

19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 927 630**

51 Int. Cl.:

H04N 21/00

(2011.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

96 Fecha de presentación y número de la solicitud europea: **08.12.2004** **E 19214057 (2)**

97 Fecha y número de publicación de la concesión europea: **15.06.2022** **EP 3641317**

54 Título: **Gestión de derechos digitales**

30 Prioridad:

08.12.2003 US 731809

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

08.11.2022

73 Titular/es:

DIVX, LLC (100.0%)
4350 La Jolla Village Drive, Suite 950
San Diego, CA 92122, US

72 Inventor/es:

BRANESS, JASON

74 Agente/Representante:

VALLEJO LÓPEZ, Juan Pedro

ES 2 927 630 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

Gestión de derechos digitales

5 Antecedentes de la invención

La presente invención se refiere, en general, al cifrado, la codificación, la transmisión y la decodificación de archivos multimedia. Más concretamente, la invención se refiere al cifrado, la codificación, la transmisión y la decodificación de archivos multimedia que pueden incluir pistas además de una única pista de audio y una única pista de vídeo.

El desarrollo de Internet ha impulsado el desarrollo de formatos de archivo para información multimedia para habilitar una generación, distribución y visualización normalizada de los archivos. Por lo general, un único archivo multimedia incluye una única pista de vídeo y una única pista de audio. Cuando se escriben datos multimedia en un medio físicamente transportable y de alto volumen, tal como un Disco de Vídeo Digital (DVD), se pueden usar múltiples archivos para proporcionar un número de pistas de vídeo, pistas de audio y pistas de subtítulo. Se pueden proporcionar archivos adicionales que contienen información que se puede usar para generar un menú interactivo.

El documento A-2003/188182(SATO ET AL), con fecha de 02-10-2003, describe una estructura de datos de un formato de archivo multimedia, un método de cifrado y un método de descifrado.

El documento "OpenOML AVI File Format Extensions", con fecha de 1 de septiembre de 1997, XP002179318, recuperado de internet del URL <http://www.rahul.net/jfm/odmlff2.pdf> el 04-10-2001, describe un formato de archivo compatible con AVI para vídeo.

25 Sumario de la invención

La invención se dirige a un archivo multimedia como se define en la reivindicación 1.

Breve descripción de los dibujos

La figura 1. es un diagrama de un sistema de acuerdo con una realización de la presente invención para codificar, distribuir y decodificar archivos.

La figura 2.0. es un diagrama de la estructura de un archivo multimedia de acuerdo con una realización de la presente invención.

La figura 2.0.1. es un diagrama de la estructura de un archivo multimedia de acuerdo con otra realización de la presente invención.

La figura 2.1. es un diagrama conceptual de un fragmento de lista de 'hdl' de acuerdo con una realización de la invención.

La figura 2.2. es un diagrama conceptual de un fragmento de 'strl' de acuerdo con una realización de la invención.

La figura 2.3. es un diagrama conceptual de la memoria asignada para almacenar un fragmento de 'DXDT' de un archivo multimedia de acuerdo con una realización de la invención.

La figura 2.3.1. es un diagrama conceptual de fragmentos de 'metadatos' que se pueden incluir en un fragmento de 'DXDT' de un archivo multimedia de acuerdo con una realización de la invención.

La figura 2.4. es un diagrama conceptual del fragmento de 'DMNU' de acuerdo con una realización de la invención.

La figura 2.5. es un diagrama conceptual de fragmentos de menú contenidos en un fragmento de 'GestorMenuWow' de acuerdo con una realización de la invención.

La figura 2.6. es un diagrama conceptual de fragmentos de menú contenidos dentro de un fragmento de 'GestorMenuWow' de acuerdo con otra realización de la invención.

La figura 2.6.1. es un diagrama conceptual que ilustra las relaciones entre los diversos fragmentos contenidos dentro de un fragmento de 'DMNU'.

La figura 2.7. es un diagrama conceptual del fragmento de lista de 'película' de un archivo multimedia de acuerdo con una realización de la invención.

La figura 2.8. es un diagrama conceptual del fragmento de lista de 'película' de un archivo multimedia de acuerdo con una realización de la invención que incluye una DRM.

La figura 2.9. es un diagrama conceptual del fragmento de 'DRM' de acuerdo con una realización de la invención.

La figura 3.0. es un diagrama de bloques de un sistema para generar un archivo multimedia de acuerdo con una realización de la presente invención.

La figura 3.1. es un diagrama de bloques de un sistema para generar un fragmento de 'DXDT' de acuerdo con una realización de la invención.

La figura 3.2. es un diagrama de bloques de un sistema para generar un fragmento de 'DMNU' de acuerdo con una realización de la invención.

La figura 3.3. es un diagrama conceptual de un modelo de medios de acuerdo con una realización de la invención.

La figura 3.3.1. es un diagrama conceptual de objetos a partir de un modelo de medios que se puede usar para generar automáticamente un menú pequeño de acuerdo con una realización de la invención.

La figura 3.4. es un diagrama de flujo de un proceso que se puede usar para refragmentar audio de acuerdo con una realización de la invención.

La figura 3.5. es un diagrama de bloques de un codificador de vídeo de acuerdo con una realización de la presente.
La figura 3.6. es un diagrama de flujo de un método de realización de una potenciación psicovisual de suavidad en una trama I de acuerdo con algunas realizaciones de la invención.

La figura 3.7. es un diagrama de flujo de un proceso para realizar una potenciación psicovisual de SAD de macrobloque de acuerdo con una realización de la invención.

La figura 3.8. es un diagrama de flujo de un proceso para un control de tasa de una pasada de acuerdo con una realización de la invención.

La figura 3.9. es un diagrama de flujo de un proceso para realizar un control de tasa de VBV de N-ésima pasada de acuerdo con una realización de la invención.

La figura 4.0. es un diagrama de flujo para un proceso para localizar la información multimedia requerida a partir de un archivo multimedia y visualizar la información multimedia de acuerdo con una realización de la invención.

La figura 4.1. es un diagrama de bloques de un descodificador de acuerdo con una realización de la invención.

La figura 4.2. es un ejemplo de un menú visualizado de acuerdo con una realización de la invención.

La figura 4.3. es un diagrama conceptual que muestra las fuentes de información usadas para generar la visualización mostrada en la figura 4.2 de acuerdo con una realización de la invención.

Descripción detallada de la invención

Haciendo referencia a los dibujos, algunas realizaciones de la presente invención son capaces de codificar, transmitir y descodificar archivos multimedia. Los archivos multimedia de acuerdo con algunas realizaciones de la presente invención pueden contener múltiples pistas de vídeo, múltiples pistas de audio, múltiples pistas de subtítulo, datos que se pueden usar para generar una interfaz de menú para acceder a los contenidos del archivo y 'metadatos' con respecto a los contenidos del archivo. Los archivos multimedia de acuerdo con varias realizaciones de la presente invención también incluyen referencias a pistas de vídeo, pistas de audio, pistas de subtítulo y 'metadatos' externos al archivo.

1. DESCRIPCIÓN DEL SISTEMA

Pasando a continuación a la figura 1, se muestra un sistema de acuerdo con una realización de la presente invención para codificar, distribuir y descodificar archivos. El sistema 10 incluye un ordenador 12, que se conecta a una otros diversos dispositivos informáticos por medio de una red 14. Los dispositivos que se pueden conectar a la red incluyen un servidor 16, un ordenador portátil 18 y un asistente personal digital (PDA) 20. En diversas realizaciones, las conexiones entre los dispositivos y la red pueden ser o bien cableadas o bien inalámbricas e implementarse usando cualquiera de diversos protocolos de interconexión de redes.

Durante el funcionamiento, el ordenador 12 se puede usar para codificar archivos multimedia de acuerdo con una realización de la presente invención. El ordenador 12 también se puede usar para descodificar archivos multimedia de acuerdo con algunas realizaciones de la presente invención y distribuir archivos multimedia de acuerdo con algunas realizaciones de la presente invención. El ordenador puede distribuir archivos usando cualquiera de diversos protocolos de transferencia de archivos, incluyendo por medio de una red entre elementos del mismo nivel. Además, el ordenador 12 puede transferir archivos multimedia de acuerdo con algunas realizaciones de la presente invención a un servidor 18, en donde se puede acceder a los archivos mediante otros dispositivos. Los otros dispositivos pueden incluir cualquier diversidad de dispositivos informáticos o incluso un dispositivo de descodificación dedicado. En la realización ilustrada, se muestran un ordenador portátil y un PDA. En otras realizaciones, descodificadores digitales de salón, ordenadores de escritorio, videoconsolas, dispositivos de electrónica de consumo y otros dispositivos se pueden conectar a la red, descargar los archivos multimedia y descodificarlos.

En una realización, los dispositivos acceden a los archivos multimedia a partir del servidor por medio de la red. En otras realizaciones, los dispositivos acceden a los archivos multimedia desde un número de ordenadores por medio de una red entre elementos del mismo nivel. En varias realizaciones, se pueden escribir archivos multimedia en un dispositivo de almacenamiento portátil tal como una unidad de disco, CD-ROM o DVD. En muchas realizaciones, los dispositivos electrónicos pueden acceder a los archivos multimedia escritos en dispositivos de almacenamiento portátiles.

2. DESCRIPCIÓN DE ESTRUCTURA DE ARCHIVO

Los archivos multimedia de acuerdo con algunas realizaciones de la presente invención se pueden estructurar para ser conformes con el Formato de Archivo de Intercambio de Recursos ('formato de archivo RIFF'), definido por Microsoft Corporation de Redmond, WA e International Business Machines Corporation de Armonk, NY. RIFF es un formato de archivo para almacenar datos multimedia e información asociada. Por lo general, un archivo RIFF tiene un encabezamiento de RIFF de 8 bytes, que identifica el archivo y proporciona la longitud residual del archivo después del encabezamiento (es decir, longitud archivo - 8). La totalidad del resto del archivo RIFF comprende "fragmentos" y "listas". Cada fragmento tiene un encabezamiento de fragmento de 8 bytes que identifica el tipo de fragmento, y que da la longitud en bytes de los datos a continuación del encabezamiento de fragmento. Cada lista tiene un encabezamiento de lista de 8 bytes que identifica el tipo de lista y que da la longitud en bytes de los datos a continuación del encabezamiento de lista. Los datos en una lista comprenden fragmentos y/u otras listas (que, a su

vez pueden comprender fragmentos y/u otras listas). A veces, también se hace referencia a las listas RIFF como "fragmentos de lista".

Un archivo AVI es una forma especial de archivo RIFF que sigue el formato de un archivo RIFF, pero incluye diversos fragmentos y listas con identificadores definidos que contienen datos multimedia en formatos particulares. El formato AVI fue desarrollado y definido por Microsoft Corporation. Por lo general, los archivos AVI se crean usando un codificador que puede emitir datos multimedia en el formato AVI. Por lo general, los archivos AVI son descodificados por cualquiera de un grupo de software conocido de forma colectiva como descodificadores de AVI.

Los formatos RIFF y AVI son flexibles en que estos solo definen fragmentos y listas que son parte del formato de archivo definido, pero permiten que los archivos también incluyan listas y/o fragmentos que están fuera de las definiciones de formato de archivo RIFF y/o AVI sin volver el archivo ilegible por un descodificador de RIFF y/o de AVI. En la práctica, los descodificadores de AVI (y, de forma similar, de RIFF) se implementan de tal modo que estos simplemente ignoran las listas y fragmentos que contienen información de encabezamiento no hallada en la definición de formato de archivo AVI. El descodificador de AVI ha de seguir leyendo a través de estas listas y fragmentos que no son AVI y, por lo tanto, el funcionamiento del descodificador de AVI se puede ralentizar pero, por lo demás, estos no tienen, en general, efecto alguno sobre, y son ignorados por, un descodificador de AVI.

Un archivo multimedia de acuerdo con una realización de la presente invención se ilustra en la figura 2.0. El archivo multimedia 30 ilustrado incluye un fragmento de conjunto de caracteres (fragmento de 'CSET') 32, un fragmento de lista de información (fragmento de lista de 'INFO') 34, un fragmento de encabezamiento de archivo (fragmento de lista de 'hdlr') 36, un fragmento de metadatos (fragmento de 'DXDT') 38, un fragmento de menú (fragmento de 'DMNU') 40, un fragmento de basura (fragmento de 'basura') 41, el fragmento de lista de película (fragmento de lista de 'película') 42, un fragmento de índice opcional (fragmento de 'idxl') 44 y un segundo fragmento de menú (fragmento de 'DMNU') 46. Algunos de estos fragmentos, y porciones de otros, se definen en el formato de archivo AVI, mientras que otros no están contenidos en el formato de archivo AVI. En muchos casos, pero no en todos ellos, el análisis posterior identifica fragmentos o porciones de fragmentos que se definen como parte del formato de archivo AVI.

Otro archivo multimedia de acuerdo con una realización de la presente invención se muestra en la figura 2.0.1. El archivo multimedia 30' es similar al mostrado en la figura 2.0. excepto por que el archivo incluye múltiples fragmentos 'RIFF' concatenados. Los fragmentos 'RIFF' pueden contener un fragmento 'RIFF' similar al mostrado en la figura 2.0. que puede excluir el segundo fragmento de 'DMNU' 46 o puede contener información de menú en la forma de un fragmento de 'DMNU' 46'.

En la realización ilustrada, los datos multimedia incluyen múltiples fragmentos 'RIFF' concatenados, en donde el primer fragmento 'RIFF' 50 incluye un fragmento de conjunto de caracteres (fragmento de 'CSET') 32, un fragmento de lista de información (fragmento de lista de 'INFO') 34, un fragmento de encabezamiento de archivo (fragmento de lista de 'hdlr') 36, un fragmento de metadatos (fragmento de 'DXDT') 38, un fragmento de menú (fragmento de 'DMNU') 40, un fragmento de basura (fragmento de 'basura') 41, el fragmento de lista de película (fragmento de lista de 'película') 42 y un fragmento de índice opcional (fragmento de 'idxl') 44. El segundo fragmento 'RIFF' 52 contiene un segundo fragmento de menú (fragmento de 'DMNU') 46'. Se pueden incluir fragmentos 'RIFF' adicionales 54, que contienen títulos adicionales, después del fragmento de menú 'RIFF' 52. Los fragmentos 'RIFF' adicionales pueden contener medios independientes en un formato de archivo conforme con AVI. En una realización, el segundo fragmento de menú 46' y los fragmentos 'RIFF' adicionales tienen unos códigos de 4 caracteres especializados (definidos en el formato AVI y analizados posteriormente) de tal modo que los primeros dos caracteres de los códigos de 4 caracteres aparecen como los segundos dos caracteres y los segundos dos caracteres de los códigos de 4 caracteres aparecen como los primeros dos caracteres.

2.1. El fragmento de 'CSET'

El fragmento de 'CSET' 32 es un fragmento definido en el formato de archivo de Intercalación de Audio y Vídeo (formato de archivo AVI), creado por Microsoft Corporation. El fragmento de 'CSET' define el conjunto de caracteres y la información de idioma del archivo multimedia. La inclusión de un fragmento de 'CSET' de acuerdo con algunas realizaciones de la presente invención es opcional.

Un archivo multimedia de acuerdo con una realización de la presente invención no usa el fragmento de 'CSET' y usa UTF-8, que es definido por el Consorcio Unicode, para el conjunto de caracteres por defecto combinado con la Especificación de Idiomas 3066 de RFC, que es definida por el Grupo de Tareas Especiales sobre Ingeniería de Internet para la información de idioma.

2.2. El fragmento de lista de 'INFO'

El fragmento de lista de 'INFO' 34 puede almacenar información que ayuda a identificar los contenidos del archivo multimedia. La lista de 'INFO' se define en el formato de archivo AVI y su inclusión en un archivo multimedia de acuerdo con algunas realizaciones de la presente invención es opcional. Muchas realizaciones que incluyen un fragmento de 'DXDT' no incluyen un fragmento de lista de 'INFO'.

2.3. El fragmento de lista de 'hdrl'

El fragmento de lista de 'hdrl' 38 se define en el formato de archivo AVI y proporciona información con respecto al formato de los datos en el archivo multimedia. En general, se requiere la inclusión de un fragmento de lista de 'hdrl' o un fragmento que contiene una información de descripción similar. El fragmento de lista de 'hdrl' incluye un fragmento para cada pista de vídeo, cada pista de audio y cada pista de subtítulo.

Un diagrama conceptual de un fragmento de lista de 'hdrl' 38 de acuerdo con una realización de la invención que incluye una única pista de vídeo 62, dos pistas de audio 64, una pista de audio externa 66, dos pistas de subtítulo 68 y una pista de subtítulo externa 70 se ilustra en la figura 2.1. La lista de 'hdrl' 60 incluye un fragmento de 'avih'. El fragmento de 'avih' 60 contiene información global para la totalidad del archivo, tal como el número de secuencias dentro del archivo y la anchura y la altura del vídeo contenido en el archivo multimedia. El fragmento de 'avih' se puede implementar de acuerdo con el formato de archivo AVI.

Además del fragmento de 'avih', la lista de 'hdrl' incluye una lista de descriptores de secuencia para cada pista de audio, de vídeo y de subtítulo. En una realización, la lista de descriptores de secuencia se implementa usando fragmentos de 'strl'. Un fragmento de 'strl' de acuerdo con una realización de la presente invención se ilustra en la figura 2.2. Cada fragmento de 'strl' sirve para describir cada pista en el archivo multimedia. Los fragmentos de 'strl' para las pistas de audio, de vídeo y de subtítulo dentro del archivo multimedia incluyen un fragmento de 'strl' que hace referencia a un fragmento de 'strh' 92, un fragmento de 'strf' 94, un fragmento de 'strd' 96 y un fragmento de 'strn' 98. La totalidad de estos fragmentos se puede implementar de acuerdo con el formato de archivo AVI. De un interés particular es el fragmento de 'strh' 92, que especifica el tipo de pista de medios, y el fragmento de 'strd' 96, que se puede modificar para indicar si el vídeo está protegido por gestión de derechos digitales. Posteriormente se proporciona un análisis de diversas implementaciones de gestión de derechos digitales de acuerdo con algunas realizaciones de la presente invención.

Los archivos multimedia de acuerdo con algunas realizaciones de la presente invención pueden contener referencias a archivos externos que contienen información multimedia tal como una pista de audio adicional o una pista de subtítulo adicional. Las referencias a estas pistas se pueden contener o bien en el fragmento de 'hdrl' o bien en el fragmento de 'basura' 41. En todo caso, la referencia se puede contener en el fragmento de 'strh' 92 de un fragmento de 'strl' 90, que hace referencia o bien a un archivo local o bien a un archivo almacenado de forma remota. El archivo al que se hace referencia puede ser un archivo AVI convencional o un archivo multimedia de acuerdo con una realización de la presente invención que contiene la pista adicional.

En algunas realizaciones adicionales, el archivo al que se hace referencia puede contener cualquiera de los fragmentos que pueden estar presentes en el archivo de referencia, incluyendo fragmentos de 'DMNU', fragmentos de 'DXDT' y fragmentos asociados con pistas de audio, de vídeo y/o de subtítulo para una presentación multimedia. Por ejemplo, un primer archivo multimedia podría incluir un fragmento de 'DMNU' (analizado con mayor detalle posteriormente) que hace referencia a una primera presentación multimedia situada dentro del fragmento de lista de 'película' del primer archivo multimedia y una segunda presentación multimedia dentro del fragmento de lista de 'película' de un segundo archivo multimedia. Como alternativa, ambos fragmentos de lista de 'película' se pueden incluir en el mismo archivo multimedia, que no es necesario que sea el mismo archivo que el archivo en el que está situado el fragmento de 'DMNU'.

2.4. El fragmento de 'DXDT'

El fragmento de 'DXDT' 38 contiene unos así denominados 'metadatos'. 'Metadatos' es un término usado para describir datos que proporcionan información acerca de los contenidos de un archivo, un documento o una radiodifusión. Los 'metadatos' almacenados dentro del fragmento de 'DXDT' de archivos multimedia de acuerdo con algunas realizaciones de la presente invención se pueden usar para almacenar una información específica del contenido tal como título, autor, propietario de los derechos de autor y reparto. Además, se pueden proporcionar detalles técnicos acerca del códec usado para codificar el archivo multimedia, tales como las opciones de CLI usadas y la distribución de cuantificador después de cada pasada.

En una realización, los metadatos se representan dentro del fragmento de 'DXDT' como una serie de sentencias, en donde cada sentencia incluye un sujeto, un predicado, un objeto y una autoridad. El sujeto es una referencia a lo que se está describiendo. El sujeto puede hacer referencia a un archivo, un artículo, una persona o una organización. El sujeto puede hacer referencia a cualquier cosa que tenga características que se puedan describir. El predicado identifica una característica del sujeto que se está describiendo. El objeto es una descripción de la característica identificada del sujeto y la autoridad identifica la fuente de la información.

Lo siguiente es una tabla que muestra un ejemplo de cómo se pueden representar diversos fragmentos de 'metadatos', como un objeto, un predicado, un sujeto y una autoridad:

Tabla 1 - Representación conceptual de 'metadatos'

Sujeto	Predicado	Objeto	Autoridad
_:archivo281	http://purl.org/dc/elements/1.1/title	Matrix	_:aut42
_:archivo281	http://xmlns.divxnetworks.com/2004/11/Cast#person	_:reparto871	_:aut42
_:archivo281	http://xmlns.divxnetworks.com/2004/11/Cast#person	_:reparto872	_:aut42
_:archivo281	http://xmlns.divxnetworks.com/2004/11/Cast#person	_:reparto873	_:aut42
_:reparto871	http://xmlns.divxnetworks.com/2004/11/cast#name	Keanu Reeves	_:aut42
_:reparto871	http://xmlns.divxnetworks.com/2004/11/cast#role	Actor	_:aut42
_:reparto871	http://xmlns.divxnetworks.com/2004/11/cast#character	Neo	_:aut42
_:reparto282	http://xmlns.divxnetworks.com/2004/11/cast#name	Andy Wachowski	_:aut42
_:reparto282	http://xmlns.divxnetworks.com/2004/11/cast#role	Director	_:aut42
_:reparto283	http://xmlns.divxnetworks.com/2004/11/cast#name	Larry Wachowski	_:aut42
_:reparto283	http://xmlns.divxnetworks.com/2004/11/cast#role	Director	_:aut42
_:archivo281	http://purl.org/dc/elements/1.1/rights	Derechos de autor 1998 Warner Brothers. Todos los derechos reservados.	_:aut42
_:archivo281	Serie	_:archivo321	_:aut42
_:archivo321	Episodio	2	_:aut42
_:archivo321	http://purl.org/dc/elements/1.1/title	Matrix Reloaded	_:aut42
_:archivo321	Serie	_:archivo122	_:aut42
_:archivo122	Episodio	3	_:aut42
_:aut42	http://xmlns.com/foaf/0.1/organization	_:foaf92	_:aut42
_:foaf92	http://xmlns.com/foaf/0.1/name	Warner Brothers	_:aut42
_:archivo281	http://xmlns.divxnetworks.com/2004/11/track#track	_:pista#dc00	_:aut42
_:pista#dc00	http://xmlns.divxnetworks.com/2004/11/track#resolution	1024 x 768	_:aut42
_:archivo281	http://xmlns.divxnetworks.com/2004/11/content#certificationLevel	HT	_:aut42
_:pista#dc00	http://xmlns.divxnetworks.com/2004/11/track#frameTypeDist	32, 1, 3, 5	_:aut42
_:pista#dc00	http://xmlns.divxnetworks.com/2004/11/track#ajustesCodec	bv1 276 -psy 0 - clave 300 -b 1 -sc 50 -pq 5 -vbv 6951200, 3145728, 2359296 - perfil 3 -nf	_:aut42

En una realización, la expresión del sujeto, predicado, objeto y autoridad se implementa usando representaciones binarias de los datos, que se puede considerar que forman Grafos con Etiquetas Dirigidos (DLG). Un DLG consiste en nodos que son o bien recursos o bien literales. Los recursos son identificadores, que o bien pueden ser conformes con una convención de nomenclatura tal como un Identificador de Recursos Universal ("URI") como se define en RFC 2396 del Grupo de Tareas Especiales sobre Ingeniería de Internet (<http://www.ietf.org/rfc/rfc2396.txt>) o bien se refieren a datos específicos del propio sistema. Los literales son representaciones de un valor real, en lugar de una referencia.

Una ventaja de los DLG es que estos permiten la inclusión de un número flexible de artículos de datos que son del mismo tipo, tales como los miembros de reparto de una película. En el ejemplo mostrado en la tabla 1, se incluyen tres miembros de reparto. No obstante, se puede incluir cualquier número de miembros de reparto. Los DLG también permiten conexiones relacionales con otros tipos de datos. En la tabla 1, hay un artículo de 'metadatos' que tiene un sujeto "_:archivo281", una "Serie" de predicado, y un objeto "_:archivo321". El sujeto "_:archivo281" indica que los 'metadatos' se refieren al contenido del archivo al que se hace referencia como "_:archivo321" (en este caso, la película "Matrix"). El predicado es "Serie", que indica que el objeto tendrá información acerca de otra película en la serie a la

que pertenece "Matrix". No obstante, " _:archivo321" no es el título o cualquier otra información específica acerca de la serie que incluye "Matrix", sino más bien una referencia a otra entrada que proporciona más información acerca de " _:archivo321". La siguiente entrada de 'metadatos', con el sujeto " _:archivo321", no obstante, incluye datos acerca de " _:archivo321", en concreto, que el Título, como es especificado por el Vocabulario Dublin Core, como es indicado por "http://purl.org/dc/elements/1.1/title" de esta secuela, es "Matrix Reloaded".

Las sentencias de 'metadatos' adicionales en la tabla 1 especifican que Keanu Reeves era un miembro del reparto que desempeñaba el papel de Neo y que tanto Larry como Andy Wachowski eran directores. También se expresa información técnica en los 'metadatos'. Las sentencias de 'metadatos' identifican que " _:archivo281" incluye la pista " _:pista#dc00". Los 'metadatos' proporcionan información que incluye la resolución de la pista de vídeo, el nivel de certificación de la pista de vídeo y los ajustes de códec. Aunque no se muestra en la tabla 1, los 'metadatos' también pueden incluir un identificador único asignado a una pista en el momento de la codificación. Cuando se usan identificadores únicos, codificar el mismo contenido múltiples veces dará como resultado un identificador diferente para cada versión codificada del contenido. No obstante, una copia de la pista de vídeo codificada conservaría el identificador de la pista a partir de la cual se copió esta.

Las entradas mostradas en la tabla 1 se pueden sustituir con otros vocabularios tales como el vocabulario de UPnP, que es definido por el foro de UPnP (véase <http://www.upnpforum.org>). Otra alternativa serían los vocabularios de Lenguaje de Sentencia de Artículos Digitales (DIDL) o de DIDL Ligerero desarrollados por la Organización Internacional de Normalización como parte del trabajo hacia la norma MPEG-21. Lo siguiente son ejemplos de predicados dentro del vocabulario de UPnP:

```
urn:schemas-upnp-org:metadata-1-0/upnp/artist
urn:schemas-upnp-org:metadata-1-0/upnp/actor
urn:schemas-upnp-org:metadata-1-0/upnp/autor
urn:schemas-upnp-org:metadata-1-0/upnp/producer
urn:schemas-upnp-org:metadata-1-0/upnp/director
urn:schemas-upnp-org:metadata-1-0/upnp/genre
urn:schemas-upnp-org:metadata-1-0/upnp/album
urn:schemas-upnp-org:metadata-1-0/upnp/playlist
urn:schemas-upnp-org:metadata-1-0/upnp/originalTrackNumber
urn:schemas-upnp-org:metadata-1-0/upnp/userAnnotation
```

La autoridad para la totalidad de los 'metadatos' es ' _:aut42'. Las sentencias de 'metadatos' muestran que ' _:aut42' es 'Warner Brothers'. La autoridad posibilita la evaluación tanto de la calidad del archivo como de las sentencias de 'metadatos' asociadas con el archivo.

Los nodos en un grafo se conectan por medio de nodos de recurso con nombre. Una sentencia de 'metadatos' consiste en un nodo de sujeto, un nodo de predicado y un nodo de objeto. De forma opcional, un nodo de autoridad se puede conectar al DLG como parte de la sentencia de 'metadatos'.

Para cada nodo, hay determinadas características que ayudan a explicar adicionalmente la funcionalidad del nodo. Los tipos posibles se pueden representar como sigue usando el lenguaje de programación ANSI C:

```
/** Tipo no válido */
#define
TIPO_IDENTIFICADOR_RDF_DESCONOCIDO    0x00

/** URI de Recurso rdf:about */
#define TIPO_IDENTIFICADOR_RDF_RECURSO  0x01

/** rdf:IdNodo, _:archivo o N Tripletas generados */
#define TIPO_IDENTIFICADOR_RDF_ANONIMO  0x02
/** URI de predicado */
#define
TIPO_IDENTIFICADOR_RDF_PREDICADO       0x03

/** rdf:li, rdf: <n> */
#define TIPO_IDENTIFICADOR_RDF_ORDINAL  0x04
/** URI de Autoridad */
#define
TIPO_IDENTIFICADOR_RDF AUTORIDAD       0x05

/** literal con formato UTF-8 */
#define TIPO_IDENTIFICADOR_RDF_LITERAL  0x06

/** Entero Literal */
#define TIPO_IDENTIFICADOR_RDF_INT      0x07
```

```

/** Datos de XML literal */
#define
TIPO_IDENTIFICADOR_RDF_XML_LITERAL      0x08

```

Un ejemplo de una estructura de datos (representada en el lenguaje de programación ANSI C) que representa los fragmentos de 'metadatos' contenidos dentro del fragmento de 'DXDT' es como sigue:

```

typedef struct StructDatosRDF
{
    Encabezamiento
    RDF            Encabezamiento;
    uint32_t       numDeSentencias;
    SentenciaRDF   sentencias[RDF_MAX_SENTENCIAS];
} DatosRDF

```

5 El fragmento de 'DatosRDF' incluye un fragmento al que se hace referencia como un fragmento de 'EncabezamientoRDF', un valor 'numDeSentencias' y una lista de fragmentos de 'SentenciaRDF'.

10 El fragmento de 'EncabezamientoRDF' contiene información acerca de la forma en la que se dio formato a los 'metadatos' en el fragmento. En una realización, los datos en el fragmento de 'EncabezamientoRDF' se pueden representar como sigue (representado en ANSI C):

```

typedef struct StructEncabezamientoRDF
{
    uint16_t       versionPrincipal;
    uint16_t       versionSecundaria;
    uint16_t       versionCorreccion;
    uint16_t       numDeEsquemas;
    EsquemaRDF     Esquemas[RDF_MAX_ESQUEMAS]
} EncabezamientoRDF;

```

15 El fragmento de 'EncabezamientoRDF' incluye un número 'versión' que indica la versión del formato de descripción de recurso para posibilitar la compatibilidad con versiones posteriores. El encabezamiento incluye un segundo número 'numDeEsquemas' que representa el número de fragmentos de 'EsquemaRDF' en la lista 'Esquemas', que también forma parte del fragmento de 'EncabezamientoRDF'. En varias realizaciones, los fragmentos de 'EsquemaRDF' se usan para posibilitar que recursos complejos se representen más eficientemente. En una realización, los datos contenidos en un fragmento de 'EsquemaRDF' se pueden representar como sigue (representado en ANSI C):

```

20
    typedef struct StructEsquemaRDF
    {
        wchar_t*     prefijo;
        wchar_t*     uri;
    } EsquemaRDF;

```

25 El fragmento de 'EsquemaRDF' incluye una primera cadena de texto tal como 'dc' identificada como 'prefijo' y una segunda cadena de texto tal como 'http://purl.org/dc/elements/11/' identificada como 'uri'. El 'prefijo' define un término que se puede usar en los 'metadatos' el lugar del 'uri'. El 'uri' es un Identificador de Recursos Universal, que puede ser conforme con un vocabulario normalizado especificado o puede ser un vocabulario específico para un sistema particular.

30 Volviendo al análisis del fragmento de 'DatosRDF'. Además de un fragmento de 'EncabezamientoRDF', el fragmento de 'DatosRDF' también incluye un valor 'numDeSentencias' y una lista 'sentencia' de fragmentos de 'SentenciaRDF'. El valor 'numDeSentencias' indica el número real de fragmentos de 'SentenciaRDF' en la lista 'sentencias' que contienen información. En una realización, los datos contenidos en el fragmento de 'SentenciaRDF' se pueden representar como sigue (representado en ANSI C):

```

    typedef struct StructSentenciaRDF {
        SujetoRDF            sujeto;
        PredicadoRDF         predicado
    }

```

```

ObjetoRDF          objeto;
AutoridadRDF       autoridad;
} SentenciaRDF;

```

Cada fragmento de 'SentenciaRDF' contiene un fragmento de 'metadatos' con respecto al archivo multimedia. Los fragmentos 'sujeto', 'predicado', 'objeto' y 'autoridad' se usan para contener los diversos componentes de los 'metadatos' descritos anteriormente.

