

19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 494 541**

51 Int. Cl.:

H04N 19/31 (2014.01)

H04N 19/597 (2014.01)

H04N 19/61 (2014.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

96 Fecha de presentación y número de la solicitud europea: **17.09.2010 E 10776851 (7)**

97 Fecha y número de publicación de la concesión europea: **25.06.2014 EP 2478703**

54 Título: **Procedimientos de codificación de video multi-pista y aparato que usa un extractor que referencia dos o más unidades NAL no consecutivas**

30 Prioridad:

16.09.2009 US 243030 P

22.09.2009 US 244827 P

11.01.2010 US 293961 P

15.01.2010 US 295261 P

24.05.2010 US 785851 P

45 Fecha de publicación y mención en BOPI de la traducción de la patente:
15.09.2014

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:

CARPINTERO LÓPEZ, Mario

ES 2 494 541 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

Procedimientos de codificación de video multi-pista y aparato que usa un extractor que referencia dos o más unidades NAL no consecutivas

Campo técnico

- 5 Esta divulgación se refiere al transporte de datos de video codificados.

Antecedentes

10 Las capacidades del video digital se pueden incorporar a una amplia gama de dispositivos, incluyendo los televisores digitales, los sistemas de difusión directa digitales, los sistemas de difusión inalámbrica, los asistentes digitales personales (PDA), los ordenadores portátiles o de sobremesa, las cámaras digitales, los dispositivos de grabación digital, los reproductores de medios digitales, los dispositivos de videojuegos, las consolas de videojuegos, los teléfonos de radio celulares o de satélite, los dispositivos de teleconferencia de video y similares. Los dispositivos de video digitales implementan técnicas de compresión de video, tales como las descritas en las normativas definidas por MPEG-2, MPEG-4, ITU-T H.263 o ITU-T H.264 / MPEG-4, Parte 10, la Codificación de Video Avanzada (AVC), y extensiones de tales normativas, para transmitir y recibir información de video digital de forma más eficiente.

15 Las técnicas de compresión de video realizan predicción espacial y/o predicción temporal para reducir o eliminar la redundancia inherente en las secuencias de video. Para la codificación de video basada en bloques, una trama de video o corte (slice) se puede dividir en macrobloques. Cada macrobloque se puede dividir adicionalmente. Los macrobloques en una trama o corte intra codificados (I) se codifican usando predicción espacial con respecto a los macrobloques vecinos. Los macrobloques en una trama o corte (P o B) inter codificados pueden usar predicción espacial con respecto a los macrobloques vecinos en la misma trama o corte o predicción temporal con respecto a otras tramas de referencia.

20 Después de que se han codificado los datos de video, los datos de video se pueden empaquetar por un multiplexor para su transmisión o almacenamiento. El MPEG-2 incluye una sección de "Sistemas" que define un nivel de transporte para muchas normativas de codificación de video. Los sistemas del nivel de transporte de MPEG-2 se pueden usar por los codificadores de video MPEG-2, u otros codificadores de video conformes con las diferentes normativas de codificación de video. Por ejemplo, el MPEG-4 describe diferentes metodologías de codificación y decodificación que las del MPEG-2, pero los codificadores de video que implementan las técnicas de la normativa MPEG-4 aún pueden usar las metodologías del nivel de transporte del MPEG-2. En general, las referencias a los "sistemas MPEG-2" se refieren al nivel de transporte de los datos de video prescrito por el MPEG-2. El nivel de transporte prescrito en el MPEG-2 también se denomina en la presente divulgación como un "flujo de transporte de MPEG-2" o simplemente un "flujo de transporte". Del mismo modo, el nivel de transporte de los sistemas MPEG-2 también incluye flujos de programa. Los flujos de transporte y los flujos de programa en general incluyen formatos diferentes para la entrega de datos similares, donde un flujo de transporte comprende uno o más "programas" que incluyen tanto datos de audio como de video, aunque los flujos de programa incluyen un programa que incluye tanto datos de audio como de video.

25 Se han hecho esfuerzos para desarrollar nuevas normativas de codificación de video basadas en H.264 / AVC. Una de tales normativas es la normativa de codificación de video escalable (SVC), que es la extensión escalable para la H.264 / AVC. Otra normativa es la codificación de video multi-visión (MVC), que convierte la extensión multi-visión a la H.264 / AVC. La especificación de los Sistemas MPEG-2 describe cómo se pueden multiplexar los flujos de datos multimedia comprimidos (video y audio) junto con los otros datos para formar un flujo de datos único adecuado para la transmisión o almacenamiento digital. La última especificación de los sistemas MPEG-2 se especifica en el documento "Information Technology-Generic Coding of Moving Pictures and Associated Audio System, Recommendation H.222.0; International Organization for Standardisation ISO / IEC JTC1/SC29/WG11; Coding of Moving Pictures and Associated Audio" de mayo de 2006. El MPEG ha diseñado recientemente la normativa de transporte de MVC sobre sistemas MPEG-2 y la última versión de esta especificación es "Study of ISO/IEC 13818 - 1:2007/FPDAM4 Transport of MVC", documento de MPEG N10572, MPEG de ISO/IEC JTC1/SC29/WG11, de Maui, Hawaii, USA, abril de 2009.

30 El borrador de la última reunión de MVC se describe en el documento JVT-AB204 "Join Draft 8.0 on Multiview Video Coding" de la 28ª reunión de JVT, Hanover, Alemania, julio de 2008 disponible en la dirección http://wftp3.itu.int/av-arch/jvt-site/2008_07_Hannover/JVT-AB204.zip. Una versión posterior integrada dentro de la normativa de AVC se describe en el documento de JVT-AD007, "Editors draft revision to ITU-T Rec. H264 | ISO/IEC 14496-10 Advanced Video Coding - in preparation for ITU-T SG 16 AAP Consent (in integrated form)" 30ª reunión de JVT, Génova, CH, febrero de 2009, disponible en http://wftp3.itu.int/av-arch/jvt-site/jvt-site/2009_01_Genova/JVT-AD007.zip.

35 El documento "Deliverable D3.MVC/SVC storage format" de Karsten Gruneberg y otros, Programa de las Tecnologías de la Información y Comunicación (ICT), Proyecto N° FP7-ICT-214063, del 29 de enero de 2009 desvela el uso de extractores que pueden referenciar una o más unidades NAL consecutivas de una pista y que se pueden des-referenciar para suministrar los datos referenciados al campo de datos en lugar de al propio extractor.

Sumario

En general la presente divulgación describe las técnicas para el uso de extractores de medios en formatos de datos de video multi-pista para formar una pista de extractores de medios. La presente divulgación modifica el formato de medios de base de la Organización Internacional para la Normalización (ISO) para usar un extractor que es capaz de referenciar una o más unidades de la capa de acceso de red (NAL) potencialmente no consecutivas. Tal extractor puede estar presente en cualquier pista de un fichero del formato de medios de base de la ISO. La presente divulgación también describe las modificaciones al formato de ficheros del Proyecto de Miembros de la Tercera Generación (3GPP) para incluir un valor de la tasa de trama como un atributo de un bloque de selección de pista. La presente divulgación describe además, con respecto a la codificación de video multi-visión (MVC) la extensión al formato de medios de base de la ISO, el uso del extractor para soportar la extracción suficiente de los puntos de operación de MVC.

En un ejemplo, un procedimiento de la codificación de datos de video incluye la construcción por un dispositivo de video fuente, de una primera pista que incluye una muestra de video que comprende una pluralidad de unidades de la capa de acceso de red (NAL) basadas en los datos de video codificados, en el que la muestra de video está incluida en una unidad de acceso, construyendo, por el dispositivo de video fuente, una segunda pista que incluye un extractor que identifica al menos una de la pluralidad de unidades NAL en la muestra de video de la primera pista, comprendiendo la al menos una de la pluralidad de unidades NAL una primera unidad NAL identificada, y en el que el extractor identifica una segunda unidad NAL de la unidad de acceso, en el que la primera unidad NAL identificada y la segunda unidad NAL identificada son no consecutivas, incluyendo la primera pista y la segunda pista en un fichero de video conforme con el formato de ficheros de medios de base de la Organización Internacional para la Normalización (ISO), y emitir el fichero de video.

En otro ejemplo, un aparato de codificación de los datos de video incluye un codificador configurado para codificar los datos de video, un multiplexor configurado para construir una primera pista que incluye una muestra de video que comprende una pluralidad de unidades de la capa de acceso de red (NAL), en base a los datos de video codificados, en el que la muestra de video está incluida en una unidad de acceso, construir una segunda pista que incluye un extractor que identifica al menos una de la pluralidad de unidades NAL en la muestra de video de la primera pista, comprendiendo la al menos una unidad de la pluralidad de unidades NAL una primera unidad NAL identificada y en el que el extractor identifica una segunda unidad NAL de la unidad de acceso, en el que la primera unidad NAL identificada y la segunda unidad NAL identificada son no consecutivas, incluir la primera pista y la segunda pista en un fichero de video conforme con el formato de ficheros de medios de base de la Organización Internacional para la Normalización (ISO), y una interfaz de salida configurada para emitir el fichero de video.

En otro ejemplo, un aparato de codificación de datos de video incluye medios para construir una primera pista que incluye una muestra de video que comprende una pluralidad de unidades de la capa de acceso de red (NAL) basadas en los datos de video codificados, en el que la muestra de video está incluida en una unidad de acceso, medios para construir una segunda pista que incluye un extractor que identifica al menos una de la pluralidad de unidades NAL en la muestra de video de la primera pista, comprendiendo la al menos una de la pluralidad de unidades NAL una primera unidad NAL identificada, y en el que el extractor identifica una segunda unidad NAL de la unidad de acceso, en el que la primera unidad NAL identificada y la segunda unidad NAL son no consecutivas, medios para incluir la primera pista y la segunda pista en un fichero de video conforme con el formato de ficheros de medios de base de la Organización Internacional para la Normalización (ISO), y medios para emitir el fichero de video.

En otro ejemplo, un medio de almacenamiento legible por ordenador que comprende instrucciones que, cuando se ejecutan, causan que un procesador del dispositivo fuente construya una primera pista que incluye una muestra de video una pluralidad de unidades de la capa de acceso de red (NAL) basadas en los datos de video codificados, en el que la muestra de video está incluida en una unidad de acceso, construir una segunda pista que incluye un extractor que identifica al menos una de la pluralidad de unidades NAL en la muestra de video de la primera pista, comprendiendo la al menos una de la pluralidad de unidades NAL una primer unidad NAL identificada, y en el que el extractor identifica una segunda unidad NAL de la unidad de acceso, en el que la primera unidad NAL identificada y la segunda unidad NAL identificada son no consecutivas, incluir la primera pista y la segunda pista en un fichero de video conforme con el formato de ficheros de medios de base de la Organización Internacional para la Normalización (ISO), y emitir el fichero de video.

En otro ejemplo, un procedimiento de decodificación de datos de video incluye recibir, por un demultiplexor de un dispositivo de destino, un fichero de video conforme con el formato de ficheros de medios de base de la Organización Internacional para la Normalización (ISO), comprendiendo el fichero de video una primera pista y una segunda pista, que incluye la primera pista una muestra de video que comprende una pluralidad de unidades de la capa de acceso de red (NAL) que corresponden a datos de video codificados, en el que la muestra de video está incluida en una unidad de acceso, e incluyendo la segunda pista un extractor que identifica al menos una de la pluralidad de unidades NAL de la primera pista, comprendiendo la al menos una de la pluralidad de unidades NAL una primera unidad NAL identificada, y en el que el extractor identifica una segunda unidad NAL de la unidad de acceso, en el que la primera unidad NAL identificada y la segunda unidad NAL identificada son no consecutivas, seleccionando la segunda pista a decodificar, y enviando los datos de video codificados de la primera unidad NAL y

la segunda unidad NAL identificada por el extractor de la segunda pista a un decodificador de video del dispositivo de destino.

En otro ejemplo, un aparato de decodificación de datos de video incluye un decodificador de video configurado para decodificar los datos de video, y un demultiplexor configurado para recibir un fichero de video conforme con el formato de ficheros de medios de base de la Organización Internacional para la Normalización (ISO), comprendiendo el fichero de video una primera pista y una segunda pista, incluyendo la primera pista una muestra de video que comprende una pluralidad de unidades de la capa de acceso de red (NAL) que corresponden a datos de video codificados, en el que la muestra de video está incluida en una unidad de acceso, e incluyendo la segunda pista un extractor que identifica al menos una de la pluralidad de unidades NAL de la primera pista, la al menos una de la pluralidad de unidades NAL que comprende una primera unidad NAL identificada y en el que el extractor identifica una segunda unidad NAL de la unidad de acceso, en el que la primera unidad NAL identificada y la segunda unidad NAL identificada son no consecutivas, seleccionar la segunda pista a decodificar y enviar datos de video codificados de la primera unidad NAL y la segunda unidad NAL identificada por el extractor de la segunda pista al decodificador de video.

En otro ejemplo, un aparato de decodificación de datos de video incluye medios para recibir, por un demultiplexor de un dispositivo de destino, un fichero de video conforme con el formato de ficheros de medios de base de la Organización Internacional para la Normalización (ISO), comprendiendo el fichero de video una primera pista y una segunda pista, incluyendo la primera pista una muestra de video que comprende una pluralidad de unidades de la capa de acceso de red (NAL) correspondientes a los datos de video codificados, en el que la muestra de video está incluida en una unidad de acceso, e incluyendo la segunda pista un extractor que identifica al menos una de la pluralidad de unidades NAL de la primera pista, comprendiendo la al menos una de la pluralidad de unidades NAL una primera unidad NAL identificada, y en el que el extractor identifica una segunda unidad NAL de la unidad de acceso, en el que la primera unidad NAL identificada y la segunda unidad NAL identificada son no consecutivas, medios para seleccionar la segunda pista a decodificar, y medios para enviar los datos de video codificados de la primera unidad NAL y la segunda unidad NAL identificada por el extractor de la segunda pista a un decodificador del dispositivo de destino.

En otro ejemplo, un medio de almacenamiento legible por ordenador se codifica con instrucciones que, cuando se ejecutan, causan que un procesador de un dispositivo de destino, una vez que recibe un fichero de video conforme con el formato de ficheros de medios de base de la Organización Internacional para la Normalización (ISO), comprendiendo el fichero de video una primera pista y una segunda pista, incluyendo la primera pista una muestra de video que comprende una pluralidad de unidades de la capa de acceso de red (NAL) correspondientes a datos de video codificados, en el que la muestra de video está incluida en una unidad de acceso, e incluyendo la segunda pista un extractor que identifica al menos una de la pluralidad de unidades NAL de la primera pista, comprendiendo la al menos una unidad de la pluralidad de unidades NAL una primera unidad NAL identificada, y en el que el extractor identifica una segunda unidad NAL de la unidad de acceso, en el que la primera unidad NAL identificada y la segunda unidad NAL identificada son no consecutivas, seleccionar la segunda pista a decodificar y enviar los datos de video codificados de la primera unidad NAL y la segunda unidad NAL identificada por el extractor de la segunda pista a un decodificador de video.

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

Breve descripción de los dibujos

La FIG. 1 es un diagrama de bloques que ilustra un sistema de ejemplo en el que un dispositivo fuente de audio / video (A/V) transporta datos de audio y de video a un dispositivo A/V de destino.

La FIG. 2 es un diagrama de bloques que ilustra una disposición de ejemplo de componentes de un multiplexor.

La FIG. 3 es un diagrama de bloques que ilustra un fichero de ejemplo que incluye una primera pista que tiene un conjunto de muestras de video y una segunda pista que tiene extractores que se refiere a un subconjunto de muestras de video de la primera pista.

La FIG. 4 es un diagrama de bloques que ilustra otro fichero de ejemplo que incluye dos pistas de extractores distintas.

La FIG.5 es un diagrama de bloques que ilustra otro fichero de ejemplo que incluye una pista de subconjuntos y dos pistas de extractor de medios.

Las FIG. 6A - 6C son diagramas de bloques que ilustran ejemplos de una caja de datos de medios de un fichero que incluye ejemplos de extractores de medios para diversas pistas de extractores de medios.

La FIG. 7 es un diagrama conceptual que ilustra un patrón de predicción de MVC de ejemplo.

Las FIG. 8 - 21 son diagramas de bloques que ilustran diversos ejemplos de estructuras de datos para extractores de medios y otras estructuras de soporte de datos que se pueden usar de acuerdo con las técnicas de la presente divulgación.

La FIG. 22 es un diagrama de bloques que ilustra un ejemplo de caja de selección de pista del Proyecto de Miembros de la Tercera Generación (3GPP) modificada para señalar los atributos adicionales para una caja de selección de pista.

La FIG. 23 es un diagrama de flujo que ilustra un procedimiento de ejemplo para usar extractores de medios de acuerdo con las técnicas de la presente divulgación.

Descripción detallada

5 Las técnicas de la presente divulgación se dirigen en general a mejorar el formato de ficheros de medios de base de la Organización Internacional para la Normalización (ISO) y extensiones del formato de ficheros de medios de base de la ISO. Las extensiones del formato de ficheros de medios de base de la ISO incluyen, por ejemplo, la codificación de video avanzada (AVC), la codificación de video escalable (SVC), la codificación de video multi-visión (MVC), y el formato de ficheros del Proyecto de Miembros de la Tercera Generación (3GPP). En general, las técnicas de la presente divulgación se pueden usar para producir una pista de extractor de medios en el formato de ficheros de medios de base de la ISO y/o las extensiones del formato de ficheros de medios de base de la ISO. Como se describe con mayor detalle más adelante, tales pistas de extractores de medios se pueden usar para soportar la adaptación en la descarga continua de video del protocolo de transporte de hipertexto (HTTP), en algunos ejemplos. En algunos ejemplos, un extractor de medios forma parte del formato de ficheros de medios de base de la ISO y/o las extensiones del formato de ficheros de medios de base de la ISO (por ejemplo, AVC, SVC, MVC, y 3GPP) para extraer muestras enteras de otra pista para formar una nueva pista de extractor de medios.

10 Estas técnicas se pueden usar por los sistemas de MPEG-2 (Grupo de Expertos de Imágenes en Movimiento), esto es, sistemas que son conformes con MPEG-2 con respecto a los detalles del nivel de transporte. El MPEG-4, por ejemplo, proporciona normativas para la codificación de video, pero en general asume que los codificadores de video que son conformes con la normativa MPEG-4 usarán los sistemas del nivel de transporte de MPEG-2. Por consiguiente, las técnicas de la presente divulgación son aplicables a codificadores de video que son conformes con MPEG-2, MPEG-4, ITU-T H.263, ITU-T H.264 / MPEG-4, o cualquier otra normativa de codificación de video que utilice los flujos de transporte de MPEG-2 y/o flujos de programa.

15 El formato de ficheros de medios de base de la ISO proporciona ficheros que incluyen una o más pistas. La normativa del formato de ficheros de medios de base de la ISO define una pista como una secuencia temporizada de muestras relacionadas. La normativa del formato de ficheros de medios de base de la ISO define una muestra como datos asociados con un único sello temporal, y proporciona ejemplos de una muestra como una trama individual de video, una serie de tramas de video en el orden de decodificación, o una sección comprimida de audio en el orden de decodificación. Las pistas especiales denominadas como pistas de indicaciones no contienen datos de medios, sino que en cambio tienen instrucciones para el empaquetamiento de una o más pistas dentro de un canal de descarga continua. La normativa del formato de ficheros de medios de base de la ISO observa que en las pistas de indicaciones, una muestra define la formación de uno o más paquetes de descarga continua.

20 Las técnicas de la presente divulgación proporcionan la creación de pistas de extractor de medios. Una pista de extractor de medios puede incluir en general uno o más extractores. Los extractores en una pista de extractores de medios se usan para identificar y extraer muestras de otra pista. De este modo, los extractores de medios en una pista de extractor de medios se pueden considerar punteros que, cuando se des-referencian, recuperan muestras desde otra pista. A diferencia de los extractores de SVC, por ejemplo, los extractores de la presente divulgación pueden referirse a una o más unidades de la capa de acceso de red (NAL) no consecutivas de otra pista. De acuerdo con las técnicas de la presente divulgación, las pistas de extractores de medios, las pistas que contienen uno o más extractores de medios y otras pistas que no incluyen un extractor de medios se pueden agrupar juntas para formar un grupo alternativo.

25 Esta divulgación usa el término "consecutivas" con respecto a las unidades NAL para describir dos o más unidades NAL que se producen en la misma pista de forma contigua. Esto es el último byte de datos en una de las unidades NAL precede inmediatamente al primer byte de datos de otra de las unidades NAL en la misma pista cuando las dos unidades NAL son consecutivas. Dos unidades NAL en la misma unidad de acceso se consideran en general como "no consecutivas" bien donde las dos unidades NAL están separadas por alguna cantidad de datos dentro de la misma pista, o donde una unidad NAL se produce en una pista mientras que la otra unidad NAL se produce en una pista diferente. La técnica de la presente divulgación proporciona un extractor que puede identificar dos o más unidades NAL no consecutivas de una unidad de acceso.

30 Además, los extractores de la presente divulgación no están limitados al SVC, sino que están incluidos en el formato de ficheros de medios de base de la ISO en general o cualquier otra extensión del formato de ficheros de medios de base de la ISO, tal como, por ejemplo, AVC, SVC, o MVC. Los extractores de la presente divulgación también se pueden incluir en el formato de ficheros del Proyecto de Miembros de la Tercera Generación (3GPP). La presente divulgación proporciona adicionalmente la modificación del formato de ficheros del 3GPP para indicar de forma explícita una tasa de trama como un atributo de una caja de selección de pista.

35 Las pistas de extractores de medios se pueden usar en el formato de ficheros de MVC, por ejemplo, para soportar la extracción de los puntos de operación. Un dispositivo servidor puede proporcionar diversos puntos de operación en un flujo de transporte de bits de la capa de transporte de MPEG-2, cada uno de los cuales corresponde a un subconjunto respectivo de vistas particulares de los datos de video de codificación de video multi-visión. Esto es, un punto de operación en general corresponde a un subconjunto de vistas de un flujo de bits. En algunos ejemplos,

cada vista de un punto de operación incluye datos de video en la misma tasa de trama. De acuerdo con las técnicas de la presente divulgación, un punto de operación se puede representar usando una pista de extractor de medios que incluye uno o más extractores que se refieren a datos de video de otras pistas y potencialmente a muestras adicionales no incluidas en otras pistas.

5 De este modo, cada punto de operación puede incluir solo las unidades NAL necesarias requeridas para la decodificación del punto de operación, para emitir un subconjunto de vistas con una tasa de trama común. La combinación de pistas de extractores con la representación completa del video de MVC puede formar una lista de reproducción de las representaciones de MVC. El uso de pistas de extractor de medios de la presente divulgación puede soportar la selección del punto de operación y conmutar por ejemplo, a puntos de operación con diversas
10 tasas de bit resultantes de la escalabilidad temporal.

Las pistas del extractor de medios de la presente divulgación también se pueden usar para formar grupos alternativos o grupos de conmutación. Esto es, en el formato de ficheros de medios de base de la ISO, las pistas se pueden agrupar juntas para formar grupos alternativos. En el ejemplo del formato de ficheros de medios de base de la ISO, las pistas de un grupo alternativo forman sustitutos viables entre sí, de modo que generalmente solo una de
15 las pistas de un grupo alternativo se reproduce o se descarga de forma continua en cualquier momento. Las pistas de un grupo alternativo deberían ser distinguibles de las otras pistas del grupo alternativo, por ejemplo, mediante atributos tales como la tasa de bit, códec, lenguaje, tamaño de los paquetes, u otras características. Las técnicas de la presente divulgación proporcionan el agrupamiento de pistas de extractores de medios, pistas que contienen extractores de medios, y/u otras pistas de video normales para formar un grupo alternativo. En ejemplos conformes
20 con la MVC, cada una de las pistas puede corresponder a un punto de operación respectivo. Esto es, cada punto de operación en la MVC se puede representar por una en particular de las pistas, por ejemplo, o bien una pista de extractores de medios o una pista que no incluye un extractor de medios. Una pista en el mismo grupo alternativo se selecciona típicamente para descargas progresivas, para adaptarse al ancho de banda disponible.

De forma similar, las pistas de extractores de medios y otras pistas se pueden agrupar juntas para formar un grupo de conmutación en el formato de ficheros de 3GPP, y se pueden usar para la selección de pista para adaptar el ancho de banda y la capacidad del decodificador en las aplicaciones de descarga continua de HTTP. El formato de ficheros de 3GPP proporciona una definición de un grupo de conmutación de pistas. Las pistas en un grupo de conmutación pertenecen al mismo grupo alternativo. Esto es, las pistas en el mismo grupo de conmutación están disponibles para conmutar durante una sesión, mientras que las pistas en diferentes grupos de conmutación no
25 están disponibles para la conmutación, de acuerdo con el formato de ficheros del 3GPP.

La FIG. 1 es un diagrama de bloques que ilustra un sistema de ejemplo 10 en el que el dispositivo fuente de audio / video (A/V) 20 transporta datos de audio y video al dispositivo de destino A/V 40. El dispositivo fuente A/V 20 se puede denominar también como un "dispositivo de video fuente". El sistema 10 de la FIG. 1 puede corresponder a un sistema de teleconferencia de video, un sistema de cliente / servidor, un sistema de difusor / receptor, o cualquier
35 otro sistema en el cual los datos de video se envían desde un dispositivo fuente, tal como un dispositivo fuente de A/V 20, a un dispositivo de destino, tal como un dispositivo de destino de A/V 40. El dispositivo de destino de A/V 40 también se puede denominar como un "dispositivo de video de destino" o un "dispositivo de cliente". En algunos ejemplos, el dispositivo fuente de A/V 20 y el dispositivo de destino de A/V 40 pueden realizar el intercambio de información bidireccional. Esto es, el dispositivo fuente de A/V 20 y el dispositivo de destino de A/V 40 pueden ser capaces tanto de codificar como decodificar (y transmitir y recibir) datos de audio y video. En algunos ejemplos, el codificador de audio 26 puede comprender un codificador de voz, también denominado como vocoder.

El dispositivo fuente de A/V 20, en el ejemplo de la FIG. 1, comprende una fuente de audio 22 y una fuente de video 24. La fuente de audio 22 puede comprender por ejemplo, un micrófono que produce señales eléctricas representativas de los datos de audio capturados para codificar por un codificador de audio 26. Como alternativa, la
45 fuente de audio 22 puede comprender un medio de almacenamiento que almacena los datos de audio grabados anteriormente, un generador de datos de audio tal como un sintetizador computarizado, o cualquier otra fuente de datos de audio. La fuente de video 24 puede comprender una cámara de video que produce datos de video a codificar por el codificador de video 28, un medio de almacenamiento codificado con datos de video grabados anteriormente, una unidad de generación de datos de video, o cualquier otra fuente de datos de video.

Los datos sin procesar de audio y video pueden comprender datos analógicos o digitales. Los datos analógicos se pueden digitalizar antes de codificarse por el codificador de audio 26 y/o el codificador de video 28. La fuente de audio 22 puede obtener datos de audio a partir de un participante orador mientras que el participante orador está hablando, y la fuente de video 24 puede obtener simultáneamente datos de video del participante orador. En otros
50 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 video 24 puede comprender un medio de almacenamiento legible por ordenador que comprende datos de video almacenados. De este modo, las técnicas descritas en la presente divulgación se pueden aplicar a emisiones en directo, descargas continuas, datos de audio y video en tiempo real o archivados, datos de audio y video pregrabados.

Las tramas de audio que corresponden a tramas de video son generalmente tramas de audio que contienen datos de audio que se capturaron por una fuente de audio 22 al mismo tiempo que los datos de video se capturan por la
60

fente de video 24 que están contenidos dentro de las tramas de video. Por ejemplo, mientras que un participante orador produce generalmente datos de audio hablando, la fuente de audio 22 captura los datos de audio y la fuente de video 24 captura los datos de video del participante orador al mismo tiempo, esto es, mientras que la fuente de audio 22 está capturando los datos de audio. Por lo tanto, una trama de audio pueden corresponder temporalmente a una o más tramas de video particulares. Por consiguiente, una trama de audio que corresponde a una trama de video generalmente corresponde a una situación en la que se capturaron datos de audio y datos de video al mismo tiempo y para los que la trama de audio y la trama de video comprenden, respectivamente, los datos de audio y los datos de video que se capturaron el mismo tiempo.

En algunos ejemplos, el codificador de audio 26 puede codificar un sello de tiempo en cada una de las tramas de audio codificada que representa un momento en el que los datos de audio se grabaron para la trama de audio codificada, y de forma similar, el codificador de video 28 puede codificar un sello temporal en cada trama de video codificada que representa el momento en el que se grabaron los datos de video para la trama de video codificada. En tales ejemplos, una trama de audio que corresponde a una trama de video puede comprender una trama de audio que comprende un sello temporal y una trama de video que comprende el mismo sello temporal. El dispositivo fuente de A/V 20 puede incluir un reloj interno desde el cual el codificador de audio 26 y/o el codificador de video 28 pueden generar los sellos temporales, o esa fuente de audio 22 y la fuente de video 24 se pueden usar para asociar los datos de audio y de video respectivamente, con un sello temporal.

En algunos ejemplos, la fuente de audio 22 puede enviar datos a un codificador de audio 26 correspondientes a un momento en el cual se grabaron los datos de audio, y la fuente de video 24 puede enviar datos a un codificador de video 28 que corresponden a un momento en el que se grabaron los datos de video. En tales ejemplos, el codificador de audio 26 puede codificar un identificador de secuencia en los datos de audio codificados para indicar un ordenamiento temporal relativo de los datos de audio codificados pero sin indicar necesariamente un momento absoluto en el que se grabaron los datos de audio, y de forma similar, el codificador de video 28 también puede usar identificadores de secuencia para indicar un ordenamiento temporal relativo de los datos de video codificados. De forma similar, en algunos ejemplos, un identificador de secuencia se puede mapear o correlacionar de otro modo con un sello temporal.

