puede describir características generales del archivo de vídeo 100. Por ejemplo, el cuadro de MVHD 106 puede incluir datos que describen cuándo se creó inicialmente el archivo de vídeo 100, cuándo se modificó por última vez el archivo de vídeo 100, una escala de tiempo para el archivo de vídeo 100, una duración de reproducción para el archivo de vídeo 100 u otros datos que describen en general el archivo de vídeo 100.

[0106] El cuadro de TRAK 108 puede incluir datos para una pista del archivo de vídeo 100. El cuadro de TRAK 108 puede incluir un cuadro de cabecera de pista (TKHD) que describe las características de la pista correspondiente al cuadro de TRAK 108. En algunos ejemplos, el cuadro de TRAK 108 puede incluir imágenes de vídeo codificadas, mientras que, en otros ejemplos, los cuadros codificados de vídeo de la pista pueden incluirse en fragmentos de película 112, a los que se puede hacer referencia mediante los datos del cuadro de TRAK 108.

[0107] En algunos ejemplos, el archivo de vídeo 100 puede incluir más de una pista. Por consiguiente, el cuadro de MOOV 104 puede incluir un número de cuadros de TRAK igual al número de pistas del archivo de vídeo 100. El cuadro de TRAK 108 puede describir las características de una pista correspondiente del archivo de vídeo 100. Por ejemplo, el cuadro de TRAK 108 puede describir información temporal y/o espacial para la pista correspondiente. Un cuadro de TRAK similar al cuadro de TRAK 108 del cuadro de MOOV 104 puede describir características de una pista de conjunto de parámetros, cuando la unidad de encapsulación 30 (FIG. 1) incluye una pista de conjunto de parámetros en un archivo de vídeo, tal como el archivo de vídeo 100. La unidad de encapsulación 30 puede señalar la presencia de mensajes de SEI a nivel de secuencia en la pista de conjunto de parámetros dentro del cuadro de TRAK que describe la pista de conjunto de parámetros.

[0108] Los cuadros de MVEX 110 pueden describir características de unos correspondientes fragmentos de película 112, por ejemplo, para señalar que el archivo de vídeo 100 incluye fragmentos de película 112, además de los datos de vídeo incluidos dentro del cuadro de MOOV 104, si los hubiera. En el contexto de la transmisión continua de datos de vídeo, las imágenes de vídeo codificadas pueden estar incluidas en los fragmentos de película 112 en lugar de en

el cuadro de MOOV 104. Por consiguiente, todas las muestras de vídeo codificadas pueden estar incluidas en fragmentos de película 112, en lugar de en el cuadro de MOOV 104.

5 **[0109]** El cuadro de MOOV 104 puede incluir un número de cuadros de MVEX 110 iguales al número de fragmentos de película 112 del archivo de vídeo 100. Cada uno de los cuadros de MVEX 110 puede describir características de uno correspondiente de los fragmentos de película 112. Por ejemplo, cada cuadro de MVEX puede incluir un cuadro de cabecera de ampliación de película (MEHD) que describe una duración temporal para el cuadro correspondiente de los fragmentos de película 112.

10 **[0110]** Como se indica anteriormente, la unidad de encapsulación 30 puede almacenar un conjunto de datos de secuencia en una muestra de vídeo que no incluye datos de vídeo codificados reales. Una muestra de vídeo puede corresponder en general a una unidad de acceso, que es una representación de una imagen codificada en una instancia de tiempo específica. En el contexto de la AVC, la imagen codificada incluye una o más unidades NAL VCL que contienen la información para construir todos los píxeles de la unidad de acceso y otras unidades NAL no VCL asociadas, tales como mensajes de SEI. Por consiguiente, la unidad de encapsulación 30 puede incluir un conjunto de datos de secuencia, que puede incluir mensajes de SEI a nivel de secuencia, en uno de los fragmentos de película 112. La unidad de encapsulación 30 puede señalar además la presencia de un conjunto de datos de secuencia y/o de mensajes de SEI a nivel de secuencia con la presencia de estos en uno de los fragmentos de película 112 dentro de uno de los cuadros de MVEX 110 correspondiente al uno de los fragmentos de película 112.

20 **[0111]** Los fragmentos de película 112 pueden incluir una o más imágenes de vídeo codificadas. En algunos ejemplos, los fragmentos de película 112 pueden incluir uno o más grupos de imágenes (GOP), cada uno de los cuales puede incluir un número de imágenes de vídeo codificadas, por ejemplo, tramas o imágenes. Así mismo, como se describe anteriormente, los fragmentos de película 112 pueden incluir conjuntos de datos de secuencia en algunos ejemplos. Cada uno de los fragmentos de película 112 puede incluir un cuadro de cabecera de fragmento de película (MFHD). El cuadro de MFHD puede describir características del fragmento de película correspondiente, tales como un número de secuencia para el fragmento de película. Los fragmentos de película 112 pueden estar incluidos por orden de número de secuencia en el archivo de vídeo 100.

30 **[0112]** El cuadro de MFRA 114 puede describir puntos de acceso aleatorio dentro de fragmentos de película 112 del archivo de vídeo 100. Esto puede ayudar a realizar búsquedas de ubicaciones temporales particulares dentro del archivo de vídeo 100. El cuadro de MFRA 114 en general es opcional y no necesita estar incluido en los archivos de vídeo de algunos ejemplos. Del mismo modo, un dispositivo cliente, tal como el dispositivo de destino 40, no tiene necesariamente que referenciar el cuadro de MFRA 114 para descodificar correctamente y visualizar los datos de vídeo del archivo de vídeo 100. El cuadro de MFRA 114 puede incluir un número de cuadros de acceso aleatorio de fragmento de pista (TFRA) igual al número de pistas del archivo de vídeo 100 o, en algunos ejemplos, igual al número de pistas de medios (por ejemplo, pistas que no son de indicación) del archivo de vídeo 100.

40 **[0113]** La FIG. 4 es un diagrama de bloques que ilustra elementos de otro archivo de vídeo de ejemplo 140, que incluye una pista de conjunto de parámetros 146 que contiene mensajes de SEI a nivel de secuencia 156. En particular, el archivo de vídeo 140 incluye el cuadro de MOOV 142 y el cuadro de MDAT 150. El archivo de vídeo 140 puede incluir además otros cuadros, tales como un cuadro de tipo de archivo y/o un cuadro de MFRA, entre otros.

45 **[0114]** En el ejemplo de la FIG. 4, el cuadro de MOOV 142 incluye la pista de datos de vídeo 144 y la pista de conjunto de parámetros 146. La pista de conjunto de parámetros 146 corresponde al conjunto de parámetros de imagen (PPS) 152, el conjunto de parámetros de secuencia (SPS) 154 y los mensajes de información de mejora complementaria a nivel de secuencia (SEI SL) 156. La pista de datos de vídeo 144 corresponde a los fragmentos de vídeo 160A-160N (fragmentos de vídeo 160). En general, los fragmentos de vídeo 160 pueden incluir una o más imágenes de vídeo codificadas.