El 'sujeto' es un fragmento de 'SujetoRDF', que representa la porción de sujeto de los 'metadatos' descritos anteriormente. En una realización, los datos contenidos dentro del fragmento de 'SujetoRDF' se pueden representar como sigue (representado en ANSI C):

```

typedef struct StructSujetoRDF {
    uint16_t      tipo;
    wchar_t*      valor;
} SujetoRDF;

```

El fragmento de 'SujetoRDF' mostrado anteriormente incluye un valor 'tipo' que indica que los datos son o bien un Recurso o bien un nodo anónimo de un fragmento de 'metadatos' y una cadena de texto de Unicode 'valor', que contiene datos que representan el sujeto del fragmento de 'metadatos'. En algunas realizaciones en las que se ha definido un fragmento de 'EsquemaRDF', el valor puede ser un término definido en lugar de una referencia directa a un recurso.

El 'predicado' en un fragmento de 'SentenciaRDF' es un fragmento de 'PredicadoRDF', que representa la porción de predicado de un fragmento de 'metadatos'. En una realización, los datos contenidos dentro de un fragmento de 'PredicadoRDF' se pueden representar como sigue (representado en ANSI C):

```

typedef struct StructPredicadoRDF
{
    uint16_t      tipo;
    wchar_t*      valor;
} PredicadoRDF;

```

El fragmento de 'PredicadoRDF' mostrado anteriormente incluye un valor 'tipo' que indica que los datos son el URI de predicado o una entrada de lista ordinal de un fragmento de 'metadatos' y una cadena de texto 'valor', que contiene datos que representan el predicado de un fragmento de 'metadatos'. En algunas realizaciones en las que se ha definido un fragmento de 'EsquemaRDF', el valor puede ser un término definido en lugar de una referencia directa a un recurso.

El 'objeto' en un fragmento de 'SentenciaRDF' es un fragmento de 'ObjetoRDF', que representa la porción de objeto de un fragmento de 'metadatos'. En una realización, los datos contenidos en el fragmento de 'ObjetoRDF' se pueden representar como sigue (representado en ANSI C):

```

typedef struct StructObjetoRDF {
    uint16_t      tipo;
    wchar_t*      idioma;
    wchar_t*      URITipoDatos;
    wchar_t*      valor;
} ObjetoRDF;

```

El fragmento de 'ObjetoRDF' mostrado anteriormente incluye un valor 'tipo' que indica que el fragmento de datos es una cadena literal de UTF-8, un entero literal o datos de XML literal de un fragmento de 'metadatos'. El fragmento también incluye tres valores. El primer valor 'idioma' se usa para representar el idioma en el que se expresa el fragmento de 'metadatos' (por ejemplo, el título de una película puede variar en diferentes idiomas). En varias realizaciones, se puede usar una representación convencional para identificar el idioma (tal como RFC 3066 - Etiquetas para la Identificación de Idiomas especificada por el Grupo de Tareas Especiales sobre Ingeniería de Internet, véase <http://www.ietf.org/rfc/rfc3066.txt>). El segundo valor 'URITipoDatos' se usa para indicar el tipo de datos que están contenidos dentro del campo de 'valor' si este no puede ser indicado de forma explícita por el campo de 'tipo'. El URI especificado por el valor URITipoDatos apunta a un Vocabulario de URI de RDF general usado para describir el tipo

particular de Datos que se usa. Diferentes formatos en los que se puede expresar el URI se describen en <http://www.w3.org/TR/rdf-concepts/#section-Datatypes>. En una realización, el 'valor' es un 'carácter ancho'. En otras realizaciones, el 'valor' puede ser cualquiera de diversos tipos de datos, de un único bit a una imagen o una secuencia de vídeo. El 'valor' contiene el fragmento de objeto de los 'metadatos'.

5 La 'autoridad' en un fragmento de 'SentenciaRDF' es un fragmento de 'AutoridadRDF', que representa la porción de autoridad de un fragmento de 'metadatos'. En una realización, los datos contenidos dentro del fragmento de 'AutoridadRDF' se pueden representar como sigue (representado en ANSI C):

```
typedef struct StructAutoridadRDF
{
    uint16_t    tipo;
    wchar_t*    valor;
} AutoridadRDF;
```

10 La estructura de datos de 'AutoridadRDF' mostrada anteriormente incluye un valor 'tipo' que indica los datos son un Recurso o un nodo anónimo de un fragmento de 'metadatos'. El 'valor' contiene los datos que representan la autoridad para los 'metadatos'. En algunas realizaciones en las que se ha definido un fragmento de 'EsquemaRDF', el valor puede ser un término definido en lugar de una referencia directa a un recurso.

15 Una representación conceptual del almacenamiento de un fragmento de 'DXDT' de un archivo multimedia de acuerdo con una realización de la presente invención se muestra en la figura 2.3. El fragmento de 'DXDT' 38 incluye un fragmento de 'EncabezamientoRDF' 110, un valor de 'numDeSentencias' 112 y una lista de fragmentos de SentenciaRDF 114. El fragmento de EncabezamientoRDF 110 incluye un valor de 'versión' 116, un valor de 'numDeEsquemas' 118 y una lista de fragmentos de 'Esquema' 120. Cada fragmento de 'SentenciaRDF' 114 incluye un fragmento de 'SujetoRDF' 122, un fragmento de 'PredicadoRDF' 124, un fragmento de 'ObjetoRDF' 126 y un fragmento de 'AutoridadRDF' 128. El fragmento de 'SujetoRDF' incluye un valor de 'tipo' 130 y un valor de 'valor' 132. El fragmento de 'PredicadoRDF' 124 también incluye un valor de 'tipo' 134 y un valor de 'valor' 136. El fragmento de 'ObjetoRDF' 126 incluye un valor de 'tipo' 138, un valor de 'idioma' 140 (mostrado en la figura como 'idio'), un valor de 'URITipoDatos' 142 (mostrado en la figura como 'Tdatos') y un valor de "valor" 144. El fragmento de 'AutoridadRDF' 128 incluye un valor de 'tipo' 146 y un valor de 'valor' 148. Aunque el fragmento de 'DXDT' ilustrado se muestra como que incluye un único fragmento de 'Esquema' y un único fragmento de 'SentenciaRDF', un experto en la materia apreciará fácilmente que se pueden usar números diferentes de fragmentos de 'Esquema' y fragmentos de 'SentenciaRDF' en un fragmento que describe 'metadatos'.

30 Como se analiza posteriormente, los archivos multimedia de acuerdo con algunas realizaciones de la presente invención se pueden modificar y actualizar de forma continua. Puede ser difícil determinar por adelantado los 'metadatos' que asociar con el propio archivo y los 'metadatos' a los que acceder de forma remota (por ejemplo, por medio de Internet). Por lo general, unos 'metadatos' suficientes están contenidos dentro de un archivo multimedia de acuerdo con una realización de la presente invención con el fin de describir los contenidos del archivo. Se puede obtener información adicional si el dispositivo que está revisando el archivo es capaz de acceder, por medio de una red, a otros dispositivos que contienen 'metadatos' a los que se hace referencia desde dentro del archivo.

40 Los métodos de representar 'metadatos' descritos anteriormente pueden ser ampliables y pueden proporcionar la capacidad de añadir y retirar diferentes campos de 'metadatos' almacenados dentro del archivo a medida que la necesidad del mismo cambia con el paso del tiempo. Además, la representación de 'metadatos' puede ser compatible con versiones posteriores entre revisiones.

45 La forma estructurada en la que se representan los 'metadatos' de acuerdo con algunas realizaciones de la presente invención posibilita que los dispositivos consulten el archivo multimedia para determinar mejor sus contenidos. Entonces, la consulta se podría usar para actualizar los contenidos del archivo multimedia, para obtener 'metadatos' adicionales con respecto al archivo multimedia, generar un menú en relación con los contenidos del archivo o realizar cualquier otra función que implique el procesamiento automático de datos representados en un formato convencional. Además, definir la longitud de cada elemento analizable de los 'metadatos' puede aumentar la facilidad con la que los dispositivos con cantidades limitadas de memoria, tales como dispositivos de electrónica de consumo, pueden acceder los 'metadatos'.

50 En otras realizaciones, los 'metadatos' se representan usando fragmentos individuales para cada fragmento de 'metadatos'. Varios fragmentos de 'DXDT' de acuerdo con la presente invención incluyen un fragmento binario que contiene 'metadatos' codificados como se ha descrito anteriormente y fragmentos adicionales que contienen fragmentos individuales de 'metadatos' a los que se da formato o bien como se ha descrito anteriormente o bien en otro formato. En algunas realizaciones en las que se incluyen 'metadatos' binarios en el fragmento de 'DXDT', los 'metadatos' binarios se pueden representar usando ASCII codificado de 64 bits. En otras realizaciones, se pueden usar otras representaciones binarias.

60

Algunos ejemplos de fragmentos individuales que se pueden incluir en el fragmento de 'DXDT' de acuerdo con la presente invención se ilustran en la figura 2.3.1. Los 'metadatos' incluyen un fragmento de 'Metadatos' 150 que puede contener un fragmento de 'MetadatosRelacionAspectoPixeles' 152a, un fragmento de 'MetadatosURICodificador' 152b, un fragmento de 'MetadatosAjustesCodec' 152c, un fragmento de 'MetadatosTipoTrama' 152d, un fragmento de 'MetadatosResolucionVideo' 152e, un fragmento de 'MetadatosPublicador' 152f, un fragmento de 'MetadatosCreador' 152g, un fragmento de 'MetadatosGenero' 152h, un fragmento de 'MetadatosHerramientaCreador' 152i, un fragmento de 'MetadatosDerechos' 152j, un fragmento de 'MetadatosTiempoEjecucion' 152k, un fragmento de 'MetadatosCuantificador' 152l, un fragmento de 'MetadatosInfoCodec' 152m, un fragmento de 'MetadatosNombreCodificador' 152n, un Fragmento de 'MetadatosTasaTramas' 152o, un fragmento de 'MetadatosFuenteEntrada' 152p, un fragmento de 'MetadatosIDArchivo' 152q, un fragmento de 'MetadatosTipo' 152r, un fragmento de 'MetadatosTitulo' 152s y/o un fragmento de 'MetadatosNivelCert' 152t.

El fragmento de 'MetadatosRelacionAspectoPixeles' 152a incluye información con respecto a la relación de aspecto de píxeles del vídeo codificado. El fragmento de 'MetadatosURICodificador' 152b incluye información con respecto al codificador. El fragmento de 'MetadatosAjustesCodec' 152c incluye información con respecto a los ajustes del códec usado para codificar el vídeo. El fragmento de 'MetadatosTipoTrama' 152d incluye información con respecto a las tramas de vídeo. El fragmento de 'MetadatosResolucionVideo' 152e incluye información con respecto a la resolución de vídeo del vídeo codificado. El fragmento de 'MetadatosPublicador' 152f incluye información con respecto a la persona u organización que publicó los medios. El fragmento de 'MetadatosCreador' 152g incluye información con respecto al creador del contenido. El fragmento de 'MetadatosGenero' 152h incluye información con respecto al género de los medios. El fragmento de 'MetadatosHerramientaCreador' 152i incluye información con respecto a la herramienta usada para crear el archivo. El fragmento de 'MetadatosDerechos' 152j incluye información con respecto a la DRM. El fragmento de 'MetadatosTiempoEjecucion' 152k incluye información con respecto al tiempo de ejecución de los medios. El fragmento de 'MetadatosCuantificador' 152l incluye información con respecto al cuantificador usado para codificar el vídeo. El fragmento de 'MetadatosInfoCodec' 152m incluye información con respecto al códec. El fragmento de 'MetadatosNombreCodificador' 152n incluye información con respecto al nombre del codificador. El Fragmento de 'MetadatosTasaTramas' 152o incluye información con respecto a la tasa de tramas de los medios. El fragmento de 'MetadatosFuenteEntrada' 152p incluye información con respecto a la fuente de entrada. El fragmento de 'MetadatosIDArchivo' 152q incluye un identificador único para el archivo. El fragmento de 'MetadatosTipo' 152r incluye información con respecto al tipo del archivo multimedia. El fragmento de 'MetadatosTitulo' 152s incluye el título de los medios y el fragmento de 'MetadatosNivelCert' 152t incluye información con respecto al nivel de certificación de los medios. En otras realizaciones, se pueden incluir fragmentos adicionales que contienen 'metadatos' adicionales. En varias realizaciones, un fragmento que contiene 'metadatos' en un formato binario como se ha descrito anteriormente se puede incluir dentro del fragmento de 'Metadatos'. En una realización, el fragmento de 'metadatos' binarios se codifica como ASCII de 64 bits.

2.5. Los fragmentos de 'DMNU'

Haciendo referencia a las figuras 2.0. y 2.0.1., se muestran un primer fragmento de 'DMNU' 40 (40') y un segundo fragmento de 'DMNU' 46 (46'). En la figura 2.0. el segundo fragmento de 'DMNU' 46 forma parte del archivo multimedia 30. En la realización ilustrada en la figura 2.0.1., el fragmento de 'DMNU' 46' está contenido dentro de un fragmento RIFF separado. En ambos casos, el primer y el segundo fragmentos de 'DMNU' contienen datos que se pueden usar para visualizar menús navegables. En una realización, el primer fragmento de 'DMNU' 40 (40') contiene datos que se pueden usar para crear un menú simple que no incluye características avanzadas tales como animaciones de segundo plano ampliadas. Además, el segundo fragmento de 'DMNU' 46 (46') incluye datos que se pueden usar para crear un menú más complejo que incluye características avanzadas tales como un segundo plano animado ampliado.

En varias realizaciones, proporcionar un menú simple y uno complejo puede posibilitar que un dispositivo elija el menú que este desea reproducir. Colocar el menor de los dos menús antes que el fragmento de lista de 'película' 42 posibilita que aquellos dispositivos de acuerdo con algunas realizaciones de la presente invención que no pueden visualizar menús salten rápidamente la información que no se puede visualizar.

En otras realizaciones, los datos requeridos para crear un único menú se dividen entre el primer y el segundo fragmentos de 'DMNU'. Como alternativa, el fragmento de 'DMNU' puede ser un único fragmento antes del fragmento de 'película' que contiene datos para un único conjunto de menús o múltiples conjuntos de menús. En otras realizaciones, el fragmento de 'DMNU' puede ser un único o múltiples fragmentos situados en otras ubicaciones de principio a fin del archivo multimedia.

En varios archivos multimedia de acuerdo con la presente invención, el primer fragmento de 'DMNU' 40 (40') se puede generar automáticamente basándose en un menú 'más rico' en el segundo fragmento de 'DMNU' 46 (46'). La generación automática de menús se analiza con mayor detalle posteriormente.

La estructura de un fragmento de 'DMNU' de acuerdo con una realización de la presente invención se muestra en la figura 2.4. El fragmento de 'DMNU' 158 es un fragmento de lista que contiene un fragmento de menú 160 y un fragmento de 'MRIF' 162. El fragmento de menú contiene la información necesaria para construir y navegar a través de los menús. El fragmento de 'MRIF' contiene información de medios que se puede usar para proporcionar subtítulos,

vídeo de segundo plano y audio de segundo plano a los menús. En varias realizaciones, el fragmento de 'DMNU' contiene información de menú que posibilita la visualización de menús en varios idiomas diferentes.

5 En una realización, el fragmento de 'MenuWow' 160 contiene la jerarquía de los objetos de fragmento de menú que se ilustran conceptualmente en la figura 2.5. En la parte de arriba de la jerarquía está el fragmento de 'GestorMenuWow' 170. El fragmento de 'GestorMenuWow' puede contener uno o más fragmentos de 'MenusIdioma' 172 y un fragmento de 'Medios' 174.

10 El uso de fragmentos de 'MenusIdioma' 172 posibilita que el fragmento de 'DMNU' 158 contenga información de menú en idiomas diferentes. Cada fragmento de 'MenusIdioma' 172 contiene la información usada para generar un conjunto completo de menús en un idioma especificado. Por lo tanto, el fragmento de 'MenusIdioma' incluye un identificador que identifica el idioma de la información asociada con el fragmento de 'MenusIdioma'. El fragmento de 'MenusIdioma' también incluye una lista de fragmentos de 'MenuWow' 175.

15 Cada fragmento de 'MenuWow' 175 contiene la totalidad de la información que se va a visualizar en la pantalla para un menú particular. Esta información puede incluir vídeo y audio de segundo plano. La información también puede incluir datos con respecto a acciones de botón que se pueden usar para acceder a otros menús o para salir del menú y comenzar a visualizar una porción del archivo multimedia. En una realización, el fragmento de 'MenuWow' 175 incluye una lista de referencias a medios. Estas referencias se refieren a información contenida en el fragmento de
20 'Medios' 174, que se analizará más adelante. Las referencias a medios pueden definir el vídeo de segundo plano y el audio de segundo plano para un menú. El fragmento de 'MenuWow' 175 también define una superposición que se puede usar para resaltar un botón específico, cuando se accede por primera vez a un menú.

Además, cada fragmento de 'MenuWow' 175 incluye un número de fragmentos de 'MenuBoton' 176. Cada fragmento
25 de 'MenuBoton' define las propiedades de un botón en pantalla. El fragmento de 'MenuBoton' puede describir cosas tales como la superposición a usar cuando el botón es resaltado por el usuario, el nombre del botón y qué hacer en respuesta a diversas acciones realizadas por un usuario que navega a través del menú. Las respuestas a acciones se definen al hacer referencia a un fragmento de 'Acción' 178. Una única acción, por ejemplo, seleccionar un botón, puede dar como resultado que se acceda a varios fragmentos de 'Acción'. En algunas realizaciones en las que el usuario es
30 capaz de interaccionar con el menú usando un dispositivo tal como un ratón que posibilita que un puntero en pantalla se mueva en torno a la visualización de una forma sin restricciones, la ubicación en pantalla de los botones se puede definir usando un fragmento de 'RectanguloMenu' 180. El conocimiento de la ubicación en pantalla del botón posibilita que un sistema determine si un usuario está seleccionando un botón, cuando se usa un dispositivo de entrada libre.

35 Cada fragmento de 'Acción' identifica uno o más de un número de diferentes diversidades de fragmentos relacionados con acciones, que pueden incluir un fragmento de 'AccionReproduccion' 182, un fragmento de 'AccionTransicionMenu' 184, un fragmento de 'AccionVolverAReproduccion' 186, un fragmento de 'AccionSeleccionAudio' 188, un fragmento de 'AccionSeleccionSubtitulo' 190 y un fragmento de 'AccionTransicionBoton' 191. Un fragmento de 'AccionReproduccion' 182 identifica una porción de cada una de las pistas de vídeo, de audio y de subtítulo dentro de
40 un archivo multimedia. El fragmento de 'AccionReproduccion' hace referencia a una porción de la pista de vídeo usando una referencia a un fragmento de 'PistaMedios' (véase el análisis posterior). El fragmento de 'AccionReproduccion' identifica pistas de audio y de subtítulo usando fragmentos de 'PistaSubtitulo' 192 y de 'PistaAudio' 194. Los fragmentos tanto de 'PistaSubtitulo' como de 'PistaAudio' contienen referencias a un fragmento de 'PistaMedios' 198. Cuando un fragmento de 'AccionReproduccion' forma la base de una acción de acuerdo con algunas realizaciones de la presente invención, las pistas de audio y de subtítulo que se seleccionan se determinan mediante los valores de variables establecidos inicialmente como valores por defecto y, entonces, son potencialmente modificadas por las interacciones de un usuario con el menú.

50 Cada fragmento de 'AccionTransicionMenu' 184 contiene una referencia a un fragmento de 'MenuWow' 175. Esta referencia se puede usar para obtener información para realizar una transición a, y para visualizar, otro menú.

Cada fragmento de 'AccionVolverAReproduccion' 186 contiene información que posibilita que un reproductor vuelva a una porción del archivo multimedia a la que se estaba accediendo antes de que el usuario suba un menú.

55 Cada fragmento de 'AccionSeleccionAudio' 188 contiene información que se puede usar para seleccionar una pista de audio particular. En una realización, la pista de audio se selecciona de entre pistas de audio contenidas dentro de un archivo multimedia de acuerdo con una realización de la presente invención. En otras realizaciones, la pista de audio se puede situar en un archivo al que se hace referencia externamente.

60 Cada fragmento de 'AccionSeleccionSubtitulo' 190 contiene información que se puede usar para seleccionar una pista de subtítulo particular. En una realización, la pista de subtítulo se selecciona de entre un subtítulo contenido dentro de un archivo multimedia de acuerdo con una realización de la presente invención. En otras realizaciones, la pista de subtítulo se puede situar en un archivo al que se hace referencia externamente.

65 Cada fragmento de 'AccionTransicionBoton' 191 contiene información que se puede usar para realizar una transición a otro botón en el mismo menú. Esto se realiza después de que se hayan realizado otras acciones asociadas con un

botón.

El fragmento de 'Medios' 174 incluye un número de fragmentos de 'FuenteMedios' 166 y fragmentos de 'PistaMedios' 198. El fragmento de 'Medios' define la totalidad de las pistas multimedia (por ejemplo, audio, vídeo, subtítulo) usadas por la característica y el sistema de menú. Cada fragmento de 'FuenteMedios' 196 identifica un fragmento RIFF dentro del archivo multimedia de acuerdo con una realización de la presente invención, que, a su vez, puede incluir múltiples fragmentos RIFF. Cada fragmento de 'PistaMedios' 198 identifica una porción de una pista multimedia dentro de un fragmento RIFF especificado por un fragmento de 'FuenteMedios'.

El fragmento de 'MRIF' 162 es, en esencia, su propio archivo multimedia pequeño que cumple con el formato RIFF. El fragmento de 'MRIF' contiene pistas de audio, de vídeo y de subtítulo que se pueden usar para proporcionar un audio y vídeo de segundo plano y superposiciones para menús. El fragmento de 'MRIF' también puede contener vídeo que se va a usar como superposiciones para indicar botones de menú resaltados. En algunas realizaciones en las que se requieren menos datos de menú, el vídeo de segundo plano puede ser una trama fija (una variación del formato AVI) o una secuencia pequeña de tramas idénticas. En otras realizaciones, se pueden usar unas secuencias de vídeo más elaboradas para proporcionar el vídeo de segundo plano.

Como se ha analizado anteriormente, los diversos fragmentos que forman parte de un fragmento de 'MenuWow' 175 y el propio fragmento de 'MenuWow' contienen referencias a pistas de medios reales. Cada una de estas referencias es, por lo general, para una pista de medios definida en el fragmento de lista de 'hdrl' de un fragmento RIFF.

Otros fragmentos que se pueden usar para crear un fragmento de 'DMNU' de acuerdo con la presente invención se muestran en la figura 2.6. El fragmento de 'DMNU' incluye un fragmento de 'GestorMenuWow' 170'. El fragmento de 'GestorMenuWow' 170' puede contener al menos un fragmento de 'MenusIdioma' 172', al menos un fragmento de 'Medios' 174' y al menos un fragmento de 'TablaTraduccion' 200.

El contenido del fragmento de 'MenusIdioma' 172' es, en gran medida, similar al del fragmento de 'MenusIdioma' 172 ilustrado en la figura 2.5. La diferencia principal es que el fragmento de 'AccionReproduccion' 182' no contiene los fragmentos de 'PistaSubtitulo' 192 y los fragmentos de 'PistaAudio' 194.

El fragmento de 'Medios' 174' es significativamente diferente del fragmento de 'Medios' 174 mostrado en la figura 2.5. El fragmento de 'Medios' 174' contiene al menos un fragmento de 'Título' 202 y al menos un fragmento de 'PistasMenu' 204. El fragmento de 'Título' se refiere a un título dentro del archivo multimedia. Como se ha analizado anteriormente, los archivos multimedia de acuerdo con algunas realizaciones de la presente invención pueden incluir más de un título (por ejemplo, múltiples episodios en una serie de televisión, una serie relacionada de características de longitud completa o simplemente una selección de características diferentes). El fragmento de 'PistasMenu' 204 contiene información con respecto a una selección de medios que se usa para crear una visualización de menú y la pista de sonido de audio y los subtítulos que acompañan a la visualización.

El fragmento de 'Título' puede contener al menos un fragmento de 'Capítulo' 206. El fragmento de 'Capítulo' 206 hace referencia a una escena dentro de un título particular. El fragmento de 'Capítulo' 206 contiene referencias a las porciones de la pista de vídeo, cada pista de audio y cada pista de subtítulo que se corresponden con la escena indicada por el fragmento de 'Capítulo'. En una realización, las referencias se implementan usando los fragmentos de 'FuenteMedios' 196' y los fragmentos de 'PistaMedios' 198' similares a los descritos anteriormente en relación con la figura 2.5. En varias realizaciones, un fragmento de 'PistaMedios' hace referencia a la porción apropiada de la pista de vídeo y cada uno de un número de fragmentos de 'PistaMedios' adicionales hace referencia a una de las pistas de audio o pistas de subtítulo. En una realización, se hace referencia a la totalidad de las pistas de audio y pistas de subtítulo que se corresponden con una pista de vídeo particular usando fragmentos de 'PistaMedios' separados.

Como se ha descrito anteriormente, los fragmentos de 'PistasMenu' 204 contienen referencias a los medios que se usan para generar los medios de audio, de vídeo y de superposición de los menús. En una realización, las referencias a la información de medios se hacen usando los fragmentos de 'FuenteMedios' 196' y los fragmentos de 'PistaMedios' 198' contenidos dentro del fragmento de 'PistasMenu'. En una realización, los fragmentos de 'FuenteMedios' 196' y los fragmentos de 'PistaMedios' 198' se implementan de la forma descrita anteriormente en relación con la figura 2.5.

El fragmento de 'TablaTraduccion' 200 se puede usar para contener cadenas de texto que describen cada título y capítulo en diversos idiomas. En una realización, el fragmento de 'TablaTraduccion' 200 incluye al menos un fragmento de 'ConsultaTraduccion' 208. Cada fragmento de 'ConsultaTraduccion' 208 se asocia con un fragmento de 'Título' 202, un fragmento de 'Capítulo' 206 o un fragmento de 'PistaMedios' 196' y contiene un número de fragmentos de 'Traducción' 210. Cada uno de los fragmentos de 'Traducción' en un fragmento de 'ConsultaTraduccion' contiene una cadena de texto que describe el fragmento asociado con el fragmento de 'ConsultaTraduccion' en un idioma indicado por el fragmento de 'Traducción'.

Un diagrama que ilustra conceptualmente las relaciones entre los diversos fragmentos contenidos dentro de un fragmento de 'DMNU' se ilustra en la figura 2.6.1. La figura muestra la contención de un fragmento por otro fragmento usando una flecha de trazo continuo. La dirección en la que apunta la flecha indica el fragmento contenidos por el

fragmento en el que se origina la flecha. Las referencias realizadas por un fragmento a otro fragmento son indicadas por una línea de trazo discontinuo, en donde el fragmento al que se hace referencia es indicado por la flecha de trazo discontinuo.

5 2.6. El fragmento de 'basura'

El fragmento de 'basura' 41 es un fragmento opcional que se puede incluir en archivos multimedia de acuerdo con algunas realizaciones de la presente invención. La naturaleza del fragmento de 'basura' se especifica en el formato de archivo AVI.

10 2.7. El fragmento de lista de 'película'

El fragmento de lista de 'película' 42 contiene un número de fragmentos de 'datos'. Algunos ejemplos de información que los fragmentos de 'datos' pueden contener son datos de audio, de vídeo o de subtítulo. En una realización, el fragmento de lista de 'película' incluye datos para al menos una pista de vídeo, múltiples pistas de audio y múltiples pistas de subtítulo.

La intercalación de fragmentos de 'datos' en el fragmento de lista de 'película' 42 de un archivo multimedia que contiene una pista de vídeo, tres pistas de audio y tres pistas de subtítulo se ilustra en la figura 2.7. Por razones de conveniencia, un fragmento de 'datos' que contiene vídeo se describirá como fragmento de 'vídeo', se hará referencia a un fragmento de 'datos' que contiene audio como fragmento de 'audio' y se hará referencia a un fragmento de 'datos' que contiene subtítulos como fragmento de 'subtítulo'. En el fragmento de lista de 'película' 42 ilustrado, cada fragmento de 'vídeo' 262 está separado del siguiente fragmento de 'vídeo' por los fragmentos de 'audio' 264 a partir de cada una de las pistas de audio. En varias realizaciones, los fragmentos de 'audio' contienen la porción de la pista de audio que se corresponde con la porción de vídeo contenida en el fragmento de 'vídeo' a continuación del fragmento de 'audio'.

Los fragmentos de 'vídeo' adyacentes también pueden estar separados de una de las pistas de subtítulo por uno o más fragmentos de 'subtítulo' 266. En una realización, el fragmento de 'subtítulo' 266 incluye un subtítulo y un tiempo de inicio y un tiempo de parada. En varias realizaciones, el fragmento de 'subtítulo' se intercala en el fragmento de lista de 'película' de tal modo que el fragmento de 'vídeo' a continuación del fragmento de 'subtítulo' incluye la porción de vídeo que tiene lugar en el tiempo de inicio del subtítulo. En otras realizaciones, el tiempo de inicio de todos los fragmentos de 'subtítulo' y de 'audio' se produce antes del tiempo de inicio equivalente del vídeo. En una realización, los fragmentos de 'audio' y de 'subtítulo' se pueden colocar dentro de un plazo de 5 segundos del fragmento de 'vídeo' correspondiente y, en otras realizaciones, los fragmentos de 'audio' y de 'subtítulo' se pueden colocar dentro de un plazo de tiempo en relación con la cantidad de vídeo que puede ser almacenado en memoria intermedia por un dispositivo capaz de visualizar el audio y vídeo dentro del archivo.

En una realización, los fragmentos de 'datos' incluyen un código de 'FOURCC' para identificar la secuencia a la que pertenece el fragmento de 'datos'. El código de 'FOURCC' consiste en un número de secuencia de dos dígitos seguido por un código de dos caracteres que define el tipo de información en el fragmento. (Raider) Un código 'FOURCC' alternativo consiste en un código de dos caracteres que define el tipo de información en el fragmento seguido por el número de secuencia de dos dígitos. Algunos ejemplos del código de dos caracteres se muestran en la siguiente tabla:

Tabla 2 - Códigos de dos caracteres seleccionados usados en códigos FOURCC

Código de dos caracteres	Descripción
db	Trama de vídeo descomprimida
dc	Trama de vídeo comprimida
dd	Info de clave de DRM para la trama de vídeo
pc	Cambio de paleta
wb	Datos de audio
st	Subtítulo (modo de texto)
sb	Subtítulo (modo de mapa de bits)
ch	Capítulo

En una realización, la estructura de los fragmentos de 'vídeo' 262 y los fragmentos de 'audio' 264 cumple con el formato de archivo AVI. En otras realizaciones, se pueden usar otros formatos para los fragmentos que especifican la naturaleza de los medios y contienen los medios codificados.

En varias realizaciones, los datos contenidos dentro de un fragmento de 'subtítulo' 266 se pueden representar como

sigue:

```
typedef struct fragmentosubtitulo
{
    FOURCC fcc;
    DWORD cb;
    STR duracion;
    STR subtítulo;
} FRAGMENTOSUBTITULO;
```

El valor 'fcc' es el código de FOURCC que indica la pista de subtítulo y la naturaleza de la pista de subtítulo (modo de texto o de mapa de bits). El valor 'cb' especifica el tamaño de la estructura. El valor 'duración' especifica el tiempo en el punto de partida y de finalización del subtítulo. En una realización, esta puede estar en la forma hh:mm:ss.xxx - hh:mm:ss.xxx. hh representa las horas, mm los minutos, ss los segundos y xxx los milisegundos. El valor 'subtítulo' contiene o bien el texto de Unicode del subtítulo en modo de texto o bien una imagen de mapa de bits del subtítulo en el modo de mapa de bits. Varias realizaciones de la presente invención usan imágenes de mapa de bits comprimidas para representar la información de subtítulo. En una realización, el campo de 'subtítulo' contiene información con respecto a la anchura, la altura y la posición en pantalla del subtítulo. Además, el campo de 'subtítulo' también puede contener información de color y los píxeles reales del mapa de bits. En varias realizaciones, se usa una codificación por longitud de ejecución para reducir la cantidad de información de píxeles requerida para representar el mapa de bits.

Los archivos multimedia de acuerdo con algunas realizaciones de la presente invención pueden incluir gestión de derechos digitales. Esta información se puede usar en vídeo sobre aplicaciones a petición. Los archivos multimedia que están protegidos por gestión de derechos digitales solo se pueden reproducir correctamente en un reproductor al que se ha concedido el derecho específico de reproducción. En una realización, el hecho de que una pista está protegida por gestión de derechos digitales se puede indicar en la información acerca de la pista en el fragmento de lista de 'hdlr' (véase la descripción anterior). Un archivo multimedia de acuerdo con una realización de la presente invención que incluye una pista protegida por gestión de derechos digitales también puede contener información acerca de la gestión de derechos digitales en el fragmento de lista de 'película'.