Las técnicas de la presente divulgación se dirigen en general al transporte de datos de multimedia codificados (por ejemplo, audio, video), y a la recepción y subsecuente interpretación y decodificación de los datos multimedia transportados. Las técnicas de la presente divulgación se pueden aplicar al transporte de datos de video de diversas normativas y extensiones tales como, por ejemplo, datos de la codificación de video escalable (SVC), la codificación de video avanzada (AVC), la capa base OSI, o la Codificación de Video Multi-visión (MVC), u otros datos de video que comprende una pluralidad de vistas. Como se muestra en el ejemplo de la FIG. 1, la fuente de video 24 puede proporcionar una pluralidad de vistas de una escena a un codificador de video 28. Las múltiples vistas de los datos de video pueden ser útiles para la generación de datos de video tridimensionales a usar por una pantalla tridimensional, tal como un estereoscopio, o una pantalla tridimensional auto-estereoscópica.

El dispositivo fuente de A/V 20 puede proporcionar un "servicio" a un dispositivo de destino A/V 40. Un servicio generalmente corresponde a un subconjunto de vistas disponibles de los datos de MVC. Por ejemplo, los datos de video multi-visión pueden estar disponibles para ocho vistas ordenadas desde cero hasta siete. Un servicio puede corresponder a un video estéreo que tiene dos vistas, mientras que otro servicio puede corresponder a cuatro vistas, y otro servicio más puede corresponder a todas las ocho vistas. En general, un servicio corresponde a cualquier combinación (esto es, cualquier subconjunto) de vistas disponibles. Un servicio también puede corresponder a una combinación de vistas disponibles así como datos de audio.

El dispositivo fuente de A/V 20, de acuerdo con las técnicas de la presente divulgación, es capaz de proporcionar servicios que corresponden a un subconjunto de vistas. En general, una vista se representa por un identificador de vista, también denominado como una "view_id". Los identificadores de vistas en general comprenden elementos de sintaxis que se pueden usar para identificar una vista. Un codificador de MVC proporciona la view_id de una vista cuando la vista está codificada. La view_id se puede usar por el decodificador de MVC para la predicción inter-visión o por otras unidades para otros fines, por ejemplo para la representación.

La predicción entre vistas es una técnica para la codificación de datos de video de MVC de una trama con referencia a una o más tramas en una localización temporal común como la trama codificada de vistas diferentes. La FIG. 7, que se trata con mayor detalle más adelante, proporciona un esquema de codificación de ejemplo para la predicción entre vistas. En general una trama codificada de datos de video de MVC se puede codificar de forma predictiva espacialmente, temporalmente y/o con referencia a tramas de otras vistas en una localización temporal común. Por consiguiente, las vistas de referencia, desde las que se predicen otras vistas, generalmente se codifican antes de las vistas para las cuales actúan como referencia las vistas de referencia, de modo que estas vistas decodificadas se puedan usar para referencia cuando se decodifican las vistas referenciadas. El orden de decodificación no corresponde necesariamente al orden de las view_id. Por lo tanto, el orden de decodificación de las vistas se describe usando índices del orden de visión. Los índices del orden de visión son índices que indican el orden de decodificación de los componentes de visión correspondientes en una unidad de acceso.

Cada uno de los flujos individuales de datos (ya sea audio o video) se denomina como un flujo elemental. Un flujo

elemental es una componente única, codificada digitalmente (posiblemente comprimida) de un programa. Por ejemplo, el video o el audio codificado parte del programa puede ser un flujo elemental. Un flujo elemental se puede convertir en un flujo elemental empaquetado (PES) antes de que multiplexarse en un flujo de programa o un flujo de transporte. Dentro del mismo programa se usa una ID de flujo para distinguir los paquetes PES que pertenecen a un flujo elemental del otro. La unidad básica de datos de un flujo elemental es un paquete flujo elemental empaquetado (PES). De este modo cada vista de los datos de video de MVC corresponde a flujos elementales respectivos. De forma similar, los datos de audio corresponden a uno o más flujos elementales respectivos.

Una secuencia de video codificada de MVC se puede separar en varios sub-flujos de bits, cada uno de los cuales es un flujo elemental. Cada uno de los sub-flujos de bits se puede identificar usando un subconjunto de view_id de MVC. En base al concepto de cada subconjunto de view_id de MVC, se define un sub-flujo de bits de video de MVC. Un sub-flujo de bits de video de MVC contiene las unidades NAL de las vistas listadas en el subconjunto de view_id de MVC. Un flujo de programa generalmente contiene solo las unidades NAL que son las de los flujos elementales. También está diseñado que cualquiera de los dos flujos elementales no puede contener una vista idéntica.

En el ejemplo de la FIG. 1, el multiplexor 30 recibe flujos elementales que comprenden datos de video procedentes del codificador de video 28 y flujos elementales que comprenden datos de audio procedentes del codificador de audio 26. En algunos ejemplos, el codificador de video 28 y el codificador de audio 26 pueden incluir cada uno empaquetadores para formar paquetes PES a partir de datos codificados. En otros ejemplos, el codificador de video 28 y el codificador de audio 26 pueden hacer interfaz cada uno con empaquetadores respectivos para formar los paquetes PES de los datos codificados. En otros ejemplos más, el multiplexor 30 puede incluir empaquetadores para la formación de paquetes PES a partir de los datos codificados de audio y video.

Un "programa" como se usa en la presente divulgación, puede comprender una combinación de datos de audio y datos de video, por ejemplo un flujo elemental de audio y un subconjunto de vistas disponibles entregadas por un servicio de dispositivo fuente de A/V 20. Cada uno de los paquetes PES incluye una stream_id que identifica el flujo elemental al que pertenece el paquete PES. El multiplexor 30 puede ensamblar flujos elementales en flujos de programa constituyentes o flujos de transporte. Un flujo de programa y un flujo de transporte son dos multiplex alternativos que tienen por objeto diferentes aplicaciones.

En general un flujo de programa incluye datos para un programa, mientras que un flujo de transporte puede incluir datos para uno o más programas. El multiplexor 30 puede codificar cualquiera o ambos de un flujo de programa o un flujo de transporte, en base al servicio que se proporcione, un medio dentro del cual se pasará el flujo, el número de programas a enviar, u otras consideraciones. Por ejemplo, cuando los datos de video se van a codificar en un medio de almacenamiento, el multiplexor 30 puede ser que forme más probablemente un flujo de programa, mientras que cuando los datos de video son para descargar de forma continua sobre una red, difundir o enviar como parte de una telefonía de video, el multiplexor 30 se puede usar más probablemente para un flujo de transporte.

El multiplexor 30 puede estar sesgado a favor de usar un flujo de programa para el almacenamiento y representación de un programa único desde un servicio de almacenamiento digital. Un flujo de programa está destinado a usarse en entornos libres de errores o en entornos menos susceptibles a encontrar errores, porque los flujos de programa son más bien susceptibles a los errores. Un flujo de programa simplemente comprende los flujos elementales que pertenecen al mismo y usualmente contiene paquetes de longitudes variables. En un flujo de programa, los paquetes PES que se deducen de la contribución de los flujos elementales se organizan en "paquetes". Un paquete comprende una cabecera de paquete, una cabecera de sistema opcional y cualquier número de paquetes PES tomados a partir de cualquiera de los flujos elementales contribuyentes, en cualquier orden. La cabecera del sistema contiene un resumen de las características del flujo de programa tales como su tasa de datos máxima, el número de flujos elementales de video y audio contribuyentes, además de la información de temporización u otra información. Un decodificador puede usar la información contenida en una cabecera del sistema para determinar si el decodificador es capaz o no de decodificar el flujo de programa.

El multiplexor 30 puede usar un flujo de transporte para la entrega simultánea de una pluralidad de programas sobre canales potencialmente propensos a errores. Un flujo de transporte es un multiplex ideado para aplicaciones multi-programa tales como la difusión, de modo que un único flujo de transporte puede acomodar muchos programas independientes. Un flujo de transporte comprende una sucesión de paquetes de transporte, siendo cada uno de los paquetes de transportes de 188 bytes de longitud. El uso de paquetes cortos de longitud fija significa que el flujo de transporte es menos susceptible a errores que el flujo de programa. Además, a cada paquete de transporte de 188 bytes de longitud se le puede dar una protección frente a errores adicional procesando el paquete a través de un procedimiento de protección de errores normalizado, tal como la codificación de Reed-Solomon. La resistencia frente a errores mejorada del flujo de transporte significa que tiene una mejor posibilidad de sobrevivir a los canales propensos a errores que se encuentran en un entorno de difusión, por ejemplo.

Puede verse que el flujo de transporte es, el mejor de los dos multiplex con su resistencia a errores aumentada y la capacidad de transportar muchos programas simultáneos. Sin embargo, el flujo de transporte es un multiplex más sofisticado que el flujo de programa y en consecuencia es más difícil de crear y de demultiplexar. El primer byte de un paquete de transporte es un byte de sincronización que tiene un valor de 0x47 (en hexadecimal 47, en binario '01000111', en decimal 71). Un flujo de transporte único puede transportar muchos programas diferentes,

comprendiendo cada programa muchos flujos elementales empaquetados. El multiplexor 30 puede usar un campo del Identificador de Paquete (PID) de trece bits para distinguir los paquetes de transporte que contienen los datos de un flujo elemental a partir de los que transportan los datos de otros flujos elementales. Es la responsabilidad del primer multiplexor asegurar que a cada flujo elemental se le adjudica un valor único de PID. El último byte de un paquete de transporte es el campo de la cuenta de continuidad. El multiplexor 30 aumenta el valor del campo de cuenta de continuidad entre paquetes de transporte sucesivos que pertenecen al mismo flujo elemental. Esto posibilita a un decodificador u otra unidad de un dispositivo de destino, tal como un dispositivo de destino de A/V 40, detectar la ganancia o pérdida de un paquete de transporte y esperar ocultar los errores que de otro modo podrían resultar de tal evento.

El multiplexor 30 recibe los paquetes PES para los flujos elementales de un programa desde el codificador de audio 26 y el codificador de video 28 y forma las unidades de la capa de acceso de red (NAL) correspondientes desde los paquetes PES. En el ejemplo de la H.264 / AVC (Codificación de Video Avanzada), los segmentos de video codificados se organizan en unidades NAL, que proporcionan una representación de video "amistosa con la red" que aborda aplicaciones tales como la telefonía de video, el almacenamiento, la difusión, o la descarga continua. Las unidades NAL se pueden clasificar por categorías para las unidades NAL de la Capa de Codificación de Video (VCL) y las unidades NAL no de VCL. Las unidades VCL contienen el dispositivo de compresión central y pueden comprender bloques, macrobloques, y/o niveles de corte. Otras unidades NAL son las unidades NAL no VCL.

El multiplexor 30 puede formar unidades NAL que comprenden una cabecera que identifica un programa al cual pertenece la NAL así como la carga útil, por ejemplo, los datos de audio, los datos de video o datos que describen el flujo de transporte o de programa al que corresponde la unidad NAL. Por ejemplo, en la H.264 / AVC una unidad NAL puede incluir una cabecera de un byte y una carga útil de un tamaño variable. En un ejemplo, una cabecera de unidad NAL comprende un elemento `priority_id`, un elemento `temporal_id`, un elemento `anchor_pic_flag`, un elemento `view_id`, un elemento `non_idr_flag` y un elemento `inter_view_flag`. En la MVC convencional, la unidad NAL definida por la H.264 se mantiene, excepto para las unidades NAL de prefijo y las unidades NAL de corte codificado de MVC, que incluyen una cabecera de la unidad NAL de MVC de 4 bytes y la carga útil de unidades NAL.

El elemento `priority_id` de una cabecera NAL se puede usar para un procedimiento de adaptación del flujo de bits de una trayectoria. El elemento `temporal_id` se puede usar para especificar el nivel temporal de la unidad NAL correspondiente, donde los diferentes niveles temporales corresponden a diferentes tasas de trama.

El elemento `anchor_pic_flag` puede indicar si una imagen es una imagen de ancla o una imagen no de ancla. Las imágenes de ancla y todas las imágenes subsiguientes en el orden de salida (esto es, el orden de representación) se pueden decodificar correctamente sin decodificar las imágenes anteriores en el orden de decodificación (esto es, el orden del flujo de bits), y de este modo se pueden usar como puntos de acceso aleatorio. Las imágenes de ancla y las imágenes no de ancla pueden tener diferentes dependencias estando ambas señalizadas en el conjunto de parámetros de secuencia. Otros indicadores se tratarán y se usarán en las siguientes secciones de este capítulo. Tal imagen de ancla también puede denominarse como un punto de acceso de GOP abierto (Grupo de Imágenes) mientras que un punto de acceso GOP cerrado también se soporta cuando el elemento `non_idr_flag` es igual a cero. El elemento `non_idr_flag` indica si una imagen es un refresco del decodificador instantáneo (IDR) o una imagen de vista IDR (V-IDR). En general, una imagen IDR, y todas las imágenes que la siguen en el orden de salida o el orden del flujo de bits, se puede decodificar correctamente sin decodificar las imágenes anteriores bien en el orden de decodificación o el orden de representación.

El elemento `view_id` comprende información de sintaxis que se puede usar para identificar una vista, que se puede usar para la interactividad de los datos dentro de un decodificador de MVC, por ejemplo, para la predicción entre vistas, y fuera de un decodificador, por ejemplo, para la representación. El elemento `inter_view_flag` puede especificar si se usa la unidad NAL correspondiente por otras vistas para la predicción entre vistas. Para conducir la información de cabecera de la unidad NAL de 4 bytes para una vista base, que puede ser conforme con la AVC, se define una unidad NAL de prefijo en la MVC. En el contexto de la MVC, la unidad de acceso de vista base incluye las unidades NAL de VCL del momento de tiempo actual de la vista así como su unidad NAL de prefijo, que contiene solo la cabecera de unidad NAL. Un decodificador de H.264 / AVC puede ignorar la unidad NAL de prefijo.

Una unidad NAL que incluye los datos de video en su carga útil puede comprender diversos niveles de granularidad de los datos de video. Por ejemplo una unidad NAL puede comprender un bloque de datos de video, un macrobloque, una pluralidad de macrobloques, un corte de datos de video, o una trama entera de datos de video.

En general, una unidad de acceso puede comprender una o más unidades NAL para la representación de una trama de datos de video, así como datos de audio correspondientes a la trama cuando tales datos de audio están disponibles. Una unidad de acceso generalmente incluye todas las unidades NAL para un instante del tiempo de salida, por ejemplo, todos los datos de audio y video para un instante del tiempo. En un ejemplo correspondiente a la H.264 / AVC, una unidad de acceso puede comprender una imagen codificada en un instante del tiempo, que se puede presentar como una imagen codificada primaria. Por consiguiente, una unidad de acceso puede comprender todas las tramas de video de una instancia temporal común, por ejemplo, todas las componentes de vistas correspondientes al tiempo X.

60

La presente divulgación también se refiere a una imagen codificada de una vista particular como un "componente de vista". Esto es, una componente de vista comprende una imagen (o trama) codificada para una vista particular de un momento particular. Por consiguiente, una unidad de acceso puede, en algunos ejemplos, comprender todas las componentes de la vista de una instancia temporal común. El orden de decodificación de las unidades de acceso no necesariamente necesitan ser las mismas que el orden de salida o de representación. Un conjunto de unidades de acceso consecutivas puede formar una secuencia de video codificada, que puede corresponder a un grupo de imágenes (GOP) u otra unidad decodificable independientemente de un flujo de bits de unidades NAL o sub-flujo de bits.

Como con la mayor parte de las normativas de codificación de video, la H.264 / AVC define la sintaxis, semántica y procedimiento de decodificación para flujos de bits libres de errores, cualesquiera de los cuales conforma un cierto perfil o nivel. La H.264 / AVC no especifica el codificador, pero el codificador está encargado de garantizar que los flujos de bits generados cumplan con la normativa para un decodificador. En el contexto de la normativa de codificación de video, un "perfil" corresponde a un subconjunto de algoritmos, características o herramientas y restricciones a aplicar a los mismos. Como se define en la normativa H.264, por ejemplo un "perfil" es un subconjunto de toda la sintaxis del flujo de bits que se especifica por la normativa de H.264. Un "nivel" correspondiente a las limitaciones del consumo de recursos del decodificador tal como, por ejemplo, la memoria del decodificador y los cálculos, que están relacionados con la resolución de las imágenes, la tasa de bits y la tasa de procesamiento de macrobloques (MB).

La normativa H.264, por ejemplo, reconoce que, dentro de los límites impuestos por la sintaxis de un perfil determinado, es aún posible requerir una gran variación en el funcionamiento de los codificadores y decodificadores dependiendo de los valores tomados por los elementos de sintaxis en el flujo de bits tales como el tamaño especificado de las imágenes decodificadas. La normativa H.264 reconoce además que, en muchas aplicaciones, no es práctico ni económico implementar un decodificador capaz de tratar con todos los usos hipotéticos de la sintaxis dentro de un perfil particular. Por consiguiente, la normativa H.264 define un "nivel" como un conjunto especificado de restricciones impuestas sobre valores de los elementos de sintaxis en el flujo de bits. Estas restricciones pueden ser simples límites sobre valores. Como alternativa, estas restricciones pueden tomar la forma de restricciones sobre combinaciones aritméticas de valores (por ejemplo, el ancho de la imagen multiplicado por la altura de la imagen multiplicado por el número de imágenes decodificadas por segundo). La normativa H.264 proporciona además que las implementaciones individuales puedan soportar un nivel diferente para cada uno de los perfiles soportados.

Un decodificador conforme a un perfil soporta normalmente todas las características definidas en el perfil. Por ejemplo, como una característica de codificación, la codificación de una imagen B no se soporta en el perfil de la línea base de la H.264 / AVC y se soporta en otros perfiles de la H.264 / AVC. Un decodificador conforme a un nivel debería ser capaz de decodificar cualquier flujo de bits que no requiera recursos más allá de las limitaciones definidas en el nivel. Las definiciones de perfiles y niveles pueden ser útiles para su interpretación. Por ejemplo, durante la transmisión de video, se pueden negociar y acordar un par de perfiles y definiciones de nivel para una sesión de transmisión global. Más específicamente, en la H.264 / AVC, un nivel puede definir, por ejemplo, las limitaciones sobre el número de macrobloques que se necesita procesar, el tamaño de la memoria intermedia de imágenes decodificadas (DPB), el tamaño de la memoria intermedia de imágenes codificadas (CPB), el intervalo de los vectores de movimiento verticales, el número máximo de vectores de movimiento por dos MB consecutivos, y si un bloque B puede tener porciones de sub-macrobloques menores de 8x8 pixel. De este modo, un decodificador puede determinar si el decodificador es capaz de decodificar adecuadamente el flujo de bits.

Los conjuntos de parámetros generalmente contienen información de cabecera de la capa de secuencia en los conjuntos de parámetros de secuencia (SPS) y la información de cabecera de la capa de imagen que cambia infrecuentemente en los conjuntos de parámetros de imagen (PPS). Con los conjuntos de parámetros, esta información que cambia infrecuentemente no se necesita repetirla para cada secuencia o imagen, por lo tanto la eficiencia de la codificación se puede mejorar. Además, el uso de los conjuntos de parámetros puede posibilitar la transmisión fuera de banda de la información de cabecera, evitando la necesidad de transmisiones redundantes para conseguir la resistencia a errores. En la transmisión fuera de banda, las unidades NAL del conjunto de parámetros se transmiten sobre un canal diferente que las otras unidades NAL.

Las técnicas de la presente divulgación involucran la inclusión de extractores en las pistas de extractores de medios. Los extractores de la presente divulgación pueden referirse a dos o más unidades NAL de otra pista en un fichero común. Esto es, un fichero puede incluir una primera pista que tiene una pluralidad de unidades NAL, y una segunda pista que incluye un extractor que identifica dos o más de la pluralidad de unidades NAL de la primera pista. En general, un extractor puede actuar como un puntero, de modo que cuando se encuentra el extractor por el demultiplexor 38, el demultiplexor 38 puede recuperar las unidades NAL identificadas por el extractor desde la primera pista y enviar esas unidades NAL al decodificador de video 48. Una pista que incluye un extractor se puede denominar como una pista de extractores de medios. Los extractores de la presente divulgación se pueden incluir en ficheros conformes con diversos formatos de ficheros, por ejemplo, al formato de ficheros de medios de base de la ISO, el formato de ficheros de la Codificación de Video Escalable (SVC), el formato de ficheros de la Codificación de Video Avanzada (AVC), el formato de ficheros del Proyecto de Miembros de la Tercera Generación (3GPP), y/o el formato de ficheros de la Codificación de Video Multi-visión (MVC).

En general, las diversas pistas de un fichero de video se pueden usar como pistas de conmutador. Esto es, el multiplexor 30 puede incluir diversas pistas para soportar diversas tasas de trama, capacidades de representación y/o capacidades de decodificación. Por ejemplo, cuando el fichero de video es conforme con el formato de ficheros de MVC, cada pista puede representar un punto de operación de MVC diferente. Por consiguiente, el demultiplexor 38 se puede configurar para seleccionar una de las pistas desde las cuales recuperar unidades NAL y descartar datos de las otras pistas, distintas de las unidades NAL identificadas por los extractores de la pista seleccionada. Esto es, cuando la pista seleccionada incluye un extractor que se refiere a unidades NAL de otra pista, el demultiplexor 38 puede extraer las unidades NAL referenciadas mientras que descarta las unidades NAL no referenciadas de la otra pista. El demultiplexor 38 puede enviar las unidades NAL extraídas al decodificador de video 48.

Usando extractores en una pista de extractor de medios, las técnicas de la presente divulgación se pueden usar para conseguir la escalabilidad temporal entre las diversas pistas de un fichero de video. En MPEG-1 y MPEG-2, por ejemplo, las imágenes codificadas como B proporcionan una escalabilidad temporal natural. Una primera pista de un fichero de video conforme con MPEG-1 o MPEG-2 puede incluir un conjunto completo de imágenes codificadas como I, imágenes codificadas como P, e imágenes codificadas como B. Una segunda pista del fichero de video puede incluir uno o más extractores que se refieren solo a las imágenes codificadas como I y las imágenes codificadas como P de la primera pista, omitiendo las referencias a las imágenes codificadas como B. Abandonando las imágenes codificadas como B, el fichero de video puede conseguir una confirmación de una representación de video de resolución media. MPEG-1 y MPEG-2 también proporcionan una capa base y el concepto de capa de mejora para codificar dos capas temporales, en el que las imágenes de capa de mejora pueden elegir, para cada dirección de predicción, una imagen bien desde la capa base o la capa de mejora como una referencia.

Otro ejemplo de la H.264 / AVC usa imágenes codificadas como B jerárquicas para soportar la escalabilidad temporal. La primera imagen de una secuencia de video en H.264 / AVC se puede denominar como una imagen de Refresco del Decodificador Instantáneo (IDR), también conocida como una imagen clave. Las imágenes clave usualmente están codificadas a intervalos regulares o irregulares que están o intra codificadas o inter codificadas usando una imagen clave anterior como referencia para la predicción compensada de movimientos. Un Grupo de Imágenes (GOP) generalmente incluye una imagen clave y todas las imágenes que están temporalmente localizadas entre la imagen clave y una imagen clave anterior. Un GOP se puede dividir en dos partes, una es la imagen clave, y la otra incluye imágenes no de clave. Las imágenes no de clave se predicen jerárquicamente por 2 imágenes de referencia, que son las imágenes más próximas del nivel temporal inferior del pasado y del futuro. Se puede asignar un valor de identificador temporal a cada imagen para indicar una posición jerárquica de la imagen. De este modo, las imágenes con valores del identificador temporal hasta N pueden formar un segmento de video con el doble de la tasa de trama de la de un segmento de video formado por imágenes con valores del identificador temporal de hasta N - 1. Por consiguiente, las técnicas de la presente divulgación también se pueden usar para conseguir una escalabilidad temporal en la H.264 / AVC teniendo una primera pista que incluye todas las unidades NAL con valores del identificador temporal hasta N, y una segunda pista que incluye uno o más extractores que se refieren a las unidades NAL de la primera pista con valores del identificador temporal hasta N - 1.

Como se ha observado anteriormente, las técnicas de la presente divulgación se pueden aplicar a ficheros de video conformes con cualquier formato de ficheros de medios de base de la ISO, un formato de ficheros de la Codificación de Video Escalable (SVC), un formato de ficheros de la Codificación de Video Avanzada (AVC), un formato de ficheros del Proyecto de Miembros de la Tercera Generación (3GPP), y/o un formato de ficheros de la Codificación de Video Multi-visión (MVC). EL Formato de Ficheros de Medios de Base de la ISO está diseñado para contener información de medios temporizados para una presentación en un formato extensible, flexible que facilita el intercambio, la gestión, la edición y la presentación de medios. El formato de Ficheros de Medios de Base de la ISO (ISO / IEC 14496 - 12:2004) se especifica en MPEG-4 Parte 12, que define una estructura general para ficheros de medios basados en el tiempo. Se usa como la base para otros formatos de ficheros en la familia tales como el formato de ficheros de AVC (ISO / IEC 14496 - 15) soporte definido para la comprensión de video de la H.264 / MPEG-4 AVC, el formato de ficheros del 3GPP, el formato de ficheros de SVC, y el formato de ficheros MVC. El formato de ficheros 3GPP y el formato de ficheros MVC son extensiones del formato de ficheros de AVC. El formato de ficheros de medios de base de la ISO contiene la temporización, la estructura, y la información de medios para las secuencias temporizadas de los datos de medios, tales como las presentaciones audio-visuales. La estructura de ficheros está orientada a objetos. Un fichero se puede descomponer en objetos básicos muy simples y la estructura de los objetos está implicada a partir de su tipo.

Los ficheros conformes con el formato de ficheros de medios de base de la ISO están formados como una serie de objetos, llamados "cajas". Los datos en el formato de ficheros de medios de base de la ISO están contenidos en cajas y no hay otros datos dentro del fichero. Esto incluye cualquier firma inicial requerida por el formato de fichero específico. La "caja" es un bloque de construcción orientado a objetos definido por un identificador de tipo único y la longitud. Usualmente, una presentación está contenida en un fichero, y la presentación de medios está auto-contenida. El contenedor de película (caja de película) contiene los metadatos de los medios y las tramas de video y audio están contenidas en el contenedor de datos de medios y podría estar en otros ficheros.

Una presentación (secuencia de movimiento) puede estar contenida en varios ficheros. Toda la información de temporización y estructura (posición y tamaño) está generalmente en el fichero de medios de base de la ISO y los

ficheros auxiliares pueden usar esencialmente cualquier formato. Esta presentación puede ser "local" para el sistema que contiene la presentación, o puede ser a través de la red u otro mecanismo de entrega del flujo.

Los ficheros pueden tener una estructura lógica, una estructura de tiempo, y una estructura física, y estas estructuras no se requiere que estén acopladas. La estructura lógica del fichero puede ser de una película que a su vez contiene un conjunto de pistas paralelas en el tiempo. La estructura de tiempo del fichero puede ser que las pistas contengan secuencias de muestras en el tiempo, y esas secuencias se mapeen dentro de la línea de tiempos de la película global por lista de edición opcional. La estructura física del fichero puede separar los datos necesarios para la lógica, el tiempo y la descomposición estructural, a partir de las propias muestras de datos de medios. Esta información estructural puede estar concentrada en una caja de película, posiblemente extendida en el tiempo por cajas de fragmentos de película. La caja de película puede documentar las relaciones lógicas y de temporización de las muestras y también puede contener punteros a donde están localizadas. Estos punteros pueden estar dentro del mismo fichero o en otro, por ejemplo, referenciado por una URL.

Cada uno de los flujos de medios puede estar contenido en una pista especializada para ese tipo de medios (audio, video, etc.), y puede estar además parametrizada por una entrada de muestras. La entrada de muestras puede contener el 'nombre' del tipo de medio exacto (el tipo de decodificador necesario para decodificar el flujo) y cualquier parametrización del decodificador necesaria. El nombre también puede tomar la forma de un código de cuatro caracteres, por ejemplo "moov" o "trak". Hay formatos de entrada de muestras definidos no solo para medios de MPEG-4, sino también para los tipos de medios usados por otras organizaciones que usan esta familia de formatos de ficheros.

El soporte para los metadatos generalmente toma dos formas. En primer lugar, los metadatos temporizados se pueden almacenar en una pista apropiada sincronizada como se desee con los datos de medios que está describiendo. En segundo lugar, puede haber un soporte general para meta-datos no temporizados fijados a la película o a una pista individual. El soporte estructural es general y permite, como en los datos de medios, el almacenamiento de los recursos de metadatos en cualquier parte en el fichero o en otro fichero. Además, estos recursos se pueden nombrar y se pueden proteger.

En el formato de ficheros de medios de base de la ISO, un agrupamiento de muestras es una asignación de cada una de las muestras en una pista para ser un miembro de un grupo de muestras. Las muestras en un grupo de muestras no se requiere que sean contiguas. Por ejemplo, cuando se presentan en el formato de ficheros de la H.264 / AVC, las muestras de video en un nivel temporal se pueden muestrear dentro de un grupo de muestras. Los grupos de muestras se pueden representar por dos estructuras de datos: una caja de SampleToGroup (sbdp) y una caja SampleGroupDescription. La caja SampleToGroup representa la asignación de las muestras a los grupos de muestras. Puede haber una instancia de la caja SampleGroupDescription para cada entrada de grupo de muestras, para describir las propiedades del grupo correspondiente.

Una pista de metadatos opcional se puede usar para etiquetar cada pista con la "característica de interés" que tenga, por lo que su valor puede ser diferente de otros miembros del grupo (por ejemplo, su tasa de bits, el tamaño de la pantalla o el lenguaje). Algunos ejemplos dentro de una pista pueden tener características especiales o se pueden identificar individualmente. Un ejemplo de la característica es el punto de sincronización (a menudo una trama I de video). Estos puntos se pueden identificar por una tabla especial en cada pista. De forma más general, la naturaleza de las dependencias entre las muestras de pistas también se puede documentar usando metadatos. Los metadatos se pueden estructurar como una secuencia de muestras de formato de ficheros, justo como una pista de video. Tal pista se puede denominar como una pista de metadatos. Cada muestra de metadatos se puede estructurar como una declaración de metadatos. Hay varias clases de declaraciones, correspondientes a las diversas cuestiones que se podrían preguntar acerca de la muestra del formato de ficheros correspondiente o sus muestras constituyentes.

Cuando se entrega un medio sobre un protocolo de descarga continua, el medio puede necesitar transformarse del modo en el que se representa en el fichero. Un ejemplo de esto es cuando los medios se transmiten sobre el Protocolo de Tiempo Real (RTP). En el fichero, por ejemplo, cada trama de video está almacenada de forma contigua como una muestra de formato de fichero. En el RTP, se deben cumplir las normas de empaquetamiento específicas para el códec usado, para colocar estas tramas en paquetes de RTP. Un servidor de descarga continua se puede configurar para calcular tal empaquetamiento en el tiempo de ejecución. Sin embargo, se soporta para la asistencia de los servidores de descarga continua. Pistas especiales llamadas pistas de indicadores se pueden colocar en los ficheros.

