

19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 371 188**

51 Int. Cl.:
H04N 7/26 (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

- 96 Número de solicitud europea: **06804241 .5**
96 Fecha de presentación: **27.09.2006**
97 Número de publicación de la solicitud: **1929784**
97 Fecha de publicación de la solicitud: **11.06.2008**

54 Título: **TRANSCODIFICADOR DIRIGIDO POR EL CONTENIDO QUE ORGANIZA LA TRANSCODIFICACIÓN MULTIMEDIA USANDO INFORMACIÓN DEL CONTENIDO.**

30 Prioridad:
27.09.2005 US 721416 P
17.10.2005 US 727643 P
17.10.2005 US 727644 P
17.10.2005 US 727640 P
24.10.2005 US 730145 P
03.04.2006 US 789048 P
04.04.2006 US 789377 P

45 Fecha de publicación de la mención BOPI:
28.12.2011

45 Fecha de la publicación del folleto de la patente:
28.12.2011

73 Titular/es:
QUALCOMM INCORPORATED
5775 MOREHOUSE DRIVE, AF-210R
SAN DIEGO CA 92121, US

72 Inventor/es:
RAVEENDRAN, Vijayalakshmi, R.;
WALKER, Gordon;
TIAN, Tao;
BHAMIDIPATI, Phanikumar;
SHI, Fang;
CHEN, Peisong;
SUBRAMANIA, Sitaraman, Ganapathy y
OGUZ, Seyfullah, Halit

74 Agente: **Carpintero López, Mario**

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).

ES 2 371 188 T3

DESCRIPCIÓN

Transcodificador dirigido por el contenido que organiza la transcodificación multimedia usando información del contenido

Reivindicación de la prioridad a tenor de 35 U.S.C § 119

5 La presente solicitud de Patente reivindica la prioridad respecto a (a) Solicitud de Patente Provisional N° 60/721.416 titulada "A VIDEO TRANSCODER FOR REAL-TIME STREAMING AND MOBILE BROADCAST APPLICATIONS" presentada el 27 de septiembre de 2005, (b) Solicitud de Patente Provisional N° 60/789.377 titulada "A VIDEO TRANSCODER FOR REAL-TIME STREAMING AND MOBILE BROADCAST APPLICATIONS" presentada el 4 de abril de 2006, (c) Solicitud Provisional N° 60/727.643 titulada "METHOD AND APPARATUS FOR SPATIO-TEMPORAL DEINTERLACING AIDED BY MOTION COMPENSATION FOR FIELD-BASED VIDEO" presentada el 17 de octubre de 2005, (d) Solicitud Provisional N° 60/727.644 titulada "METHOD AND APPARATUS FOR SHOT DETECTION IN VIDEO STREAMING" presentada el 17 de octubre de 2005, (e) Solicitud Provisional N° 60/727.640 titulada "A METHOD AND APPARATUS FOR USING AN ADAPTIVE GOP STRUCTURE IN VIDEO STREAMING" presentada el 17 de octubre de 2005, (f) Solicitud Provisional N° 60/730.145 titulada "INVERSE TELECINE ALGORITHM BASED ON STATE MACHINE" presentada el 24 de octubre de 2005, and (g) Solicitud Provisional N° 60/789.048 titulada "SPATIO-TEMPORAL DEINTERLACING AIDED BY MOTION COMPENSATION FOR FIELD-BASED MULTIMEDIA DATA" presentada el 3 de abril de 2006. Todas las siete solicitudes de patente provisional están asignadas al asignatario de la presente.

Referencia de solicitudes en tramitación junto con la presente para la Patente

20 La presente Solicitud de Patente se relaciona con la Solicitud de Patente de Estados Unidos N° 11/373.577 titulada "CONTENT CLASSIFICATION FOR MULTIMEDIA PROCESSING" presentada el 10 de marzo de 2006, transferida al cesionario de la presente.

Antecedentes

Campo

25 La presente solicitud se dirige a aparatos y procedimientos para la transcodificación de video de datos de video para su transmisión continua en tiempo real y, más particularmente, para la transcodificación de datos de video para la transmisión continua en tiempo real en aplicaciones de emisiones para móviles.

Antecedentes

30 La compresión de video eficiente es útil en muchas aplicaciones multimedia tal como la transmisión continua de video inalámbrica y la videotelefonía, debido a los limitados recursos en ancho de banda y la variabilidad del ancho de banda disponible. Ciertas normas de codificación de video, tales como la MPEG-4 (ISO/IEC), H.264 (ITU) o codificaciones de video similares proporcionan una codificación de alta eficiencia muy adecuada para aplicaciones tales como las emisiones inalámbricas. Algunos datos multimedia, por ejemplo, las presentaciones de televisión digital, se codifican en general de acuerdo con otras normas tales como la MPEG-2. En consecuencia, se usan transcodificadores para transcodificar o convertir los datos multimedia codificados de acuerdo con una norma (por ejemplo, MPEG-2) en otra norma (por ejemplo, H.264) previamente a la emisión inalámbrica.

Las mejoras en los códec optimizados en velocidad podrían ofrecer ventajas en la resistencia frente a errores, recuperación de errores y escalabilidad. Más aún, el uso de información determinada a partir de los datos multimedia en sí podría ofrecer también mejoras adicionales para la codificación, incluyendo la resistencia frente a errores, recuperación de errores y escalabilidad. Mihaela van der Schaar et col. "A Hybrid Temporal SNR Fine-Granular Scalability for Internet Video", IEEE Transactions on circuits and systems for video technology, IEEE Service Center, Piscataway, NJ, EE.UU., vol. 11, n° 3, marzo de 2001 desvela un marco de codificación de video escalable y novedoso y un procedimiento de compresión correspondiente para la transmisión continua de video en Internet. En consecuencia, existe una necesidad de un transcodificador que proporcione un procesamiento y compresión de datos multimedia altamente eficiente que use la información determinada a partir del dato multimedia en sí, sea escalable y sea resistente frente a errores para su uso en muchas aplicaciones de datos multimedia que incluyen la emisión para móviles de información multimedia en transmisión continua.

Sumario

50 Cada parte del contenido inventivo basada en aparatos y procedimientos de transcodificación descrita e ilustrada tiene varios aspectos, ninguno de los cuales es responsable por sí solo de sus atributos deseables. Sin limitar el ámbito de esta divulgación, se explicarán ahora brevemente las características más relevantes. Después de considerar esta explicación y particularmente después de la lectura de la sección titulada "Descripción detallada" se entenderá como las características de esta transcodificación dirigida por el contenido proporciona mejoras para los aparatos y procedimientos de procesamiento de datos multimedia.

Los aspectos inventivos descritos en el presente documento se exponen en las reivindicaciones adjuntas en relación con el uso de información de contenido para varios procedimientos de codificación de datos multimedia y en varios módulos o componentes de un codificador, por ejemplo, un codificador usado en un transcodificador. Un transcodificador puede organizar la transcodificación de datos multimedia usando información del contenido. La información del contenido se puede recibir desde otra fuente, por ejemplo, metadatos que se reciben con el video. El transcodificador se puede configurar para generar información de contenido por medio de una variedad de diferentes operaciones de procesamiento. En algunos aspectos, el transcodificador genera una clasificación del contenido de los datos multimedia, que se usa a continuación en uno o más procesos de codificación. En algunos aspectos, un transcodificador dirigido por el contenido, puede determinar la información de contenido espacial y temporal de los datos multimedia y usar la información del contenido para una codificación de calidad uniforme relacionada con el contenido a través de los canales y clasificación del contenido en base a la compresión/asignación de bits.

En algunos aspectos, se obtiene o calcula la información del contenido (por ejemplo, metadatos, métrica del contenido y/o una clasificación del contenido) de los datos multimedia y se proporciona a continuación a los componentes del transcodificador para su uso en el procesamiento de los datos multimedia para su codificación. Por ejemplo, un preprocesador puede usar cierta información del contenido para la detección de cambio de escena, la realización de telecine inverso ("IVTC"), desentrelazado, compensación de movimiento y supresión de ruido (por ejemplo transformada de ondícula 2D) y reducción del ruido espacio temporal, por ejemplo eliminación de artificiosidades, desanillado, supresión de bloques y/o supresión de ruido. En algunos aspectos, un preprocesador puede también usar la información de contenido para una reducción del muestreo, por ejemplo determinando las áreas "seguras" y de "manejo de la acción" apropiadas cuando se reduce el muestreo desde la definición estándar (SD) a la de un Cuarto de Adaptador Gráfico de Video (QVGA).

En algunos aspectos, un codificador incluye un módulo de clasificación del contenido que se configura para calcular la información del contenido. El codificador puede usar la clasificación del contenido para un control de la tasa de bits (por ejemplo asignación de bits) en la determinación de los parámetros de cuantificación (QP) para cada MB, para estimación del movimiento, por ejemplo, realizando una estimación de movimiento (ME) en el color, realizando una predicción del vector de movimiento (MV), escalabilidad al proporcionar una capa base y una capa de mejora y para resistencia frente a errores mediante el uso de la clasificación del contenido para afectar a la jerarquía de precisión y a esquemas de resistencia frente a errores que incluyen, por ejemplo refresco intra adaptativo, procesos de eliminación de límites y proporcionar datos de cuadro I redundantes en una capa de mejora. En algunos aspectos, el transcodificador usa la clasificación del contenido en coordinación con un multiplexador de datos para el mantenimiento de la calidad óptima de los datos multimedia a través de los canales. En algunos aspectos, el codificador puede usar información de clasificación del contenido para forzar a que aparezcan periódicamente cuatros I en los datos codificados para permitir una conmutación de canal rápida. Tales implementaciones pueden hacer uso también de bloques I que se pueden requerir en los datos codificados para resistencia frente a errores, tal como la conmutación de acceso aleatorio y la resistencia frente a errores (basada en, por ejemplo, la clasificación del contenido) se pueden combinar de modo efectivo a través de la jerarquía de predicción para mejorar la eficiencia de la codificación mientras aumenta la robustez frente a los errores.

En un aspecto un procedimiento de procesamiento de datos multimedia incluye la recepción de datos multimedia y la codificación de los datos multimedia en un primer grupo de datos y un segundo grupo de datos en base al contenido de los datos multimedia, estando configurado el primer grupo de datos para su decodificación independientemente del segundo grupo de datos y en el que el primer y el segundo grupos de datos se codifican con diferentes niveles de calidad. En un aspecto del primer aspecto, el primer grupo de datos comprende cuadros I y cuadros P y el segundo grupo de datos comprende cuadros I, cuadros P y cuadros B. En otro aspecto, el primer grupo de datos comprende una capa base y el segundo grupo de datos comprende una capa de mejora. En un tercer aspecto, el procedimiento comprende además la clasificación del contenido de los datos multimedia en el que la codificación se basa en la clasificación del contenido. En un cuarto aspecto, la codificación comprende la determinación de un primer parámetro de cuantificación de los datos multimedia para la codificación del primer grupo de datos y la determinación de un segundo parámetro de cuantificación para la codificación del segundo grupo de datos, en el que la determinación del primer y el segundo parámetro de cuantificación se basa en la clasificación del contenido. En un quinto aspecto, la codificación comprende la asignación de una tasa de bits para al menos una parte de los datos multimedia en base a la clasificación del contenido. En un sexto aspecto, la codificación incluye además el uso de la clasificación del contenido para detectar cambios de escena y en la determinación de si incluir cuadros I en el primer grupo de datos y en el segundo grupo de datos en base a los cambios de escena detectados. En un séptimo aspecto, la codificación incluye la determinación de una tasa de cuadros para la codificación de los datos multimedia en base a la clasificación del contenido. En un octavo aspecto, la codificación incluye la realización de una estimación de movimiento de los datos multimedia en base a la clasificación del contenido. En un noveno aspecto, el procedimiento incluye también la determinación de una primera tasa de cuadros para la codificación del primer grupo de datos y la determinación de una segunda tasa de cuadros para la codificación del segundo grupo de datos, en el que la primera tasa de cuadros es menor que la segunda tasa de cuadros. En un décimo aspecto, la codificación incluye la realización de un procesamiento de resistencia frente a errores en los datos multimedia en base a la clasificación del contenido. En un undécimo aspecto, la codificación incluye la codificación del primer grupo de datos y del segundo grupo de datos de modo que el primer grupo de datos se pueda decodificar para formar unos datos multimedia que se pueden visualizar si el segundo grupo de datos no está disponible y el primer grupo de datos y el

segundo grupo de datos se pueden decodificar en combinación para formar unos datos multimedia que se pueden visualizar si tanto el primer grupo de datos como el segundo grupo de datos están disponibles. En un duodécimo aspecto, el primer parámetro de cuantificación incluye un primer tamaño de etapa para la codificación de datos y el segundo parámetro de cuantificación incluye un segundo tamaño de etapa para la codificación de datos, en el que el primer tamaño de etapa es mayor que el segundo tamaño de etapa. En un decimotercero aspecto, el procedimiento comprende además la clasificación del contenido de los datos multimedia y en el que la codificación se basa en la clasificación del contenido y en el que la codificación comprende la reducción del ruido en los datos multimedia en base a la clasificación del contenido. En un decimocuarto aspecto, la reducción de ruido incluye la realización de una eliminación de artificiosidades. En un decimoquinto aspecto, la reducción de ruido comprende el procesamiento de al menos una parte de los datos multimedia con un filtro de desanillado, en el que la fuerza del filtro de desanillado se basa en el contenido de los datos multimedia. En un decimosexto aspecto, la reducción de ruido incluye el procesamiento del menos una parte de los datos multimedia con un filtro de supresión de bloques, en el que la fuerza del filtro de supresión de bloques se basa en el contenido de los datos multimedia. En un decimoséptimo aspecto, la reducción del ruido incluye el filtrado de frecuencias seleccionadas de los datos multimedia. En un decimoctavo aspecto, la fuerza del filtro de desanillado se basa en la clasificación del contenido de los datos multimedia. En un decimonoveno aspecto, la fuerza del filtro de supresión de bloques se basa en la clasificación del contenido de los datos multimedia. En un vigésimo aspecto, la codificación comprende la reducción del muestreo de los datos multimedia. Finalmente, en un vigésimo primer aspecto, la codificación incluye la asociación de un nivel de calidad con los datos multimedia y el uso del nivel de calidad y la información del contenido de los datos multimedia para determinar una tasa de bits para la codificación de los datos multimedia.

En un segundo aspecto, un aparato para el procesamiento de datos multimedia incluye un codificador configurado para recibir datos multimedia y codificar los datos multimedia en un primer grupo de datos y un segundo grupo de datos en base al contenido de los datos multimedia, configurado el primer grupo de datos para su decodificación de modo independiente del segundo grupo de datos y en el que el primer y el segundo grupos de datos se codifican con diferentes niveles de calidad. En un aspecto del primer aspecto, el codificador incluye un módulo de clasificación del contenido configurado para determinar una clasificación del contenido de los datos multimedia y el módulo de codificación se configura adicionalmente para codificar los datos multimedia en base a la clasificación del contenido. En un segundo aspecto, el codificador se configura además para determinar un primer parámetro de cuantificación de los datos multimedia para la codificación del primer grupo de datos y un segundo parámetro de cuantificación para la codificación del segundo grupo de datos en el que la determinación del primer y segundo parámetros de cuantificación se basa en la clasificación del contenido de los datos multimedia. En un tercer aspecto, el codificador comprende un módulo de estimación del movimiento configurado para realizar una estimación del movimiento de los datos multimedia en base a la clasificación del contenido y generar información de compensación de los datos de movimiento y en el que el módulo de codificación se configura además para usar la información de compensación del movimiento para codificar los datos multimedia. En un cuarto aspecto, el codificador incluye también un módulo de cuantificación para la determinación de un parámetro de cuantificación de los datos multimedia en base a la clasificación del contenido y se configura además el codificador para usar el parámetro de cuantificación para codificar los datos multimedia. En un quinto aspecto, el codificador incluye también un módulo de asignación de bits configurado para proporcionar una tasa de bits para al menos una parte de los datos multimedia en base a la clasificación del contenido. En un sexto aspecto, el codificador incluye también un módulo de detección de cambio de escena configurado para detectar cambios de escena y el módulo de codificación se configura además para incluir cuadros I en los datos multimedia codificados en base a los cambios de escena detectados. En un séptimo aspecto el codificador incluye también un módulo de tasa de cuadros configurado para determinar una tasa de cuadros de los datos multimedia en base a la clasificación del contenido y en el que el módulo de codificación codifica los datos multimedia en base a la tasa de cuadros. En un octavo aspecto, el codificador se configura también para codificar el primer grupo de datos y el segundo grupo de datos en base a la clasificación del contenido. En un noveno aspecto, el codificador se configura también para realizar un procesamiento de errores sobre los datos multimedia en base a la clasificación del contenido.

En un tercer aspecto, un aparato para el procesamiento de datos multimedia incluye medios para la recepción de datos multimedia y medios para la codificación de los datos multimedia en un primer grupo de datos codificados y un segundo grupo de datos codificados en base al contenido de los datos multimedia, configurado el primer grupo de datos para su decodificación de modo independiente del segundo grupo de datos y en el que el primero y el segundo grupo de datos se codifican con diferentes niveles de calidad. En un aspecto del primer aspecto, el medio receptor comprende un codificador. En otro aspecto, los medios para codificación comprenden un codificador. En un tercer aspecto, los medios de codificación comprenden medios para la determinación de una clasificación del contenido de los datos multimedia en el que los medios de codificación codifican los datos multimedia sobre la clasificación del contenido. En un cuarto aspecto, los medios de codificación comprenden un transcodificador que comprende un codificador.

En un cuarto aspecto, un medio que pueda leer una máquina comprende instrucciones que tras su ejecución hacen que la máquina reciba datos multimedia y codifique los datos multimedia en un primer grupo de datos codificados y un segundo grupo de datos codificados en base al contenido de los datos multimedia, configurado el primer grupo de datos para su decodificación de modo independiente del segundo grupo de datos y en el que el primero y el segundo grupo de datos se codifican con diferentes niveles de calidad. En un aspecto del cuarto aspecto, el medio que pueda

5 leer un ordenador comprende además instrucciones para generar una clasificación del contenido indicativa del contenido de los datos multimedia, en el que la codificación de los datos multimedia en un primer grupo de datos codificados y un segundo grupo de datos codificados incluye la codificación de los datos multimedia en base a la clasificación del contenido. En otro aspecto, la codificación comprende la determinación de un primer parámetro de cuantificación de los datos multimedia para la codificación del primer grupo de datos y la determinación de un segundo parámetro de cuantificación para la codificación del segundo grupo de datos en el que la determinación del primer y segundo parámetros de cuantificación se basa en la clasificación del contenido. En un tercer aspecto, las instrucciones para codificar los datos multimedia incluyen instrucciones para asignar una tasa de bits para al menos una parte de los datos multimedia en base al contenido de los datos multimedia.

10 En un quinto aspecto, un procesador comprende una configuración para recibir datos multimedia y codificar los datos multimedia en un primer grupo de datos codificados y un segundo grupo de datos codificados en base al contenido de los datos multimedia, configurado el primer grupo de datos para su decodificación de modo independiente del segundo grupo de datos y en el que el primer y el segundo grupo de datos se codifican con niveles de calidad diferentes. En un aspecto del quinto aspecto, el procesador comprende además una configuración para generar una clasificación del contenido indicativa del contenido de los datos multimedia en el que la codificación comprende la codificación de los datos multimedia en base a la clasificación del contenido. En otro aspecto, el procesador comprende además una configuración para determinar un primer parámetro de cuantificación de los datos multimedia para la codificación del primer grupo de datos y un segundo parámetro de cuantificación para la codificación del segundo grupo de datos, en el que el primer y el segundo parámetros de cuantificación se basan en la clasificación del contenido.

Breve descripción de los dibujos

La FIGURA 1A es un diagrama de bloques en un sistema de emisión de medios que incluye un transcodificador para la transcodificación entre diferentes formatos de video.

25 La FIGURA 1B es un diagrama de bloques de un codificador configurado para codificar datos multimedia y proporcionar un primer grupo de datos codificados y un segundo grupo de datos codificados.

La FIGURA 1C es un diagrama de bloques de un procesador configurado para codificar datos multimedia.

La FIGURA 2 es un diagrama de bloques de un ejemplo de transcodificador del sistema de la FIGURA 1.

La FIGURA 3 es un diagrama de flujo que ilustra el funcionamiento de un analizador sintáctico usado con el transcodificador de la FIGURA 2.

30 La FIGURA 4 es un diagrama de flujo que ilustra el funcionamiento de un decodificador usado con el transcodificador de la FIGURA 2.

La FIGURA 5 es un diagrama de tiempos del sistema que ilustra una secuencia de operaciones realizadas por el transcodificador de la FIGURA 2.

35 La FIGURA 6 es un diagrama de flujo que ilustra una secuencia de operaciones y funciones de un preprocesador que se puede usar en el transcodificador de la FIGURA 2.

La FIGURA 7 es un diagrama de bloques de un codificador en 2 pasos de ejemplo que se puede usar en el transcodificador de la FIGURA 2.

La FIGURA 8 ilustra un ejemplo de un gráfico de clasificación que ilustra un aspecto de cómo asociar valores de textura y movimiento con la clasificación del contenido.

40 La FIGURA 9 es un diagrama de flujo que ilustra una operación de ejemplo para clasificación del contenido, tal como para su uso en el codificador de la FIGURA 7.

La FIGURA 10 es un diagrama de flujo que ilustra la operación de un control de tasa tal como para su uso con el codificador de la FIGURA 7.

45 La FIGURA 11 es un diagrama de flujo que ilustra la operación de un estimador de movimientos de ejemplo, tal como para su uso con el codificador de la FIGURA 7.

La FIGURA 12 es un diagrama de flujo que ilustra la operación de una función de codificador de decisión de modo de ejemplo, tal como para su uso con el codificador de la FIGURA 7.

La FIGURA 13 es un diagrama de flujo que ilustra una operación de ejemplo que efectúa la escalabilidad para su uso en el codificador de la FIGURA 7.

50 La FIGURA 14 es un diagrama de flujo que ilustra una operación de ejemplo que efectúa un flujo de datos de tasa-distorsión como ocurre en el codificador de la FIGURA 7, por ejemplo.

- La FIGURA 15 es un gráfico que ilustra la relación entre la complejidad de la codificación, los bits asignados y calidad visual humana.
- La FIGURA 16 es un gráfico que ilustra una fórmula de detección de escenas no lineal.
- 5 La FIGURA 17 es un diagrama de bloques del sistema que ilustra medios para la recepción de datos multimedia y medios para la codificación de los datos multimedia recibidos.
- La FIGURA 18 es un diagrama que ilustra un proceso de desentrelazado que usa estimación/compensación de movimiento.
- La FIGURA 19 es un diagrama de bloques de un sistema de comunicación multimedia.
- 10 La FIGURA 20 es un diagrama que ilustra la organización de un flujo de bits de video en una capa de mejora y una capa base.
- La FIGURA 21 es un diagrama que ilustra la alineación de porciones a los límites de los cuadros de video.
- La FIGURA 22 es un diagrama de bloques que ilustra la jerarquía de predicción.
- La FIGURA 23 es un diagrama de flujo de proceso que ilustra un procedimiento de codificación de datos multimedia en base a la información del contenido.
- 15 La FIGURA 24 es un diagrama de flujo de proceso que ilustra un procedimiento de codificación de datos multimedia de modo que se alineen los límites de los datos en base al nivel de información del contenido.
- La FIGURA 25 es un gráfico que ilustra un área de acción segura y un área de título seguro de un cuadro de datos.
- La FIGURA 26 es un gráfico que ilustra un área de acción segura de un cuadro de datos.
- 20 La FIGURA 27 es un diagrama de flujo de proceso que ilustra un proceso de codificación de datos multimedia usando refresco intra adaptativo en base a la información de contenido multimedia.
- La FIGURA 28 es un diagrama de flujo de proceso que ilustra un proceso de codificación de datos multimedia usando cuadros I redundantes en base a la información del contenido multimedia.
- La FIGURA 29 ilustra vectores de compensación del movimiento entre un cuadro actual y un cuadro previo MV_P y un carro actual y un cuadro siguiente MV_N .
- 25 La FIGURA 30 es un diagrama de flujo de proceso que ilustra una detección de toma.
- La FIGURA 31 es un diagrama de flujo de proceso que ilustra la codificación de las capas base y de mejora.
- La FIGURA 32 es un esquema que ilustra la codificación de un macro bloque.
- La FIGURA 33 es un esquema que ilustra módulos para la codificación de una capa base y una capa de mejora.
- La FIGURA 34 muestra un ejemplo de un proceso de selector del coeficiente de capa base y de capa de mejora.
- 30 La FIGURA 35 muestra otro ejemplo de un proceso de selector del coeficiente de una capa base y una capa de mejora.
- La FIGURA 36 muestra otro ejemplo de un proceso de selector del coeficiente de una capa base y una capa de mejora.
- 35 La FIGURA 37 es un diagrama de flujo de proceso que ilustra la codificación de datos multimedia en base a la información del contenido.
- La FIGURA 38 es un diagrama que ilustra posibles decisiones del sistema en un proceso de telecine inverso.
- La FIGURA 39 ilustra los límites a ser filtrados en un macro bloque mediante un proceso de supresión de bloques.
- La FIGURA 40 es un diagrama que ilustra un proceso que desentrelazado espacio-temporal.
- La FIGURA 41 ilustra un ejemplo para un nuevo muestreo de polifase 1D.
- 40 La FIGURA 42 es un diagrama de flujo que ilustra un ejemplo de una estructura GOP adaptativa en la transmisión continua de video.
- Se indica que, donde es apropiado, las numeraciones similares se refieren a partes similares a través de las diversas vistas de los dibujos.

Descripción detallada

La siguiente descripción detallada se dirige a ciertos aspectos explicados en la presente divulgación. Sin embargo, la invención se puede realizar en una multitud de formas diferentes. La referencia en la presente especificación a “un aspecto” significa que se incluye un rasgo, estructura o característica particular descrito en conexión con el aspecto en al menos un aspecto. Las apariciones de las frases “en un aspecto”, “de acuerdo con un aspecto” o “en algunos aspectos”, en varios lugares en la especificación no se refieren necesariamente todas al mismo aspecto ni son aspectos separados alternativos mutuamente exclusivos de otros aspectos. Más aún, se describen varias características que se pueden exhibir por algunos aspectos y no por otros. De modo similar, se describen varios requisitos que pueden ser requisitos para algunos aspectos pero no para otros aspectos.

La siguiente descripción incluye detalles para proporcionar una comprensión global de los ejemplos. Sin embargo, se entenderá por un experto en la materia que los ejemplos se pueden poner en práctica incluso si cada detalle de un proceso o dispositivo en un ejemplo o aspecto no se describe o ilustra en el presente documento. Por ejemplo, los componentes eléctricos se pueden mostrar en diagramas de bloques que no ilustran cada conexión eléctrica o cada elemento eléctrico del componente para no oscurecer los ejemplos con detalles innecesarios. En otros casos, tales componentes, otras estructuras y técnicas se pueden mostrar en detalle para explicar adicionalmente los ejemplos.

La presente divulgación se refiere al control de la codificación y aparatos y procedimientos de transcodificación que usan información del contenido de los datos multimedia que se codifican. La “información del contenido” o “contenido” (de los datos multimedia) son términos amplios que significan la información en relación al contenido de datos multimedia y pueden incluir, por ejemplo, metadatos, métrica calculada a partir de los datos multimedia e información relacionada con el contenido asociada con una o más métricas, por ejemplo una clasificación del contenido. La información del contenido se puede proporcionar a un codificador o ser determinada por un codificador, dependiendo de la aplicación particular. La información del contenido se puede usar para muchos aspectos de la codificación de los datos multimedia, incluyendo la detección de cambio de escena, procesamiento temporal, reducción de ruido espacio-temporal, reducción del muestreo, determinación de las tasas de bits para cuantificación, escalabilidad, resistencia frente a errores, mantenimiento de la calidad multimedia óptima a través de los canales de emisión y cambio rápido de canal. Usando uno o más de estos aspectos, un transcodificador puede organizar el procesamiento de datos multimedia y producir datos multimedia codificados en relación al contenido. Las descripciones y figuras en el presente documento que describen los aspectos de la transcodificación pueden ser aplicables también a aspectos de la codificación y a aspectos de la decodificación.