Un fragmento de lista de 'película' de un archivo multimedia de acuerdo con una realización de la presente invención que incluye una pista de vídeo, múltiples pistas de audio, al menos una pista de subtítulo e información que posibilita la gestión de derechos digitales se ilustra en la figura 2.8. El fragmento de lista de 'película' 42' es similar al fragmento de lista de 'película' mostrado en la figura 2.7. con la adición de un fragmento de 'DRM' 270 antes de cada fragmento de vídeo 262'. Los fragmentos de 'DRM' 270 son fragmentos de 'datos' que contienen información de gestión de derechos digitales, que puede ser identificada por un código de FOURCC 'nndd'. Los primeros dos caracteres 'nn' se refieren al número de pista y los segundos dos caracteres son 'dd' para señalar que el fragmento contiene información de gestión de derechos digitales. En una realización, el fragmento de 'DRM' 270 proporciona la información de gestión de derechos digitales para el fragmento de 'vídeo' 262' a continuación del fragmento de 'DRM'. Un dispositivo que intenta reproducir la pista de vídeo protegida por gestión de derechos digitales usa la información en el fragmento de 'DRM' para descodificar la información de vídeo en el fragmento de 'vídeo'. Por lo general, se interpreta que la ausencia de un fragmento de 'DRM' antes de un fragmento de 'vídeo' quiere decir que el fragmento de 'vídeo' está desprotegido.

En un sistema de cifrado de acuerdo con una realización de la presente invención, los fragmentos de vídeo solo están parcialmente cifrados. En donde se usa un cifrado parcial, los fragmentos de 'DRM' contienen una referencia a la porción de un fragmento de 'vídeo' que se cifra y una referencia a la clave que se puede usar para descifrar la porción cifrada. Las claves de descifrado se pueden situar en un encabezamiento de 'DRM', que es parte del fragmento de 'strd' (véase la descripción anterior). Las claves de descifrado se aleatorizan y se cifran con una clave maestra. El encabezamiento de 'DRM' también contiene información que identifica la clave maestra.

Una representación conceptual de la información en un fragmento de 'DRM' se muestra en la figura 2.9. El fragmento de 'DRM' 270 puede incluir un valor de 'trama' 280, un valor de 'estado' 282, un valor de 'desplazamiento' 284, un valor de 'número' 286 y un valor de 'clave' 288. El valor de 'trama' se puede usar para hacer referencia a la trama cifrada de vídeo. El valor de 'estado' se puede usar para indicar si se cifra la trama, el valor de 'desplazamiento' 284 apunta al inicio del bloque cifrado dentro de la trama y el valor de 'número' 286 indica el número de bytes cifrados en el bloque. El valor de 'clave' 288 hace referencia a la clave de descifrado que se puede usar para descifrar el bloque.

2.8. El fragmento de 'idx1'

El fragmento de 'idx1' 44 es un fragmento opcional que se puede usar para indexar los fragmentos de 'datos' en el fragmento de lista de 'película' 42. En una realización, el fragmento de 'idx1' se puede implementar como se especifica en el formato AVI. En otras realizaciones, el fragmento de 'idx1' se puede implementar usando estructuras de datos que hacen referencia a la ubicación dentro del archivo de cada uno de los fragmentos de 'datos' en el fragmento de lista de 'película'. En varias realizaciones, el fragmento de 'idx1' identifica cada fragmento de 'datos' mediante el número de pista de los datos y el tipo de los datos. Los códigos de FOURCC a los que se ha hecho referencia

anteriormente se pueden usar para este fin.

3. Codificación de un archivo multimedia

5 Algunas realizaciones de la presente invención se pueden usar para generar archivos multimedia en un número de formas. En un caso, algunos sistemas de acuerdo con algunas realizaciones de la presente invención pueden generar archivos multimedia a partir de archivos que contienen pistas de vídeo, pistas de audio y pistas de subtítulo separadas. En tales casos, se puede crear e insertar en el archivo otra información, tal como información de menú y 'metadatos'.

10 Otros sistemas de acuerdo con algunas realizaciones de la presente invención se pueden usar para extraer información a partir de los diversos archivos en un Disco de Vídeo Digital ("DVD") y crear un único archivo multimedia de acuerdo con una realización de la presente invención. En donde la fuente inicial de la información es un DVD, algunos sistemas de acuerdo con algunas realizaciones de la presente invención pueden usar un códec para obtener una compresión más grande y pueden refragmentar el audio de tal modo que los fragmentos de audio se corresponden con los fragmentos de vídeo en el archivo multimedia recién creado. Además, la información de menú en el sistema de menú de DVD se puede analizar y usarse para generar información de menú incluida en el archivo multimedia.

Otras realizaciones pueden generar un nuevo archivo multimedia mediante la adición de contenido adicional a un archivo multimedia existente de acuerdo con una realización de la presente invención. Un ejemplo de adición de contenido adicional sería la adición de una pista de audio adicional al archivo tal como una pista de audio que contiene un comentario (por ejemplo, comentarios del director, relato creado a posteriori de un vídeo de vacaciones). La información de pista de audio adicional intercalada en el archivo multimedia también podría estar acompañada por una modificación de la información de menú en el archivo multimedia para habilitar la reproducción de la nueva pista de audio.

25 3.1. Generación usando pistas de datos almacenadas

Un sistema de acuerdo con una realización de la presente invención para generar un archivo multimedia se ilustra en la figura 3.0. El componente principal del sistema 350 es el intercalador 352. El intercalador recibe fragmentos de información y los intercala para crear un archivo multimedia de acuerdo con una realización de la presente invención en el formato descrito anteriormente. El intercalador también recibe información con respecto a 'metadatos' a partir de un gestor de metadatos 354. El intercalador emite un archivo multimedia de acuerdo con algunas realizaciones de la presente invención a un dispositivo de almacenamiento 356.

35 Por lo general, los fragmentos proporcionados al intercalador se almacenan en un dispositivo de almacenamiento. En varias realizaciones, la totalidad de los fragmentos se almacenan en el mismo dispositivo de almacenamiento. En otras realizaciones, los fragmentos se pueden proporcionar al intercalador a partir de diversos dispositivos de almacenamiento o generarse y proporcionarse al intercalador en tiempo real.

40 En la realización ilustrada en la figura 3.0., ya se han generado el fragmento de 'DMNU' 358 y el fragmento de 'DXDT' 360, y se almacenan en dispositivos de almacenamiento. La fuente de vídeo 362 se almacena en un dispositivo de almacenamiento y se descodifica usando un descodificador de vídeo 364 y, entonces, se codifica usando un codificador de vídeo 366 para generar un fragmento de 'vídeo'. Las fuentes de audio 368 también se almacenan en dispositivos de almacenamiento. Se generan fragmentos de audio mediante la descodificación de la fuente de audio usando un descodificador de audio 370 y, entonces, codificando el audio descodificado usando un codificador de audio 372. Los fragmentos de 'subtítulo' se generan a partir de subtítulos de texto 374 almacenados en un dispositivo de almacenamiento. Los subtítulos se proporcionan a un primer transcodificador 376, que convierte cualquiera de un número de formatos de subtítulo en un formato de mapa de bits sin procesar. En una realización, el formato de subtítulo almacenado puede ser un formato tal como SRT, SUB o SSA. Además, el formato de mapa de bits puede ser el de un mapa de bits de cuatro bits que incluye una tabla de consulta de paleta de color. La tabla de consulta de paleta de color incluye una identificación de profundidad de color de 24 bits para cada uno de los dieciséis códigos de color de cuatro bits posibles. Un único archivo multimedia puede incluir más de una tabla de consulta de paleta de color (véase el código de FOURCC de paleta "pc" en la tabla 2 anterior). El mapa de bits de cuatro bits permite de este modo que cada menú tenga 16 colores simultáneos diferentes tomados de una paleta de 16 millones de colores. En algunas realizaciones alternativas, se usan números diferentes de bits por píxel y profundidades de color diferentes. La salida del primer transcodificador 376 se proporciona a un segundo transcodificador 378, que comprime el mapa de bits. En una realización, se usa una codificación por longitud de ejecución para comprimir el mapa de bits. En otras realizaciones, se usan otros formatos de compresión adecuados.

60 En una realización, las interfaces entre los diversos codificadores, descodificadores y transcodificadores son conformes con las normas de DirectShow especificadas por Microsoft Corporation. En otras realizaciones, no es necesario que el software usado para realizar la codificación, la descodificación y la transcodificación cumpla con tales normas.

65 En la realización ilustrada, se muestran componentes de procesamiento separados para cada fuente de medios. En otras realizaciones, los recursos se pueden compartir. Por ejemplo, se podría usar un único descodificador de audio y

codificador de audio para generar fragmentos de audio a partir de la totalidad de las fuentes. Por lo general, la totalidad del sistema se puede implementar en un ordenador usando software y conectarse a un dispositivo de almacenamiento tal como una unidad de disco duro.

Con el fin de utilizar el intercalador de la forma descrita anteriormente, el fragmento de 'DMNU', el fragmento de 'DXDT', los fragmentos de 'vídeo', los fragmentos de 'audio' y los fragmentos de 'subtítulo' de acuerdo con algunas realizaciones de la presente invención se han de generar y proporcionarse al intercalador. El proceso de generar cada uno de los diversos fragmentos en un archivo multimedia de acuerdo con una realización de la presente invención se analiza con mayor detalle posteriormente.

3.2. Generación de un fragmento de 'DXDT'

El fragmento de 'DXDT' se puede generar en cualquiera de un número de formas. En una realización, se introducen 'metadatos' en estructuras de datos por medio de una interfaz gráfica de usuario y, entonces, se analizan para dar un fragmento de 'DXDT'. En una realización, los 'metadatos' se expresan como una serie de sentencias de sujeto, de predicado, de objeto y de autoridad. En otra realización, las sentencias de 'metadatos' se expresan en cualquiera de diversos formatos. En varias realizaciones, cada sentencia de 'metadatos' se analiza para dar un fragmento separado. En otras realizaciones, varias sentencias de 'metadatos' en un primer formato (tal como expresiones de sujeto, de predicado, de objeto, de autoridad) se analizan para dar un primer fragmento y otras sentencias de 'metadatos' en otros formatos se analizan para dar fragmentos separados. En una realización, las sentencias de 'metadatos' se escriben en un archivo de configuración de XML y el archivo de configuración de XML se analiza para crear los fragmentos dentro de un fragmento de 'DXDT'.

Una realización de un sistema para generar un fragmento de 'DXDT' a partir de una serie de sentencias de 'metadatos' contenidas dentro de un archivo de configuración de XML se muestra en la figura 3.1. El sistema 380 incluye un archivo de configuración de XML 382, que se puede proporcionar a un analizador 384. El archivo de configuración de XML incluye los 'metadatos' codificados como XML. El analizador analiza el XML y genera un fragmento de 'DXDT' 386 al convertir la sentencia de 'metadatos' en fragmentos que se escriben en el fragmento de 'DXDT' de acuerdo con cualquiera de los formatos de fragmento de 'metadatos' descritos anteriormente.

3.3. Generación de un fragmento de 'DMNU'

Un sistema que se puede usar para generar un fragmento de 'DMNU' de acuerdo con una realización de la presente invención se ilustra en la figura 3.2. El sistema de generación de fragmentos de menú 420 requiere como entrada un modelo de medios 422 e información de medios. La información de medios puede adoptar la forma de una fuente de vídeo 424, una fuente de audio 426 y una fuente de superposición 428.

La generación de un fragmento de 'DMNU' usando las entradas al sistema de generación de fragmentos de menú implica la creación de un número de archivos intermedios. El modelo de medios 422 se usa para crear un archivo de configuración de XML 430 y la información de medios se usa para crear un número de archivos AVI 432. El archivo de configuración de XML es creado por un transcodificador de modelos 434. Los archivos AVI 432 se crean al intercalar la información de vídeo, de audio y de superposición usando un intercalador 436. La información de vídeo se obtiene mediante el uso de un descodificador de vídeo 438 y un codificador de vídeo 440 para descodificar la fuente de vídeo 424 y recodificar esta de la forma analizada posteriormente. La información de audio se obtiene mediante el uso de un descodificador de audio 442 y un codificador de audio 444 para descodificar el audio y codificar este de la forma descrita posteriormente. La información de superposición se genera usando un primer transcodificador 446 y un segundo transcodificador 448. El primer transcodificador 446 convierte la superposición en una representación gráfica tal como un mapa de bits convencional y el segundo transcodificador toma la información gráfica y le da formato según se requiere para su inclusión en el archivo multimedia. Una vez que se han generado el archivo de XML y los archivos AVI que contienen la información requerida para construir los menús, el generador de menús 450 puede usar la información para generar un fragmento de 'DMNU' 358'.

3.3.1. El modelo de menú

En una realización, el modelo de medios es un modelo orientado a objetos que representa la totalidad de los menús y sus subcomponentes. El modelo de medios organiza los menús en una estructura jerárquica, lo que permite que los menús se organicen mediante una selección de idioma. Un modelo de medios de acuerdo con una realización de la presente invención se ilustra en la figura 3.3. El modelo de medios 460 incluye un objeto de 'GestorMedios' de nivel más alto 462, que se asocia con un número de objetos de 'MenusIdioma' 463, un objeto de 'Medios' 464 y un objeto de 'TablaTraduccion' 465. El 'GestorMenu' también contiene el idioma de menú por defecto. En una realización, el idioma por defecto puede ser indicado por el código de idioma de dos letras 639 de ISO.

Los objetos de 'MenusIdioma' organizan la información para diversos menús mediante una selección de idioma. La totalidad de los objetos de 'Menú' 466 para un idioma dado se asocian con el objeto de 'MenusIdioma' 463 para ese idioma. Cada objeto de 'Menú' se asocia con un número de objetos de 'Botón' 468 y hace referencia a un número de objetos de 'PistaMedios' 488. Los objetos de 'PistaMedios' 488 a los que se hace referencia indicados el vídeo de

segundo plano y el audio de segundo plano para el objeto de 'Menú' 466.

Cada objeto de 'Botón' 468 se asocia con un objeto de 'Acción' 470 y un objeto de 'Rectángulo' 484. El objeto de 'Botón' 468 también contiene una referencia a un objeto de 'PistaMedios' 488 que indica la superposición que se va a usar cuando se resalta el botón en una visualización. Cada objeto de 'Acción' 470 se asocia con un número de objetos que pueden incluir un objeto de 'TransicionMenu' 472, un objeto de 'TransicionBoton' 474, un objeto de 'VolverAReproduccion' 476, un objeto de 'SeleccionSubtitulo' 478, un objeto de 'SeleccionAudio' 480 y un objeto de 'AccionReproduccion' 482. Cada uno de estos objetos define la respuesta del sistema de menú a diversas entradas procedentes de un usuario. El objeto de 'TransicionMenu' contiene una referencia a un objeto de 'Menú' que indica un menú al que se debería realizar una transición en respuesta a una acción. El objeto de 'TransicionBoton' indica un botón que se debería resaltar en respuesta a una acción. El objeto de 'VolverAReproduccion' puede dar lugar a que un reproductor reanude la reproducción de una característica. Los objetos de 'SeleccionSubtitulo' y de 'SeleccionAudio' contienen referencias a los objetos de 'Título' 487 (analizados posteriormente). El objeto de 'AccionReproduccion' contiene una referencia a un objeto de 'Capítulo' 492 (analizado posteriormente). El objeto de 'Rectángulo' 484 indica la porción de la pantalla ocupada por el botón.

El objeto de 'Medios' 464 indica la información de medios a la que se hace referencia en el sistema de menú. El objeto de 'Medios' tiene un objeto de 'PistasMenu' 486 y un número de objetos de 'Título' 487 asociado con este. El objeto de 'PistasMenu' 486 hace referencia a unos objetos de 'PistaMedios' 488 que son indicativos de los medios usados para construir los menús (es decir, audio de segundo plano, vídeo de segundo plano y superposiciones).

Los objetos de 'Título' 487 son indicativos de una presentación multimedia y tienen un número de objetos de 'Capítulo' 492 y objetos de 'FuenteMedios' 490 asociados con los mismos. Los objetos de 'Título' también contienen una referencia a un objeto de 'ConsultaTraduccion' 494. Los objetos de 'Capítulo' son indicativos de un determinado punto en una presentación multimedia y tienen un número de objetos de 'PistaMedios' 488 asociados con los mismos. Los objetos de 'Capítulo' también contienen una referencia a un objeto de 'ConsultaTraduccion' 494. Cada objeto de 'PistaMedios' asociado con un objeto de 'Capítulo' es indicativo de un punto en una pista o bien de audio, o bien de vídeo o bien de subtítulo de la presentación multimedia y hace referencia a un objeto de 'FuenteMedios' 490 y un objeto de 'ConsultaTraduccion' 494 (analizados posteriormente).

El objeto de 'TablaTraduccion' 465 agrupa un número de cadenas de texto que describen las diversas partes de las presentaciones multimedia indicadas por los objetos de 'Título', los objetos de 'Capítulo' y los objetos de 'PistaMedios'. El objeto de 'TablaTraduccion' 465 tiene un número de objetos de 'ConsultaTraduccion' 494 asociado con este. Cada objeto de 'ConsultaTraduccion' es indicativo de un objeto particular y tiene un número de objetos de 'Traducción' 496 asociado con este. Cada uno de los objetos de 'Traducción' es indicativo de una cadena de texto que describe el objeto indicado por el objeto de 'ConsultaTraduccion' en un idioma particular.

Un modelo de objeto de medios se puede construir usando un software configurado para generar los diversos objetos descritos anteriormente y para establecer las asociaciones y referencias requeridas entre los objetos.

3.3.2. Generación de un archivo de XML

Un archivo de configuración de XML se genera a partir del modelo de menú, que representa la totalidad de los menús y sus subcomponentes. El archivo de configuración de XML también identifica la totalidad de los archivos de medios usados por los menús. El XML se puede generar mediante la implementación de una aplicación analizadora apropiada que analiza el modelo de objeto para dar código de XML.

En otras realizaciones, una aplicación de edición de vídeo puede proporcionar a un usuario una interfaz de usuario que posibilita la generación directa de un archivo de configuración de XML sin crear un modelo de menú.

En algunas realizaciones en las que otro sistema de menú es la base del modelo de menú, tal como un menú de DVD, los menús pueden ser purgados por el usuario para eliminar opciones de menú en relación con un contenido no incluido en el archivo multimedia generado de acuerdo con la práctica de la presente invención. En una realización, esto se puede realizar al proporcionar una interfaz gráfica de usuario que posibilita la eliminación de objetos del modelo de menú. En otra realización, el purgado de los menús se puede lograr al proporcionar una interfaz gráfica de usuario o una interfaz de texto que puede editar el archivo de configuración de XML.

3.3.3. La información de medios

Cuando se genera el fragmento de 'DMNU', la información de medios proporcionada al generador de menús incluye los datos requeridos para proporcionar el vídeo de segundo plano, el audio de segundo plano y las superposiciones de primer plano para los botones especificados en el modelo de menú (véase la descripción anterior). En una realización, se usa una aplicación de edición de vídeo tal como VideoWave distribuida por Roxio, Inc. de Santa Clara, CA para proporcionar las pistas de medios de origen que representan las superposiciones de selección de vídeo, de audio y de botón para cada menú individual.

3.3.4. Generación de archivos AVI intermedios

Como se ha analizado anteriormente, las pistas de medios que se usan como el vídeo de segundo plano, el audio de segundo plano y las superposiciones de botón de primer plano se almacenan en un único archivo AVI para uno o más menús. Los fragmentos que contienen las pistas de medios en un archivo AVI de menú se pueden crear mediante el uso de software diseñado para intercalar pistas de vídeo, de audio y de superposición de botón. Los fragmentos de 'audio', de 'vídeo' y de 'superposición' (es decir, fragmentos de 'subtítulo' que contienen información de superposición) se intercalan en un archivo conforme con el formato AVI usando un intercalador.

Como se ha mencionado anteriormente, se puede crear un archivo AVI separado para cada menú. En otras realizaciones, se podrían usar otros formatos de archivo o un único archivo para contener la información de medios usada para proporcionar la información de audio de segundo plano, de vídeo de segundo plano y de superposición de primer plano.

3.3.5. Combinación del archivo de configuración de XML y los archivos AVI

En una realización, un ordenador está configurado para analizar información a partir del archivo de configuración de XML para crear un fragmento de 'MenuWow' (descrito anteriormente). Además, el ordenador puede crear el fragmento de 'MRIF' (descrito anteriormente) usando los archivos AVI que contienen los medios para cada menú. Entonces, el ordenador puede completar la generación del fragmento de 'DMNU' mediante la creación de las referencias necesarias entre el fragmento de 'MenuWow' y los fragmentos de medios en el fragmento de 'MRIF'. En varias realizaciones, la información de menú se puede cifrar. El cifrado se puede lograr mediante el cifrado de la información de medios contenida en el fragmento de 'MRIF' de una forma similar a la descrita posteriormente en relación con fragmentos de 'vídeo'. En otras realizaciones, se usan diversas técnicas de cifrado alternativas.

3.3.6. Generación automática de menús a partir del modelo de objeto

Haciendo referencia de nuevo a la figura 3.3., un menú que contiene menos contenido que el menú completo se puede generar automáticamente a partir del modelo de menú al simplemente examinar los objetos de 'Título' 487 asociados con el objeto de 'Medios' 464. Los objetos usados para generar automáticamente un menú de acuerdo con una realización de la invención se muestran en la figura 3.3.1. Un software puede generar un archivo de configuración de XML para un menú simple que simplemente posibilita la selección de una sección particular de una presentación multimedia y la selección de las pistas de audio y de subtítulo a usar.

3.3.7. Generación de fragmentos de 'DXDT' y de 'DMNU' usando un único archivo de configuración

Algunos sistemas de acuerdo con varias realizaciones de la presente invención son capaces de generar un único archivo de configuración de XML que contiene tanto 'metadatos' como información de menú y de usar el archivo de XML para generar los fragmentos de 'DXDT' y de 'DMNU'. Estos sistemas obtienen el archivo de configuración de XML usando la información de 'metadatos' y el modelo de objeto de menú. En otras realizaciones, no es necesario que el archivo de configuración esté en XML.

3.4. Generación de fragmentos de 'audio'

Los fragmentos de 'audio' en el fragmento de lista de 'película' de archivos multimedia de acuerdo con algunas realizaciones de la presente invención se pueden generar mediante la decodificación de una fuente de audio y, entonces, la codificación de la fuente para dar fragmentos de 'audio' de acuerdo con la práctica de la presente invención. En una realización, los fragmentos de 'audio' se pueden codificar usando un códec mp3.

3.4.1. Refragmentación de audio

En donde la fuente de audio se proporciona en fragmentos que no contienen información de audio que se corresponde con los contenidos de un fragmento de 'vídeo' correspondiente, entonces algunas realizaciones de la presente invención pueden refragmentar el audio. Un proceso que se puede usar para refragmentar el audio se ilustra en la figura 3.4. El proceso 480 implica identificar (482) un fragmento de 'vídeo', identificar (484) la información de audio que acompaña al fragmento de 'vídeo' y extraer (486) la información de audio a partir de los fragmentos de audio existentes para crear (488) un nuevo fragmento de 'audio'. El proceso se repite hasta que se toma la decisión (490) de que se ha refragmentado la totalidad de la fuente de audio. Punto en el que se ha completado la refragmentación del audio (492).

3.5. Generación de fragmentos de 'vídeo'

Como se ha descrito anteriormente, el proceso de creación de fragmentos de vídeo puede implicar decodificar la fuente de vídeo y codificar el vídeo decodificado para dar fragmentos de 'vídeo'. En una realización, cada fragmento de 'vídeo' contiene información para una única trama de vídeo. El proceso de decodificación simplemente implica tomar vídeo en un formato particular y decodificar el vídeo a partir de ese formato a un formato de vídeo convencional, que se puede descomprimir. El proceso de codificación implica tomar el vídeo convencional, codificar el vídeo y

generar fragmentos de 'vídeo' usando el vídeo codificado.

Un codificador de vídeo de acuerdo con una realización de la presente invención se ilustra conceptualmente en la figura 3.5. El codificador de vídeo 500 preprocesa 502 la información de vídeo convencional 504. Entonces se realiza una estimación de movimiento 506 sobre el vídeo preprocesado para proporcionar una compensación de movimiento 508 al vídeo preprocesado. Se realiza una transformada de coseno discreta (transformación DCT) 510 sobre el vídeo de movimiento compensado. A continuación de la transformación DCT, se cuantifica 512 el vídeo y se realiza una predicción inversa 514. Entonces se genera una secuencia de bits comprimida 516 mediante la combinación de una versión de código 518 de textura del vídeo con una codificación de movimiento 520 generada usando los resultados de la estimación de movimiento. Entonces, la secuencia de bits comprimida se usa para generar los fragmentos de 'vídeo'.

Con el fin de realizar una estimación de movimiento 506, el sistema ha de tener conocimiento de cómo la trama de vídeo previamente procesada será descodificada por un dispositivo de descodificación (por ejemplo, cuando el vídeo comprimido se descomprime para su visionado por un reproductor). Esta información se puede obtener mediante la cuantificación inversa 522 de la salida del cuantificador 512. Entonces, se puede realizar una DCT inversa 524 sobre la salida del cuantificador inverso y el resultado colocarse en un almacén de tramas 526 para el acceso al mismo durante el proceso de estimación de movimiento.

Los archivos multimedia de acuerdo con algunas realizaciones de la presente invención también pueden incluir un número de potenciaciones psicovisuales 528. Las potenciaciones psicovisuales pueden ser métodos de compresión de vídeo basándose en las percepciones humanas de la visión. Estas técnicas se analizan más adelante y, en general, implican modificar el número de bits usados por el cuantificador para representar diversos aspectos del vídeo. Otros aspectos del proceso de codificación también pueden incluir potenciaciones psicovisuales.

En una realización, la totalidad del sistema de codificación 500 se puede implementar usando un ordenador configurado para realizar las diversas funciones descritas anteriormente. Posteriormente se proporcionan algunos ejemplos de implementaciones detalladas de estas funciones.

3.5.1. Preprocesamiento

Las operaciones de preprocesamiento 502 que son realizadas de forma opcional por un codificador 500 de acuerdo con una realización de la presente invención pueden usar un número de técnicas de procesamiento de señal para mejorar la calidad del vídeo codificado. En una realización, el preprocesamiento 502 puede implicar uno o todos de entre desentrelazado, reducción de ruido espacial/temporal y redimensionamiento. En algunas realizaciones en las que se usan todas estas tres técnicas de preprocesamiento, por lo general el desentrelazado se realiza en primer lugar, seguido por la reducción de ruido espacial/temporal y el redimensionamiento.

3.5.2. Estimación y compensación de movimiento

Un codificador de vídeo de acuerdo con una realización de la presente invención puede reducir el número de píxeles requeridos para representar una pista de vídeo al buscar píxeles que se repiten en múltiples tramas. En esencia, por lo general cada trama en un vídeo contiene muchos de los mismos píxeles que la anterior a esta. El codificador puede realizar varios tipos de búsquedas para coincidencias en píxeles entre cada trama (como macrobloques, píxeles, medios píxeles y cuartos de píxel) y elimina estas redundancias siempre que sea posible sin reducir la calidad de imagen. Usando una estimación de movimiento, el codificador puede representar la mayor parte de la instantánea simplemente mediante el registro de los cambios que han tenido lugar desde la última trama en lugar de almacenar la totalidad de la instantánea para cada trama. Durante la estimación de movimiento, el codificador divide la trama que está analizando en cuadrícula uniforme de bloques, a la que se hace referencia a menudo como 'macrobloques'. Para cada 'macrobloque' en la trama, el codificador puede intentar hallar un bloque coincidente en la trama previa. El proceso de intentar hallar bloques coincidentes se denomina 'búsqueda de movimiento'. El movimiento del 'macrobloque' se puede representar como un vector bidimensional, es decir, una representación (x, y). El algoritmo de búsqueda de movimiento se puede realizar con diversos grados de precisión. Una búsqueda de píxeles completos es una en la que el codificador intentará localizar bloques coincidentes al recorrer gradualmente la trama de referencia en una cualquiera de las dimensiones, un píxel de cada vez. En una búsqueda de medios píxeles, el codificador busca un bloque coincidente al recorrer gradualmente la trama de referencia en una cualquiera de las dimensiones, medio píxel de cada vez. El codificador puede usar cuartos de píxel, otras fracciones de píxel, o búsquedas que implican una granularidad de más de un píxel.

La realización de codificador ilustrada en la figura 3.5. realiza una estimación de movimiento de acuerdo con una realización de la presente invención. Durante la estimación de movimiento, el codificador tiene acceso al vídeo preprocesado 502 y la trama previa, que se almacena en un almacén de tramas 526. La trama previa se genera al tomar la salida del cuantificador, realizando una cuantificación inversa 522 y una transformación DCT inversa 524. La razón de realizar las funciones inversas es de tal modo que la trama en el almacén de tramas es como aparecería cuando es descodificada por un reproductor de acuerdo con una realización de la presente invención.

La compensación de movimiento se realiza al tomar los bloques y los vectores generados como resultado de la

estimación de movimiento. El resultado es una aproximación de la imagen codificada que se puede poner en coincidencia con la imagen real al proporcionar información de textura adicional.

3.5.3. Transformada de coseno discreta

La DCT y la DCT inversa realizadas por el codificador ilustrado en la figura 3.5. son de acuerdo con la norma especificada en ISO/IEC 14496-2:2001(E), Anexo A.1 (transformadas de codificación).

3.5.3.1. Descripción de transformada

La DCT es un método de transformación de un conjunto de puntos de datos en el dominio espacial en una representación en el dominio de la frecuencia. En el caso de la compresión de vídeo, una DCT bidimensional convierte bloques de imagen en una forma en la que las redundancias son más fácilmente aprovechables. Un bloque en el dominio de la frecuencia puede ser una matriz dispersa que se comprime fácilmente mediante codificación por entropía.

3.5.3.2. Potenciones psicovisuales para transformada

Los coeficientes de DCT se pueden modificar para mejorar la calidad de la imagen cuantificada mediante la reducción del ruido de cuantificación en áreas en donde este es inmediatamente evidente para un espectador humano. Además, el tamaño de archivo se puede reducir mediante el aumento del ruido de cuantificación en porciones de la imagen en donde este no es discernible fácilmente por un espectador humano.

Los codificadores de acuerdo con una realización de la presente invención pueden realizar aquello a lo que se hace referencia como potenciación psicovisual 'lenta'. La potenciación psicovisual 'lenta' analiza bloques de la imagen de vídeo y decide si permitir algo de ruido, se pueden ahorrar algunos bits sin deteriorar la apariencia del vídeo. El proceso usa una métrica por bloque. Se hace referencia al proceso como un proceso 'lento', debido a que este realiza una cantidad considerable de cómputo para evitar los artefactos de aparición de bloques o de aparición de anillos.

Otras realizaciones de codificadores de acuerdo con algunas realizaciones de la presente invención implementan una potenciación psicovisual 'rápida'. La potenciación psicovisual 'rápida' es capaz de controlar en dónde aparece ruido dentro de un bloque y puede dar forma al ruido de cuantificación.

Las potenciones psicovisuales tanto 'lenta' como 'rápida' se analizan con mayor detalle posteriormente. Se pueden realizar otras potenciones psicovisuales de acuerdo con algunas realizaciones de la presente invención que incluyen potenciones que controlan el ruido en los bordes de imagen y que buscan concentrar unos niveles más altos de ruido de cuantificación en áreas de la imagen en donde este no es inmediatamente evidente para la visión humana.

3.5.3.3. Potenciación psicovisual 'lenta'

La potenciación psicovisual 'lenta' analiza bloques de la imagen de vídeo y determina si permitir algo de ruido, puede ahorrar bits sin deteriorar la apariencia del vídeo. En una realización, el algoritmo incluye dos fases. La primera implica la generación de una imagen diferenciada para los píxeles de luminancia de entrada. La imagen diferenciada se genera de la forma descrita posteriormente. La segunda fase implica modificar los coeficientes de DCT antes de la cuantificación.

3.5.3.3.1. Generación de imagen diferenciada

Cada píxel p'_{xy} de la imagen diferenciada se computa a partir de los píxeles de origen descomprimidos, p_{xy} , de acuerdo con lo siguiente:

$$P'_{xy} = \max(|p_{x+1y} - p_{xy}|, |p_{x-1y} - p_{xy}|, |p_{x+1y} - p_{xy}|, |p_{x-1y} - p_{xy}|)$$

en donde

p'_{xy} estará en el intervalo de 0 a 255 (suponiendo vídeo de 8 bits),

3.5.3.3.2. Modificación de coeficientes de DCT

La modificación de los coeficientes de DCT puede implicar el cómputo de un factor de aparición de anillos de bloque, el cómputo de la energía de bloque y la modificación real de los valores de coeficiente.

3.5.3.3.3. Cómputo de factor de aparición de anillos de bloque

Para cada bloque de la imagen, se calcula un "factor de aparición de anillos" basándose en la región local de la imagen diferenciada. En algunas realizaciones en las que el bloque se define como un bloque 8 x 8, el factor de aparición de

anillos se puede determinar usando el siguiente método.