Las pistas de indicadores contienen instrucciones generales para los servidores de descarga continua sobre cómo formar los flujos de paquetes desde las pistas de medios para un protocolo específico. Debido a que la forma de estas instrucciones es independiente de los medios, puede que los servidores no necesiten revisar cuando se introducen nuevos códec. Además el software de la codificación y la edición se pueden ignorar por los servidores de descarga continua. Una vez que termina la edición sobre un fichero, se puede usar un elemento de software llamado un indicador para añadir pistas de indicaciones al fichero, antes de colocarlo en un servidor de descarga continua. Como ejemplo, hay un formato de la pista de indicaciones definido para flujos de RTP en la especificación del formato de ficheros de MP4.

- El 3GP (formato de ficheros del 3GPP) es un formato de contenedor multimedia definido por el Proyecto de Miembros de la Tercera Generación (3GPP) para los servicios multimedia del UMTS de 3G. Típicamente se usa sobre teléfonos móviles 3G y otros dispositivos capaces de 3G, pero también se puede reproducir sobre algunos teléfonos y dispositivos 2G y 4G. El formato de ficheros de 3GPP se basa en el formato de ficheros de medios de base de la ISO. El último 3GP se especifica en el documento TS26.244 del 3GPP, "Transparent end-to-end packet switched streaming service (PSS); 3GPP file format (3GP)". El formato de ficheros del 3GPP almacena flujos de video como MPEG-4 Parte 2 o H.263 o (MPEG-4 Parte 10) (AVC / H.264). El 3GPP permite el uso de códec AMR y H.263 en el formato de ficheros de medios de base de la ISO (MPEG-4 Parte 12), porque el 3GPP especifica el uso de la Entrada de Muestras y campos de plantillas en el formato de ficheros de medios de base de la ISO así como la definición de nuevas cajas a las que se refieren los códec. Para el almacenamiento de la información específica de medios de MPEG-4 en ficheros 3GP, la especificación de 3GP se refiere a MP4 y el formato de ficheros de AVC, que también se basan en el formato de ficheros de medios de base de la ISO. Las especificaciones del formato de ficheros de MP4 y AVC describen el uso del contenido de MPEG-4 en el formato de ficheros de medios de base de la ISO.
- El formato de ficheros de SVC, como una extensión del formato de ficheros de AVC, tiene nuevas estructuras de extractor y nivel. Los extractores son punteros que proporcionan información acerca de la posición y el tamaño de los datos de codificación de video en la muestra con tiempo de decodificación igual en otra pista. Esto permite la construcción de una jerarquía de pistas directamente en el dominio de la codificación. Una pista de extractor en la SVC está enlazada con una o más pistas base, desde las que extrae los datos en el tiempo de ejecución. Un extractor es un puntero no comparable con una cabecera de unidad NAL con extensiones de SVC. Si la pista usada para la extracción contiene datos de la codificación de video a una tasa de trama diferente, entonces el extractor también contiene un desplazamiento del tiempo de decodificación para asegurar la sincronía entre pistas. En el tiempo de ejecución, el extractor tiene que reemplazarse por los datos a los que apunta, antes de que el flujo se pase al decodificador de video.
- Debido a que las pistas del extractor en SVC están estructuradas como las pistas de codificación de video, pueden representar el subconjunto que necesitan en diferentes modos. Una pista de extractores de SVC contiene solo instrucciones sobre cómo extraer los datos desde otra pista. En el formato de ficheros de SVC, también hay agregadores, que pueden agregar la unidad NAL dentro de una muestra junto como una unidad NAL, incluyendo la agregación de unidades NAL en una capa dentro de un agregador. El extractor en SVC está diseñado para extraer un cierto intervalo de bytes desde una muestra o un agregador, o solo una unidad NAL entera, pero no múltiples unidades NAL, especialmente las que no son consecutivas en una muestra. En el formato de ficheros de SVC, podría haber muchos puntos de operación de video. Los niveles se diseñan para agrupar las muestras en una o más pistas para un punto de operación.
- El formato de ficheros de MVC también soporta una pista de extractores, que extrae las unidades NAL de diferentes vistas para formar un punto de operación, que es un subconjunto de vistas en una cierta tasa de trama. El diseño de la pista de extractores de MVC es similar al extractor en el formato de ficheros de SVC. Sin embargo, no se soporta el uso de las pistas de extractor de MVC para formar un grupo alternativo. Para soportar la selección de pistas se propone la siguiente propuesta del MPEG de P. Frojdh, A. Norkin y C. Priddle, del MPEG "File format sub-track selection and switching", de la ISO / IEC JTC1/SC29/WG11 MPEG M16665, Londres, Reino Unido. Esta propuesta intenta posibilitar el concepto de grupo de alternativo / conmutación en un nivel de sub-pista.
- Un mapa de grupo de muestras es una extensión al grupo de muestras. En el Mapa de grupo de muestras, cada entrada de grupo (de muestras) tiene su descripción de "groupID" que realmente es un mapa para una view_id, posiblemente después de agregar unidades NAL en una vista dentro de una unidad NAL. En otras palabras, cada una de las entradas de grupo de muestras tiene sus vistas de contenidos listadas en el valor ScalableNALUMapEntry. El grouping_type de este grupo de muestras es "scrm".
- La descarga progresiva es un término usado para describir la transferencia de ficheros de medios digitales desde un servidor a un cliente, usualmente usando el protocolo HTTP. Cuando se inicia desde un ordenador el consumidor puede comenzar la reproducción del medio antes de que se complete la descarga. La diferencia clave entre la descarga continua de un medio y la descarga progresiva es cómo se reciben los datos del medio digital y se almacenan por el dispositivo de usuario final que está accediendo al medio digital. Un reproductor de medios que es capaz de reproducir por descarga progresiva se basa en que los metadatos localizados en la cabecera del fichero estén intactos y una memoria local intermedia del fichero de medios digital según se descarga desde el servidor web. En el punto en el que una cantidad especificada de datos se hace disponible para el dispositivo de reproducción local, el medio comenzará a reproducirse. Esta cantidad especificada de memoria intermedia se incorpora dentro del fichero por el productor del contenido en la configuración del codificador y se refuerza por las configuraciones de memoria intermedia adicionales impuestas por el reproductor de medios.
- En el 3GPP, el transporte de HTTP/TCP/IP se soporta para la descarga y la descarga progresiva de los ficheros 3GP. Además el uso de HTTP para la descarga continua de video tiene algunas ventajas, y los servicios de descarga continua de video basados en el HTTP se están haciendo populares. Algunas ventajas de la descarga continua del HTTP incluyen que existen componentes de Internet y protocolos que se pueden usar, de modo que no se necesitan nuevos esfuerzos para desarrollar nuevas técnicas para el transporte de datos de video sobre una red.

Otros protocolos de transporte, por ejemplo el formato de la carga útil de RTP, requieren dispositivos de red intermedios, por ejemplo, cajas intermedias, que se encargan del formato de medios y el contexto de señalización. También, la descarga continua de HTTP puede ser conducida por el cliente, lo que evita muchos problemas de control. Por ejemplo, para explotar todas las características para obtener un funcionamiento óptimo, el servidor puede mantener el seguimiento del tamaño y el contenido de los paquetes que aún no se han confirmado, el servidor también puede analizar la estructura de ficheros y reconstruir el estado de la memoria intermedia del cliente para tomar decisiones de conmutación / estrechamiento óptimo de RD. Además, las restricciones sobre las variaciones del flujo de bits se pueden satisfacer para seguir cumpliendo con los perfiles negociados. El HTTP no necesariamente requiere nuevas implementaciones hardware o software en el servidor Web que tiene implementado el HTTP 1.1. La descarga continua de HTTP también proporciona un TCP amistoso y cortafuegos transversal. Las técnicas de la presente divulgación pueden mejorar la descarga continua de HTTP de los datos de video para superar problemas relacionados con el ancho de banda, por ejemplo, proporcionando la adaptación de tasa de bits.

Las normativas de compresión de video tales como ITU-T H.261, H.262, H.263, MPEG-1, MPEG-2 y H.264 / MPEG-4 parte 10 hacen uso de la predicción temporal compensada del movimiento para reducir la redundancia temporal. El codificador usa una predicción compensada de movimiento a partir de algunas imágenes codificadas anteriormente (también denominadas en este documento como tramas) para predecir las imágenes codificadas actuales de acuerdo con vectores de movimiento. Hay tres tipos importantes de imágenes en la codificación típica de video. Hay imágenes intra codificadas (las "imágenes I" o "tramas I"), imágenes Predichas (las "imágenes P" o "tramas P") e imágenes de predicción Bi-direccional ("imágenes B" o "tramas B"). Los bloques de imágenes P pueden ser intra codificados o predichos con referencia a otra imagen. En una imagen B, los bloques se pueden predecir a partir de una o dos imágenes de referencia, o se pueden intra codificar. Estas imágenes de referencia se podrían localizar antes o después de la imagen actual en el orden temporal.

De acuerdo con la normativa de codificación H.264, como ejemplo, las imágenes B usan dos listas de imágenes de referencias codificadas anteriormente, la lista 0 y la lista 1. Esas dos listas pueden contener cada una las imágenes codificadas del pasado y/o del futuro en orden temporal. Los bloques en una imagen B se pueden predecir en uno de varios modos: predicción compensada por movimiento desde una imagen de referencia de la lista 0, la predicción compensada por movimiento desde una imagen de referencia de la lista 1, o una predicción compensada por movimiento desde la combinación de las imágenes de referencia de ambas, la lista 0 y la lista 1. Para obtener la combinación de ambas imágenes de referencia de la lista 0 y la lista 1, se obtienen dos áreas de referencia compensadas en movimiento a partir de las imágenes de referencia de la lista 0 y la lista 1 respectivamente. Su combinación se usará para predecir el bloque actual.

Los bloques más pequeños de video pueden proporcionar una mejor resolución y se pueden usar para las localizaciones de una trama de video que indica altos niveles de detalle. En general, los macrobloques y diversas particiones, a veces denominadas como sub-bloques, se pueden considerar bloques de video. Además, un corte se puede considerar que es una pluralidad de bloques de video, tales como macrobloques y/o sub-bloques. Cada corte puede ser una unidad decodificable independientemente de una trama de video. Como alternativa, las propias tramas pueden ser unidades decodificables, u otras porciones de una trama se pueden definir como unidades decodificables. El término "unidad codificada" o "unidad de codificación" se puede referir a cualquier unidad decodificable independientemente de una trama de video tal como una trama entera, un corte de una trama, un grupo de imágenes (GOP) también denominado como una secuencia, u otra unidad decodificable independientemente definida de acuerdo con las técnicas de codificación aplicables.

El término macrobloque se refiere a una estructura de datos para la codificación de una imagen y/o datos de video de acuerdo con un matriz de pixel bidimensional que comprende 16 x 16 pixel. Cada pixel comprende una componente de crominancia y una componente de luminancia. Por consiguiente, el macrobloque puede definir cuatro bloques de luminancia, comprendiendo cada uno una matriz bidimensional de 8 x 8 pixel, dos bloques de crominancia, comprendiendo cada uno una matriz bidimensional de 16 x 16 pixel, y una cabecera que comprende información de sintaxis, tal como un patrón de bloque codificado (CBP), un modo de codificación (por ejemplo, los modos de codificación intra (I) o inter (P o B)), un tamaño de partición para particiones de un bloque intra codificado (por ejemplo, 16 x 16, 16 x 8, 8 x 16, 8 x 8, 8 x 4, 4 x 8, o 4 x 4), o uno o más vectores de movimiento para un macrobloque inter codificado.

El codificador de video 28, el decodificador de video 48, el codificador de audio 28, el decodificador de audio 46, el multiplexor 30 y el demultiplexor 38 se puede implementar cada uno como cualquiera de una diversidad de circuiterías de codificador o decodificador, como sea aplicable, tal como uno o más microprocesadores, procesadores de señal digital (DSP), circuitos integrados específicos de la aplicación (ASIC), redes de puertas programables en campo (FPGA), circuitería de lógica discreta, software, hardware, firmware o cualquier combinación de los mismos. Cada uno del codificador de video 28 y el decodificador de video 48 se pueden incluir en uno o más codificadores o decodificadores, cualquiera de los cuales puede estar integrado como parte de un codificador / decodificador (CODEC) de video combinado. Del mismo modo, cada uno del codificador de audio 26 y el decodificador de audio 46 pueden estar incluidos en uno o más codificadores o decodificadores, cualquiera de los cuales puede estar integrado como parte de un codificador / decodificador de audio (CODE) combinado. Un aparato que incluye el codificador de video 28, el decodificador de video 48, el codificador de audio 26, el decodificador de audio 46, el multiplexor 30 y/o el demultiplexor 38 puede comprender un circuito integrado, un microprocesador, y/o

un dispositivo de comunicaciones inalámbricas, tal como un teléfono celular.

De acuerdo con las técnicas de la presente divulgación, el multiplexor 30 puede ensamblar unidades NAL dentro de pistas de un fichero de video conforme con el formato de ficheros de medios de base de la ISO o una derivación del mismo (por ejemplo, SVC, AVC, MVC, o 3GPP), e incluir una pista de extractores de medios que identifica una o más unidades NAL potencialmente no consecutivas de otra pista, y pasar el fichero de video a la interfaz de salida 32. La interfaz de salida 32 puede comprender, por ejemplo, un transmisor, un transceptor, un dispositivo para escribir datos a un medio legible por ordenador tal como, por ejemplo, una unidad óptica, una unidad de medio magnética (por ejemplo una unidad de disco flexible), un puerto del bus serie universal (USB), una interfaz de red u otra interfaz de salida. La interfaz de salida 32 saca la unidad NAL o la unidad de acceso a un medio legible por ordenador 34, por ejemplo, un medio transitorio, tal como una señal de transmisión o una onda portadora, o un medio de almacenamiento de ordenador tal como un medio magnético, un medio óptico, una memoria o una unidad flash.

La interfaz de entrada 36 recupera los datos desde el medio legible por ordenador 34. La interfaz de entrada 36 puede comprender, por ejemplo, una unidad óptica, una unidad de medio magnético, 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 NAL o la unidad de acceso al demultiplexor 38. El demultiplexor 38 puede demultiplexar un flujo de transporte o flujo de programa en los flujos PES constituyentes, desempaquetar los flujos PES para recuperar los datos codificados, y enviar los datos codificados a cualquier decodificador de audio 46 o decodificador de video 48, dependiendo de si los datos codificados son parte de un flujo de audio o de video, por ejemplo, como se indica por las cabeceras de paquetes PES del flujo. El demultiplexor 38 puede seleccionar inicialmente una de las pistas incluidas en un fichero de video recibido, y a continuación pasar solo los datos de la pista seleccionada y los datos de las otras pistas referenciadas por los extractores de la pista seleccionada al decodificador de video 48, descartando los datos de otras pistas no referenciadas por un extractor de la pista seleccionada. El decodificador de audio 46 decodifica los datos de audio codificados y envía los datos de audio decodificados a la salida de audio 42, mientras que el decodificador de video 48 decodifica los datos de video codificados y envía los datos de video decodificados, que pueden incluir una pluralidad de vistas de un flujo, a la salida de video 44. La salida de video 44 puede comprender una pantalla que usa una pluralidad de vistas de una escena, por ejemplo, una pantalla de estereoscopio o auto-estereoscopio que presenta cada vista de una escena simultáneamente.

La FIG. 2 es un diagrama de bloques que ilustra una disposición de ejemplo de componentes del multiplexor 30 (FIG. 1). En el ejemplo de la FIG. 2, un multiplexor 30 incluye una unidad de gestión de flujo 60, una interfaz de entrada de video 80, una interfaz de entrada de audio 82, una interfaz de salida del flujo multiplexado 84, y tablas de información específica del programa 88. La unidad de gestión del flujo 60 incluye el constructor de unidades NAL 62, la unidad de búsqueda del identificador de flujo (flujo ID) 66, la unidad de generación de pistas 64, y la unidad de generación de extractores 68.

En el ejemplo de la FIG. 2, la interfaz de entrada de video 80 y la interfaz de entrada de audio 82 incluyen empaquetadores respectivos para formar las unidades PES a partir de los datos de video codificados y los datos de audio codificados. En otros ejemplos, los empaquetadores de video y/o audio pueden estar presentes externos al multiplexor 30. Con respecto al ejemplo de la FIG. 2, la interfaz de entrada de video 80 puede formar paquetes PES a partir de datos de video codificados recibidos desde el codificador de video 28 y la interfaz de entrada de audio 82 puede formar paquetes PES a partir de los datos codificados recibidos desde el codificador de audio 26.

Después de que el constructor de unidades NAL 62 construye las unidades NAL, el constructor de unidades NAL 62 envía las unidades NAL a la unidad de generación de pistas 64. La unidad de generación de pistas 64 recibe las unidades NAL y ensambla un fichero de video que incluye las unidades NAL en una o más pistas del fichero de video. La unidad de generación de pistas 64 puede ejecutar además la unidad de generación de extractores 68 para generar extractores para una o más pistas de extractores construidas por la unidad de generación de pistas 64. Cuando se determina que una o más unidades NAL pertenecen a múltiples pistas, en lugar de duplicar la unidad NAL entre las pistas, la unidad de generación de extractores 68 puede construir un extractor para una pista que se refiere a la unidad NAL. De este modo, el multiplexor 30 puede evitar la duplicación de datos entre pistas lo que puede reducir el consumo de ancho de banda cuando se transmite el fichero de video.

A continuación se tratan diversos ejemplos de estructuras de datos y componentes para un extractor. En general, un extractor puede incluir un valor de identificador de pista que se refiere a una pista en la que se incluye una unidad NAL referenciada y uno o más identificadores de unidades NAL que identifican las unidades NAL referenciadas por el extractor. En algunos ejemplos, los identificadores de las unidades NAL pueden referirse a un intervalo de bits o bytes en la pista referenciada por el valor del identificador de pista correspondiente a las unidades NAL identificadas. En algunos ejemplos, los identificadores de unidades NAL pueden referirse individualmente a cada unidad NAL identificada por el extractor, por ejemplo, para identificar unidades NAL no consecutivas. En algunos ejemplos, los identificadores de unidades NAL pueden referirse a las unidades NAL en base a un desplazamiento desde la localización temporal o espacial del extractor en la pista de extractores de medios.

La unidad de generación de pistas 64 puede, en algunos ejemplos, incluir unidades NAL adicionales en una pista de extractores de medios. Esto es, una pista de extractores de medios puede incluir tanto unidades NAL como

extractores. Por consiguiente, en algunos ejemplos, la unidad de generación de pistas 64 puede construir un fichero de video que tiene una primera pista que incluye solo unidades NAL y una segunda pista que incluye uno o más extractores que se refieren a todas o un subconjunto de las unidades NAL de la primera pista. Además, en algunos ejemplos, la unidad de generación de pistas 64 puede incluir unidades NAL adicionales en la segunda pista que no están incluidas en la primera pista. De la misma forma, las técnicas de la presente divulgación se pueden extender a una pluralidad de pistas. Por ejemplo, la unidad de generación de pistas 64 puede construir una tercera pista que puede referirse a unidades NAL de la primera pista y/o unidades NAL de la segunda pista y puede incluir adicionalmente unidades NAL no incluidas en la primera o segunda pistas.

La FIG. 3 es un diagrama de bloques que ilustra un fichero de ejemplo 100 que incluye una primera pista que tiene un conjunto de muestras de video y una segunda pista que tiene extractores que se refieren a un subconjunto de muestras de video de la primera pista. En el ejemplo de la FIG. 3, el fichero 100 incluye una caja MOOV 102 y una caja de datos de medios (MDAT) 110. La caja MOOV 102 corresponde a una caja de película, que el formato de ficheros de medios de base de la ISO define como una caja contenedor cuyas sub-cajas definen los metadatos para una presentación. La caja MDAT 104 corresponde a una caja de datos de medios, que el formato de ficheros de medios de base de la ISO define como una caja que puede mantener los datos reales para una presentación.

En el ejemplo de la FIG. 3, la caja MOOV 102 incluye la pista de subconjuntos completos 104 y la pista de extractores de medios 106. El formato de ficheros de medios de base de la ISO define una "pista" como una secuencia temporizada de muestras relacionadas en un fichero de medios de base de la ISO. El formato de ficheros de medios de base de la ISO observa además para los datos de medios, que una pista corresponde a una secuencia de imágenes o audio muestreado.

La caja MDAT 110, en el ejemplo de la FIG. 3, incluye la muestra codificada como I 112, las muestras codificadas como P 114, las muestras codificadas como B 116, y las muestras codificadas como B 118. Las muestras codificadas como B 116 y las muestras codificadas como B 118 se considera que son diferentes niveles de codificación jerárquica. En el ejemplo de la FIG. 3, las muestras codificadas como B 116 se pueden usar como referencia para las muestras codificadas como B 118, y por lo tanto las muestras codificadas como B 118 pueden estar en un nivel de codificación jerárquico que es inferior del nivel de codificación jerárquico de las muestras codificadas como B 116. El orden de representación de las muestras puede diferir del orden jerárquico (también denominado como un orden de decodificación), y el orden en el que las muestras se incluyen en la caja MDAT 110. Por ejemplo, las muestras codificadas como I 112 pueden tener un valor del orden de representación de 0 y un valor del orden de decodificación de 0, las muestras codificadas como P 114 pueden tener un valor del orden de representación de 2 y un valor del orden de decodificación de 1, las muestras codificadas como B 116 pueden tener un valor del orden de representación de 1 y un valor del orden de decodificación de 2, y las muestras codificadas como B 118 pueden tener un valor del orden de representación de 4 y un valor del orden de decodificación de 3. La pista 1 puede incluir muestras adicionales, por ejemplo, una muestra con un valor del orden de representación de 3 y un valor del orden de decodificación de 4.

Cada una de la muestras codificadas como I 112, las muestras codificadas como P 114, las muestras codificadas como B 116 y las muestras codificadas como B 118 pueden corresponder a diversas unidades NAL o unidades de acceso. El formato de ficheros de medios de base de la ISO define una "muestra" como todos los datos asociados con un único sello temporal, por ejemplo, una trama individual de video, una serie de tramas de video en el orden de decodificación, o una sección comprimida de audio en el orden de decodificación. La pista de subconjuntos completos 104, en el ejemplo de la FIG. 3 incluye metadatos que se refieren a la muestra codificada como I 112, las muestras codificadas como P 114, las muestras codificadas como B 116 y las muestras codificadas como B 118.

La caja MDAT 110 incluye adicionalmente el extractor 120, el extractor 122, y el extractor 124. De ese modo los extractores 120 - 124 están incluidos en una caja de datos de película, que en general incluiría muestras de datos. En el ejemplo de la FIG. 3, el extractor 120 se refiere a la muestra codificada como I 112, el extractor 122 se refiere a las muestras codificadas como P 114, y el extractor 124 se refiere a las muestras codificadas como B 118. Puede haber dos o más unidades NAL correspondientes a la muestra codificada como I 112, las muestras codificadas como P 114 y/o las muestras codificadas como B 118, y las unidades NAL pueden ser no consecutivas. De acuerdo con las técnicas de la presente divulgación, los extractores 120 - 124 pueden sin embargo identificar cada una de las unidades NAL de la muestra correspondiente, incluso aunque puede haber dos o más unidades NAL no consecutivas en la muestra correspondiente. La pista de extractores de medios 106, en el ejemplo de la FIG. 3, incluye metadatos que se refieren al extractor 120, al extractor 122 y al extractor 124.

Cada uno de los extractores 120 - 124 también pueden incluir valores del orden de representación y valores del orden de decodificación. Por ejemplo el extractor 120 puede tener un valor del orden de representación de 0 un valor del orden de decodificación de 0, el extractor 122 puede tener un valor del orden de representación de 1 y un valor del orden de decodificación de 1, el extractor 124 puede tener un valor del orden de representación de 2 y un valor del orden de decodificación de 2. En algunos ejemplos, los valores de representación y/o de decodificación pueden saltar ciertos valores, por ejemplo, para compaginar los valores de la muestra identificada.

La pista de subconjuntos completos 104 y la pista de extractores de medios 106 pueden formar un grupo alternativo, de modo que el demultiplexor 38 (FIG. 1) puede seleccionar bien la pista de subconjuntos completos 104 o la pista

de extractores de medios 106 a decodificar por el decodificador de video 48. Con respecto al ejemplo de MVC, la pista de subconjuntos completos 104 puede corresponder a un primer punto de operación y la pista de extractores de medios 106 puede corresponder a un segundo punto de operación. Con respecto al ejemplo de 3GPP, la pista de subconjuntos completos 104 y la pista de extractores de medios 106 pueden formar un grupo conmutador. De este modo, la pista de subconjuntos completos 104 y la pista de extractores de medios 106 se pueden usar para adaptar la disponibilidad de ancho de banda y la capacidad del decodificador, por ejemplo en aplicaciones de descarga continua de HTTP.

Cuando se selecciona la pista de subconjuntos completos 104, el demultiplexor 38 puede enviar las muestras correspondientes a la pista de subconjuntos completos 104 (por ejemplo, la muestra codificada como I 112, las muestras codificadas como P 114, las muestras codificadas como B 116 y las muestras codificadas como B 118) al decodificador de video 48. Cuando se selecciona la pista de extractores de medios 106, el demultiplexor 38 puede enviar las muestras correspondientes a la pista de extractores de medios 106, incluyendo muestras identificadas por los extractores de medios correspondientes a la pista de extractores de medios 106, al decodificador de video 48. De este modo cuando se selecciona la pista de extractores de medios 106, el demultiplexor 38 puede enviar la muestra codificada como I 112, las muestras codificadas como P 114, las muestras codificadas como B 118 al decodificador de video 48, que puede recuperar el demultiplexor 38 a partir de la pista de subconjuntos completos 104 desreferenciando el extractor 120, el extractor 122 y el extractor 124.

La FIG. 4 es un diagrama de bloques que ilustra otro fichero de ejemplo 140 que incluye dos pistas distintas de extractores 146, 148. Aunque en la FIG. 4 se ilustran dos pistas de extractores, en general un fichero puede incluir cualquier número de pistas de extractores. En el ejemplo de la FIG. 4 el fichero 140 incluye la caja MOOV 142 y la caja MDAT 150. La caja MOOV 142 incluye la pista de subconjuntos completos 144 y las pistas de extractores de medios 146, 148. La caja MDAT 150 incluye muestras de datos y extractores para las diversas pistas, por ejemplo la muestra codificada como I 152, las muestras codificadas como P 154, las muestras codificadas como B 156 y, las muestras codificadas como B 158 y los extractores 160 - 168.

En el ejemplo de la FIG. 4, los extractores 160 - 164 corresponden a la pista de extractores de medios 146, mientras que los extractores 166 - 168 corresponden a la pista de extractores de medios 148. En este ejemplo, el extractor 160 de la pista de extractores de medios 146 identifica las muestras codificadas como I 152, el extractor 162 identifica las muestras codificadas como P 154, y el extractor 164 identifica las muestras codificadas como B 156. En este ejemplo, el extractor 166 identifica las muestras codificadas como I 152, mientras que el extractor 162 identifica las muestras codificadas como P 154. El ejemplo de la FIG. 4 muestra un ejemplo en el que dos o más extractores de diversas pistas de extractores de medios se refieren a la misma muestra de una pista de subconjuntos completos.

Las pistas de extractores de medios se pueden usar para representar subconjuntos temporales de un flujo de video, que es decodificable y alternar / conmutar una pista de la pista que contiene el flujo de bits original de resolución temporal completa, por ejemplo la pista de subconjuntos completos 144. La pista de subconjuntos completos 144 puede, por ejemplo representar un flujo de video de 30 tramas por segundo (FPS). En algunos ejemplos, que no incluyen un cierto nivel jerárquico de imágenes B en un su-flujo de bits, la tasa de trama del sub-flujo de bits se puede reducir a la mitad o reducir en alguna otra fracción. Por ejemplo, la pista de extractores de medios 146, que no incluye las muestras codificadas como B 158, puede tener una tasa de trama que se reduce a la mitad, con relación a la pista de subconjuntos completos 144. Por ejemplo, la pista de extractores de medios 146 puede tener una tasa de trama de 15 FPS. Del mismo modo, la pista de extractores de medios 148 puede tener una tasa de trama que se puede reducir a la mitad en comparación con la pista de extractores de medios 146 omitiendo tanto las muestras codificadas como B 156 como las muestras codificadas como B 158, y de este modo tener una tasa de 7,5 FPS.

La FIG. 5 es un diagrama de bloques que ilustra otro fichero de ejemplo 180 que incluye una pista de subconjuntos 188 y dos pistas de extractores de medios 184, 186. La caja MOOV 182 del fichero 180 incluye la pista de subconjuntos 188 y las pistas de extractores de medios 184, 186, mientras que la caja MDAT 190 incluye la muestra codificada como I 192, las muestras codificadas como P 194, las muestras codificadas como B 202, las muestras codificadas como B 208 y los extractores 198, 200, 204, 206 y 210.

Como se ha tratado anteriormente, una pista de extractores de medios puede incluir extractores que se refieren a muestras de otra pista. Además, una pista de extractores de medios puede incluir además muestras de video adicionales que no se incluyen en otra pista. En el ejemplo de la FIG. 5, la pista de subconjuntos 188 incluye la muestra codificada como I 192, y las muestras codificadas como P 194. La pista de extractores de medios 186 incluye los extractores 198, 200 y adicionalmente incluye las muestras codificadas como B 202. De forma similar, la pista de extractores de medios 184 incluye los extractores 204, 206, 210 y adicionalmente las muestras codificadas como B 208.

En el ejemplo de la FIG. 5, la pista de extractores de medios 186 incluye muestras codificadas de datos de video (muestras codificadas como B 202), y la pista de extractores de medios 184 incluye el extractor 210 que se refiere a las muestras de la pista de extractores de medios 186 que incluye las muestras codificadas. Esto es, en el ejemplo de la FIG. 5, el extractor 210 se refiere a muestras codificadas como B 202. Por consiguiente, la pista de extractores de medios 184 puede representar una resolución temporal completa de un flujo de bits, mientras que la pista de

extractores de medios 186 y la pista de subconjuntos 188 pueden representar subconjuntos del flujo de bits de resolución temporal completa. Esto es, la pista de extractores de medios 186 y la pista de subconjuntos 188 pueden tener resoluciones temporales inferiores (por ejemplo, tasas de trama más bajas) que la resolución temporal completa representada por la pista de extractores de medios 184.