50 **[0115]** En el ejemplo de la FIG. 4, los datos para PPS 152, SPS 154, mensajes de SEI a nivel de secuencia 156 y fragmentos de vídeo 160 se almacenan en el cuadro de MDAT 150. El PPS 152, el SPS 154, y los mensajes de SEI a nivel de secuencia 156 pueden denominarse colectivamente como un conjunto de parámetros de secuencia. En algunos ejemplos, la pista de conjunto de parámetros 146 puede incluir más de un conjunto de parámetros de imagen. El PPS 152, el SPS 154 y mensajes de SEI a nivel de secuencia 156 en general describen características de fragmentos de vídeo 160. Estas características pueden ayudar a un dispositivo cliente, como el dispositivo de destino 40, a seleccionar, descodificar y/o representar fragmentos de vídeo 160.

60 **[0116]** Como ejemplo, cuando el archivo de vídeo 140 corresponde al formato de archivo AVC, los mensajes de SEI a nivel de secuencia 156 pueden incluir uno o más de los mensajes de SEI de rectángulo de barrido panorámico, mensajes de SEI de información de vídeo estéreo y/o mensajes de SEI de la disposición de empaquetado de tramas. Para mencionar otro ejemplo, cuando el archivo de vídeo 140 corresponde a un formato de archivo SVC, los mensajes de SEI a nivel de secuencia 156 pueden incluir mensajes de SEI de información de ajustabilidad a escala. Para mencionar otro ejemplo, cuando el archivo de vídeo 140 corresponde a un formato de archivo MVC, los mensajes de SEI a nivel de secuencia 156 pueden incluir uno o más de los mensajes de SEI de información de ajustabilidad a

escala de la vista, mensajes de SEI de información de adquisición multivista y/o mensajes de SEI de información de escena multivista.

5 **[0117]** Un archivo de vídeo puede incluir un número de pistas de medios similares a la pista de datos de vídeo 144. Cada una de las pistas de datos de vídeo puede tener características diferentes, que pueden corresponder a diversos requisitos de descodificación y/o representación. El dispositivo de destino 40 puede seleccionar una de las pistas de medios basándose, entre otras cosas, en las características de los fragmentos de vídeo para la pista de medios señalizada por el conjunto de datos de secuencia. Por ejemplo, el dispositivo de destino 40 puede seleccionar una de las pistas basándose en las capacidades de descodificación y/o representación del descodificador de vídeo 48 y la salida de vídeo 44, respectivamente. El dispositivo de destino 40 puede determinar las características de los fragmentos de vídeo 160 basándose en la pista de conjunto de parámetros 146, que incluye el PPS 152, el SPS 154 y los mensajes de SEI a nivel de secuencia 156.

15 **[0118]** El dispositivo de destino 40 puede configurarse para recuperar el conjunto de parámetros de secuencia, por ejemplo, el cuadro de MOOV 104, el de los fragmentos de película 112 que incluye el conjunto de parámetros de secuencia, la muestra de vídeo que incluye el conjunto de parámetros de secuencia pero ningún dato de vídeo codificado u otro elemento del archivo de vídeo 100 antes de cualquier muestra codificada descrita por el conjunto de parámetros de secuencia. De forma adicional o alternativa, el dispositivo de origen 20 puede configurarse para proporcionar el conjunto de parámetros de secuencia al dispositivo de destino 40 antes de proporcionar cualquier imagen de vídeo codificada descrita por el conjunto de parámetros de secuencia al dispositivo de destino 40. Además, el dispositivo de origen 20 puede configurarse para rechazar una petición para uno de los fragmentos de película 112 del dispositivo de destino 40 cuando el dispositivo de destino 40 no ha solicitado primero datos del conjunto de datos de secuencia para el fragmento de película 112 solicitado.

25 **[0119]** La FIG. 5 es un diagrama de bloques que ilustra un conjunto de ejemplos 198 de presentaciones 182, 190 que incluyen archivos 3GPP 186, 194 que incluyen conjuntos de datos de secuencia respectivos para las presentaciones respectivas. Es decir, el conjunto 198 puede incluir presentaciones adicionales, pero la FIG. 5 representa dos presentaciones 182, 190 con fines de ejemplo.

30 **[0120]** La descripción de presentación de medios (MPD) 180 incluye datos que describen presentaciones del conjunto 198, por ejemplo, presentación 182, 190. La MPD 180 puede describir, por ejemplo, un localizador uniforme de recursos (URL) y/o un nombre de recurso uniforme (URN) de una presentación. La MPD 180 puede describir aún más las características de cada una de las presentaciones en el conjunto 198. Por ejemplo, la MPD 180 puede incluir conjuntos de datos de secuencia para cada una de las presentaciones. La MPD 180 también puede describir un número de archivos de vídeo 3GPP almacenados en cada una de las presentaciones.

35 **[0121]** En el ejemplo de la FIG. 5, la presentación 182 incluye datos de cabecera 184, archivo 3GPP con el conjunto de datos de secuencia 186 y archivos 3GPP 188A-188N (archivos 3GPP 188). De manera similar, en este ejemplo, la presentación 190 incluye datos de cabecera 192, archivo 3GPP con el conjunto de datos de secuencia 194 y archivos 3GPP 196A-196N (archivos 3GPP 196). Aunque la letra "N" se usa para designar cardinalidades tanto para los archivos 3GPP 188 como para los archivos 3GPP 196, N no es necesariamente el mismo para las presentaciones 182 y 190. Es decir, puede haber un número diferente de archivos 3GPP en la presentación 182 y la presentación 190, así como en cualquiera de las presentaciones del conjunto 198.

45 **[0122]** Los datos de cabecera 184 pueden describir ciertas características de la presentación 182, por ejemplo, un número de archivos 3GPP en la presentación 182, direcciones de memoria correspondientes al comienzo de cada uno de los archivos 3GPP 188, longitudes de los archivos 3GPP 188 (por ejemplo, en bytes), duraciones de los archivos 3GPP 188 (por ejemplo, en segundos, minutos, etc.), capacidades de descodificación y/o representación esperadas para que un dispositivo cliente descodifique y represente correctamente archivos 3GPP 188 u otros datos. La presentación 182 puede incluir datos adicionales, por ejemplo, como se describe en la norma del formato de archivo de medios de base de la ISO y/o extensiones del mismo, como el estándar 3GPP. Los archivos 3GPP 188 pueden incluir imágenes de vídeo codificadas. El archivo 3GPP con el conjunto de datos de secuencia 186 incluye un conjunto de datos de secuencia que describe imágenes de vídeo codificadas de una secuencia de imágenes de vídeo codificadas, por ejemplo, como se incluye en uno o más de los archivos de 3GPP 188.

55 **[0123]** En el ejemplo de la FIG. 5, las presentaciones del conjunto 198 pueden corresponder a diversas presentaciones de la misma secuencia de datos de vídeo. Es decir, cada una de las presentaciones puede ser similar durante la reproducción, por ejemplo, cuando se descodifica y se representa. Sin embargo, los datos de vídeo pueden codificarse o presentarse de manera diferente dentro de diferentes presentaciones del conjunto 198. Por ejemplo, las muestras de vídeo de la presentación 190 pueden tener una resolución más alta o más baja, codificadas con una tasa de bits más alta o más baja, esperar un tamaño de búfer de descodificador particular, tener una frecuencia de trama particular, esperar una velocidad de descodificación de macrobloques particular o diferenciarse entre sí basándose en otras características. Una unidad de encapsulación, tal como la unidad de encapsulación 30 (FIG. 1) puede colocar un conjunto de datos de secuencia en un archivo 3GPP dentro de una presentación, tal como el archivo 3GPP 186 de la presentación 182 y el archivo 3GPP 194 de la presentación 190. De forma adicional o alternativa, la unidad de

encapsulación 30 puede almacenar datos de secuencia para una o más presentaciones del conjunto 198 en la MPD 180.