Inicialmente, se determina un umbral basándose en los valores de los píxeles de luminancia máxima y mínima dentro del bloque 8 x 8:

$$5 \quad \text{umbral}_{\text{bloque}} = \text{suelo} ((\text{máx}_{\text{bloque}} - \text{mín}_{\text{bloque}}) / 8) + 2$$

La imagen diferenciada y el umbral se usan para generar una correlación los píxeles "planos" en las proximidades del bloque. El potencial de que cada bloque tenga un umbral diferente evita la creación de una correlación píxeles planos para la totalidad de la trama. La correlación se genera como sigue:

$$10 \quad \text{plano}_{xy} = 1 \text{ cuando } p'_{xy} < \text{umbral}_{\text{bloque}}$$

$$15 \quad \text{plano}_{xy} = 0 \text{ de lo contrario}$$

La correlación de píxeles planos se filtra de acuerdo con una operación lógica simple:

$$20 \quad \text{plano}'_{xy} = 1 \text{ cuando } \text{plano}_{xy} = 1 \text{ y } \text{plano}_{x-1y} = 1 \text{ y } \text{plano}_{xy-1} = 1 \text{ y } \text{plano}_{x-1y-1} = 1$$

$$\text{plano}'_{xy} \text{ de lo contrario}$$

Entonces, los píxeles planos en la correlación filtrada se cuentan a lo largo de la región 9 x 9 que cubre el bloque 8 x 8.

$$\text{recuentoplano}_{\text{bloque}} = \sum \text{plano}'_{xy} \text{ para } 0 = x = 8 \text{ y } 0 = y = 8$$

25 El riesgo de artefactos de aparición de anillos visibles se puede evaluar usando la siguiente expresión:

$$\text{riesgoaparicionanillos}_{\text{bloque}} = ((\text{recuentoplano}_{\text{bloque}} - 10) \times 256 + 20) / 40$$

Entonces, el factor de aparición de anillos del bloque 8 x 8 se puede obtener usando la siguiente expresión:

$$30 \quad \text{Factoraparicionanillos} = 0 \text{ cuando } \text{riesgoaparicionanillos} > 255$$

$$= 255 \text{ cuando } \text{riesgoaparicionanillos} < 0$$

$$= 255 - \text{riesgoaparicionanillos} \text{ de lo contrario}$$

35 3.5.3.3.4. Cómputo de energía de bloque

La energía para los bloques de la imagen se puede calcular usando el siguiente procedimiento. En varias realizaciones, se usan bloques 8 x 8 de la imagen.

40 Se realiza una DCT directa sobre la imagen de origen:

$$T = fDCT(S)$$

45 en donde S son los 64 valores de luminancia de imagen de origen del bloque 8 x 8 en cuestión y T es la versión transformada de la misma porción de la imagen de origen.

La energía en una posición de coeficiente particular se define como el cuadrado del valor de ese coeficiente:

$$50 \quad e_k = t_k^2 \text{ para } 0 = k = 63$$

en donde t_k es el k -ésimo coeficiente del bloque transformado T .

3.5.3.3.5. Modificación de coeficientes

55 La modificación de los coeficientes de DCT se puede realizar de acuerdo con el siguiente proceso. En varias realizaciones, el proceso se realiza para cada coeficiente de DCT de CA no nula antes de la cuantificación. La magnitud de cada coeficiente se cambia en un delta pequeño, determinándose el valor del delta de acuerdo con técnicas psicovisuales.

60 La modificación de coeficientes de DCT de cada coeficiente de CA no nulo c_k se realiza al calcular una energía basándose en las energías local y de bloque usando la siguiente fórmula:

$$\text{energía}_k = \text{máx} (a_k \times e_k, 0, 12 \times \text{energíatotal})$$

65 en donde a_k es una constante cuyo valor depende de la posición de coeficiente como se describe en la siguiente tabla:

Tabla 3 - Tabla de coeficientes

0,0	1,0	1,5	2,0	2,0	2,0	2,0	2,0
1,0	1,5	2,0	2,0	2,0	2,0	2,0	2,0
1,5	2,0	2,0	2,0	2,0	2,0	2,0	2,0
2,0	2,0	2,0	2,0	2,0	2,0	2,0	2,0
2,0	2,0	2,0	2,0	2,0	2,0	2,0	2,0
2,0	2,0	2,0	2,0	2,0	2,0	2,0	2,0
2,0	2,0	2,0	2,0	2,0	2,0	2,0	2,0
2,0	2,0	2,0	2,0	2,0	2,0	2,0	2,0

La energía se puede modificar de acuerdo con el factor de aparición de anillos del bloque usando la siguiente relación:

$$energía'_k = factor_{aparicionanillos} \times energía_k$$

El valor resultante se desplaza y se recorta antes de usarse como una entrada a una tabla de consulta (LUT).

$$e_k = \min(1023, 4 \times energía'_k)$$

$$d_k = LUT_i \text{ en donde } i = e_k$$

La tabla de consulta se computa como sigue:

$$LUT_i = \min(suelo(k_{textura} \times ((i + 0,5) / 4)^{1/2} + k_{plano} \times desplazamiento), 2 \times Q_p)$$

El valor ' $desplazamiento$ ' depende de un cuantificador, Q_p , como se describe en la siguiente tabla:

Tabla 4 - $desplazamiento$ como una función de los valores de Q_p

Q_p	$desplazamiento$	Q_p	$desplazamiento$
1	-0,5	16	8,5
2	1,5	17	7,5
3	1,0	18	9,5
4	2,5	19	8,5
5	1,5	20	10,5
6	3,5	21	9,5
7	2,5	22	11,5
8	4,5	23	10,5
9	3,5	24	12,5
10	5,5	25	11,5
11	4,5	26	13,5
12	6,5	27	12,5
13	5,5	28	14,5
14	7,5	29	13,5
15	6,5	30	15,5
		31	14,5

Las variables $k_{textura}$ y k_{plano} controlan la intensidad del efecto psicovisual en regiones planas y con textura, respectivamente. En una realización, estas toman valores en el intervalo de 0 a 1, con 0 significando sin efecto y 1 queriendo decir pleno efecto. En una realización, los valores para $k_{textura}$ y k_{plano} se establecen como sigue:

Luminancia:

$$k_{textura} = 1,0$$

$$k_{plano} = 1,0$$

Crominancia:

5 $k_{textura} = 1,0$

$$k_{plano} = 0,0$$

10 La salida a partir de la tabla de consulta (d_k) se usa para modificar la magnitud del coeficiente de DCT mediante un proceso aditivo:

$$c'_k = c_k - \min(d_k, |c_k|) \times \text{sgn}(c_k)$$

15 Por último, el coeficiente de DCT c_k es sustituido por el coeficiente modificado c'_k y se transmite hacia adelante para la cuantificación.

3.5.3.4. Potenciación psicovisual 'rápida'

20 Se puede realizar una potenciación psicovisual 'rápida' sobre los coeficientes de DCT mediante el cómputo de una correlación de 'importancia' para los píxeles de luminancia de entrada y, entonces, la modificación de los coeficientes de DCT.

3.5.3.4.1. Cómputo de una correlación de 'importancia'

25 Una correlación de 'importancia' se puede generar al calcular un valor de 'importancia' para cada píxel en el lugar de la luminancia de la trama de vídeo de entrada. En varias realizaciones, el valor de 'importancia' aproxima la sensibilidad del ojo humano a cualquier distorsión situada en ese píxel particular. La correlación de 'importancia' es una distribución de valores de 'importancia' de píxel.

30 La 'importancia' de un píxel se puede determinar al calcular, en primer lugar, el intervalo dinámico de un bloque de píxeles que rodean al píxel (d_{xy}). En varias realizaciones, el intervalo dinámico de un bloque 3 x 3 de píxeles centrado en la ubicación de píxel (x, y) se computa al restar el valor del píxel más oscuro en el área del valor del píxel más claro en el área.

35 La 'importancia' de un píxel (m_{xy}) se puede obtener a partir del intervalo dinámico del píxel como sigue:

$$m_{xy} = 0,08 / \max(d_{xy}, 3) + 0,001$$

3.5.3.4.2. Modificación de coeficientes de DCT

40 En una realización, la modificación de los coeficientes de DCT implica la generación de matrices de energía de función de base y tablas de consulta de delta.

3.5.3.4.3. Generación de matrices de energía de función de base

45 Se puede usar un conjunto de matrices de energía de función de base en la modificación de los coeficientes de DCT. Estas matrices contienen valores constantes que se pueden computar antes de la codificación. Se usa una matriz 8 x 8 para cada una de las 64 funciones de base DCT. Cada matriz describe cómo cada píxel en un bloque 8 x 8 se verá afectado por una modificación de su coeficiente correspondiente. La matriz de k-ésima energía de función de base se obtiene al tomar una matriz 8 x 8 A_k con el coeficiente correspondiente establecido a 100 y los otros coeficientes establecidos a 0.

$$a_{kn} = 100 \text{ si } n = k \\ = 0 \text{ de lo contrario}$$

55 en donde
n representa la posición de coeficiente dentro de la matriz 8 x 8; $0 = n = 63$

60 Se realiza una DCT inversa sobre la matriz para dar una matriz 8 x 8 adicional A'_k . Los elementos de la matriz (a'_{kn}) representan la k-ésima función de base DCT.

$$A'_k = iDCT(A_k)$$

Cada valor en la matriz transformada se eleva entonces al cuadrado:

65 $b_{kn} = a'^2_{kn}$ para $0 = n = 63$

El proceso se lleva a cabo 64 veces para producir las matrices de energía de función de base B_k , $0 = k = 63$, comprendiendo cada una 64 valores naturales. Cada valor de matriz es una medida de cuánto un píxel en la n -ésima posición en el bloque 8×8 se verá afectado por cualquier error o modificación del coeficiente k .

3.5.3.4.4. Generación de tabla de consulta de delta

Se puede usar una tabla de consulta (LUT) para agilizar el cómputo del delta de modificación de coeficientes. Los contenidos de la tabla se pueden generar de una forma que depende de la intensidad deseada de la potenciación psicovisual 'rápida' y el parámetro de cuantificador (Q_p).

Los valores de la tabla de consulta se pueden generar de acuerdo con la siguiente relación:

$$LUT_i = \text{mín} (\text{suelo} (128 \times k_{\text{textura}} \times \text{intensidad} / (i + 0,5) + k_{\text{plano}} \times \text{desplazamiento} + 0,5), 2 \times Q_p)$$

en donde

i es la posición dentro de la tabla, $0 = i = 1023$.

intensidad y desplazamiento dependen del cuantificador, Q_p , como se describe en la siguiente tabla:

Tabla 5 - Relación entre valores de intensidad y desplazamiento y el valor de Q_p

Q_p	intensidad	desplazamiento	Q_p	intensidad	desplazamiento
1	0,2	-0,5	16	2,0	8,5
2	0,6	1,5	17	2,0	7,5
3	1,0	1,0	18	2,0	9,5
4	1,2	2,5	19	2,0	8,5
5	1,3	1,5	20	2,0	10,5
6	1,4	3,5	21	2,0	9,5
7	1,6	2,5	22	2,0	11,5
8	1,8	4,5	23	2,0	10,5
9	2,0	3,5	24	2,0	12,5
10	2,0	5,5	25	2,0	11,5
11	2,0	4,5	26	2,0	13,5
12	2,0	6,5	27	2,0	12,5
13	2,0	5,5	28	2,0	14,5
14	2,0	7,5	29	2,0	13,5
15	2,0	6,5	30	2,0	15,5
			31	2,0	14,5

Las variables k_{textura} y k_{plano} controlan la intensidad del efecto psicovisual en regiones planas y con textura, respectivamente. En una realización, estas toman valores en el intervalo de 0 a 1, con 0 significando sin efecto y 1 queriendo decir pleno efecto. En una realización, los valores para k_{textura} y k_{plano} se establecen como sigue:

Luminancia:

$$k_{\text{textura}} = 1,0$$

$$k_{\text{plano}} = 1,0$$

Crominancia:

$$k_{\text{textura}} = 1,0$$

$$k_{\text{plano}} = 0,0$$

3.5.3.4.5. Modificación de coeficientes de DCT

Los coeficientes de DCT se pueden modificar usando los valores calculados anteriormente. En una realización, cada coeficiente de DCT de CA no nula se modifica de acuerdo con el siguiente procedimiento antes de la cuantificación.

Inicialmente, un valor de 'energía' (e_k) se computa al tomar el producto escalar de la matriz de energía de función de base correspondiente y el bloque 8×8 apropiado a partir de la correlación de importancia. Esta 'energía' es una medida de cómo los errores de cuantificación en el coeficiente particular serían percibidos por un espectador humano. Esta es la suma del producto de la importancia de píxel y la energía de función de base de píxel:

$$e_k = M \cdot B_k$$

en donde

M contiene los valores de correlación de importancia del bloque 8 x 8; y
 B_k es la k -ésima matriz de energía de función de base.

El valor de 'energía' resultante se desplaza y se recorta antes de usarse como un índice (d_k) en la tabla de consulta de delta.

$$e'_k = \min [1023, \text{suelo} (e_k / 32768)]$$

$$d_k = LUT_i$$

en donde

$$i = e'_k$$

La salida de la tabla de consulta de delta se usa para modificar la magnitud del coeficiente de DCT mediante un proceso aditivo:

$$c'_k = c_k - \min (d_k, |c_k|) \times \text{sign} (c_k)$$

El coeficiente de DCT c_k se sustituye con el c'_k modificado y se transmite hacia adelante para la cuantificación.

3.5.4. Cuantificación

Los codificadores de acuerdo con algunas realizaciones de la presente invención pueden usar un cuantificador convencional tal como al cuantificador definido por la Unión Internacional de Telecomunicaciones como Codificación de Vídeo para Comunicaciones de Tasa de Bits Baja, Recomendación de UIT-T H.263, 1996.

3.5.4.1. Potenciaciones psicovisuales para cuantificación

Algunos codificadores de acuerdo con algunas realizaciones de la presente invención, usan una potenciación psicovisual que aprovecha los efectos psicológicos de la visión humana para lograr una compresión más eficiente. El efecto psicovisual se puede aplicar a nivel de trama y a nivel de macrobloque.

3.5.4.2. Potenciaciones psicovisuales a nivel de trama

Cuando se aplica a nivel de trama, la potenciación es parte del algoritmo de control de tasa y su objetivo es ajustar la codificación de tal modo que una cantidad dada de tasa de bits se usa del mejor modo para asegurar la máxima calidad visual como es percibida por los ojos humanos. La potenciación psicovisual de tasa de tramas está motivada por la teoría de que la visión humana tiende a ignorar los detalles cuando la acción es alta y que la visión humana tiende a percibir el detalle cuando una imagen es estática. En una realización, la cantidad de movimiento se determina al mirar la suma de diferencias absolutas (SAD) para una trama. En una realización, el valor de SAD se determina mediante la suma de las diferencias absolutas de los píxeles de luminancia de ubicación conjunta de dos bloques. En varias realizaciones, se usan las diferencias absolutas de bloques de píxeles 16 x 16. En algunas realizaciones que abordan desplazamientos de píxel fraccionarios, se realiza una interpolación como se especifica en la norma MPEG-4 (una norma de ISO/IEC desarrollada por el Grupo de Expertos en Imágenes en Movimiento de ISO/IEC), antes de que se haya calculado la suma de diferencias absolutas.

La potenciación psicovisual a nivel de trama solo es de aplicación a las tramas P de la pista de vídeo y se basa en el valor de SAD de la trama. Durante la codificación, el módulo psicovisual mantiene un registro del SAD promedio (es decir, $\overline{SAD}$) de la totalidad de las tramas P de la pista de vídeo y la distancia promedio del SAD de cada trama con respecto a su SAD global (es decir, $\overline{DSAD}$). El promediado se puede realizar usando un algoritmo de promedio móvil exponencial. En una realización, el algoritmo de control de tasa de una pasada descrito anteriormente se puede usar como el periodo de promediado en el presente caso (véase la descripción anterior).

Para cada trama P de la pista de vídeo codificada, el cuantificador de trama Q (obtenido a partir del módulo de control de tasa) tendrá una corrección psicovisual aplicada al mismo. En una realización, el proceso implica calcular una relación R usando la siguiente fórmula:

$$R = \frac{SAD - \overline{SAD}}{\overline{DSAD}} - I$$

en donde

I es una constante y se establece actualmente a 0,5. La R se recorta para quedar dentro de la cota de $[-1, 1]$.

Entonces, el cuantificador se ajusta de acuerdo con la relación R , por medio del cálculo mostrado posteriormente:

$$Q_{adj} = Q \cdot [Q \cdot (1 + R \cdot S_{trama})]$$

en donde

S_{trama} es una constante de intensidad para las potenciaciones psicovisuales a nivel de trama.

La constante S_{trama} determina lo fuerte que puede ser un ajuste para la psicovisual a nivel de trama. En una realización del códec, está disponible la opción de ajustar S_{trama} a 0,2, 0,3 o 0,4.

3.5.4.3. Potenciaciones psicovisuales a nivel de macrobloque

Los codificadores de acuerdo con algunas realizaciones de la presente invención que utilizan una potenciación psicovisual en el intento a nivel de macrobloque para identificar los macrobloques que son prominentes para la calidad visual del vídeo para un espectador humano e intentan codificar esos macrobloques con una calidad más alta. El efecto de las potenciaciones psicovisuales a nivel de macrobloque es quitar bits de las partes menos importantes de una trama y aplicar los mismos a partes más importantes de la trama. En varias realizaciones, se logran potenciaciones usando tres tecnologías, que se basan en la suavidad, el brillo y el SAD de macrobloque. En otras realizaciones, se puede usar cualquiera de las técnicas, solas o en combinación con otra de las técnicas u otra técnica enteramente.

En una realización, todas las tres potenciaciones psicovisuales a nivel de macrobloque descritas anteriormente comparten un parámetro común, S_{MB} , que controla la intensidad de la potenciación psicovisual a nivel de macrobloque. Entonces, el cuantificador máximo y mínimo para los macrobloques se obtienen a partir del parámetro de intensidad y el cuantificador de trama Q_{trama} por medio de los cálculos mostrados posteriormente:

$$Q_{MBMáx} = \frac{Q_{trama}}{(1 - S_{MB})},$$

y

$$Q_{MBMín} = Q_{trama} \cdot (1 - S_{MB})$$

en donde

$Q_{MBMáx}$ es el cuantificador máximo

$Q_{MBMín}$ es el cuantificador mínimo

Los valores $Q_{MBMáx}$ y $Q_{MBMín}$ definen las cotas superior e inferior para los cuantificadores de macrobloque para la totalidad de la trama. En una realización, se proporciona la opción de ajustar el valor S_{MB} a cualquiera de los valores 0,2, 0,3 y 0,4. En otras realizaciones, se pueden utilizar otros valores para S_{MB} .

3.5.4.3.1. Potenciación de brillo

En algunas realizaciones en las que se realiza una potenciación psicovisual basándose en el brillo de los macrobloques, el codificador intenta codificar los macrobloques más brillantes con una calidad más grande. La base teórica de esta potenciación es que las partes relativamente oscuras de la trama son, más o menos, ignoradas por los espectadores humanos. Esta potenciación psicovisual de macrobloque se aplica a las tramas I y a las tramas P de la pista de vídeo. Para cada trama, el codificador examina la totalidad de la trama en primer lugar. Se calcula el brillo promedio (BR) y también se calcula la diferencia promedio de brillo a partir del (DBR) promedio. Entonces, estos valores se usan para desarrollar dos umbrales ($T_{BRInferior}$, $T_{BRSuperior}$), que se pueden usar como indicadores para si se debería aplicar la potenciación psicovisual:

$$T_{BRInferior} = \overline{BR} - \overline{DBR}$$

$$T_{BRInferior} = \overline{BR} + (\overline{BR} - T_{BRInferior})$$

Entonces, la potenciación de brillo se aplica basándose en los dos umbrales usando las condiciones expuestas posteriormente para generar un cuantificador previsto (Q_{MB}) para el macrobloque:

$$Q_{MB} = Q_{MBMín} \text{ cuando } BR > T_{BRSuperior}$$

$$Q_{MB} = Q_{trama} \text{ cuando } T_{BRInferior} < BR \leq T_{BRSuperior},$$

y

$$Q_{MB} = Q_{MBM\acute{a}x} \text{ cuando } BR < T_{BRInferior}$$

en donde

BR es el valor de brillo para ese macrobloque particular

En algunas realizaciones en las que el codificador es conforme con la norma MPEG-4, la técnica de potenciación de brillo psicovisual a nivel de macrobloque no puede cambiar, de un macrobloque al siguiente, el cuantificador en más de ± 2 . Por lo tanto, el Q_{MB} calculado puede requerir una modificación basándose en el cuantificador usado en el macrobloque previo.

3.5.4.3.2. Potenciación de suavidad

Los codificadores de acuerdo con algunas realizaciones de la presente invención que incluyen una potenciación psicovisual de suavidad, modifican el cuantificador basándose en la variación espacial de la imagen que se está codificando. El uso de una potenciación psicovisual de suavidad puede estar motivado por la teoría de que la visión humana tiene una sensibilidad aumentada a los artefactos de cuantificación en partes suaves de una imagen. Por lo tanto, la potenciación psicovisual de suavidad puede implicar aumentar el número de bits para representar porciones más suaves de la imagen y disminuir el número de bits en donde hay un grado alto de variación espacial en la imagen.

En una realización, la suavidad de una porción de una imagen se mide como la diferencia promedio en la luminancia de píxeles en un macrobloque con respecto al brillo del macrobloque ($\overline{DR}$). Un método de realización de una potenciación psicovisual de suavidad en una trama I de acuerdo con algunas realizaciones de la presente invención se muestra en la figura 3.6. El proceso 540, implica examinar la totalidad de la trama para calcular (542) $\overline{DR}$. Entonces, el umbral para aplicar la potenciación de suavidad, T_{DR} , se puede obtener (544) usando el siguiente cálculo:

$$T_{DR} = \frac{\overline{DR}}{2}$$

La siguiente potenciación de suavidad se realiza (546) basándose en el umbral.

$$Q_{MB} = Q_{trama} \text{ cuando } DR \geq T_{DR},$$

y

$$Q_{MB} = Q_{MBMin} \text{ cuando } DR < T_{DR}$$

en donde

Q_{MB} es el cuantificador previsto para el macrobloque

DR es el valor de desviación para el macrobloque (es decir, luminancia media - brillo medio)

Algunas realizaciones que codifican archivos de acuerdo con la norma MPEG-4 están limitadas, como se ha descrito anteriormente, en que el cambio de cuantificador a nivel de macrobloque puede ser, como mucho, ± 2 de un macrobloque al siguiente.

3.5.4.3.3. Potenciación de SAD de macrobloque

Los codificadores de acuerdo con algunas realizaciones de la presente invención pueden utilizar una potenciación psicovisual de SAD de macrobloque. Se puede usar una potenciación psicovisual de SAD de macrobloque para aumentar el detalle para macrobloques estáticos y permitir un detalle disminuido en porciones de una trama que se usan en una escena de acción alta.

Un proceso para realizar una potenciación psicovisual de SAD de macrobloque de acuerdo con una realización de la presente invención se ilustra en la figura 3.7. El proceso 570 incluye inspeccionar (572) la totalidad de una trama I para determinar el SAD promedio (es decir, $\overline{MBSAD}$) para la totalidad de los macrobloques en la totalidad de la trama y también se obtiene la diferencia promedio del SAD de un macrobloque con respecto al promedio (es decir, $\overline{MBSAD}$). En una realización, ambos de estos macrobloques se promedian a lo largo de los macrobloques codificados entre tramas (es decir, los macrobloques codificados usando una compensación de movimiento u otras dependencias en tramas de vídeo codificadas previas). Entonces, se obtienen (574) dos umbrales para aplicar la potenciación de SAD de macrobloque a partir de estos promedios usando las siguientes fórmulas:

$$T_{MBSADInferior} = \overline{MBSAD} - \overline{DMBSAD},$$

y

$$T_{MBSADSuperior} = \overline{MBSAD} + \overline{DMBSAD},$$

en donde

$T_{MBSADInferior}$ es el umbral inferior

$T_{MBSADSuperior}$ es el umbral superior, que puede estar acotado por 1024 si es necesario

Entonces, la potenciación de SAD de macrobloque se aplica (576) basándose en estos dos umbrales de acuerdo con las siguientes condiciones:

$$Q_{MB} = Q_{MBMáx} \text{ cuando } MBSAD > T_{MBSADSuperior},$$

$$Q_{MB} = Q_{trama} \text{ cuando } T_{MADInferior} \leq MBSAD \leq T_{MBSADSuperior}$$

$$Q_{MB} = Q_{MBMín} \text{ cuando } MBSAD < T_{MSADInferior}$$

en donde

Q_{MB} es el cuantificador previsto para el macrobloque

MBSAD es el valor de SAD para ese macrobloque particular

Algunas realizaciones que codifican archivos de acuerdo con la especificación de MPEG-4 están limitadas, como se ha descrito anteriormente, en que el cambio de cuantificador a nivel de macrobloque puede ser, como mucho, ± 2 de un macrobloque al siguiente.

3.5.5. Control de tasa

La técnica de control de tasa usada por un codificador de acuerdo con una realización de la presente invención puede determinar cómo usa el codificador la tasa de bits asignada para codificar una secuencia de vídeo. Por lo general, un codificador buscará codificar a una tasa de bits predeterminada y la técnica de control de tasa es responsable de hacer coincidir la tasa de bits generada por el codificador tan cerca como es posible de la tasa de bits predeterminada. La técnica de control de tasa también puede buscar asignar la tasa de bits de una forma que asegure la más alta calidad visual de la secuencia de vídeo cuando se descodifica la misma. Gran parte del control de tasa se realiza mediante el ajuste del cuantificador. El cuantificador determina lo finamente que codifica el codificador la secuencia de vídeo. Un cuantificador menor dará como resultado una calidad más alta y un consumo de bits más alto. Por lo tanto, el algoritmo de control de tasa busca modificar el cuantificador de una forma que equilibre los intereses en conflicto de la calidad de vídeo y el consumo de bits.

Los codificadores de acuerdo con algunas realizaciones de la presente invención pueden utilizar cualquiera de diversas técnicas de control de tasa diferentes. En una realización, se usa una técnica de control de tasa de una única pasada. En otras realizaciones, se usa una técnica de control de tasa de pasada doble (o múltiple). Además, se puede realizar un control de tasa de 'memoria intermedia de vídeo verificada' según se requiera. Algunos ejemplos específicos de estas técnicas se analizan posteriormente. No obstante, se puede usar cualquier técnica de control de tasa en un codificador de acuerdo con la práctica de las presentes invenciones.

3.5.5.1. Control de tasa de una pasada

Una realización de una técnica de control de tasa de una pasada de acuerdo con una realización de la presente invención busca permitir picos altos de tasa de bits para escenas de movimiento alto. En varias realizaciones, la técnica de control de tasa de una pasada busca aumentar la tasa de bits lentamente en respuesta a un aumento en la cantidad de movimiento en una escena y disminuir rápidamente la tasa de bits en respuesta a una reducción en el movimiento en una escena.

En una realización, el algoritmo de control de tasa de una pasada usa dos periodos de promediado para realizar un seguimiento de la tasa de bits. Un promedio a largo plazo para asegurar la convergencia de tasa de bits global y un promedio a corto plazo para habilitar una respuesta a variaciones en la cantidad de acción en una escena.

Una técnica de control de tasa de una pasada de acuerdo con una realización de la presente invención se ilustra en la figura 3.8. La técnica de control de tasa de una pasada comienza (582) mediante la inicialización (584) del codificador con una tasa de bits deseada, la tasa de tramas de vídeo y otros diversos parámetros (analizados más adelante). Se almacena una variable en coma flotante, que es indicativa del cuantificador. Si una trama requiere una cuantificación (586), entonces la variable en coma flotante se recupera (588) y el cuantificador se obtiene mediante el redondeo de la variable en coma flotante al entero más cercano. Entonces, la trama se codifica (590). Se hacen

observaciones durante la codificación de la trama que habilitan la determinación (592) de un valor de cuantificador nuevo. El proceso decide (594) realizar una repetición a menos que no haya ninguna trama más. Punto en el que se ha completado la codificación (596).

5 Como se ha analizado anteriormente, el codificador se inicializa (584) con diversos parámetros. Estos parámetros son la 'tasa de bits', la 'tasa de tramas', el 'Máx Intervalo entre Tramas Clave', el 'Cuantificador Máximo', el 'Cuantificador Mínimo', el 'periodo de promediado', el 'periodo de reacción' y la 'relación arriba/abajo'. Lo siguiente es un análisis de cada uno de estos parámetros.

10 3.5.5.1.1. La 'tasa de bits'

El parámetro de 'tasa de bits' establece la tasa de bits objetivo de la codificación.

15 3.5.5.1.2. La 'tasa de tramas'

La 'tasa de tramas' define el periodo entre tramas de vídeo.

3.5.5.1.3. El 'Máx Intervalo entre Tramas Clave'

20 El 'Máx Intervalo entre Tramas Clave' especifica el intervalo máximo entre las tramas clave. Las tramas clave se insertan automáticamente de forma normal en el vídeo codificado cuando el códec detecta un cambio de escena. En circunstancias en las que una escena continúa durante un intervalo largo sin siquiera un corte, las tramas clave se pueden insertar y asegurar que el intervalo entre tramas clave es siempre menor que o igual al 'Máx Intervalo entre Tramas Clave'. En una realización, el parámetro de 'Máx Intervalo entre Tramas Clave' se puede establecer a un valor
25 de 300 tramas. En otras realizaciones, se pueden usar otros valores.

3.5.5.1.4. El 'Cuantificador Máximo' y el 'Cuantificador Mínimo'

30 Los parámetros de 'Cuantificador Máximo' y de 'Cuantificador Mínimo' establecen la cota superior e inferior del cuantificador usado en la codificación. En una realización, las cotas de cuantificador se establecen a unos valores de entre 1 y 31.

3.5.5.1.5. El 'periodo de promediado'

35 El parámetro de 'periodo de promediado' controla la cantidad de vídeo que se considera cuando se modifica el cuantificador. Por lo general, un periodo de promediado más largo dará como resultado que el vídeo codificado tenga una tasa global más precisa. En una realización, se usa un 'periodo de promediado' de 2000. Aunque, en otras realizaciones, se pueden usar otros valores.

40 3.5.5.1.6 El 'periodo de reacción'

El parámetro de 'periodo de reacción' determina lo rápido que el codificador se adapta a los cambios en el movimiento en escenas recientes. Un valor de 'periodo de reacción' más largo puede dar como resultado unas escenas de movimiento alto de mejor calidad y unas escenas de movimiento bajo de calidad peor. En una realización, se usa un
45 'periodo de reacción' de 10. Aunque, en otras realizaciones, se pueden usar otros valores.

3.5.5.1.7. La 'relación arriba/abajo'

50 El parámetro de 'relación arriba/abajo' controla la sensibilidad relativa para el ajuste de cuantificador como reacción a las escenas de movimiento alto o bajo. Por lo general, un valor más grande da como resultado unas escenas de movimiento alto de calidad más alta y un consumo de bits aumentado. En una realización, se usa una 'relación arriba/abajo' de 20. Aunque en otras realizaciones, se pueden usar otros valores.

3.5.5.1.8. Cálculo del valor de cuantificador

55 Como se ha analizado anteriormente, la técnica de control de tasa de una pasada implica el cálculo de un valor de cuantificador después de la codificación de cada trama. Lo siguiente es una descripción de una técnica de acuerdo con una realización de la presente invención que se puede usar para actualizar el valor de cuantificador.

60 El codificador mantiene dos promedios móviles exponenciales que tienen periodos iguales al 'periodo de promediado' ($P_{promedio}$) y el 'periodo de reacción' ($P_{reaccion}$) un promedio móvil de la tasa de bits. Los dos promedios móviles exponenciales

se pueden calcular de acuerdo con la relación:

65

$$A_t = A_{t-1} \cdot \frac{P-T}{P} + B \cdot \frac{T}{P}$$

en donde

- 5 A_t es el promedio en el caso t ;
 A_{t-1} es el promedio en el caso $t-1$ (habitualmente, el promedio en la trama previa);
 T representa el periodo de intervalo (habitualmente, el tiempo de trama); y
 P es el periodo promedio, que puede ser o bien $P_{promedio}$ y o bien $P_{reacción}$.

- 10 Entonces, el promedio móvil calculado anteriormente se ajusta a una tasa de bits al dividir por el intervalo de tiempo entre el caso actual y el último caso en el vídeo, usando el siguiente cálculo:

$$R_t = A_t \frac{1}{T}$$

15 en donde

- R_t es la tasa de bits;
 A_t es uno cualquiera de los promedios móviles; y
 T es el intervalo de tiempo entre el caso actual y el último caso (este es, habitualmente, el inverso de la tasa de tramas).
20

El codificador puede calcular la tasa de bits objetivo ($R_{objetivo}$) de la siguiente trama como sigue:

$$R_{objetivo} = R_{global} + (R_{global} - R_{promedio})$$

25 en donde

- R_{global} es la tasa de bits global establecida para la totalidad del vídeo; y
 $R_{promedio}$ es la tasa de bits promedio usando el periodo de promediado largo.
30

En varias realizaciones, la tasa de bits objetivo está acotada inferiormente por un 75 % de la tasa de bits global. Si la tasa de bits objetivo cae por debajo de esa cota, entonces se forzará que está suba hasta la cota para asegurar la calidad del vídeo.