El aparato y procedimientos del transcodificador se refieren a la transcodificación desde un formato a otro y se describe específicamente en el presente documento en relación a la transcodificación de video MPEG-2 a un formato mejorado, H.264 escalable para la transmisión a través de canales inalámbricos a dispositivos móviles, ilustrativo de algunos aspectos. Sin embargo, la descripción de la transcodificación de video MPEG-2 al formato H.264 no se pretende que limite el ámbito de la invención, sino que es meramente de ejemplo de algunos aspectos de la invención. Los aparatos y procedimientos desvelados proporcionan una arquitectura altamente eficiente que soporta una codificación flexible frente a errores con acceso aleatorio y capacidad de capas y que se puede aplicar asimismo a la transcodificación y/o codificación de formatos de video distintos del MPEG-2 y el H.264.

“Datos multimedia” o simplemente “multimedia” como se usa en el presente documento, es una expresión amplia que incluye datos de video (que pueden incluir datos de audio), datos de audio o datos tanto de video como de audio. “Datos de video” o “video” como se usa en el presente documento es una expresión amplia que se refiere a datos basados en cuadros o basados en campos, que incluyen una o más imágenes o secuencias de imágenes relacionadas, texto de contenido, información de imagen y/o datos de audio y se pueden usar para referirse a datos multimedia (por ejemplo, las expresiones se pueden usar de modo intercambiable) a menos que se especifique lo contrario.

Se describen a continuación ejemplos de varios componentes de un transcodificador y ejemplos de procesos que pueden usar información del contenido para la codificación de datos multimedia.

Sistema de emisión multimedia

La FIGURA 1A es un diagrama de bloques que ilustra un flujo de datos de algunos aspectos de un sistema de emisión de datos multimedia 100. En el sistema 100, un proveedor de datos multimedia 106 comunica datos multimedia codificados 104 a un transcodificador 200. Los datos multimedia codificados 104 se reciben por el transcodificador 200, que procesa los datos multimedia 104 en datos multimedia en bruto en el bloque 110. El procesamiento en el bloque 110 decodifica y analiza los datos multimedia codificados 104 y procesa adicionalmente los datos multimedia para prepararlos para su codificación en otro formato. Los datos multimedia decodificados se proporcionan al bloque 112 en el que se codifican los datos multimedia a un formato o norma multimedia predeterminado. Una vez se han codificado los datos multimedia, en el bloque 114 se preparan para su transmisión, por medio, por ejemplo, de un sistema de emisión inalámbrica (por ejemplo una red de emisiones de teléfono celular o por medio de otra red de comunicaciones). En algunos aspectos, los datos multimedia recibidos 104 se han codificado de acuerdo con la norma MPEG-2. Después de que los datos multimedia transcodificados 104 se hayan

decodificado, el transcodificador 200 codifica los datos multimedia a una norma H.264.

La FIGURA 1B es un diagrama de bloques de un transcodificador 130 que se puede configurar para realizar el procesamiento de los bloques 110 y 112 de la FIGURA 1A. El transcodificador 130 se puede configurar para recibir datos multimedia, decodificar y analizar los datos multimedia en transmisiones continuas elementales en paquetes (por ejemplo, subtítulos, audio, metadatos, video en bruto, datos de CC y marcas del tiempo de presentación), codificarlas en un formato deseado y proporcionar datos codificados para un procesamiento adicional o transmisión. El transcodificador 130 se puede configurar para proporcionar datos codificados a dos o más grupos de datos, por ejemplo, un primer grupo de datos codificados y un segundo grupo de datos codificados, que se denomina como codificación en capas. En algunos ejemplos de aspectos, los varios grupos de datos (o capas) en un esquema de codificación en capas se pueden codificar con diferentes niveles de calidad y formatear de modo que los datos codificados en un primer grupo de datos sean de la calidad más baja (por ejemplo proporcionar un nivel de calidad visual inferior cuando se visualizan) que los datos codificados en un segundo grupo de datos.

La FIGURA 1C es un diagrama de bloques de un procesador 140 que se puede configurar para transcodificar datos multimedia y se puede configurar para realizar una parte o todo el procesamiento representado en los bloques 110 y 112 de la FIGURA 1A. El procesador 140 puede incluir los módulos 124a..n, realizar uno o más de los procesos de transcodificación descritos en el presente documento, incluyendo la decodificación, análisis, preprocesamiento y codificación y usar información del contenido para el procesamiento. El procesador 140 incluye también una memoria interna 122 y se puede configurar para comunicarse con una memoria externa 120, tanto directamente como indirectamente a través de otro dispositivo. El procesador 140 incluye también un módulo de comunicaciones 126 configurado para comunicarse con uno o más dispositivos externos al procesador 140, que incluyen la recepción de datos multimedia y el suministro de datos codificados, tal como datos codificados en un primer grupo de datos y datos codificados en un segundo grupo de datos. En algunos ejemplos de aspectos, los varios grupos de datos (o capas) en un esquema de codificación en capas se pueden codificar con diferentes niveles de calidad y formatear de modo que los datos codificados en un primer grupo de datos sean de una calidad más baja (por ejemplo proporcionen un nivel de calidad visual inferior cuando se visualizan) que los datos codificados en un segundo grupo de datos.

El transcodificador 130 o el preprocesador 140 (configurado para la transcodificación) los componentes de los mismos y los procesos contenidos en ellos, se pueden implementar mediante hardware, software, firmware, middleware, microcódigo o cualquier combinación de los mismos. Por ejemplo, un analizador, decodificador, preprocesador o codificador pueden ser de componentes independientes, incorporados como hardware, firmware, middleware en un componente de otro dispositivo o ser implementados en microcódigo o software que se ejecuta en un procesador o una combinación de los mismos. Cuando se implementa en software, firmware, middleware o microcódigo, el código de programa o segmentos de códigos que realizan la compensación de movimiento, clasificación de tomas y procesos de codificación se pueden almacenar en un medio que pueda leer una máquina tal como un medio de almacenamiento. Un segmento de código puede representar un procedimiento, una función, un subprograma, un programa, una rutina, una subrutina, un módulo, un paquete de software, una clase o cualquier combinación de instrucciones, estructuras de datos o sentencias de programas. Un segmento de código se puede conectar a otro segmento de código o a un circuito de hardware mediante el pase y/o recepción de información, datos, argumentos, parámetros o contenido de memoria.

Ejemplo ilustrativo de arquitectura de transcodificador

La FIGURA 2 ilustra un diagrama de bloques de un ejemplo de un transcodificador que se puede usar por el transcodificador 200 ilustrado en el sistema de emisión multimedia 100 de la FIGURA 1. El transcodificador 200 comprende un analizador/decodificador 202, un preprocesador 226, un codificador 228 y una capa de sincronización 240, descrita adicionalmente a continuación. El transcodificador 200 se configura para usar información del contenido de los datos multimedia 104 para uno o más aspectos del proceso de transcodificación como se describe en el presente documento. La información del contenido se puede obtener desde una fuente externa al transcodificador 200, a través de metadatos multimedia o ser calculada por el transcodificador, por ejemplo, por el preprocesador 226 o el codificador 228. Los componentes mostrados en la FIGURA 2 son ilustrativos de los componentes que se pueden incluir en un transcodificador que usa información del contenido para uno o más procesos de transcodificación. En una implementación particular, uno o más de los componentes del transcodificador 200 se pueden excluir o se pueden incluir componentes adicionales. Adicionalmente, se describen partes del transcodificador y procesos de transcodificación de modo que permitan a alguien experto en la técnica poner en práctica la invención aunque cada detalle de un proceso o dispositivo pueda no estar descrito en el presente documento.

La FIGURA 5 ilustra un diagrama de tiempos como una ilustración gráfica de las relaciones temporales de la operación de varios componentes y/o procesos del transcodificador 200. Como se muestra en la FIGURA 5, la transmisión continua de video codificado 104 (datos multimedia codificados), tal como video MPEG-2, se reciben primero en un tiempo cero (0) arbitrario por el analizador 205 (FIGURA 2). A continuación, la transmisión continua de video se analiza 501, demultiplexa 502 y decodifica 503 tal como por el analizador 205 en combinación con el decodificador 214. Como se ilustra, estos procesos pueden ocurrir en paralelo, con un ligero desplazamiento de tiempo, para proporcionar una salida de transmisión continua del procesamiento de datos al preprocesador 226

(FIGURA 2). En un momento T_1 504 una vez que el preprocesador 226 ha recibido suficientes datos del decodificador 214 para comenzar la salida de los resultados del procesamiento, las etapas de procesamiento restantes se convierten en secuenciales por naturaleza, ocurriendo el primer paso de codificación 505, el segundo paso de codificación 506 y la recodificación 507 en secuencia después del preprocesamiento hasta la finalización de la recodificación en el momento T_f 508.

El transcodificador 200 descrito en el presente documento se puede configurar para transcodificar una variedad de datos multimedia y muchos de los procesos se aplican a cualquier tipo de dato multimedia que se transcodifique. Aunque algunos de los ejemplos proporcionados en el presente documento se refieren particularmente a la transcodificación de datos MPEG-2 en datos H.264, esos ejemplos no significan que limiten la divulgación a tales datos. Los aspectos de la codificación descritos a continuación se pueden aplicar a la transcodificación de cualquier norma de datos multimedia adecuada a otra norma adecuada de datos multimedia.

Analizador/decodificador

Con referencia de nuevo a la FIGURA 2, el analizador/decodificador 202 recibe datos multimedia 104. El analizador/decodificador 202 incluye un analizador de transmisión continua de transporte ("analizador") 205 que recibe los datos multimedia 104 y analiza los datos en una transmisión continua elemental (ES) de video 206, una ES de audio 208, marcas de tiempo de presentación (PTS) 210 y otros datos tales como subtítulos 212. Una ES transporta un tipo de datos (video o audio) desde un codificador de video o audio simple. Por ejemplo, una ES de video comprende los datos de video para una secuencia de datos, incluyendo la secuencia la cabecera y todas las partes de la secuencia. Una transmisión continua elemental en paquetes, o PES, consiste en una única ES que se ha convertido en paquetes, cada uno comenzando típicamente con una cabecera de paquete añadida. Una transmisión continua PES contiene solamente un tipo de datos desde un origen, por ejemplo desde un codificador de video o audio. Los paquetes PES tienen longitud variable, que no se corresponde a la longitud fija de paquete de los paquetes de transporte y puede ser mucho más largo que un paquete de transporte. Cuando los paquetes de transporte se forman a partir de una transmisión continua PES, se puede colocar la cabecera PES al comienzo del contenido útil del paquete de transporte, siguiendo inmediatamente a la cabecera del paquete de transporte. El contenido del paquete PES restante llena el contenido útil de paquetes de transporte sucesivos hasta que se ha usado completamente el paquete PES. El paquete de transporte final se puede rellenar hasta una longitud fija, por ejemplo, mediante su relleno con bytes, por ejemplo bytes = 0xFF (todos unos).

El analizador 205 comunica la ES de video 206 a un decodificador 214 que es parte del analizador/decodificador 202 mostrado aquí. En otras configuraciones el analizador 205 y el decodificador 214 son componentes separados. Las PTS 210 se envían a un generador de PTS de transcodificador 215, que puede generar marcas de tiempo de presentación separadas particulares para el transcodificador 200 para su uso en la disposición de los datos a ser enviados desde el transcodificador 200 a un sistema de emisión. El generador de PTS de transcodificador 215 se puede configurar para proporcionar datos a una capa de sincronismo 240 del transcodificador 200 para coordinar la sincronización de los datos emitidos.

La FIGURA 3 ilustra un diagrama de flujo de un ejemplo de un proceso 300 que puede seguir el analizador 205 cuando analiza las varias transmisiones continuas elementales en paquetes descritos anteriormente. El proceso 300 comienza en el bloque 302 cuando se reciben los datos multimedia 104 desde un proveedor de contenido 106 (FIGURA 1). El proceso 300 prosigue en el bloque 304 en el que se realiza la inicialización del analizador 205. La inicialización se puede activar mediante un comando de adquisición generado independientemente 306. Por ejemplo, un proceso que sea independiente del analizador 205 y se base en una planificación de televisión recibida externamente y la información de alineación del canal puede generar el comando de adquisición 306. Adicionalmente, se pueden introducir los descriptores de memoria intermedia 308 de la transmisión continua (TS) de transporte en tiempo real para ayudar tanto en la inicialización como en el procesamiento principal.

Como se ilustra en el bloque 304, la inicialización puede incluir la adquisición de una verificación de sintaxis de comandos, realización de un primer paso PSI/PSIP/SI (información específica de programa / protocolo de información de programa y sistema / información del sistema) de procesamiento, la realización de un procesamiento específicamente relacionado tanto con el comando de adquisición como con la verificación de la consistencia PSI/PSIP/SI, la asignación de una memoria intermedia PES para cada PES y el establecimiento de tiempos (por ejemplo, para la alineación con el instante deseado de comienzo de la adquisición). La memoria intermedia de PES mantiene los datos ES analizados y comunica cada dato ES analizado a un decodificador de audio correspondiente 216, codificador de prueba 220, decodificador 214 o generador de PTS de transcodificador 215.

Tras la inicialización, el proceso 300 prosigue en el bloque 310 para el procesamiento principal de los datos multimedia 104 recibidos. El procesamiento del bloque 310 puede incluir un filtrado del identificador de paquete (PID) objetivo, supervisión y procesamiento PSI/PSIP/SI continuo y un proceso de tiempos (por ejemplo, para conseguir una duración de adquisición deseada) de modo que los datos multimedia entrantes se pasen a las memorias intermedias de PES apropiadas. Como resultado del procesamiento se generan los datos multimedia del bloque 310, un descriptor del programa y una indicación de 'leída' de la memoria intermedia de PES, que tendrán una interfaz con el decodificador 214 (FIGURA 2) como se describe en el presente documento a continuación.

Después del bloque 310, el proceso 300 prosigue en el bloque 314, en el que tiene lugar la finalización de las operaciones de análisis, incluyendo la generación de un interruptor del temporizador y la liberación de las memorias intermedias del PES a continuación de su consumo. Se indica que las memorias intermedias del PES existirán para todas las transmisiones continuas elementales relevantes del programa citado en su descriptor tal como transmisiones de audio, video y subtítulos.

Con referencia de nuevo a la FIGURA 2, el analizador 205 envía la ES de audio 208 a un decodificador de audio 216 que corresponde a la implementación del transcodificador y proporciona el texto codificado 216 a la capa de sincronismo 240 y la decodificación de la información de audio. La información de subtítulos 212 se suministra a un codificador de texto 220. Los datos de subtítulo oculto (CC) 218 desde un decodificador 214 se proporcionan también al codificador del texto 220, que codifica la información de subtítulos 212 y los datos de CC 218 en un formato afectado por el transcodificador 200.

El analizador/decodificador 202 incluye también el decodificador 214, que recibe la ES de video 206. El decodificador 214 puede generar metadatos asociados con los datos de video, decodifica la transmisión continua elemental en paquetes de video codificada en video en bruto 224 (por ejemplo, en un formato de definición estándar) y procesa los datos de subtítulo oculto de video en la transmisión continua de ES de video.

La FIGURA 4 muestra un diagrama de flujo que ilustra un ejemplo de un proceso de decodificación 400 que se puede realizar por el decodificador 214. El proceso 400 comienza con la entrada de datos de transmisión continua elemental de video 206 en el bloque 402. El proceso 400 prosigue en el bloque 404 en el que se inicializa el decodificador. La inicialización puede incluir un número de tareas, incluyendo la detección de una cabecera de secuencia de video (VSH), la realización de un primer paso del procesamiento de la VSH, la secuencia de video (VS) y la Extensión de Visualización VS (incluyendo el formato de video, colores primarios y coeficientes de matriz) y la asignación de memorias intermedias de datos para almacenar respectivamente la imagen decodificada, los datos asociados y los datos de subtítulo oculto (CC). Adicionalmente, se introduce la información de 'leída' de la memoria intermedia PES de video 406 proporcionada por el analizador 205 (por ejemplo, que se puede generar por el proceso 300 en el bloque 310 de la FIGURA 3).

Tras la inicialización en el bloque 404, el proceso 400 prosigue en el bloque 408 en el que se realiza el procesamiento principal de la ES de video por parte del decodificador 214. El procesamiento principal incluye la extracción la información de 'leída' de la memoria intermedia del PES de video o "interfaz" sobre la disponibilidad de nuevos datos, la decodificación de la ES de video, la reconstrucción y almacenamiento de los datos de píxeles en la sincronización de los límites de la imagen, la generación de metadatos de video y a/v y su almacenamiento en los límites de la imagen y el almacenamiento de los datos de CC en los límites de imagen. Los resultados del bloque 410 del procesamiento principal 408 incluye la generación de descriptores de secuencia, descriptores de memoria intermedia de imagen decodificada, descriptores de memoria intermedia de metadatos y descriptores de memoria intermedia de datos de CC.

Después del procesamiento principal 408, el proceso 400 prosigue en el bloque 412 en el que realiza un proceso de finalización. El proceso de finalización puede incluir la determinación de las condiciones de finalización, incluyendo que no aparezcan nuevos datos durante una duración particular por encima de un umbral predeterminado, la detección de una secuencia de código final y/o la detección de una señal de finalización explícita. El proceso de finalización puede incluir además la liberación de las memorias intermedias de la imagen decodificada, los metadatos asociados y los datos de CC posteriormente a su consumo por un procesador que se describirá a continuación. El proceso 400 finaliza en el bloque 414 en el que entra en un estado de espera a que se reciba como entrada una ES de video.

Preprocesador

La FIGURA 2, y con más detalle la FIGURA 6, ilustran un aspecto de muestra de un preprocesador 226 que puede usar información del contenido para una o más operaciones de procesamiento. El preprocesador 226 recibe metadatos 222 y datos de video 224 en "bruto" decodificados desde el analizador/decodificador 202. El preprocesador 226 se configura para realizar ciertos tipos de procesamiento sobre los datos de video 224 y los metadatos 222 y proporcionar multimedia procesada (por ejemplo, cuadros de referencia de la capa base, cuadros de referencia de la capa de mejora, información de ancho de banda, información del contenido) y video al decodificador 228. Tal procesamiento de los datos multimedia puede mejorar la calidad visual, desescalonado y eficiencia de compresión de los datos. En general, el preprocesador 226 recibe secuencias de video proporcionadas por el decodificador 214 en el analizador/decodificador 202 y convierte las secuencias de video en secuencias de video progresivas para su procesamiento adicional (por ejemplo codificación) por el codificador 228. En algunos aspectos, el preprocesador 226 se puede configurar para numerosas operaciones, incluyendo telecine inverso, desentrelazado, filtrado (por ejemplo, eliminación de artificiosidades, desanillado, supresión de bloques y supresión de ruido), redimensionamiento (por ejemplo, reducción de la resolución espacial a partir de la definición estándar hasta un Cuarto de Matriz Gráfica de Video (QVGA)) y generación de una estructura GOP (por ejemplo calculando la generación de un mapa de complejidad, detección de cambio de escena y detección de desvanecimiento/flash).

El preprocesador 226 puede usar los metadatos del decodificador para afectar a una o más de las operaciones de

preprocesamiento. Los metadatos pueden incluir información que se relaciona con, describe o clasifica el contenido de los datos multimedia (“información de contenido”); en particular los metadatos pueden incluir una clasificación del contenido. En algunos aspectos, los metadatos no incluyen la información del contenido deseada para operaciones de codificación. En tales casos el preprocesador 226 se puede configurar para determinar la información del contenido y usar la información del contenido para operaciones de preprocesamiento y/o proporcionar la información del contenido a otros componentes del transcodificador 200, por ejemplo el decodificador 228. En algunos aspectos, el preprocesador 226 puede usar dicha información del contenido para influenciar la división del GOP, determinar el tipo de filtrado apropiado y/o determinar los parámetros de codificación que se comunican a un codificador.

La FIGURA 6 muestra un ejemplo ilustrativo de varios en bloques de proceso que se pueden incluir en el preprocesador 226 e ilustran el procesamiento que se puede realizar por el preprocesador 226. En este ejemplo, el preprocesador 226 recibe metadatos y video 222, 224 y proporciona datos de salida 614 que comprenden metadatos y video (procesados) al codificador 228. Típicamente, hay tres tipos de video que se pueden recibir. Primero, el video recibido puede ser video progresivo, en el que no se requiere desentrelazado. Segundo, los datos de video pueden ser video de telecine, video entrelazado convertido a partir de secuencias de película de 24 imágenes por segundo, en cuyo caso se puede requerir una operación de telecine inverso. Tercero, el video puede ser video no de telecine entrelazado. El preprocesador 226 puede procesar estos tipos de video como se describe a continuación.

En el bloque 601, el preprocesador 226 determina si los datos de video recibidos 222, 224 son video progresivo. En algunos casos, esto se puede determinar a partir de los metadatos si los metadatos contienen tal información o mediante el procesamiento de los datos de video en sí. Por ejemplo, un proceso de telecine inverso, descrito a continuación, puede determinar si el video recibido 222 es video progresivo. Si lo es, el proceso prosigue en el bloque 607 en el que se realizan las operaciones de filtrado (por ejemplo supresión de ruido) sobre el video para reducir el ruido, tal como el ruido blanco gaussiano. Si los datos de video 222, 224 no son video progresivo, en el bloque 600 el proceso prosigue al bloque 604 a un detector de fase 604.

El detector de fase 604 distingue entre video que se originó en un telecine y el que comienza en un formato de emisión estándar. Si se toma la decisión de que el video era de telecine (la ruta de decisión SÍ que sale del detector de fase 604), el video de telecine se devuelve a su formato original en el telecine inverso 606. Las tramas redundantes se identifican y se eliminan y los campos derivados del mismo cuadro de video se recomponen en una imagen completa. Dado que la secuencia de las imágenes de película reconstruidas fueron registradas fotográficamente a intervalos regulares 1/24 de segundo, el proceso de estimación del movimiento realizado en un divisor GOP 612 o el decodificador 228 es más preciso usando las imágenes de telecine invertidas en lugar de los datos de telecine, que tienen una base de tiempos irregular.

En un aspecto, el detector de fase 604 toma ciertas decisiones después de la recepción de un cuadro de video. Estas decisiones incluyen: (i) si el video presente es de una salida de telecine y la fase de extracción 3:2 es una de las 5 fases P₀, P₁, P₂, P₃ y P₄ mostradas a la FIGURA 38 y (ii) el video se generó como NTSC convencional. Esa decisión se indica como la fase P₅. Esas decisiones aparecen como salidas del detector de fase 604 mostrado en la FIGURA 2. La ruta desde el detector de fase 604 etiquetada “SÍ” actúa sobre el telecine inverso 606, indicando que se le ha suministrado la fase de adaptación correcta de modo que puede clasificar los campos que se formaron a partir de la misma imagen fotográfica y combinarlos. La ruta desde el detector de fase 604 etiquetada “NO” actúa de modo similar sobre el desentrelazador 605 para separar un cuadro NTSC aparente en campos para su procesamiento óptimo. El detector de fase 604 puede analizar continuamente los cuadros de video debido a que se pueden recibir en cualquier momento diferentes tipos de video. Como ejemplo, se puede insertar video de acuerdo con la norma NTSC en el video como un anuncio. Después del telecine inverso, el video progresivo resultante se envía a un supresor de ruido (filtro) 607 que se puede usar para reducir el ruido blanco gaussiano.

Cuando se reconoce video NTSC convencional (el recorrido NO desde el detector de fase 601), se transmite al desentrelazador 605 para compresión. El desentrelazador 605 transforma los campos entrelazados en video progresivo y se pueden realizar entonces las operaciones de supresión de ruido sobre el video progresivo. Se describe a continuación un ejemplo ilustrativo del proceso de desentrelazado.

Los dispositivos de video analógicos tradicionales como los televisores reproducen el video de una manera entrelazada, es decir tales dispositivos transmiten las líneas de barrido numeradas pares (campos pares) y líneas de barrido numeradas impares (campos impares). Desde el punto de vista del muestreo de la señal, esto es equivalente a un submuestreo espacio-temporal en un patrón descrito por:

$$F(x, y, n) = \begin{cases} \Theta(x, y, n), & \text{si } y \bmod 2 = 1 \text{ para campos pares,} \\ \Theta(x, y, n), & \text{si } y \bmod 2 = 0 \text{ para campos impares,} \\ \text{En caso contrario, borrado,} & \end{cases} \quad [1]$$

en la que Θ se refiere a la imagen de cuadro original, F indica el campo de entrelazado y (x, y, n) representan las posiciones horizontal, vertical y temporal de un píxel respectivamente.

Sin pérdida de generalidad, se puede suponer que $n = 0$ es un campo par a lo largo de esta divulgación de modo que la Ecuación 1 anterior se simplifica como

$$F(x, y, n) = \begin{cases} \Theta(x, y, n), & \text{si } y \bmod 2 = n \bmod 2, \\ \text{En caso contrario, borrado,} & \end{cases} \quad [2]$$

5 Dado que la reducción no se realiza en la dimensión horizontal, el patrón de submuestreado se puede representar en la siguiente coordenada $n \sim y$.

El objetivo de un desentrelazador es transformar el video entrelazado (una secuencia de campos) en cuadros progresivos no entrelazados (una secuencia de cuadros). En otras palabras, interpolar los campos pares e impares para "recuperar" o generar los cuadros de imagen completos. Esto se puede representar mediante la Ecuación 3:

$$F_o(x, y, n) = \begin{cases} F(x, y, n), & y \bmod 2 = n \bmod 2, \\ F_i(x, y, n), & \text{en otro caso,} \end{cases} \quad [3]$$

10 en la que F_i representa los resultados del desentrelazado para píxeles perdidos.

La FIGURA 40 es un diagrama de bloques que ilustra ciertos aspectos de un aspecto de un desentrelazador 605 que usa el filtrado Wmed y estimación de movimiento para generar un cuadro progresivo a partir de datos multimedia entrelazados. La parte superior de la FIGURA 40 muestra un mapa de intensidad de movimiento 4002 que se puede generar usando información a partir el campo actual, dos campos previos (Campo PP y Campo P) y dos campos posteriores (Campo Siguiente y Campo Siguiente al Siguiente). El mapa de intensidad de movimiento 4002 clasifica o divide, el cuadro actual en dos o más niveles de movimiento diferentes y se puede generar mediante un filtrado espacio temporal, descrito con más detalle en el presente documento a continuación. En algunos aspectos, el mapa de intensidad de movimiento 4002 se genera para identificar áreas estáticas, áreas de movimiento lento y áreas de movimiento rápido como se describe con referencia a las Ecuaciones 4-8 a continuación. Un filtro espacio temporal, por ejemplo el filtro Wmed 4004, filtra los datos multimedia entrelazados usando criterios en base al mapa de intensidad del movimiento y produce un cuadro de desentrelazado provisional espacio temporal. En algunos aspectos, el proceso de filtrado Wmed involucra una proximidad horizontal de $[-1, 1]$, una proximidad vertical de $[-3, 3]$ de cinco campos adyacentes, que se representan por los cinco campos (Campo PP, Campo P, Campo Actual, Campo Siguiente, Campo Siguiente al Siguiente) ilustrados en la FIGURA 40, en la que Z^{-1} representa un retardo de un campo. En relación al Campo Actual, el Campo Siguiente y el Campo P son campos sin paridad y el Campo PP y el Campo Siguiente al Siguiente son campos de paridad. La proximidad usada para el filtrado espacio-temporal se refiere a la localización espacial y temporal de campos y píxeles realmente usados durante la operación de filtrado y se puede ilustrar como una apertura, como se muestra, por ejemplo, en las Figuras 6 y 7.

30 El desentrelazador 605 puede incluir también un supresor de ruido (filtro supresor de ruido) 4006 configurado para filtrar el cuadro de desentrelazado provisional espacio temporal generado mediante el filtro Wmed 4004. La supresión de ruido del cuadro de desentrelazado provisional espacio temporal hace el proceso de búsqueda de movimiento posterior más preciso especialmente si la secuencia de datos multimedia entrelazados de origen está contaminada con ruido blanco. Esto también puede al menos en parte retirar el escalonado entre filas pares e impares en una imagen Wmed. El supresor de ruido 4006 se puede implementar como una variedad de filtros que incluyen un reductor de ruido en base a una reducción de las ondículas y un filtro de ondículas de Wiener. Un reductor de ruido se puede usar para eliminar el ruido del cuadro Wmed candidato antes de que se procese adicionalmente usando información de compensación de movimiento y puede eliminar el ruido que está presente en el cuadro Wmed y retener la señal presente independientemente del contenido de frecuencia de la señal. Se pueden usar varios tipos de filtros de supresión de ruido, incluyendo los filtros de ondículas. Las ondículas son una clase de funciones usadas para localizar una señal dada tanto en los dominios del espacio como del escalado. La idea fundamental detrás de las ondículas es analizar la señal en diferentes escalas o resoluciones de modo que pequeños cambios en la representación de la ondícula produzcan un cambio correspondientemente pequeño en la señal original.