5 **[0124]** La FIG. 6 es un diagrama de flujo que ilustra un procedimiento de ejemplo para encapsular un conjunto de datos de secuencia dentro de un archivo de vídeo y proporcionar el archivo de vídeo desde un dispositivo de origen a un dispositivo de destino. Aunque se describe con respecto a los componentes del dispositivo de origen 20 y el dispositivo de destino 40 (FIG. 1) con fines de ejemplo y explicación, debe entenderse que cualquier dispositivo adecuado puede implementar las técnicas de la FIG. 6.

10 **[0125]** Inicialmente, la unidad de encapsulación 30 puede recibir una secuencia de imágenes de vídeo codificadas (210). Un codificador, como el codificador de vídeo 28, puede haber incluido mensajes de SEI a nivel de secuencia con muestras codificadas. De forma adicional o alternativa, la unidad de encapsulación 30 puede crear conjuntos de parámetros por separado del codificador de vídeo 28. En cualquier caso, la unidad de encapsulación 30 puede desacoplar mensajes de SEI a nivel de secuencia de imágenes de vídeo codificadas descritas por los mensajes de SEI a nivel de secuencia (212).

15 **[0126]** Es decir, la unidad de encapsulación 30 puede crear un archivo de vídeo que incluye mensajes de SEI a nivel de secuencia e imágenes de vídeo codificadas descritas por los mensajes de SEI a nivel de secuencia (214). Al hacerlo, sin embargo, la unidad de encapsulación 30 puede almacenar los mensajes de SEI a nivel de secuencia por separado de las imágenes de vídeo codificadas descritas por los mensajes de SEI a nivel de secuencia. De esta manera, los conjuntos de parámetros pueden transmitirse y procesarse por separado de las imágenes de vídeo codificadas. Por ejemplo, de acuerdo con las técnicas de esta divulgación, la unidad de encapsulación 30 puede almacenar los mensajes de SEI a nivel de secuencia en una pista de conjunto de parámetros del archivo de vídeo creado. Para mencionar otro ejemplo, la unidad de encapsulación 30 puede almacenar los mensajes de SEI a nivel de secuencia en un fragmento de pista que no incluye imágenes de vídeo codificadas, un archivo 3GPP separado que no incluye imágenes de vídeo codificadas, una estructura de datos similar a una imagen de vídeo codificada pero eso sí en realidad no incluye datos de vídeo codificados, en una descripción de presentación de medios que describe el archivo de vídeo, o de algún modo separa los mensajes de SEI a nivel de secuencia de las imágenes de vídeo codificadas descritas por los mensajes de SEI a nivel de secuencia.

20 **[0127]** La unidad de encapsulación 30 puede generar a continuación el archivo de vídeo (216). Por ejemplo, la unidad de encapsulación 30 puede hacer que el dispositivo de origen 20 escriba el archivo de vídeo en un medio de almacenamiento, tal como, por ejemplo, un disco óptico, un disco flexible, una unidad de memoria flash, un disco duro, una unidad de estado sólido u otro medio de almacenamiento. Dichos medios de almacenamiento pueden transportarse físicamente al dispositivo de destino 40. De forma alternativa, el dispositivo de origen 20 puede transmitir el archivo de vídeo al dispositivo de destino 40, por ejemplo, a través de radiodifusión, transmisión de red u otras técnicas de transmisión. En cualquier caso, el dispositivo de destino 40 puede recibir en última instancia el archivo de vídeo (218).

25 **[0128]** En algunos ejemplos, el dispositivo de origen 20 puede proporcionar porciones distintas del archivo de vídeo al dispositivo de destino 40, por ejemplo, en respuesta a una o más peticiones GET HTTP o GET parcial emitidas por el dispositivo de destino 40 al dispositivo de origen 20. El dispositivo de destino 40 puede emitir una primera petición GET HTTP o GET parcial al dispositivo de origen 20 para recuperar un conjunto de datos de secuencia, por ejemplo, todo o una porción de una pista de conjunto de parámetros que incluye mensajes de SEI a nivel de secuencia y una segunda (o más) petición o peticiones GET HTTP o GET parcial para recuperar imágenes de vídeo codificadas descritas por el conjunto de datos de secuencia.

30 **[0129]** Después de recibir el archivo de vídeo, el dispositivo de destino 40 puede descodificar el archivo de vídeo basándose en los mensajes de SEI a nivel de secuencia (220). Es decir, el descodificador de vídeo 48 puede usar datos de los mensajes de SEI a nivel de secuencia para ayudar en el proceso de descodificación. Del mismo modo, la pantalla de vídeo 44 puede usar los mensajes de SEI a nivel de secuencia para ayudar en la representación de los datos de vídeo descodificados, con el fin de visualizar los datos de vídeo descodificados (222). Por ejemplo, el descodificador de vídeo 48 y/o la pantalla de vídeo 44 (o una unidad de representación o un motor de representación del mismo) pueden usar los mensajes de SEI a nivel de secuencia para determinar cómo adaptar los datos de vídeo a una resolución nativa de la pantalla de vídeo 44, determinar si una imagen es una imagen izquierda o derecha (por ejemplo, para datos de vídeo estéreo), determinar cómo se empaquetan las tramas, determinar la información de ajustabilidad a escala para los datos de vídeo, determinar la información de adquisición multivista y/o determinar la información de escena multivista.

35 **[0130]** La FIG. 7 es un diagrama de bloques que ilustra otro ejemplo de un archivo de vídeo 240. En el ejemplo de la FIG. 7, el archivo de vídeo 240 incluye el cuadro de película (MOOV) 242 y el cuadro de datos de película (MDAT) 250. El cuadro de MOOV 242 incluye una pista de datos de vídeo y un número de pistas de conjuntos de parámetros 246A-246C. Aunque se ilustran tres pistas de conjuntos de parámetros en el ejemplo de la FIG. 7, debe entenderse que un archivo de vídeo en general puede incluir un número cualquiera de pistas de conjuntos de parámetros.

65

[0131] La pista de datos de vídeo 244 corresponde a un número de diversos fragmentos de vídeo 254A-254C, cada uno de los cuales puede almacenarse en el cuadro de MDAT 250. En este ejemplo, cada conjunto de fragmentos de vídeo 254A, 254B y 254C puede tener información diferente a nivel de secuencia. Por ejemplo, la información a nivel de secuencia puede cambiar entre los fragmentos de vídeo 254A, 254B y 254C. Por consiguiente, cada uno de los fragmentos de vídeo 254A, 254B y 254C puede tener un conjunto de datos de secuencia respectivo. En este ejemplo, el conjunto de datos de secuencia 252A se refiere a los fragmentos de vídeo 254A, el conjunto de datos de secuencia 252B se refiere a los fragmentos de vídeo 254B y el conjunto de datos de secuencia 252C se refiere a los fragmentos de vídeo 254C.