- 35 Entonces, el codificador actualiza el cuantificador interno basándose en la diferencia entre $R_{objetivo}$ y $R_{reacción}$. Si $R_{reacción}$ es menor que $R_{objetivo}$, entonces hay una verosimilitud de que la trama previa fuera de una complejidad relativamente baja. Por lo tanto, el cuantificador se puede disminuir al realizar el siguiente cálculo:

$$Q'_{interno} = Q_{interno} \cdot \left(1 - \frac{1}{P_{reacción}}\right)$$

- 40 Cuando $R_{reacción}$ es mayor que $R_{objetivo}$, hay una verosimilitud significativa de que la trama previa poseyera un nivel relativamente alto de complejidad. Por lo tanto, el cuantificador se puede aumentar al realizar el siguiente cálculo:

$$Q'_{interno} = Q_{interno} \cdot \left(1 + \frac{1}{SP_{reacción}}\right)$$

45 en donde

S es la 'relación arriba/abajo'.

3.5.5.1.9. Codificación de VOP-B

- 50 El algoritmo descrito anteriormente también se puede aplicar a la codificación de VOP-B. Cuando se habilita VOP-B en la codificación, el cuantificador para el VOP-B (Q_B) se elige basándose en el cuantificador del VOP-P (Q_P) siguiendo el VOP-B. El valor se puede obtener de acuerdo con las siguientes relaciones:

$$Q_B = 2, Q_P \text{ para } Q_P \leq 4$$

$$Q_B = 5 + \frac{3}{4} \cdot Q_P \text{ para } 4 < Q_P \leq 20$$

$$Q_B = Q_P \text{ para } Q \geq 20$$

60

3.5.5.2. Control de tasa de dos pasadas

Los codificadores de acuerdo con una realización de la presente invención que usan una técnica de control de tasa de dos (o de múltiples) pasadas pueden determinar las propiedades de una secuencia de vídeo en una primera pasada y, entonces, codificar la secuencia de vídeo con conocimiento de las propiedades de la totalidad de la secuencia. Por lo tanto, el codificador puede ajustar el nivel de cuantificación para cada trama basándose en su complejidad relativa en comparación con otras tramas en la secuencia de vídeo.

Una técnica de control de tasa de dos pasadas de acuerdo con una realización de la presente invención, el codificador realiza una primera pasada en la que el vídeo se codifica de acuerdo con la técnica de control de tasa de una pasada descrita anteriormente y se registra la complejidad de cada trama (se puede usar cualquiera de diversas métricas diferentes para medir la complejidad). La complejidad promedio y, por lo tanto, el cuantificador promedio (Q_{ref}) se pueden determinar basándose en la primera. En la segunda pasada, la secuencia de bits se codifica con cuantificadores determinados basándose en los valores de complejidad calculados durante la primera pasada.

3.5.5.2.1. Cuantificadores para los VOP-I

El cuantificador Q para los VOP-I se establece a $0,75 \times Q_{ref}$, con la condición de que la siguiente trama no sea un VOP-I. Si la siguiente trama también es un VOP-I, la Q (para la trama actual) se establece a $1,25 \times Q_{ref}$.

3.5.5.2.2. Cuantificadores para los VOP-P

El cuantificador para los VOP-P se puede determinar usando la siguiente expresión.

$$Q = F^{-1} \left\{ F(Q_{ref}) \cdot \left[\overline{C_{complejidad}} / C_{complejidad} \right]^k \right\}$$

en donde

$C_{complejidad}$ es la complejidad de la trama;

$\overline{C_{complejidad}}$ es la complejidad promedio de la secuencia de vídeo;

$F(x)$ es una función que proporciona el número por el que se ha de multiplicar la complejidad de la trama para dar el número de bits requerido para codificar la trama usando un cuantificador con un valor de cuantificación x ;

$F^{-1}(x)$ es la función inversa de $F(x)$; y k es el parámetro de intensidad.

La siguiente tabla define una realización de una función $F(Q)$ que se puede usar para generar el factor por el que se ha de multiplicar la complejidad de una trama con el fin de determinar el número de bits requerido para codificar la trama usando un codificador con un cuantificador Q .

Tabla 6 - Valores de $F(Q)$ con respecto a Q .

Q	$F(Q)$	Q	$F(Q)$
1	1	9	0,013
2	0,4	10	0,01
3	0,15	11	0,008
4	0,08	12	0,0065
5	0,05	13	0,005
6	0,032	14	0,0038
7	0,022	15	0,0028
8	0,017	16	0,002

Si se elige que el parámetro de intensidad k sea 0, entonces el resultado es un cuantificador constante. Cuando se elige que el parámetro de intensidad sea 1, el cuantificador es proporcional a $C_{complejidad}$. Varios codificadores de acuerdo con algunas realizaciones de la presente invención tienen un parámetro de intensidad k igual a 0,5.

3.5.5.2.3. Cuantificadores para los VOP-B

El cuantificador Q para los VOP-B se puede elegir usando la misma técnica para elegir el cuantificador para los VOP-B en la técnica de una pasada descrita anteriormente.

3.5.5.3. Control de tasa de memoria intermedia de vídeo verificada

El número de bits requeridos para representar una trama puede variar dependiendo de las características de la secuencia de vídeo. La mayor parte de los sistemas de comunicación operan a una tasa de bits constante. Un problema que se puede encontrar con las comunicaciones de tasa de bits variable es la asignación de recursos

suficientes para manejar los picos en el uso de recursos. Varios codificadores de acuerdo con algunas realizaciones de la presente invención codifican vídeo con una vista para evitar el desbordamiento de una memoria intermedia de vídeo de descodificador, cuando la tasa de bits de la comunicación de tasa de bits variable experimenta un aumento rápido.

Los objetivos del control de tasa de verificador de memoria intermedia de vídeo (VBV) pueden incluir generar vídeo que no superará la memoria intermedia de un descodificador cuando se transmite. Además, puede ser deseable que el vídeo codificado coincida con una tasa de bits objetivo y que el control de tasa produzca vídeo de alta calidad.

Los codificadores de acuerdo con varias realizaciones de la presente invención proporcionan una elección de al menos dos técnicas de control de tasa de VBV. Se hace referencia a una de las técnicas de control de tasa de VBV como control de tasa causal y se hace referencia a la otra técnica como control de tasa de N-ésima pasada.

3.5.5.3.1. Control de tasa causal

Se puede usar un control de tasa de VBV causal junto con una técnica de control de tasa de una pasada y genera salidas simplemente basándose en los valores de cuantificador actual y previo.

Un codificador de acuerdo con una realización de la presente invención incluye un control de tasa causal que implica establecer el cuantificador para la trama n (es decir, Q_n) de acuerdo con la siguiente relación.

$$\frac{1}{Q'_n} = \frac{1}{Q'_{n-1}} + X_{tasa\ de\ bits} + X_{velocidad} + X_{tamaño}$$

$$\frac{1}{Q_n} = \frac{1}{Q'_n} + X_{deriva}$$

en donde

Q'_n es el cuantificador estimado por el control de tasa de una única pasada;

$X_{tasa\ de\ bits}$ se calcula mediante la determinación de una tasa de bits objetivo basándose en la deriva con respecto a la tasa de bits deseada;

$X_{velocidad}$ se calcula basándose en el tiempo estimado hasta que la memoria intermedia de VBV se ha desbordado o subdesbordado;

$X_{tamaño}$ solo se aplica sobre el resultado de los VOP-P y se calcula basándose en la tasa a la cual el tamaño de los VOP-P comprimidos está cambiando con el paso del tiempo;

X_{deriva} es la deriva con respecto a la tasa de bits deseada.

En varias realizaciones, se puede forzar al control de tasa de VBV causal a omitir tramas e insertar relleno para respetar el modelo de VBV. Si una trama comprimida contiene inesperadamente demasiados o demasiado pocos bits, entonces esta se puede omitir o rellenarse.

3.5.5.3.2. Control de tasa de VBV de N-ésima pasada

Se puede usar un control de tasa de VBV de N-ésima pasada junto con una técnica de control de tasa de múltiples pasadas y esta usa información obtenida durante el análisis previo de la secuencia de vídeo. Los codificadores de acuerdo con varias realizaciones de la presente invención realizan un control de tasa de VBV de N-ésima pasada de acuerdo con el proceso ilustrado en la figura 3.9. El proceso 600 comienza con la primera pasada, durante la cual se realiza un análisis (602). Se realiza una generación de correlación (604) y se genera una estrategia (606). Entonces se realiza el control de tasa de n-ésima pasada (608).

3.5.5.3.3. Análisis

En una realización, la primera pasada usa alguna forma de control de tasa causal y se registran datos para cada trama con respecto a cosas tales como la duración de la trama, el tipo de codificación de la trama, el cuantificador usado, los bits de movimiento producidos y los bits de textura producidos. Además, también se puede registrar información global tal como los ajustes de escala de tiempo, de resolución y de códec.

3.5.5.3.4. Generación de correlación

Se usa información a partir del análisis para generar una correlación la secuencia de vídeo. La correlación puede especificar el tipo de codificación usado para cada trama (I/B/P) y puede incluir datos para cada trama con respecto a la duración de la trama, la complejidad de movimiento y la complejidad de textura. En otras realizaciones, la correlación también puede contener información que posibilita una predicción mejor de la influencia del cuantificador y otros parámetros sobre el tamaño de trama comprimida y la distorsión perceptual. En varias realizaciones, se realiza una

generación de correlación después de que se haya completado la $N - 1$ -ésima pasada.

3.5.5.3.5. *Generación de estrategia*

La correlación se puede usar para planear una estrategia en lo que respecta a cómo funcionará el control de tasa de N -ésima pasada. Se puede planear el nivel ideal de la memoria intermedia de VBV después de que se haya codificado cada trama. En una realización, la generación de estrategia da como resultado una información para cada trama que incluye el tamaño de trama comprimida deseado, un cuantificador de trama estimado. En varias realizaciones, la generación de estrategia se realiza después de la generación de correlación y antes de la N -ésima pasada.

En una realización, el proceso de generación de estrategia implica el uso de un proceso iterativo para simular el codificador y determinar los valores de cuantificador deseados para cada trama al intentar mantener el cuantificador tan cerca como es posible de la mediana del valor de cuantificador. Se puede usar una búsqueda binaria para generar un cuantificador de base para la totalidad de la secuencia de vídeo. El cuantificador de base es el valor constante que da lugar a que el simulador logre la tasa de bits objetivo deseada. Una vez que se ha hallado el cuantificador de base, el proceso de generación de estrategia implica la consideración de las restricciones de VBV. En una realización, se usa un cuantificador constante si este no modifica las restricciones de VBV. En otras realizaciones, el cuantificador se modula basándose en la complejidad de movimiento en las tramas de vídeo. Esto se puede ampliar adicionalmente para incorporar un enmascaramiento a partir de cambios de escenas y otros efectos temporales.

3.5.5.3.6. *Control de tasa de N -ésima pasada dentro de lazo*

En una realización, el control de tasa de N -ésima pasada dentro de lazo usa la estrategia y usa la correlación para realizar la mejor predicción posible de la influencia del cuantificador y otros parámetros sobre el tamaño de trama comprimida y la distorsión perceptual. Puede existir una facultad discrecional limitada para apartarse de la estrategia para adoptar una estrategia correctiva a corto plazo. Por lo general, seguir la estrategia evitará el quebrantamiento del modelo de VBV. En una realización, el control de tasa de N -ésima pasada dentro de lazo usa un lazo de control PID. La realimentación en el lazo de control es la deriva acumulada con respecto a la tasa de bits ideal.

Aunque la generación de estrategia no implica omitir tramas, el control de tasa N -ésimo dentro de lazo puede omitir tramas si, de lo contrario, la memoria intermedia de VBV se subdesbordara. De forma similar, el control de tasa de N -ésima pasada dentro de lazo puede solicitar que se inserte relleno de vídeo para evitar el desbordamiento de VBV.

3.5.6. *Predicciones*

En una realización, la predicción de CA/CC se realiza de una forma que es conforme con la norma a la que se hace referencia como ISO/IEC 14496-2:2001 (E), sección 7.4.3. (predicción de CC y de CA) y 7.7.1. (predicción de CC y de CA de campo).

3.5.7. *Codificación de textura*

Un codificador de acuerdo con una realización de la presente invención puede realizar una codificación de textura de una forma que es conforme con la norma a la que se hace referencia como ISO/IEC 14496-2:2001(E), anexo B (códigos de longitud variable) y 7.4.1. (descodificación de longitud variable).

3.5.8. *Codificación de movimiento*

Un codificador de acuerdo con una realización de la presente invención puede realizar una codificación de movimiento de una forma que es conforme con la norma a la que se hace referencia como ISO/IEC 14496-2:2001(E), anexo B (códigos de longitud variable) y 7.6.3. (descodificación de vector de movimiento).

3.5.9. *Generación de fragmentos de 'vídeo'*

La pista de vídeo se puede considerar una secuencia de tramas 1 a N . Algunos sistemas de acuerdo con algunas realizaciones de la presente invención son capaces de codificar la secuencia para generar una secuencia de bits comprimida. Se da formato a la secuencia de bits mediante la segmentación de la misma en los fragmentos 1 a N . Cada trama de vídeo n tiene un fragmento n correspondiente.

Los fragmentos se generan al adjuntar bits de la secuencia de bits al fragmento n hasta que este, junto con los fragmentos 1 a $n - 1$, contienen una información suficiente para que un descodificador de acuerdo con una realización de la presente invención descodifique la trama de vídeo n . En los casos en los que una información suficiente se contiene en los fragmentos 1 a $n - 1$ para generar una trama de vídeo n , un codificador de acuerdo con algunas realizaciones de la presente invención puede incluir un fragmento marcador. En una realización, el fragmento marcador es una trama P no codificada con una información de temporización idéntica a la de la trama previa.

3.6. Generación de fragmentos de 'subtítulo'

Un codificador de acuerdo con una realización de la presente invención puede tomar subtítulos en uno de una serie de formatos convencionales y, entonces, convierte los subtítulos en mapas de bits. Entonces, la información en los mapas de bits se comprime usando una codificación por longitud de ejecución. Entonces se da formato a los mapas de bits codificados por longitud de ejecución para dar un fragmento, que también incluye información con respecto al tiempo de inicio y el tiempo de parada para el subtítulo particular contenido dentro del fragmento. En varias realizaciones, también se puede incluir, en el fragmento, información con respecto al color, el tamaño y la posición del subtítulo en la pantalla. En la pista de subtítulo se pueden incluir fragmentos que establecen la paleta para los subtítulos y que indican que la paleta ha cambiado. Se puede usar cualquier aplicación capaz de generar un subtítulo en un formato de subtítulo convencional para generar el texto de los subtítulos. Como alternativa, se puede usar software para convertir el texto introducido por un usuario directamente en información de subtítulo.

3.7. Intercalación

Una vez que el intercalador ha recibido la totalidad de los fragmentos descritos anteriormente, el intercalador construye un archivo multimedia. La construcción del archivo multimedia puede implicar la creación de un fragmento de 'CSET', un fragmento de lista de 'INFO', un fragmento de 'hdlr', un fragmento de lista de 'película' y un fragmento de idx1. Algunos métodos de acuerdo con algunas realizaciones de la presente invención para crear estos fragmentos y para generar archivos multimedia se describen posteriormente.

3.7.1. Generación de un fragmento de 'CSET'

Como se ha descrito anteriormente, el fragmento de 'CSET' es opcional y puede ser generado por el intercalador de acuerdo con la Especificación de Formato de Contenedor de AVI.

3.7.2. Generación de un fragmento de lista de 'INFO'

Como se ha descrito anteriormente, el fragmento de lista de 'INFO' es opcional y puede ser generado por el intercalador de acuerdo con la Especificación de Formato de Contenedor de AVI.

3.7.3. Generación del fragmento de lista de 'hdlr'

El fragmento de lista de 'hdlr' es generado por el intercalador basándose en la información en los diversos fragmentos proporcionados al intercalador. El fragmento de lista de 'hdlr' hace referencia a la ubicación dentro del archivo de los fragmentos a los que se hace referencia. En una realización, el fragmento de lista de 'hdlr' usa desplazamientos de archivo con el fin de establecer referencias.

3.7.4. Generación del fragmento de lista de 'película'

Como se ha descrito anteriormente, el fragmento de lista de 'película' se crea mediante la codificación de pistas de audio, de vídeo y de subtítulo para crear fragmentos de 'audio', de 'vídeo' y de 'subtítulo' y, entonces, la intercalación de estos fragmentos. En varias realizaciones, el fragmento de lista de 'película' también puede incluir información de gestión de derechos digitales.

3.7.4.1. Intercalación del vídeo/el audio/los subtítulos

Se puede usar diversas reglas para intercalar los fragmentos de audio, de vídeo y de subtítulo. Por lo general, el intercalador establece un número de colas para cada una de las pistas de vídeo y de audio. El intercalador determina qué cola se debería escribir en el archivo de salida. La selección de cola se puede basar en el periodo de intercalación al escribir desde la cola que tiene el número más bajo de periodos de intercalación escritos. El intercalador puede tener que aguardar la totalidad de un periodo de intercalación para estar presente en la cola antes de que se pueda escribir el fragmento en el archivo.

En una realización, los fragmentos de 'audio', de 'vídeo' y de 'subtítulo' generados se intercalan de tal modo que los fragmentos de 'audio' y de 'subtítulo' se encuentran dentro del archivo antes de los fragmentos de 'vídeo' que contienen información con respecto a las tramas de vídeo con las que se corresponden los mismos. En otras realizaciones, los fragmentos de 'audio' y de 'subtítulo' se pueden situar después de los fragmentos de 'vídeo' con los que se corresponden los mismos. Las diferencias de tiempo entre la ubicación de los fragmentos de 'audio', de 'vídeo' y de 'subtítulo' depende en gran medida de las capacidades de almacenamiento en memoria intermedia de reproductores que se usan para hacer funcionar los dispositivos. En algunas realizaciones en las que el almacenamiento en memoria intermedia es limitado o desconocido, el intercalador intercala los fragmentos de 'audio', de 'vídeo' y de 'subtítulo' de tal modo que los fragmentos de 'audio' y de 'subtítulo' se encuentran entre fragmentos de 'vídeo', en donde el fragmento de 'vídeo' que inmediatamente sigue al fragmento de 'audio' y de 'subtítulo' contiene la primera trama de vídeo que se corresponde con el audio o el subtítulo.

3.7.4.2. Generación de información de DRM

En algunas realizaciones en las que se usa una DRM para proteger el contenido de vídeo de un archivo multimedia, la información de DRM se puede generar de forma concurrente con la codificación de los fragmentos de vídeo. A medida que se genera cada fragmento, el fragmento se puede cifrar, y generarse un fragmento de DRM que contiene información con respecto al cifrado del fragmento de vídeo.

3.7.4.3. Intercalación de la información de DRM

Un intercalador de acuerdo con una realización de la presente invención intercala un fragmento de DRM que contiene información con respecto al cifrado de un fragmento de vídeo antes del fragmento de vídeo. En una realización, el fragmento de DRM para el fragmento de vídeo n está situado entre el fragmento de vídeo $n - 1$ y el fragmento de vídeo n . En otras realizaciones, la separación de la DRM antes y después del fragmento de vídeo n depende de la cantidad de almacenamiento en memoria intermedia proporcionado dentro del dispositivo que descodifica el archivo multimedia.

3.7.5. Generación del fragmento de 'idx1'

Una vez que se ha generado el fragmento de lista de 'película', la generación del fragmento de 'idx1' es un proceso simple. El fragmento de 'idx1' se crea mediante la lectura de la ubicación dentro del fragmento de lista de 'película' de cada fragmento de 'datos'. Esta información se combina con información leída a partir del fragmento de 'datos' con respecto a la pista a la que pertenece el fragmento de 'datos' y el contenido del fragmento de 'datos'. Entonces, la totalidad de esta información se inserta en el fragmento de 'idx1' de una forma apropiada para cualquiera de los formatos descritos anteriormente que sea el que se está usando para representar la información.

4. Transmisión y distribución de archivo multimedia

Una vez que se ha generado un archivo multimedia, el archivo se puede distribuir a lo largo de cualquiera de diversas redes. El hecho de que, en muchas realizaciones, los elementos requeridos para generar una presentación multimedia y menús, entre otras cosas, están contenidos dentro de un único archivo simplifica la transferencia de la información. En varias realizaciones, el archivo multimedia se puede distribuir por separado de la información requerida para descifrar los contenidos del archivo multimedia.

En una realización, se proporciona contenido multimedia a un primer servidor y se codifica para crear un archivo multimedia de acuerdo con la presente invención. Entonces, el archivo multimedia se puede situar o bien en el primer servidor o bien en un segundo servidor. En otras realizaciones, la información de DRM se puede situar en el primer servidor, el segundo servidor o un tercer servidor. En una realización, se puede consultar el primer servidor para determinar la ubicación del archivo multimedia codificado y/o para determinar la ubicación de la información de DRM.

5. Descodificación de archivo multimedia

Se puede acceder a la información a partir de un archivo multimedia de acuerdo con una realización de la presente invención mediante un ordenador configurado usando software apropiado, un reproductor dedicado que tiene cableado físico para acceder a información a partir del archivo multimedia o cualquier otro dispositivo capaz de analizar un archivo AVI. En varias realizaciones, los dispositivos pueden acceder a la totalidad de la información en el archivo multimedia. En otras realizaciones, un dispositivo puede ser incapaz de acceder a la totalidad de la información en un archivo multimedia de acuerdo con una realización de la presente invención. En una realización particular, un dispositivo no es capaz de acceder a parte alguna de la información descrita anteriormente que se almacena en fragmentos que no se especifican en el formato de archivo AVI. En algunas realizaciones en las que no se puede acceder a toda la información, por lo general el dispositivo descartará aquellos fragmentos que no son reconocidos por el dispositivo.

Por lo general, un dispositivo que es capaz de acceder a la información contenida en un archivo multimedia de acuerdo con una realización de la presente invención es capaz de realizar un número de funciones. El dispositivo puede visualizar una presentación multimedia que implica la visualización de vídeo en un visualizador visual, generar audio a partir de una de potencialmente un número de pistas de audio en un sistema de audio y visualizar subtítulos a partir de potencialmente una de un número de pistas de subtítulo. Varias realizaciones también pueden visualizar menús en un visualizador visual al tiempo que se reproduce el audio y/o vídeo acompañante. Estos menús de visualización son interactivos, con características tales como botones seleccionables, menús desplegados y submenús. En algunas realizaciones, los artículos de menú pueden apuntar a un contenido de audio/vídeo fuera del archivo multimedia al que se está accediendo en la actualidad. El contenido exterior o bien se puede situar de forma local con respecto al dispositivo que accede al archivo multimedia, o bien este se puede situar de forma remota, tal como a lo largo de una red de área local, de área extensa o pública. Muchas realizaciones también pueden buscar uno o más archivos multimedia de acuerdo con 'metadatos' incluidos dentro del archivo o archivos multimedia o 'metadatos' a los que se hace referencia mediante uno o más de los archivos multimedia.

5.1. Visualización de presentación multimedia

Dada la capacidad de los archivos multimedia de acuerdo con algunas realizaciones de la presente invención para soportar múltiples pistas de audio, múltiples pistas de vídeo y múltiples pistas de subtítulo, la visualización de una presentación multimedia usando un archivo multimedia de este tipo que combina vídeo, audio y/o subtítulos puede requerir la selección de una pista de audio, pista de vídeo y/o pista de subtítulo particular o bien a través de un sistema de menú visual o un sistema de menú desplegable (el funcionamiento de los cuales se analiza posteriormente) o bien por medio de los ajustes por defecto del dispositivo usado para generar la presentación multimedia. Una vez que se han seleccionado una pista de audio, una pista de vídeo y, potencialmente, una pista de subtítulo, puede proseguir la visualización de la presentación multimedia.

Un proceso para localizar la información multimedia requerida a partir de un archivo multimedia que incluye DRM y visualizar la información multimedia de acuerdo con una realización de la presente invención se ilustra en la figura 4.0. El proceso 620 incluye obtener la clave de cifrado requerida para descifrar el encabezamiento de DRM (622). Entonces, se usa la clave de cifrado para descifrar (624) el encabezamiento de DRM y el primer fragmento de DRM está situado (626) dentro del fragmento de lista de 'película' del archivo multimedia. La clave de cifrado requerida para descifrar el fragmento de 'DRM' se obtiene (628) a partir de la tabla en el encabezamiento de 'DRM' y se usa la clave de cifrado para descifrar un fragmento de vídeo cifrado. Entonces, el fragmento de audio requerido y cualquier fragmento de subtítulo requerido acompañan al fragmento de vídeo se descodifican (630) y la información de audio, de vídeo y cualquiera información de subtítulo se presentan (632) por medio del visualizador y el sistema de sonido.

En varias realizaciones, la pista de audio elegida puede incluir múltiples canales para proporcionar un audio de sonido estéreo o envolvente. Cuando se elige una pista de subtítulo para su visualización, se puede realizar una determinación en lo que respecta a si la trama de vídeo previa incluía un subtítulo (esta determinación se puede realizar en cualquiera de diversas formas que logra el resultado de identificar un fragmento de 'subtítulo' previo que contenía información de subtítulo que se debería visualizar a lo largo de la trama de vídeo actualmente descodificada). Si el subtítulo previo incluía un subtítulo y la información de temporización para el subtítulo indica que el subtítulo se debería visualizar con la trama actual, entonces el subtítulo se superpone sobre la trama de vídeo descodificada. Si la trama previa no incluía un subtítulo o la información de temporización para el subtítulo en la trama previa indica que el subtítulo no se debería visualizar junto con la trama actualmente descodificada, entonces se busca un fragmento de 'subtítulo' para la pista de subtítulo seleccionada. Si se localiza un fragmento de 'subtítulo', entonces el subtítulo se superpone sobre el vídeo descodificado. El vídeo (incluyendo cualquier subtítulo superpuesto) se visualiza entonces con el audio acompañante.

Volviendo al análisis de la figura 4.0., el proceso determina (634) si hay fragmento de DRM adicional alguno. Si lo hay, entonces se localiza (626) el siguiente fragmento de DRM y el proceso continúa hasta que no queda fragmento de DRM adicional alguno. Punto en el que se ha completado la presentación de las pistas de audio, de vídeo y/o de subtítulo (636),

En varias realizaciones, un dispositivo puede buscar una porción particular de la información multimedia (por ejemplo, una escena particular de una película con una pista de audio acompañante particular y, de forma opcional, una pista de subtítulo acompañante particular) usando información contenida dentro del fragmento de 'hdl' de un archivo multimedia de acuerdo con la presente invención. En muchas realizaciones, la descodificación del fragmento de 'vídeo', el fragmento de 'audio' y/o el fragmento de 'subtítulo' se puede realizar en paralelo con otras tareas.

Un ejemplo de un dispositivo capaz de acceder a información a partir del archivo multimedia y visualizar vídeo junto con una pista de audio particular y/o una pista de subtítulo particular es un ordenador configurado de la forma descrita anteriormente usando software. Otro ejemplo es un reproductor de DVD equipado con un códec que incluye estas capacidades. En otras realizaciones, cualquier dispositivo configurado para situar o seleccionar (ya sea de forma deliberada o arbitraria) fragmentos de 'datos' que se corresponden con pistas de medios particulares y descodificar esas pistas para la presentación es capaz de generar una presentación multimedia usando un archivo multimedia de acuerdo con la práctica de la presente invención.

En varias realizaciones, un dispositivo puede reproducir información multimedia a partir de un archivo multimedia en combinación con información multimedia a partir de un archivo externo. Por lo general, un dispositivo de este tipo lo haría al abastecer una pista de audio o pista de subtítulo desde un archivo local al que se hace referencia en un archivo multimedia del tipo descrito anteriormente. Si el archivo al que se hace referencia no se almacena de forma local y el dispositivo se interconecta en red con la ubicación en la que se almacena el dispositivo, entonces el dispositivo puede obtener una copia local del archivo. El dispositivo accedería entonces a ambos archivos, estableciendo una canalización de vídeo, una canalización de audio y una canalización de subtítulo (si así se requiere) en las que las diversas pistas de multimedia se alimentan a partir de las diferentes fuentes de archivos.

5.2. Generación de menús

Un descodificador de acuerdo con una realización de la presente invención se ilustra en la figura 4.1. El descodificador 650 procesa un archivo multimedia 652 de acuerdo con una realización de la presente invención al proporcionar el archivo a un desmultiplexor 654. El desmultiplexor extrae el fragmento de 'DMNU' a partir del archivo multimedia y extrae la totalidad de los fragmentos de 'Menusldioma' a partir del fragmento de 'DMNU' y los proporciona a un analizador de menú 656. El desmultiplexor también extrae la totalidad de los fragmentos de 'Medios' a partir del

fragmento de 'DMNU' y los proporciona a un representador de medios 658. El analizador de menú 656 analiza información a partir de los fragmentos de 'Menuldioma' para construir una máquina de estados que representa la estructura de menú definida en el fragmento de 'Menuldioma'. La máquina de estados que representa la estructura de menú se puede usar para proporcionar visualizaciones al usuario y para responder a órdenes de usuario. La máquina de estados se proporciona a un controlador de estado de menú 660. El controlador de estado de menú realiza un seguimiento del estado actual de la máquina de estados de menú y recibe órdenes procedentes del usuario. Las órdenes procedentes del usuario pueden dar lugar a una transición de estado. La visualización inicial proporcionada a un usuario y cualquier actualización a la visualización que acompaña a una transición de estado de menú se pueden controlar usando una interfaz de reproductor de menú 662. La interfaz de reproductor de menú 662 se puede conectar al controlador de estado de menú y el representador de medios. La interfaz de reproductor de menú da instrucciones al representador de medios acerca de qué medios se deberían extraer a partir de los fragmentos de medios y proporcionarse al usuario por medio del reproductor 664 conectado al representador de medios. El usuario puede dotar de instrucciones al reproductor usando un dispositivo de entrada tal como un teclado, ratón o control remoto. En general, el archivo multimedia dicta el menú inicialmente visualizado al usuario y las instrucciones del usuario dictan el audio y el vídeo visualizado a continuación de la generación del menú inicial. El sistema ilustrado en la figura 4.1. se puede implementar usando un ordenador y software. En otras realizaciones, el sistema se puede implementar usando circuitos integrados para funciones específicas o una combinación de software y firmware.

Un ejemplo de un menú de acuerdo con una realización de la presente invención se ilustra en la figura 4.2. La visualización de menú 670 incluye cuatro áreas de botón 672, un vídeo de segundo plano 674, incluyendo un título 676 y un puntero 678. El menú también incluye un audio de segundo plano (no mostrado). El efecto visual creado por la visualización puede ser engañoso. Por lo general, la apariencia visual de los botones es parte del vídeo de segundo plano y los propios botones son regiones definidas simplemente del vídeo de segundo plano que tienen acciones particulares asociadas con los mismos, cuando la región es activada por el puntero. Por lo general, el puntero es una superposición.

La figura 4.3. ilustra conceptualmente la fuente de la totalidad de la información en la visualización mostrada en la figura 4.2. El vídeo de segundo plano 674 puede incluir un título de menú, la apariencia visual de los botones y el segundo plano de la visualización. La totalidad de estos elementos, y elementos adicionales, pueden parecer estáticos o animados. El vídeo de segundo plano se extrae mediante el uso de información contenida en un fragmento de 'PistaMedios' 700 que indica la ubicación del vídeo de segundo plano dentro de una pista de vídeo 702. El audio de segundo plano 706 que puede acompañar al menú se puede situar usando un fragmento de 'PistaMedios' 708 que indica la ubicación del audio de segundo plano dentro de una pista de audio 710. Como se ha descrito anteriormente, el puntero 678 es parte de una superposición 713. La superposición 713 también puede incluir gráficos que parecen resaltar la porción del vídeo de segundo plano que aparece como un botón. En una realización, la superposición 713 se obtiene usando un fragmento de 'PistaMedios' 712 que indica la ubicación de la superposición dentro de una pista de superposición 714. La forma en la que el menú interactúa con un usuario es definida por los fragmentos de 'Acción' (no mostrados) asociados con cada uno de los botones. En la realización ilustrada, se ilustra un fragmento de 'AccionReproduccion' 716. El fragmento de 'AccionReproduccion' hace indirectamente referencia (no se muestran los otros fragmentos a los que se hace referencia mediante el fragmento de 'AccionReproduccion') a una escena dentro de una presentación multimedia contenida dentro del archivo multimedia (es decir, una pista de audio, de vídeo y, posiblemente, una de subtítulo). El fragmento de 'AccionReproduccion' 716 hace referencia en última instancia a la escena usando un fragmento de 'PistaMedios' 718, que indica la escena dentro de la pista de características. También se hace referencia a un punto en una pista de audio seleccionada o por defecto y, potencialmente, una pista de subtítulo.