45 Una reducción de las ondículas o un filtro de ondículas de Wiener se puede aplicar también como el supresor de ruido. La reducción de las ondículas consiste en una transformación de la ondícula de la señal con ruido, seguida por una reducción de los coeficientes de ondícula pequeños a cero (o el valor más pequeño), en tanto que se dejan los coeficientes grandes sin afectar. Finalmente, se realiza una transformación inversa para obtener la señal estimada.

50 El filtrado supresor de ruido refuerza la precisión de la compensación de movimiento en entornos ruidosos. La supresión de ruido por reducción de la ondícula puede involucrar la reducción en el dominio de transformadas de la ondícula, y típicamente comprende tres etapas: una transformada de ondícula directa lineal, una supresión de ruido por reducción no lineal y una transformada de ondícula inversa lineal. El filtro de Wiener es un filtro óptimo MSE que se puede usar para mejorar las imágenes degradadas por la adición de ruido y difuminado. Tales filtros se conocen en general en la técnica y se describen, por ejemplo en "Ideal spatial adaptation by wavelet shrinkage", referenciado

anteriormente y por el documento de S. P. Ghael, A. M. Sayeed y R. G. Baraniuk, "Improvement Wavelet denoising via empirical Wiener filtering" Proceedings of SPIE, vol. 3169, págs. 389-399, San Diego, julio de 1997.

En algunos aspectos, un filtro de supresión de ruido se basa en un aspecto de un filtro de ondícula B-spline cúbico biortogonal (4, 2). Un filtro así se puede definir mediante las siguientes transformadas directa e inversa

$$5 \quad h(z) = \frac{3}{4} + \frac{1}{2}(z + z^{-1}) + \frac{1}{8}(z + z^{-2}) \quad (\text{transformada directa}) \quad [4]$$

y

$$g(z) = \frac{5}{4}z^{-1} - \frac{5}{32}(1 + z^{-2}) - \frac{3}{8}(z + z^{-3}) - \frac{3}{32}(z^2 + z^{-4}) \quad (\text{transformada inversa}) \quad [5]$$

10 La aplicación de un filtro de supresión de ruido puede aumentar la precisión de la compensación de movimiento en un entorno ruidoso. Las implementaciones de tales filtros se describen adicionalmente en "Ideal spatial adaptation by wavelet shrinkage", D.L. Donoho y I.M. Johnstone, Biometrika, vol. 8, págs. 425-455, 1994.

15 La parte inferior de la FIGURA 40 ilustra un aspecto para la determinación de la información de movimiento (por ejemplo candidatos para el vector de movimiento, estimación de movimiento, compensación de movimiento) de datos multimedia entrelazados. En particular, la FIGURA 40 ilustra una estimación de movimiento y un esquema de compensación de movimiento que se usa para generar un cuadro progresivo provisional compensado en movimiento del cuadro seleccionado y combinado a continuación con el cuadro provisional Wmed para formar un cuadro progresivo "final" resultante, mostrado como el marco al actual desentrelazado 4014. En algunos aspectos, los candidatos (o estimaciones) del vector de movimiento ("MV") de los datos multimedia entrelazados se proporcionan al desentrelazador desde los estimadores externos del movimiento y se usan para proporcionar un punto de arranque para el estimador y compensador de movimiento bidireccional ("ME/MC") 4018. En algunos aspectos, un selector de candidatos a MV 4022 usa los MV determinados previamente para los bloques contiguos para candidatos a MV de los bloques que se están procesando, tal como los MV de los bloques procesados previamente, por ejemplo los bloques en un cuadro previo desentrelazado 4020. La compensación del movimiento se puede realizar bidireccional, en base al cuadro desentrelazado previo 70 y a un cuadro Wmed siguiente (por ejemplo futuro) 4008. Se mezclan, o combinan un cuadro Wmed actual 4010 y un cuadro actual compensado en movimiento ("MC") 25 4016, mediante un combinador 4012. Se proporciona un cuadro actual desentrelazado resultante 4014, ahora un cuadro progresivo, al ME/MC 4018 para su uso como un cuadro previo desentrelazado 4020 y también se comunica externamente al desentrelazador 605 para un procesamiento posterior.

30 Es posible desacoplar los esquemas de predicción de desentrelazado que comprenden una interpolación entre campos a partir de la interpolación entre campos con un esquema de desentrelazado Wmed + MC. En otras palabras, el filtrado Wmed espacio temporal se puede usar principalmente para finalidades de interpolación entre campos, en tanto que la interpolación entre campos se puede realizar durante la compensación del movimiento. Esto reduce la relación de señal a ruido de pico del resultante Wmed, aunque la calidad visual después de que se aplique la compensación del movimiento es más satisfactoria, debido que los píxeles malos de las decisiones del modo de predicción imprecisa entre campos se eliminarán del proceso de filtrado Wmed.

35 Tras el telecine inverso apropiado o proceso de desentrelazado, en el bloque 608, se procesa el video progresivo para una supresión del escalonado y remuestreo (por ejemplo redimensionamiento). En algunos aspectos del remuestreo, se implementa un remuestreador de polifase para redimensionamiento del tamaño de imagen. En un ejemplo de la reducción de muestreo, la relación entre la imagen original y la redimensionada puede ser p/q , en el que p y q son enteros primos entre sí. El número total de fases que es p . La frecuencia de corte del filtro de polifase en algunos aspectos es 0,6 para factores de redimensionamiento alrededor de 0,5. La frecuencia de corte no se ajusta exactamente a la relación de redimensionamiento para reforzar la respuesta en alta frecuencia de la secuencia redimensionada. Esto permite inevitablemente algún escalonado. Sin embargo, es bien conocido que el ojo humano prefiere imágenes definidas aunque un poco escalonadas a imágenes libres de escalonamiento y borrosas.

45 La FIGURA 41 ilustra un ejemplo de un remuestreado polifásico, que muestra que Las fases en la relación de redimensionamiento es $\frac{3}{4}$. La frecuencia de corte ilustrada en la FIGURA 41 es también $\frac{3}{4}$. Los píxeles originales se ilustran en la figura anterior con ejes verticales. Se dibuja también una función de sincronismo centrada alrededor de los ejes para representar la forma de onda de filtrado. Debido a que la frecuencia de corte se eligió para que fuese exactamente la misma que la relación de remuestreo, los ceros de la función de sincronismo sólo solapan las posiciones de los píxeles después del redimensionado, ilustrado la FIGURA 41 con cruces. Para hallar un valor de píxeles después del redimensionado, la contribución se puede acumular a partir de los píxeles originales como se muestra en la siguiente ecuación:

$$50 \quad v(x) = \sum_{i=-\infty}^{\infty} u(i) \times \text{senc}(\pi f_c(i-x)) \quad [6]$$

en la que f_c es la frecuencia de corte. El filtro polifásico 1-D anterior se puede aplicar tanto a la dimensión horizontal, la dimensión vertical.

Otro aspecto del remuestreo (redimensionamiento) es tener en cuenta el sobrebarrido. En una señal de televisión NTSC, una imagen tiene 486 líneas de barrido y en video digital podría tener 720 píxeles en cada línea de barrido. Sin embargo, no toda la imagen completa es visible en la televisión debido a desajustes entre el tamaño y el formato de pantalla. La parte de la imagen que no es visible se denomina sobrebarrido.

Para ayudar a las emisoras a poner información útil en el área visible por tantos televisores como sea posible, la Sociedad de Ingeniería de Imágenes en Movimiento y Televisión (SMPTE) definió tamaños específicos del cuadro de acción denominadas área segura de acción y área segura de títulos. Véanse las prácticas recomendadas del SMPTE RP 27.3-1989 sobre *Specifications for Safe Action and Safe Title Areas Test Pattern for Television Systems*. El área segura de acción se define por el SMPTE como el área en la que “debe tener lugar toda acción significativa”. El área segura de títulos se define como el área en la que “se puede confinar toda la información útil para asegurar la visibilidad en la mayoría de los receptores de televisión domésticos”.

Por ejemplo, con referencia a la FIGURA 25, el área segura de datos 2510 ocupa el centro del 90% de la pantalla, dando un 5% de borde a todo alrededor. El área segura de títulos 2505 ocupa el centro del 80% de la pantalla, dando un 10% de borde. Con referencia ahora a la FIGURA 26, debido a que la área segura de títulos es tan pequeña, para añadir más contenido en la imagen, algunas emisoras pondrán texto en el área segura de acción, lo que está dentro del cuadro rectangular blanco 2615.

Normalmente se pueden ver bordes negros en el sobrebarrido. Por ejemplo, en la FIGURA 26 aparecen bordes negros en el lado superior 2620 en el lado inferior 2625 de la imagen. Estos bordes negros se pueden eliminar en el sobrebarrido, debido a que el video H.264 usa extensión de los límites en la estimación del movimiento. Los bordes negros extendidos pueden aumentar la parte residual. Conservadoramente, los límites pueden cortarse en un 2% y hacer a continuación el redimensionamiento. Se pueden generar en consecuencia los filtros para redimensionamiento. El truncado se realiza para eliminar el sobrebarrido antes de una reducción de muestreo polifásica.

Con referencia de nuevo a la FIGURA 6, el video progresivo prosigue entonces al bloque 610 en el que se realizan las operaciones de supresión de bloques y desanillado. Tienen lugar comúnmente dos tipos de artificiosidades, “bloques” y “anillado”, en las aplicaciones de compresión de video. Las artificiosidades de bloques ocurren debido a que los algoritmos de compresión dividen cada trama en bloques (por ejemplo bloques de 8x8). Cada bloque se reconstruye con algunos pequeños errores y los errores en los bordes de un bloque contrastan a menudo con los errores en los bordes de los bloques contiguos, haciendo visibles los límites entre bloques. Por el contrario, las artificiosidades de anillos aparecen como distorsiones alrededor de los bordes de las características de la imagen. Las artificiosidades de anillos suceden debido a que el codificador descarta demasiada información en la cuantificación de los coeficientes DCT de alta frecuencia. En algunos ejemplos ilustrativos, tanto el supresión de bloques como el desanillado puede usar filtros FIR (respuesta de impulsos finita) paso bajo para ocultar estas artificiosidades visibles.

En un ejemplo el procesamiento del supresión de bloques, un filtro de supresión de bloques se puede aplicar a todos los bordes de los bloques 4x4 de un cuadro, excepto los bordes en los límites del cuadro y cualquier borde para el que se deshabilite el proceso de filtro de supresión de bloques. El proceso de filtrado se realizará en base a macro bloques después de la finalización del proceso de construcción del cuadro con todos los macro bloques procesados en un cuadro para incrementar las direcciones de macro bloques. Para cada macro bloque, se filtran primero los bordes verticales, de izquierda a derecha y a continuación se filtran los bordes horizontales de arriba a abajo. El filtro de supresión de bloques de luma se realiza en cuatro bordes de 16 muestras y el proceso del filtro de supresión de bloques para cada componente de croma se realiza en dos bordes de 8 muestras, para la dirección horizontal y para la dirección vertical, como se muestra en la FIGURA 39. Los valores de muestra por encima y a la izquierda del macro bloque actual que ya pueden haber sido modificados por la operación del proceso de supresión de bloques sobre los macro bloques previos se deberán usar como entrada para el proceso de filtro de supresión de bloques en el macro bloque actual y se puede modificar adicionalmente durante el filtrado del macro bloque actual. Los valores de muestra modificados durante el filtrado de los bordes verticales se pueden usar como entrada para el filtrado de los bordes horizontales para el mismo macro bloque. Un proceso de supresión de bloques se puede invocar para los componentes de luma y croma por separado.

En un ejemplo del proceso de desanillado, se puede aplicar de modo adaptativo un filtro 2-D para suavizar las áreas próximas a los bordes. Los píxeles de los bordes se someten a poco o ningún filtrado para evitar el desenfoque.

Divisor GOP

Tras la supresión de bloques y desanillado, se procesa el video progresivo por un divisor GOP 612. El posicionamiento del GOP puede incluir la detección de cambios de toma, generación de mapas de complejidad (por ejemplo, mapas de ancho de banda temporal, espacial) y división GOP adaptativa. Se describe cada uno de estos a continuación.

A. Detección de cambio de escena

La detección de tomas se refiere a la determinación de cuando un cuadro en un grupo de imágenes (GOP) exhibe datos que indican que ha tenido lugar un cambio de escena. En general, dentro de un GOP. Los cuadros pueden no tener cambios significativos en cualquiera de dos o tres (o más) cuadros adyacentes o puede haber cambios lentos o cambios rápidos. Naturalmente, estas clasificaciones de cambio de escena se pueden dividir adicionalmente en un nivel mayor de cambios dependiendo de una aplicación o necesidad específica.

La detección de cambios de toma o escena es importante para una codificación de video eficiente. Típicamente, cuando no está cambiando un GOP significativamente, un cuadro I al comienzo de un GOP tiene seguido por un número de cuadros predictivos que pueden codificar suficientemente el video de modo que la decodificación posterior y visualización del video sea visualmente aceptable. Sin embargo, cuando está cambiando una escena, tanto abruptamente como lentamente, se pueden necesitar cuadros I adicionales y una codificación menos predictiva (cuadros P y cuadros B) para producir posteriormente resultados decodificados visualmente aceptables.

Se describen a continuación los sistemas y procedimientos de detección y codificación de tomas que mejoran el rendimiento de los sistemas de codificación existentes. Tales aspectos se pueden implementar en el divisor de GOP 612 del preprocesador 226 (FIGURA 7) o incluirse en un dispositivo codificador que puede funcionar con o sin un preprocesador. Tales aspectos utilizan estadísticas (o métricas) que incluyen comparaciones estadísticas entre cuadros adyacentes de datos de video para determinar si ha tenido lugar un cambio de escena abrupto, una escena está cambiando lentamente o si hay luces de flash de cámaras en la escena que pueden hacer la codificación del video especialmente compleja. Las estadísticas se pueden obtener a partir de un preprocesador y enviarse entonces a un dispositivo de codificación o se pueden generar en un dispositivo de codificación (por ejemplo, mediante un procesador configurado para realizar una compensación de movimiento). Las estadísticas resultantes ayudan en la decisión de la detección de cambio de escena. En un sistema que no transcodifique, a menudo existe un preprocesador adecuado o un procesador configurable. Si el preprocesador realiza un desentrelazado ayudado por compensación de movimiento, están disponibles las estadísticas de compensación del movimiento y listas para su uso. En tales sistemas, un algoritmo de detección de tomas puede aumentar ligeramente la complejidad del sistema.

El ejemplo ilustrativo de un detector de tomas descrito en el presente documento sólo necesita utilizar estadísticas de un cuadro previo, un cuadro actual y un cuadro siguiente y en consecuencia tiene una latencia muy baja. El detector de tomas diferencia varios tipos diferentes de eventos de toma, incluyendo el cambio de escena abrupto, el desvanecimiento cruzado y otros cambios de escena lentos y luces de flash de cámaras. Mediante la determinación de los tipos diferentes de eventos de tomas con diferentes estrategias en el codificador, se mejoran la eficiencia de codificación y la calidad visual.

Se puede usar la detección de cambio de escena para cualquier sistema de codificación de video para que conserve bits inteligentemente mediante la inserción de un cuadro I a intervalos fijos. En algunos aspectos, la información de contenido obtenida por el preprocesador (por ejemplo, o bien incorporada en metadatos o calculada por el preprocesador 226) se puede usar para la detección de cambio de escena. Por ejemplo, dependiendo de la información del contenido, se pueden ajustar dinámicamente los valores de umbral u otros criterios descritos a continuación para tipos diferentes de contenido de video.

La codificación de video funciona normalmente en un grupo de imágenes (GOP) estructurado. Un GOP comienza normalmente con un cuadro intra codificado (cuadro I), seguido por una serie de cuadros P (predictivo) o B (bidireccional). Típicamente un cuadro I puede almacenar todos los datos requeridos para visualizar el cuadro, un cuadro B confía en los datos en los cuadros precedentes y siguientes (por ejemplo conteniendo solamente los datos que cambian desde el cuadro precedente o es diferente de los datos en el cuadro siguiente) y un cuadro P contiene datos que han cambiado desde el cuadro precedente. En el uso común, los cuadros I se intercalan con cuadros P y cuadros B en el video codificado. En términos de tamaño (por ejemplo número de bits usados para codificar el cuadro), los cuadros I son típicamente mucho mayores que los cuadros P que a su vez son mayores que los cuadros B.

Para un eficiente proceso de codificación, transmisión y decodificación, la longitud de un GOP debería ser suficientemente larga para reducir la pérdida de eficiencia de los grandes cuadros I y suficientemente corta como para combatir del desajuste entre codificador y decodificador o impedimentos del canal. Además, se pueden codificar internamente macro bloques (MB) en cuadros P por la misma razón.

La detección de cambio de escena se puede usar para que un codificador de video determine una longitud del GOP apropiada e inserte cuadros I en base a la longitud del GOP, en lugar de insertar un cuadro I frecuentemente no necesario a intervalos fijos. En un sistema de video de transmisión continua práctico, el canal de comunicación queda afectado por errores de bits o pérdida de paquetes. Dónde colocar los cuadros I o MB I puede impactar significativamente en la calidad del video decodificada y la experiencia de visionado. Un esquema de codificación es usar cuadros de intra codificación para imágenes o partes de imágenes que tengan un cambio significativo frente a las imágenes o partes de imágenes previas localizadas conjuntamente. Normalmente estas regiones no se pueden predecir de modo efectivo y con eficiencia con una estimación del movimiento y la codificación se puede realizar más eficientemente si tales regiones se eximen de las técnicas de codificación de cuadros intra (por ejemplo usando

codificación de cuadros B y cuadros P). En el contexto de los impedimentos de canal, esas regiones es probable que sufran de propagación de errores, que se pueden reducir o eliminar (o estar cerca de hacerlo) mediante la codificación de cuadros intra.

Partes del GOP de video se puede clasificar en dos o más categorías, en el que cada región puede tener un criterio de codificación de cuadros intra diferente que puede depender de la implementación particular. Como un ejemplo, el video se puede clasificar en tres categorías: cambios de escena abruptos, desvanecimiento cruzado y otros cambios de escena lentos y luces de flash de cámaras.

Los cambios de escena abruptos incluyen cuadros que son significativamente diferentes del cuadro previo, normalmente producidos por una operación de cámara. Dado que el contenido de estos cuadros es diferente del cuadro previo, los cuadros de cambio de escena abruptos se deberían codificar como cuadros I.

El desvanecimiento cruzado y otros cambios de escena lentos incluyen el cambio lento de escenas, normalmente producidos por un procesamiento por ordenador de las tomas de cámara. La mezcla gradual de dos escenas diferentes puede resultar más placentera a los ojos humanos, pero supone un reto a la codificación de video. La compensación de movimiento no puede reducir la tasa de bits de esos cuadros de modo efectivo, y se pueden actualizar más MB intra para estos cuadros.

Las luces de flash de cámara o eventos de flash de cámara, tienen lugar cuando el contenido de un cuadro incluye flashes de cámara. Tales flashes son de duración relativamente corta (por ejemplo un cuadro) y extremadamente brillantes de modo que los píxeles en un cuadro que represente los flashes exhiben una luminancia inusualmente alta con relación al área correspondiente en un cuadro adyacente. Las luces de flashes de cámara desplazan la luminancia de una imagen brusca y rápidamente. Normalmente la duración de un flash de cámara es más corto que la duración del enmascaramiento temporal del sistema de visión humana (HVS), que se define típicamente como de 44 ms. Los ojos humanos no son sensibles a la calidad de estas ráfagas cortas de brillo y por lo tanto se pueden codificar toscamente. Debido a que los cuadros de luces de flash no se pueden manejar de modo efectivo con compensación de movimiento y son malos candidatos para predicción de futuros cuadros, la codificación tosca de estos cuadros no reduce la eficiencia de la codificación de los cuadros futuros. Las escenas clasificadas como luces de flash no se deberían usar para predecir otros cuadros debido a luminancia artificialmente alta y otros cuadros no se pueden usar de modo efectivo para predecir estos cuadros por la misma razón. Una vez identificados, estos cuadros se pueden retirar debido a que pueden requerir una cantidad de procesamiento relativamente alta. Una opción es retirar los cuadros de luces de flash de cámara y codificar un coeficiente de DC en su lugar; tal solución es simple, computacionalmente rápida y ahorra muchos bits.

Cuando se detecta cualquiera de las categorías anteriores de cuadros, se declara un evento de toma. La detección de toma no solamente útil para mejorar la calidad de la codificación, también puede ayudar en la identificación de la búsqueda de indexado del contenido de video. Un aspecto ilustrativo de un proceso de detección de escena se describe en el presente documento a continuación. En este ejemplo, un proceso de detección de tomas calcula primero información, o métrica, para un cuadro seleccionado que se está procesando para la detección de tomas. Las métricas pueden incluir información de estimación de movimiento bidireccional y compensación de procesamiento del video y otras métricas basadas en la luminancia.

Para realizar una estimación/compensación de movimiento bidireccional, se puede reprocesar una secuencia de video con un compensador de movimiento bidireccional que adapta cada bloque de 8x8 del cuadro actual con bloques en dos de los cuadros más adyacentes de los cuadros contiguos, uno en el pasado y uno en el futuro. El compensador de movimiento produce vectores de movimiento y diferencia métricas para cada bloque. La Figura 29 es una ilustración que muestra un ejemplo del ajuste de los píxeles de un cuadro actual C a un cuadro pasado P y a un cuadro futuro (o siguiente) N, y representa los vectores de movimiento de los píxeles ajustados (vector de movimiento pasado MV_P , vector de movimiento futuro MV_N). Se describe en el presente documento en general a continuación una descripción de una generación del vector de movimiento bidireccional y la codificación relacionada con referencia a la FIGURA 32.

Después de determinar la información de movimiento bidireccional (por ejemplo información de movimiento que identifica los MB (mejor ajustados) en los cuadros adyacentes correspondientes), se puede generar métrica adicional (por ejemplo mediante un compensador de movimiento en el divisor del GOP 612 u otro componente adecuado) mediante varias comparaciones del cuadro actual con el cuadro siguiente y el cuadro previo. El compensador de movimiento puede producir una métrica de diferencia para cada bloque. La métrica de diferencia puede ser una suma de diferencia de cuadrados (SSD) o una suma de la diferencia absoluta (SAD). Sin pérdida de generalidad, aquí se usa la SAD como un ejemplo.

Para cada trama, se calcula una relación SAD, también denominada como una "relación de contraste", como sigue:

$$\gamma = \frac{\varepsilon + SAD_P}{\varepsilon + SAD_N} \quad [6]$$

en la que SAD_P y SAD_N son la suma de las diferencias absolutas de la métrica de diferencia directa e inversa,

respectivamente. Se debería indicar que el denominador contiene un número positivo pequeño ε para equilibrar el efecto de la unidad en el denominador. Por ejemplo, si el cuadro previo, el cuadro actual y el cuadro siguiente son idénticos, la búsqueda de movimiento conduciría a $SAD_P = SAD_N = 0$. En este caso, el cálculo anterior genera $\gamma = 1$ en lugar de 0 o infinito.

- 5 Se puede calcular para cada cuadro un histograma de luminancia. Típicamente las imágenes multimedia tienen una profundidad de luminancia de 8 bits (por ejemplo el número de "bins"). La profundidad de luminancia usada para el cálculo del histograma de luminancia de acuerdo con algunos aspectos se puede establecer en 16 para obtener el histograma. En otros aspectos, la profundidad de luminancia se puede establecer hasta un número apropiado que puede depender del tipo de datos que se están procesando, la potencia de cálculo disponible u otros criterios predeterminados. En algunos aspectos, la profundidad de luminancia se puede establecer dinámicamente en base a una métrica calculada o recibida, tal como el contenido de los datos.

La ecuación a continuación ilustra un ejemplo del cálculo de una diferencia de histograma de luminancia (λ):

$$\lambda = \frac{\sum_{i=1}^{16} |N_{Pi} - N_{Ci}|}{N} \quad [7]$$

- 15 en la que N_{Pi} es el número de bloques en el i^o bin para el cuadro previo y N_{Ci} es el número de bloques en el i^o bin para el cuadro actual y N es el número de bloques total en un cuadro. Si la diferencia de histograma de luminancia del cuadro previo y actual son completamente diferentes (o disjuntos), entonces $\lambda = 2$.

Usando esta información, se calcula una métrica de diferencia de cuadro (D) como sigue:

$$D = \frac{\gamma_C}{\gamma_P} + A\lambda(2\lambda + 1) \quad [8]$$

en la que A es una constante elegida por aplicación, $\gamma_C = \frac{\varepsilon + SAD_P}{\varepsilon + SAD_N}$ y $\gamma_P = \frac{\varepsilon + SAD_{PP}}{\varepsilon + SAD_C}$.

- 20 El cuadro seleccionado (actual) se clasifica como un cuadro de cambio de escena abrupto si la métrica de diferencia del cuadro satisface el criterio mostrado en la Ecuación 9:

$$D = \frac{\gamma_C}{\gamma_P} + A\lambda(2\lambda + 1) \geq T_1 \quad [9]$$

en la que A es una constante elegida por aplicación, y T_1 es un umbral.

- 25 En una simulación de ejemplo muestra que, fijando $A = 1$ y $T_1 = 5$, se consigue un buen rendimiento de detección. Si el cuadro actual es un cuadro de cambio de escena abrupta, entonces γ_C debería ser grande y γ_P debería ser pequeña. Se puede usar la relación $\frac{\gamma_C}{\gamma_P}$ en lugar de γ_C solamente de modo que se normalice la métrica al nivel de actividad del contexto.

- 30 Se debería indicar que el criterio anterior usa la diferencia de histograma de luminancia λ en una forma no lineal. La FIGURA 16 ilustra que $\lambda^*(2\lambda+1)$ es una función inversa. Cuando λ es pequeña (por ejemplo cercana a cero) apenas se preenfatisa. Cuanto más grande se haga λ , mayor énfasis se realiza por la función. Con este pre-énfasis, para cualquier λ mayor de 1,4, se detecta un cambio de escena abrupto si el umbral T_1 se establece en 5.

El cuadro actual se determina que es un desvanecimiento cruzado o un cambio de escena lento si la métrica de intensidad de escena D satisface el criterio mostrado en la Ecuación 5:

$$T_2 \leq D < T_1 \quad [10]$$

- 35 para un cierto número de cuadros continuos, en el que T_1 es el mismo umbral usado anteriormente y T_2 es otro valor de umbral.

- 40 Un evento de luz de flash normalmente produce que el histograma de luminancia salte a un lado más brillante. En este aspecto ilustrativo de cámara, la estadística del histograma de luminancia se usa para determinar si el cuadro actual comprende luces de flash de cámara. Un proceso de detección de toma puede determinar si la luminancia del cuadro actual menos es mayor que la luminancia del cuadro previo en un cierto umbral T_3 y la luminancia del cuadro actual es mayor que la luminancia del siguiente cuadro en el umbral T_3 , como se muestra las Ecuaciones 11 y 12.

$$\bar{Y}_C - \bar{Y}_P \geq T_3 \quad [11]$$

$$\bar{Y}_C - \bar{Y}_N \geq T_3 \quad [12]$$

Si no se satisface el criterio anterior, no se clasifica el cuadro actual como que comprende luces de flash de cámara. Si se satisface el criterio, el proceso de detección de toma determina si una métrica de diferencia hacia atrás SAD_P y la métrica de diferencia hacia delante SAD_N son mayores que un cierto umbral T_4 , como se ilustra en las Ecuaciones a continuación:

$$SAD_P \geq T_4 \quad [13]$$

$$SAD_N \geq T_4 \quad [14]$$

en la que \bar{Y}_C es la luminancia media del cuadro actual, \bar{Y}_P es la luminancia mediante el cuadro previo, \bar{Y}_N es la luminancia media del siguiente cuadro y SAD_P y SAD_N son las métricas de diferencia hacia adelante y hacia atrás asociadas con el cuadro actual.