[0132] El conjunto de datos de secuencia para un conjunto dado de fragmentos de vídeo puede aplicarse a cada uno de los fragmentos de vídeo en la secuencia. Por ejemplo, el conjunto de datos de secuencia 252A puede aplicarse a cada uno de los fragmentos de vídeo 254A. Sin embargo, una o más características de los fragmentos de vídeo 254B pueden diferir de los fragmentos de vídeo 254A. Por consiguiente, el conjunto de datos de secuencia 252B puede describir las características de los fragmentos de vídeo 254B.

[0133] Cada conjunto de datos de secuencia puede corresponder a una pista respectiva de conjunto de parámetros. En el ejemplo de la FIG. 7, la pista de conjunto de parámetros 246A corresponde al conjunto de datos de secuencia 252A, la pista de conjunto de parámetros 246B corresponde al conjunto de datos de secuencia 252B, y la pista de conjunto de parámetros 246C corresponde al conjunto de datos de secuencia 252C. Cada pista de conjunto de parámetros puede incluir datos que describen un tiempo durante el cual el conjunto de datos de secuencia correspondiente se aplica a fragmentos de vídeo de la pista de datos de vídeo 244. Por ejemplo, la pista de conjunto de parámetros 246A puede incluir datos que indican que el conjunto de datos de secuencia 252A ya no se aplica después del tiempo correspondiente al último de los fragmentos de vídeo 254A. Esta información de temporización puede ser el tiempo UTC o corresponder a una duración de la reproducción.

[0134] En algunos ejemplos, un archivo de vídeo puede incluir una pista adicional de conjunto de parámetros que corresponde a todos los fragmentos de vídeo 254A-254C. En dichos ejemplos, un dispositivo de destino puede utilizar datos tanto de la pista de conjunto de parámetros que se aplica a todos los fragmentos de vídeo como datos de la pista de conjunto de parámetros que se aplica particularmente a un fragmento de vídeo dado.

[0135] Aunque se ilustra con fines de ejemplo como incluido en el cuadro de MOOV 242, debe entenderse que los conjuntos de datos de secuencia 252 pueden incluirse en cualquiera de las otras ubicaciones descritas en esta divulgación. Por ejemplo, los conjuntos de datos de secuencia pueden incluirse en fragmentos de vídeo particulares, archivos 3GPP particulares, dentro de una descripción de presentación de medios, u otra ubicación o ubicaciones dentro de, o asociadas con, un archivo de vídeo.

[0136] 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 se pueden almacenar en, o transmitir por, un medio legible por ordenador, como una o más instrucciones o código, y ejecutar 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 tal 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) medios de almacenamiento tangibles legibles por ordenador que son no transitorios o (2) un medio de comunicación tal como una señal o una onda portadora. Los medios de almacenamiento de datos pueden ser cualquier medio disponible al que se puede acceder desde uno o más ordenadores o uno o más procesadores para recuperar instrucciones, código y/o estructuras de datos para la implementación de las técnicas descritas en esta divulgación. Un producto de programa informático puede incluir un medio legible por ordenador.

[0137] 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 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 debe entender que los medios de almacenamiento legibles por ordenador y los medios de almacenamiento de datos no incluyen conexiones, ondas portadoras, señales ni otros medios transitorios, sino que, en cambio, se refieren a medios de almacenamiento tangibles no transitorios. Los discos, como se usan en el presente documento, incluyen el disco compacto (CD), el disco láser, el disco óptico, el disco versátil digital (DVD), el disco flexible y el disco Blu-ray, donde algunos discos reproducen habitualmente los datos magnéticamente, mientras que otros discos reproducen los datos ópticamente con láseres. Las combinaciones de lo anterior también se deben incluir dentro del alcance de los medios legibles por ordenador.

5 [0138] 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. Por consiguiente, 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. Así mismo, en algunos aspectos, la funcionalidad descrita en el presente documento puede proporcionarse dentro de hardware especializado y/o módulos de software configurados para la codificación y la descodificación, o incorporarse en un códec combinado. Además, las técnicas se podrían implementar por completo en uno o más circuitos o elementos lógicos.

15 [0139] Las técnicas de la presente divulgación se pueden implementar en una amplia variedad de dispositivos o aparatos, incluyendo un teléfono inalámbrico, un circuito integrado (IC) o un conjunto de IC (por ejemplo, un conjunto de chips). En la presente 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 su lugar, 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 incluya uno o más procesadores como se describe anteriormente, junto con software y/o firmware adecuados.

20 [0140] Se han descrito diversos ejemplos. Estos y otros ejemplos están dentro del alcance de las siguientes reivindicaciones.

REIVINDICACIONES

1. Un procedimiento para emitir de datos de vídeo encapsulados que comprende una secuencia de imágenes de vídeo codificadas, el procedimiento que comprende:

5 generar uno o más mensajes de información de mejora complementaria, SEI, que están separados de las imágenes de vídeo codificadas, en el que los mensajes de SEI describen imágenes de vídeo codificadas en una secuencia respectiva de la secuencia de imágenes de vídeo codificadas e incluyen elementos comunes a más de una de las imágenes de vídeo codificadas en la secuencia respectiva;

10 encapsular los mensajes de SEI en un primer archivo que comprende un segmento de inicialización para una representación de medios, el segmento de inicialización encapsula además al menos uno de entre un conjunto de parámetros de secuencia, SPS, y/o un conjunto de parámetros de imagen, PPS, para la secuencia de imágenes de vídeo codificadas, en el que el primer archivo y el segmento de inicialización no incluyen ningún dato medios y en el que el primer archivo se ajusta al formato de archivo de medios de base de la ISO o una extensión del mismo;

15 encapsular la secuencia de imágenes de vídeo codificadas en uno o más archivos de vídeo separados del primer archivo, y en el que uno o más archivos de vídeo separados del primer archivo también se ajustan al formato de archivo de medios de base de la ISO o una extensión del mismo;

20 recibir una petición de acuerdo con un protocolo de transmisión continua por la red para un conjunto de datos de secuencia, en el que la petición comprende una petición HTTP GET y

25 enviar el primer archivo en respuesta a la petición antes de emitir uno cualquiera de los uno o más archivos de vídeo separados.

2. El procedimiento de la reivindicación 1, en el que los mensajes de SEI comprenden mensajes de SEI a nivel de secuencia aplicables a dos o más unidades de acceso, en el que las dos o más unidades de acceso comprenden la secuencia de imágenes de vídeo codificadas, y en el que generar los uno o más mensajes de SEI comprende separar los mensajes de SEI a nivel de secuencia de las unidades de acceso.

3. El procedimiento de la reivindicación 1, en el que los mensajes de SEI se definen como al menos uno de entre un tipo de mensaje de SEI de la codificación de vídeo avanzada (AVC)/H.264, extensión de la codificación de vídeo ajustable a escala (SVC) de H.264 y extensión de la codificación de vídeo multivista (MVC) de H.264.

4. El procedimiento de la reivindicación 1, en el que el primer archivo que contiene los mensajes de SEI comprende al menos una porción de una pista de conjunto de parámetros, en el que la pista de conjunto de parámetros comprende además al menos uno de entre los datos del conjunto de parámetros de secuencia y los datos del conjunto de parámetros de imagen para la secuencia de imágenes de vídeo codificadas.

5. El procedimiento de la reivindicación 1, en el que generar los mensajes de SEI comprende generar información de señalización dentro de al menos uno de entre los mensajes de SEI que señala una duración de la secuencia de imágenes de vídeo codificadas correspondientes a los mensajes de SEI.