A medida que el usuario introduce órdenes usando el dispositivo de entrada, la visualización se puede actualizar no solo en respuesta a la selección de áreas de botón, sino también simplemente debido a que el puntero está situado dentro de un área de botón. Como se ha analizado anteriormente, por lo general, la totalidad de la información de medios usada para generar los menús está situada dentro del archivo multimedia y, más concretamente, dentro de un fragmento de 'DMNU'. Aunque, en otras realizaciones, la información se puede situar en cualquier otra parte dentro del archivo y/o en otros archivos.

5.3. Acceso a los Metadatos

Los 'metadatos' son un método normalizado de representar información. La naturaleza normalizada de 'Metadatos' posibilita que a los datos accedan, y que estos sean entendidos por, procesos automáticos. En una realización, los 'metadatos' se extraen y se proporcionan a un usuario para su visionado. Varias realizaciones posibilitan que se inspeccionen los archivos multimedia en un servidor para proporcionar información con respecto a los hábitos de visionado y a las preferencias de visionado de un usuario. Tal información podría ser usada por aplicaciones de software para recomendar otros archivos multimedia cuyo visionado puede disfrutar un usuario. En una realización, las recomendaciones se pueden basar en los archivos multimedia contenidos en servidores de otros usuarios. En otras realizaciones, un usuario puede solicitar un archivo multimedia y el archivo se puede situar mediante un motor de búsqueda y/o agentes inteligentes que inspeccionan los 'metadatos' de archivos multimedia en diversas ubicaciones. Además, el usuario puede elegir entre diversos archivos multimedia que contienen una presentación multimedia particular basándose en 'metadatos' con respecto a la forma en la que se codificó cada una de las diferentes

versiones de la presentación.

5 En varias realizaciones, se puede acceder a los 'metadatos' de archivos multimedia de acuerdo con algunas realizaciones de la presente invención para fines de catálogo o para crear un menú simple para acceder al contenido del archivo.

10 Aunque la descripción anterior contiene muchas realizaciones específicas de la invención, estas no se deberían interpretar como limitaciones al alcance de la invención, sino más bien como un ejemplo de una realización de la misma. Por ejemplo, un archivo multimedia de acuerdo con una realización de la presente invención puede incluir una única presentación multimedia o múltiples presentaciones multimedia. Además, un archivo de este tipo puede incluir uno o más menús y cualquier diversidad de tipos diferentes de 'metadatos'. Por consiguiente, el alcance de la invención no debería ser determinado por las realizaciones ilustradas, sino por las reivindicaciones adjuntas.

REIVINDICACIONES

1. Un archivo multimedia generado mediante un sistema de cifrado, estando dicho archivo adaptado para gestión de derechos digitales DRM y para su distribución de forma separada de la información requerida para descifrar los contenidos del archivo multimedia, comprendiendo el archivo multimedia:

fragmentos de vídeo que están solo parcialmente cifrados, constando cada uno de dichos fragmentos de vídeo de una porción cifrada y una porción no cifrada;

fragmentos de 'DRM' que contienen cada uno una referencia a la porción cifrada de un fragmento de vídeo de dichos fragmentos de vídeo y una referencia a una clave de descifrado para descifrar la porción cifrada de dicho fragmento de vídeo, en donde las claves de descifrado se aleatorizan y se cifran con una clave maestra; y

un fragmento ('strd', 96) que comprende un encabezamiento de 'DRM', estando dichas claves de descifrado situadas en dicho encabezamiento de 'DRM', conteniendo dicho encabezamiento de 'DRM' información para identificar la clave maestra.

2. El archivo multimedia de la reivindicación 1, en donde:

el fragmento de 'DRM' (270) puede incluir un valor de 'trama' (280), un valor de 'estado' (282), un valor de 'desplazamiento' (284), un valor de 'número' (286) y un valor de 'clave' (288);

en donde el valor de 'trama' (280) se puede usar para hacer referencia a la trama cifrada de vídeo;

en donde el valor de 'estado' (282) se puede usar para indicar si la trama está cifrada;

en donde el valor de 'desplazamiento' (284) apunta al inicio del bloque cifrado dentro de la trama;

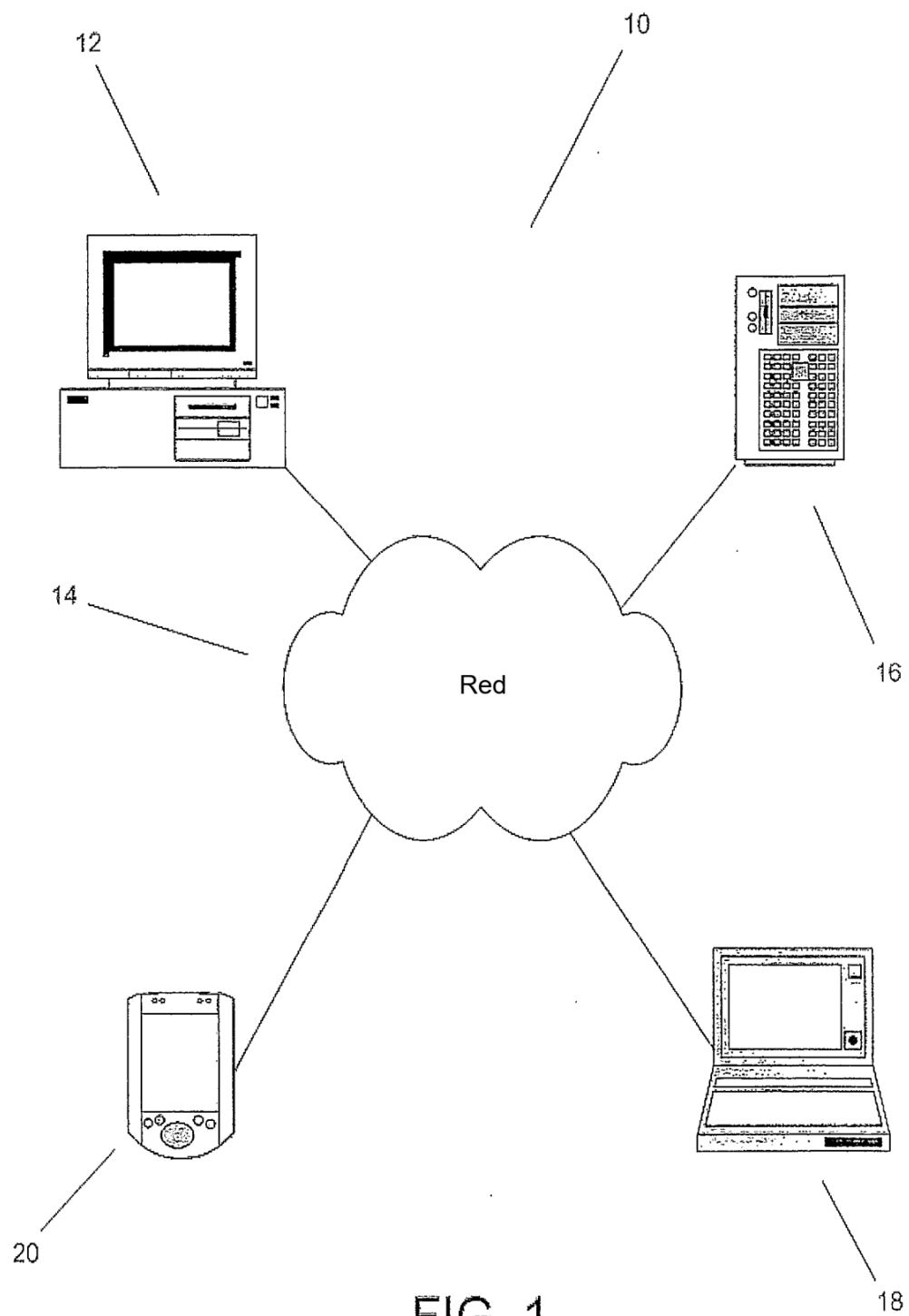
en donde el valor de 'número' (286) indica el número de bytes cifrados en el bloque; y

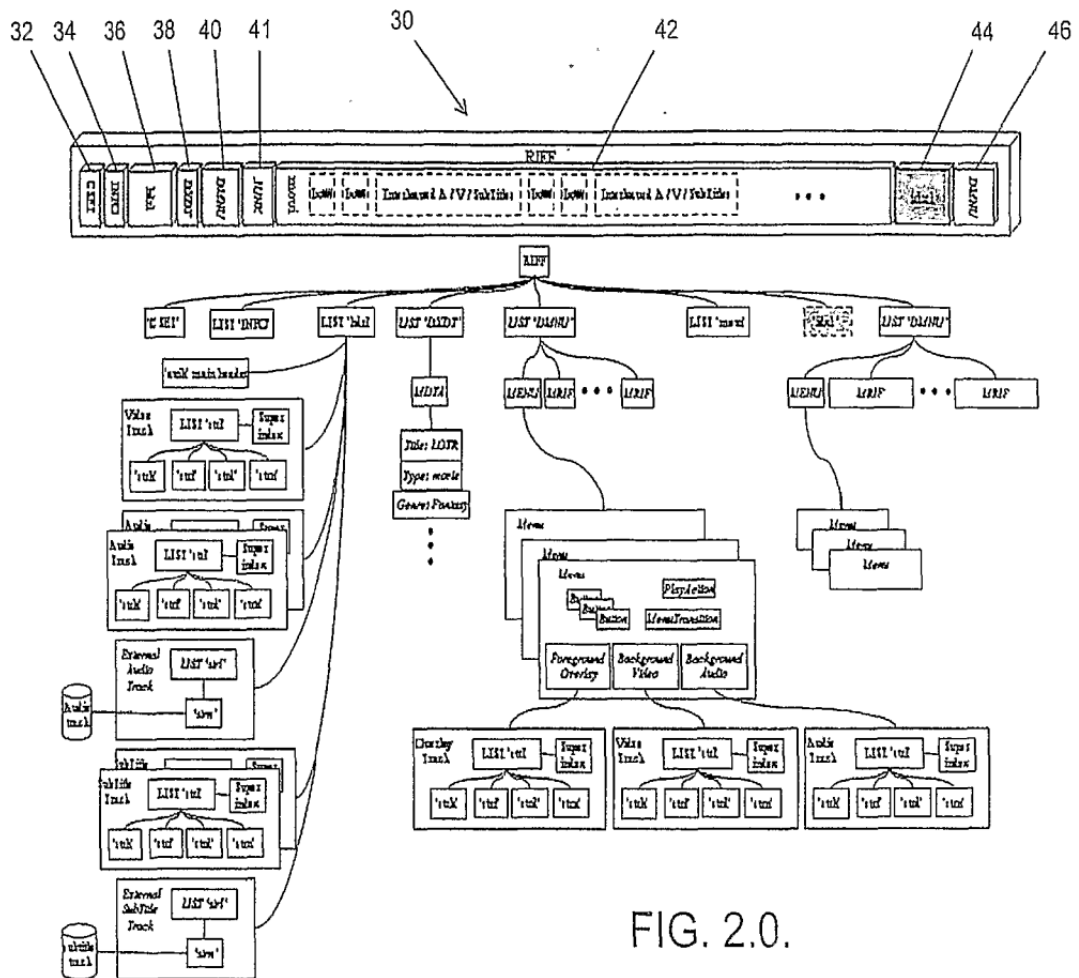
en donde el valor de 'clave' (288) hace referencia a la clave de descifrado que se puede usar para descifrar el bloque.

3. El archivo multimedia de la reivindicación 1, en donde a medida que se genera cada fragmento de vídeo, dicho fragmento de vídeo se puede cifrar y se genera un fragmento de DRM que contiene información con respecto al cifrado del fragmento de vídeo.

4. El archivo multimedia de la reivindicación 1, que comprende además un intercalador configurado para intercalar un fragmento de DRM que contiene información con respecto al cifrado de un fragmento de vídeo antes del fragmento de vídeo.

5. El archivo multimedia de la reivindicación 1, en donde el fragmento de DRM para el fragmento de vídeo n está situado entre el fragmento de vídeo $n-1$ y el fragmento de vídeo n .





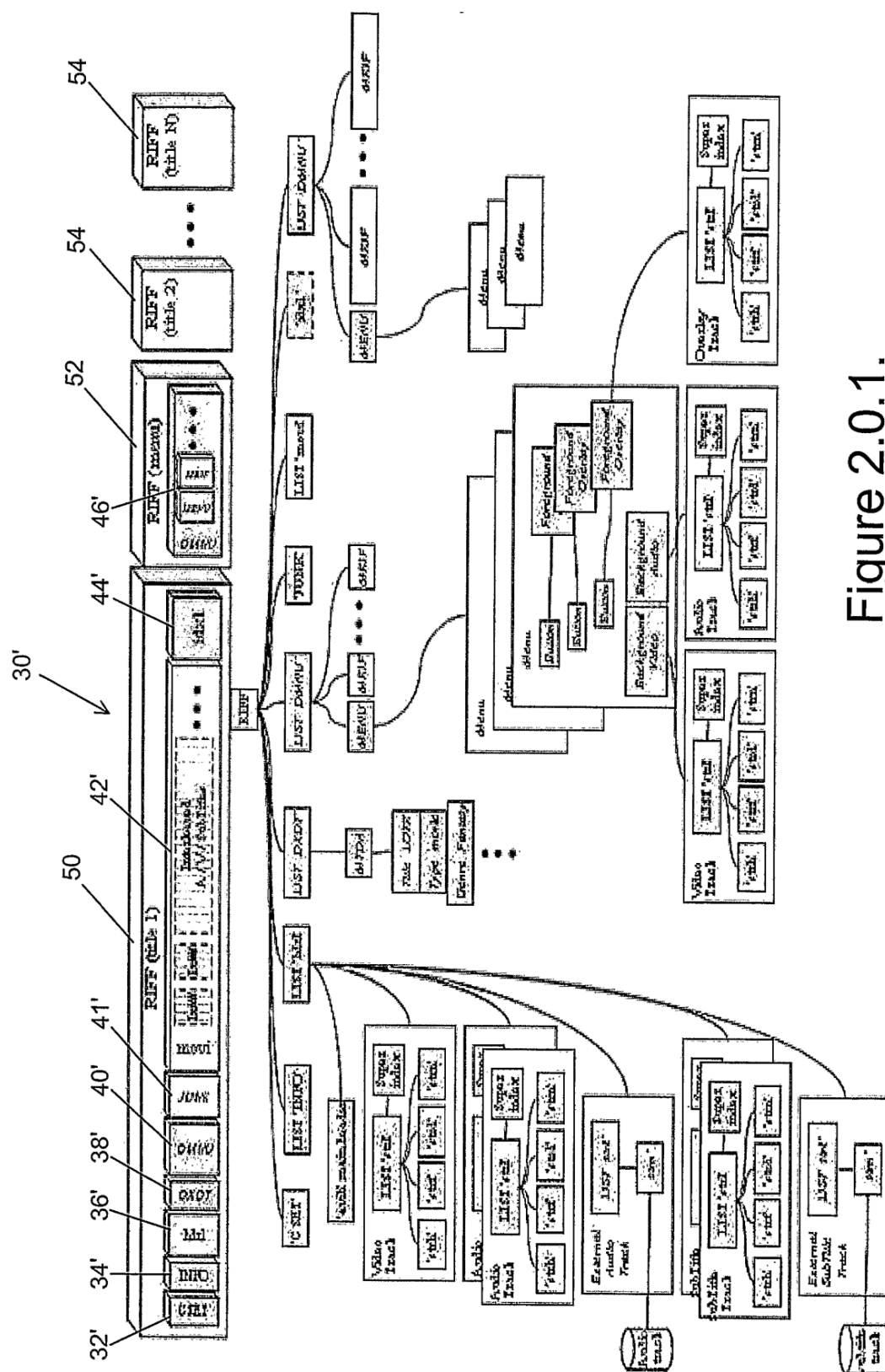


Figure 2.0.1.

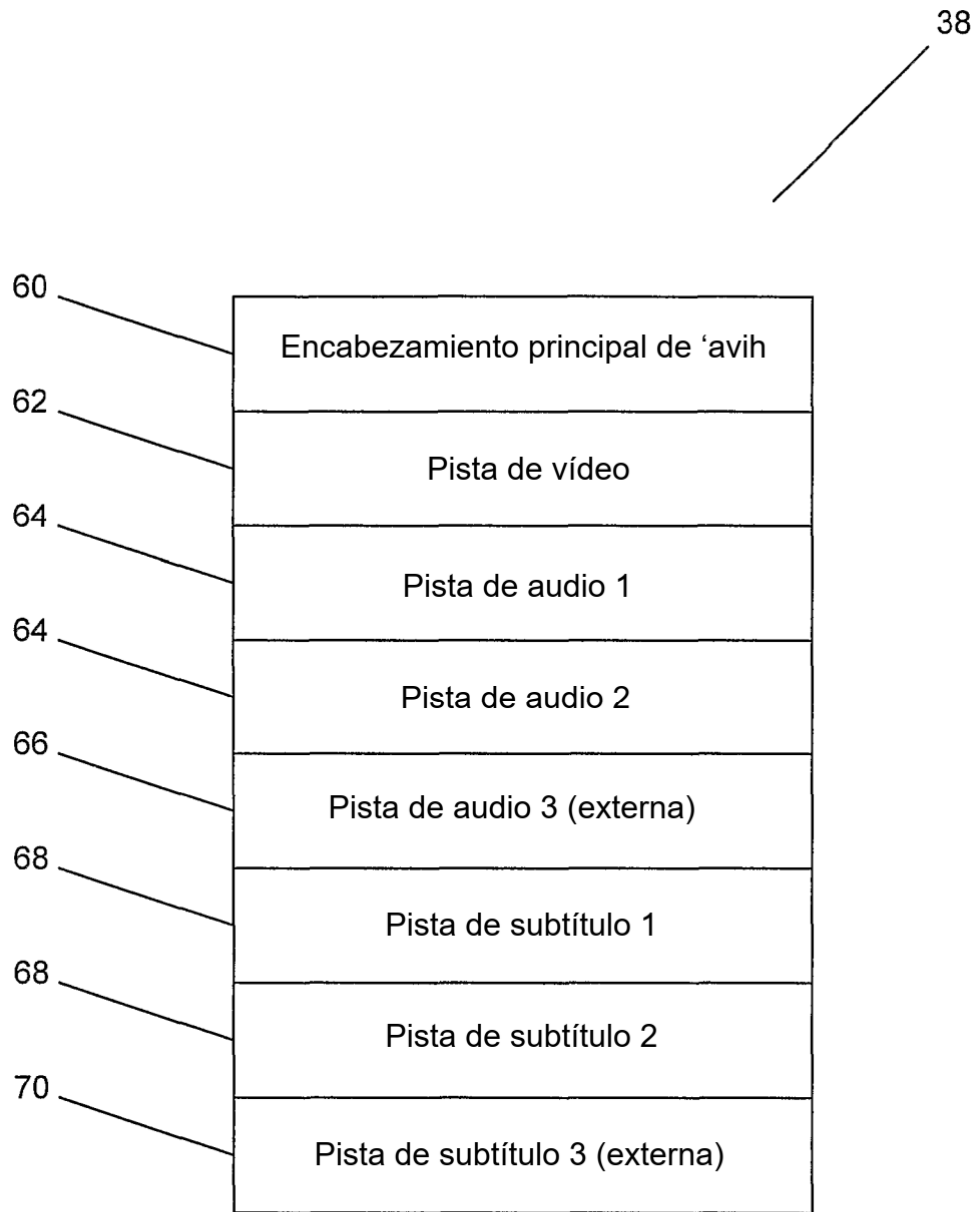


Fig. 2.1.

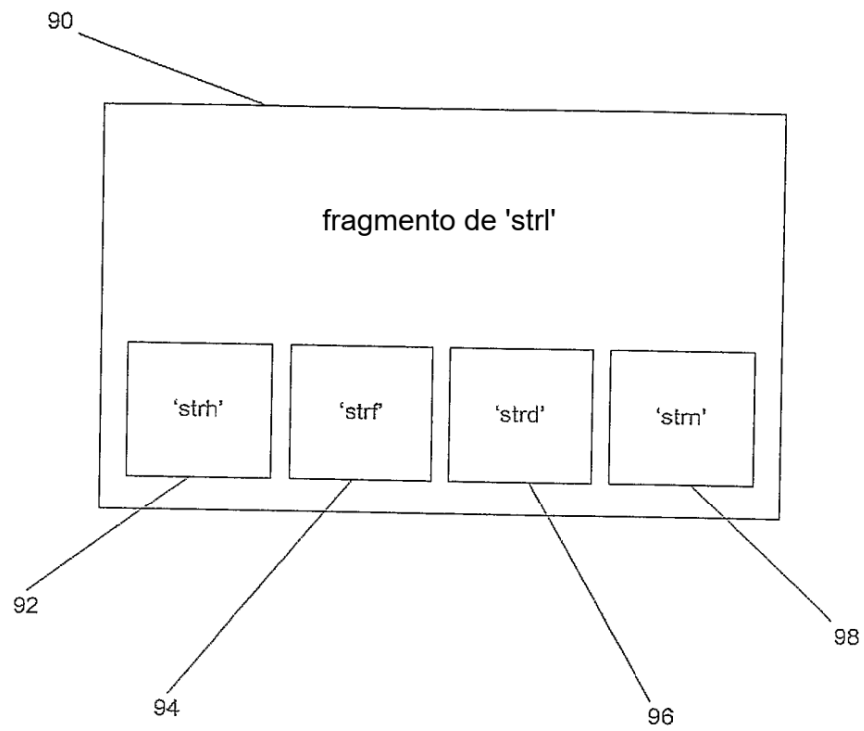


FIG. 2.2.

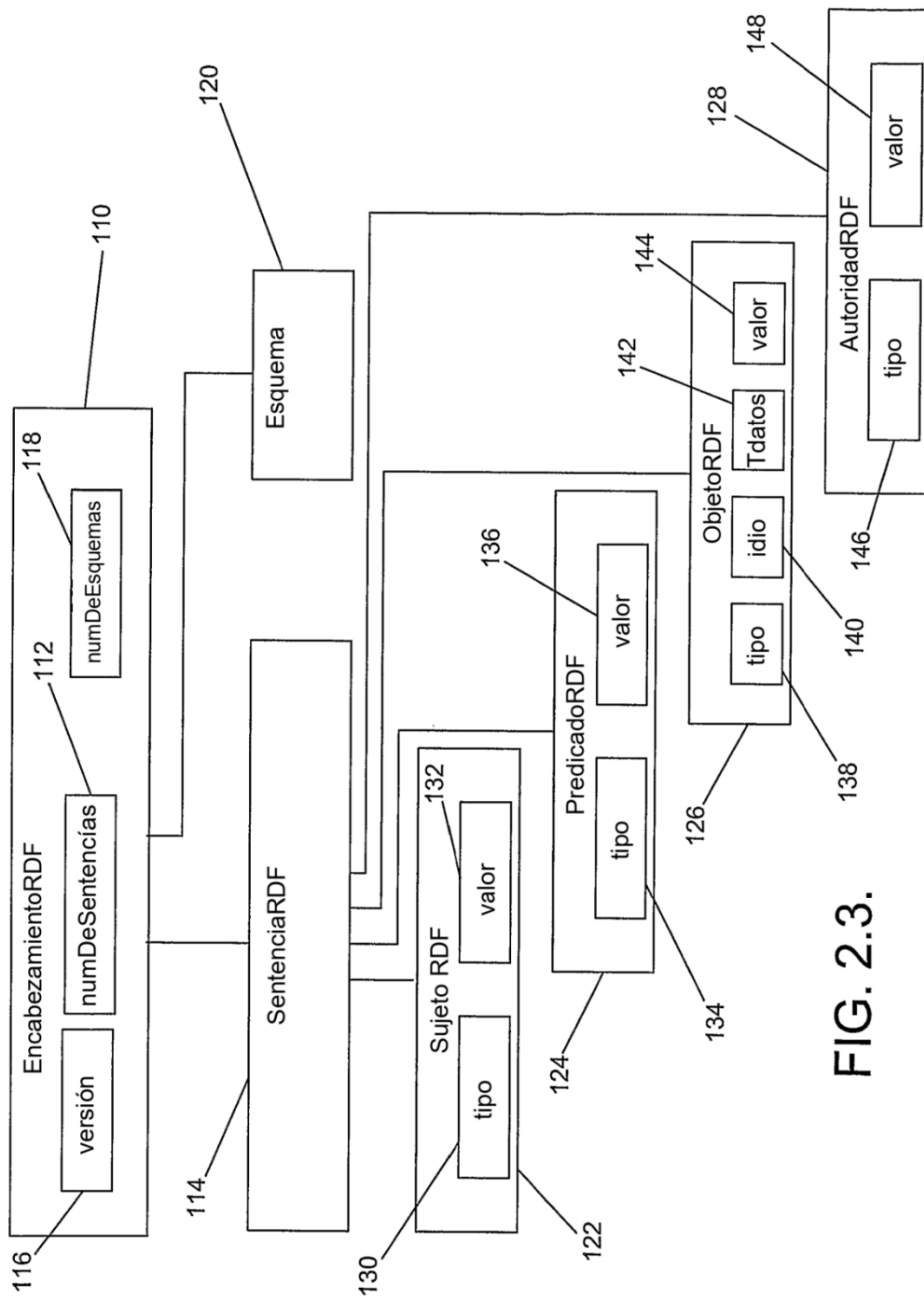


FIG. 2.3.

FIG. 2.3.1.

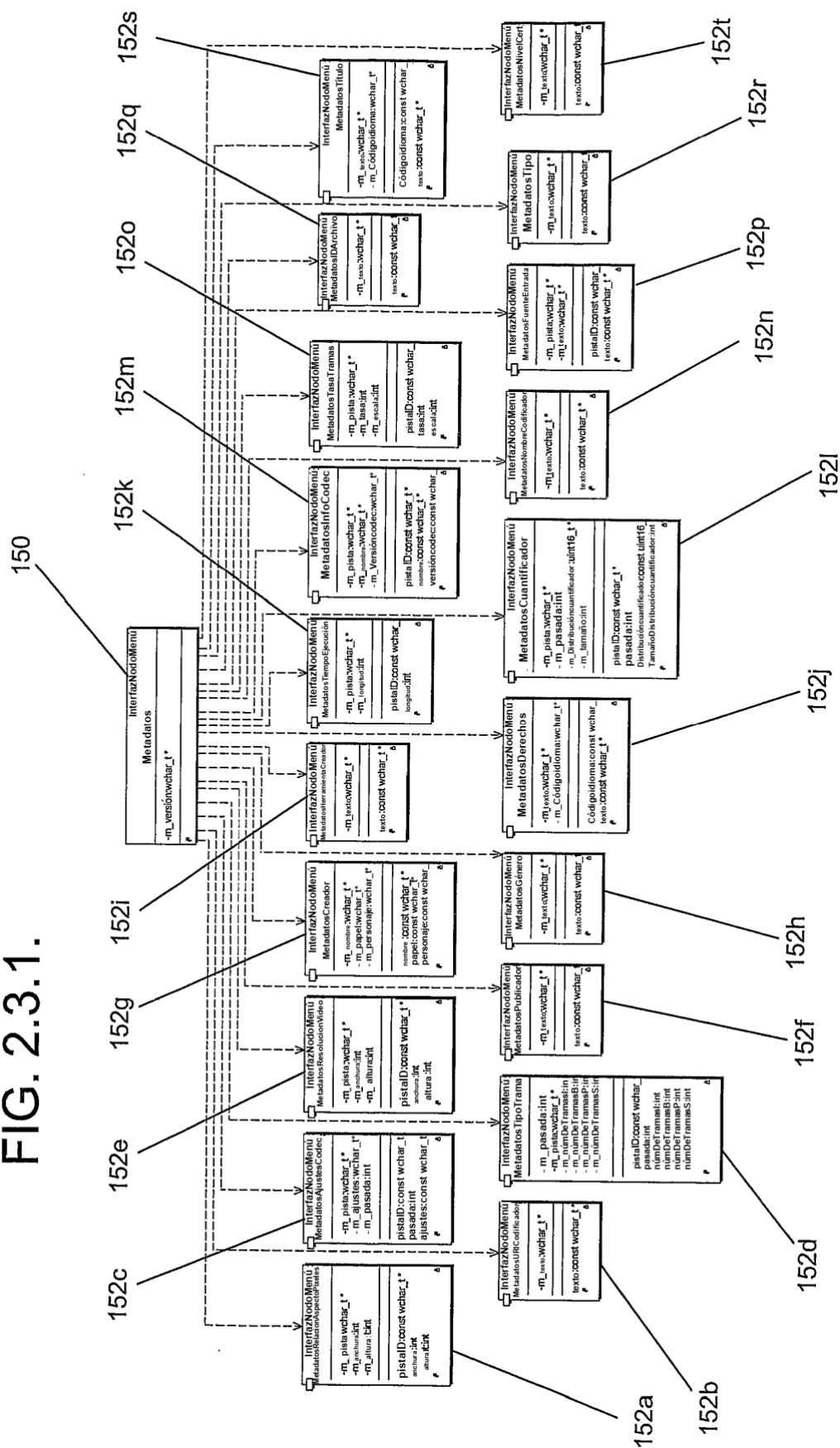
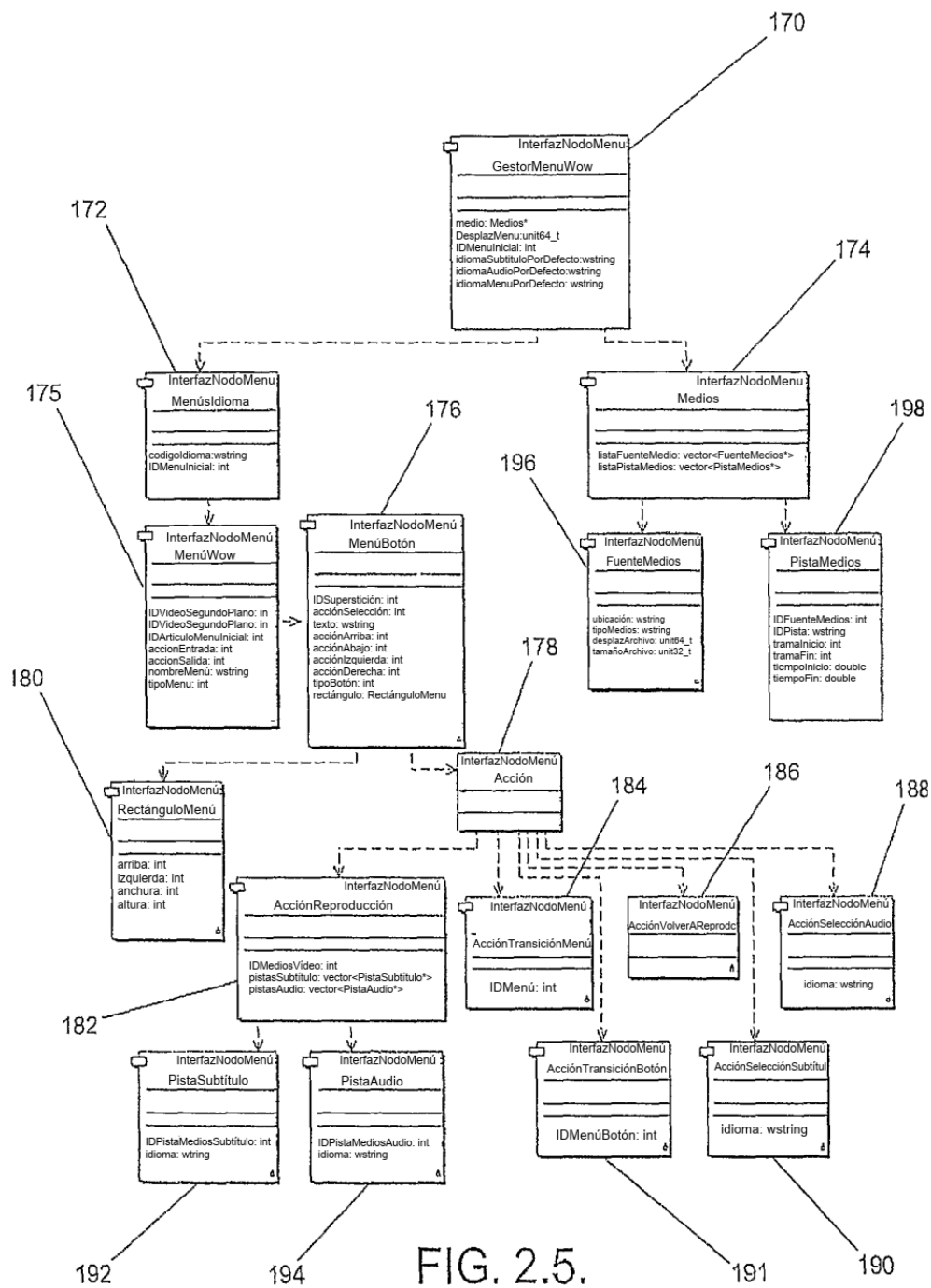




FIG. 2.4.