El proceso de detección de toma determina eventos de flash de cámara determinando primero si la luminancia de un cuadro actual es mayor que la luminancia del cuadro previo y la luminancia del cuadro siguiente. Si no, el cuadro no es un evento de flash de cámara; pero si es sí, podría serlo. El proceso de detección de tomas puede evaluar entonces si la métrica de diferencia hacia atrás es mayor que un umbral T_3 y si la métrica de diferencia hacia delante es mayor que un umbral T_4 ; si ambas condiciones se satisfacen, el proceso de detección de tomas clasifica el cuadro actual como que tiene luces de flash de cámara. Si el criterio no se satisface, el cuadro no se clasifica como ningún tipo de eventos de toma o se le puede dar una clasificación por defecto que identifica la codificación a ser realizada en el cuadro (por ejemplo salto de cuadro codificado como un cuadro I).

Algunos valores de ejemplo para T_1 , T_2 , T_3 y T_4 se han mostrado anteriormente. Típicamente, estos valores de umbral se seleccionan por medio de pruebas de una implementación particular de detección de tomas. En algunos aspectos, uno más de los valores de umbral T_1 , T_2 , T_3 y T_4 están predeterminados y tales valores se incorporan en el clasificador de tomas en el dispositivo de codificación. En algunos aspectos, uno más de los valores de umbral T_1 , T_2 , T_3 y T_4 se pueden fijar durante el procesamiento (por ejemplo dinámicamente) en base al uso de información (por ejemplo metadatos) suministrada al clasificador de tomas o en base a la información calculada por el clasificador de tomas por sí mismo.

La codificación del video usando la información de detección de tomas se realiza típicamente en el codificador, pero se describe aquí para completar la divulgación de la detección de tomas. Con referencia a la FIGURA 30, un proceso de codificación 301 puede usar la información de detección de tomas para codificar el video en base a la toma detectada en la secuencia de cuadros. El proceso 301 prosigue en el bloque 303 y hace una comprobación para ver si el cuadro actual se clasifica como un cambio de escena abrupto. Si es así, en el bloque 305 del cuadro actual se puede codificar como un cuadro I y se puede determinar un límite de GOP. Si no, el proceso 301 prosigue en el bloque 307; si el cuadro actual se clasifica como una parte de una escena que cambia lentamente en el bloque 309 el cuadro actual y otros cuadros en la escena que cambia lentamente se puede codificar como un cuadro predictivo (por ejemplo un cuadro P o un cuadro B). El proceso 301 prosigue entonces hacia el bloque 311 en el que comprueba si el cuadro actual se clasificará como una escena de luz de flash que comprende flashes de cámara. Si es así, en el bloque 313 el cuadro se puede identificar para un procesamiento especial, por ejemplo, su eliminación o codificación con un coeficiente de DC para el cuadro; si no es así, no se realiza ninguna clasificación del cuadro actual y el cuadro actual se puede codificar de acuerdo con otros criterios, codificado como un cuadro I o retirado.

En el aspecto descrito anteriormente, la cantidad de diferencia entre el cuadro a ser comprimido y sus dos cuadros adyacentes se indica por la métrica de diferencia de cuadro D. Si se detecta una cantidad significativa de cambio unidireccional en la luminancia, significa un efecto de desvanecimiento cruzado en el cuadro. Cuanto más prominente sea el desvanecimiento cruzado, mayor ganancia se obtendrá mediante el uso de cuadros B. En algunos aspectos, se usa una métrica de diferencia de cuadro modificada como se muestra en la ecuación a continuación.

$$D_1 = \begin{cases} \left(1 - \alpha + 2\alpha \frac{|d_P - d_N|}{d_P + d_N}\right) \times D, & \text{si } Y_P - \Delta \geq Y_C \geq Y_N + \Delta \text{ ó } Y_P + \Delta \leq Y_C \leq Y_N - \Delta, \\ D, & \text{en caso contrario} \end{cases} \quad [15]$$

en la que $d_P = |Y_C - Y_P|$ y $d_N = |Y_C - Y_N|$ son la diferencia de luma entre el cuadro actual y el cuadro previo y la diferencia de luma entre el cuadro actual y el siguiente cuadro, respectivamente, Δ representa una constante que se puede determinar en una experimentación normal dado que puede depender de la implementación y α es una variable de ponderación que tiene un valor entre 0 y 1.

B. Generación del mapa de ancho de banda

El preprocesador 226 (FIGURA 6) se puede configurar también para generar un mapa de ancho de banda que se puede usar para la codificación de datos multimedia. En algunos aspectos, un módulo de clasificación del contenido

712 en el codificador 228 (FIGURA 7) genera el mapa de ancho de banda en su lugar.

La cualidad visual humana V puede ser una función tanto de la complejidad de codificación C como de los bits asignados B (también denominado como ancho de banda). La FIGURA 5 es un gráfico que ilustra esta relación. Se debería indicar que la métrica de complejidad de codificación C considera frecuencias espaciales y temporales desde el punto de vista de la visión humana. Para distorsiones más sensibles a los ojos humanos, el valor de complejidad es correspondientemente más alto. Puede asumirse típicamente que V es decreciente monótonamente en C y monótonamente creciente en B .

Para conseguir una calidad visual constante, se asigna un ancho de banda (B_i) al objeto i^o (cuadro o MB) a ser codificado que satisface los criterios expresados en las dos ecuaciones inmediatamente continuación:

$$B_i = B(C_i, V) \quad [16]$$

$$B = \sum_i B_i \quad [17]$$

En las dos ecuaciones inmediatamente anteriores, C_i es la complejidad de codificación del objeto i^o , B es el ancho de banda total disponible y V es la cualidad visual conseguida para un objeto. La cualidad visual humana es difícil de formular como una ecuación. Por lo tanto, la ecuación establecida anteriormente no se define con precisión. Sin embargo, si se supone que el modelo 3-D es continuo en todas las variables, la relación de ancho de banda (B_i/B) se puede tratar como inalterada dentro de la proximidad de un par (C, V) . La relación de ancho de banda β_i se define en la ecuación mostrada a continuación:

$$\beta_i = B_i / B \quad [18]$$

La asignación de bits se puede definir entonces como se expresa en las ecuaciones a continuación:

$$\beta_i = \beta(C_i) \quad [19]$$

$$1 = \sum_i \beta_i \quad \text{para } (C_i, V) \in \delta(C_0, V_0)$$

en la que δ indica la "vecindad".

La complejidad de codificación queda afectada por la sensibilidad visual humana, tanto espacial como temporal. El modelo de visión humana de Girod es un ejemplo de modelo que se puede usar para definir la complejidad espacial. Este modelo considera la frecuencia espacial local y la iluminación ambiente. La métrica resultante se denomina D_{csat} . En un punto de procesamiento en el proceso, si una imagen ha de ser intra-codificada o inter-codificada no es conocido y se generan las relaciones de ancho de banda para ambas. Los bits se asignan de acuerdo con la relación entre β_{INTRA} de diferentes objetos de video. Para imágenes intra-codificadas, la relación de ancho de banda se expresa en la ecuación a continuación:

$$\beta_{INTRA} = \beta_{0INTRA} \log_{10}(1 + \alpha_{INTRA} Y^2 D_{csat}) \quad [20]$$

En la ecuación anterior, Y es el componente de luminancia media de un macro bloque, α_{INTRA} es un factor de ponderación para el cuadrado de la luminancia y el término D_{csat} que la sigue, β_{0INTRA} es un factor de normalización para garantizar $1 = \sum_i \beta_i$. Por ejemplo, un valor para $\alpha_{INTRA} = 4$ consigue una calidad visual buena. La información

del contenido (por ejemplo una clasificación del contenido) se puede usar para establecer α_{INTRA} en un valor que corresponde a un nivel de calidad visual buena deseado para el contenido particular del video. En un ejemplo, si el contenido del video comprende una emisión de noticias con una "cabeza parlante", el nivel de calidad visual se puede fijar más bajo debido a que la imagen de información o parte que se puede visualizar del video puede ser considerado como menos importante que la parte de audio y se pueden asignar menos bits para codificar los datos. En otro ejemplo, si el contenido de video comprende un evento deportivo, la información de contenido se puede usar para fijar α_{INTRA} en un valor que corresponde a un nivel de calidad visual más alto debido a que las imágenes visualizadas pueden ser más importantes para un observador y en consecuencia se pueden asignar más bits para codificar los datos.

Para comprender esta relación, se debería indicar que se asigna ancho de banda logarítmicamente con la complejidad de la codificación. El término cuadrado de la luminancia Y^2 refleja el hecho de que los coeficientes con magnitudes más grandes usan más bits para codificarse. Para impedir que el logaritmo entre en valores negativos, se añade la unidad al término en el paréntesis. Se pueden usar también logaritmos con otras bases.

La complejidad temporal se determina mediante la medición de una métrica de diferencia de cuadro, que mide la diferencia entre dos cuadros consecutivos teniendo cuenta la cantidad de movimiento (vectores de movimiento). Con una métrica de diferencia de cuadro tal como la suma de las diferencias absolutas (SAD).

La asignación de bits para imágenes inter-codificadas pueden considerar la complejidad espacial así como la temporal. Esto se expresa a continuación:

$$\beta_{INTER} = \beta_{0INTER} \log_{10} \left(1 + \alpha_{INTER} \cdot SSD \cdot D_{csat} \exp \left(-\gamma \|MV_P + MV_N\|^2 \right) \right) \quad [21]$$

En la ecuación anterior MV_P y MV_N son los vectores de movimiento hacia adelante y hacia atrás del MB actual (véase la FIGURA 29). Se puede indicar que Y^2 en la fórmula del ancho de banda de intra-codificación se sustituye por la suma de las diferencias cuadradas (SSD). Para comprender el papel de $\|MV_P + MV_N\|^2$ en la ecuación anterior, nótese las siguientes características del sistema visual humano: las áreas que experimentan un movimiento suave, predecible (un pequeño $\|MV_P + MV_N\|^2$) atraen la atención y puede ser seguidas por el ojo y típicamente no pueden tolerar más distorsión que en las regiones fijas. Sin embargo, las áreas que experimentan un movimiento rápido o impredecible (un gran $\|MV_P + MV_N\|^2$) no pueden ser seguidas y puede tolerarse una cuantificación significativa. Los experimentos muestran que $\alpha_{INTRA} = 1$, $\gamma = 0,001$ consigue una buena calidad visual.

C. División del GOP adaptativo

En otro ejemplo ilustrativo del procesamiento que se puede realizar por el preprocesador 226, el Divisor de GOP 612 de la FIGURA 6 puede cambiar también adaptativamente la composición de un grupo de imágenes codificadas juntas y se explica con referencia un ejemplo que usa MPEG2. Algunas normas de compresión de video antiguas (por ejemplo MPEG2) no requieren que un GOP tenga una estructura regular, aunque se puede imponer una. La secuencia MPEG2 comienza siempre con un cuadro I, es decir uno que se ha codificado sin referencia a imágenes previas. El formato de GOP MPEG2 se predispone normalmente en el codificador mediante la fijación del espaciado en el GOP de las imágenes P o predictivas que siguen al cuadro I. Los cuadros P son imágenes que en parte se han predicho a partir de imágenes I o P previas. Los cuadros entre el cuadro I de comienzo y los cuadros P sucesivos se codifican como cuadros B. Un cuadro "B" (B se refiere a bidireccional) puede usar las imágenes previa y siguiente I o P tanto individualmente como simultáneamente como referencia. El número de bits necesarios para codificar un cuadro I como media excede el número de bits necesarios para codificar un cuadro P; de la misma manera el número de bits necesarios para codificar un cuadro P excede en media el necesario para un cuadro B. Un cuadro que se salta, si se usa, no requerirá ningún bit para su representación.

El concepto subyacente al uso de cuadros P y B y, en los algoritmos de compresión más recientes, el salto de cuadros para reducir la tasa de datos necesarios para representar el video es la eliminación de la redundancia temporal. Cuando la redundancia temporal es alta —es decir, hay un pequeño cambio de imagen a imagen— el uso de imágenes P, B o saltadas representa eficientemente la transmisión continua de video, debido a que las imágenes I o P decodificadas anteriormente se usan posteriormente como referencias para decodificar otras imágenes P o B.

La división del GOP adaptativo se basa en el uso de este concepto adaptativamente. Las diferencias entre cuadros se cuantifican y se toma una decisión automáticamente para representar la imagen mediante cuadros I, P, B o saltado después de que se hayan realizado las pruebas adecuadas sobre las diferencias cuantificadas. Una estructura adaptativa tiene ventajas no disponibles en una estructura de GOP fija. Una estructura fija ignorará la posibilidad de que ha tenido lugar un pequeño cambio en el contenido; un procedimiento adaptativo permitiría que se inserten muchos más cuadros B entre cada I y P o dos cuadros P, produciendo de ese modo el número de bits necesarios para representar adecuadamente la secuencia de cuadros. A la inversa cuando el cambio del contenido de video es significativo, la eficiencia de los cuadros P se reduce grandemente debido a que la diferencia entre los cuadros predichos y de referencia es demasiado grande. Bajo estas condiciones, el ajuste de los objetos puede caer fuera de las regiones de búsqueda de movimiento o la similitud entre los objetos de ajuste se reduce debido a la dispersión producida por los cambios en el ángulo de la cámara. En ese punto los cuadros P o el I que ésta adyacente al cuadro P se deberían elegir como más cercanos entre sí y se deberían insertar menos cuadros B. Un GOP dijo no puede realizar ese ajuste.

En el sistema desvelado en el presente documento, estas condiciones se detectan automáticamente. La estructura del GOP es flexible y se realiza para adaptarse a estos cambios en el contenido. El sistema evalúa una métrica de diferencia de cuadro, que se puede considerar como la medición de la distancia entre cuadros, con las mismas propiedades aditivas de distancia. Conceptualmente, dados los cuadros F_1 , F_2 y F_3 que tienen distancias entre cuadros d_{12} y d_{23} , la distancia entre F_1 y F_3 se toma como que es al menos $d_{12} + d_{23}$. Las asignaciones de cuadros se realizan en base a la métrica similar a distancia.

El divisor de GOP opera mediante la asignación de tipos de imagen a cuadros según se reciben. El tipo de imagen indica el procedimiento de predicción que se puede requerir en la codificación de cada bloque:

Las imágenes I se codifican sin referencia a otras imágenes. Dado que permanecen solitarias proporcionan puntos de acceso en la transmisión continua de datos cuando puede comenzar la decodificación. Se asigna un tipo de codificación I a un cuadro si la "distancia" a su cuadro predecesor excede un umbral de cambio de escena.

Las imágenes P pueden usar las imágenes I o P previas para una predicción compensada de movimiento. Usan

bloques en los campos previos o cuadros que se pueden desplazar desde el bloque que se predice como una base para la codificación. Después de que se reste el bloque de referencia del bloque que se está considerando, se codifica el bloque residual, usando típicamente la transformada de coseno discreta para la eliminación de la redundancia espacial. Se asigna un tipo de modificación P a un cuadro si la "distancia" entre él y el último cuadro asignado para ser un cuadro P excede un segundo umbral, que es típicamente menor que el primero.

Las imágenes de cuadros B pueden usar las imágenes previas y siguientes, P o I para una compensación de movimiento como se ha descrito anteriormente. Un bloque en una imagen B puede ser predicho hacia adelante, hacia atrás o bidireccionalmente o podría ser codificado internamente sin referencia a otros cuadros. En H.264 un bloque de referencia puede ser una combinación lineal de hasta 32 bloques de otros tantos cuadros. Si el cuadro no se puede asignar para ser un tipo I o P, se asigna como un tipo B, si la "distancia" desde él a su predecesor inmediato es mayor que un tercer umbral, que es típicamente menor que el segundo umbral.

Si el cuadro no se puede asignar para convertirse en un cuadro B codificado, se asigna un estatus de "cuadro saltado". Este cuadro se pueden saltar debido que es virtualmente una copia de un cuadro previo.

La evaluación de una métrica que cuantifica la diferencia entre cuadros adyacentes en el orden de visualización es la primera parte que tiene lugar en este procesamiento. Esta métrica es la distancia a la que se ha referido anteriormente; con ella, cada cuadro se evalúa para su tipo propio. Por ello, el espaciado entre los cuadros I y el P adyacente o dos cuadros P consecutivos, puede ser variable. El cálculo de la métrica comienza por el procesamiento de los cuadros de video con un compensador de movimiento basado en bloques, siendo un bloque la unidad básica de compresión de video, compuesta normalmente por 16x16 píxeles, aunque son posibles otros tamaños de bloques tales como 8x8, 4x4 y 8x16. Para cuadros que consisten en dos campos desentrelazados, la compensación del movimiento se puede realizar en base a un campo, teniendo lugar la búsqueda de los bloques de referencia en los campos más que en los cuadros. Para un bloque en el primer campo del cuadro actual se encuentra un bloque de referencia hacia adelante en los campos del cuadro que lo siguen; de la misma manera se halla un bloque de referencia hacia atrás en campos del cuadro que precede inmediatamente al campo actual. Los bloques actuales se montan en un campo compensado. El proceso continúa con el segundo campo del cuadro. Los dos campos compensados se combinan para formar un cuadro compensado hacia adelante y hacia atrás.

Para cuadros creados en el telecine inverso 606, la búsqueda de bloques de referencia es en base solamente a cuadros, dado que sólo hay cuadros de película reconstruidos. Se hallan dos bloques de referencia y dos diferencias, hacia adelante y hacia atrás, conduciendo también a un cuadro compensado hacia adelante y hacia atrás. En resumen, el compensador de movimiento produce vectores de movimiento y métrica de diferencia para cada bloque; pero un bloque es parte de un campo NTSC en el caso de que se esté procesando la salida del desentrelazador 605 y sea parte de un cuadro de película si se procesa la salida del telecine inverso. Nótese que las diferencias en la métrica se evalúan entre un bloque en el campo o cuadro que está siendo considerado y un bloque que se ajusta mejor a él, tanto en un campo o cuadro precedente como en un campo o cuadro que lo sigue inmediatamente, dependiendo de si se está evaluando la diferencia hacia adelante o hacia atrás. Sólo los valores de luminancia entran en este cálculo.

La etapa de compensación de movimiento garantiza de ese modo dos conjuntos de diferencias. Éstas están entre bloques de valores actuales de luminancia y los valores de luminancia en los bloques de referencia tomados a partir de cuadros que están inmediatamente por delante e inmediatamente detrás del cuadro uno en el tiempo. El valor absoluto de cada diferencia hacia adelante y cada diferencia hacia atrás se determina para cada píxel y se suma cada uno por separado a través del cuadro completo. Ambos campos incluyen en las dos sumas cuando se procesan los campos NTSC desentrelazados que comprenden un cuadro. En esta forma se hallan SAD_P y SAD_N , los valores absolutos sumados de las diferencias hacia adelante y hacia atrás.

Para cada cuadro se calcula una relación SAD usando la relación,

$$\gamma = \frac{\varepsilon + SAD_P}{\varepsilon + SAD_N} \quad [22]$$

en la que SAD_P y SAD_N son los valores absolutos sumados de las diferencias hacia adelante y hacia atrás, respectivamente. Se añade un número positivo pequeño ε al numerador para impedir el error de "división por cero". Se añade un término ε similar al denominador, reduciendo adicionalmente la sensibilidad de γ cuando tanto SAD_P como SAD_N están cerca de cero.

En un aspecto alternativo, la diferencia puede ser el SSD, la suma de las diferentes cuadradas y SAD, la suma de las diferencias absolutas, o la SATD en la que los bloques de los valores de píxeles se transforma mediante la aplicación de la Transformada de Coseno Discreta bidimensional a ellos antes de que se tomen las diferencias en los elementos del bloque. Las sumas se evalúan a través del área del video, aunque se puede usar un área más pequeña en otros aspectos.

Se calcula también el histograma de luminancia de cada cuadro según se recibe (sin compensación de movimiento). El histograma opera en el coeficiente DC, es decir el coeficiente (0,0), en la matriz de coeficientes de 16x16 que es

el resultado de la aplicación de la Transformada de Coseno Discreta bidimensional al bloque de valores de luminancia si están disponibles. De modo equivalente se puede usar el valor medio de los 256 valores de luminancia en el bloque 16x16 en el histograma. Para imágenes cuya profundidad de luminancia es de ocho bits, el número de bins se establece en 16. La siguiente métrica evalúa la diferencia de histograma

$$5 \quad \lambda = \frac{1}{4} \sum_{i=1}^{16} |N_{Pi} - N_{Ci}| \quad [23]$$

en la anterior, N_{Pi} es el número de bloques del cuadro previo en el i^{o} bin y N_{Ci} es el número de bloques del cuadro actual que pertenecen en el i^{o} bin, N es el número total de bloques en un cuadro.

Estos resultados intermedios se reúnen para formar la métrica de diferencia del cuadro actual como

$$D = \frac{\gamma_C}{\gamma_P} + \lambda(2\lambda + 1) \quad [24]$$

10 en la que γ_C es la relación SAD en base al cuadro actual y γ_P es la relación SAD en base al cuadro previo. Si una escena tiene un movimiento suave y su histograma de luma apenas cambia, entonces $D \approx 1$. Si el cuadro actual visualiza un cambio de escena abrupto entonces γ_C será grande y γ_P debería ser pequeño. Se usa la relación $\frac{\gamma_C}{\gamma_P}$ en lugar de γ_C en solitario de modo que se normaliza la métrica al nivel de actividad del contexto.

15 La FIGURA 42 ilustra el proceso de asignación de tipos de compresión a cuadros. D , la diferencia de cuadro actual definida en la Ecuación 19, es la base para decisiones tomadas con respecto a las asignaciones de cuadros. Como indica el bloque de decisión 4202, si el cuadro bajo consideración es el primero en una secuencia, se sigue la trayectoria de decisión marcada SÍ al bloque 4206, declarando de ese modo el cuadro como un cuadro I. Las diferencias de cuadro acumuladas se establecen en cero en el bloque 4208 y el proceso vuelve (en el bloque 4210) al bloque de comienzo. Si el cuadro que está siendo considerado no es el primer cuadro en una secuencia, se sigue la trayectoria marcada NO desde el bloque 4202 en el que se tomó la decisión y en el bloque de prueba 4204 se prueba la diferencia de cuadro actual contra el umbral de cambio de escena. Si la diferencia de cuadro actual es mayor que ese umbral, se sigue la trayectoria de decisión marcada SÍ al bloque 4206, conduciendo de nuevo a la asignación de un cuadro I.

25 Si la diferencia de cuadro actual es menor que el umbral de cambio de escena, se sigue la trayectoria NO al bloque 4212 en el que se añade la diferencia de cuadro actual a la diferencia de cuadro acumulada. Continuando a través del diagrama de flujo, en el bloque de decisión 4214 se compara la diferencia de cuadro acumulada con el umbral t , que es en general menor que el umbral de cambio de escena. Si la diferencia de cuadro acumulada es mayor que t , el control se transfiere al bloque 4216 y se asigna el cuadro como un cuadro P; la diferencia de cuadros acumulada se repone entonces a cero en la etapa 4218. Si la diferencia de cuadros acumulada es menor que t , el control se transfiere desde el bloque 4214 al bloque 4220. Allí la diferencia de cuadros actual se compara con τ , que es menor que t . Si la diferencia de cuadros actual es más pequeña que τ , se asigna el cuadro para ser saltado en el bloque 4222 y entonces el proceso retorna; si la diferencia de cuadro actual es mayor que τ , se asigna el cuadro para ser un cuadro B en el bloque 4226.

Codificador

35 Volviendo a referirnos a la FIGURA 2, el transcodificador 200 incluye un codificador 228 que recibe metadatos procesados y video en bruto del preprocesador 226. Los metadatos pueden incluir cualquier información recibida originalmente en el video de origen 104 y cualquier información calculada por el preprocesador 226. El codificador 228 incluye un codificador de un primer paso 230, un codificador de un segundo paso 232 y un recodificador 234. El codificador 228 recibe también una entrada desde el control del transcodificador 231 que puede proporcionar información (por ejemplo metadatos, información de resistencia frente a errores, información del contenido, información de la tasa de bits codificadas, información de equilibrio entre la capa base y la capa de mejora e información de cuantificación) desde el codificador del segundo paso 232 al codificador del primer paso 230, el recodificador 234 así como el preprocesador 226. El codificador 228 codifica el video recibido usando la información del contenido recibida del preprocesador 226 y/o la información del contenido que se genera por el codificador 228 por sí mismo, por ejemplo, mediante el módulo de clasificación del contenido 712 (FIGURA 7).

La FIGURA 7 ilustra un diagrama de bloques de módulos funcionales que se pueden incluir en un codificador en dos pasos de ejemplo que se puede usar para el codificador 228 que se ilustra en la FIGURA 2. Se muestran en la FIGURA 7 varios aspectos de los módulos funcionales, aunque la FIGURA 7 y la descripción en el presente documento no se dirigen necesariamente a toda la funcionalidad que se puede incorporar en un codificador. En consecuencia, ciertos aspectos de los módulos funcionales se describen a continuación siguiendo a la explicación de la codificación de las capas base y de mejora a continuación.

Codificación de la capa base y de la capa de mejora

El codificador 228 puede ser un codificador escalable en SNR, que puede codificar el video en bruto y los metadatos desde el preprocesador 226 en un primer grupo de datos codificados, también denominados en el presente documento como una capa base y uno o más grupos adicionales de datos codificados, también denominados en el presente documento como capas de mejora. Un algoritmo de codificación genera los coeficientes de la capa base y de la capa de mejora que, cuando se decodifican, se pueden combinar en el decodificador cuando están ambas capas disponibles para la decodificación. Cuando no están disponibles ambas capas, la codificación de la capa base le permite ser decodificada como una única capa.

Se describe con referencia a la FIGURA 31 un aspecto de dicho proceso de codificación de capa múltiple. En el bloque de 321, se codifica un cuadro I con macro bloques completamente intra-codificados (MB intra-codificados). En H.264, los MB intra-codificados en cuadros I se decodifican con una predicción espacial completamente aprovechada, que proporciona una cantidad significativa de ganancia de codificación. Hay dos submodos: intra 4x4 e intra 16x16. Si la capa base ha de aprovechar la ganancia de codificación proporcionada por la previsión espacial, entonces la capa base necesita estar codificada y decodificada antes de la codificación y decodificación de la capa de mejora. Se usa una codificación y decodificación de los cuadros I en dos pasos. En la capa base, un parámetro de cuantificación de la capa base QP_b permite a los coeficientes de transformación un tamaño de salto de cuantificación tosco. La diferencia en cuanto a píxeles entre el cuadro original y el cuadro de la capa base reconstruido se codifica en la capa de mejora. La capa de mejora usa un parámetro de cuantificación QP_e que permite un tamaño de salto de cuantificación más fino. Los medios de codificación, tal como codificador 228 de la FIGURA 2 pueden realizar la codificación en el bloque 321.

En el bloque 323, un codificador codifica los datos de la capa base y los datos de la capa de mejora para los cuadros P y/o B en el GOP que están siendo procesados. Los medios de codificación, tal como el codificador 228 puede realizar la codificación en el bloque 323. En el bloque 325, el proceso de codificación comprueba si hay más cuadros P o B para codificar. Los medios de codificación, tal como el codificador escalable en SNR 228 pueden realizar el acto 325. Si permanecen más cuadros P o B, se repite la etapa 323 hasta que todos los cuadros en el GOP hayan acabado de ser codificados. Los cuadros P y B están compuestos de macro bloques intra-codificados (MB intra-codificados), aunque puede haber MB intra-codificados en los cuadros P y B como se explicará a continuación.

Para que el decodificador distinga entre los datos de la capa base y de la capa de mejora, el codificador 228 codifica una información de cabecera adicional, en el bloque 327. Los tipos de información de cabecera adicional incluyen, por ejemplo, datos que identifican el número de capas, datos que identifican una capa como una capa base, datos que identifican una capa como una capa de mejora, datos que identifican las interrelaciones entre capas (tal como, la capa 2 es una capa de mejora para la capa base 1, o la capa 3 es una capa de mejora para la capa base 2) o datos que identifican una capa como una capa de mejora final en una cadena de capas de mejora. La información de cabecera adicional se puede contener en cabeceras conectadas con los datos de la capa base y/o de mejora a la que pertenecen, o estar contenidos en mensajes de datos separados. Los medios de codificación, tal como el codificador 228 de la FIGURA 2 pueden realizar el proceso en el bloque 327.