6. Un procedimiento para recibir datos de vídeo codificados, el procedimiento que comprende:

solicitar un conjunto de datos de secuencia utilizando una petición HTTP GET;

50 recibir, en respuesta a la petición HTTP GET, mensajes de información de mejora complementaria encapsulada, SEI, en un primer archivo que comprende un segmento de inicialización para una representación de medios, en el que los mensajes de SEI describen una secuencia de imágenes de vídeo codificadas de uno o más archivos de vídeo separados del primer archivo, en el que los mensajes de SEI describen imágenes de vídeo codificadas de la secuencia de imágenes de vídeo codificadas e incluyen elementos comunes a más de una de las imágenes de vídeo codificadas, el segmento de inicialización además comprende encapsulaciones de al menos uno de entre un conjunto de parámetros de secuencia, SPS, y/o conjunto de parámetros de imagen, PPS, en el que el segmento de inicialización no incluye ningún dato de medios y en el que el primer archivo se ajusta al formato de archivo de medios de base de la ISO o una extensión de los mismos;

60 después de recibir los mensajes de SEI encapsulados del primer archivo, recibir la secuencia encapsulada de imágenes de vídeo codificadas en uno o más archivos de vídeo separados del primer archivo y en el que uno o más archivos de vídeo separados del primer archivo también se ajustan al formato de archivo de medios de base de la ISO o una extensión del mismo; y

65 representar la secuencia de imágenes de vídeo codificadas en base a los mensajes de SEI después de que se haya descodificado la secuencia de imágenes de vídeo codificadas.

7. El procedimiento de la reivindicación 6, que comprende además:

5 enviar una segunda petición para una o más de la secuencia encapsulada de imágenes de vídeo codificadas solo después de recibir los mensajes de SEI encapsulados, en el que recibir la secuencia encapsulada de imágenes de vídeo codificadas comprende recibir las una o más de la secuencia encapsulada de imágenes de vídeo codificadas en respuesta a la segunda petición.

10 8. El procedimiento de la reivindicación 7, en el que los mensajes de SEI comprenden un primer conjunto de mensajes de SEI, y en el que la secuencia de imágenes de vídeo codificadas comprende una primera secuencia de imágenes de vídeo codificadas, el procedimiento que comprende además:

15 enviar una tercera petición para un segundo conjunto de mensajes de SEI después de recibir la primera secuencia de imágenes de vídeo codificadas;

recibir el segundo conjunto de mensajes de SEI en respuesta a la tercera petición;

20 enviar una cuarta petición para una segunda secuencia de imágenes de vídeo codificadas descritas por el segundo conjunto de mensajes de SEI;

recibir la segunda secuencia de imágenes de vídeo codificadas después de recibir el segundo conjunto de mensajes de SEI; y

25 representar la segunda secuencia de imágenes de vídeo codificadas en base al segundo conjunto de mensajes de SEI después de que se haya descodificado la segunda secuencia de imágenes de vídeo codificadas.

30 9. El procedimiento de la reivindicación 6, que comprende además extraer los mensajes de SEI de al menos una de entre una pista de un conjunto de parámetros, un fragmento de pista que no incluye imágenes de vídeo codificadas y un archivo de formato del Proyecto de Asociación de Tercera Generación (3GPP) que no incluye datos de vídeo codificado.

35 10. El procedimiento de la reivindicación 6, que comprende además analizar y descodificar los mensajes de SEI, en el que los mensajes de SEI se codifican por separado de las unidades de acceso que comprenden la secuencia de imágenes de vídeo codificadas.

11. Un dispositivo que comprende uno o más medios para realizar el procedimiento de cualquiera de las reivindicaciones 1 a 10.

40 12. El dispositivo de la reivindicación 11, en el que los uno o más medios comprenden uno o más procesadores.

13. Un medio legible por ordenador que tiene almacenadas en el mismo instrucciones que cuando se ejecutan hacen que uno o más procesadores ejecuten el procedimiento de cualquiera de las reivindicaciones 1 a 10.