- 5 De acuerdo con las técnicas de la presente divulgación, el formato de ficheros H.264 / AVC se puede modificar para incluir pistas de extractores que se pueden extraer como cualquier subconjunto temporal de conformación de la pista que contiene el flujo de bits original de resolución temporal completa. Para el H.264/AVC que soporta la codificación jerárquica de imágenes B (o P), asumiendo que hay N niveles temporales, cada uno de los flujos de bits que incluyen muestras desde el nivel 0 al k ($k < N$) se pueden extraer definiendo la pista de extractores correspondiente.
- 10 De este modo, para el mismo video, podría haber N pistas (incluyendo N - 1 pistas de extractores) que forman un grupo de alternar/conmutar. Los extractores pueden estar asociados con un nivel jerárquico temporal correspondiente al nivel jerárquico temporal de las muestras identificadas por los extractores. Por ejemplo, el valor de identificador temporal que especifica el nivel temporal de las muestras también se puede señalar en el extractor.

15 Las FIG. 6A - 6C son diagramas de bloques que ilustran ejemplos de una caja MDAT 220 de un fichero que incluye ejemplos de extractores de medios para diversas tasas pistas de extractores de medios. Cada una de las FIG. 6A - 6C representa la muestra de ancla 222 que incluyen la muestra de vista 0 224A, la muestra de vista 2 226A, la muestra de vista 1 228A, la muestra de vista 4 230A y la muestra de vista 3 232A, y la muestra no de ancla 223 que incluye la muestra de vista 0 224B, la muestra de vista 2 226B, la muestra de vista 1 228B, la muestra de vista 4 230B, y la muestra de vista 3 232B. La elipse además de la muestra no de ancha 223 indica que las muestras adicionales se pueden incluir en la caja MDAT 220. Cada una de las muestras de ancla y no de ancla puede formar colectivamente una primera pista el fichero. En un ejemplo, la pista de extractores de medios para cada conjunto de extractores del fichero representada en las FIG 6A - 6C puede corresponder a un punto de operación separado de un fichero de video conforme al formato de ficheros de MVC, de acuerdo con las técnicas de la divulgación. De este modo, las técnicas de la presente divulgación se pueden usar para generar una o más pistas de extractores de medios correspondientes a un punto de operación de un fichero de video conforme al formato de ficheros de MVC.

20

25

Las FIG. 6A - 6C representan los extractores 240, 244, 250 de diversas pistas de extractores de medios, donde los extractores 240, 244, 250 se incluirían cada uno en la caja de MDAT 220, pero se ilustran en figuras separadas con fines de claridad. Esto es, cuando está completamente ensamblada, la caja MDAT 220 puede incluir cada uno de los conjuntos de extractores 240, 244 y 250.

- 30 Las FIG. 6A - 6C proporcionan un ejemplo de un fichero que incluye una pista que contiene extractores de medios así como muestras de video real. Diversas muestras pueden estar contenidas separadamente en diferentes pistas de acuerdo con los diferentes niveles temporales. Para cada nivel temporal, una pista particular puede contener todas las muestras de video así como extractores para las pistas con niveles temporales más bajos. Las muestras de video (unidades NAL) se pueden separar en diferentes pistas, mientras que la pista con una tasa de trama mayor puede tener extractores apuntando a las otras pistas. De este modo, es posible tener fragmentos de películas que contienen muestras de solo un nivel temporal y un fragmento de película puede contener posiblemente extractores apuntando a otros fragmentos. En este caso, los fragmentos de películas de diferentes pistas, pero durante el mismo periodo de tiempo, se podrían intercalar aumentando el orden del nivel temporal.
- 35

40 La FIG. 6A proporciona un ejemplo de extractores 240 que incluye los extractores 242A - 242N correspondientes a una pista de extractores de medios. En este ejemplo, el extractor 242A se refiere a ambas la muestra 224A de la vista 0 de la muestra de ancla 222. El extractor 242N se refiere a la muestra 224B de la vista 0 de la muestra no de ancla 223. En general, un extractor del conjunto de extractores 240 se refiere a la muestra correspondiente de la vista 0, en el ejemplo de la FIG. 6A. Cada uno de los extractores 242A - 242N corresponde a una pista de extractores de medios común, que puede pertenecer a un grupo de conmutar y/o un grupo de alternar. La pista de extractores de medios puede corresponder además a un punto de operación individual, por ejemplo, un punto de operación incluyendo la vista 0.

45

En algunos ejemplos, para el video estéreo codificado usando MVC, puede haber tres puntos de operación, incluyendo un punto de operación que soporta emitir dos vistas, y un segundo punto de operación que soporta emitir solo una vista (por ejemplo solo la vista 0 o la vista 1). El tercer punto de operación podría ser un punto de operación que saca la vista 1. Dependiendo de la relación de predicción, el tercer punto de operación puede incluir solo las unidades NAL de VCL y las unidades NAL asociadas no de VCL en la vista 1, todas las unidades NAL de la vista 0 y la vista 1, o las unidades NAL en la vista 1 así como las unidades NAL de ancla (esto es, las unidades NAL de los componentes de vistas de ancla). En tal caso de estéreo, los ejemplos de las técnicas desveladas pueden proporcionar que los otros dos puntos de operación se pueden representar por dos pistas de extractores. Estas dos pistas de extractores pueden formar un grupo de conmutar y, junto con la pista de video original, estas tres pistas pueden formar un grupo de alternar.

50

55

Esta divulgación proporciona técnicas para modificar el formato de ficheros de MVC para incluir pistas de extractores de medios MVC. En general, las pistas de video MVC, incluyendo las pistas de extractores de medios MVC, con el mismo número de vistas para la salida se pueden caracterizar como grupos de conmutar. Todos los puntos de operación representados por las pistas de un fichero pueden pertenecer a un grupo de alternar de una presentación

60

de video de MVC. Las vistas de cada una de las muestras de ancla 222 y muestras no de ancla 223 pueden formar una pista de subconjuntos completos, por ejemplo un punto de operación incluyendo todas las vistas disponibles.

5 Un extractor se puede referir a una parte continua de una muestra, por ejemplo como se muestra con respecto a los extractores 246A - 246N en la FIG. 6B. En el ejemplo de la FIG. 6B, el extractor 246A se refiere a la muestra 224A de la vista 0 y la muestra 226A de la vista 2. La estructura de datos que representa el extractor 246A puede especificar un intervalo de bytes para las vistas identificadas, una vista de comienzo y una vista de fin, una vista de comienzo y varias vistas posteriores, u otra representación de una serie continua de vistas identificadas por el extractor. El conjunto de extractores 244 puede corresponder a otra pista del extractor de medios que puede corresponder a su vez a un punto de operación separado de MVC.

10 Dos extractores se pueden referir también a dos partes (por ejemplo, dos vistas no continuas) de una muestra, por ejemplo como se muestra con respecto a los extractores 254A, 256A en la FIG. 6C. Por ejemplo, la muestra de extractor 252A incluye el extractor 254A que se refiere a la muestra 224A de la vista 0 y la muestra 226A de la vista 2, así como el extractor 254B que se refiere la muestra 230A de la vista 4. De este modo la muestra representada por la muestra del extractor 252A puede corresponder a una muestra de extractor que se refiere a muestras de vistas no consecutivas. De forma similar, la muestra de extractor 252N, en el ejemplo de la FIG. 6C, incluye el extractor 256A que se refiere a la muestra 224B de la vista 0 y la muestra 226B de la vista 2, así como el extractor 256B que se refiere a la muestra 230B de la vista 4.

Los extractores también se pueden definir con respecto a muestras de ancla y muestras no de ancla, donde los extractores definidos con respecto a muestras de ancla se pueden referir a diferentes vistas que los extractores definidos con respecto a muestras no de ancla.

Las pistas de extractores de medios de MVC mencionadas anteriormente en el formato de ficheros de medios de base de la ISO o el formato de ficheros de MVC pueden ser instancias de pistas de metadatos que se pueden implementar con una funcionalidad de extracción similar y se pueden usar para representar pistas de alternar y/o conmutar de una pista de video normal.

25 En ejemplos que usan el formato de ficheros de MVC, un flujo de bits completo puede estar contenido en una pista y todos los otros puntos de operación posibles se pueden representar por las pistas de extractor, cada uno de los cuales puede señalar, por ejemplo, varias vistas para emitir, los valores de identificador de vistas de las vistas para emitir, el ancho de banda requerido para la transmisión, y la tasa de trama.

30 La FIG. 7 es un diagrama conceptual que ilustra un patrón de predicción de MVC de ejemplo. En el ejemplo de la FIG. 7, se ilustran ocho vistas (que tienen ID de vistas de "S0" a "S7") y doce localizaciones temporales (de "T0 a "T11") para cada vista. Esto es, cada fila en la FIG. 7 corresponde a una vista, mientras que cada columna indica una localización temporal.

Aunque la MVC tiene una vista llamada base que es decodificable por los decodificadores de H.264 / AVC y se podría soportar un par de vistas estéreo también por la MVC, la ventaja de la MVC es que podría soportar un ejemplo que usa más de dos vistas como una entrada de video de 3D y decodificar este video 3D representado por las vistas múltiples. Un procesador de un cliente que tiene un decodificador de MVC puede esperar contenidos de video de 3D sin vistas múltiples. Un componente de vista de ancla y un componente de vista no de ancla en una vista pueden tener dependencias de vista diferentes. Por ejemplo, los componentes de vista de ancla en la vista S2 dependen de componentes de vista en la vista S0. Sin embargo, los componentes de vista no de ancla en la vista S2 no dependen de los componentes de vista en otras vistas.

45 Las tramas en la FIG. 7 se indican para cada fila y cada columna en la FIG. 7 usando un bloque sombreado que incluye una letra, que designa si la trama correspondiente está intra codificada (esto es, una trama I), o inter codificada en una dirección (esto es, como una trama P) o en múltiples direcciones (esto es, como una trama B). En general, las predicciones se indican por flechas, donde la trama apuntada usa el objeto desde el que se apunta para referencia de predicción. Por ejemplo, la trama P de la vista S2 en la localización temporal T0 se predice desde la trama I de la vista S0 en la localización temporal T0.

50 Como con la codificación de video de una única vista, las tramas de una secuencia de video de codificación de video multi-visión se pueden codificar predictivamente con respecto a las tramas en diferentes localizaciones temporales. Por ejemplo, la trama b de la vista S0 en la localización temporal T1 tiene una flecha apuntándola desde la trama I de la vista S0 en la localización temporal T0, indicando que la trama b está predicha desde la trama I. Adicionalmente, sin embargo, en el contexto de la codificación de video multi-visión, las tramas se pueden predecir inter vistas. Estos es, un componente de vista puede usar los componentes de vista de otras vistas para referencia. En la MVC, por ejemplo, la predicción inter vistas se realiza como si la componente de vista en otra vista fuese una referencia inter predicción. Las referencias potenciales inter vistas se señalizan en la extensión de MVC del Conjunto de Parámetros de Secuencia (SPS) y se puede modificar por procedimiento de construcción de listas de imágenes de referencia, que posibilita el ordenamiento flexible de otras referencias inter-predicción o inter-vistas. La Tabla 1 a continuación proporciona una definición de ejemplo para un conjunto de parámetros de secuencia de extensión de MVC.

TABLA 1

seq parameter_set_mvc_extensión () {	C	Descriptor
num_views_minus1	0	ue(v)
for (i = 0; i<= num_views_minus1; i++)		
view_id [i]	0	ue(v)
for (i = 1; i<= num_views_minus1; i++) {		
num_anchor_refs_10 [i]	0	ue(v)
for (j = 0; j < num_anchor_refs_10[i]; j++)		
anchor_refs_10 [i][j]	0	ue(v)
num_anchor_refs_11 [i]	0	ue(v)
for (j = 0; j < num_anchor_refs_11 [i]; j++)		
anchor_ref_11 [i][j]	0	ue(v)
}		
for (i = 1; i<= num_views_minus1; i++) {		
num_non_anchor_refs_10 [i]	0	ue(v)
for (j = 0; j < num_non_anchor_refs_10[i]; j++)		
non_anchor_ref_10 [i][j]	0	ue(v)
num_non_anchor_refs_11[i]	0	ue(v)
for (j = 0; j < num_non_anchor_refs_11 [i]; j++)		
non_anchor_ref_11 [i][j]	0	ue(v)
}		
num_level_values_signalled_minus 1	0	ue(v)
for (i = 0; j<= num_level_values_signalled_minus1; i++) {		
level_idc [i]	0	u(8)
num_applicable_ops_minus1 [i]	0	ue(v)
for (j = 0; j<= num_applicable_ops_minus1[i]; j++){		
applicable_op_temporal id[i][j]	0	u(3)
applicable_op_num_target_views_minus1 [i][j]	0	ue(v)
for(k = 0; k <= applicable_op_num_target_views_minus1 [i][j]; k++)		
applicable_op_target_view_id[i][j][k]	0	ue(v)
applicable_op_num_views_minus1 [i][j]	0	ue(v)
}		
}		
}		

La FIG. 7 proporciona diversos ejemplos de la predicción inter vistas. Las tramas de la vista S1, en el ejemplo de la FIG. 7 se ilustran de modo que se predicen a partir de tramas en localizaciones temporales diferentes de la vista S1, así como inter vistas predichas desde tramas de las tramas de vistas S0 y S2 en las mismas localizaciones temporales. Por ejemplo, la trama b de la vista S1 en la localización temporal T1 se predice a partir de cada una de las tramas B de la vista S1 en las localizaciones temporales T0 y T2, así como las tramas b de las vistas S0 y S2 en la localización temporal T1.

En el ejemplo de la FIG. 7, la letra mayúscula "B" y la letra minúscula "b" están destinadas a indicar diferentes relaciones jerárquicas entre tramas, en lugar de diferentes metodologías de codificación. En general, las tramas con la letra mayúscula "B" son relativamente más altas en la jerarquía de predicción que las tramas de la letra minúscula "b". Esto es, en el ejemplo de la FIG. 7, las tramas "b" se codifican con referencia a las tramas "B". Se pueden añadir niveles jerárquicos adicionales que tienen tramas adicionales codificadas de forma bidireccional que se pueden referir a las tramas "b" de la FIG. 7. La FIG. 7 también ilustra variaciones en la jerarquía de predicción usando diferentes niveles de sombreado, donde una mayor cantidad de sombreado de las tramas (esto es, relativamente más oscuro) indica que son más altas en la jerarquía de predicción que las tramas que tienen un menor sombreado (esto es, relativamente más claro). Por ejemplo, todas las tramas I en la FIG. 7 se ilustran con un sombreado total, mientras que las tramas P tienen un sombreado algo más claro, y las tramas B (y las tramas b con letra minúscula) tiene diversos niveles de sombreado relativos entre sí, pero siempre más claros que el sombreado de las tramas P y las tramas I.

En general, la jerarquía de predicción está relacionada con los índices del orden de visión, en que las tramas relativamente más altas en la jerarquía de predicción se deberían decodificar antes que las tramas de decodificación que son relativamente más bajas en la jerarquía, de modo que las tramas relativamente más altas en la jerarquía se pueden usar como tramas de referencia durante la decodificación de las tramas relativamente más bajas en la jerarquía. Un índice de orden de visión es un índice que indica el orden de decodificación de los componentes de vistas en una unidad de acceso. Los índices del orden de visión están implicados en la expansión de MVC de SPS, como se especifica en el Anexo H de H.264 / AVC (MVC modificada). En el SPS, para cada índice i se señala la view_id correspondiente. La decodificación de las componentes de vistas seguirá el orden ascendente del índice de orden de visión. Si están presentes todas las vistas, entonces los índices del orden de visión están en un orden consecutivo desde 0 al num_views_minus_1.

De este modo, las tramas usadas como tramas de referencia se pueden decodificar antes de la decodificación de las tramas que se codifican con referencia a las tramas de referencia. Un índice del orden de visión es un índice que indica el orden de decodificación de los componentes de vistas en una unidad de acceso. Para cada índice del orden de visión i, se señala la view_id correspondiente. La decodificación de las componentes de vistas sigue el orden ascendente de los índices del orden de visión. Si todas las vistas están presentes, entonces el conjunto de índices del orden de visión comprende un conjunto ordenado consecutivamente desde cero a uno menos que el número total de vistas.

Para ciertas tramas a iguales niveles de jerarquía, el orden de decodificación puede que no tenga importancia relativa para los otros. Por ejemplo, la trama I de la vista S0 en la localización temporal T0 se usa como una trama de referencia para la trama P de la vista S2 en la localización temporal T0, que se usa a su vez como trama de referencia para la trama P de la vista S4 en la localización temporal T0. Por consiguiente la trama I de la vista S0 en la localización temporal T0 se debería decodificar antes de la trama P de la vista S2 en la localización temporal T0, que se debería decodificar antes que la trama P de la vista S4 en la localización temporal T0. Sin embargo, entre las vistas S1 y S3, el orden de decodificación no importa, porque las vistas S1 y S3 no descansan en cada una de las otras para la predicción, sino más bien se predicen solo a partir de vistas que son más altas en la jerarquía de predicción. Además, la vista S1 se puede decodificar antes que la vista S4, siempre que la vista S1 se decodifique después que las vistas S0 y S2.

De este modo, se puede usar un ordenamiento jerárquico para describir las vistas S0 hasta S7. Supongamos que la notación SA > SB significa que la vista SA se debería decodificarse antes que la vista SB, Usando esta notación S0 > S2 > S4 > S6 > S7, en el ejemplo de la FIG. 7, También con respecto al ejemplo de la FIG. 7, S0 > S1, S2 > S1, S2 > S3, S4 > S3, S4 > S5 y S6 > S5. Cualquier orden de decodificación para las vistas que no viola estos requisitos es posible. Por consiguiente, son posibles muchos órdenes diferentes de decodificación, con solo ciertas limitaciones. Dos ejemplos de órdenes de decodificación están presentes a continuación, aunque se debería entender que son posibles muchos otros órdenes de decodificación. En un ejemplo, ilustrado en la Tabla 2 a continuación, las vistas se decodifican siempre que sea posible.

TABLA 2

ID de Visión	S0	S1	S2	S3	S4	S5	S6	S7
Índice del orden de visión	0	2	1	4	3	6	5	7

El ejemplo de la Tabla 2 reconoce que la vista S1 se puede decodificar inmediatamente después de que las vistas S0 y S2 se hayan decodificado, la vista S3 se puede decodificar inmediatamente después de que se hayan decodificado las vistas S2 y S4, y la vista S5 se puede decodificar después de que se hayan decodificado las vistas S4 y S6.

- 5 La tabla 3 dada a continuación proporciona otro ejemplo de orden de decodificación en el que el orden de decodificación es tal que cualquier vista que se use como referencia para otra vista se decodifica antes que las vistas que no se usan como referencia para cualquier otra vista.

TABLA 3

ID de Visión	S0	S1	S2	S3	S4	S5	S6	S7
Índice del orden de visión	0	5	1	6	2	7	3	4

- 10 El ejemplo de la Tabla 3 reconoce que las tramas de las vistas S1, S3, S5 y S7 no actúan como tramas de referencia para tramas de cualesquiera otras vistas, y por lo tanto las vistas S1, S3, S5 y S7 se decodifican después de las tramas de las vistas que se usan como tramas de referencia, esto es, las vistas S0, S2, S4 y S6, en el ejemplo de la FIG. 7. Con relación a cada una de las otras, las vistas S1, S3, S5 y S7 se pueden decodificar en cualquier orden. Por consiguiente, en el ejemplo de la Tabla 3, la vista S7 se decodifica antes de cada una de las vistas S1, S3 y S5.

- 15 Para ser claros, puede haber una relación jerárquica entre las tramas de cada vista así como las localizaciones temporales de las tramas de cada vista. Con respecto al ejemplo de la FIG. 7, las tramas en la localización temporal T0 son o bien intra predichas o predichas entre vistas a partir de tramas de otras vistas en la localización temporal T0. De forma similar, las tramas en la localización temporal T8 son o bien intra predichas o predichas entre vistas a partir de tramas de otras vistas en la localización temporal T8. Por consiguiente, con respecto a una jerarquía temporal, las localizaciones temporales T0 y T8 están en la parte superior de la jerarquía temporal.

- Las tramas en la localización temporal T4, en el ejemplo de la FIG. 7, están más bajas en la jerarquía temporal que las tramas de las localizaciones temporales T0 y T8 debido a que las tramas de la localización temporal T4 están codificadas como B con referencia a tramas de localizaciones temporales T0 y T8. Las tramas en las localizaciones temporales T2 y T6 son más bajas en la jerarquía temporal que las tramas en la localización temporal T4. Finalmente, las tramas en las localizaciones temporales T1, T3, T5 y T7 son más bajas en la jerarquía temporal que las tramas de las localizaciones temporales T2 y T6.

- En MVC, un subconjunto de un flujo de bits global se puede extraer para formar un sub-flujo de bits que está aún conforme con la MVC. Hay muchos posibles sub-flujos de bits que pueden requerir aplicaciones específicas, en base a, por ejemplo a un servicio proporcionado por un servidor, la capacidad, el soporte y las capacidades de los decodificadores de uno o más clientes y/o la preferencia de uno o más clientes. Por ejemplo, un cliente podría requerir solo tres vistas, y podría haber dos escenarios. En un ejemplo un cliente puede requerir una experiencia de visión suave y podría preferir vistas con los valores de view_id S0, S1 y S2, mientras que otro cliente puede requerir escalabilidad de las vistas y preferir las vistas con valores de view_id S0, S2, y S4. Si originalmente las view_id se ordenan con respecto al ejemplo de la Tabla 9, los valores de índices del orden de visión son {0, 1, 2} y {0, 1, 4} en esos dos ejemplos, respectivamente. Obsérvese que ambos de estos sub-flujos de bits se pueden decodificar como flujos de bits de MVC independientes y se pueden soportar simultáneamente.

- Puede haber muchos sub-flujos de bits de MVC que sean decodificables por decodificadores de MVC. En teoría, cualquier combinación de vistas que satisfaga las siguientes dos propiedades se pueden decodificar por un decodificador de MVC que cumpla con un cierto perfil o nivel: (1) las componentes de vistas en cada unidad de acceso están ordenadas en un orden creciente del índice del orden de visión, y (2) para cada vista en la combinación, también se incluyen en la combinación sus vistas dependientes.

- Con respecto a las técnicas de la presente divulgación, se pueden representar diversos sub-flujos de bits de MVC usando pistas de extractores de medios y/o pistas de muestras de video puras. Cada una de estas pistas puede corresponder a un punto de operación de MVC.

- 45 Las FIG. 8 - 21 son diagramas de bloques que ilustran diversos ejemplos de estructuras de datos para extractores de medios y otras estructuras de soporte de datos que se pueden usar de acuerdo con las técnicas de la presente divulgación. Los diversos extractores de medios de las FIG. 8 - 22 incluyen diversas características, como se tratará con detalle más adelante. En general, cualquiera de los extractores de medios de las FIG. 8 - 21 se pueden incluir en una pista de extractores de medios de un fichero conforme con el formato de ficheros de medios de base de la ISO o una extensión del formato de ficheros de medios de base de la ISO para identificar muestras codificadas del fichero. En general, se puede usar un extractor de medios para extraer una o más muestras completas desde una pista referenciada. Las FIG. 8 - 12 son ejemplos de extractores de medios que son capaces de identificar una caja de muestras de video de otra pista. Como se muestra en la FIG. 13, otro modo de implementar el extractor es posibilitar el agrupamiento de muestras, de muestras desde otra pista. Para dar un soporte más específico a la escalabilidad temporal se puede señalar un identificador temporal, como se muestra en la FIG. 14. Las FIG. 8 a 14

no muestran extractores que extraigan unidades NAL no consecutivas, y como tal no caen dentro del ámbito de la presente aplicación. La FIG. 16, 22 son ejemplos de extractores de medios para MVC y son capaces de extraer una o más unidades NAL potencialmente no consecutivas desde cada caja de muestras de video (unidad de acceso). Diversos ejemplos de extractores se basan en desplazamientos y longitudes de bytes en un fichero o unidad de acceso, mientras que otros ejemplos se pueden basar puramente en los índices de todas las unidades NAL, de este modo no es necesaria la señalización de los intervalos de bytes necesarios. El mecanismo de señalización de los extractores con los índices de todas las unidades NAL también se pueden extender a un formato de ficheros de SVC.

Los ejemplos de las FIG. 8 - 21 también se pueden aplicar al formato de ficheros del 3GPP directamente como extensiones del formato de ficheros de 3GPP. Los elementos y conceptos de una o más de las FIG. 8 - 21 también se pueden combinar con elementos de otras de las FIG. 8 - 22 para formar otros extractores. Aunque ciertas de las FIG. 8 - 21 se describen con respecto a un formato de fichero particular, en general, los ejemplos de las FIG. 8 - 21 se pueden usar con respecto a cualquier formato de fichero con características similares, por ejemplo el formato de ficheros de medios de base de la ISO o extensiones de los formatos de ficheros de medios de base de la ISO. Para facilitar el uso de los extractores propuestos en el 3GPP, la caja de selección de pistas de 3GPP se puede extender para incluir más características para cada una de las pistas alternas (extraídas), tal como el identificador temporal, el número de vistas a representar y el número de vistas a decodificar, como se muestra en el ejemplo de la FIG. 21.

La FIG. 8 es un diagrama de bloques que ilustra un extractor de medios de ejemplo 300 que ilustra el formato de un extractor de medios. En el ejemplo de la FIG. 8, el extractor de medios 300 incluye el índice de referencia de pista 302 y un valor de desplazamiento de muestra 304. El extractor de medios 300 puede corresponder a la definición de una estructura de datos que se puede solicitar dentro de una pista de extractores de medios, de acuerdo con las técnicas de la presente divulgación. El multiplexor 30 se puede configurar para incluir un extractor conforme con el ejemplo de extractor de medios 300 en una pista de extractores de medios de un fichero de video para identificar una unidad NAL de una pista diferente del fichero de video. El demultiplexor 38 se puede configurar para recuperar la unidad NAL identificada usando el extractor conforme al extractor de medios 300.

El índice de referencia de pista 302 puede corresponder a un identificador de una pista en la que está presente una unidad NAL identificada. A cada pista de un fichero de video se le puede asignar un índice único, para diferenciar las pistas del fichero de video. El índice de referencia de pistas 302 puede especificar el índice de la referencia de pista a usar para encontrar la pista desde la que extraer los datos. La muestra en esa pista desde el que se extraen los datos puede estar temporalmente alineada (en la línea de tiempos de decodificación de medios, usando la tabla de tiempo para muestras, ajustada por un desplazamiento especificado por el valor de desplazamiento de muestra 304) conteniendo la muestra el extractor. En algunos ejemplos, la primera pista de un fichero de video tiene un valor de índice de '1', y por lo tanto el multiplexor 30 puede asignar un valor de '1' al valor del índice de referencia de pista 302 para referirse a la primera pista de un fichero de video. Un valor de '0' para el valor del índice de referencia de pista se puede reservar para un uso futuro.

El valor del desplazamiento de muestras 304 define un valor de desplazamiento desde la localización temporal del extractor de medios 300 en la pista de extractores de medios para una unidad NAL identificada de la pista referenciada por el índice de referencia de pista 302. Esto es, el valor de desplazamiento muestra 304 da el índice relativo de la muestra en la pista enlazada que se usa como la fuente de información. Un valor de cero para el valor del desplazamiento de muestras 304 se refiere a la muestra con el mismo tiempo de decodificación, o el más próximo precedente para la muestra que contiene al extractor, la muestra 1 (uno) en la siguiente muestra, la muestra -1 (menos 1) es la muestra anterior, y así sucesivamente. Cuando un extractor de medios, conforme con el extractor de medios se usa en H.263 o en MPEG-4 parte 2, por ejemplo, el extractor de medios se puede usar para extraer un subconjunto temporal de la pista de video referenciada por el índice de referencia de pista 302.

El seudocódigo siguiente proporciona una definición de ejemplo de una clase de extractor de medios similar al extractor de medios 300.

```
class aligned (8) MediaExtractor () {
    unsigned int (8) track_ref_index;
    signed int (8) sample_offset;
}
```

El multiplexor 30 y el demultiplexor 38 pueden solicitar un objeto de datos del extractor de medios usando el extractor de medios definido en el seudocódigo de ejemplo anterior. Por consiguiente, el demultiplexor 38, por ejemplo, se puede referir al extractor de medios solicitado en caso de recuperación de datos desde una pista seleccionada para recuperar los datos identificados desde otra pista referenciada por el extractor de medios solicitado.

En el seudocódigo de ejemplo, la clase MediaExtractor () está alineado a bytes. Esto es, cuando un extractor se solicita desde la clase del MediaExtractor (), el extractor estará alineado sobre una frontera de ocho bytes. La variable "trak_ref_index" corresponde al valor del índice de referencia de pista 302, y en este seudocódigo de ejemplo, corresponde a un valor entero de ocho bytes sin signo. La variable "sample_offset" corresponde a un valor

de desplazamiento de muestras 304 y, en este ejemplo, a un valor entero de ocho bytes con signo.

La FIG. 9 es un diagrama de bloques que ilustra otro ejemplo de extractor de medios 310. El extractor de medios 310 incluye el índice de referencia de pista 314 y el valor del desplazamiento de muestra 316, y además, incluye una cabecera de muestras 312. El índice de referencia de pista 314 y el valor del desplazamiento de muestra 316 pueden generalmente incluir datos similares al índice de referencia de pista 302 y el valor del desplazamiento de muestras 304 (FIG. 8).