Para tener una decodificación de capa simple, los coeficientes de las dos capas se combinan antes de la cuantificación inversa. Por lo tanto los coeficientes de las dos capas han de ser generados de modo interactivo; en caso contrario esto podría introducir una cantidad significativa de cabecera adicional. Una razón para el incremento de la cabecera adicional es que la codificación de la capa base y la codificación de la capa de mejora podrían usar referencias temporales diferentes. Se necesita un algoritmo para generar los coeficientes de la capa base y de la capa de mejora, que se puede combinar en el decodificador antes de la cuantificación cuando están disponibles ambas capas. A la vez, el algoritmo debería proporcionar un video de capa base aceptable cuando la capa de mejora no esté disponible o el decodificador decida no decodificar la capa de mejora por razones tales como, por ejemplo, ahorro de energía. Los detalles de un ejemplo ilustrativo de un proceso así se explicarán adicionalmente a continuación en el contexto de una breve explicación de la codificación predictiva estándar inmediatamente a continuación.

Los cuadros P (o cualquier sección intra-codificada) pueden aprovechar la redundancia temporal entre una región en una imagen actual y una región de previsión del ajuste mejor en una imagen de referencia. La localización de la región de previsión del ajuste mejor en el cuadro de referencia se puede codificar en un vector de movimiento. La diferencia entre la región actual y la región de previsión de referencia del ajuste mejor es conocida como error residual (o error de predicción).

La FIGURA 32 es una ilustración de un ejemplo de un proceso de construcción de un cuadro P en, por ejemplo, MPEG-4. El proceso 331 es una ilustración más detallada de un proceso de ejemplo que podría tener lugar en el bloque 323 de la FIGURA 31. El proceso 331 incluye la imagen actual 333 formada por 5 x 5 macro bloques, en el que el número de macro bloques en este ejemplo es arbitrario. Un macro bloque está formado por 16 x 16 píxeles. Los píxeles se pueden definir por un valor de luminancia de 8 bits (Y) y dos valores de crominancia de 8 bits (Cr y Cb). En MPEG los componentes Y, Cr y Cb se pueden almacenar en un formato 4:2:0, en el que los componentes Cr y Cb se inframuestran en 2 en las direcciones X e Y. De ahí que, cada macro bloque podría consistir en 256 componentes Y, 64 componentes Cr y 64 componentes Cb. El macro bloque 335 de la imagen actual 333 se predice a partir de la imagen de referencia 337 en un punto en el tiempo diferente al de la imagen actual 333. Se realiza una búsqueda en la imagen de referencia 337 para localizar el macro bloque del ajuste mejor 339 que esté más cercano,

en términos de los valores de Y, Cr y Cb al macro bloque actual 335 que está siendo codificado. La localización del macro bloque del ajuste mejor 339 en la imagen de referencia 337 se codifica en el vector de movimiento 341. La imagen de referencia 337 puede ser un cuadro I o un cuadro P que un decodificador habrá reconstruido previamente a la construcción de la imagen actual 333. El macro bloque del ajuste mejor 339 se resta del macro bloque actual 335 (se calcula una diferencia para cada uno de los componentes Y, Cr y Cb) dando como resultado un error residual 343. El error residual 323 se codifica con una Transformada de Coseno Discreta (DCT) 2D 345 y se cuantifica 347 a continuación. La cuantificación 347 se puede realizar para proporcionar una compresión espacial mediante, por ejemplo, la asignación de menos bits a los coeficientes de alta frecuencia mientras que se asignan más bits a los coeficientes de baja frecuencia. Los coeficientes de cuantificación del error residual 343, junto con el vector de movimiento 341 y la imagen de referencia 333 que identifican la información, se codifica en la información que representa el macro bloque actual 335. La información codificada se puede almacenar en la memoria para un uso futuro o utilizada para finalidades de, por ejemplo, corrección de error o mejora de imagen, o transmitido a través de la red 349.

Los coeficientes cuantificados codificados del error residual 343, junto con el vector de movimiento codificado 341 se pueden usar para reconstruir el macro bloque actual 335 en el codificador para su uso como parte de un cuadro de referencia para una estimación y compensación del movimiento posterior. El codificador puede emular el proceso de un decodificador para esta reconstrucción del cuadro P. La emulación del decodificador dará como resultado que tanto el codificador como el decodificador trabajan con la misma imagen de referencia. El proceso de reconstrucción, tanto si se realiza en codificador, para una inter-codificación adicional como en un decodificador, se presenta en el presente documento. La reconstrucción de un cuadro P se puede comenzar después de que el cuadro de referencia (o una parte de una imagen o cuadro que está siendo referenciada) se reconstruya. Los coeficientes cuantificados codificados se decuantifican 351 y entonces se realiza una DCT inversa 2D, o IDCT, 353 dando como resultado el error residual decodificado o reconstruido 355. El vector de movimiento codificado 341 se decodifica y usa para localizar el macro bloque del ajuste mejor ya reconstruido 357 en la imagen de referencia ya reconstruida 337. El error residual reconstruido 355 se añade entonces al macro bloque del ajuste mejor reconstruido 357 para formar el macro bloque reconstruido 359. El macro bloque reconstruido 359 se puede almacenar en la memoria, visualizar independientemente o en una imagen con otros macro bloques reconstruidos, o ser procesado adicionalmente para mejora de la imagen.

Los cuadros B (o cualquier sección codificada con predicción bidireccional) pueden aprovechar la redundancia temporal entre una región en una imagen actual y una región de predicción del ajuste mejor en una imagen previa y una región de predicción del ajuste mejor en una imagen posterior. La región de predicción del ajuste mejor posterior y la región de predicción del ajuste mejor previa se combinan para formar una región de predicción bidireccional combinada. La diferencia entre región de la imagen actual y la región de predicción bidireccional combinada del ajuste mejor es un error residual (o error de predicción). Las localizaciones de la región de predicción del ajuste mejor en la imagen de referencia posterior y la región de predicción del ajuste mejor en la imagen de referencia previa se pueden codificar en dos vectores de movimiento.

La FIGURA 33 ilustra un ejemplo de un proceso de codificador para la codificación de los coeficientes de la capa base y de la capa de mejora que se puede realizar por el codificador 228. Las capas base y de mejora se codifican para proporcionar flujos de bits escalables en SNR. La FIGURA 33 representa un ejemplo para la codificación de coeficientes de error residual inter MB tal como se haría en la etapa 323 de la FIGURA 31. Sin embargo se podrían usar procedimientos similares para codificar los coeficientes intra MB asimismo. Los medios de codificación tal como el componente codificador 228 de la FIGURA 2 pueden realizar el proceso ilustrado en la FIGURA 33 y la etapa 323 de la FIGURA 32. Los datos de video originales (a ser codificados) 361 (los datos de video comprenden información de luma y croma en este ejemplo) se introducen en un bucle de macro bloque de ajuste mejor de la capa base 363 y un bucle de macro bloque del ajuste mejor de la capa de mejora 365. El objeto de ambos bucles 363 y 365 es minimizar el error residual que se calcula en los sumadores 367 y 369 respectivamente. Los bucles 363 y 365 se pueden realizar en paralelo, como se muestra, o secuencialmente. Los bucles 363 y 365 incluyen lógica para la búsqueda en las memorias intermedias 371 y 373, respectivamente, que contienen cuadros de referencia, para identificar el macro bloque del ajuste mejor que minimice el error residual entre el macro bloque del ajuste mejor y los datos originales 361 (las memorias intermedias 371 y 373 pueden ser la misma memoria intermedia). Los errores residuales del bucle 363 y del bucle 365 serán diferentes dado que el bucle de la capa base 363 utilizará generalmente un tamaño de etapa de cuantificación más tosco (un valor QP más alto) que el bucle de la capa de mejora 365. Los bloques de transformación 375 y 377 transforman los errores residuales para cada bucle.

Los coeficientes transformados se separan entonces en coeficientes de la capa base y de la capa de mejora en el selector 379. La separación del selector 379 puede tomar varias formas, como se explica a continuación. Una característica común de las técnicas de separación es que el coeficiente de la capa de mejora, C'_{mej} , se calcula de modo que sea un refinamiento diferencial al coeficiente de la capa base C'_{base} . El cálculo de la capa de mejora para ser un refinamiento de la capa base permite a un decodificador decodificar el coeficiente de la capa base por sí mismo y tener una representación razonable de la imagen o combinar los coeficientes de la capa base y de mejora y tener una representación refinada de la imagen. Los coeficientes seleccionados por el selector 379 se cuantifican por los cuantificadores 381 y 383. Los coeficientes cuantificados \tilde{C}'_{base} y \tilde{C}'_{mej} (calculados con los cuantificadores 381 y 383 respectivamente) se pueden almacenar en la memoria o transmitir a través de una red a un decodificador.

Para ajustar la reconstrucción del macro bloque en un decodificador, el decuantificador 385 decuantifica los coeficientes de error residual de la capa base. Los coeficientes de error residual decuantificados se transforman a la inversa 387 se añaden 389 al macro bloque del ajuste mejor hallado en la memoria intermedia 371, dando como resultado un macro bloque reconstruido que se ajusta a lo que se reconstruirá en el decodificador. El cuantificador 383, decuantificador 391, transformador inverso 393, sumador 397 y memoria intermedia 373 realizan cálculos similares en el bucle de mejora 365 a los que se realizaron en el bucle de la capa base 363. Además, el sumador 393 se usa para combinar los coeficientes de la capa de mejora y de la capa base decuantificados en la reconstrucción de la capa de mejora. El cuantificador y decuantificador de la capa de mejora utilizará generalmente un tamaño de etapa de cuantificador más fino (un QP más bajo) que en la capa base.

Las FIGURAS 34, 35 y 36 muestran ejemplos de los procesos de selector de coeficiente de la capa base y de la capa de mejora que se pueden emplear en el selector 379 de la FIGURA 33. Los medios de selección tal como el codificador 228 de la FIGURA 2 pueden realizar los procesos representados en las FIGURAS 34, 35 y 36. Usando la FIGURA 34 como ejemplo, los coeficientes transformados se dividen en coeficientes de la capa base y de mejora como se muestra en las siguientes ecuaciones:

$$C'_{base} = \begin{cases} 0, & \text{si } C_{base} \text{ y } C_{mej} \text{ tienen signos opuestos} \\ \min(C_{base}, C_{mej}), & \text{en caso contrario} \end{cases} \quad [25]$$

$$C'_{mej} = C_{mej} - Q_b^{-1} (Q_b (C'_{base})) \quad [26]$$

en la que la función "min" puede ser tanto un mínimo matemático como una magnitud mínima de los 2 argumentos. La Ecuación 25 se representa como el bloque 401 y la Ecuación 26 se representa como el sumador 510 en la FIGURA 34. En la Ecuación 26, Q_b se refiere al cuantificador de la capa base 381 y Q_b^{-1} se refiere al decuantificador de la capa base 385. La Ecuación 26 convierte el coeficiente de la capa de mejora en un refinamiento diferencial del coeficiente de la capa base calculado con la Ecuación 25.

La FIGURA 35 es una ilustración de otro ejemplo de un selector del coeficiente de la capa base y de la capa de mejora 379. En este ejemplo la Ecuación (.) contenida en el bloque 405 representa lo siguiente.

$$C'_{base} = \begin{cases} C_{base}, & \text{si } |Q_b^{-1} Q_b (C_{base}) - C_{mej}| < |C_{mej}| \\ 0, & \text{en caso contrario} \end{cases} \quad [27]$$

El sumador 407 calcula el coeficiente de la capa de mejora como se muestra en las siguientes dos ecuaciones:

$$C'_{mej} = C_{mej} - Q_b^{-1} (Q_b (C'_{base})) \quad [28]$$

en la que C'_{base} viene dada por la Ecuación 27.

La FIGURA 36 es una ilustración de otro ejemplo de un selector de una capa base y de una capa de mejora 379. En este ejemplo, el coeficiente de la capa base no se cambia y la capa de mejora es igual a la diferencia entre el coeficiente de la capa base cuantificado/decuantificado y el coeficiente de la capa de mejora original.

Además de los coeficientes de error residual de la capa base y de mejora, el decodificador necesita información que identifique como se codifican los MB. Los medios de codificación tal como el componente codificador 228 de la FIGURA 2 pueden codificar la información de la cabecera adicional que puede incluir un mapa de partes intra-codificadas e inter-codificadas, tal como, por ejemplo un mapa de MB en el que los macro bloques (o submacro bloques) se identifican como que están intra-codificados o inter-codificados (también identificando qué tipo de inter-codificación, incluyendo, por ejemplo la directa, inversa o bidireccional) y que partes de los cuadros inter-codificados se referencian. En un aspecto de ejemplo, el mapa de MB y los coeficientes de la capa base se codifican en la capa base y los coeficientes de la capa de mejora se codifican en la capa de mejora.

Los cuadros P y los cuadros B pueden contener MB intra-codificados así como inter MB. Es común para codificadores de video híbridos el uso de una optimización de distorsión de tasa (RD) para decidir codificar ciertos macro bloques en cuadros P o B como MB intra-codificados. Para tener una decodificación de capa simple en donde los MB intra-codificados no dependen de los inter MB de la capa de mejora, no se usa ningún inter MB contiguo para predicción espacial de los MB intra-codificados de la capa base. Para mantener la complejidad de cálculo sin cambio para la decodificación de la capa de mejora, para los MB intra-codificados en los cuadros P o B de la capa base, se podría saltar el refinamiento en la capa de mejora.

Los MB intra-codificados en cuadros P o B puede requerir muchos más bits que los inter MB. Por esta razón, los MB intra-codificados en cuadros P o B se podrían codificar solamente en la calidad de la capa base con un QP más alto. Esto introducirá algún deterioro en la calidad de video, pero este deterioro debería pasar desapercibido si se refina en un último cuadro con los coeficientes inter MB de la capa base y de mejora como se ha explicado anteriormente. Dos razones hacen de este deterioro desapercibido. El primero es una característica del sistema visual humano (HVS) y el otro es que los inter MB refinan los intra MB. Con objetos que cambian de posición desde el primer

cuadro al segundo cuadro, algunos píxeles en el primer cuadro no son visibles en el segundo cuadro (información a ser cubierta) y algunos píxeles en el segundo cuadro son visibles por primera vez (información descubierta). Los ojos humanos no son sensibles a la información visual descubierta y a ser cubierta. De modo que para la información descubierta, incluso si se codifica con una calidad más baja, los ojos pueden no notar la diferencia. Si la misma información permanece en el siguiente cuadro P, habrá una alta probabilidad de que el siguiente cuadro P en la capa de mejora pueda refinarlo debido a que la capa de mejora tiene un QP más bajo.

Otra técnica común que introducen los MB intra-codificados en cuadros P o B es conocida como Intra Refresco. En este caso, algunos MB se codifican como MB intra-codificados, incluso aunque la optimización de R-D estándar dictaría que deberían ser MB inter-codificados. Estos MB intra-codificados, contenidos en la capa base, se pueden codificar o bien con un QP_b o bien con un QP_e . Si se usa el QP_e para la capa base, entonces no se necesita ningún refinamiento en la capa de mejora. Si se usa QP_b para la capa base, entonces puede ser necesario un refinamiento, en caso contrario en la capa de mejora, la caída de calidad será notable. Dado que las inter-codificaciones son más eficientes que las intra-codificaciones en el sentido de la eficiencia de codificación, estos refinamientos en la capa de mejora serán inter-codificados. En esta forma, los coeficientes de la capa base no se usarán para la capa de mejora. Por lo tanto, se obtiene una mejora de calidad en la capa de mejora sin introducir nuevas operaciones.

Los cuadros B se usan comúnmente en las capas de mejora debido a las elevadas cualidades de compresión que ofrecen. Sin embargo, los cuadros B pueden tener que referenciar MB intra-codificados de un cuadro P. Si los píxeles del cuadro B fuesen a ser codificados en una calidad de capa de mejora, pueden requerir demasiados bits debido a la calidad más baja de los MB intra-codificados del cuadro P, como se ha explicado anteriormente. Aprovechando las cualidades del HVS, como se ha explicado anteriormente, los MB de cuadro B podrían codificarse con una calidad más baja cuando referencian MB intra-codificados de cuadros P de calidad más baja.

Un caso extremo de MB intra-codificados de cuadros P o B es cuando todos los MB en un cuadro P o B se codifican en el modo intra debido a la presencia de un cambio de escena en el video que se está codificando. En este caso, el cuadro completo se puede codificar en la calidad de capa base sin ningún refinamiento en la capa de mejora. Si tiene lugar un cambio de escena en un cuadro B y se supone que los cuadros B sólo se codifican en la capa de mejora, entonces el cuadro B se podría codificar en la calidad de la capa base o simplemente eliminarse. Si tiene lugar un cambio de escena en un cuadro P, no se necesitaría ningún cambio, pero el cuadro P podía ser eliminado o codificado en una calidad de capa base. La codificación en capas escalable se describe adicionalmente en la Solicitud de Patente de Estados Unidos Pendiente junto con la presente N^o [expediente de representante n^o de referencia 050078] titulada "SCALABLE VIDEO CODING WITH TWO LAYER ENCODING AND SINGLE LAYER DECODING" y propiedad del asignatario de la presente.

Parte del codificador de primer paso

La FIGURA 7 muestra un ejemplo ilustrativo del codificador 228 de la FIGURA 2. Los bloques mostrados ilustran varios procesamientos de codificador que se pueden incluir en el codificador 228. En este ejemplo, el codificador 228 incluye una parte de primer paso 706 por encima de una línea de demarcación 704, y una parte de segundo paso 706 (que incluye la funcionalidad de un codificador de segundo paso 232 y recodificador 234 de la FIGURA 2) debajo de la línea 704.

El codificador 228 recibe metadatos y video en bruto desde el preprocesador 226. Los metadatos pueden incluir cualquier metadato recibido o calculado por el preprocesador 226, incluyendo metadatos relacionados con la información del contenido del video. La parte del primer paso 702 del codificador 228 ilustra procesos de ejemplo que se pueden incluir en una codificación del primer paso 702, que se describe a continuación en términos de su funcionalidad. Como conocerá un experto en la técnica, tal funcionalidad se puede realizar de varias formas (por ejemplo mediante hardware, software, firmware o una combinación de los mismos).

La Figura 7 ilustra un módulo de intra refresco adaptativo (AIR). El módulo AIR 710 proporciona una entrada a un módulo de instanciación del cuadro I 708 que instancia un cuadro I en base a los metadatos. La parte del primer paso 702 puede incluir también un módulo de clasificación del contenido 712 configurado para recibir los metadatos y el video y determinar la información de contenido en relación con el video. La información del contenido se puede proporcionar a un módulo de asignación de bits y control de tasas 714, que recibirá también los metadatos y el video. El módulo de asignación de bits de control 714 determina la información de control de la tasa de bits y la proporciona al módulo de decisión de modo 715. La información del contenido del video se puede proporcionar a un módulo de intra-modelo (distorsión) 716, que proporciona la información de distorsión de intra-codificación al módulo de decisión de modo 715 y una escalabilidad de tasa-distorsión para el módulo de las capas base y de mejora 718. Se proporcionan el video los metadatos a un módulo de estimación de movimiento (distorsión) 720 que proporciona información de distorsión de inter-codificación al módulo de escalabilidad de tasa-distorsión para las capas base y de mejora 718. El módulo de escalabilidad de tasa-distorsión para las capas base y de mejora 718 determina la información de tasa de una distorsión de escalabilidad usando las estimaciones de distorsión a partir del módulo de estimación de movimiento 720 y del módulo de distorsión del intra modelo 716 que se proporcionan al módulo de decisión de modo 715. El módulo de decisión de modo 715 recibe también entradas desde el módulo de ordenación porciones/MB 722. El módulo de ordenación porciones/MB 722 recibe entradas desde un módulo de resistencia frente a errores 740 (mostrado en la parte del segundo paso 706) y proporciona información sobre la alineación de

las partes de video (porciones) codificables independientemente con límites de unidad de acceso para resistencia a los errores al módulo de decisión de modo 715. El módulo de decisión de modo 715 determina la información del modo de codificación en base a sus entradas y proporciona el "mejor" modo de codificación a la parte del segundo paso 706. Se describe a continuación una explicación ilustrativa adicional de algunos ejemplos de dicha codificación de la parte del primer paso 702.

Como se ha establecido anteriormente, el módulo de clasificación de contenido 712 recibe los metadatos y el video en bruto suministrados por el preprocesador 226. En algunos ejemplos, el preprocesador 226 calcula la información del contenido partir de los datos multimedia y proporciona la información del contenido al módulo de clasificación de contenido 712 (por ejemplo los metadatos), que puede usar la información del contenido para determinar una clasificación de contenido para los datos multimedia. En algunos aspectos, el módulo de clasificación de contenido 712 se configura para determinar varias informaciones de contenido a partir de los datos multimedia y puede configurarse también para determinar una clasificación de contenido.

El módulo de clasificación de contenido 712 se puede configurar para determinar una clasificación de contenido diferente para el video que tenga diferente tipo de contenido. La clasificación de contenido diferente puede dar como resultado diferentes parámetros usados en aspectos de la codificación de los datos multimedia, por ejemplo, la determinación de la tasa de bits (por ejemplo la asignación de bits) para la determinación de los parámetros de cuantificación, estimación de movimiento, escalabilidad, resistencia a los errores, mantenimiento de la calidad de datos multimedia óptima a través de los canales y para esquemas de cambio de canal rápido (por ejemplo, forzando cuadros I periódicamente para permitir un cambio de canal rápido). De acuerdo con un ejemplo, el codificador 228 se configura para determinar una utilización de tasa-distorsión (R-D) y asignaciones de tasas de bits en base a la clasificación del contenido. La determinación de la clasificación de contenido permite a los datos multimedia ser comprimidos hasta un nivel de calidad dado que corresponde a una tasa de bits deseada en base a una clasificación del contenido. También, y mediante la clasificación del contenido de los datos multimedia (por ejemplo, la determinación de la clasificación del contenido en base al sistema visual humano), la calidad perceptual resultante de los datos multimedia comunicados en una pantalla de un dispositivo receptor se hace dependiendo del contenido de video.

Como un ejemplo de procedimiento que realiza el módulo de clasificación del contenido 712 para clasificar el contenido, la FIGURA 9 muestra un proceso 900 que ilustra un proceso de ejemplo mediante el que el módulo de clasificación del contenido 712 puede operar. Como se muestra, el proceso 900 comienza en el bloque de entrada 902 en el que el módulo de clasificación del contenido 712 recibe datos multimedia en bruto y meta datos. El proceso 900 prosigue entonces al bloque 904 en el que el módulo de clasificación del contenido 712 determina la información espacial y la información temporal de los datos multimedia. En algunos aspectos, la información espacial y temporal se determina mediante enmascaramiento espacial y temporal (por ejemplo filtrado). La información espacial y temporal se puede determinar en base a los metadatos que incluyen datos de cambio de escena y suavizado de los vectores de movimiento (MV). El proceso 900 prosigue entonces al bloque 912 que realiza estimaciones de complejidad espacial, complejidad temporal y sensibilidad. El proceso 900 prosigue entonces al bloque 916 en el que el contenido de los datos multimedia se clasifica en base a los resultados de los datos espaciales, temporales y de sensibilidad determinados en los bloques 904 y 912. También en el bloque 916, se puede seleccionar una curva tasa-distorsión (R-D) particular y/o se pueden actualizar los datos de la curva R-D. El proceso 900 prosigue entonces al bloque de salida 918, en el que la salida puede incluir un mapa de complejidad-distorsión o un valor que indica la actividad espacial y temporal (por ejemplo una clasificación de contenido) y/o las curvas R-D seleccionadas. Volviendo a referirnos a la FIGURA 7, el módulo de clasificación del contenido 712 proporciona una salida a un módulo de asignación de bits de control de tasa 714, un módulo de intra modelo (distorsión) 716 y también para el módulo de instanciación de cuadros I 708, explicado anteriormente.

Información del contenido

El módulo de clasificación del contenido 712 se puede configurar para calcular una variedad información del contenido a partir de los datos multimedia, incluyendo una variedad de métricas relacionada con el contenido, incluyendo complejidad espacial, complejidad temporal, valores de relación de contraste, desviaciones estándar y métricas de diferencia de cuadros, descritas adicionalmente a continuación.

El módulo de clasificación del contenido 712 se puede configurar para determinar la complejidad espacial y la complejidad temporal de los datos multimedia y también para asociar un valor de textura a la complejidad espacial y un valor de movimiento a la complejidad temporal. El módulo de clasificación del contenido 712 recibe información de contenido preprocesada en relación al contenido de los datos multimedia que están siendo codificados desde el preprocesador 226 o, alternativamente, el preprocesador 226 puede configurarse para calcular la información del contenido. Como se ha descrito anteriormente, la información del contenido puede incluir, por ejemplo, uno o más valores D_{csat} , valores de relación de contraste, vectores de movimiento (MV) y suma de las diferencias absolutas (SAD).

En general, los datos multimedia incluyen una o más secuencias de imágenes o cuadros. Cada cuadro se puede repartir en bloques de píxeles para su procesamiento. La complejidad espacial es un término amplio que describe generalmente una medición del nivel de detalles espaciales dentro de un cuadro. Las escenas con áreas

principalmente planas o sin cambios o pocos cambios de iluminación y crominancia tendrán una baja complejidad espacial. La complejidad espacial se asocia con la textura de los datos de video. La complejidad espacial se basa en, en este aspecto, una métrica de sensibilidad visual humana denominada D_{csat} , que se calcula para cada bloque como una función de la frecuencia espacial local e iluminación ambiente. Los expertos en la técnica serán conscientes de las técnicas para el uso de patrones de frecuencia espacial e iluminación y características de contraste de las imágenes visuales para aprovecharse del sistema visual humano. Son conocidas un cierto número de métricas de sensibilidad para aprovecharse de las limitaciones de perspectiva del sistema visual humano y se podrían usar con el procedimiento descrito en el presente documento.

La complejidad temporal es un término amplio que se usa para describir en general una medición del nivel de movimiento en un dato multimedia como se referencia entre cuadros en una secuencia de cuadros. Las escenas (por ejemplo secuencias de cuadros de datos de video) con poco o ningún movimiento tienen una baja complejidad temporal. La complejidad temporal se puede calcular para cada macro bloque y se puede basar en el valor del D_{csat} , vectores de movimiento y la suma de las diferencias de píxeles absolutas entre un cuadro y otro cuadro (por ejemplo un cuadro de referencia).

La métrica de diferencia de cuadros da una medición de la diferencia entre dos cuadros consecutivos teniendo en cuenta la cantidad de movimiento (por ejemplo, el vector de movimiento o MV) junto con la energía residual representada como la suma de la diferencia absoluta (SAD) entre un predictor y el marco bloque actual. La diferencia de cuadros proporciona también una medición de la eficiencia de predicción bidireccional o unidireccional.

Un ejemplo de una métrica de diferencia de cuadros en base a la información de movimiento recibida desde un preprocesador que realiza potencialmente un desentrelazado compensado por el movimiento es como sigue. El desentrelazador realiza una estimación de movimiento bidireccional y por ello está disponible un vector de movimiento bidireccional y una información de SAD. Se puede deducir una diferencia de cuadros representada por SAD_MV para cada macro bloque como sigue:

$$SAD_MV = \log_{10} [SAD * \exp(-\min(1, MV))] \quad [29]$$

en la que $MV = \text{Raíz_cuadrada}(MV_x^2 + MV_y^2)$, $SAD = \min(SAD_N, SAD_P)$, en la que SAD_N es la SAD calculada a partir del cuadro de referencia hacia atrás y SAD_P es la SAD calculada a partir del cuadro de referencia hacia adelante.

Otro enfoque para la estimación de una diferencia de cuadros se describió anteriormente con referencia las Ecuaciones 6-8. Se puede calcular una relación de SAD (o relación de contraste) γ como se ha descrito anteriormente en la Ecuación 6. Se puede determinar también un histograma de luminancia de cada cuadro, siendo calculada la diferencia de histograma λ usando la Ecuación 7. La métrica de diferencia de cuadros D se puede calcular como se muestra en la Ecuación 8.

En un ejemplo ilustrativo, se utiliza una relación de contraste y una métrica de diferencia de cuadros de la siguiente manera para obtener una clasificación del contenido de video, que podría predecir fiablemente las características en una secuencia de video dada. Aunque se describe aquí como que sucede en el codificador 228, un preprocesador 226 se puede configurar también para determinar una clasificación del contenido (u otra información del contenido) y pasar la clasificación del contenido al codificador 228 por medio de metadatos. El proceso descrito en el ejemplo a continuación clasifica el contenido en ocho clases posibles, de modo similar a la clasificación obtenida a partir de un análisis en base a la curva R-D. El proceso de clasificación produce la salida de un valor en el intervalo entre 0 y 1 para cada supercuadro dependiendo de la complejidad de la escena y el número de sucesos de cambio de escena en ese supercuadro. El módulo de clasificación del contenido en el preprocesador puede ejecutar las siguientes etapas (1) – (5) para cada supertrama para obtener una métrica de clasificación del contenido a partir del contraste de cuadros y los valores de diferencia de cuadros.