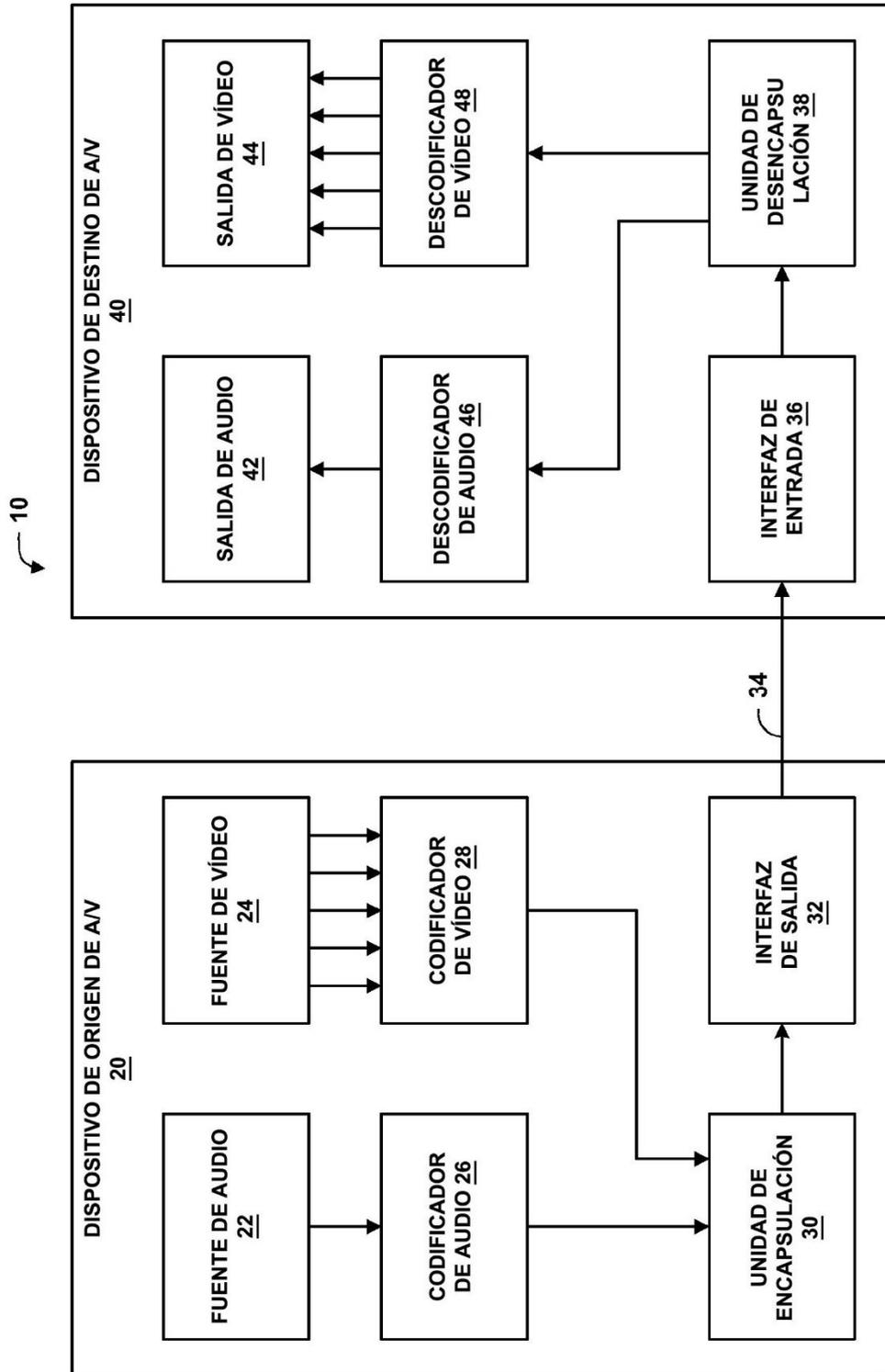


FIG. 1

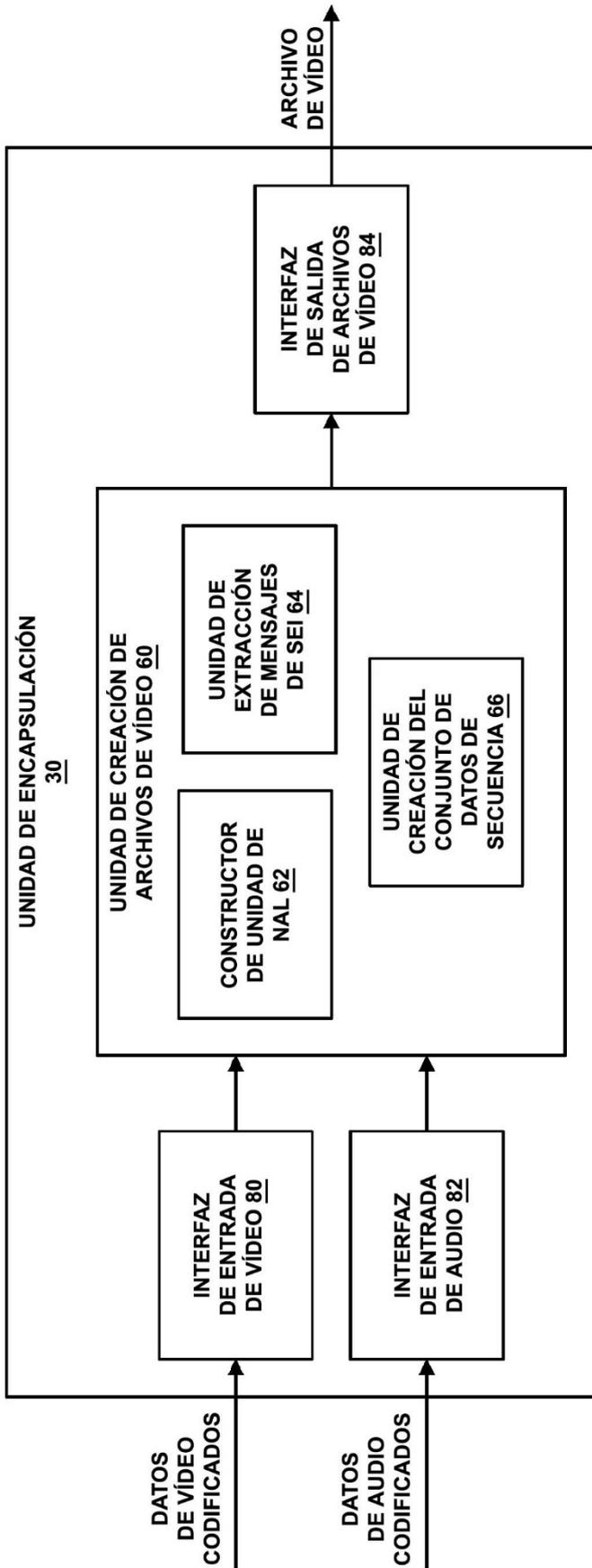


FIG. 2