La cabecera de muestra 312 en un ejemplo correspondiente para la H.264 / AVC, se puede construir de acuerdo con las cabeceras de unidades NAL de una muestra de video referenciada por el extractor de medios 310. La cabecera de muestra 312 puede contener un byte de datos con tres elementos de sintaxis: `forbiden_zero_bit_nal_ref_idc` (que puede comprender 3 bits), `nal_unit_type` (que puede comprender 5 bits). El valor de "nal_unit_type" puede ser 29 (o cualquier otro número reservado) y los otros elementos de sintaxis pueden ser los mismos que los elementos de sintaxis en la muestra de video identificada. Para ejemplos conformes con MPEG-4 parte 2 visual, la cabecera de muestra 312 puede comprender un código de cuatro bytes, que pueden incluir el prefijo de código de comienzo "0x00 00 01" y el código de comienzo de '0xC5' o cualquier otro número reservado), donde "0x" indica que el valor siguiente a "0x" es un valor hexadecimal. Para H.263, la cabecera de muestra 312 también puede incluir un código de comienzo alineado a byte que es diferente del código de comienzo de las muestras de video normales. La cabecera de muestras 312 se puede usar por el demultiplexor 38 para el fin de sincronización, de modo que el extractor se puede considerar como una muestra de video normal.

El seudocódigo siguiente proporciona una definición de ejemplo de una clase de extractor de medios similar al extractor de medios 310:

```
class aligned (8) MediaExtractor () {
    SampleHeader ()
    unsigned int (8) track_ref_index;
    signed int (8) sample_offset;
}
```

El multiplexor 30 y el demultiplexor 38 pueden solicitar un objeto de datos de extractor de medios usando el extractor de medios definido en el seudocódigo de ejemplo anterior. Por consiguiente, el demultiplexor 38, por ejemplo, puede referirse al extractor de medios solicitado cuando recupera datos desde una pista seleccionada para recuperar datos identificados desde otra pista referenciada por el extractor de medios solicitado.

La FIG. 10, es un diagrama de bloques que ilustra un extractor de medios de ejemplo 320 que identifica las unidades NAL señalizando el intervalo de bytes de las unidades NAL identificadas, dentro del extractor. El extractor de medios 320 incluye la cabecera de la muestra 322, que puede ser similar a la cabecera de muestra 312, y el índice de referencia de pista 324, que puede ser similar al índice de referencia de pista 302. Más que un valor de desplazamiento de muestra, sin embargo, el ejemplo del extractor de medios 320 incluye el valor del desplazamiento de datos 326 y el valor de la longitud de datos 328.

El valor de desplazamiento de datos 326 puede describir el punto de comienzo de los datos identificados por el extractor de medios 320. Esto es, el valor de desplazamiento de datos 326 puede comprender un valor representativo del desplazamiento para el primer byte dentro de la pista identificada por el valor del índice de la pista 324 a copiar. El valor de la longitud de datos 328 puede describir el número de bytes a copiar, y por consiguiente, puede ser equivalente a la longitud de la muestra referenciada (o muestras cuando se referencian múltiples unidades NAL).

El siguiente seudocódigo proporciona una definición de ejemplo de una clase de extractor de medios similar al extractor de medios 320:

```
class aligned (8) MediaExtractor () {
    SampleHeader ()
    unsigned int (8) track_ref_index;
    unsigned int (32) data_offset;
    signed int(32) data_length;
}
```

El multiplexor 30 y el demultiplexor 38 pueden solicitar un objeto de datos de extractor de medios usando el extractor de medios definido en el seudocódigo de ejemplo anterior. Por consiguiente, el demultiplexor 38, por ejemplo, se puede referir al extractor de medios solicitado cuando recupera datos desde una pista seleccionada para recuperar datos identificados desde otra pista referenciada por el extractor de medios solicitado.

La FIG. 11, es un diagrama de bloques que ilustra un extractor de medios de ejemplo 340 que contiene bits reservados para su futura extensibilidad. El extractor de medios 340 incluye el índice de referencia de pista 342 y el valor de desplazamiento de muestra 346, que pueden ser similares al extractor de medios 302 y el valor de desplazamiento de muestra 304 respectivamente. Además, el extractor de medios 340 incluye bits reservados 344,

que pueden comprender bits reservados usados para extensiones futuras para el extractor de medios. El siguiente seudocódigo proporciona una definición de una clase de ejemplo de un extractor de medios similar al extractor de medios 340.

```

5      class aligned (8) MediaExtractor () {
          unsigned int (8) track_ref_index;
          unsigned int (8) reserved_bits;
          signed int(8) sample_offset;
        }

```

10 El multiplexor 30 y el demultiplexor 38 pueden solicitar un objeto de datos de extractor de medios usando el extractor de medios definido en el seudocódigo de ejemplo anterior. Por consiguiente, el demultiplexor 38, por ejemplo, se puede referir al extractor de medios solicitado cuando recupera datos desde una pista seleccionada para recuperar datos identificados desde otra pista referenciada por el extractor de medios solicitado.

15 La FIG. 12, es un diagrama de bloques que ilustra un extractor de medios de ejemplo 350, que usa un valor de identificador de pista, en lugar de un valor de índice de referencia de pista. El uso de un valor de identificador de pista para identificar una pista se puede referir a la presentación de la caja de referencia de pista en el formato de ficheros de medios de base de la ISO. El ejemplo del extractor de medios 350 incluye el identificador de pista 352, bits reservados 354, y el valor de desplazamiento de muestra 356. Los bits reservados 354 son opcionales, como se indica por la línea discontinua alrededor de los bits reservados 354. Esto es, algunos ejemplos pueden incluir los bits reservados 354, mientras que otros ejemplos pueden omitir los bits reservados 354. El valor del desplazamiento de muestra 356 puede ser similar al valor del desplazamiento de muestras 304.

20 El identificador de pista 352 especifica la ID de pista de la pista desde la que extraer los datos. La muestra en la pista desde la que se extraen los datos puede estar temporalmente alineada (en la línea de tiempos de la decodificación de medios, usando la tabla de tiempo para muestras, ajustada por un desplazamiento especificado por el desplazamiento de muestra 356) conteniendo la muestra el extractor de medios 350. A la primera referencia de pista se puede asignar un valor de identificador de 1. El valor de 0 puede estar reservado para uso futuro y extensiones.

25 El siguiente seudocódigo proporciona una definición de ejemplo de una clase de extractor de medios similar al extractor de medios 350:

```

30      class aligned (8) MediaExtractor () {
          unsigned int (8) track_id;
          unsigned int (8) reserved_bits;
          signed int(8) sample_offset;
        }

```

35 El multiplexor 30 y el demultiplexor 38 pueden solicitar un objeto de datos de extractor de medios usando el extractor de medios definido en el seudocódigo de ejemplo anterior. Por consiguiente, el demultiplexor 38, por ejemplo, se puede referir al extractor de medios solicitado cuando recupera datos desde una pista seleccionada para recuperar datos identificados desde otra pista referenciada por el extractor de medios solicitado.

40 La FIG. 13, es un diagrama de bloques que ilustra un grupo de muestras del extractor de medios de ejemplo 360. El multiplexor 30 puede incluir el grupo de muestras del extractor de medios 360 en una caja de tipo mensajes (que tiene el identificador de tipo "MSG"), en un contenedor de caja de tabla de muestras. El multiplexor 30 se puede configurar para incluir cero o un objeto del grupo de muestras del extractor de medios 360 en la caja de mensajes. En el ejemplo de la FIG. 13, el grupo de muestras de extractor de medios 360 incluye el índice de referencia de pista 362, el tipo de grupo 364, la cuenta del número de grupo 366, los bits reservados 368 y los índices de descripción de grupo 370.

45 El índice de referencia de pista 362 especifica el índice de la referencia de pista usado para encontrar la pista desde la cual se extraen los datos desde los grupos de muestras bajo ciertos criterios. Esto es, el índice de referencia de pistas 362 identifica la pista desde la cual se extraen los datos identificados por el extractor de medios, en un modo similar al índice de referencia de la pista 302.

50 El valor del tipo de grupo 364 identifica el tipo de grupo de muestra al cual corresponde el grupo de muestras del extractor de medios 360. El valor del tipo de grupo 364 generalmente identifica los criterios usados para formar los grupos de muestras del grupo de muestreo y enlaza los criterios con una tabla de descripción del grupo de muestras con el mismo valor para el tipo de grupo en la pista identificada por el índice de referencia de pista 362. El valor del tipo de grupo 364 puede comprender un valor entero. De este modo, el valor de tipo de grupo del grupo de muestras del extractor de medios 360 puede ser el mismo que el tipo de grupo de la pista al que se refiere el índice de referencia de pista 362. Como alternativa, para un subconjunto temporal de video, el valor del tipo de grupo 34 se puede definir como "vtst", el grupo de muestras del extractor de medios se puede definir solo para ese tipo de grupo y la tabla de sintaxis no necesitaría el elemento de sintaxis de "grouping_type".

El valor de cuenta del número de grupo 366 puede describir el número de grupos de muestras en la pista de extractores de medios incluyendo el grupo de muestras del extractor de medios 360. Un valor de cero para el valor de cuenta del número de grupos 366 puede representar que todos los grupos de muestras bajo el criterio referenciado por el valor del tipo de grupo 364 se usan para formar la pista del extractor de medios. El índice de descripción de grupo 368 define un índice de la entrada del grupo de muestras que se usa para formar la pista del extractor de medios en la tabla de descripción del grupo de muestras.

De acuerdo con las técnicas de la presente divulgación, se puede usar un procedimiento de ensamblado para colocar todas las muestras en las entradas del grupo de muestras de modo que las muestras están temporalmente ordenadas, de modo que una muestra A que sigue a una muestra B en la pista del extractor de medios indica que la muestra A sigue a la muestra B en la pista referenciada por el índice de referencia de pista 362.

El siguiente pseudocódigo proporciona una definición de ejemplo de una clase de grupo de muestras de extractor de medios similar al grupo de muestras del extractor de medios 360:

```

class aligned (8) MedEtrSampleGroup () {
    unsigned int (8) track_ref_index;
    unsigned int (32) grouping_tipo;
    unsigned int (32) group_number_count;
    for (i = 0; i < group_number_count; i++)
        unsigned int (32) group_description_index;
}

```

El multiplexor 30 y el demultiplexor 38 pueden solicitar un objeto de datos de extractor de medios usando el extractor de medios definido en el pseudocódigo de ejemplo anterior. Por consiguiente, el demultiplexor 38, por ejemplo, se puede referir al extractor de medios solicitado cuando recupera datos desde una pista seleccionada para recuperar los datos identificados desde otra pista referenciada por el extractor de medios solicitado.

La FIG. 14, es un diagrama de bloques que ilustra un extractor de medios de ejemplo 380 que se puede usar en el contexto de ficheros de video conformes con el formato de ficheros de AVC. El ejemplo de extractor de medios 380 incluye el índice de referencia de pista 382, el valor del identificador temporal 384, los bits reservados 386, y el valor de desplazamiento de muestras 388. El índice de referencia de pista 382 y el valor de desplazamiento de muestras 388 se pueden usar de forma similar al índice de referencia de pistas 302 y el valor del desplazamiento de muestras 304, respectivamente. Los bits reservados 386 se pueden reservar para usos futuros, y no se asignan un valor semántico en este momento.

El valor del identificador temporal 384 especifica el nivel temporal de una muestra a extraer por el extractor de medios 380. En un ejemplo, el nivel temporal está en el intervalo de 0 a 7, inclusive. Como se ha tratado anteriormente, las imágenes codificadas pueden corresponder al nivel temporal, donde el nivel temporal generalmente describe la jerarquía de codificación entre las tramas. Por ejemplo, las tramas de clave (también denominadas como tramas de ancla) se pueden asignar al más alto nivel temporal, mientras que las tramas que no se usan como tramas de referencia se pueden asignar a niveles temporales relativamente inferiores. De este modo, el extractor de medios 380 puede identificar las muestras extraídas a partir de la pista referenciada por el índice de referencia de pista 382 por referencia al nivel temporal de las muestras, en lugar de identificar explícitamente las propias muestras. Una pista de extractores de medios con extractores de medios hasta un mayor valor que el definido por el valor del identificador temporal 384 puede corresponder a un punto de operación con una tasa de trama más alta.

El pseudocódigo siguiente proporciona una definición de ejemplo de una clase de extractor de medios similar al extractor de medios 380:

```

class aligned (8) MediaExtractor () {
    unsigned int (8) track_ref_index;
    unsigned int (3) temporal_id;
    unsigned int (5) reserved_bits;
    signed int (8) sample_offset;
}

```

El multiplexor 30 y el demultiplexor 38 pueden solicitar un objeto de datos de extractor de medios usando el extractor de medios definido en el pseudocódigo de ejemplo anterior. Por consiguiente, el demultiplexor 38, por ejemplo, se puede referir al extractor de medios solicitado cuando recupera datos desde una pista seleccionada para recuperar datos identificados desde otra pista referenciada por el extractor de medios solicitado.

La FIG. 15, es un diagrama de bloques que ilustra un extractor de medios de MVC de ejemplo 420 que se puede usar para modificar la MVC para incluir las pistas de extractores de medios. El ejemplo de extractor de medios 420 incluye una cabecera de unidad NAL opcional 422, el índice de referencia de pista 424, el desplazamiento de muestra 426, la cuenta de conjuntos de bytes continuos 428, y un bucle de valores incluyendo valores de desplazamiento de datos 430 y valores de la longitud de datos 432. El extractor de medios de MVC 420 se puede

usar para extraer un número de unidades NAL de un subconjunto de componentes de vistas desde una pista particular. El ejemplo de extractor de medios de MVC 420 puede saltar componentes de vistas en una pista cuando se extraen datos desde una muestra de la pista referenciada.

5 Cuando está presente, la cabecera de la unidad NAL 422 puede reflejar la cabecera de la unidad NAL de las unidades NAL identificadas por el extractor de medios de MVC 420. Esto es, los elementos de sintaxis de la cabecera de unidades NAL 422 se pueden generar de acuerdo con la sintaxis de la cabecera de la unidad NAL en el extractor o el procedimiento de generación de agregador definido en el formato de ficheros de MVC. En algunos ejemplos, el extractor puede no necesitar la cabecera de la unidad NAL 422, por ejemplo cuando una serie de extractores se genera para incluir las cabeceras de unidades NAL relacionadas.

10 El valor del índice de referencia de pista 424 especifica el índice de la referencia de pista a usar para encontrar la pista desde la cual extraer los datos. La muestra en la pista desde la cual se extraen los datos puede estar temporalmente alineada en la línea de tiempos de decodificación de medios, ajustada por un desplazamiento especificado por el valor del desplazamiento de muestras 426, que contiene la muestra el extractor de medios de MVC 420. La primera referencia de pista se puede designar para recibir un valor de índice de uno y el valor de cero para el valor del índice de referencia de pista puede estar reservado.

15 El valor del desplazamiento de muestras 426 define un desplazamiento, relativo para la localización temporal del extractor de medios de MVC 420, de la muestra a extraer que está localizada en la pista referenciada por el valor del índice de referencia de pista 424. Un valor de cero para el valor de desplazamiento de muestra 426 indica que la muestra a extraer está en la misma localización temporal, un uno negativo indica una muestra anterior, un uno positivo indica una muestra siguiente y así sucesivamente.

La cuenta de conjuntos de bytes continuos 428 describe el número de conjuntos de bytes continuos de la muestra de la pista desde la que se extraen los datos. Si la cuenta de conjuntos de bytes continuos 428 tiene un valor de cero, toda la muestra referenciada en la pista es para extraer. Los conjuntos de bytes continuos también se pueden referenciar como porciones separadas de una muestra.

25 Los valores del desplazamiento de datos 430 y los valores de longitud de datos 432 se producen en un bucle. En general, el número de iteraciones del bucle, esto es, el número de valores de desplazamientos de datos 430 y los valores de longitud de datos 432 están relacionados con el número de porciones de una muestra a extraer (por ejemplo, un número de conjuntos de bytes continuos). De este modo, se pueden extraer dos o más porciones de una muestra usando el extractor de medios de MVC 420. Para cada porción de una muestra a extraer, el valor correspondiente de los valores de desplazamiento de datos 430 indica el comienzo de la porción (por ejemplo, un primer byte de la porción, con relación el primer byte de la muestra) y el valor correspondiente de los valores de longitud de datos 432 indica la longitud, por ejemplo el número de bytes a copiar. En algunos ejemplos, un valor de cero para uno de los valores de la longitud de datos 432 puede indicar que todos los bytes restantes en la muestra se van a copiar, es decir, que la porción corresponde al byte indicado por el valor correspondiente de los valores de desplazamiento de datos 430 y todos los bytes contiguos hasta el final de la muestra.

El seudocódigo siguiente proporciona una definición de ejemplo de una clase de extractor de medios similar al extractor de medios de MVC 420:

```

class aligned (8) MediaExtractorMVC () {
  NALUnitHeader (); // se puede omitir en algunos ejemplos
40  unsigned int (3) track_ref_index;
  signed int (8) sample_offset;
  unsigned int(8) continuous_byte_set_count;
  for (i = 0; i < continuous_byte_set_count; i++) {
    unsigned int ((lengthSizeMinusOne + 1) * 8)
45    data_offset;
    unsigned int ((lengthSizeMinusOne + 1) * 8)
    data_length;
  }
}

```

50 El multiplexor 30 y el demultiplexor 38 pueden solicitar un objeto de datos de extractor de medios usando el extractor de medios definido en el seudocódigo de ejemplo anterior. Por consiguiente, el demultiplexor 38, por ejemplo, se puede referir al extractor de medios solicitado cuando recupera datos desde una pista seleccionada para recuperar datos identificados desde otra pista referenciada por el extractor de medios solicitado.

55 La FIG. 16, es un diagrama de bloques que ilustra otro ejemplo del extractor de medios de MVC 440 que se puede usar para modificar la MVC para incluir pistas de extractores de medios. El ejemplo de extractor de medios de MVC 440 identifica las unidades NAL particulares para extracción, en oposición a bytes específicos de una muestra como se describe con respecto al ejemplo de la FIG. 15. En el ejemplo de la FIG. 16, el extractor de medios de MVC 440 incluye una cabecera de unidad NAL opcional 442, el índice de referencia de pista 444, el desplazamiento de

muestra 446, la cuenta de conjuntos de NALU (unidad NAL) continuos 448, y un bucle de valores de desplazamiento de NALU 450 y número de unidades de NAL continuas 452. La cabecera de la unidad NAL 442, el índice de referencia de pista 444 y un valor de desplazamiento de muestra 446 se definen generalmente del mismo modo que la cabecera de unidad NAL 422, el índice de referencia 424 y el valor del desplazamiento de muestra 426, respectivamente.

La cuenta de conjuntos de NALU continuos 448 describe un número de unidades NAL continuas de la muestra de la pista desde la que se extraen los datos. En algunos ejemplos, si este valor se fija a cero, se extrae toda la muestra referenciada en la pista.

Los valores de desplazamiento de NALU 450 y los números de NALUS continuas 452 se producen en un bucle. En general, hay tantas instancias de valores de desplazamiento de NALU y números de NALU continuas como conjuntos de NALUS continuos, como se define por la cuenta de conjuntos de NALUS continuos 448. Cada valor de desplazamiento de NALU describe el desplazamiento de una unidad NAL correspondiente en la muestra de la pista desde la que se extraen los datos. Las unidades NAL que comienzan desde este desplazamiento de las unidades NAL se pueden extraer usando este extractor. Cada número de valor de NALU continuas describe el número entero de unidades NAL referenciadas únicas a copiar al conjunto correspondiente de unidades NAL.

El seudocódigo siguiente proporciona una definición de ejemplo de una clase de extractor de medios similar al extractor de medios de MVC 440:

```

Class aligned (8) MediaExtractorMVC () {
    NALUnitHeader(); // se puede omitir en algunos ejemplos
    unsigned int (8) track_ref_index;
    signed int (8) sample_offset;
    unsigned int (8) continuous_NALU_set_count;
    for (i = 0 ; i < continuous_NALU_Set_Count; i++) {
        unsigned int ((lengthSizeMinusOne + 1) * 8)
            NALU_offset;
        unsigned int((lengthSizeMinusOne + 1) * 8)
            num_continuous_NALUs
    }
}

```

El multiplexor 30 y el demultiplexor 38 pueden solicitar un objeto de datos de extractor de medios usando el extractor de medios definido en el seudocódigo de ejemplo anterior. Por consiguiente, el demultiplexor 38, por ejemplo, se puede referir al extractor de medios solicitado cuando recupera datos desde una pista seleccionada para recuperar datos identificados desde otra pista referenciada por el extractor de medios solicitado.

La FIG. 17, es un diagrama de bloques que ilustra otro ejemplo del extractor de medios de MVC 460 que agrega unidades NAL en la misma componente de vista cuando hay más de una unidades NAL para una componente de vista. El extractor de medios de MVC 460 se puede usar a continuación para extraer las componentes de vistas identificadas. En el ejemplo de la FIG. 17, el extractor de medios de MVC 460 incluye una cabecera de unidad NAL opcional 462, el índice de referencia de pista 464, el desplazamiento de muestra 466, la cuenta de conjuntos de vistas continuas 468, y un bucle de valores del desplazamiento de componentes de vistas 470 y las cuentas de componentes de vistas 472. La cabecera de la unidad NAL 462, el índice de referencia de pista 464, y el valor de desplazamiento de muestra 466 se definen generalmente del mismo modo que la cabecera de la unidad NAL 422, el índice de referencia de pistas 424 y el valor de desplazamiento de muestras 426, respectivamente.

La cuenta de conjuntos de vistas continuas 468 define un número de componentes de vistas continuas de una muestra identificada en la pista identificada por el índice de referencia de pista 464 desde el cual se extraen los datos. El multiplexor 30 puede fijar el valor de conjuntos de vistas continuas 468 a cero para indicar que la muestra referenciada entera en la pista es para extraer.

Los valores de desplazamientos de la componentes de vistas 470 y las cuentas de componentes de vistas 472 ocurren en un bucle. En general, hay tantas iteraciones del bucle como el valor de la cuenta de conjuntos de vistas continuas 468 y cada bucle corresponde a uno de los conjuntos de vistas continuas. Cada uno de los valores de desplazamiento de componentes de vistas 470 indica el desplazamiento de la primera componente de vista en la muestra de la pista desde la que se extraen los datos para un conjunto de vistas continuas correspondientes. Las componentes de vistas que comienzan desde este desplazamiento de las componentes de vistas se pueden extraer a continuación usando el extractor de medios de MVC 460. Cada una de las cuentas de componentes de vistas 472 describe el número de componentes de vistas referenciadas enteras en la muestra para copiar al conjunto de vistas continuas correspondiente.

El seudocódigo siguiente proporciona una definición de ejemplo de una clase de extractor de medios similar al extractor de medios de MVC 460:

```

class aligned (8) MediaExtractorMVC () {
    NALUnitHeader (); // se puede omitir en algunos ejemplos
    unsigned int (8) track_ref_index;
    signed int (8) sample_offset;
5    unsigned int (8) continuous_view_set_count;
    for (i = 0 ; i < continuous_view_set_count; i++) {
        unsigned int ((lengthSizeMinusOne + 1) * 8)
            view_component_offset;
10        unsigned int ((lengthSizeMinusOne + 1) * 8)
            view_component_count
    }
}

```

El multiplexor 30 y el demultiplexor 38 pueden solicitar un objeto de datos de extractor de medios usando el extractor de medios definido en el pseudocódigo de ejemplo anterior. Por consiguiente, el demultiplexor 38, por ejemplo, se puede referir al extractor de medios solicitado cuando recupera datos desde una pista seleccionada para recuperar datos identificados desde otra pista referenciada por el extractor de medios solicitado.

La FIG. 18, es un diagrama de bloques que ilustra otro ejemplo del extractor de medios de MVC 480 que se puede usar para referirse a diversas pistas. En el ejemplo de la FIG. 18, el extractor de medios de MVC 480 incluye una cabecera de unidad NAL opcional 482, la cuenta de conjuntos de vistas continuas 484, y un bucle de valores de desplazamientos de muestra 486, los valores de índices de referencia de pista 488, los valores de desplazamientos de las componentes de vistas 490, y las cuentas de componentes de vistas 492. La cabecera de la unidad NAL 482 se puede definir de forma similar a la cabecera de unidad NAL 422 y se puede omitir en algunos ejemplos.

La cuenta de conjuntos de vistas continuas 484 da el número de componentes de vistas continuas de la muestra de la pista de extractores de medios, con índice de referencia de pista de track_ref_index, a partir del cual se extraen los datos. El valor de track_ref_index puede especificar el índice de la referencia de pista a usar para encontrar la pista desde la cual se extraen los datos. Las componentes de vistas en la pista desde las cuales se extraen los datos pueden estar alineadas temporalmente (en la línea de tiempos de decodificación de medios, usando la tabla de tiempo para muestras, ajustadas por un desplazamiento especificado por el valor correspondiente de los valores de desplazamiento de muestras 486) conteniendo la muestra el MediaExtractorMVC. La primera referencia de pista puede tener el valor índice 1; el valor 0 puede estar reservado para un uso futuro.

El ejemplo de extractor de medios de MVC 480 incluye cada uno de los valores de desplazamientos de muestras 486, los valores de índices de referencia de pista 488, los valores de desplazamientos de componentes de vistas 490, y las cuentas de componentes de vistas 492 en un bucle. Cada iteración del bucle corresponde a una pista particular desde la cual se extraen los datos para una muestra correspondiente al extractor de medios de MVC 480.

Los valores de desplazamientos de muestras 486 definen el índice relativo de la muestra en la pista referenciada por el valor correspondiente de los valores de índices de referencia de pista 488, que se puede usar como la fuente de información. La muestra 0 (cero) es la muestra en la pista identificada por el valor correspondiente de los valores de índices de referencia de pista 488 con el mismo tiempo, o el más próximo anterior, de decodificación para la muestra que contiene el extractor de medios de MVC 480, la muestra 1 (uno) es la siguiente muestra, la muestra -1 (menos 1) es la muestra anterior, y así sucesivamente.

Cada uno de los valores de índices de referencia de pista 488 especifica el índice de la referencia de pista a usar para encontrar la pista a partir de la cual se extraen los datos para la iteración correspondiente del bucle. Usando múltiples valores de índices de referencia de pista, el extractor de medios de MVC 480 puede extraer datos a partir de múltiples pistas diferentes.

Cada uno de los valores de desplazamientos de componentes de vistas 490 describe el desplazamiento de la primera componente de vista en la muestra de la pista, con índice de referencia de pista que corresponde al valor correspondiente de los valores de índices de referencia de pista 488 de esta iteración del bucle, desde la cual se extraen los datos. Los componentes de vistas que comienzan desde este desplazamiento de componentes de vistas se pueden extraer usando el extractor de medios de MVC 480. En algunos ejemplos, se puede construir un extractor de medios similar a los de las FIG. 15 - 17 que tienen una estructura de bucle anidado, en el que un bucle exterior itera sobre las pistas desde las cuales se extraerán las muestras y un bucle interior itera sobre las muestras a extraer desde las pistas correspondientes. Cada una de las cuentas de componentes de vistas 492 describe un número de componentes de vistas referenciadas en la muestra de la pista con un índice de referencia de pista correspondiente al valor actual de los valores de índices de referencia de pista 488 en esta iteración del bucle.

El pseudocódigo siguiente proporciona una definición de ejemplo de una clase de extractor de medios similar al extractor de medios de MVC 480:

```

class aligned (8) MediaExtractorMVC () {
    NALUnitHeader (); // se puede omitir en algunos ejemplos
    unsigned int (8) continuous_view_set_count;
    for (i = 0 ; i < continuous_view_set_count; i++) {
5         signed int (8) sample_offset;
            unsigned int (8) track_ref_index;
            unsigned int ((lengthSizeMinusOne + 1) * 8)
                view_component_offset;
10         unsigned int((lengthSizeMinusOne + 1) * 8)
            view_component_count
        }
    }
}

```

15 El multiplexor 30 y el demultiplexor 38 pueden solicitar un objeto de datos de extractor de medios usando el extractor de medios definido en el seudocódigo de ejemplo anterior. Por consiguiente, el demultiplexor 38, por ejemplo, se puede referir al extractor de medios solicitado cuando recupera datos desde una pista seleccionada para recuperar datos identificados desde otra pista referenciada por el extractor de medios solicitado.

20 La FIG. 19, es un diagrama de bloques que ilustra otro ejemplo del extractor de medios de MVC 500 que señala la duración del extractor. El extractor de medios de MVC 500 puede proporcionar una o más ventajas cuando diferentes muestras en la pista del extractor de medios comparten los mismos elementos de sintaxis de los extractores. En el ejemplo de la FIG. 19, el extractor de medios de MVC 500 incluye la cuenta de muestras 502, la cuenta de muestras de vistas continuas 504, los valores de desplazamientos de muestras 506, los índices de referencia de pistas 508, los desplazamientos de componentes de vistas 510 y las cuentas de componentes de vistas 512.

25 La cuenta de conjuntos de vistas continuas 504, los valores de desplazamientos de muestras 506, los índices de referencia de pistas 508, los desplazamientos de componentes de vistas 510, y las cuentas de componentes de vistas 512 se pueden definir en general de acuerdo con los valores correspondientes de la cuenta de conjuntos de vistas continuas 484, los valores de desplazamientos de muestras 486, los índices de referencias de pistas 488, los desplazamientos de componentes de vistas 490, y las cuentas de componentes de vistas 492. La cuenta de muestras 502 puede definir el número de muestras continuas en la pista del extractor de medios que contiene el extractor de medios de MVC 500 que usa el mismo extractor de medios.

30 El seudocódigo siguiente proporciona una definición de ejemplo de una clase de extractor de medios similar al extractor de medios de MVC 500:

```

class aligned (8) MediaExtractorMVC () {
    unsigned int (8) sample_count;
35     unsigned int (8) continuous_view_set_count;
    for (i = 0 ; i < continuous_view_set_count; i++) {
        signed int (8) sample_offset;
        unsigned int (8) track_ref_index;
        unsigned int ((lengthSizeMinusOne + 1) * 8)
40         view_component_offset;
        unsigned int((lengthSizeMinusOne + 1) * 8)
            view_component_count
    }
}

```

45 El multiplexor 30 y el demultiplexor 38 pueden solicitar un objeto de datos de extractor de medios usando el extractor de medios definido en el seudocódigo de ejemplo anterior. Por consiguiente, el demultiplexor 38, por ejemplo, se puede referir al extractor de medios solicitado cuando recupera datos desde una pista seleccionada para recuperar datos identificados desde otra pista referenciada por el extractor de medios solicitado.

50 La FIG. 20, es un diagrama de bloques que ilustra otro ejemplo del extractor de medios de MVC 520 que define un conjunto de diferente de extractores. Para cada muestra en una pista de extractores de medios, la muestra puede o bien usar uno o más del conjunto de extractores o una referencia a los extractores. Esto es, se puede definir un conjunto de extractores de medios similar al extractor de medios de MVC 520 y cada muestra puede o bien usar uno o más del conjunto de extractores o una referencia a los extractores para identificar una muestra de otra pista.