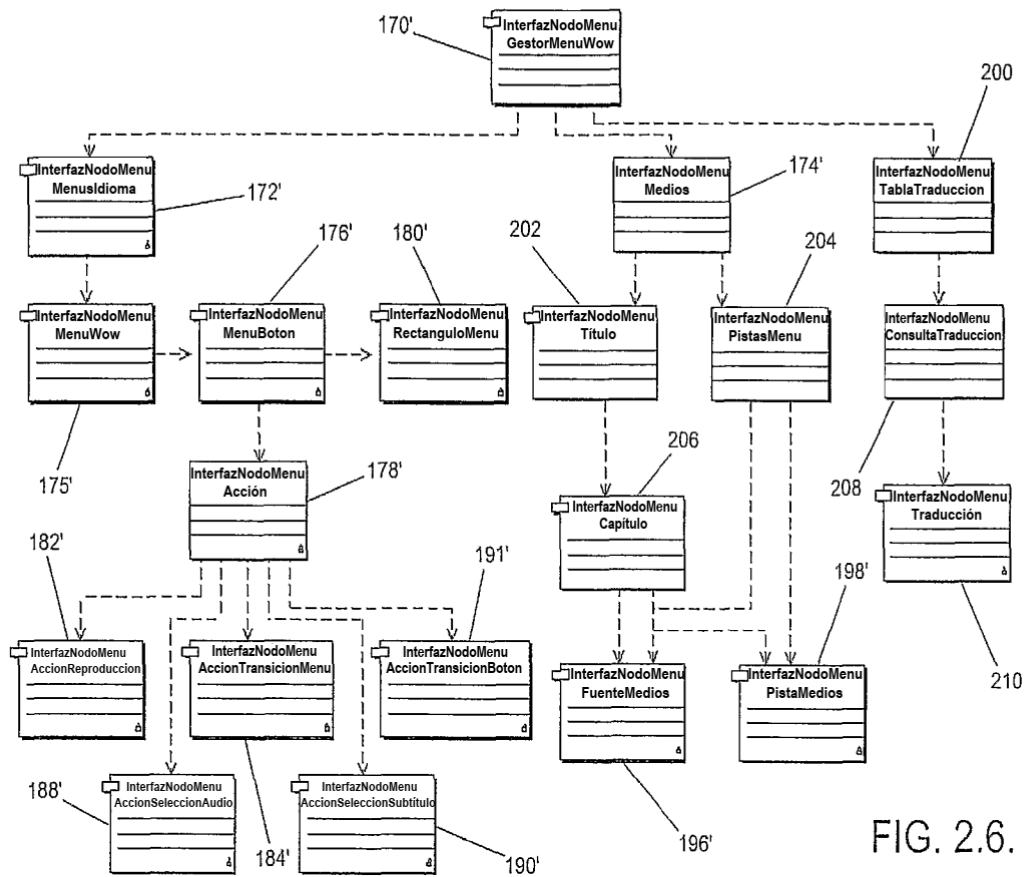


FIG. 2.6.

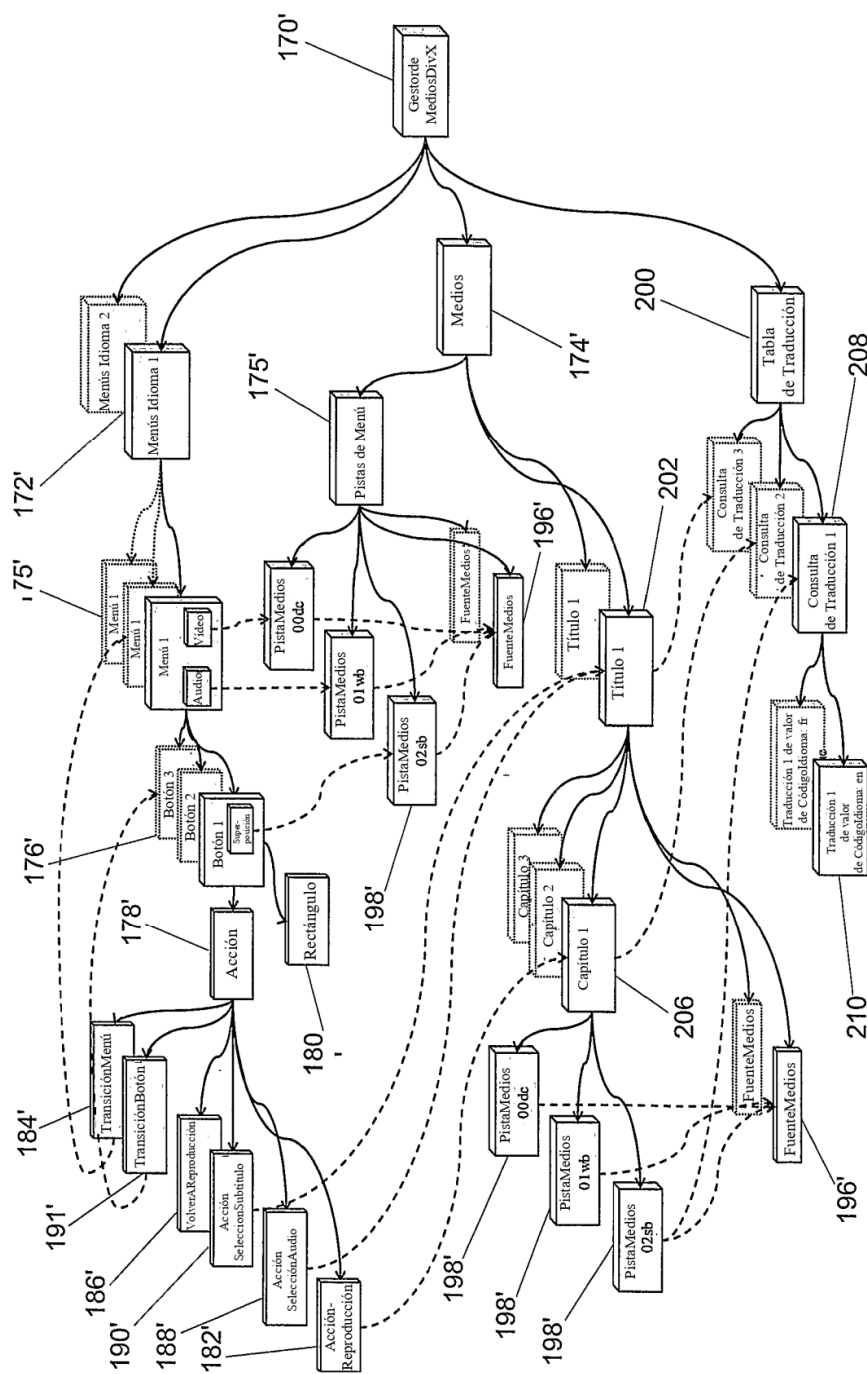
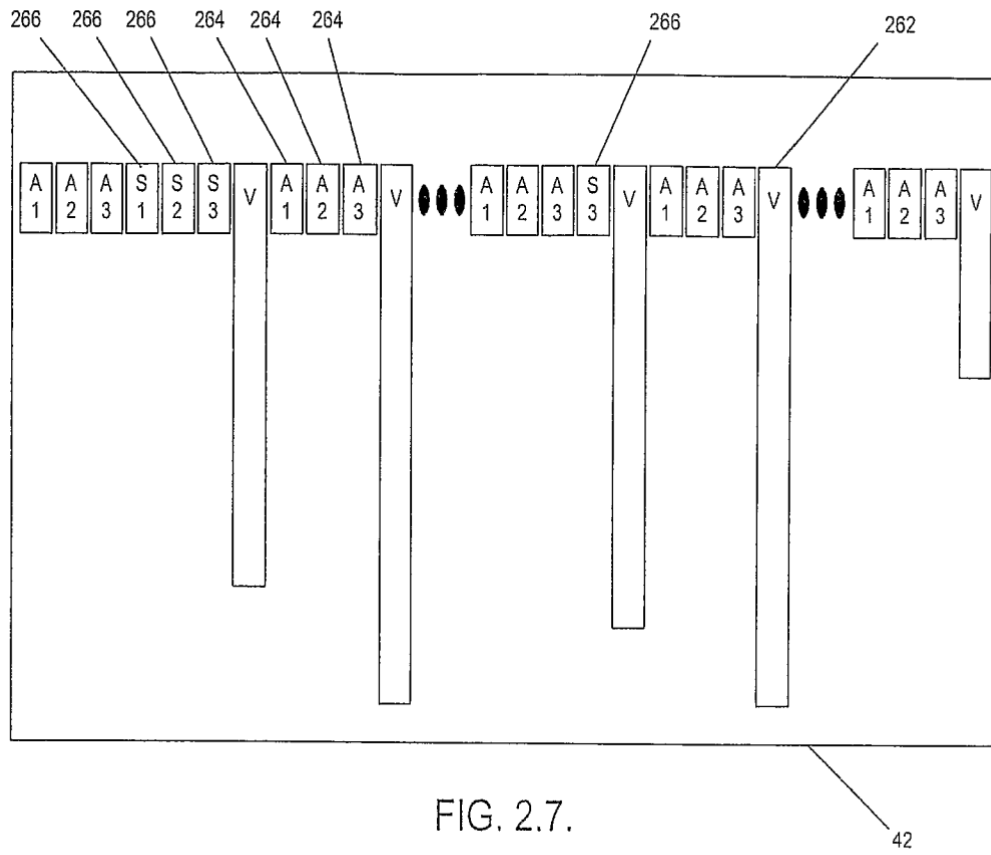
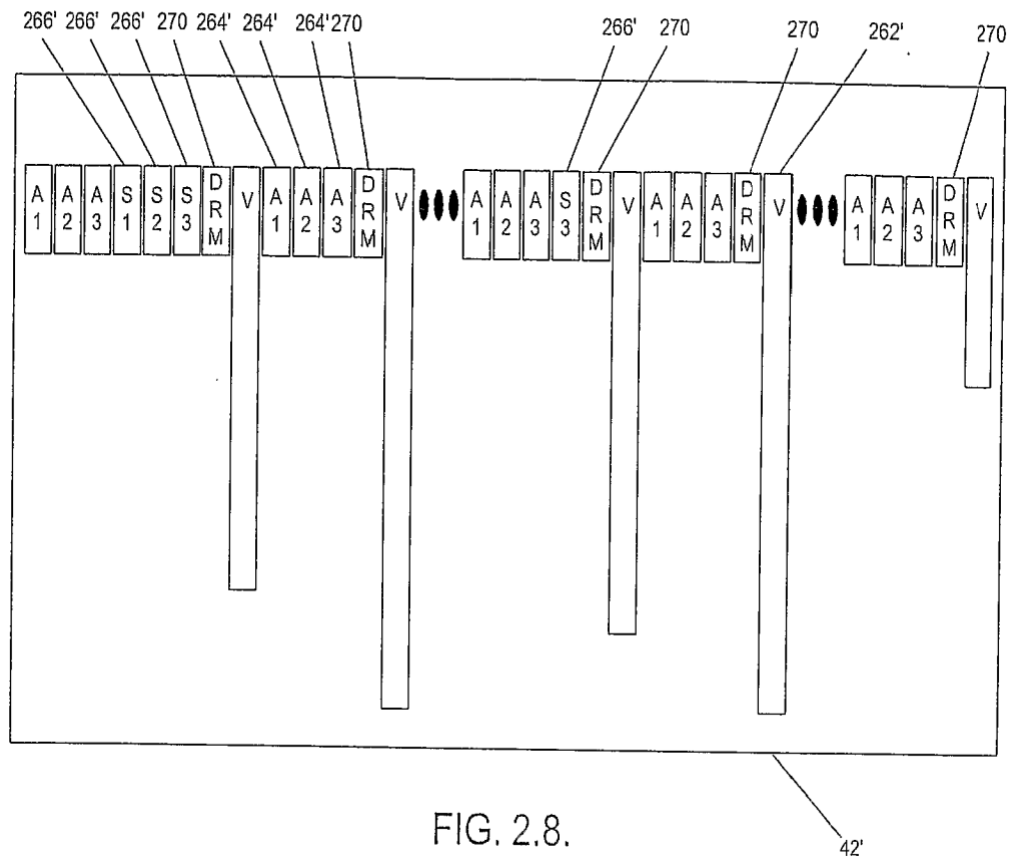


FIG. 2.6.1.





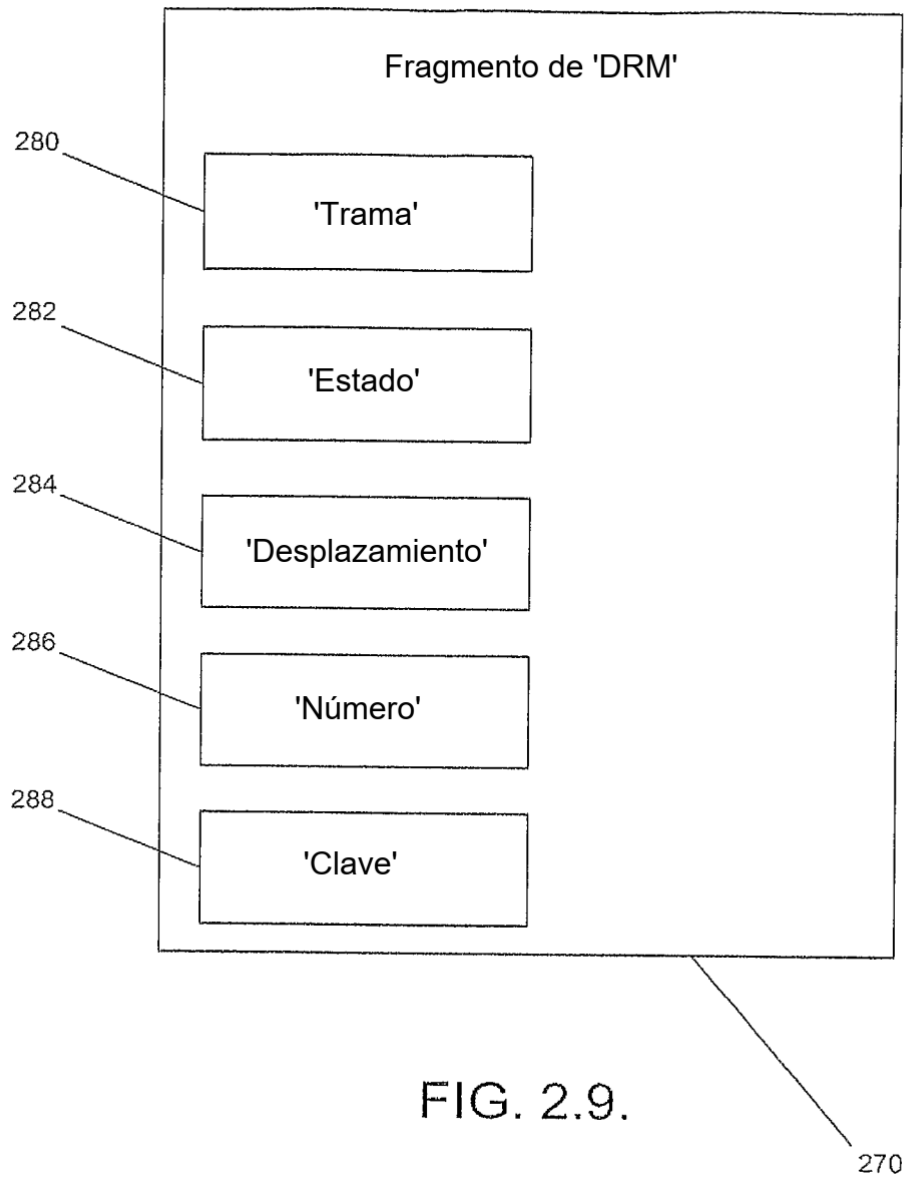
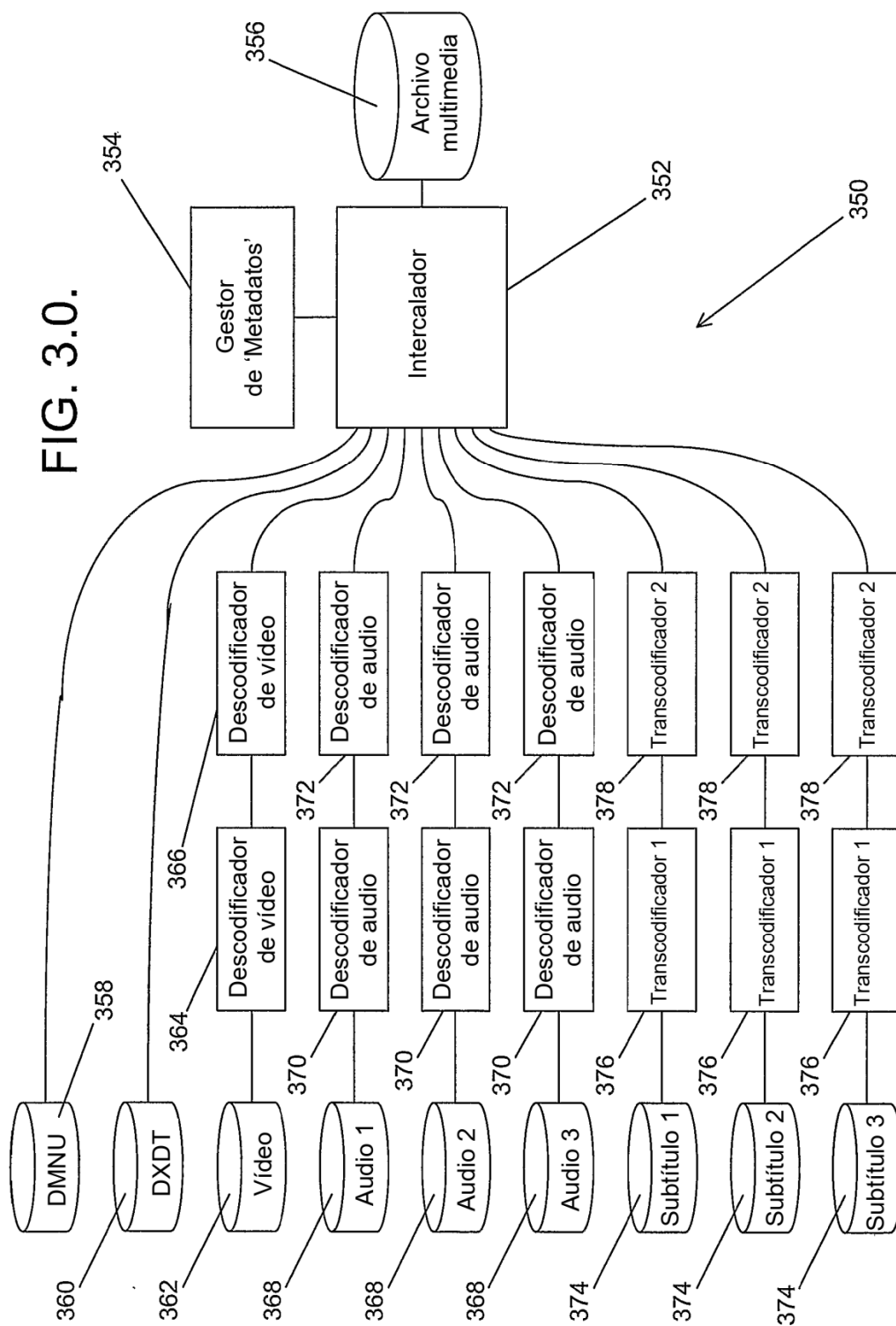


FIG. 3.0.



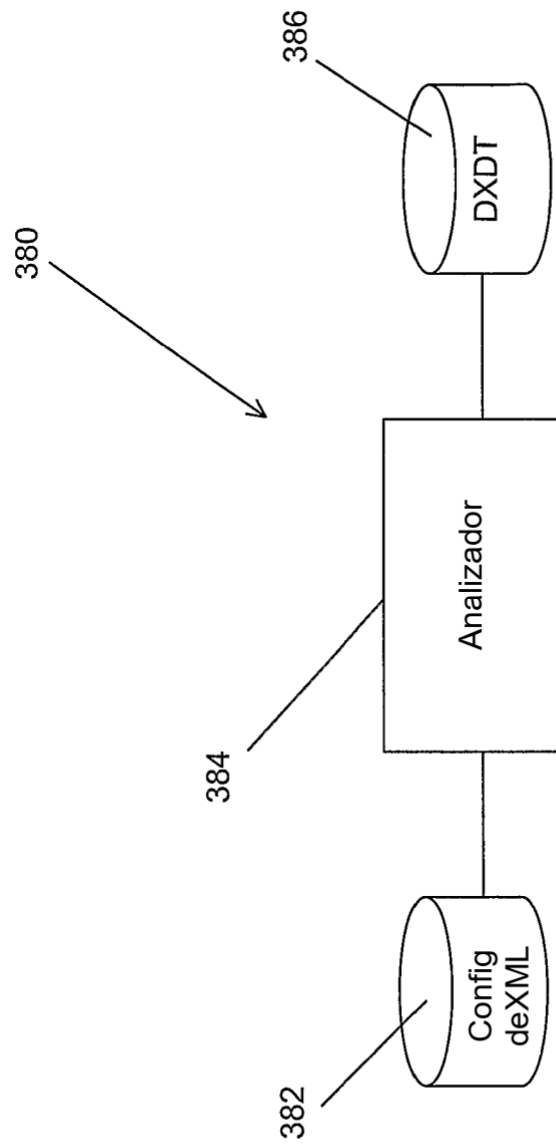


FIG. 3.1.

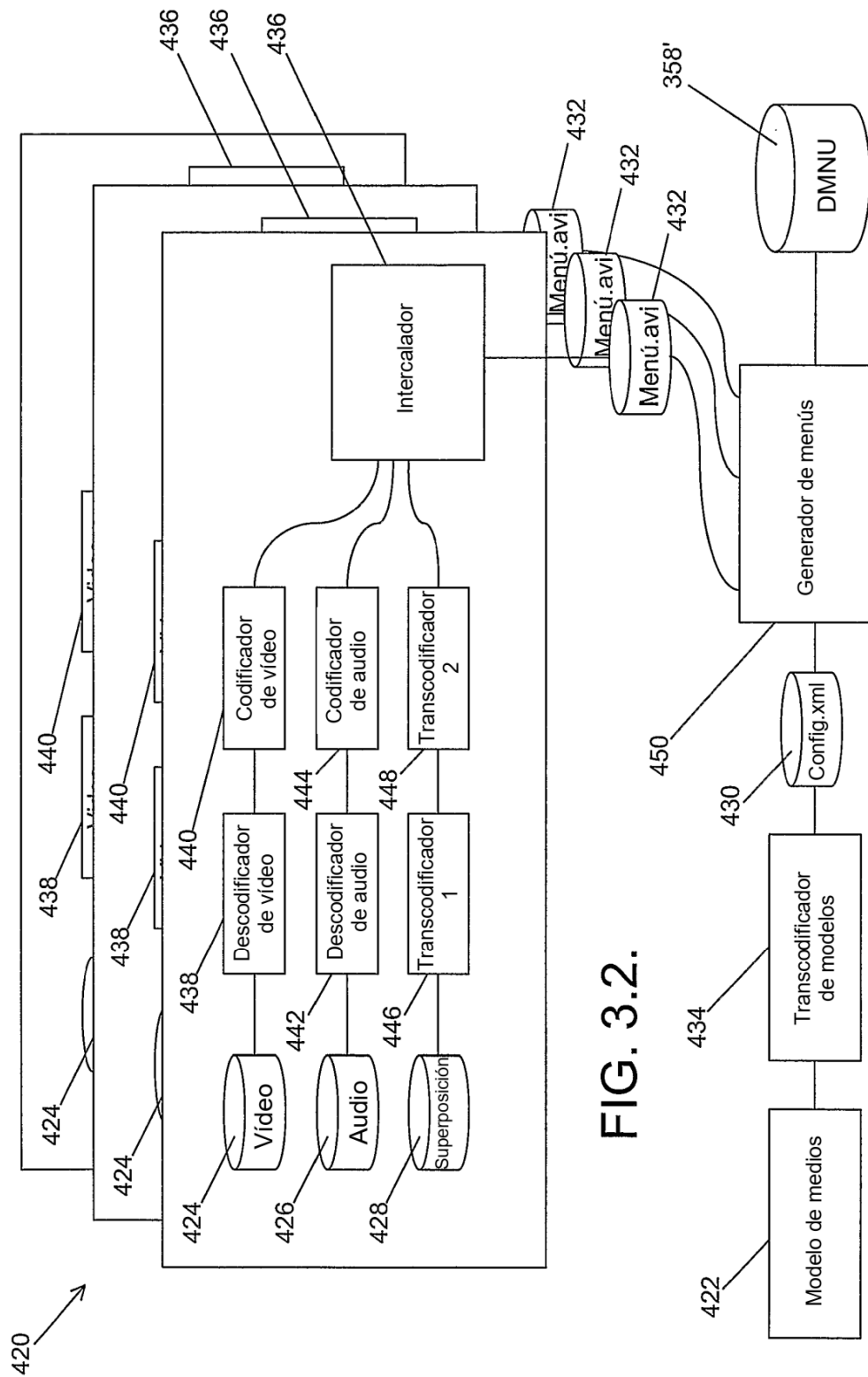


FIG. 3.2.

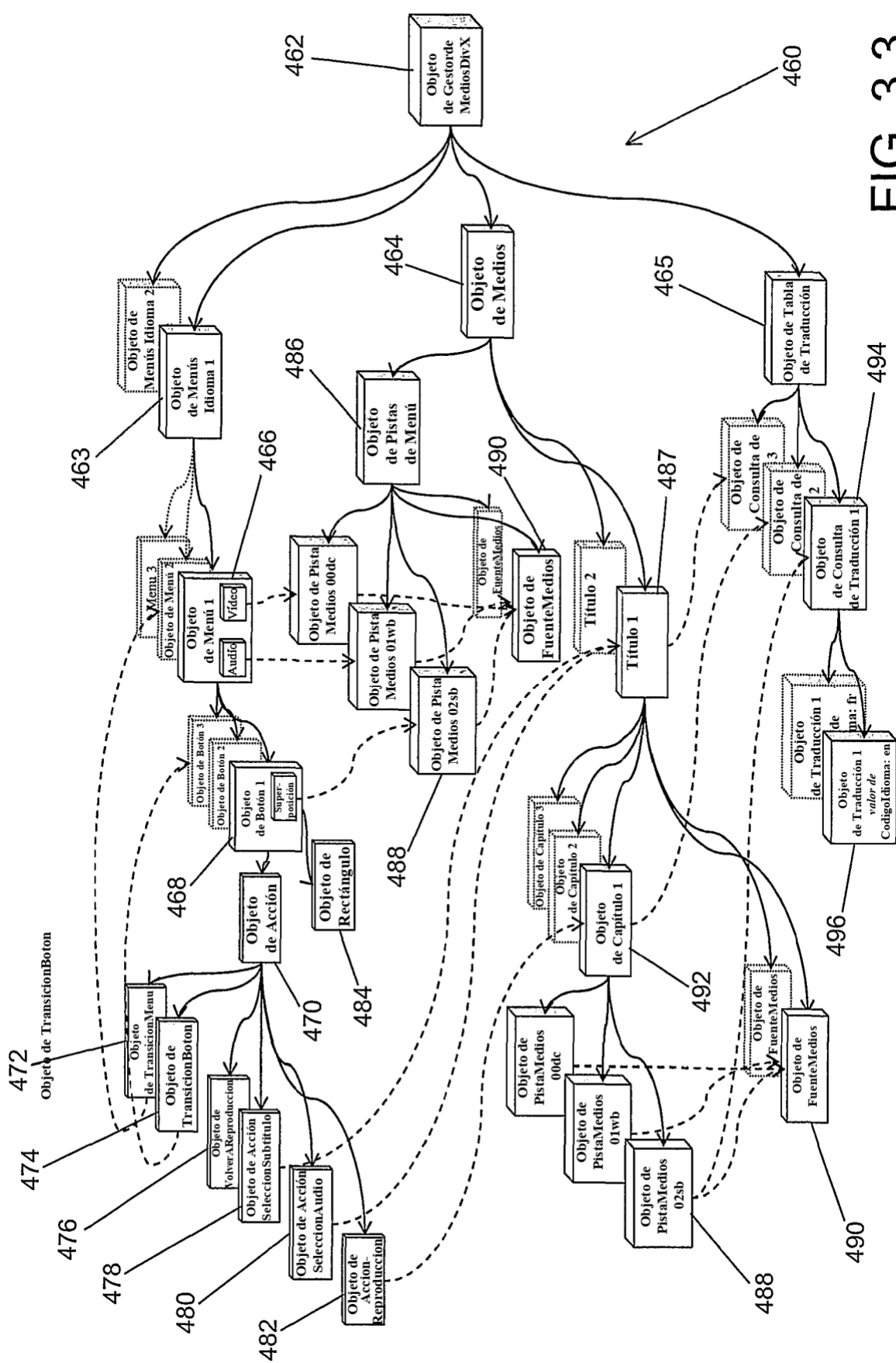


FIG. 3.3.

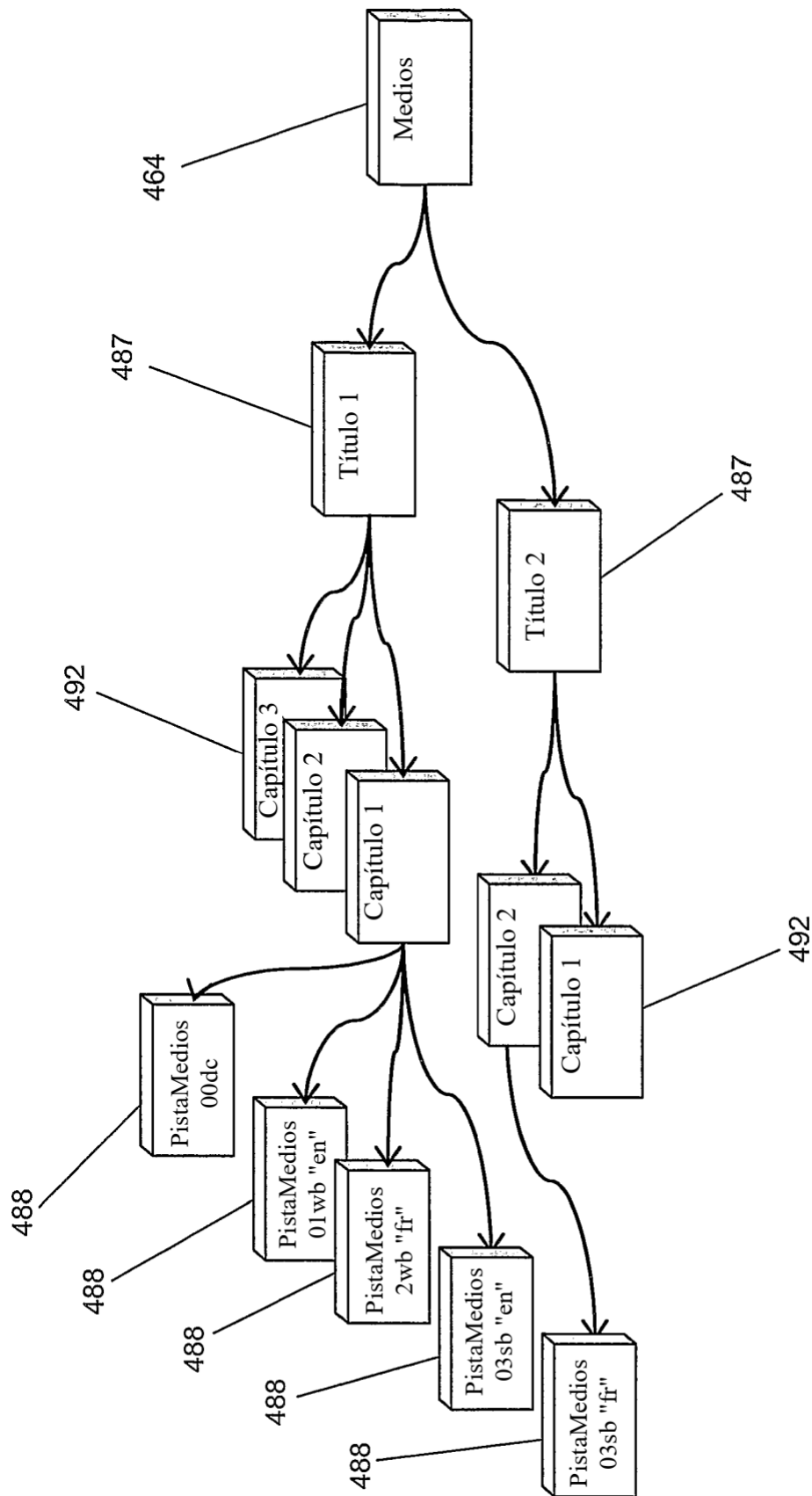


FIG. 3.3.1.