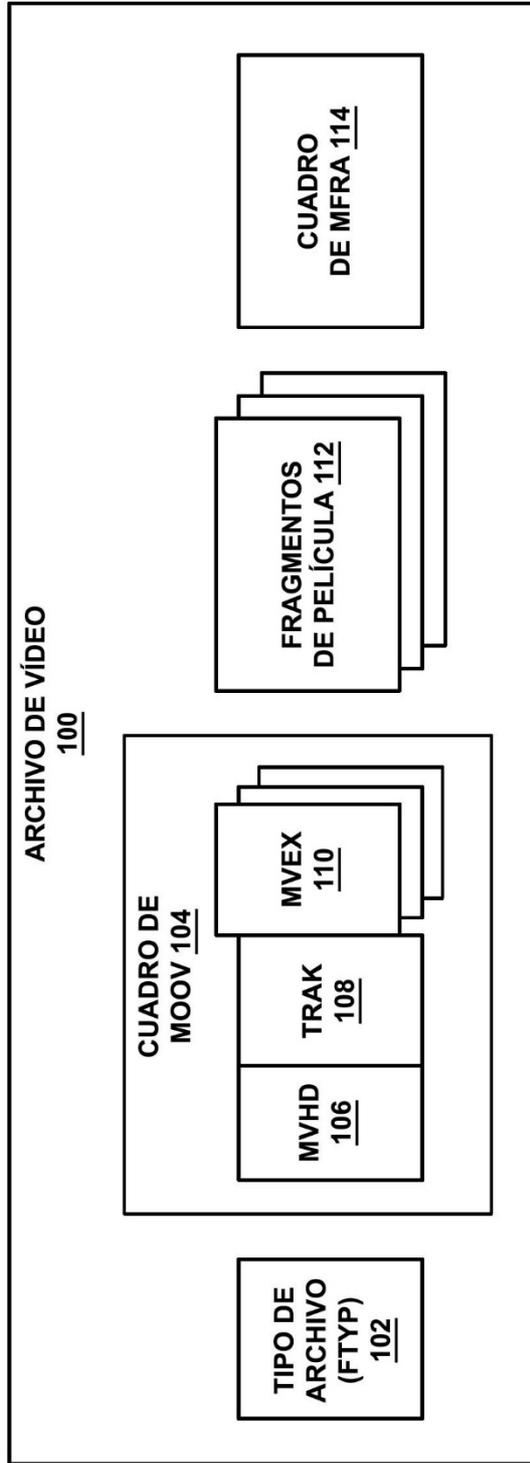


FIG. 3

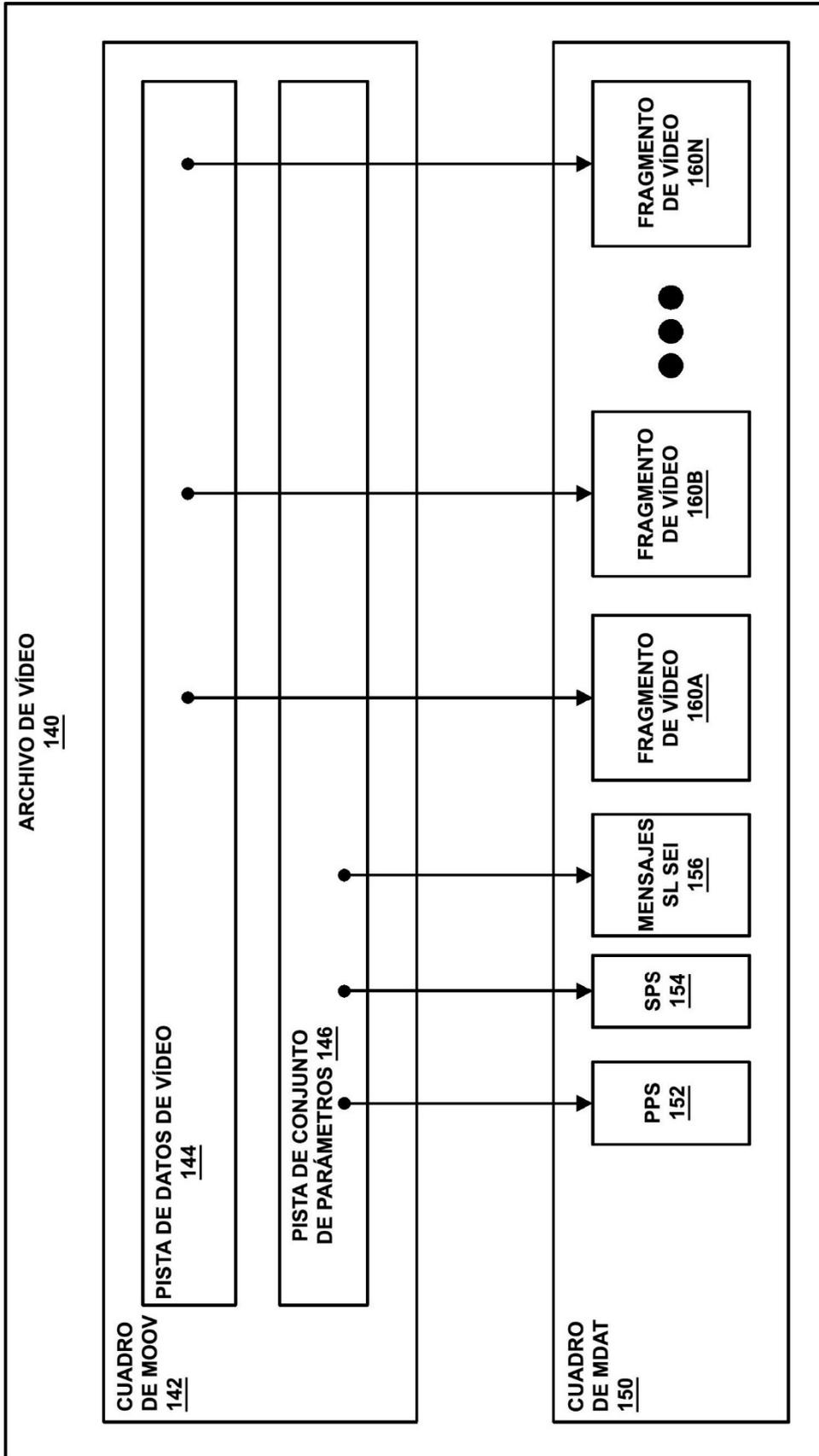


FIG. 4

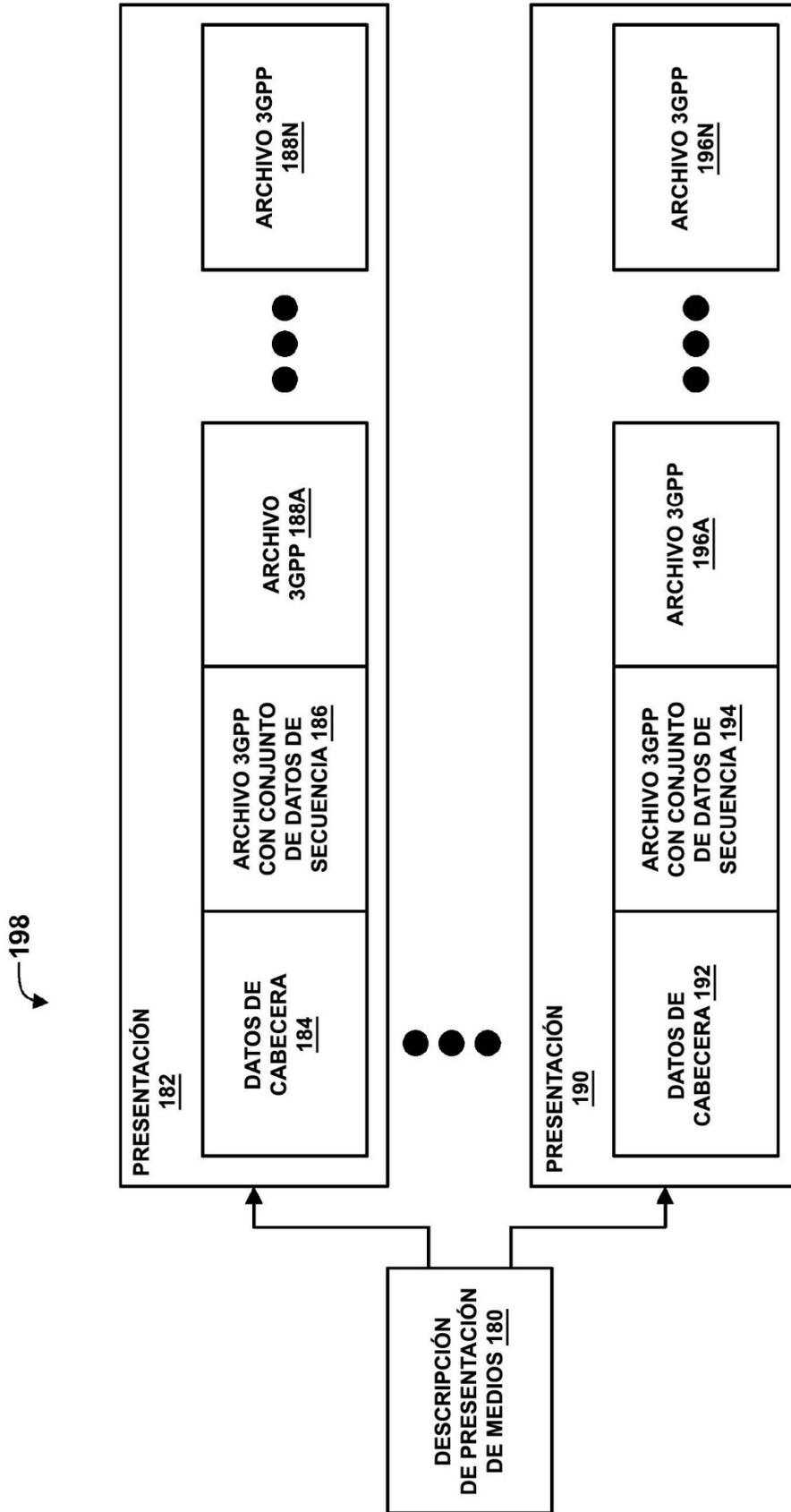