1. Calcular el Contraste de Cuadro Medio y la Desviación de Contrastes de Cuadro a partir de los valores de contraste de macro bloques.

2. Normalizar el Contraste de Cuadros y los valores de Diferencia de Cuadros usando los valores obtenidos de simulaciones, que son 40 y 5 respectivamente.

3. Calcular una métrica de clasificación del contenido usando, por ejemplo, la ecuación generalizada:

$$CCMétrica = CCW1 * Media_Contraste_Cuadro_I + CCW2 * Media_Diferencia_Cuadro - CCW3 * Desviación_Contraste_I^2 * \exp(CCW4 * Desviación_Diferencia_Cuadro^2) \quad [30]$$

en la que CCW1, CCW2, CCW3 y CCW4 son factores de ponderación. En este ejemplo, los valores eligen para ser 0,2 para CCW1, 0,9 para CCW2, 0,1 para CCW3 y -0,00009 para CCW4.

4. Determinar el número de cambios de escena en el súper cuadro. En general, un súper cuadro se refiere a un grupo de imágenes o cuadros que se pueden visualizar en un periodo de tiempo particular. Típicamente, el periodo de tiempo es un segundo. En algunos aspectos, un súper cuadro comprende 30 cuadros (para video de 30

imágenes por segundo). En otros aspectos un súper cuadro comprende 24 cuadros (video de 24 imágenes por segundo). Dependiendo del número de cambios de escena, se puede obtener la ejecución de uno de los siguientes casos.

5 (a) No Cambios de Escena: cuando no hay cambios de escena en un súper cuadro, la métrica es completamente dependiente solo de los valores de diferencia de cuadros como se muestra en la siguiente ecuación:

$$\begin{aligned} \text{CCMétrica} = & (\text{CCW2} + (\text{CCW1}/2)) * \text{Media_Diferencia_Cuadro} - \\ & - (\text{CCW3} - (\text{CCW1}/2)) * 1 * \exp(\text{CCW4} * \text{Desviación_Diferencia_Cuadro}^2) \end{aligned} \quad [31]$$

10 (b) Un Único Cambio de Escena: Cuando hay un único cambio de escena observado en el súper cuadro, se usaría la ecuación por defecto para calcular la métrica, como se muestra a continuación:

$$\begin{aligned} \text{CCMétrica} = & \text{CCW1} * \text{Media_Contraste_Cuadro_I} + \text{CCW2} * \text{Media_Diferencia_Cuadro} - \\ & - \text{CCW3} * \text{Desviación_Contraste_I}^2 * \exp(\text{CCW4} * \text{Desviación_Diferencia_Cuadro}^2) \end{aligned} \quad [32]$$

15 (c) Dos Cambios de Escena: Cuando se observa que hay al menos 2 cambios de escena en el súper cuadro dado, se acuerda más peso para el último súper cuadro que para el primero dado que el primero se refrescará en cualquier caso por el último rápidamente, como se muestra en la siguiente ecuación:

$$\begin{aligned} \text{CCMétrica} = & 0.1 * \text{Media_Contraste_Cuadro_I1} + \text{CCW1} * \text{Media_Contraste_Cuadro_I2} + \\ & + (\text{CCW2}-0.1) * \text{Media_Diferencia_Cuadro} - \\ & - \text{CCW3} * \text{Desviación_Contraste_I1}^2 * \text{Desviación_Contraste_I2}^2 * \\ & * \exp(\text{CCW4} * \text{Desviación_Diferencia_Cuadro}^2) \end{aligned} \quad [33]$$

20 (d) Hay Tres o más Cambios de Escena: si en el súper cuadro dado se observa que tiene más de 3 cuadros I (digamos N), se da más peso al último cuadro I y a todos los demás cuadros I se les da un peso de 0,05, como se muestra la siguiente ecuación:

$$\begin{aligned} \text{CCMétrica} = & 0.05 * \text{Media_Contraste_Cuadro_I}_{(1,\dots,N-1)} + \text{CCW1} * \text{Media_Contraste_Cuadro_I}_{(N)} + \\ & + (\text{CCW2}-(0.05*(N-1))) * \text{Media_Diferencia_Cuadro} - \\ 25 & - \text{CCW3} * \text{Desviación_Contraste_I}_{(N)}^2 * \text{Desviación_Contraste_I}_{(1,\dots,N-1)}^2 * \\ & * \exp(\text{CCW4} * \text{Desviación_Diferencia_Cuadro}^2) \end{aligned} \quad [34]$$

5. Se puede usar una corrección para la métrica en el caso de escenas de movimiento bajo cuando la diferencia de cuadros media es menor de 0,05. Se añadiría un desplazamiento de 0,33 (DESPLAZAMIENTOCC) a la CCMétrica.

30 El módulo de clasificación del contenido 702 usa el valor de D_{csat} , vectores de movimiento y/o la suma de las diferencias absolutas para determinar un valor que indica una complejidad espacial para el macro bloque (o la cantidad designada de datos de video). La complejidad temporal se determina mediante la medición de la métrica de diferencia de cuadros (la diferencia entre dos cuadros consecutivos teniendo en cuenta la cantidad de movimiento, con vectores de movimiento y la suma de las diferencias absolutas entre los cuadros).

35 En algunos aspectos, el módulo de clasificación del contenido 712 se puede configurar para generar un mapa de ancho de banda. Por ejemplo, la generación del mapa de ancho de banda se puede realizar por el módulo de clasificación del contenido 712 si el preprocesador 226 no genera un mapa de ancho de banda.

Determinación de los valores de textura y movimiento

40 Para cada macro bloque en los datos multimedia, el módulo de clasificación del contenido 712 asocia un valor de textura con la complejidad espacial y un valor de movimiento con la complejidad temporal. El valor de textura se refiere a los valores de luminiscencia de los datos multimedia, en el que un valor de textura bajo indica pequeños cambios en los valores de luminiscencia de los píxeles contiguos a los datos y un valor de textura alto indica grandes cambios en los valores de luminiscencia de los píxeles contiguos a los datos. Una vez que se calculan los valores de textura y movimiento, el módulo de clasificación del contenido 712 determina una clasificación del contenido mediante la consideración tanto de la información de movimiento como de textura. El módulo de clasificación del contenido 712 asocia la textura para la clasificación de los datos de video con un valor de textura relativo, por ejemplo textura "Baja", textura "Media" o textura "Alta", que generalmente indican la complejidad de los valores de luminancia de los macro bloques. También, el módulo de clasificación del contenido 712 asocia el valor de movimiento calculado para la clasificación de los datos de video con un valor de movimiento relativo, por ejemplo, movimiento "Bajo", movimiento "Medio" o movimiento "Alto" que generalmente indican la cantidad de movimiento de

los macro bloques. En aspectos alternativos, se pueden usar menos o más categorías para el movimiento y la textura. A continuación, se puede determinar una métrica de clasificación del contenido mediante la consideración de los valores de textura y movimiento asociados.

5 La FIGURA 8 ilustra un ejemplo de un gráfico de clasificación que ilustra cómo se asocian los valores de textura y movimiento con una clasificación del contenido. Una persona experta la técnica estará familiarizado con muchas formas de implementar tal gráfico de clasificación, por ejemplo, en una tabla de búsqueda o una base de datos. El gráfico de clasificación se genera en base a evaluaciones predeterminadas del contenido del dato de video. Para determinar la clasificación del dato de video, se referencia un valor de textura de "Bajo", "Medio" o "Alto" (en el "eje x") de modo cruzado con un valor de movimiento de "Bajo", "Medio" o "Alto" (en el "eje y"). Se asigna a los datos de video una clasificación del contenido indicada por el bloque de intersección. Por ejemplo, un valor de textura "Alto" y un valor de movimiento "Medio" da como resultado una clasificación de siete (7). La FIGURA 8 ilustra varias combinaciones de valores de textura y movimiento relativos que se asocian con ocho clasificaciones de contenido diferentes, en este ejemplo. En otros aspectos, se pueden usar más o menos clasificaciones. Una descripción adicional de un aspecto ilustrativo de la clasificación del contenido se desvela en la Solicitud de Patente de Estados Unidos pendiente junto con la presente N° de Solicitud de Patente de Estados Unidos N° 11/300 73.577 titulada "Content Classification for Multimedia Processing" presentada el 10 de marzo de 2006, asignada al asignatario de la presente.

Asignación de bits de control de tasa

20 Como se ha descrito el presente documento, una clasificación del contenido de los datos multimedia se puede usar en los algoritmos de codificación para mejorar de modo efectivo la gestión de los bits mientras se mantiene una constante en la calidad de video perceptiva. Por ejemplo, la métrica de clasificación se puede usar en algoritmos para la detección de cambio de escena, la codificación del control de asignación de la tasa de bits y la conversión hacia arriba de la tasa de cuadros (FRUC). Los sistemas de compresor/descompresor (códec) y los algoritmos de procesamiento de señal digital se usan comúnmente en comunicaciones de datos de video y se puede configurar para conservar el ancho de banda, pero hay un compromiso entre la calidad y la conservación del ancho de banda. Los mejores códec proporcionan la mejor conservación del ancho de banda en tanto que producen la menor degradación de la calidad de video.

30 En un ejemplo ilustrativo, el módulo de asignación de bits de control de tasa 714 usa la clasificación del contenido para determinar una tasa de bits (por ejemplo el número de bits asignados para la codificación de los datos multimedia) y almacena la tasa de bits en la memoria para su uso por otros procesos y componentes del codificador 228. Una tasa de bits determinada a partir de la clasificación de los datos de video puede ayudar a conservar el ancho de banda en tanto proporciona datos multimedia con un nivel de calidad consistente. En un aspecto, se puede asociar una tasa de bits diferente a cada una de las ocho clasificaciones de contenido diferentes y a continuación se usa esa tasa de bits para codificar los datos multimedia. El efecto resultante es que aunque se asignan diferentes clasificaciones de contenido de datos multimedia a un número diferente de bits para codificación, la calidad percibida es similar o consistente cuando se ve en una pantalla.

40 En general, los datos multimedia con una clasificación de contenido más alta son indicativos de un nivel más alto de movimiento y/o textura y se les asignan más bits cuando se codifican. Los datos multimedia con una clasificación más baja (indicativa de una textura y movimiento menores) tienen menos bits asignados. Para datos multimedia de una clasificación de contenido particular, la tasa de bits se puede determinar en base a un nivel de calidad percibida objetivo seleccionada para la visión de los datos multimedia. La determinación de la calidad de los datos multimedia se puede determinar por humanos que ven y puntúan los datos multimedia. En algunos aspectos alternativos, las estimaciones de la calidad de datos multimedia se pueden realizar por sistemas de prueba automáticos que usan, por ejemplo, algoritmos de relación señal a ruido. En un aspecto, un conjunto de niveles de calidad estándar (por ejemplo cinco) y una tasa de bits correspondiente necesarios para conseguir ese nivel de calidad particular se determinan previamente para los datos multimedia de cada clasificación de contenido. Para determinar un conjunto de niveles de calidad, los datos multimedia de una clasificación de contenido particular se pueden evaluar mediante la generación de una Puntuación de Opinión Media (MOS) que proporciona la indicación numérica de la calidad visual percibida de los datos multimedia cuando se codifican usando una cierta tasa de bits. La MOS se puede expresar como un número único en el intervalo de 1 a 5 en el que 1 es la calidad más baja percibida y 5 la calidad más alta percibida. En otros aspectos, la MOS puede tener más de cinco o menos de cinco niveles de calidad, y se pueden usar diferentes descripciones de cada nivel de calidad.

55 La determinación de la calidad de datos multimedia se puede determinar por humanos que ven y puntúan los datos multimedia. En algunos aspectos alternativos, las estimaciones de la calidad de los datos multimedia se pueden realizar por sistemas de prueba automáticos que usan, por ejemplo, algoritmos de relación señal a ruido. En un aspecto, un conjunto de niveles de calidad estándar (por ejemplo cinco) y una tasa de bits correspondiente necesarios para conseguir ese nivel de calidad particular se determinan previamente para los datos multimedia de cada clasificación de contenido.

60 Conociendo la relación entre el nivel de calidad percibida visualmente y la tasa de bits para los datos multimedia de una cierta clasificación de contenido se puede determinar mediante la selección de un nivel de calidad objetivo (por

ejemplo deseado). El nivel de calidad objetivo usado para determinar la tasa de bits se puede seleccionar previamente, seleccionar por un usuario, seleccionar a través de un proceso automático o un proceso semiautomático se requiera una entrada de un usuario o desde otro proceso o ser seleccionada dinámicamente por el dispositivo o sistema de codificación en base a criterios predeterminados. Un nivel de calidad objetivo se puede seleccionar en base a, por ejemplo, el tipo de aplicación de codificación o el tipo de dispositivo cliente que recibirá los datos multimedia.

En el ejemplo ilustrado en la FIGURA 7, el módulo de asignación de bits de control de tasa 714 recibe tanto datos desde el módulo de clasificación del contenido 712 como de los metadatos directamente del preprocesador 226. El módulo de asignación de bits de control de tasa 714 reside en la parte del primer paso del codificador 228 y un módulo de ajuste fino del control de tasa 738 reside en la parte del segundo paso 706. Este aspecto del control de la tasa en dos pasos se configura de modo que el primer paso (módulo de asignación de los bits de control de tasa 714) realiza un asignación de bits adaptativa al contexto con una anticipación del súper cuadro (por ejemplo, un objetivo a largo plazo de las tasas de bits promedio de 256 kbps) y limita la tasa de pico y el segundo paso (módulo de ajuste fino del control de tasa 738) refina los resultados del primer paso en una escalabilidad en dos capas que realiza la adaptación de la tasa. El control de la tasa funciona en cuatro niveles: (1) nivel de GOP - controla la distribución de bits de los cuadros I, P, B y F para que no sean uniformes dentro de un GOP; (2) nivel de súper cuadro - controla los límites firmes en el tamaño de súper cuadro máximo; (3) nivel de cuadro - controla los requisitos de bits de acuerdo con la complejidad espacial y temporal de los cuadros de datos multimedia, que se basan en la información del contenido (por ejemplo una clasificación del contenido) y (4) nivel de macro bloque - controla la asignación de bits de los macro bloques en base a los mapas de complejidad espacial y temporal, que se basan en la información del contenido (por ejemplo una clasificación del contenido).

Un diagrama de flujo de ejemplo del funcionamiento del módulo de control de tasa 714 se ilustra en la FIGURA 10. Como se muestra en la FIGURA 10, el proceso 1000 comienza en un bloque de entrada 1002. El módulo de control de tasa 714 recibe varias entradas, ninguna de las cuales se ilustra necesariamente por la FIGURA 7. Por ejemplo, la información de entrada puede incluir metadatos desde el preprocesador 226, una tasa de bits objetivo, un tamaño de memoria intermedia del codificador (o, como equivalente, el tiempo de retardo máximo para el control de tasa), un retardo de control de tasa inicial y una información de tasa de cuadro. La información de entrada adicional puede incluir entradas en el nivel de grupo de imágenes (GOP), incluyendo, por ejemplo, un tamaño del súper cuadro máximo, longitud y distribución de cuadros P/B del GOP (incluyendo información de cambio de escena), disposición de las capas base y de mejora deseada, una métrica de complejidad-distorsión para imágenes en el GOP para unos 30 cuadros futuros. Otra información de entrada incluye entradas en el nivel de imagen, incluyendo el mapa de complejidad-distorsión para la imagen actual (recibida desde el módulo de clasificación del contenido 712), parámetros de cuantificación (QP) y división en bits de los 30 últimos cuadros (ajustados a través de una ventana deslizante). Finalmente, la información de entrada en el nivel de macro bloque (MB) incluye, por ejemplo, la diferencia absoluta media (MAD) o macro bloques no localizados conjuntamente (MB) en una imagen de referencia y un patrón de bloques codificados (CBP) de macro bloques después de la cuantificación (si se saltarán o no).

Después de las entradas en el bloque 1002, el proceso 1000 prosigue en el bloque 1004 para la inicialización de la codificación del flujo de bits. Simultáneamente, se realiza una inicialización de la memoria intermedia 1006. A continuación, se inicializa un GOP como se muestra en el bloque 1008, con la asignación de bits del GOP 1010 recibida como parte de la inicialización. Después de la inicialización del GOP, el flujo prosigue en el bloque 1012, en el que se inicializa una porción. Esta inicialización incluye una actualización de los bits de cabecera como se muestra por el bloque 1014. Después de que se realicen las inicializaciones del bloque 1004, 1008 y 1012, se lleva a cabo el control de tasa (RC) para una unidad básica o macro bloque (MB) como se muestra por el bloque 1016. Como parte de la determinación del control de tasa de un macro bloque en el bloque 1016, se reciben entradas por medio de las interfaces en el codificador 228. Estas entradas pueden incluir la asignación de bits de macro bloque (MB) 1018, una actualización de los parámetros del modelo cuadrático 1020 y una actualización de la desviación absoluta media a partir de los parámetros medios ("MAD", una estimación robusta de la dispersión) 1022. A continuación el proceso 1000 prosigue en el bloque 1024 para la ejecución de operaciones después de la codificación de una imagen 1024. Este procedimiento incluye la recepción de una actualización de los parámetros de la memoria intermedia como se muestra por el bloque 1026. El proceso 1000 prosigue entonces al bloque de salida 1028 en donde el módulo de control de tasa 714 produce la salida de los parámetros de cuantificación QP para cada macro bloque MB para ser usada por un módulo de decisión de modo 715 como se muestra en la FIGURA 7.

Estimación de movimiento

El módulo de estimación del movimiento 730 recibe entradas de metadatos y video bruto del preprocesador 226 y proporciona una salida que puede incluir el tamaño del bloque, métrica de distorsión de los vectores movimiento e identificadores del cuadro de referencia a un módulo de decisión de modo 715. La FIGURA 11 ilustra una operación de ejemplo del módulo de estimación del movimiento 720. Como se muestra, el proceso 1100 comienza con una entrada 1102. En el nivel de cuadro, el módulo 720 recibe la entrada de la ID del cuadro de referencia y los vectores de movimiento. En el nivel de macro bloque, la entrada 1102 incluye píxeles de entrada y píxeles del cuadro de referencia. El proceso 1100 continúa en la etapa 1104 en la que se realizan la estimación de movimiento de color (ME) y la previsión del vector de movimiento. Para llevar a cabo estos procesos, se reciben varias entradas incluyendo vectores de movimiento MPEG-2, y vectores de movimiento de luma MV 1106, suavizado del vector de

movimiento 1108 y vectores de movimiento no causales 1110. A continuación, el proceso 1100 prosigue en el bloque 1112 en donde se realiza un algoritmo o metodología de búsqueda del vector de movimiento, tal como los procedimientos de búsqueda hexagonal o en rombo. Las entradas al proceso en el bloque 1112 pueden incluir la suma de la diferencia absoluta (SAD), la suma de la diferencia cuadrada (SSD) y/o otras métricas como se muestra por el bloque 1114. Una vez que se realiza la búsqueda del vector de movimiento, el proceso 1100 prosigue en el bloque de finalización 1116, en donde se realiza el procesamiento de la finalización. El proceso 1100 finaliza entonces en el bloque de salida 1118, que conduce a una salida del tamaño de bloque, vector de movimiento (MV), métrica de distorsión e identificadores del Cuadro de Referencia.

Escalabilidad de R-D para la capa base y de mejora

La FIGURA 13 ilustra un diagrama de flujo de ejemplo de un proceso de escalabilidad 1300 que se puede realizar por el módulo de R-D de escalabilidad 718. El proceso 1300 comienza en el bloque de arranque 1302 y prosigue en el bloque 1304 en el que el módulo de R-D de escalabilidad 718 recibe una entrada desde el módulo de estimación de movimiento 720 realiza una estimación del movimiento. La estimación del movimiento confía en la entrada de cuadros de referencia de la capa base, cuadros de referencia de la capa de mejora y el cuadro original a ser codificado como se indica por el bloque 1306. Tal información se puede calcular por el divisor del GOP y comunicarse al módulo de R-D de escalabilidad 718 por medio, por ejemplo, de metadatos. El proceso 1300 prosigue en el bloque 1308 para determinar la información de escalabilidad de los datos de la capa base y de los datos de la capa de mejora. La codificación de la capa base se realiza a continuación como se muestra en el bloque 1310, seguida por la codificación de la capa de mejora en el bloque 1312. La codificación de la capa de mejora puede usar los resultados de codificación de la capa base para tener una predicción entre capas como entrada, como se ilustra por el bloque 1314, por ello se realiza temporalmente después de la codificación de la capa base. Esto se describe adicionalmente en la Solicitud de Patente de Estados Unidos pendiente junto con la presente N° [expediente/referencia de representante n° 050078] titulada "SCALABLE VIDEO CODING WITH TWO LAYER ENCODING AND SINGLE LAYER DECODING". Después de que está completa la codificación, el proceso 1300 finaliza en el bloque 1316.

Ordenación Porción/Macro bloque

La parte del primer paso 302 incluye también un módulo de ordenación de porción/macro bloque 722, que recibe una entrada desde un módulo de resistencia frente a errores 740 en la parte del segundo paso y proporciona una información de alineación de porciones al módulo de decisión de modo 715. Las porciones son recortes de datos de video codificado decodificables independientemente (decodificación de entropía). Las unidades de acceso (AU) son cuadros de video codificados que comprende cada uno un conjunto de unidades NAL que contiene siempre exactamente una imagen codificada primaria. Además de la imagen codificada primaria, una unidad de acceso puede contener también una o más imágenes codificadas redundantes u otras unidades NAL que no contienen porciones de divisiones de datos en porciones de una imagen codificada. La decodificación de una unidad de acceso siempre da como resultado una imagen decodificada.

Los cuadros puede ser bloques multiplexados por división de tiempo o paquetes de capa física (denominados cápsulas TDM) que ofrecen la más elevada diversidad de tiempos. Un súper cuadro corresponde a una unidad de tiempo (por ejemplo, 1 seg.) y contiene cuatro cuadros. La alineación de los límites de porción y AU con los límites de cuadro en el dominio del tiempo da como resultado la separación y localización más eficiente de los datos corruptos. Durante un desvanecimiento profundo, la mayor parte de los datos contiguos en una cápsula TDM está afectada por errores. Debido a la diversidad de tiempos, las cápsulas TDM restantes tienen una alta probabilidad de estar intactas. Los datos no corruptos se pueden utilizar para recuperar y ocultar los datos perdidos de la cápsula TDM afectada. Una lógica similar se aplica al multiplexado en el dominio de la frecuencia (FDM) en el que la diversidad de frecuencia se consigue a través de la separación de subportadoras de frecuencia que modulan símbolos de datos. Adicionalmente, se aplica una lógica similar a la diversidad espacial (por medio de la separación en antenas transmisoras y receptoras) y otras formas aplicadas frecuentemente en redes inalámbricas.

Para alinear las porciones y las AU a los cuadros, la creación del bloque de código (FEC) exterior y el encapsulado de la capa MAC se debería asimismo alinear. La FIGURA 20 ilustra la organización de los datos de video codificados o de un flujo de bits de video en porciones y AU. El video codificado puede estar formado por uno o más flujos de bits, por ejemplo el flujo de bits de la capa base y el flujo de bits de la capa de mejora cuando se aplica la codificación de video en capas.

El flujo de bits de video comprende las AU como se ilustra en la FIGURA 20 por el Cuadro 1' 2005, el Cuadro 3' 2010 y el Cuadro M' 2015. Las AU se componen de porciones de datos, como se ilustra por la Porción 1 2020, Porción 2 2025 y Porción N 2030. Cada comienzo de porción se identifica por un código de comienzo y proporciona una adaptación a la red. En general, el cuadro I o las AU intra-codificadas son grandes, seguidas de cuadros P o cuadros predichos hacia adelante seguidos por cuadros B. La codificación de una AU en múltiples porciones incurre en un coste de sobrecarga significativa en términos de la tasa de bits codificada debido a que la predicción espacial a través de las porciones está restringida y las cabeceras de porción contribuyen también a la sobrecarga. Debido a que los límites de las porciones son puntos de resincronización, la restricción de los paquetes de la capa física contiguos a porciones controla los errores dado que cuando se corrompe un PLP, el error está confinado a la porción

en el PLP mientras que si el PLP contenía múltiples porciones o partes de múltiples porciones, el error impactará en todas las porciones o partes de porciones en el PLP.

5 Dado que los cuadros I son típicamente grandes, por ejemplo, del orden de decenas de kbits, la sobrecarga debida a múltiples porciones no es una gran proporción del tamaño del cuadro I total o de la tasa de bits total. También, teniendo más porciones en una AU intra-codificada permite una resincronización mejor y más frecuente y una ocultación del error espacial más eficiente. También, los cuadros I transportan la información más importante en el flujo de bits de video dado que los cuadros P y B se predicen fuera de los cuadros I. Los cuadros I sirven también como puntos de acceso aleatorio para la adquisición de un canal.

10 Con referencia ahora a la FIGURA 21, la alineación cuidadosa de los cuadros I a los límites de cuadros y las porciones con una AU I asimismo a los límites de cuadro, permite un control de error más eficiente, una protección de error (dado que si una porción que pertenecía al Cuadro 1 2105 se pierde, las porciones que pertenecen al Cuadro 2 2110 tienen una alta probabilidad de estar intactas debido a que el Cuadro 2 2110 tiene una separación de tiempo significativa desde el Cuadro 1 2105) la recuperación de errores se puede realizar por medio de la resincronización y ocultación de error.

15 Debido a que los cuadros P son típicamente de un tamaño del orden de unos pocos kbits, la alineación de las porciones a un cuadro P y un número entero de cuadros P a los límites del cuadro permite una resistencia frente a errores sin una pérdida perjudicial de eficiencia (por similares razones a la de los cuadros I). Se puede emplear en tales aspectos la ocultación del error temporal. Alternativamente, dispersando cuadros P consecutivos de modo que lleguen en diferentes cuadros proporciona una diversidad de tiempo añadida entre los cuadros P, que puede ser debida a que la ocultación temporal se basa en vectores de movimiento y datos de los cuadros I o P previamente reconstruidos. Los cuadros B pueden ser extremadamente pequeños (centenares de bits) a moderadamente grandes (unos pocos miles de bits). De ahí que la alineación de un número entero de cuadros B a los límites del cuadro es deseable para conseguir una resistencia frente a errores sin una pérdida perjudicial de eficiencia.

Módulo de decisión de modo

25 La FIGURA 2 ilustra algunos ejemplos de la operación del módulo de decisión de modo 715. Como se muestra, el proceso 1200 comienza en un bloque de entrada 1202. En un ejemplo ilustrativo, las varias entradas de información al módulo de decisión de modo 715 incluyen el tipo de porción, el coste 4x4 intra, el coste 16x16 intra el coste 8x8 intraUV, modo 16x16 intraY, modo intraUV, datos del vector de movimiento (MVD), parámetros de cuantificación (QP), SpPredMB4x4Y, SpPredMB16x16Y, SpPredMB8x8U, SpPredMB8x8V, marcador de distorsión de tasa, píxeles YMB en bruto, píxeles UMB en bruto y píxeles VMB en bruto. El proceso 1200 prosigue entonces al bloque 1204 que codifica la inicialización, que se puede iniciar mediante una señal de entrada o una interfaz que dirige la inicialización del codificador como se indica por el bloque 1206. La inicialización puede incluir el establecimiento de los modos permitidos (incluyendo salto, director), el ajuste de las ponderaciones de modo (si se requiere, por omisión serán iguales ponderaciones para todos los modos) y el establecimiento de las memorias intermedias. Tras la inicialización, el proceso 1200 prosigue en el bloque 1208 en donde se realiza el procesamiento principal para la decisión de modo, incluyendo: el cálculo del coste del modo de macro bloque (MB) para cada modo permitido, la ponderación de cada coste del modo de MB con un factor de ponderación y la selección de un modo de coste del modo de MB mínimo. Las entradas involucradas en estas operaciones incluyen la estimación del movimiento (por ejemplo MVD y predicción) y predicción espacial (por ejemplo costes intra y previsiones) como se ilustra por los bloques 1210 y 1212. La interfaz con el módulo de decisión de modo 715 es una codificación de entropía en el bloque 1214 que, entre otras cosas, mejora la tasa de compresión. El proceso 1200 prosigue con el bloque 1216 en donde las memorias intermedias se adaptan para pasar información al codificador de la parte del segundo paso 706. Finalmente, el proceso 1200 prosigue en el bloque 1218 en donde se puede comunicar el modo de codificación "mejor" al codificador de la parte del segundo paso 706.