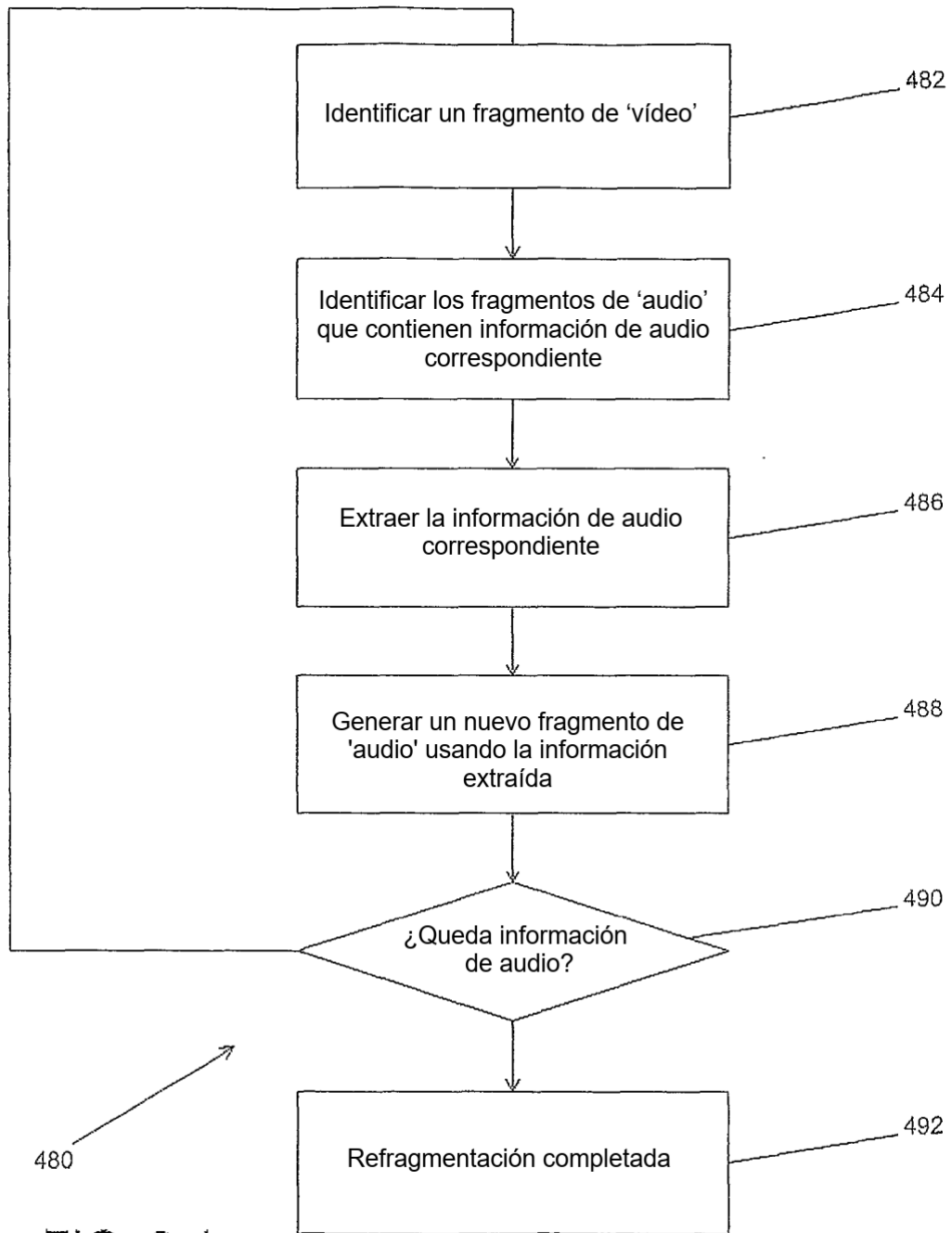
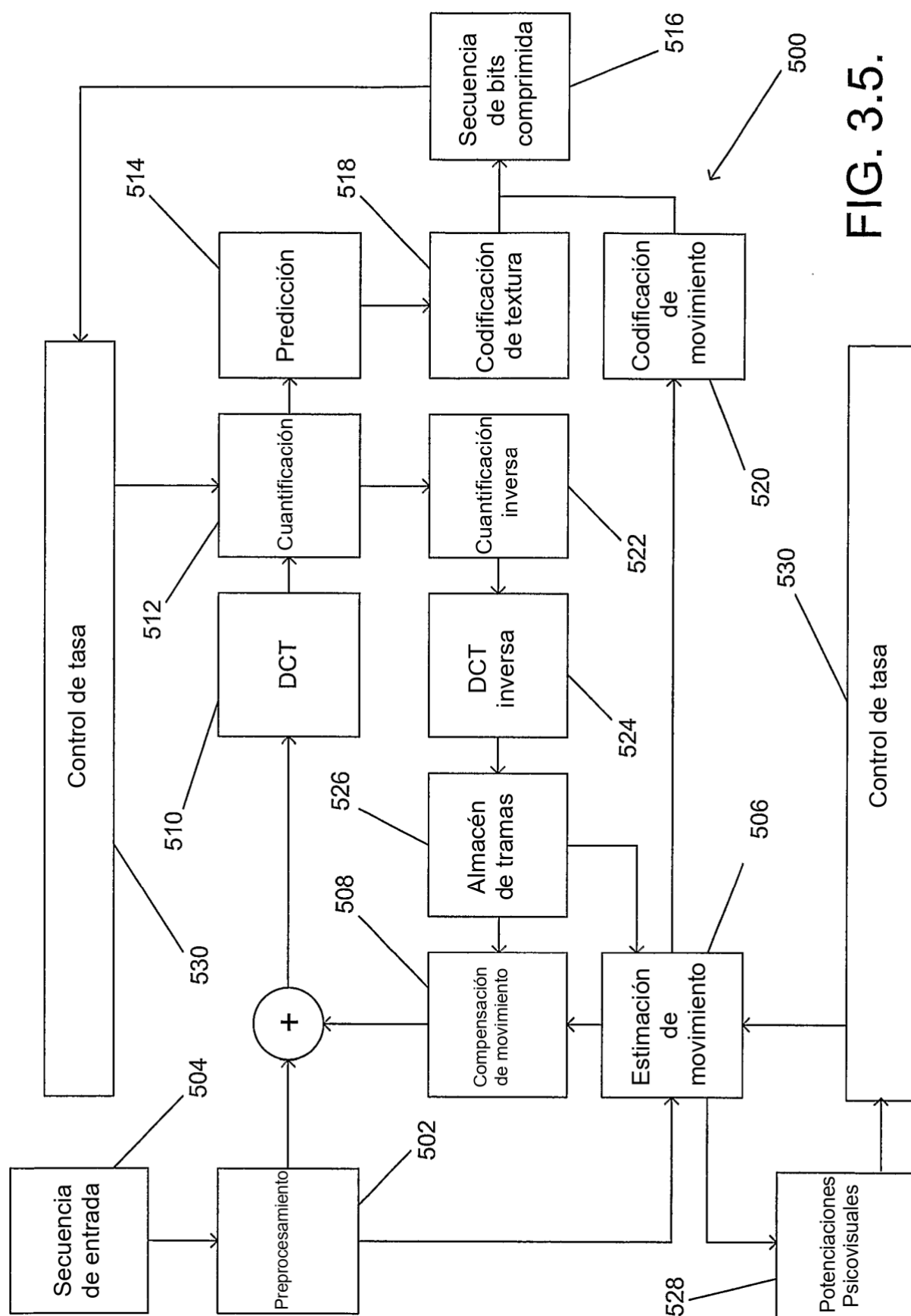


FIG. 3.4.



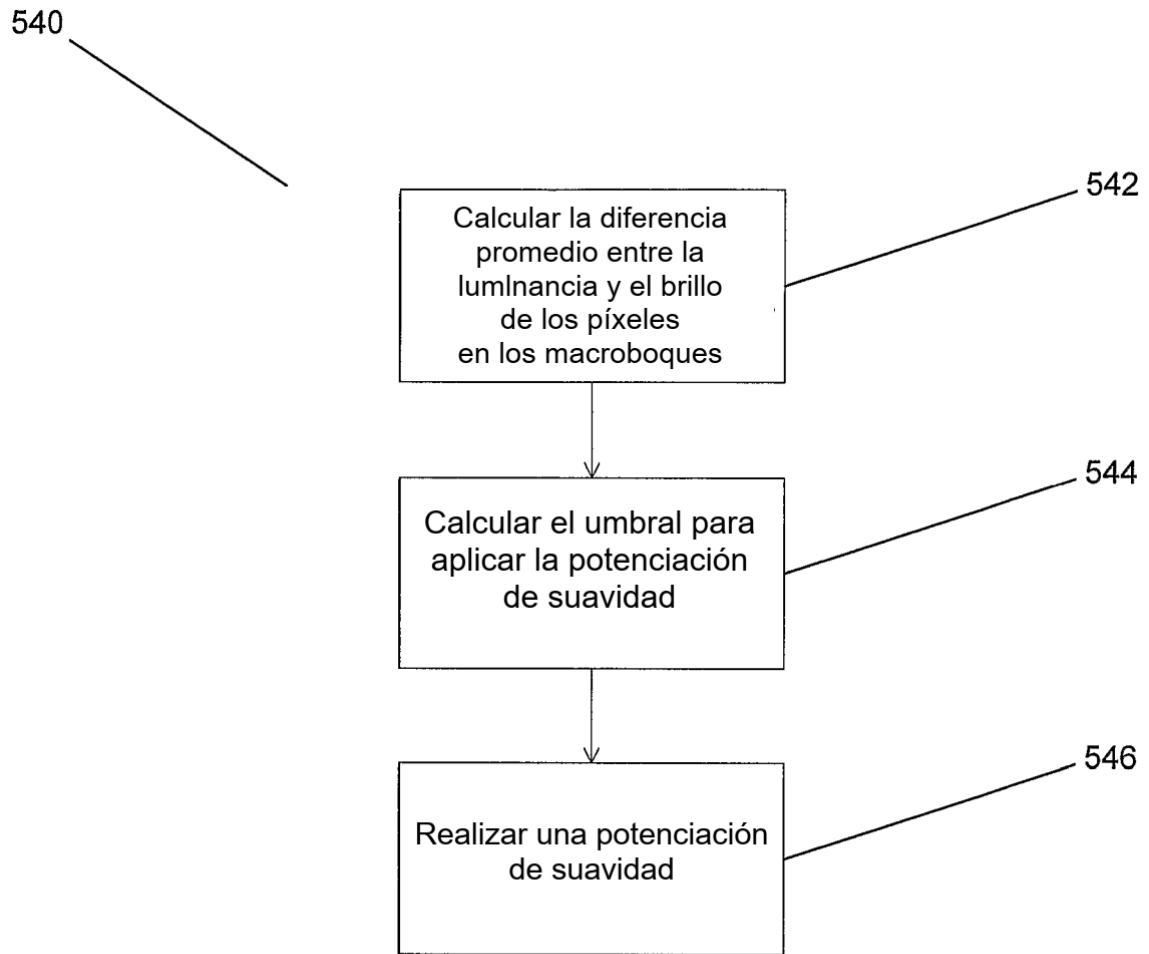


FIG. 3.6.

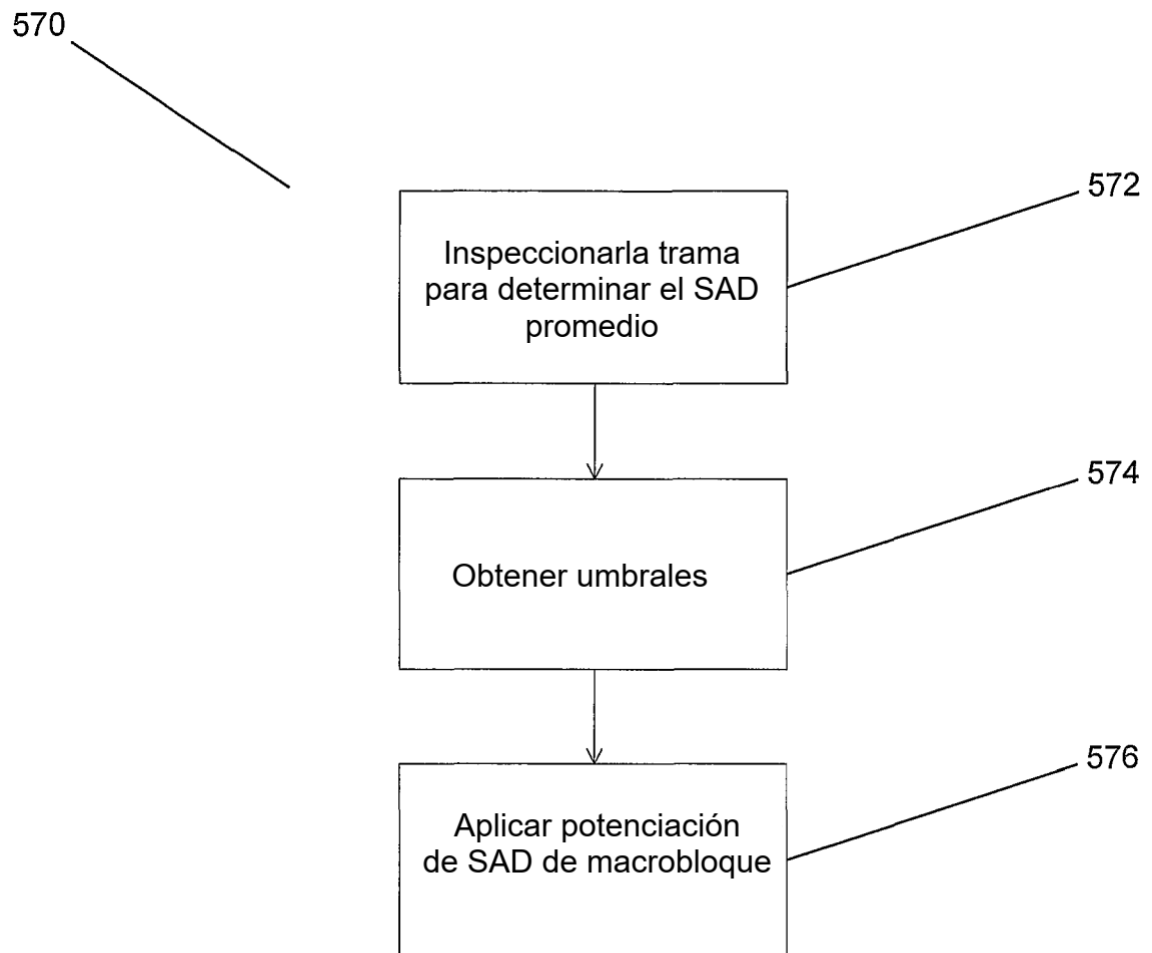


FIG. 3.7.

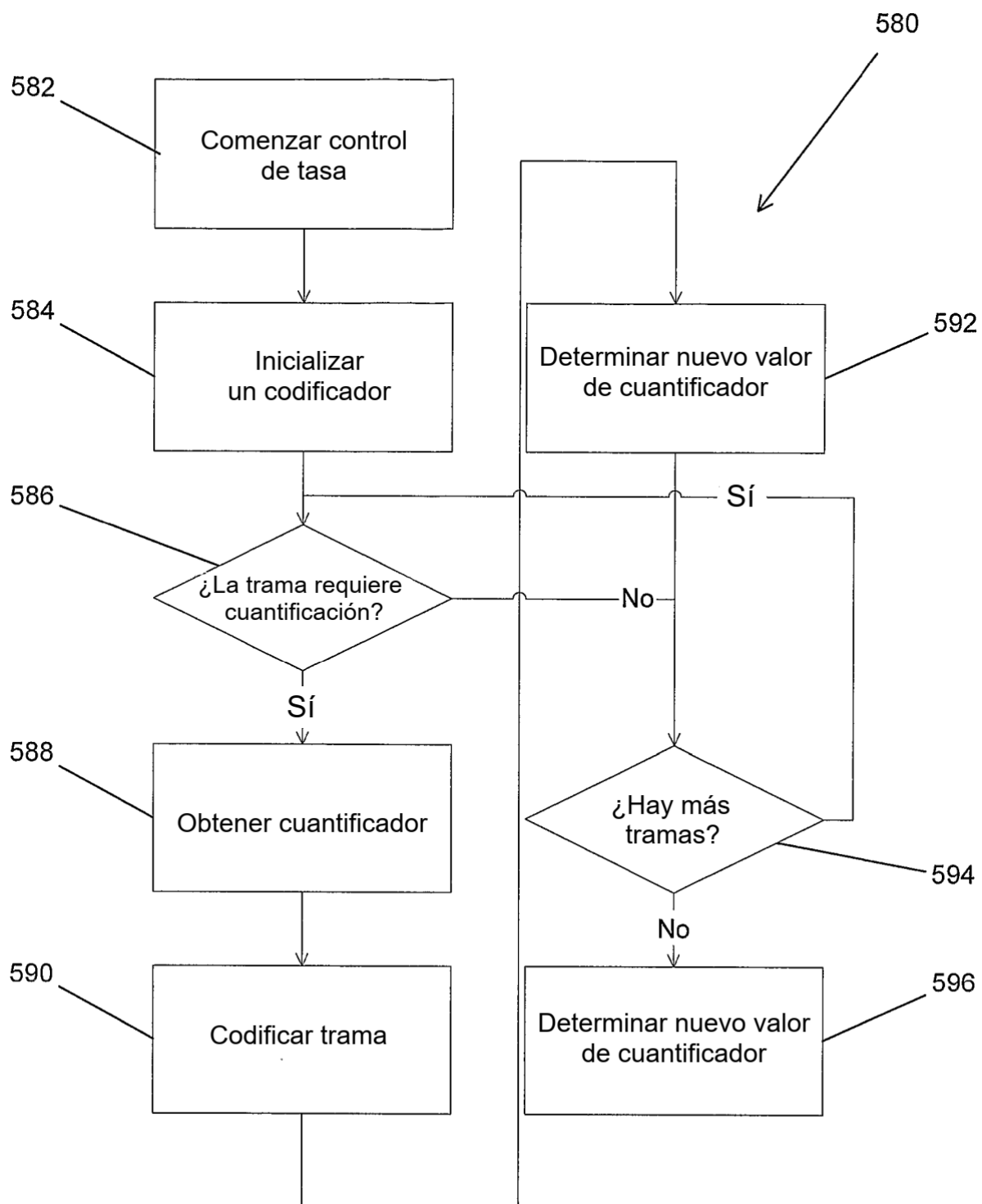


FIG. 3.8.

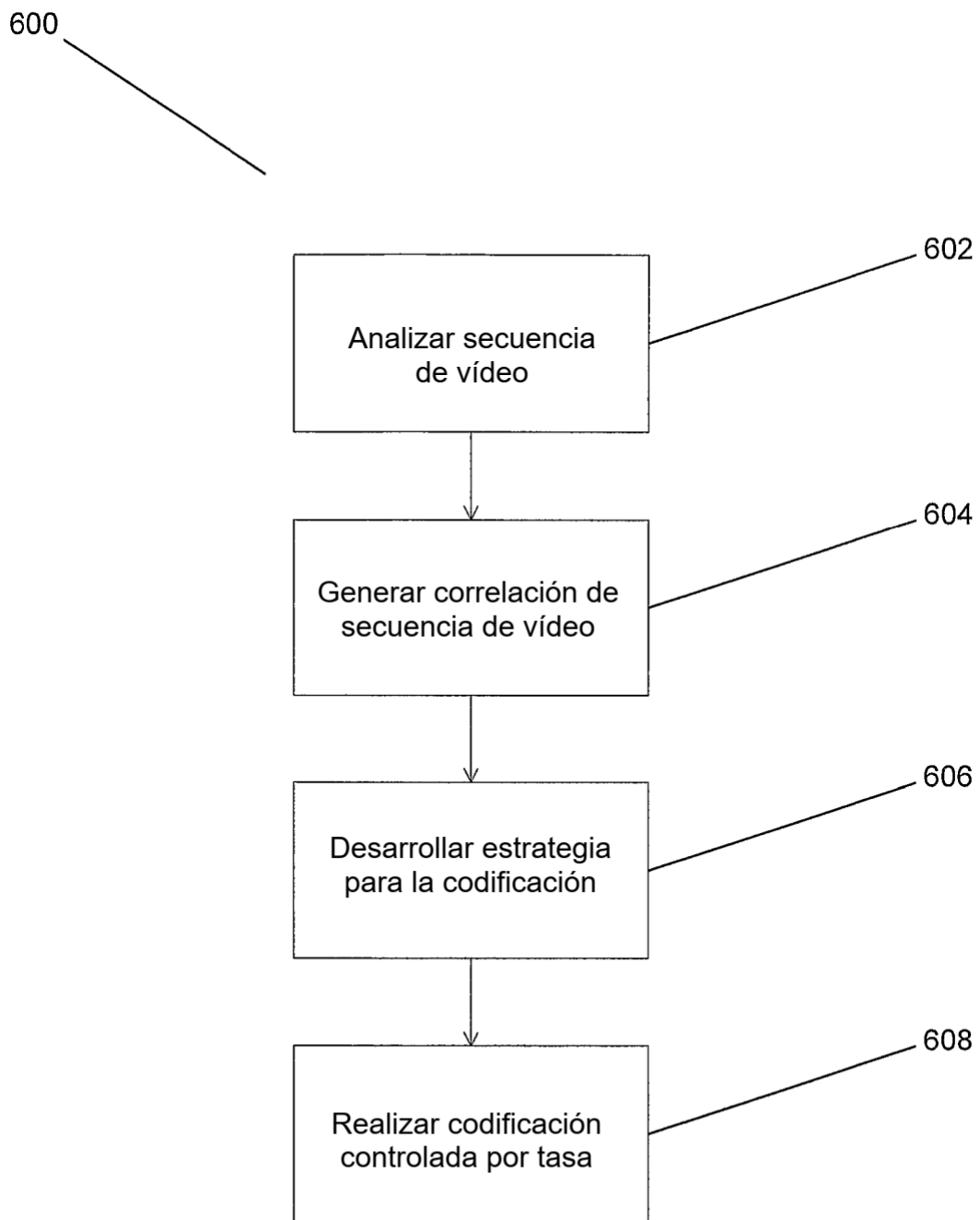


FIG. 3.9.

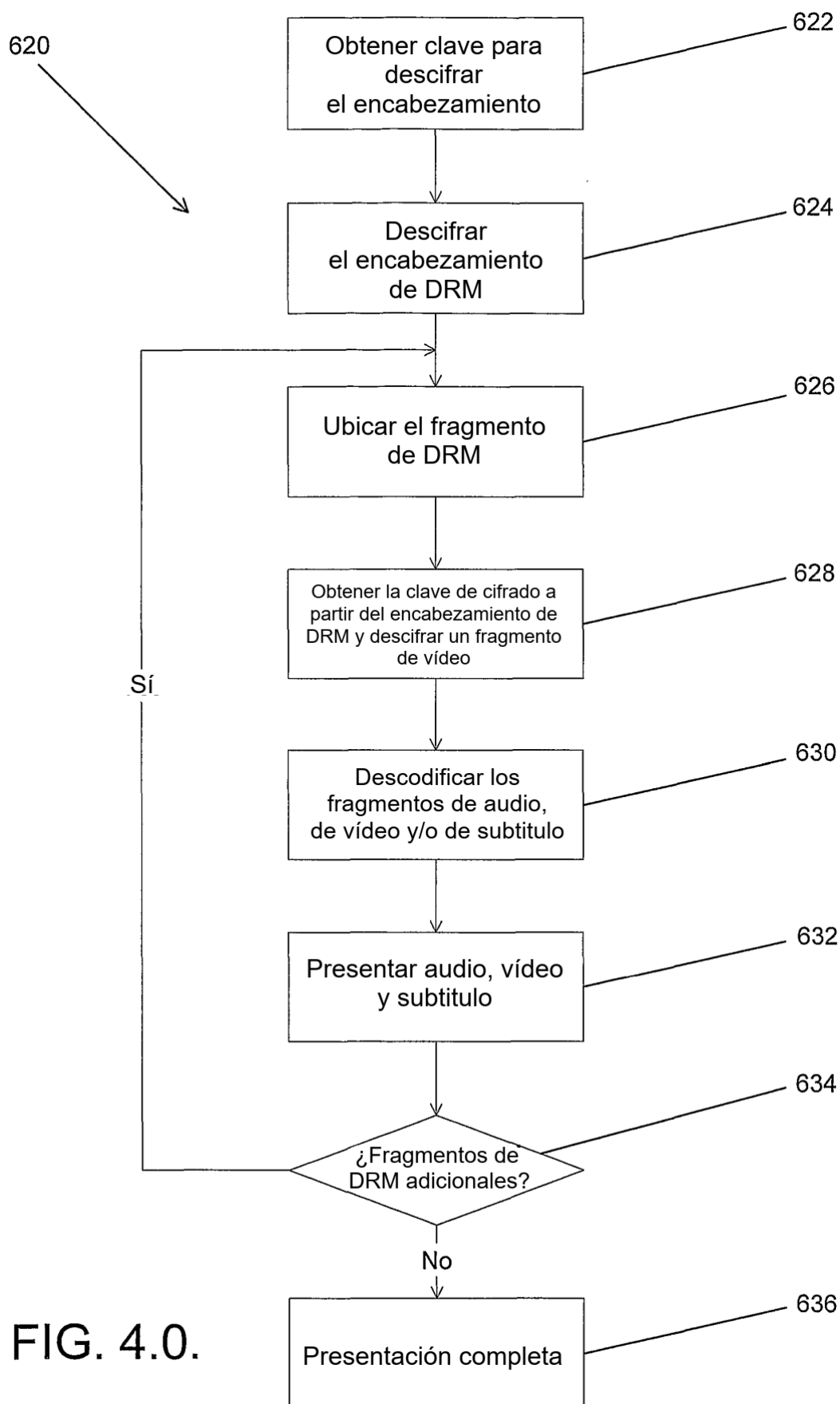


FIG. 4.0.

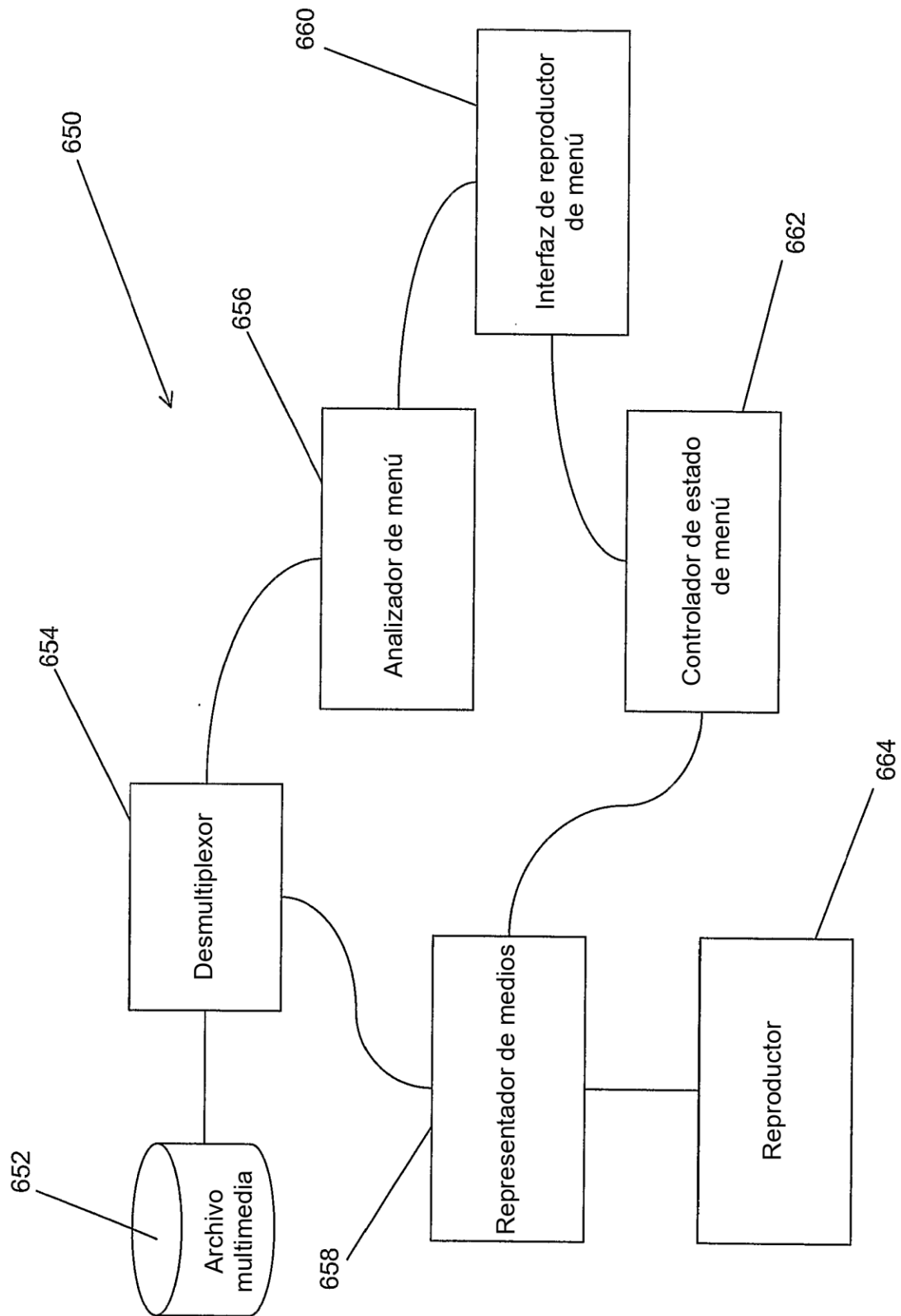


FIG. 4.1.

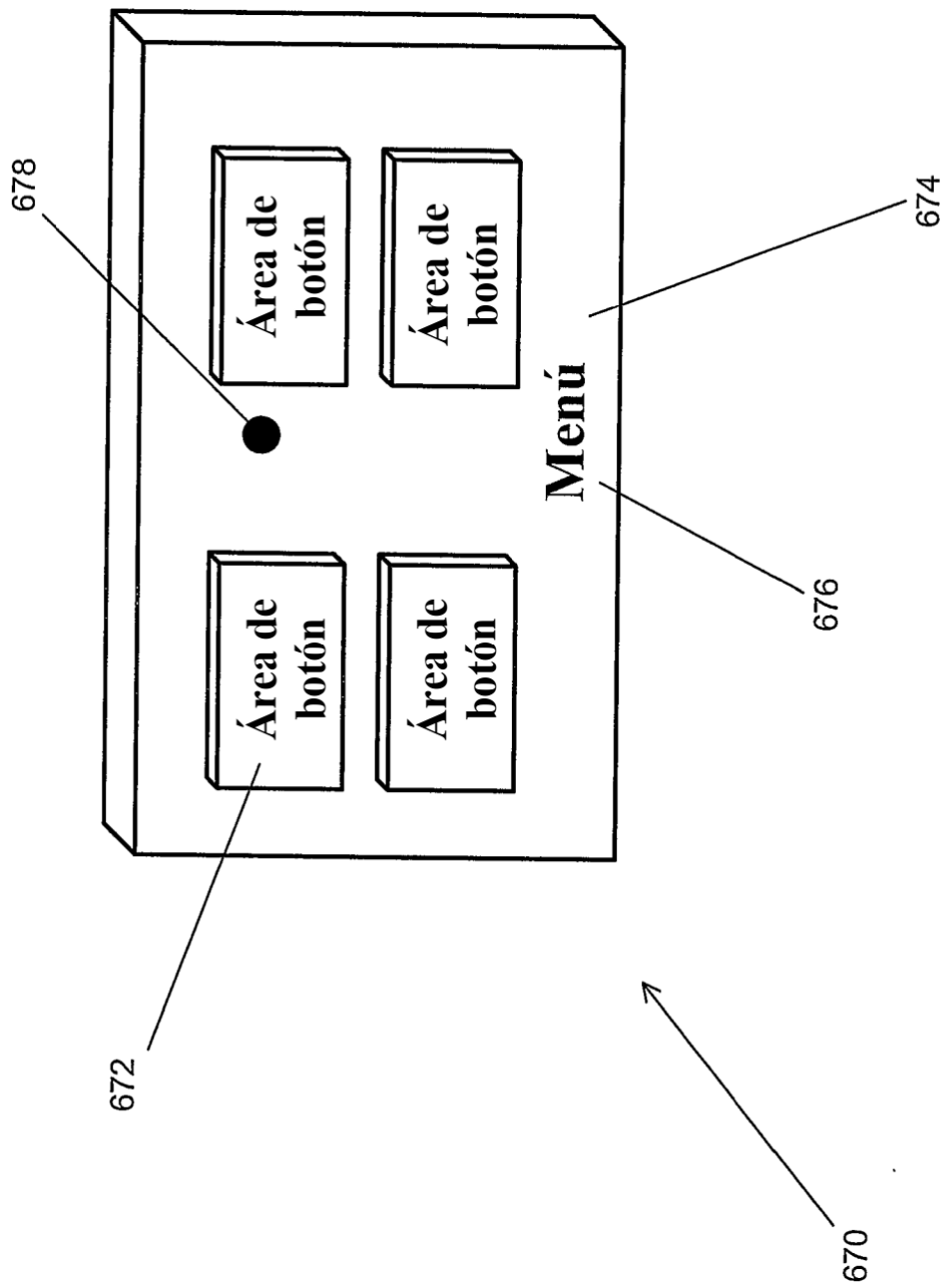


FIG. 4.2.