FIG. 5

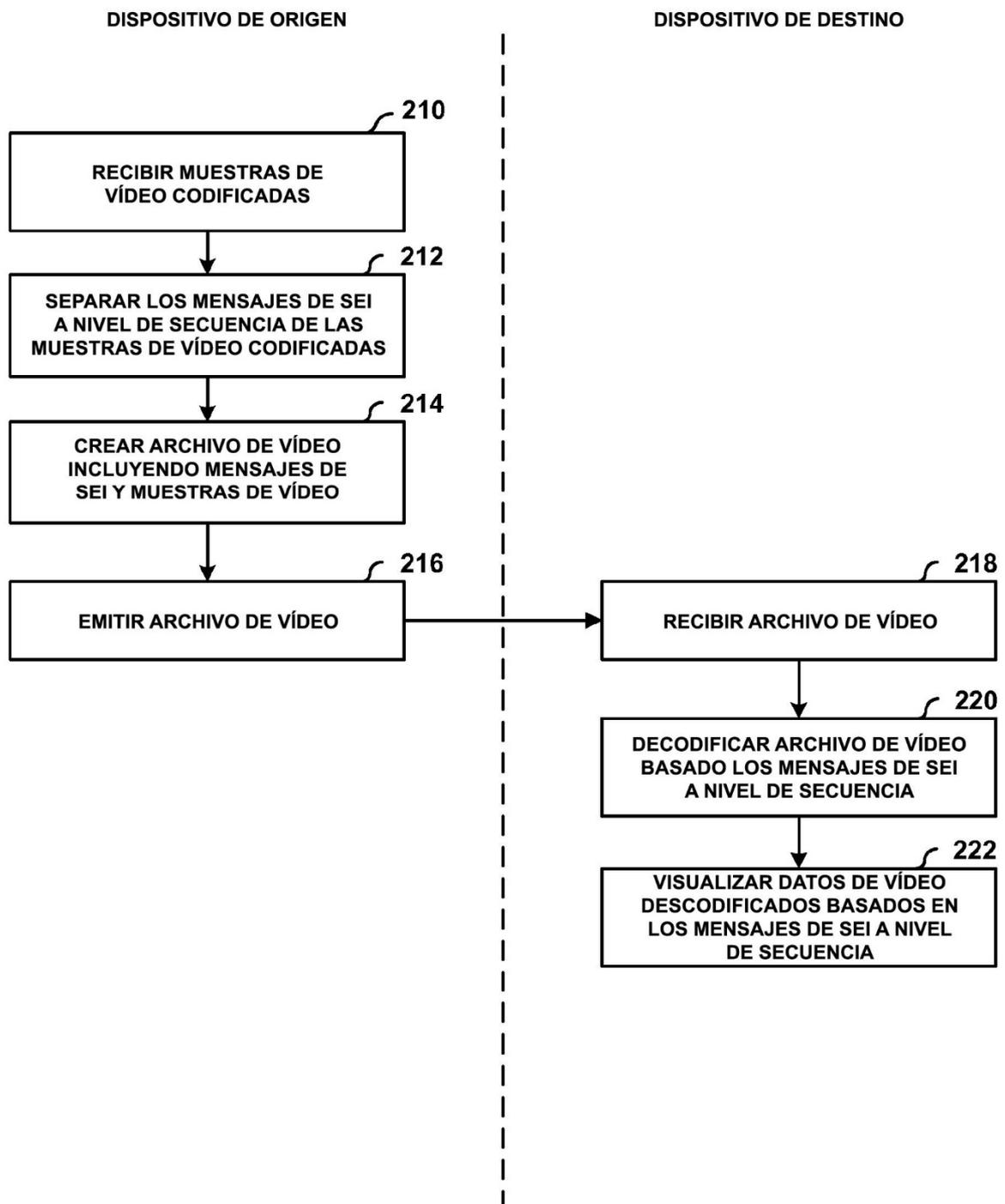


FIG. 6

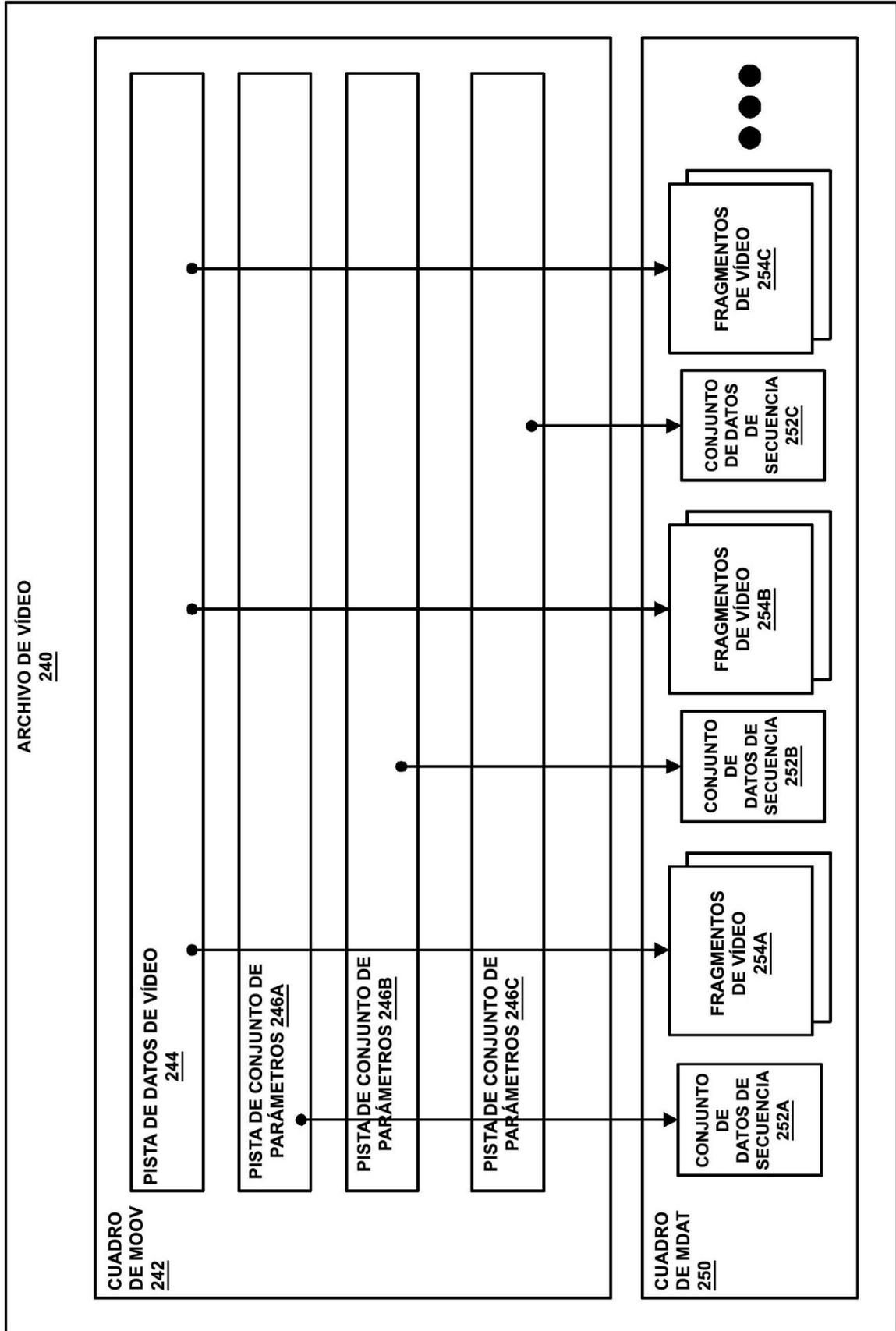


FIG. 7