Codificador de la parte del segundo paso

Con referencia de nuevo a la FIGURA 7, la parte del segundo paso 706 del codificador 228 incluye un módulo codificador del segundo paso 232 para la realización del segundo paso de la codificación. El codificador del segundo paso 232 recibe la salida desde el módulo de decisión de modo 715. El codificador del segundo paso 232 incluye un módulo de Cuantificación del MC/Transformada 726 y un codificador de Zig Zag (ZZ)/Entropía 728. Los resultados del codificador de segundo paso 232 se sacan a un módulo de escalabilidad 730 y un módulo de empaquetado del flujo de bits 731, que produce la salida hacia una capa base y de mejora codificadas para la transmisión por el transcodificador 200 por medio de una capa de sincronización 240 (ilustrada en la FIGURA 2). Como se muestra, en la FIGURA 2, se indica que las capas base y de mejora del codificador de segundo paso 232 y del recodificador 234 se montan juntas mediante la capa de sincronización 240 en un PES en paquetes 242 que incluye las capas base y de mejora, un PES de datos 244 (por ejemplo, CC y otros datos de texto), y un PES de audio 246. Se indica que el codificador de audio 236 recibe información de audio del codificador 218 y, a su vez, codifica la información y produce la salida de la información codificada 238 hacia la capa de sincronismo 240.

Recodificador

Con referencia de nuevo la FIGURA 7, la parte del segundo paso del codificador 706 incluye también un recodificador 234, que corresponde al decodificador 234 de la FIGURA 2, el recodificador 234 recibe también la salida de la parte del primer paso 702 e incluye unas partes del Cuantificación del MC/Transformada 726 y codificación de ZZ/Entropía 728. Adicionalmente, el módulo de escalabilidad 730 produce la salida hacia el recodificador 234. El recodificador 234 produce la salida de la capa base y de mejora resultantes desde el recodificador hacia el módulo de empaquetado del flujo de bits 731 para la transmisión a un sincronizador (por ejemplo la capa de sincronización 240 mostrada en la FIGURA 2). El codificador 228 del ejemplo en la FIGURA 7 incluye también un módulo de ajuste fino del control de tasas 738 que proporciona una realimentación del empaquetado del flujo de bits tanto para el módulo de Cuantificación del MC/Transformada 234 en el codificador del segundo paso 232 como para el módulo de ZZ/Entropía 736 en el recodificador 234 para ayudar al ajuste de la codificación del segundo paso (por ejemplo, para aumentar la eficiencia de la compresión).

Módulo de resistencia a los errores

El ejemplo de codificador 228 ilustrado en la FIGURA 7 incluye también un módulo de resistencia frente a errores 740 en la parte del segundo paso 706. El módulo de resistencia frente a errores 740 comunica con el módulo de empaquetado del flujo de bits 700 31 y el módulo de ordenación porción/MB 722. El módulo de resistencia frente a errores 740 recibe metadatos desde el preprocesador 228 y selecciona un esquema de resistencia frente a errores, por ejemplo, la alineación de la porción y las unidades de acceso a los límites del cuadro, jerarquía predictiva y intra refresco adaptativo. La selección del esquema de resistencia frente a errores se puede basar en la información recibida en los metadatos o en la información comunicada al módulo de resistencia frente a errores desde el módulo de empaquetado del flujo de bits 731 y el módulo de ordenación porción/MB 722. El módulo de resistencia frente a errores 740 proporciona información al módulo de ordenación porción/macro bloques (MB) en la parte del primer paso 702 para implementar los procesos de resistencia frente a errores seleccionados. Las transmisiones de video a través de entornos propensos a errores pueden emplear estrategias y algoritmos de resistencia frente a errores que puede dar como resultado la presentación de unos datos más claros y con menos errores a un usuario de la visión. La descripción de la resistencia frente a errores a continuación puede aplicarse a cualquier aplicación individual o combinación de las existentes o futuras, capas de transporte y físicas u otras tecnologías. Los algoritmos de robustez frente a errores integran una comprensión de las propiedades de susceptibilidad frente a errores y las capacidades de protección de errores entre las capas OSI en conjunto con las propiedades deseables del sistema de comunicación tal como baja latencia y alto rendimiento. El procesamiento de la resistencia frente a errores se puede basar en la información del contenido de los datos multimedia, por ejemplo, en la clasificación del contenido de los datos multimedia. Una de las ventajas principales es la recuperabilidad de los errores de desvanecimiento y de multi-trayecto del canal. Los enfoques de resistencia frente a errores descritos a continuación pertenecen específicamente a procesos que se pueden incorporar en el codificador 228 (por ejemplo en particular en el módulo de resistencia frente a errores 740 y en el módulo de ordenación porción/MB 722) y se puede extender genéricamente a la comunicación de datos en entornos proclives a errores.

Resistencia frente a errores

Para un sistema de compresión híbrido basado en la predicción, los cuadros intra-codificados se codifican independientemente sin ninguna predicción temporal. Los cuadros inter-codificados se pueden predecir temporalmente a partir de cuadros pasados (cuadros P) y cuadros futuros (cuadros B). El mejor predictor se puede identificar a través de un proceso de búsqueda en el cuadro de referencia (uno o más) y se usa una medida de distorsión tal como la SAD para identificar el mejor ajuste. La región codificada predictiva del cuadro actual puede ser un bloque de tamaño y forma variables (16x16, 32x32, 8x4, etc.) o un grupo de píxeles identificados como un objeto a través, por ejemplo, de la segmentación.

La predicción temporal se extiende típicamente sobre muchos cuadros (por ejemplo 10 a decenas de cuadros) y se finaliza cuando un cuadro se codifica como un cuadro I, estando típicamente el GOP definido por la frecuencia de cuadros I. Para una máxima eficiencia de codificación, un GOP es una escena, por ejemplo, los límites del GOP se alinean con los límites de la escena y los cuadros de cambio de escena se codifican como cuadros I. Las secuencias de movimiento bajo se componen de un fondo relativamente estático y el movimiento se limita generalmente a un objeto en primer plano. Los ejemplos de contenido de tales secuencias de movimiento bajo incluyen programas de noticias y predicciones del tiempo en donde no más del 30% del contenido más visto es de esta naturaleza. En secuencias de movimiento bajo, la mayoría de las regiones se intra-codifican y los cuadros predichos se refieren hacia atrás al cuadro I a través de los cuadros predichos intermedios.

Con referencia a la FIGURA 22, el bloque intra-codificado 2205 en el cuadro I es el predictor para el bloque intra-codificado 2210 en el cuadro codificado (o AU) P1. En este ejemplo, la región de estos bloques es una parte fija del fondo. A través de la predicción temporal consecutiva, la sensibilidad del bloque intra-codificado 2205 a los errores crece hasta que es un predictor "bueno" que también implica que su "importancia" es más alta. Adicionalmente, el bloque intra-codificado 2205, en virtud de su cadena de predicción temporal denominada la cadena de predicción, persiste más tiempo en la pantalla (durante la duración de la escena en el ejemplo en la figura).

La jerarquía de predicción se define como el árbol de bloques creados en base a su nivel de importancia o medición de persistencia con el padre en la parte superior (bloque intra-codificado 2205) y los hijos en la parte inferior. Nótese que el bloque inter-codificado 2215 en P1 está en el segundo nivel de jerarquía y así sucesivamente. Las hojas son bloques que finalizan una cadena de predicción.

- 5 La jerarquía de predicción se puede crear para secuencias de video independientemente del tipo de contenido (asimismo para música y deportes y no solamente noticias) y se puede aplicar a la comprensión de video (y datos) basada en predicción en general (esto se aplica a todos aspectos descritos en esta solicitud). Una vez que se establece la jerarquía de predicción, los algoritmos de resistencia frente a errores tal como el intra refresco adaptativo, descrito a continuación, se puede aplicar más efectivamente. La medida de la importancia se puede basar en la posibilidad de recuperar un bloque dado de los errores tal como a través de operaciones de ocultación y aplicación de intra refresco adaptativo para mejorar la resistencia de los flujos de bits codificados a los errores. Una estimación de la medición de importancia se puede basar en el número de veces que se usa un bloque como un previsor también denominada como una métrica de persistencia. La métrica de persistencia se usa también para mejorar la eficiencia de codificación mediante la detención de la propagación del error de predicción. La métrica de persistencia también aumenta la asignación de bits para los bloques con importancia más alta.

Intra refresco adaptativo

- El intra refresco adaptativo es una técnica de resistencia a los errores que se puede basar en la información del contenido de los datos multimedia. En un proceso de intra refresco, algunos MB se intra-codifican incluso aunque la optimización R-D estándar dictaría que deberían ser MB inter-codificados. El AIR emplea un intra refresco ponderado por movimiento para introducir MB intra-codificados en cuadros P o B. Estos MB intra-codificados, contenidos en la capa base, se pueden codificar o bien con un QP_b o bien con un QP_e . Si se usa el QP_e para la capa base, entonces no se necesita ningún refinamiento en la capa de mejora. Si se usa QP_b para la capa base, entonces puede ser apropiado un refinamiento, en caso contrario en la capa de mejora, la caída de calidad será notable. Dado que la inter-codificación es más eficiente que la intra-codificación en el sentido de la eficiencia de codificación, estos refinamientos en la capa de mejora serán inter-codificados. En esta forma, los coeficientes de la capa base no se usarán para la capa de mejora, y se mejora de calidad en la capa de mejora sin introducir nuevas operaciones.

- En algunos aspectos, el intra refresco adaptativo se puede basar en la información del contenido de los datos multimedia (por ejemplo una clasificación del contenido) en lugar de, o además de, en base a la ponderación del movimiento. Por ejemplo, si la clasificación del contenido es relativamente alta (por ejemplo, escenas que tengan una complejidad espacial y temporal alta) el intra refresco adaptativo puede introducir relativamente más MB intra-codificados en cuadros P o B. Alternativamente, si la clasificación del contenido era relativamente baja (indicando una escena menos dinámica con baja complejidad espacial y/o temporal), el intra refresco adaptativo puede introducir menos MB intra-codificados en los cuadros P y B. Tales métricas y procedimientos para la mejora de la resistencia a los errores se pueden aplicar no solamente en el contexto de comunicaciones multimedia inalámbricas sino hacia la compresión de datos y el procesamiento multimedia en general (por ejemplo en la generación de gráficos).

Cuadro de cambio de canal

- Un cuadro de cambio de canal (CSF) se define en el presente documento como un término amplio que describe un cuadro de acceso aleatorio insertado en una localización apropiada en una transmisión continua de emisión para una adquisición de canal rápida y por ello un cambio de canal rápido entre transmisiones continuas en un múltiplex de emisión. Los cuadros de cambio de canal aumentan también la robustez frente a errores, dado que proporcionan datos redundantes que se pueden usar si el cuadro primario se transmite con un error. Un cuadro I o un cuadro I progresivo tal como el cuadro de refresco de decodificador progresivo en H.264 sirve típicamente como un punto de acceso aleatorio. Sin embargo, frecuentes cuadros I (o cortos GOP, más cortos que las duraciones de las escenas) da como resultado una reducción significativa en la eficiencia de la compresión. Dado que los bloques de intra-codificación se pueden requerir para resistencia frente a errores, el acceso aleatorio y la resistencia a los errores se puede combinar de modo efectivo a través de la jerarquía de predicción para mejorar la eficiencia de codificación mientras se mantiene la robustez frente a errores.

- La mejora del cambio de acceso aleatorio y robustez frente a errores se puede conseguir en concierto y se puede basar en la información del contenido tal como una clasificación del contenido. Para secuencias de movimiento bajo, las cadenas de predicción son largas y una parte significativa de la información requerida para reconstruir un súper cuadro o escena está contenida en el cuadro I que apareció al comienzo de la escena. Los errores de canal tienden a ser a ráfagas y cuando un desvanecimiento aparece y falla el FEC y la codificación del canal, hay un error residual fuerte que falla en su ocultación. Esto es particularmente severo para secuencias de poco movimiento (y por ello de una baja tasa de bits) dado que la cantidad de datos codificados no es suficientemente significativa para proporcionar una buena diversidad de tiempos dentro del flujo de bits de video y debido a que éstas son secuencias altamente comprensibles que generan cada bit útil para la reconstrucción. Las secuencias de movimiento alto son más robustas frente a errores debido a la naturaleza del contenido —más información nueva en cada cuadro aumenta el número de bloques intra-codificados que se pueden decodificar independientemente y son inherentemente más resistentes frente a errores—. El intra refresco adaptativo en base a jerarquía de predicción

5 consigue un alto rendimiento para secuencias de movimiento alto y la mejora del rendimiento no es significativa para secuencias de movimiento bajo. Por ello, un cuadro de cambio de canal que contiene la mayor parte del cuadro I es una buena fuente de diversidad para secuencias de movimiento bajo. Cuando aparece un error en un súper cuadro, la decodificación de los cuadros consecutivos comienza desde el CSF que recupera la información perdida debido a la predicción y se consigue una resistencia frente a los errores.

10 En el caso de secuencias de alto movimiento tal como secuencias que tengan una clasificación de contenido relativamente alta (por ejemplo 6-8), el CSF puede consistir en bloques que persisten en el SF —aquellos que son buenos previsores—. Todas las otras regiones del CSF no han de ser codificadas dado que son bloques que tienen cadenas de predicción cortas que implica que se finalizan con intra bloques. Por ello, el CSF aun sirve para recuperarse de la información perdida debido a la predicción cuando ocurre un error. Los CSF para secuencias de movimiento bajo están a la altura del tamaño de los cuadros I y se pueden codificar con una tasa de bits más baja por medio de una cuantificación más fuerte, con la que los CSF para secuencias de movimiento alto son mucho más pequeños que los correspondientes cuadros I.

15 La resistencia frente a errores en base a la jerarquía de predicción puede funcionar bien con escalabilidad y puede conseguir una codificación en capas altamente eficiente. La escalabilidad para soportar la modulación jerárquica en tecnologías de capas físicas, puede requerir la partición de los datos del flujo de bits de video con relaciones de ancho de banda específicas. Estas pueden no ser siempre las relaciones ideales para una escalabilidad óptima (por ejemplo, con la menor sobrecarga). En algunos aspectos, se usa una escalabilidad de 2 capas con una relación de ancho de banda 1:1. La división del flujo de bits de video en 2 capas de igual tamaño puede no ser tan eficiente para secuencias de movimiento bajo. Para secuencias de movimiento bajo, la capa base que contiene toda la información de cabecera y metadatos es más grande que la capa de mejora. Sin embargo, dado que los CSF para secuencias de movimiento bajo son más grandes, ellos encajan claramente en el ancho de banda restante en la capa de mejora.

20 Las secuencias de movimiento alto tienen suficiente información residual que se puede conseguir una partición de los datos en 1:1 con la menor sobrecarga. Adicionalmente, un cuadro de cambio de canal para tales secuencias es mucho más pequeño para secuencias de movimiento alto. Por ello, la resistencia frente a errores en base a la jerarquía de predicción puede funcionar bien asimismo con la escalabilidad para secuencias de movimiento alto. La extensión de los conceptos explicados anteriormente para clips de movimiento moderado es posible en base a las descripciones de estos algoritmos y los conceptos propuestos se aplican para la codificación de video en general.

Multiplexor

30 En algunos aspectos del codificador, se puede usar un multiplexor para la codificación de múltiples transmisiones continuas multimedia producidas por el decodificador y usarse para preparar los bits codificados para emisión. Por ejemplo, en el aspecto ilustrativo del codificador 228 mostrado en la FIGURA 2, la capa de sincronismo 240 comprende un multiplexor. El multiplexor se puede implementar para proporcionar el control de asignación de tasa de bits. La complejidad estimada se puede proporcionar al multiplexor, que puede entonces asignar el ancho de banda disponible para una colección de canales de video multiplexados de acuerdo con la complejidad de codificación anticipada por esos canales de video, lo que permite entonces que permanezca la calidad de un canal particular relativamente constante incluso si el ancho de banda para la colección de las transmisiones continuas de video multiplexadas es relativamente constante. Esto proporciona un canal con una colección de canales para tener una tasa de bits variable y una calidad visual relativamente constante, más que una tasa de bits relativamente constante y una calidad visual variable.

45 La FIGURA 18 es un diagrama de bloques que ilustra un sistema de codificación múltiple de transmisiones continuas o canales multimedia 1802. Las transmisiones continuas multimedia 1802 se codifican mediante los codificadores respectivos 1804, que están en comunicación con un multiplexor (MUX) 1806, que a su vez está en comunicación con un medio de transmisión 1808. Por ejemplo, las transmisiones continuas multimedia 1802 pueden corresponder a varios canales de contenido, tal como canales de noticias, canales de deporte, canales de películas y otros similares. Los codificadores 1804 codifican las transmisiones continuas multimedia 1802 en el formato de codificación especificado por el sistema. Mientras que se describe en el contexto de la codificación de transmisiones continuas de video, los principios y ventajas de las técnicas desveladas son aplicables en general a las transmisiones continuas multimedia que incluyen, por ejemplo, transmisiones de audio. Las transmisiones continuas multimedia codificadas se proporcionan a un multiplexor 1806, que combina las varias transmisiones continuas multimedia codificadas y envía la transmisión continua combinada al medio de transmisión 1808 para su transmisión.

50 El medio de transmisión 1808 puede corresponder a una variedad de medios, tales como, pero sin limitarse a, comunicación por satélite digital, tal como DirecTV®, cable digital, comunicaciones de Internet cableadas e inalámbricas, redes ópticas, redes de teléfonos celulares y otros similares. El medio de transmisión 1808 puede incluir, por ejemplo, la modulación en una frecuencia de radio (RF). Típicamente, debido a las limitaciones espectrales y otras similares, el medio de transmisión tiene un ancho de banda limitado y los datos desde el multiplexor 1806 al medio de transmisión se mantienen con una tasa de bits (CBR) relativamente constante.

55 En sistemas convencionales, el uso de una tasa de bits constante (CBR) a la salida del multiplexor 1806 puede requerir que las transmisiones continuas multimedia o de video codificadas que se introducen en el multiplexor 1806

sean también de CBR. Como se ha descrito en los antecedentes, el uso del CBR cuando se codifica contenido de video puede dar como resultado una calidad visual variable, que típicamente es indeseable.

En el sistema ilustrado, dos o más de los codificadores 1804 comunican una complejidad de codificación anticipada de los datos de entrada. Uno o más de los codificadores 1804 pueden recibir un control de la tasa de bits adaptada desde el multiplexor 1806 en respuesta. Esto permite a un codificador 1804 que espera codificar video relativamente complejo recibir una tasa de bits más alta o un ancho de banda más alto (más bits por cuadro) para esos cuadros del video en una forma de tasa de bits casi variable. Esto permite a la transmisión continua multimedia 1802 ser codificada con una calidad visual consistente. El ancho de banda extra que se usa por el codificador particular 1804 que codifica video relativamente complejo viene de los bits que se habrían usado en otro caso para la codificación de otras transmisiones continuas de video 1804 si los codificadores se implementasen para funcionar con tasas de bits constantes. Esto mantiene la salida del multiplexor 1806 en una tasa de bits constante (CBR).

Mientras que una transmisión continua multimedia individual 1802 puede ser relativamente a "ráfagas", esto es, variable en el ancho de banda usado, la suma acumulativa de las transmisiones continuas de video múltiples puede ser menos a ráfagas. La tasa de bits de los canales que se codifican con video menos complejo puede ser reasignada por, por ejemplo, el multiplexor 1806, a los canales que están codificando video relativamente complejo y esto puede mejorar la calidad visual de las transmisiones continuas de video combinadas como un todo.

Los codificadores 1804 proporcionan al multiplexor 1806 una indicación de la complejidad de un conjunto de cuadros de video a ser codificados y multiplexados juntos. La salida del multiplexor 1806 debería proporcionar una salida que no sea más alta que la tasa de bits especificada para el medio de transmisión 1808. Las indicaciones de la complejidad se pueden basar en la clasificación del contenido como se ha explicado anteriormente para proporcionar un nivel de calidad seleccionado. El multiplexor 1806 analiza las indicaciones de complejidad y proporciona a los varios codificadores 1804 un número asignado de bits o ancho de banda y los codificadores 1804 usan esta información para codificar los cuadros de video en el conjunto. Esto permite a un conjunto de cuadros de video tener individualmente una tasa de bits variable y conseguir aún una tasa de bits constante como grupo.

La clasificación del contenido se puede usar también en la habilitación de una conversión basada en la calidad de multimedia en general para cualquier compresor genérico. La clasificación del contenido y los procedimientos y aparatos descritos en el presente documento se pueden usar en el procesamiento multimedia en base a la calidad y/o en base al contenido de cualquier dato multimedia. Un ejemplo es su uso en la compresión de multimedia en general por cualquier compresor genérico. Otro ejemplo es en la descompresión o decodificación en cualquier compresor o decodificador o post-procesador tal como en las operaciones de interpolación, remuestreo, mejora, restauración y presentación.

Con referencia ahora a la FIGURA 19, un sistema de comunicación de video típico incluye un sistema de compresión de video que consiste en un codificador de video y un decodificador de video conectados mediante una red de comunicación. Las redes inalámbricas son una clase de redes proclives a errores en las que los canales de comunicación exhiben un desvanecimiento logarítmico-normal o ensombrecido y desvanecimiento multi-trayecto en escenarios móviles además de la pérdida de trayectoria. Para combatir los errores del canal y proporcionar comunicaciones fiables para los datos de la capa de aplicación, el modulador de RF incluye una corrección de error directa que incluyen intercaladores y codificación de canal tal como la convolucional o la turbo codificación.

La compresión de video reduce la redundancia en el origen del video y aumenta la cantidad de información transportada en cada bit de datos de video codificados. Esto aumenta el impacto en la calidad cuando incluso una pequeña parte del video codificado se pierde. La predicción espacial y temporal inherente en los sistemas de compresión de video agrava la pérdida y produce errores que se propagan resultando en artificiosidades visibles en el video reconstruido. Los algoritmos de resistencia frente a errores en el codificador de video y los algoritmos de recuperación de errores en el decodificador de video mejoran la robustez frente a errores del sistema de compresión de video.

Típicamente, el sistema de compresión de video es agnóstico respecto a la red subyacente. Sin embargo, en redes proclives a errores, la integración o alineación de los algoritmos de protección de error en la capa de aplicación con el FEC y la codificación del canal en las capas de enlace/físicas es altamente deseable y proporciona la mayor eficiencia en la mejora del rendimiento frente a errores del sistema.

La FIGURA 14 ilustra un ejemplo de un flujo de datos de tasa-distorsión que puede suceder en el codificador 228 para codificar cuadros. El proceso 1400 comienza en el inicio 1402 y prosigue en el bloque de decisión 1404 en el que recibe la entrada del detector de cambio de escena 1410 desde el preprocesador 226 (por ejemplo por medio de metadatos) y se adquiere la entrada de resistencia frente a errores 1406. Si la información indica un cuadro seleccionado como un cuadro I, el proceso intra-codifica el cuadro. Si la información indica que el cuadro seleccionado es un cuadro P o B, el proceso usa intra-codificación e (Inter) codificación de la estimación del movimiento para codificar el cuadro.

Después de que sucede una condición afirmativa para las condiciones del bloque 1404 el proceso 1400 prosigue en un bloque de preparación 1414 en el que la tasa R se fija en el valor $R = R_{cal}$, la calidad objetivo deseada en base a

las curvas R-D. El ajuste se recibe desde el bloque de datos 1416 que comprende las curvas R-D. El proceso 1400 prosigue entonces al bloque 1418 en donde se realiza la asignación de bits de control de tasa {QPi} en base a la información de actividad e imagen/video (por ejemplo la clasificación del contenido) desde un proceso de clasificación del contenido en el bloque 1420.

5 El bloque de asignación del bit de control de tasa 1418 se usa, a su vez, para la estimación del movimiento en el bloque 1422. La estimación del movimiento 1422 puede recibir también la entrada de metadatos desde el preprocesador 1412, el suavizador del detector de movimiento (MPEG-2 + Histórico) desde el bloque 1424 y múltiples cuadros de referencia (macro bloques MB causales + no causales) desde el bloque 1426. El proceso 1400 prosigue entonces al bloque 1428 donde se determinan los cálculos de tasa para los modos intra-codificados por parte de la asignación de bits de control de tasa (QPi). El proceso 1400 prosigue a continuación al bloque 1430 en donde se determinan los parámetros de modo y cuantificación. La decisión de modo del bloque 1430 se realiza en base a la entrada de estimación de movimiento del bloque 1422, la entrada de resistencia a los errores 1406 y la escalabilidad R-D, que se determina en el bloque 1432. Una vez que se decide el modo, el flujo prosigue al bloque 1432. Se indica que el flujo desde el bloque 1430 al 1432 sucede cuando los datos se pasan desde las partes del primer paso al segundo paso del codificador.

En el bloque 1432, se realiza la transformación y cuantificación por el segundo paso del codificador 228. El proceso de transformada/cuantificación se ajusta o se afina como se indica con el bloque 1444. Este proceso de transformada/cuantificación puede estar influenciado por el módulo de ajuste final del control de tasas (FIGURA 7). El proceso 1400 prosigue entonces en el bloque 1434 para una clasificación de zigzag y codificación de entropía para producir la capa base codificada. La codificación de zigzag prepara los datos cuantificados en un formato eficiente para codificación. La codificación de entropía es una técnica de compresión que usa una serie de códigos de bits para representar un conjunto de símbolos posibles. La capa mejorada resultante del bloque de transformada/cuantificación 1432 se envía también a un sumador 1436, que resta la capa base y envía el resultado a un codificador ZZ/entropía 1438 para la capa mejorada, como se ha descrito previamente con referencia a las FIGURAS 31-36. Se indica además, la capa mejorada se realimenta (véase la línea 1440 de actualización de la tasa verdadera) para la actualización de la clasificación del contenido en 1420 de tasa verdadera y una operación para la determinación de los históricos a largo y corto plazo de las tasas de bits para su uso en el control de la tasa.

La FIGURA 17 es un diagrama de bloques de alto nivel de un sistema de codificación multimedia. El sistema de codificación multimedia incluye medios para la recepción de datos multimedia como se ilustra por el módulo para la recepción de datos multimedia 1705. Tales medios pueden comprender, por ejemplo, un transcodificador, un codificador, un preprocesador, un procesador configurado para recibir datos multimedia, un receptor. Más específicamente, los medios de recepción pueden comprender componentes y módulos descritos en el presente documento para recibir datos multimedia, incluyendo, en varios ejemplos, el transcodificador 200. El sistema de codificación comprende también medios para la codificación de datos multimedia como se ilustra por el módulo para la codificación de datos multimedia 1710. Tales medios de codificación 1710 pueden comprender el transcodificador 200, el codificador 228 o el preprocesador 226.

Las FIGURAS 23, 24, 27 y 28 son diagramas de flujo de proceso que ejemplifican procedimientos de codificación de datos multimedia que realizan los aspectos descritos en el presente documento. La FIGURA 23 es un diagrama de flujo de proceso que ilustra un proceso 2300 de codificación de datos multimedia en base a la información del contenido. El bloque 2305 del proceso 2300 recibe los datos multimedia codificados y en el bloque 2310 el proceso 2300 decodifica los datos multimedia. En el bloque 2315, el proceso 2300 determina la información del contenido asociada con los datos multimedia decodificados. En el bloque 2320, el proceso 2300 codifica los datos multimedia en base a la información del contenido.

La FIGURA 24 es un diagrama de flujo de proceso que ilustra un proceso 2400 de codificación de datos multimedia de modo que alinea los límites de los datos en base al nivel de información del contenido. En el bloque 2405, el proceso 2400 obtiene información del contenido asociada con los datos multimedia, lo que se puede realizar, por ejemplo, por el preprocesador 216 o el módulo de clasificación del contenido 712 mostrado en la FIGURA 7. En el bloque 2410, el proceso 2400 codifica los datos multimedia de modo que alinea los límites de los datos en base a la información del contenido. Por ejemplo, los límites de porciones y los límites de unidades de acceso se alinean con los límites de cuadros en base a la clasificación del contenido de los datos multimedia que están siendo codificados. Los datos codificados están disponibles entonces para un procesamiento posterior y/o transmisión a un dispositivo móvil y el proceso 2400 finaliza.