55 El ejemplo del extractor de medios de MVC 520 incluye el valor del identificador de extractor 522, el valor del desplazamiento de muestra 524, el valor del índice de referencia 526, la cuenta de conjuntos de vistas continuas 528, y un bucle que incluye los desplazamientos de componentes de vistas 530 y las cuentas de componentes de vistas 532. El valor de desplazamiento de muestras 524, la cuenta de conjuntos de vistas continuas 528, los desplazamientos de componentes de vistas 530 y las cuentas de componentes de vistas 532 se pueden definir de acuerdo con los valores correspondientes de la cuenta de conjuntos de vistas continuas 484, los valores de

desplazamientos de muestras 486, los desplazamientos de componentes de vistas 490, y las cuentas de componentes de vistas 492. El valor del índice de referencia de pista 526 se puede definir de acuerdo con, por ejemplo, el índice de referencia de pista 464.

El valor del identificador de extractor 522 define un identificador del extractor, esto es, el extractor de medios de MVC 520. A los extractores en la misma pista de extractores de medios se asignan diferentes valores de identificador de extractores, de modo que una muestra en la pista de extractores de medios se puede referir al valor del identificador de extractor a usar como el extractor de medios. También se puede definir una caja de extractores de referencia para incluir un número de extractores y un identificador de extractores de referencia. El número del valor de extractores puede proporcionar el número de extractores usados para copiar los datos para la muestra en la pista de extractores. Cuando el número del valor de extractores es igual a cero, se puede usar el extractor que tiene un identificador de extractor predeterminado, por ejemplo, un identificador de extractor igual a cero. El identificador del extractor de referencia puede proporcionar el identificador de extractor del extractor usado para copiar los datos para la muestra en la pista de extractores. Esta caja se puede incluir en una muestra de la pista de extractores de medios.

El seudocódigo siguiente proporciona una definición de ejemplo de una clase de extractores de medios similar al extractor de medios de MVC 520:

```

class aligned (8) MediaExtractorMVC () {
    unsigned int ((lengthSizeMinusOne + 1) * 8) extractor_id;
    signed int (8) sample_offset;
    unsigned int (8) track_ref_index;
    for (i = 0 ; i < continuous_view_set_count; i++) {
        unsigned int((lengthSizeMinusOne + 1) * 8)
            view_component_offset;
        unsigned int((lengthSizeMinusOne + 1) * 8)
            view_component_count;
    }
}

```

El multiplexor 30 y el demultiplexor 38 pueden solicitar un objeto de datos de extractor de medios usando el extractor de medios definido en el seudocódigo de ejemplo anterior. Por consiguiente, el demultiplexor 38, por ejemplo, se puede referir al extractor de medios solicitado cuando recupera datos desde una pista seleccionada para recuperar datos identificados desde otra pista referenciada por el extractor de medios solicitado.

El seudocódigo siguiente proporciona una definición de ejemplo de una clase de caja de extractores de referencia para la caja de extractores de referencia descrita anteriormente:

```

class aligned (8) RefExtractorMVC () {
    unsigned int ((lengthSixeMinusOne + 1) * 8)
        num_extractor;
    for (i = 0 ; i < num_extractor; i++)
        ref_extractor_id;
}

```

La FIG. 21, es un diagrama de bloques que ilustra otro ejemplo del extractor de medios de MVC 550 que se forma usando un mapa de grupos de muestras. El ejemplo del extractor de medios de MVC 550 especifica los grupos de unidades NAL desde una serie de entradas de muestras, cada una de las cuales contribuye a las unidades NAL continuas en el mapa de grupos de muestras. En el ejemplo de la FIG. 22, el extractor de medios de MVC 550 incluye la cuenta del grupo de NALU 552 y un bucle que incluye índices de pista 554, los índices de descripción de grupo 556, las muestras de mapa de comienzo de NALU 558, y las cuentas de vistas de NALU 560.

La cuenta de grupos de NALU 552 especifica el número de grupos de unidades NAL desde una entrada de mapa de grupos de muestras en una pista de referencia. Los índices de referencia de pista 554 especifican cada uno el índice de referencia de la pista a usar para encontrar la pista desde la cual se extraen los datos para la iteración correspondiente del bucle. Los índices de descripción de grupo 556 especifican cada uno el índice la entrada de mapa de grupos de muestras que se usa para formar el grupo de unidades NAL para la iteración correspondiente del bucle. Las muestras de mapa de comienzo de NALU 558 especifican cada una el desplazamiento de la unidad NAL en el grupo de muestras de mapa con un índice de entrada del mapa de muestras del índice correspondiente de los índices de descripción de grupo 556 en la iteración correspondiente del bucle. Las cuentas de vistas de NALU 560 especifican el número de unidades NAL continuas, a extraer dentro del extractor de medios. En el mapa de grupos de muestras con un índice de entrada del mapa de muestras del índice correspondiente de los índices de descripción de grupo 556 en la iteración correspondiente del bucle.

El seudocódigo siguiente proporciona una definición de ejemplo de una clase de extractor de medios similar al extractor de medios de MVC 550:

```

class aligned (8) MedEtrMapSampleGroup () {
  unsigned int (32) NALU_group_count;
5   for (i =0; i< NALU_group_count; i++) {
      unsigned int (8) track_ref_index;
      unsigned int (32) group_description_index;
      unsigned int (8) NALU_start_map_sample;
10      unsigned int (8) NALU_view_count;
  }
}

```

El multiplexor 30 y el demultiplexor 38 pueden solicitar un objeto de datos de extractor de medios usando el extractor de medios definido en el seudocódigo de ejemplo anterior. Por consiguiente, el demultiplexor 38, por ejemplo, se puede referir al extractor de medios solicitado cuando recupera datos desde una pista seleccionada para recuperar datos identificados desde otra pista referenciada por el extractor de medios solicitado.

Las técnicas de la presente divulgación pueden incluir un procedimiento de ensamblado para disponer las componentes de vistas de muestras en un grupo de muestras. Las componentes de vistas en las muestras de las entradas de grupo de muestras se ordenan en un modo oportuno de modo que la componente de vista A en una muestra A está siguiendo a una componente de vista en una muestra B en la pista de extractores de medios si la muestra A sigue a la muestra B en una pista original (con el índice de un índice de referencia de pista); una componente de vista en la muestra A está siguiendo una componente de vista en la muestra B en la pista de extractores de medios si la muestra A tiene un tiempo de decodificación anterior que la muestra B; dos componentes de vistas en la misma muestra de una pista están siguiendo el orden de presentación en la tabla de sintaxis del grupo de muestras del mapa de extractores de medios; dos componentes de vistas en la misma muestra de una pista están siguiendo el orden original si pertenecen al mismo grupo de unidad NAL, es decir se extraen por el elemento de sintaxis del mismo bucle en el grupo del mapa de muestras de extractores de medios; y dos componentes de vistas están siguiendo el orden de los índices del orden de visión como se especifica en la caja de identificadores de vistas en el formato de ficheros de MVC si se extraen desde muestras en diferentes pistas pero con el mismo sello temporal.

La FIG. 22 es un diagrama de bloques que ilustra un ejemplo de caja de selección de pistas de 3GPP modificada 390 para señalar atributos adicionales para la caja de selección de pistas. La normativa más reciente del 3GPP, como la de este documento, especifica una AttributeList que incluye atributos que describen el lenguaje, el ancho de banda, el códec, el tamaño de pantalla, el tamaño máximo de paquete, y el tipo de medio. La lista de atributos 392 de la caja de selección de pistas de 3GPP 390 incluye el valor del lenguaje 394, el valor del ancho de banda 396, el valor de códec 398, y el valor del tamaño de pantalla 400, que señalan estos atributos de acuerdo con la normativa de 3GPP existente. Además, las técnicas de la presente divulgación pueden modificar la caja de selección de pistas de 3GPP existente para incluir el valor de la tasa de trama 406, el valor del identificador temporal 408, y en algunos casos, el valor del número de vista de pantalla 410 y el valor de la lista de vistas de salida 412.

El valor del lenguaje 394 define un valor de LANG de un tipo de grupo de un atributo de "alt-group" en el SDP del nivel de sesión como se define en la cláusula 5.3.3.4 de la normativa de 3GPP existente. El valor del ancho de banda 396 define un valor de un atributo "b=AS" en el SDP del nivel de medios. El valor de códec 398 define un valor de SampleEntry en la caja de descripción de muestras de una pista de medios. El valor del tamaño de pantalla 400 define los campos de ancho y alto de un valor de MP4VisualSampleEntry y el valor de H.263SampleEntry en una pista de medios. El valor del tamaño de paquete máximo 402 define un valor para el campo MaxPacketSize en RTPHintSampleEntry, por ejemplo en una pista de indicaciones de RTP. El valor del tipo de medio 404 describe un HandlerType en una caja de Manejador de una pista de medios. En general, estos valores corresponden a la normativa de 3GPP existente.

El valor de la tasa de trama 406 describe la tasa de trama de una pista de video o pista de extractor de medios correspondiente a la caja de selección de pistas de 3GPP 390. El valor del identificador temporal 408 corresponde al identificador temporal de la pista de video correspondiente a la caja de selección de pista de 3GPP 390, y puede depender de las pistas con menores valores de identificador temporal. En algunos ejemplos, el multiplexor 30 puede indicar que el valor del valor de identificador temporal 408 no se especifica estableciendo el valor a un valor pre-configurado "no especificado", por ejemplo, 8. En general, el multiplexor 30 puede indicar que el valor del valor de identificador temporal 408 para una pista no de video no se especifica. En algunos ejemplos, el multiplexor 30 también puede indicar que el valor del valor de identificador temporal 408 no se especifica cuando la pista de video correspondiente no contiene extractores de medios y/o no se referencia por otra pistas como un subconjunto temporal.

En ejemplos para los que la MVC se considera en el 3GPP, el multiplexor 30 puede incluir atributos adicionales del valor del número de vista de representación 410 y el valor de la lista de vistas de salida 412. En tales ejemplos, el multiplexor 30 puede omitir el valor del identificador temporal 408. El valor del número de vistas de representación

410 describe un número de vistas que se sacan para la pista correspondiente. El número de la vista a emitir y el número de la vista a decodificar no son necesariamente iguales, por ejemplo, cuando una vista a representar está codificada con referencia a una vista que no se ha representado. El valor de la lista de salida 412 puede definir una lista de N identificadores de vistas que identifican las N vistas a emitir.

5 La FIG. 23 es un diagrama de flujo que ilustra un procedimiento de ejemplo para usar los extractores de medios de acuerdo con las técnicas de la presente divulgación. Inicialmente, un dispositivo fuente, tal como un dispositivo fuente de A/V 20 (FIG. 1) construye una pista de video para un fichero conforme a un formato de fichero de acuerdo con las técnicas de la presente divulgación. Esto es, el multiplexor 30 ensambla los datos de video codificado en la pista de modo que la pista de video incluye muestras de video codificado que incluyen una o más unidades NAL
10 (600). El multiplexor 30 también construye un extractor que referencia algunas o todas de la una o más unidades NAL de la pista de video (602) y construye una pista de extractores que incluye el extractor (604). Además, el multiplexor 30 puede incluir muestras de video codificadas en la pista de extractores de medios y pistas adicionales incluyendo las muestras de video codificadas y/o los extractores de medios.

15 El multiplexor 30 puede emitir a continuación el fichero (606). El fichero se puede emitir como una señal a través de un transmisor, transceptor, interfaz de red, modem u otro medio de salida de señal, o el fichero se puede emitir a un medio de almacenamiento a través de una interfaz hardware, tal como una interfaz USB, un grabador de medios magnético, un grabador óptico, u otra interfaz hardware.

20 El dispositivo de destino de A/V 40 puede recibir finalmente el fichero (608), por ejemplo recibiendo la señal o leyendo el medio de almacenamiento. El demultiplexor 38 puede seleccionar una de las dos (o más) pistas a decodificar (610). El demultiplexor 38 puede seleccionar una de las pistas en base a las capacidades de decodificación del decodificador de video 48, las capacidades de representación de la salida de video 44, u otros criterios. Cuando se selecciona una pista de extractores, el demultiplexor 38 puede recuperar las unidades NAL referenciadas por los extractores en la pista de extractores desde la pista en la que se almacenan las muestras de video codificadas identificadas por los extractores.

25 El demultiplexor 38 puede descartar las muestras de video codificadas (u otras unidades NAL) que no están en la pista seleccionada y no están identificadas por al menos un extractor en la pista seleccionada. Esto es, el demultiplexor 38 puede evitar enviar tales muestras de video codificadas al decodificador de video 48, de modo que el decodificador de video 48 no necesita encargarse de la decodificación de los datos de video no usados.

30 En uno o más ejemplos, las funciones descritas se pueden implementar en hardware, software, firmware, o cualquier combinación de los mismos. Si se implementa en software, las funciones se pueden almacenar o transmitir sobre un medio legible por ordenador como una o más instrucciones o código. El medio legible por ordenador puede incluir medios de almacenamiento legibles por ordenador tales como los medios de almacenamiento de datos o medios de comunicación incluyendo cualquier medio que facilite la transferencia de un programa de ordenador desde un sitio a otro. El medio de almacenamiento de datos puede ser cualquier medio disponible que se puede acceder por uno o
35 más ordenadores o uno o más procesadores para recuperar las instrucciones, código y/o estructuras de datos para la implementación de las técnicas descritas en la presente divulgación. A modo de ejemplo, y no de limitación, tales medios de almacenamiento legibles por ordenador pueden comprender RAM, ROM, EEPROM, CD-ROM u otro almacenamiento de disco óptico, almacenamiento de disco magnético, u otros dispositivos de almacenamiento magnético, memoria flash, o cualquier otro medio que se puede usar para almacenar el código de programa deseado en la forma de instrucciones o estructuras de datos y que se puede acceder por un ordenador. También cualquier conexión se denomina adecuadamente un medio legible por ordenador. Por ejemplo, si las instrucciones se transmiten desde un sitio web, servidor u otra fuente remota usando un cable coaxial, cable de fibra óptica, par
40 trenzado, línea de abonado digital (DSL), o tecnologías inalámbricas tales como infrarrojos, radio, y microondas, entonces el cable coaxial, cable de fibra óptica, par trenzado, DSL, o tecnologías inalámbricas tales como infrarrojos, radio, y microondas están incluidas en la definición de medio. Se debería entender, sin embargo, que un medio de almacenamiento legible por ordenador y un medio de almacenamiento de datos no incluyen las conexiones, ondas portadoras, señales u otros medios transitorios. Los discos magnéticos y los discos ópticos, usados en este documento, incluyen el disco compacto (CD), el disco láser, disco óptico, disco versátil digital (DVD), disco flexible y disco Blu-ray donde los discos magnéticos usualmente reproducen datos magnéticamente, mientras que los discos
45 ópticos reproducen datos ópticamente con láseres. Las combinaciones de los anteriores también se deberían incluir dentro el ámbito de los medios legibles por ordenador.

50 Las instrucciones codificadas en un medio legible por ordenador se pueden ejecutar por uno o más procesadores, tales como uno o más procesadores de señal digital (DSP), microprocesadores de propósito general, circuitos integrados de aplicación específica (ASIC), redes de lógica programable en campo (FPGA), u otros integrados equivalentes o circuitería lógica discreta. Por consiguiente, el término "procesador" como se usa en este documento se puede referir a cualquiera de las estructuras anteriores o cualquier otra estructura adecuada para la implementación de las técnicas descritas en este documento. Además, en algunos aspectos, la funcionalidad descrita en este documento se puede proporcionar dentro de hardware dedicado y/o módulos software configurados para codificar y decodificar, o incorporar en un códec combinado. También las técnicas se podrían implementar
60 completamente en uno o más circuitos o elementos lógicos.

- Las técnicas de la presente divulgación se pueden implementar en una amplia diversidad de dispositivos o aparatos, incluyendo un teléfono inalámbrico, un circuito integrado (IC) o un conjunto de IC (por ejemplo un conjunto de chips). Diversos componentes, módulos o unidades se describen en la presente divulgación para enfatizar aspectos funcionales de los dispositivos configurados para realizar las técnicas desveladas, pero no necesariamente requieren la realización por diferentes unidades de hardware. Más bien, como se ha descrito anteriormente, las diversas unidades se pueden combinar en una unidad hardware de códec o proporcionar por una colección de unidades hardware inter-operativas, incluyendo uno o más procesadores como se ha descrito anteriormente, en conjunción con software y/o firmware adecuados.
- 5
- 10 Se han descrito diversos ejemplos. Estos y otros ejemplos están dentro del alcance de las siguientes reivindicaciones.

REIVINDICACIONES

1. Un procedimiento de codificación de datos de video, comprendiendo el procedimiento:

5 construir, por un dispositivo de video fuente (20), una primera pista que incluye una muestra de video que comprende una pluralidad de unidades de la capa de acceso de red (NAL) basadas en los datos de video codificados, en el que la muestra de video está incluida en una unidad de acceso;
 10 construir, por un dispositivo de video fuente (20), una segunda pista que incluye un extractor (420; 440; 460; 480; 500; 520; 550) que identifica al menos una de la pluralidad de unidades NAL en la muestra de video de la primera pista, comprendiendo la al menos una de la pluralidad de unidades NAL una primera unidad NAL identificada, y en el que el extractor (420; 440; 460; 480; 500; 520; 550) identifican una segunda unidad NAL de la unidad de acceso;
 incluyendo la primera pista y la segunda pista en un fichero de video conforme con el formato de ficheros de medios de base de la Organización Internacional para la Normalización (ISO); y emitir el fichero de video

15 **caracterizado porque** la primera unidad NAL identificada y la segunda unidad NAL identificada son no consecutivas.

2. El procedimiento de la reivindicación 1, en el que construir la segunda pista comprende además incluir una o más unidades NAL adicionales en la segunda pista, basadas en los datos codificados, que no están incluidas en la pluralidad de unidades NAL de la primera pista.

20 3. El procedimiento de la reivindicación 2, que comprende además construir una tercera pista que incluye un primer extractor (420; 440; 460; 480; 500; 520; 550) que identifica una o más de la pluralidad de unidades NAL de la primera pista y un segundo extractor (420; 440; 460; 480; 500; 520; 550) que identifica al menos una o más de las unidades NAL de la segunda pista.

4. El procedimiento de la reivindicación 3, en el que construir la tercera pista comprende además incluir una o más unidades NAL en la tercera pista que no están incluidas en la primera pista y la segunda pista.

25 5. El procedimiento de la reivindicación 1, en el que construir la segunda pista comprende construir el extractor (420; 440; 460; 480; 500; 520; 550) para identificar cada una de la pluralidad de unidades NAL de la muestra de video de la primera pista, y en el que el extractor (420; 440; 460; 480; 500; 520; 550) causa que un dispositivo de destino extraiga cada una de la pluralidad de unidades NAL de la muestra de video como un conjunto.

30 6. El procedimiento de la reivindicación 1, en el que construir la segunda pista comprende construir el extractor (420; 440; 460; 480; 500; 520; 550) para identificar la una o más de la pluralidad de unidades NAL de la muestra de video especificando un intervalo de bytes de la una o más de la pluralidad de unidades NAL de la muestra de video en la primera pista del fichero de video.

35 7. El procedimiento de la reivindicación 1, en el que la primera pista y la segunda pista forman un grupo de conmutación, de modo que o bien la primera pista o la segunda pista es seleccionable para la decodificación por un dispositivo de destino en base a las características de cada pista.

8. El procedimiento de la reivindicación 7, en el que la construcción de la segunda pista comprende:

señalar una tasa de trama de la segunda pista
 señalar un identificador temporal de la muestra de video de la primera pista para la segunda pista; y
 en el que cuando la segunda pista comprende más de una vista, construir la segunda pista comprende además:
 40 señalar un valor representativo de un número de vistas a representar después de la decodificación de la segunda pista;
 señalar uno o más valores de identificador de vista para las vistas a representar para la segunda pista; y
 señalar un valor representativo de un número de vistas a decodificar para la segunda pista.

9. Un aparato de codificación de los datos de video, comprendiendo el aparato:

45 medios (20) para construir una primera pista que incluye una muestra de video que comprende una pluralidad de unidades de la capa de acceso de red (NAL) basadas en los datos de video codificados, en el que la muestra de video está incluida en una unidad de acceso;
 medios (20) para construir una segunda pista que incluye un extractor (420; 440; 460; 480; 500; 520; 550) que
 50 identifica al menos una de la pluralidad de unidades NAL en la muestra de video de la primera pista, comprendiendo la al menos una de la pluralidad de unidades NAL una primera unidad NAL identificada, y en el que el extractor (420; 440; 460; 480; 500; 520; 550) identifica una segunda unidad NAL de la unidad de acceso;
 medios para incluir la primera pista y la segunda pista en un fichero de video conforme con el formato de ficheros de medios de base de la Organización Internacional para la Normalización (ISO); y
 medios (32) para emitir el fichero de video,

caracterizado porque la primera unidad NAL identificada y la segunda unidad NAL identificada son no consecutivas.

10. Un procedimiento de decodificación de datos de video, comprendiendo el procedimiento:

5 recibir por un demultiplexor (38) de un dispositivo de destino (40), un fichero de video conforme con el formato de ficheros de medios de base de la Organización Internacional para la Normalización (ISO), comprendiendo el fichero de video una primera pista y una segunda pista, incluyendo la primera pista una muestra de video que comprende una pluralidad de unidades de la capa de acceso de red (NAL) correspondientes a datos de video codificados, en el que la muestra de video está incluida en una unidad de acceso, e incluyendo la segunda pista un extractor (420; 440; 460; 480; 500; 520; 550) que identifica al menos una de la pluralidad de unidades NAL de la primera pista, comprendiendo al menos una de la pluralidad de unidades NAL una primera unidad NAL identificada, y en el que el extractor (420; 440; 460; 480; 500; 520; 550) identifica una segunda unidad NAL de la unidad de acceso;

10 seleccionar la segunda pista a decodificar; y

15 enviar los datos de video codificados de la primera unidad NAL y la segunda unidad NAL identificadas por el extractor (420; 440; 460; 480; 500; 520; 550) de la segunda pista a un decodificador de video (48) del dispositivo de destino (40),

caracterizado porque la primera unidad NAL identificada y la segunda unidad NAL son no consecutivas.

11. El procedimiento de la reivindicación 10 que comprende además descartar cada una de la pluralidad de unidades NAL de la primera pista que no están identificadas por el extractor (420; 440; 460; 480; 500; 520; 550) de la segunda pista.

12. El procedimiento de la reivindicación 10, en el que la segunda pista comprende además una o más unidades NAL que no están incluidas en la primera pista, comprendiendo el procedimiento además enviar los datos de video codificados de la una o más unidades NAL de la segunda pista al decodificador de video (48).

13. El procedimiento de la reivindicación 10, en el que el fichero de video comprende además una tercera pista que incluye una pluralidad de unidades NAL correspondientes a datos de video codificados, comprendiendo además el procedimiento enviar los datos de video codificados de la pluralidad de unidades NAL de la tercera pista al decodificador de video (48).

14. Un aparato (40) de decodificación de datos de video, comprendiendo el aparato:

30 medios (36) para recibir un fichero de video conforme con el formato de ficheros de medios de base de la Organización Internacional de Normalización (ISO), comprendiendo el fichero de video una primera pista y una segunda pista, incluyendo la primera pista muestras de video que comprenden una pluralidad de unidades de la capa de acceso de red (NAL) correspondientes a datos de video codificados, en el que la muestra de video está incluida en una unidad de acceso, e incluyendo la segunda pista un extractor (420; 440; 460; 480; 500; 520; 550) que identifica al menos una de la pluralidad de unidades NAL de la primera pista, comprendiendo la al menos una de la pluralidad de unidades NAL una primera unidad NAL identificada y en el que el extractor (420; 440; 460; 480; 500; 520; 550) identifica una segunda unidad NAL de la unidad de acceso;

35 medios para seleccionar la segunda pista a decodificar; y

medios para enviar datos de video codificados de la primera unidad NAL y la segunda unidad NAL identificadas por el extractor (420; 440; 460; 480; 500; 520; 550) de la segunda pista a un decodificador de video del aparato.

caracterizado porque la primera unidad NAL identificada y la segunda unidad NAL son no consecutivas.

15. Un medio de almacenamiento legible por ordenador que comprende instrucciones que, cuando se ejecutan, causan que un procesador realice el procedimiento de acuerdo con una cualquiera de las reivindicaciones 1 a 8 o 10 a 13.

45

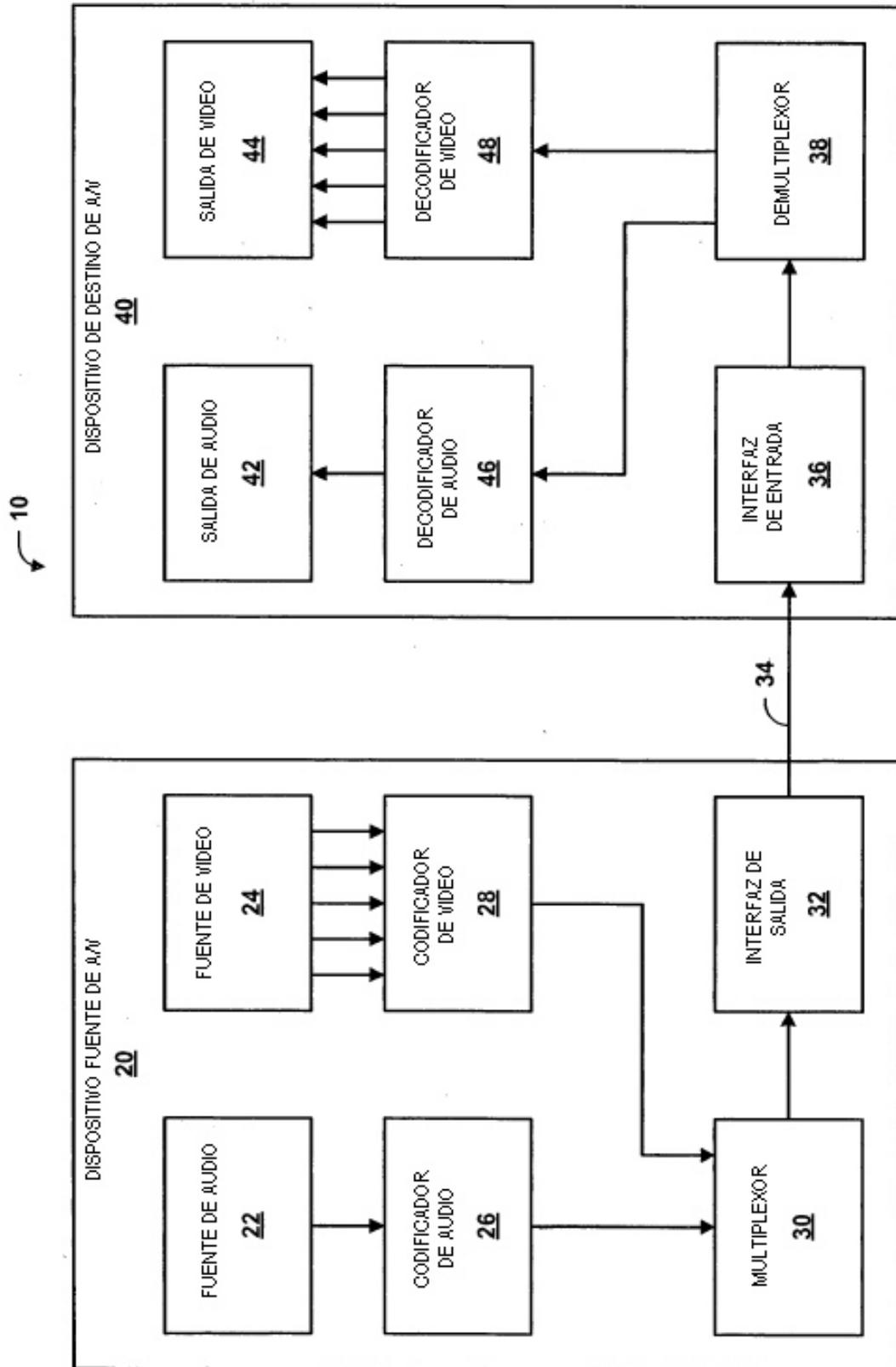


FIG. 1

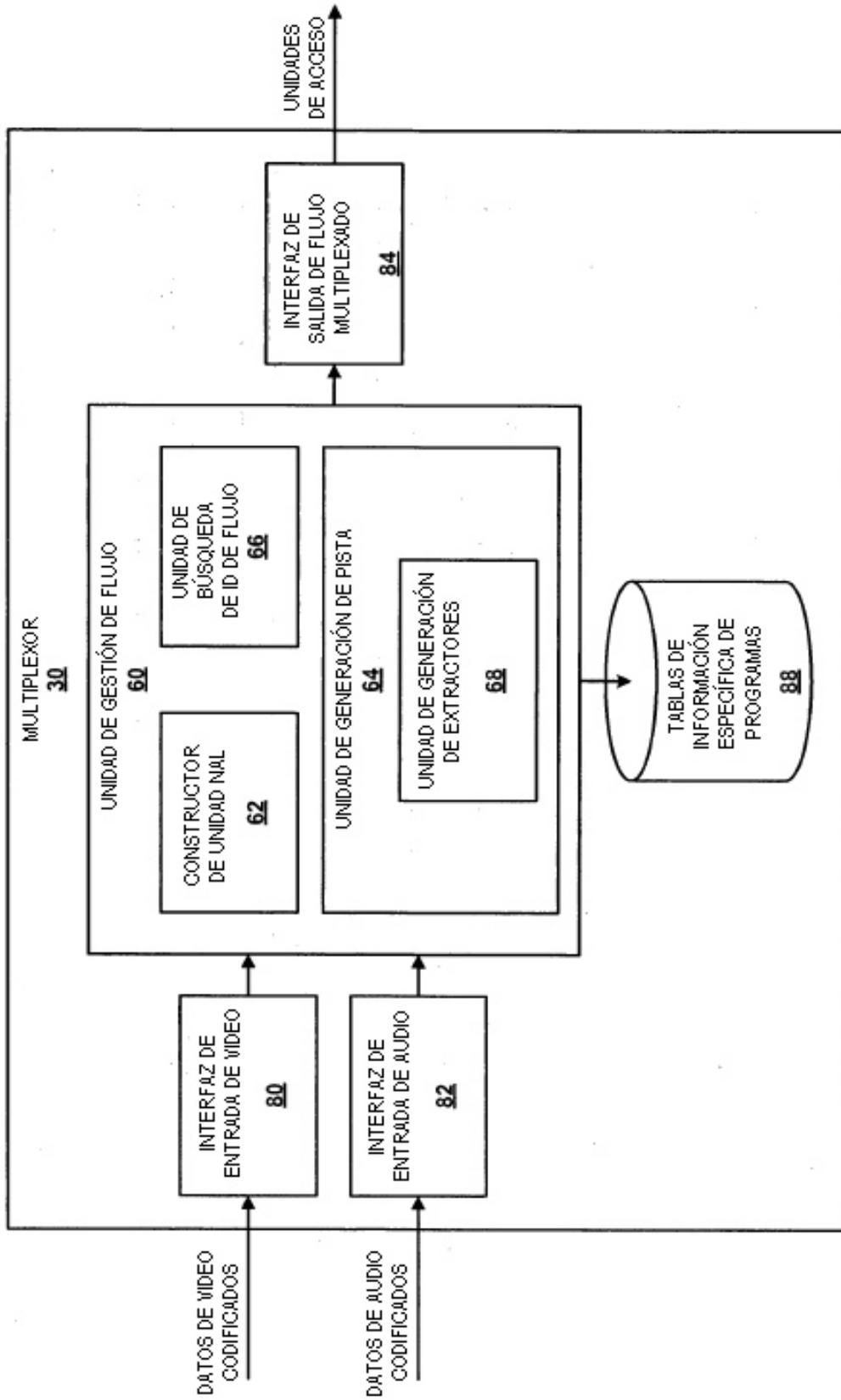


FIG. 2

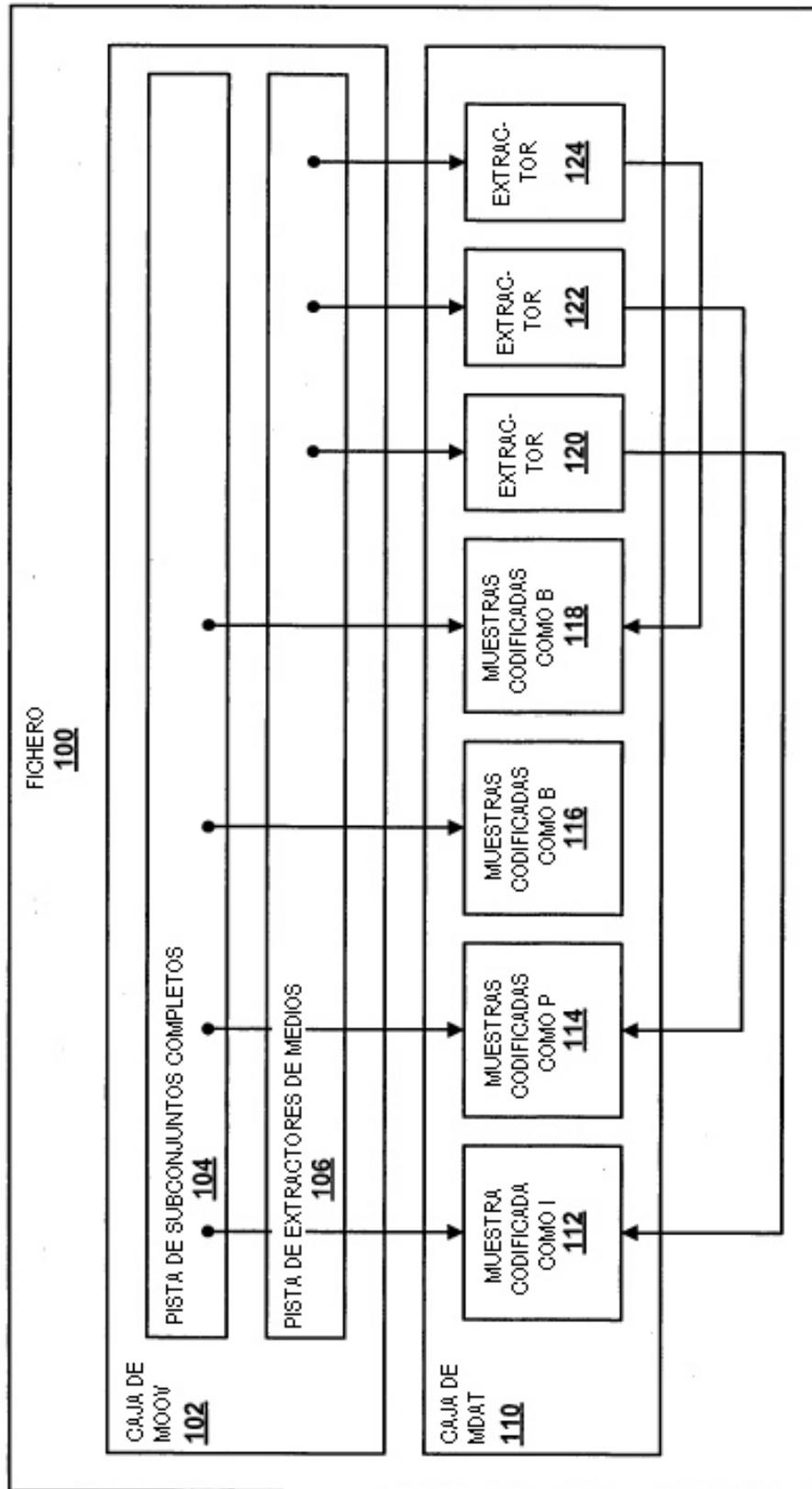


FIG. 3

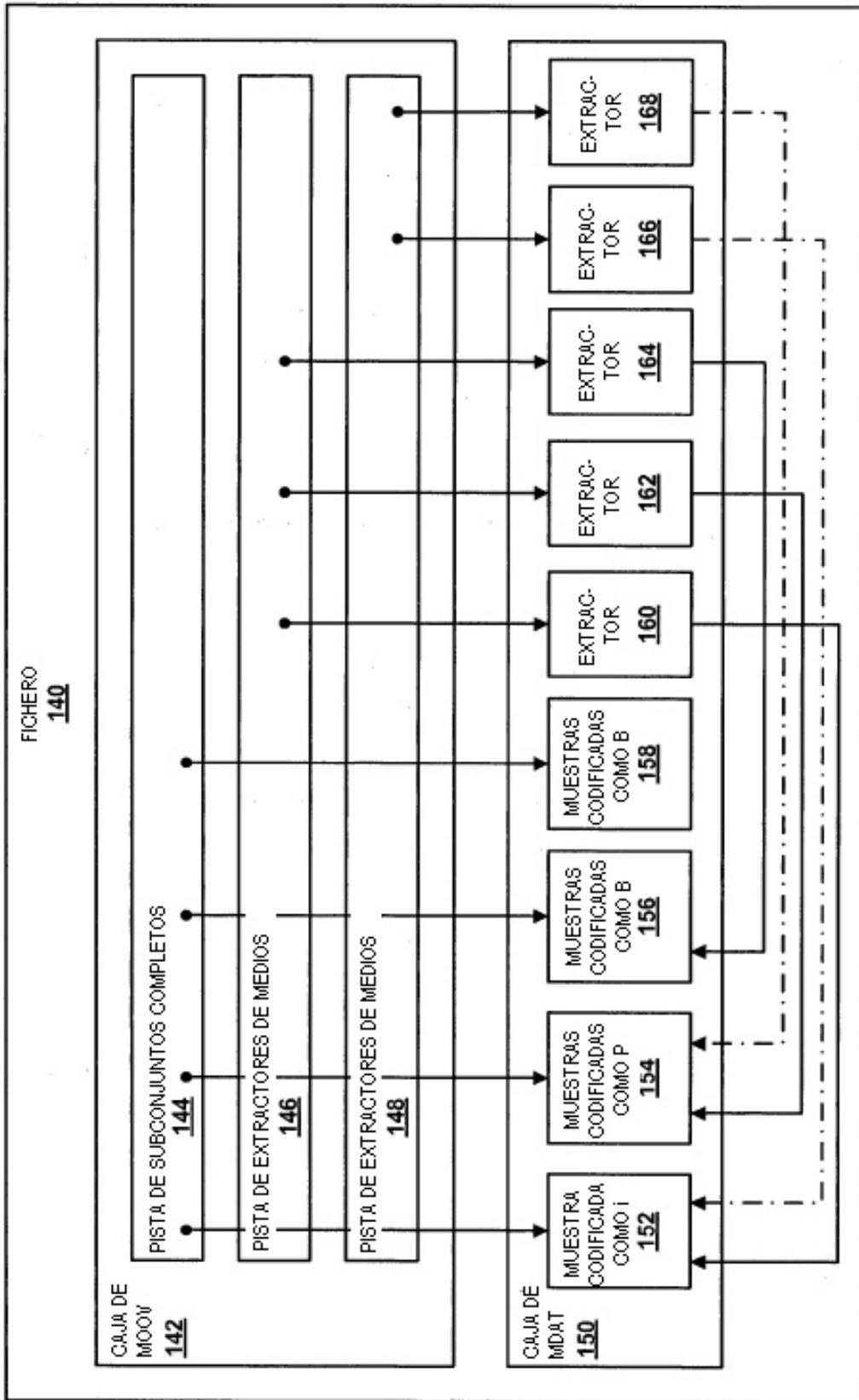


FIG. 4

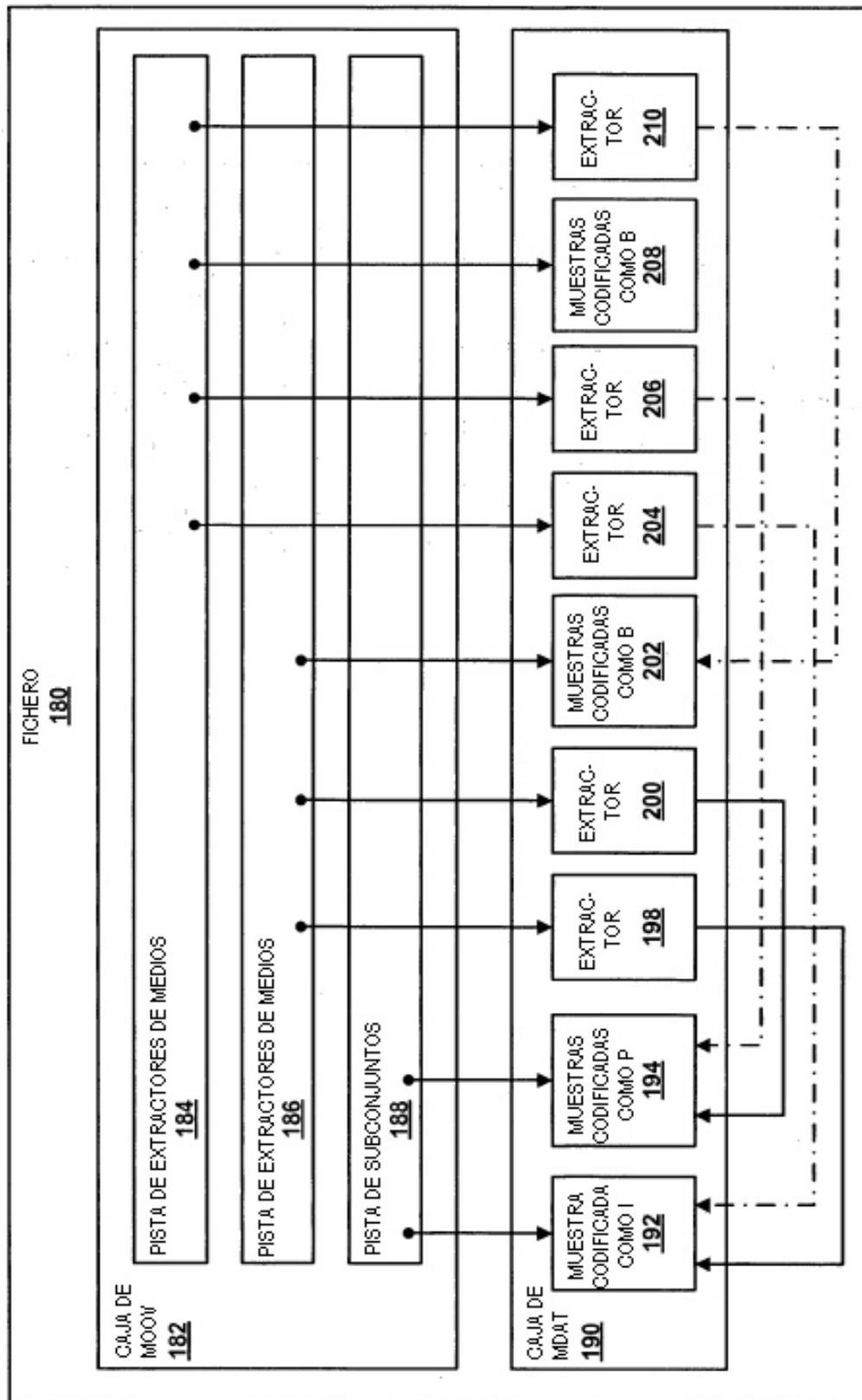


FIG. 5

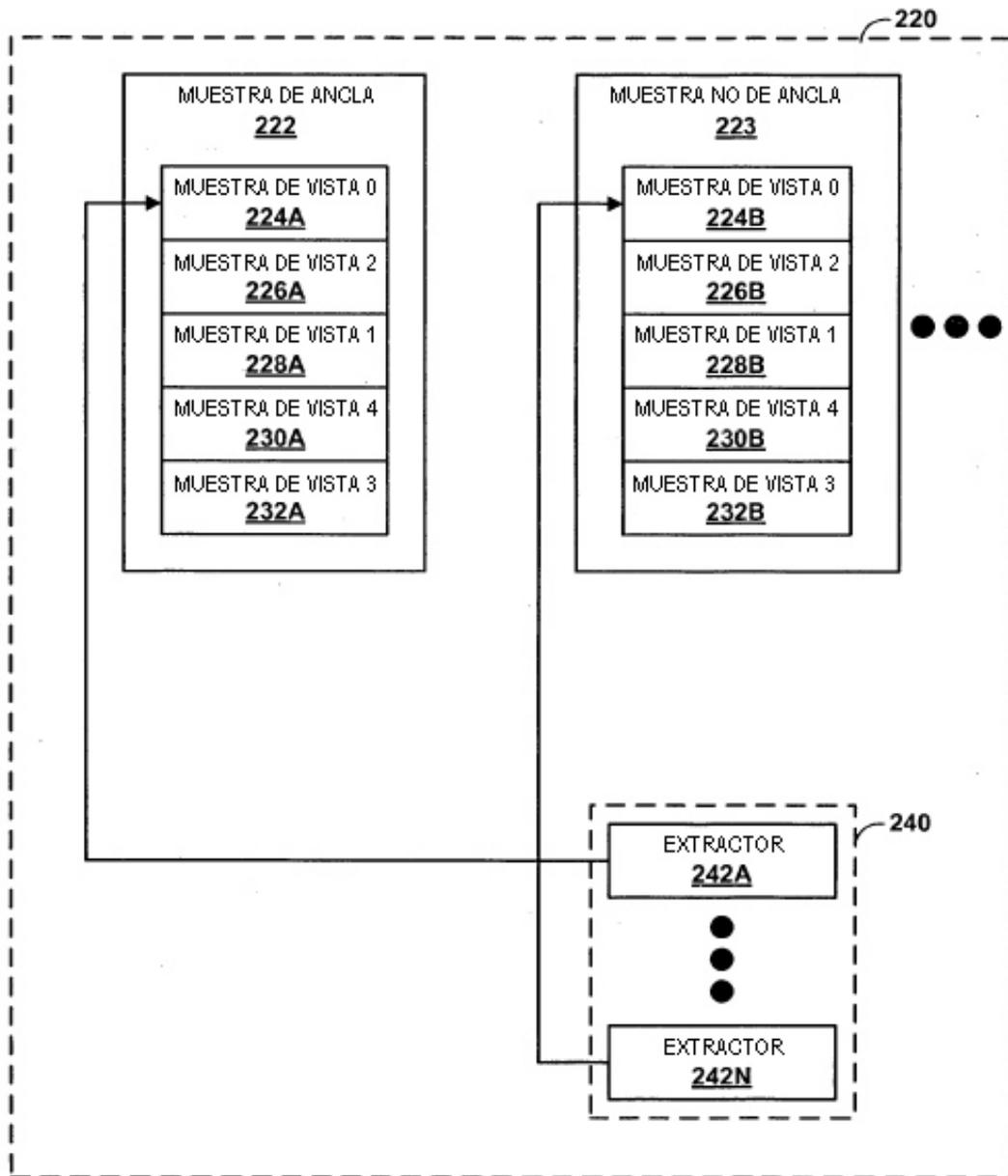


FIG. 6A

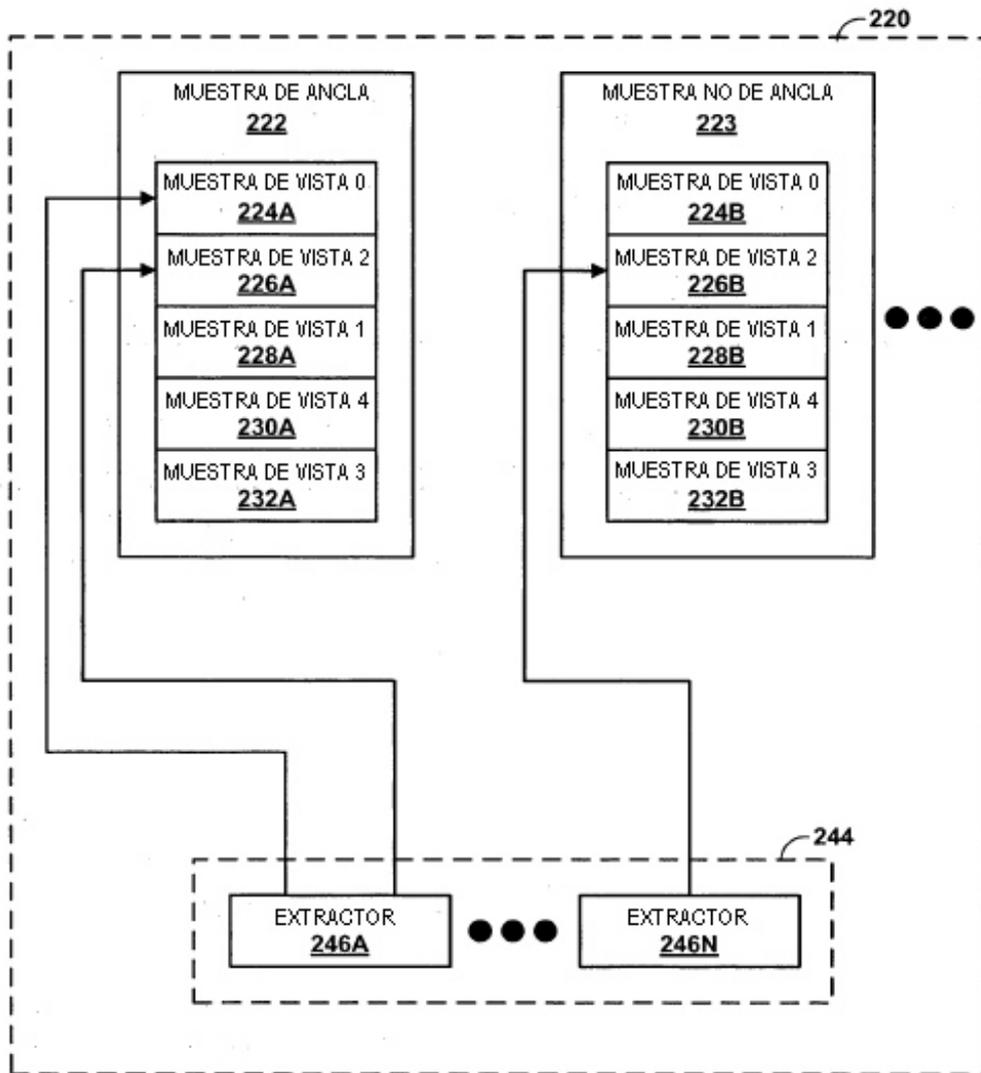


FIG. 6B

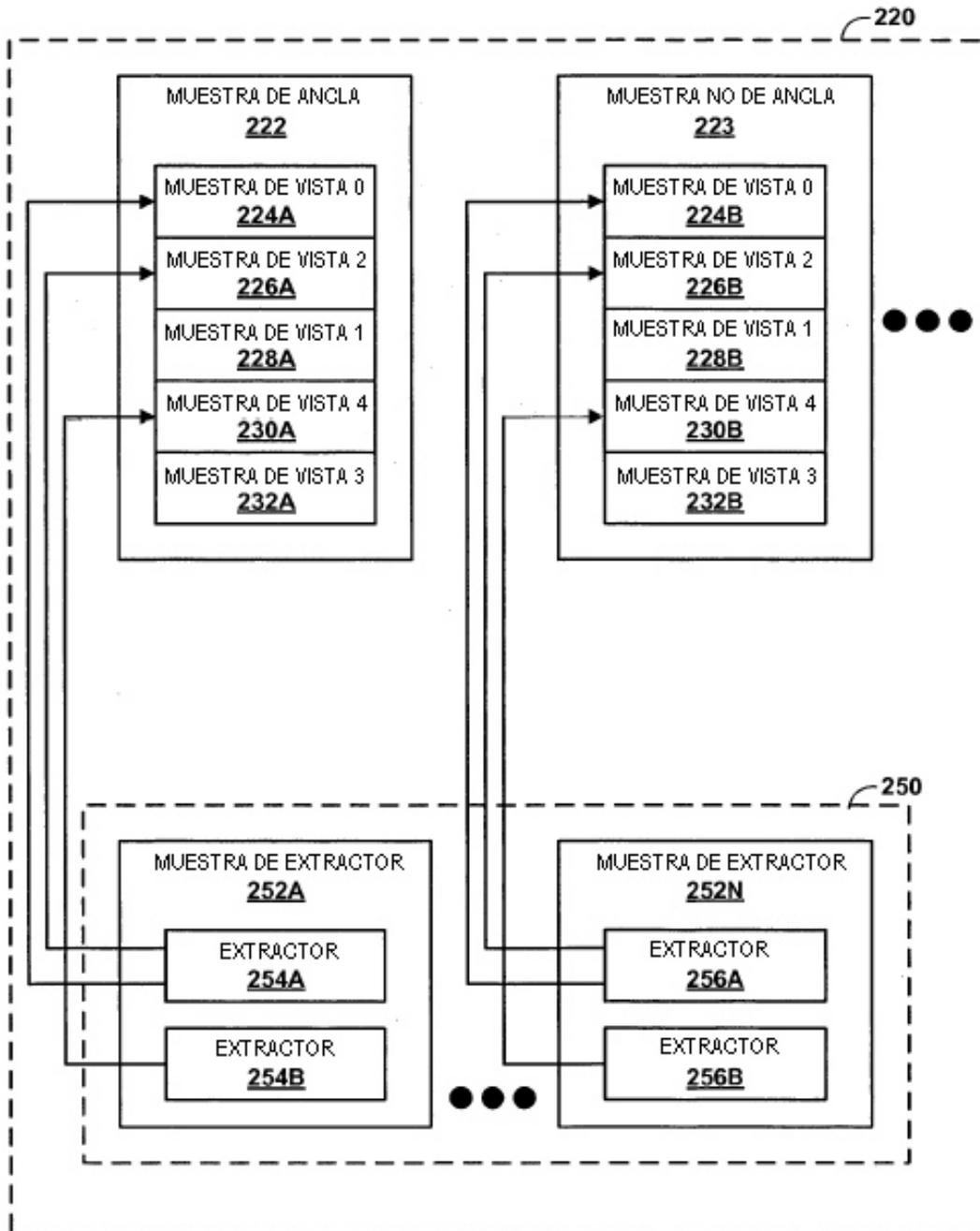


FIG. 6C

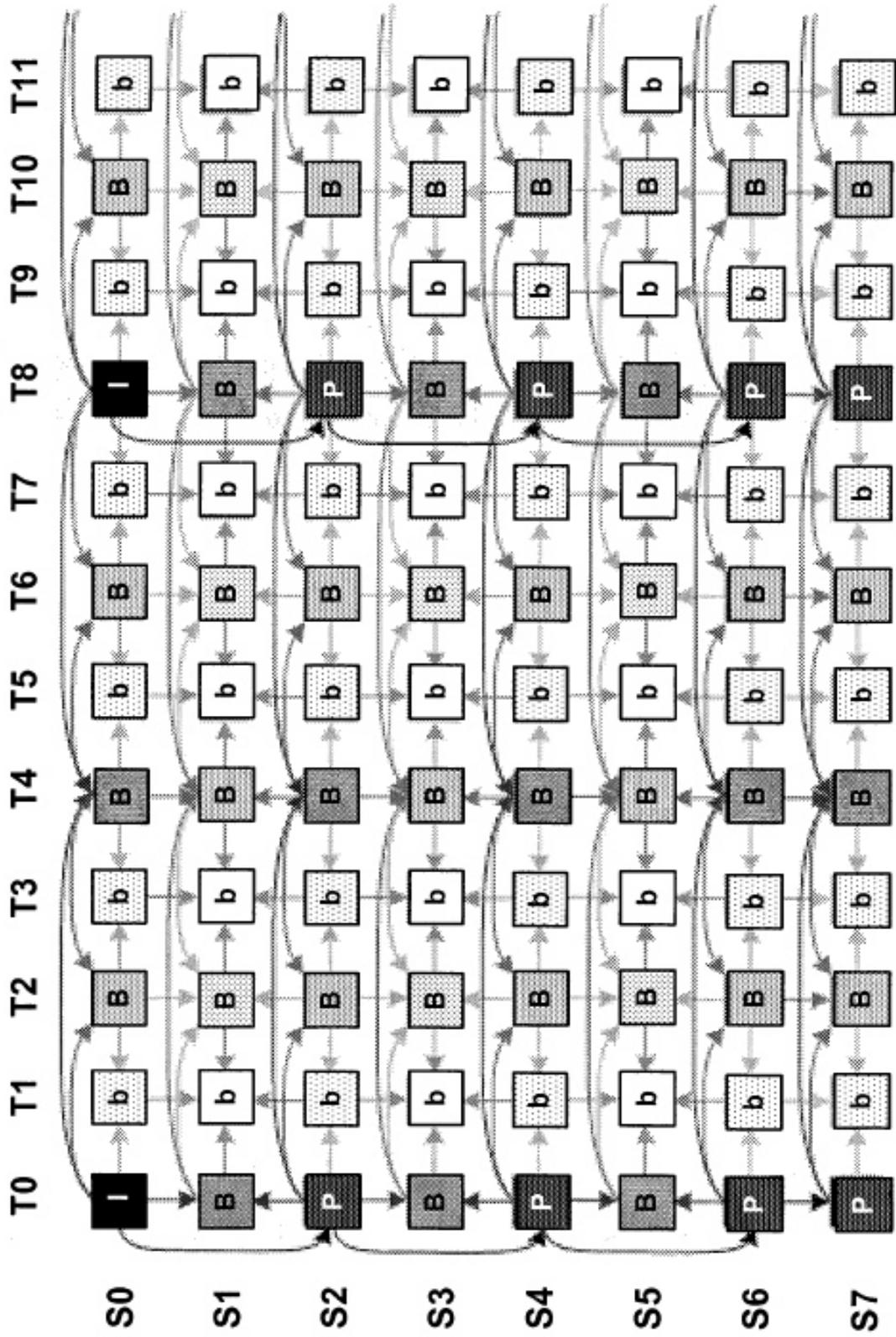


FIG. 7

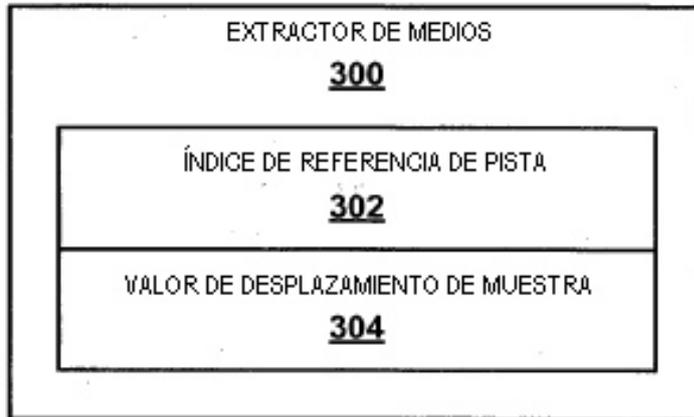


FIG. 8

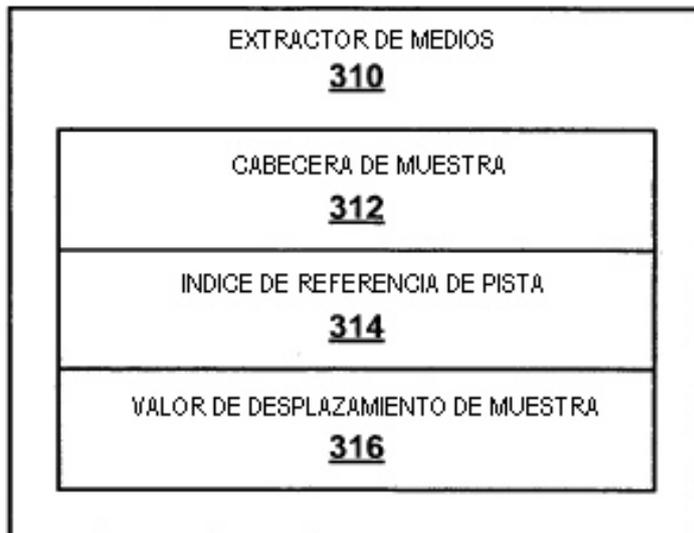


FIG. 9

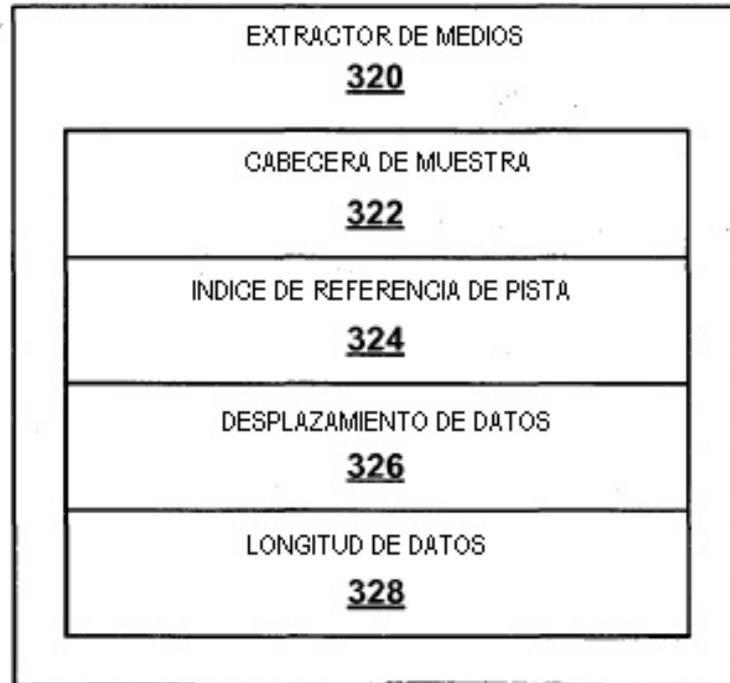


FIG. 10

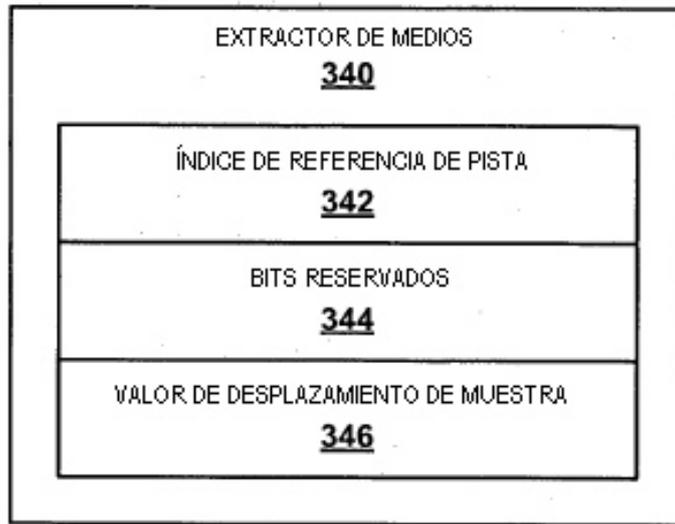


FIG. 11

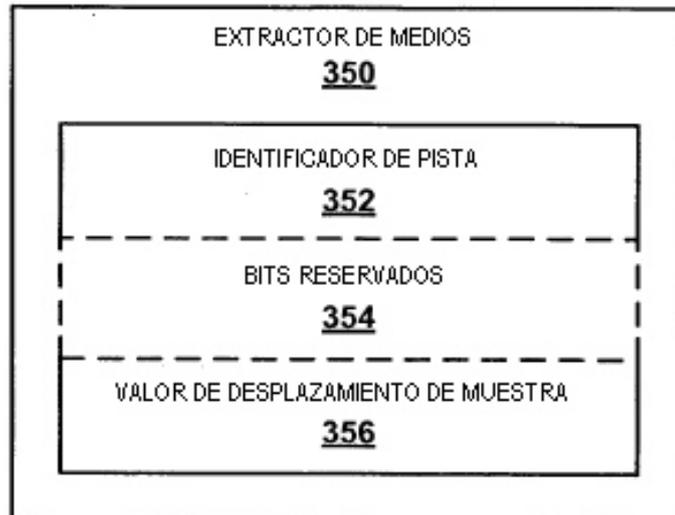


FIG. 12

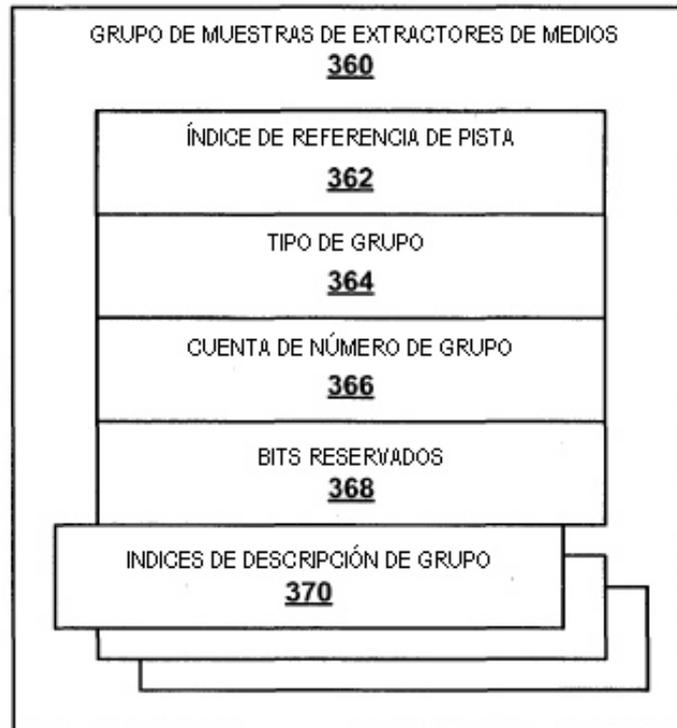


FIG. 13

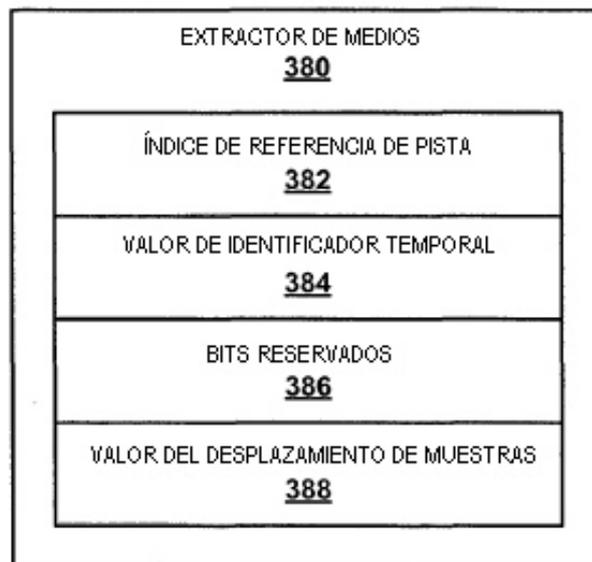


FIG. 14

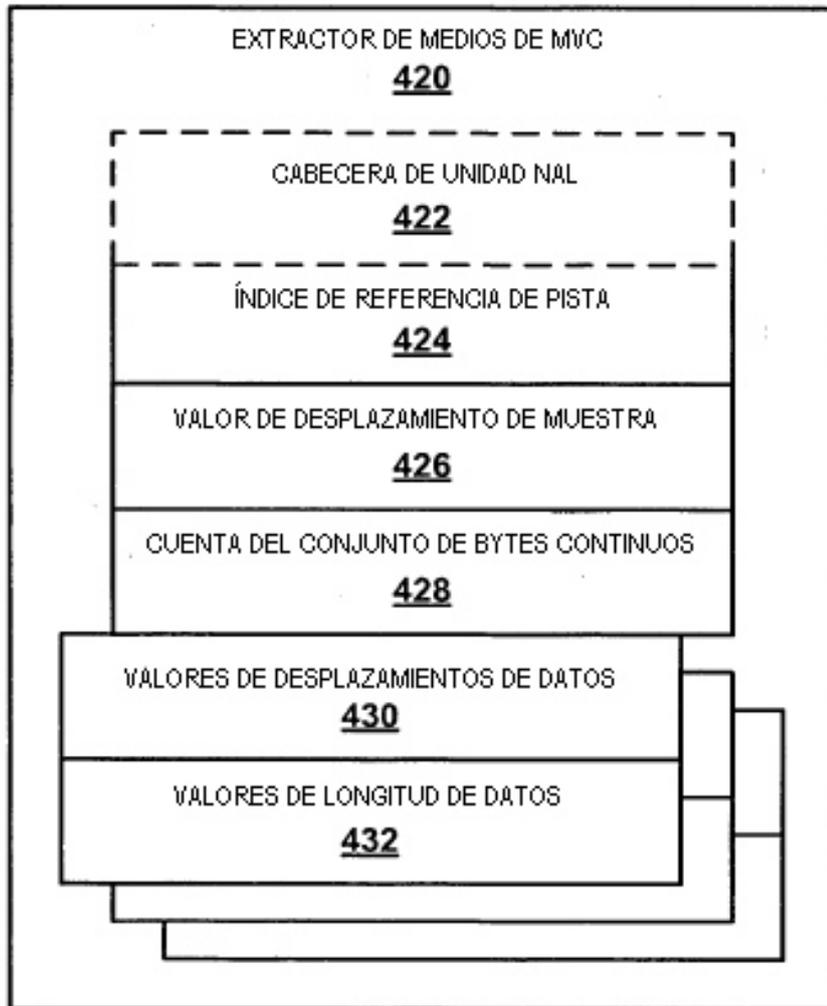


FIG. 15

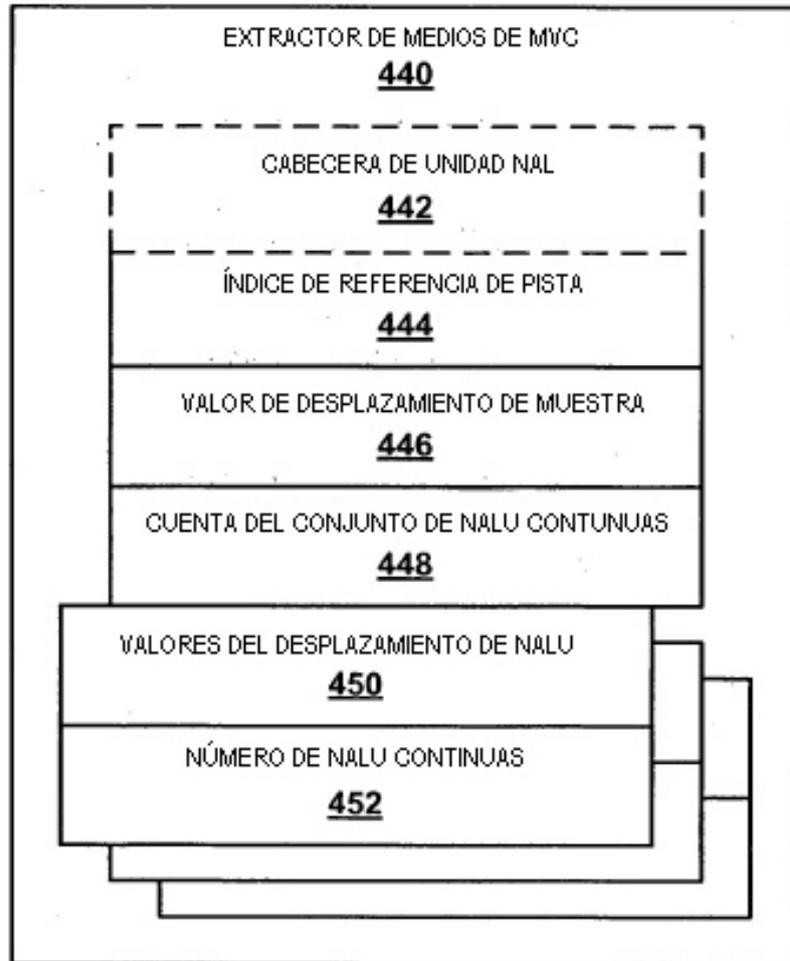


FIG. 16

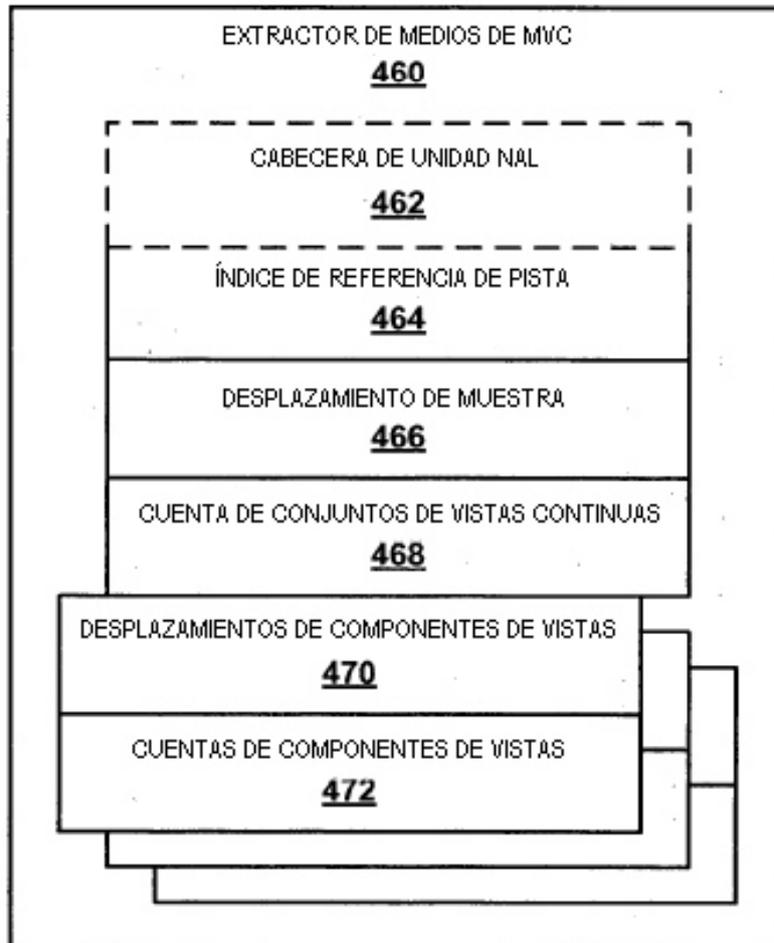


FIG. 17

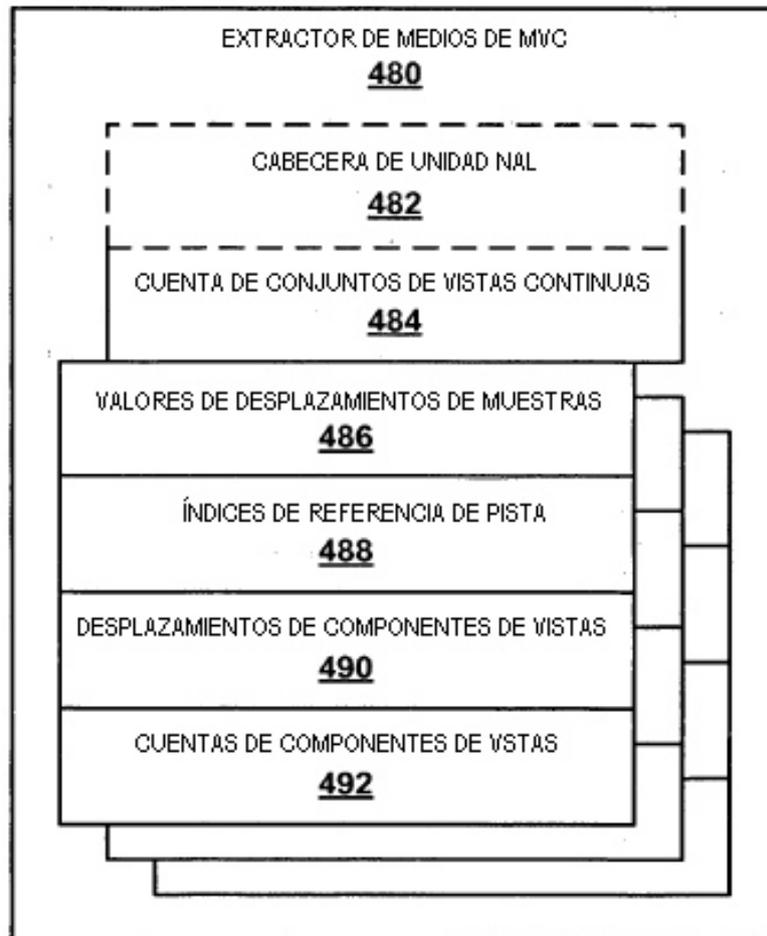


FIG. 18

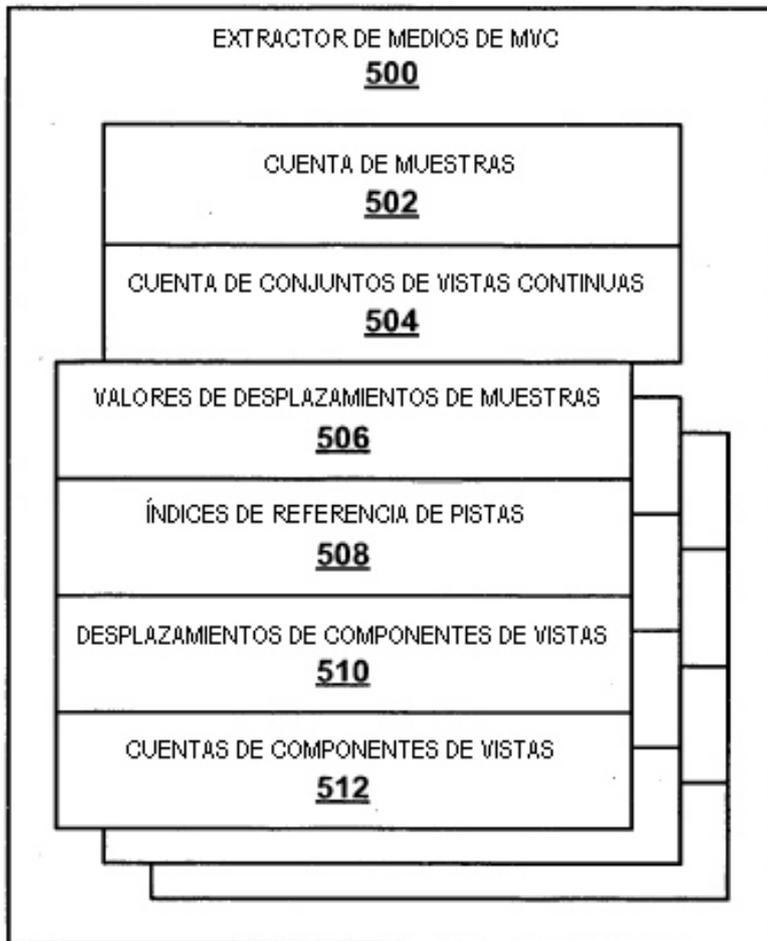


FIG. 19

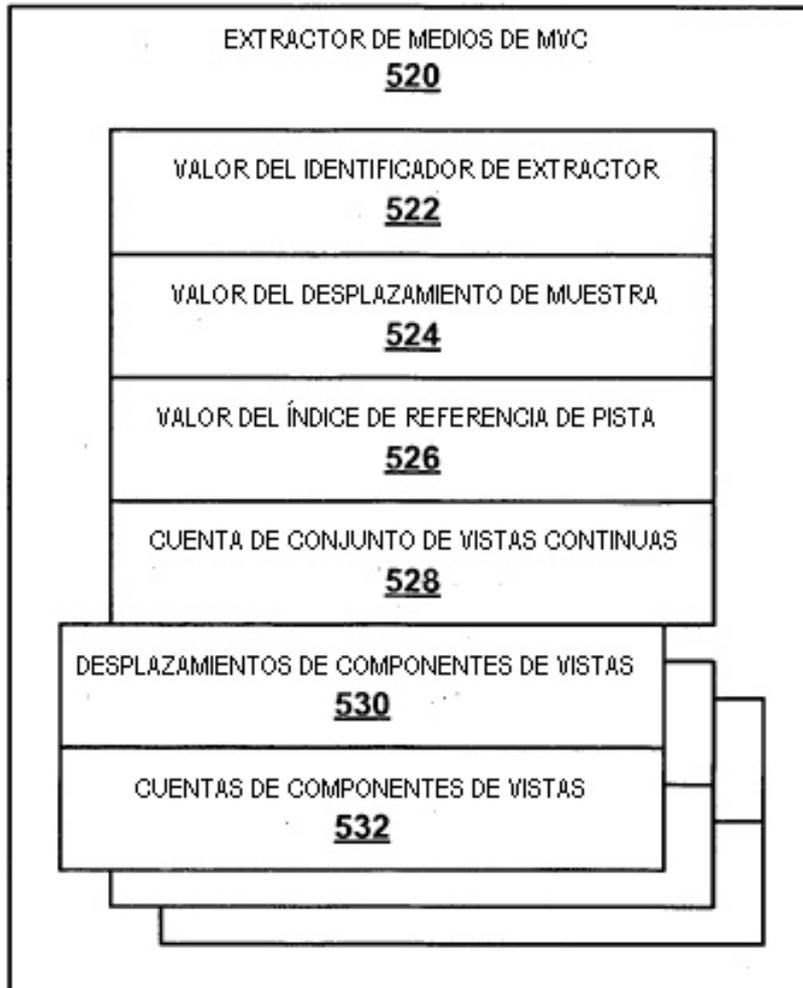


FIG. 20

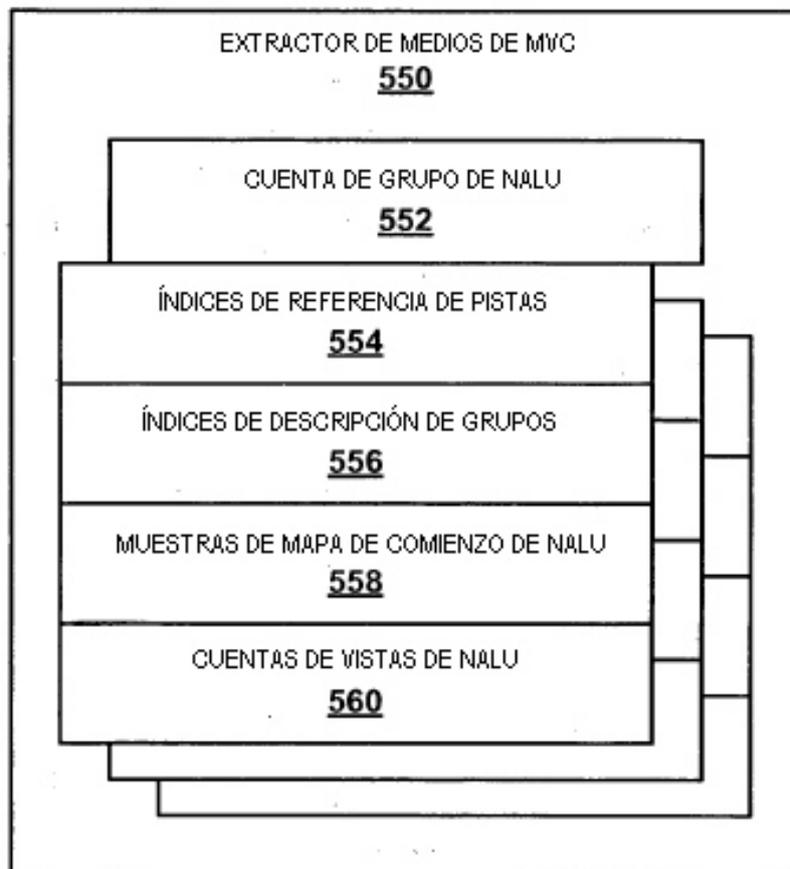


FIG. 21

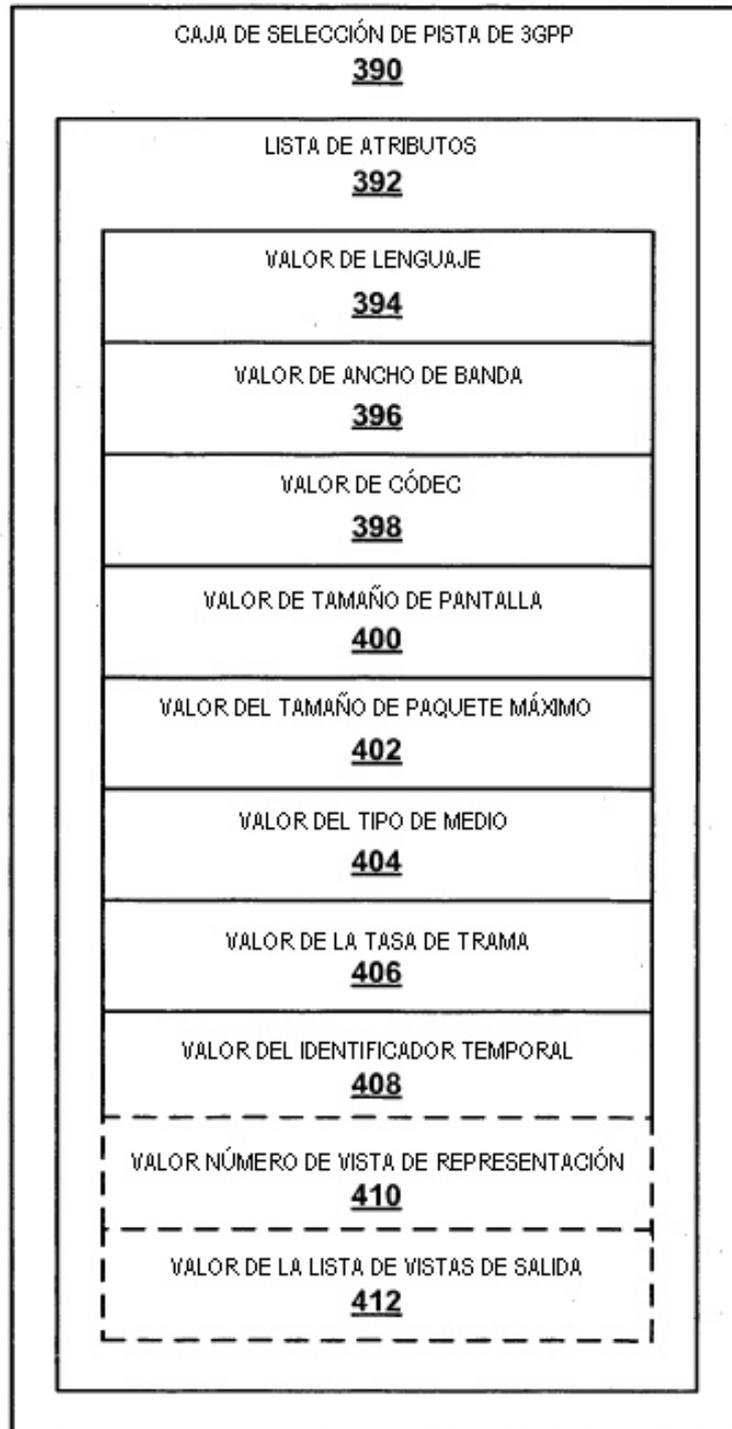


FIG. 22

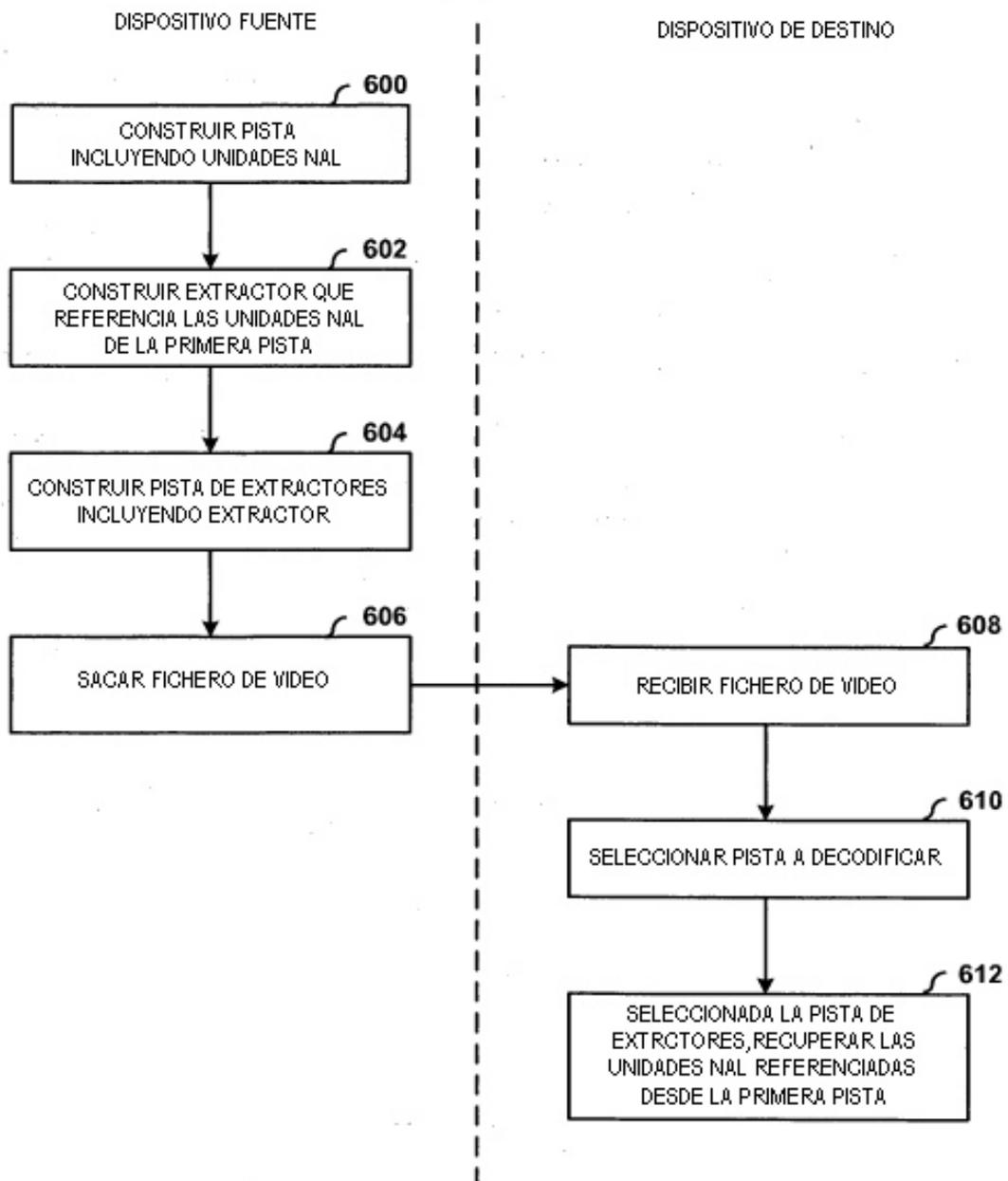


FIG. 23