La FIGURA 27 es un diagrama de flujo de proceso que ilustra un proceso 2700 para la codificación de datos usando un esquema de intra refresco adaptativo en base a la información del contenido. Cuando el proceso 2700 comienza los datos multimedia se han obtenido. En el bloque 2705, el proceso 2700 obtiene la información del contenido de los datos multimedia. La obtención de la información del contenido se puede realizar por, por ejemplo, el preprocesador 226 o el módulo de clasificación del contenido 712 como se ha descrito anteriormente. El proceso 2700 prosigue en el bloque 2710, en el que codifica los datos multimedia usando un esquema de resistencia frente a errores de intra refresco adaptativo, en el que el esquema de resistencia frente a errores de intra refresco adaptativo se basa en la información del contenido. La funcionalidad del bloque 2700 se puede realizar por el codificador 228. Los datos codificados se ponen a disposición de un procesamiento posterior y transmisión y el proceso de 2700

finaliza a continuación.

La FIGURA 28 es un diagrama de flujo de proceso que ilustra un proceso de codificación de datos multimedia usando cuadros I redundantes en base a la información del contenido multimedia. Cuando el proceso 2800 comienza los datos multimedia están disponibles para su procesamiento. En el bloque 2805, el proceso 2800 obtiene la información del contenido de los datos multimedia. Como se ha descrito anteriormente, esto se puede realizar por, por ejemplo, el preprocesador 226 y/o el codificador 228. En el bloque 2810, el proceso 2800 codifica los datos multimedia de modo que inserte uno o más cuadros I adicionales de los datos codificados en base a la información del contenido. Esto se puede realizar por el codificador 228 como se ha descrito anteriormente en conexión con el esquema de resistencia frente a errores, insertando los cuadros I en la capa base o en la capa de mejora dependiendo del esquema de resistencia frente a errores empleado. Después del bloque 2810, los datos codificados están disponibles para un procesamiento posterior y/o transmisión a un dispositivo móvil.

Se debería indicar que los procedimientos descritos en el presente documento se pueden implementar en una variedad de hardware de comunicación, procesadores y sistemas conocidos por los expertos en la técnica. Por ejemplo, el requisito general para que el cliente opere como se ha descrito en el presente documento es que el cliente tenga una pantalla para visualizar el contenido de la información, un procesador para controlar la operación del cliente y una memoria para almacenar datos y programas relacionados con la operación del cliente. En un aspecto, el cliente es un teléfono celular. En otro aspecto, el cliente es un ordenador portátil que tenga capacidad de comunicaciones. El otro aspecto más, el cliente es un ordenador personal que tenga capacidad de comunicaciones. Además, se puede incorporar hardware tal como un receptor GPS en el cliente para implementar los varios aspectos. Los varios bloques lógicos, lógica, módulos y circuitos ilustrativos descritos en conexión con los aspectos desvelados en el presente documento se pueden implementar o realizar con un procesador de finalidad general, un procesador de señal digital (DSP), un circuito integrado de aplicación específica (ASIC), una matriz de puertas programable en campo (FPGA) u otros dispositivos lógicos programables, puertas discretas o lógica de transistores, componentes de hardware discretos o cualquier combinación de los mismos diseñados para realizar las funciones descritas en el presente documento. Un procesador de finalidad general puede ser un microprocesador pero, en la alternativa, el procesador puede ser cualquier procesador, controlador, microcontrolador o máquina de estado convencional. Un procesador puede ser implementado también como la combinación de dispositivos de cálculo, por ejemplo, una combinación de un DSP y un microprocesador, una pluralidad de microprocesadores, uno más microprocesadores en conjunto con un núcleo de DSP o cualquier otra de tales combinaciones.

Las varias lógicas, bloques lógicos, módulos y circuitos ilustrativos descritos en conexión con los aspectos desvelados en el presente documento se pueden implementar o realizar con un procesador de finalidad general, procesador de señal digital (DSP), un circuito integrado de aplicación específica (ASIC), una matriz de puertas programable en campo (FPGA) u otros dispositivos lógicos programables, puertas discretas o lógica de transistores, componentes de hardware discretos o cualquier combinación de los mismos diseñada para realizar las funciones descritas en el presente documento. Un procesador de finalidad general puede ser un microprocesador pero, en la alternativa, el procesador puede ser cualquier procesador, controlador, microcontrolador o máquina de estado convencional. Un procesador se puede implementar también como la combinación de dispositivos de cálculo, por ejemplo, una combinación de un DSP y un microprocesador, una pluralidad de microprocesadores, uno más microprocesadores en conjunto con un núcleo de DSP o cualquier otra de tales combinaciones.

Los procedimientos y aparatos desvelados proporcionan la transcodificación de datos de video codificados en un formato a datos de video codificados en otro formato en donde la codificación se basa en el contenido de los datos de video y la codificación es resistente frente a errores. Los procedimientos o algoritmos descritos en conexión con los ejemplos desvelados en presente documento se pueden realizar directamente en hardware, en un módulo de software ejecutado por un procesador, firmware o en una combinación de dos o más de estos. Un módulo de software puede residir en una memoria RAM, memoria flash, memoria ROM, memoria de EPROM, memoria EEPROM, registradores, un disco duro, un disco extraíble, un CD-ROM o cualquier otra forma de medios de almacenamiento conocida en la técnica. Un medio de almacenamiento de ejemplo se conecta al procesador, de modo que el procesador pueda leer la información desde, y escribir la información a, el medio de almacenamiento. En la alternativa, en medio de almacenamiento puede ser parte integral del procesador. El procesador y el medio de almacenamiento pueden residir en un ASIC. El ASIC puede residir en un terminal de usuario. En la alternativa, el procesador y el medio de almacenamiento pueden residir como componentes discretos en un terminal de usuario.

Los ejemplos descritos anteriormente son meramente de ejemplo y los expertos en la técnica pueden hacer ahora numerosos usos de, y separaciones de, los ejemplos descritos anteriormente sin separarse de los conceptos inventivos desvelados en el presente documento. Varias modificaciones a estos ejemplos puede ser fácilmente evidentes para los expertos en la técnica y los principios genéricos definidos en el presente documento se pueden aplicar a otros ejemplos, por ejemplo en un servicio de mensajería instantánea o en cualquier aplicación de comunicación de datos inalámbrica general.

Por ello, el alcance de la divulgación no se pretende que se limite a los ejemplos mostrados en el presente documento sino que esté de acuerdo con el alcance más amplio consistente con los principios y características novedosas desveladas en el presente documento. La expresión "de ejemplo" se usa exclusivamente en el presente documento para indicar "que sirve como un ejemplo, caso o ilustración". Cualquier ejemplo descrito en el presente

documento como "de ejemplo" no se ha de interpretar necesariamente como preferido o ventajoso frente a otros ejemplos. En consecuencia, los aspectos novedosos descritos en el presente documento se han de definir únicamente mediante el alcance de las siguientes reivindicaciones.

REIVINDICACIONES

1. Un procedimiento de procesamiento de datos multimedia que comprende:

la recepción de datos multimedia y
 la codificación de los datos multimedia (361) en una primera capa de datos codificados y una segunda capa
 de datos codificados en base al contenido de los datos multimedia, estando configurada la primera capa de
 datos codificados para ser codificable de modo independiente de la segunda capa de datos codificados y
 en el que la primera capa de datos se codifica con una calidad menor que la segunda capa de datos,
caracterizado porque el procedimiento comprende:
 la codificación de la primera capa de datos mediante las etapas de:
 transformación de los coeficientes de error residuales de la primera capa de datos, siendo el error residual la
 diferencia entre los datos multimedia que se están codificando (361) y el macro bloque de ajuste mejor de un
 cuadro de referencia temporal;
 el cálculo de un coeficiente de la primera de capa de datos, C'_{base} , a partir de los coeficientes de error residual
 transformados, C_{base} y
 la cuantificación del coeficiente de la primera capa de datos, C'_{base} , mediante un primer parámetro de
 cuantificación, Q_b ;
 y la codificación de la segunda capa de datos mediante las etapas de:
 la transformación de los coeficientes de error residual en la segunda capa de datos, siendo el error residual la
 diferencia entre los datos multimedia que se están codificando (361) y el macro bloque de ajuste mejor de un
 cuadro de referencia temporal;
 el cálculo de un coeficiente de la segunda capa de de datos, C'_{mej} , a partir de los coeficientes de error residual
 transformados, C_{mej} y
 la cuantificación del coeficiente de la segunda capa de datos, C'_{mej} , mediante un segundo parámetro de
 cuantificación, en el que:

$$C'_{base} = \begin{cases} 0, & \text{si } C_{base} \text{ y } C_{mej} \text{ tienen signos opuestos} \\ \min(C_{base}, C_{mej}), & \text{en caso contrario} \end{cases}$$

o

$$C'_{base} = \begin{cases} C_{base}, & \text{si } |Q_b^{-1}Q_b(C_{base}) - C_{mej}| < |C_{mej}| \\ 0, & \text{en caso contrario} \end{cases}$$

y

$$C'_{mej} = C_{mej} - Q_b^{-1}(Q_b(C'_{base}))$$

en la que Q_b^{-1} es el parámetro de descuantificación de la primera capa de datos y en el que el primer
 parámetro de cuantificación se ajusta para ser más alto que el segundo parámetro de cuantificación, de modo
 que la primera capa de datos codificados se puede decodificar para formar unos datos multimedia que se
 pueden visualizar si la segunda capa de datos codificada no está disponible y la primera capa de datos
 codificados y la segunda capa de datos codificados se pueden decodificar en combinación para formar unos
 datos multimedia que se pueden visualizar si tanto la primera capa de datos codificada como la segunda capa
 de datos codificada están disponibles.

2. El procedimiento de la reivindicación 1, en el que la primera capa de datos comprende cuadros I, intra-codificados,
 y cuadros P, predictivos y en el que la segunda capa de datos comprende cuadros I, intra-codificados, cuadros P,
 predictivos y cuadros B, bidireccionales.

3. El procedimiento de la reivindicación 1, en el que la primera capa de datos comprende una capa base y la
 segunda capa de datos comprende una capa de mejora.

4. El procedimiento de la reivindicación 1, en el que la codificación comprende la asignación de una tasa de bits para
 al menos una parte de los datos multimedia en base a la clasificación del contenido.

5. El procedimiento de la reivindicación 1, en el que la codificación comprende además:

el uso de la clasificación del contenido para detectar cambios de escena y
 la determinación de si incluir cuadros I, intra-codificados en la primera capa de datos y en la segunda capa de
 datos en base a los cambios de escena detectados.

6. El procedimiento de la reivindicación 1, en el que la codificación comprende además la determinación de una tasa
 de cuadros para la codificación de los datos multimedia en base a la clasificación del contenido.

7. El procedimiento de la reivindicación 1, en el que la codificación comprende la realización de la estimación de movimiento de los datos multimedia en base a la clasificación del contenido.
8. El procedimiento de la reivindicación 1, que comprende además la determinación de una primera tasa de cuadros para la codificación de la primera capa de datos y la determinación de una segunda tasa de cuadros para la codificación de la segunda capa de datos, en la que la primera tasa de cuadros es menor que la segunda tasa de cuadros.
9. El procedimiento de la reivindicación 1, en el que la codificación comprende la realización de un procesamiento de resistencia a los errores en los datos multimedia en base a la clasificación del contenido.
10. El procedimiento de la reivindicación 1, en el que el primer parámetro de cuantificación comprende un primer tamaño de paso para la codificación de datos y en el que el segundo parámetro de cuantificación comprende un segundo tamaño de paso para la codificación de los datos, en el que el primer tamaño de salto es mayor que el segundo tamaño de paso.
11. El procedimiento de la reivindicación 1, en el que la codificación comprende la reducción de ruido en los datos multimedia en base a la clasificación del contenido.
12. El procedimiento de la reivindicación 11, en el que la reducción de ruido comprende la realización de una eliminación de artificiosidades.
13. El procedimiento de la reivindicación 11, en el que la reducción de ruido comprende el procesamiento de al menos una parte de los datos multimedia con un filtro de desanillado, en el que la intensidad del filtro de desanillado se basa en el contenido de los datos multimedia.
14. El procedimiento de la reivindicación 11, en el que la reducción de ruido comprende el procesamiento de al menos una parte de los datos multimedia con un filtro de supresión de bloques, en el que la intensidad del filtro de supresión de bloques se basa en el contenido de los datos multimedia.
15. El procedimiento de la reivindicación 11, en el que la reducción de ruido comprende el filtrado de frecuencias seleccionadas de los datos multimedia.
16. El procedimiento de la reivindicación 13, en el que la intensidad del filtro de desanillado se basa en la clasificación del contenido de los datos multimedia.
17. El procedimiento de la reivindicación 14, en el en el que la intensidad del filtro de supresión de bloques se basa en la clasificación del contenido de los datos multimedia.
18. El procedimiento de la reivindicación 1, en el que la codificación comprende la reducción del muestreo de los datos multimedia.
19. El procedimiento de la reivindicación 1, en el que la codificación comprende la asociación de un nivel de calidad con los datos multimedia y el uso del nivel de calidad y la información del contenido de los datos multimedia para determinar una tasa de bits para la codificación de los datos multimedia.
20. Un aparato para procesamiento de datos multimedia que comprende un codificador (228) configurado para recibir datos multimedia y codificar los datos multimedia en una primera capa de datos codificados y una segunda capa de datos codificados en base al contenido de los datos multimedia, configurada la primera capa de datos codificados para ser decodificarle de modo independiente de la segunda capa de datos y en la que la primera capa de datos se codifica con una calidad más baja que la segunda capa de datos; el aparato **caracterizado porque** el codificador (328) se configura para codificar la primera capa de datos mediante:
- la transformación de los coeficientes de error residual de la primera capa de datos, siendo el error residual la diferencia entre los datos multimedia que se están codificando (361) y el macro bloque de ajuste mejor de un cuadro de referencia temporal;
- el cálculo de un coeficiente de la primera de capa de datos, C'_{base} , para los coeficientes de error residual transformados, C_{base} y
- la cuantificación del coeficiente de la primera capa de datos, C'_{base} , mediante un primer parámetro de cuantificación, Q_b ; y en el que el codificador se configura para codificar la segunda capa de datos mediante:
- la transformación de los coeficientes de error residual de la segunda capa de datos, siendo el error residual la diferencia entre los datos multimedia que se están codificando (361) y el macro bloque de ajuste mejor de un cuadro de referencia temporal;
- el cálculo de un coeficiente de la segunda capa de de datos, C'_{mej} , a partir de los coeficientes de error residual transformados, C_{mej} y
- la cuantificación del coeficiente de la segunda capa de datos, C'_{mej} , mediante un segundo parámetro de cuantificación, en el que:

$$C'_{base} = \begin{cases} 0, & \text{si } C_{base} \text{ y } C_{mej} \text{ tienen signos opuestos} \\ \min(C_{base}, C_{mej}), & \text{en caso contrario} \end{cases}$$

o

$$C'_{base} = \begin{cases} C_{base}, & \text{si } |Q_b^{-1}Q_b(C_{base}) - C_{mej}| < |C_{mej}| \\ 0, & \text{en caso contrario} \end{cases}$$

y

$$5 \quad C'_{mej} = C_{mej} - Q_b^{-1}(Q_b(C'_{base}))$$

en la que Q_b^{-1} es el parámetro de descuantificación de la primera capa de datos y en el que los codificados se configura además para determinar el primer parámetro de cuantificación del segundo parámetro de cuantificación, siendo él el primer parámetro de cuantificación más alto que el segundo parámetro de cuantificación, de modo que la primera capa de datos codificados se puede decodificar para formar unos datos multimedia que se pueden visualizar si la segunda capa de datos codificada no está disponible y la primera capa de datos codificados y la segunda capa de datos codificados se pueden decodificar en combinación para formar unos datos multimedia que se pueden visualizar si tanto la primera capa de datos codificada como la segunda capa de datos codificada están disponibles.

15 21. El aparato de la reivindicación 20, en el que el codificador (228) comprende un módulo de estimación del movimiento (720) configurado para realizar la estimación del movimiento de los datos multimedia en base a la clasificación del contenido y generar una información de compensación del movimiento de los datos y en el que el módulo de codificación (228) se configura además para usar la información de compensación del movimiento para codificar los datos multimedia.

20 22. El aparato de la reivindicación 20, en el que el codificador (228) comprende además un módulo de cuantificación (726) para la determinación del primer y del segundo parámetros de cuantificación.

23. El aparato de la reivindicación 20, en el que el codificador (228) comprende además un módulo de asignación de bits (714) configurado para proporcionar una tasa de bits para al menos una parte de los datos multimedia en base a la clasificación del contenido.

25 24. El aparato de la reivindicación 20, en el que el codificador (228) comprende además un módulo de detección de cambio de escena configurado para detectar cambios de escena y en el que el módulo de codificación (228) se configura además para incluir cuadros I, intra-codificados en los datos multimedia codificados en base a los cambios de escena detectados.

30 25. El aparato de la reivindicación 20, en el que el codificador (228) comprende además un módulo de tasa de cuadros configurado para determinar la tasa de cuadros de los datos multimedia en base a la clasificación del contenido y en el que el módulo de codificación (228) codifica los datos multimedia en base a la tasa de cuadros.

26. El aparato de la reivindicación 20, en el que el codificador (228) se configura además para realizar un procesamiento de errores en los datos multimedia en base a la clasificación del contenido.

27. Un medio que pueda leer una máquina que comprende instrucciones que tras su ejecución hacen que una máquina realice el procedimiento de la reivindicación 1.

35 28. Un procesador que comprende una configuración para realizar el procedimiento de la reivindicación 1.

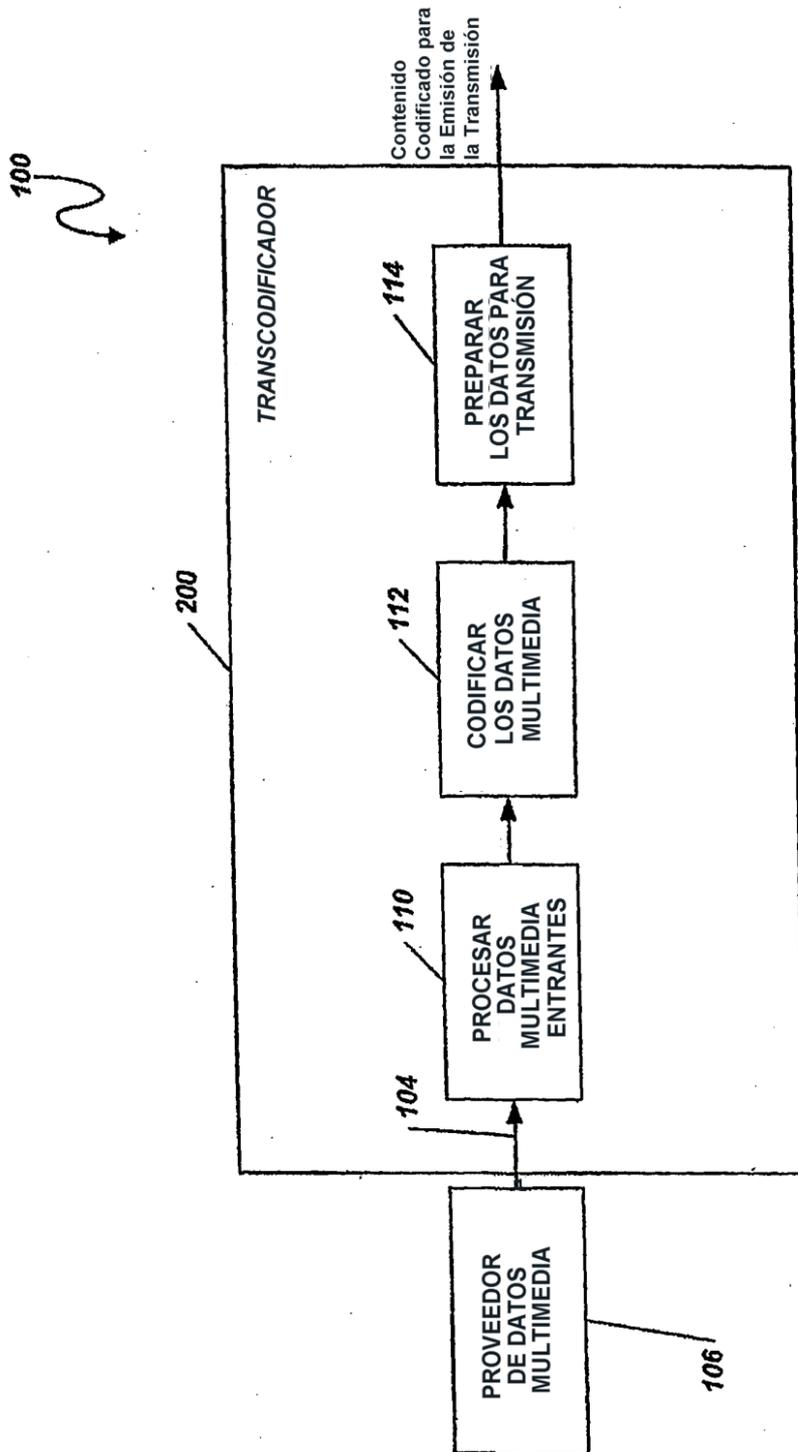


FIG. 1A

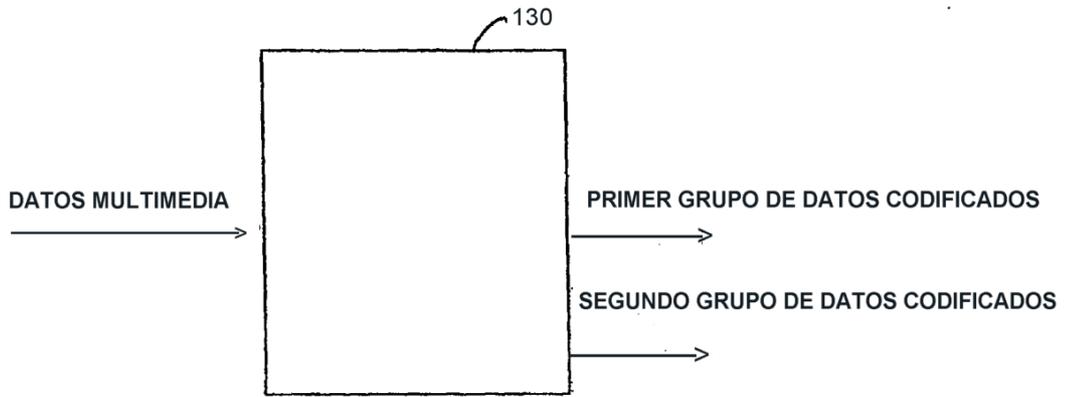


FIG. 1B

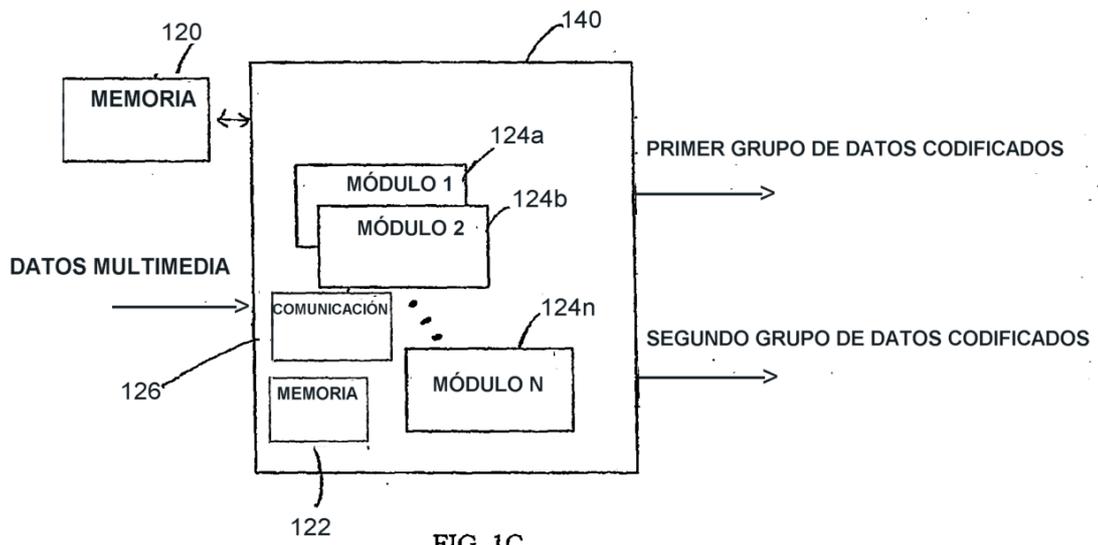


FIG. 1C

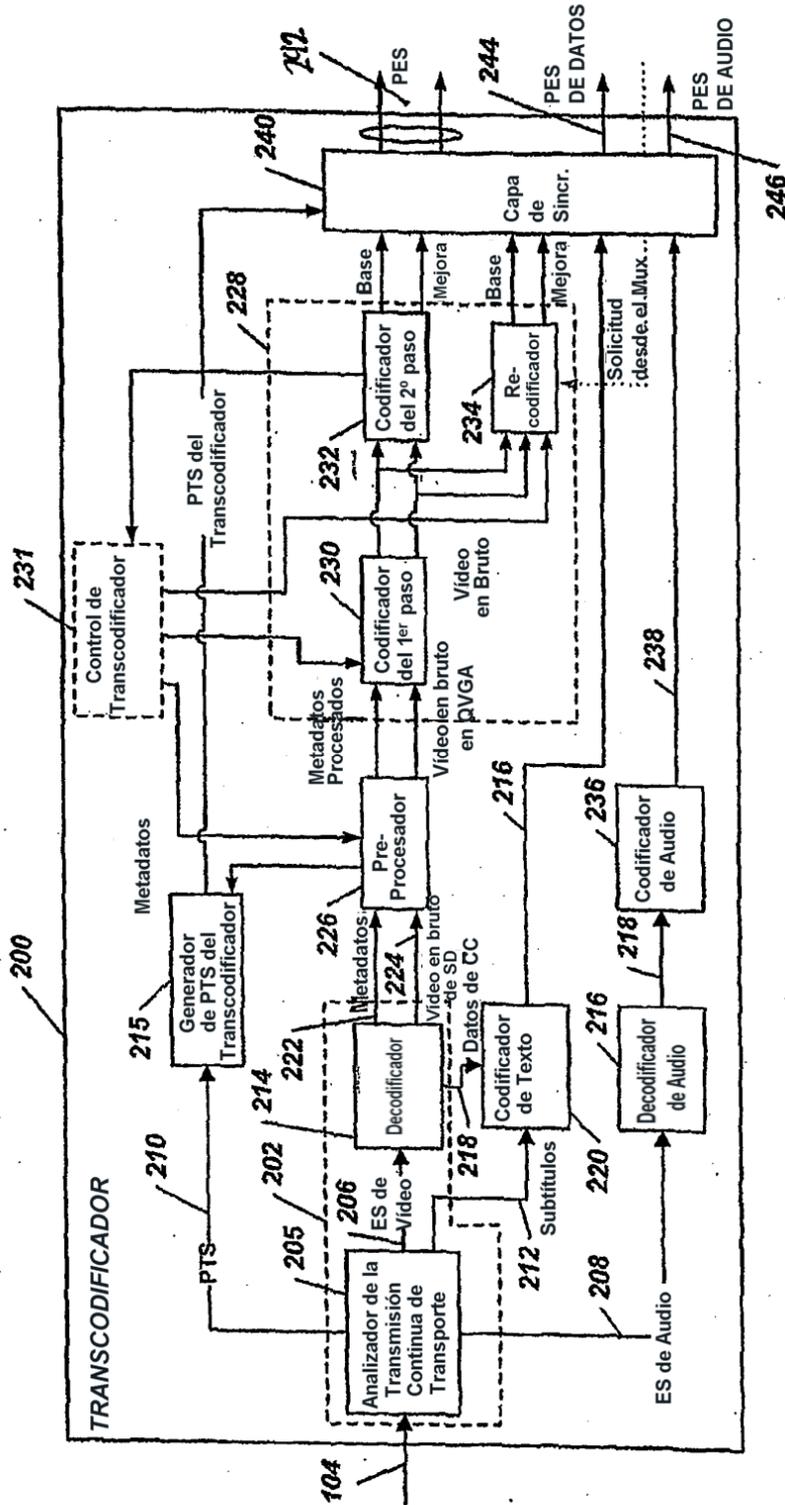


FIG. 2

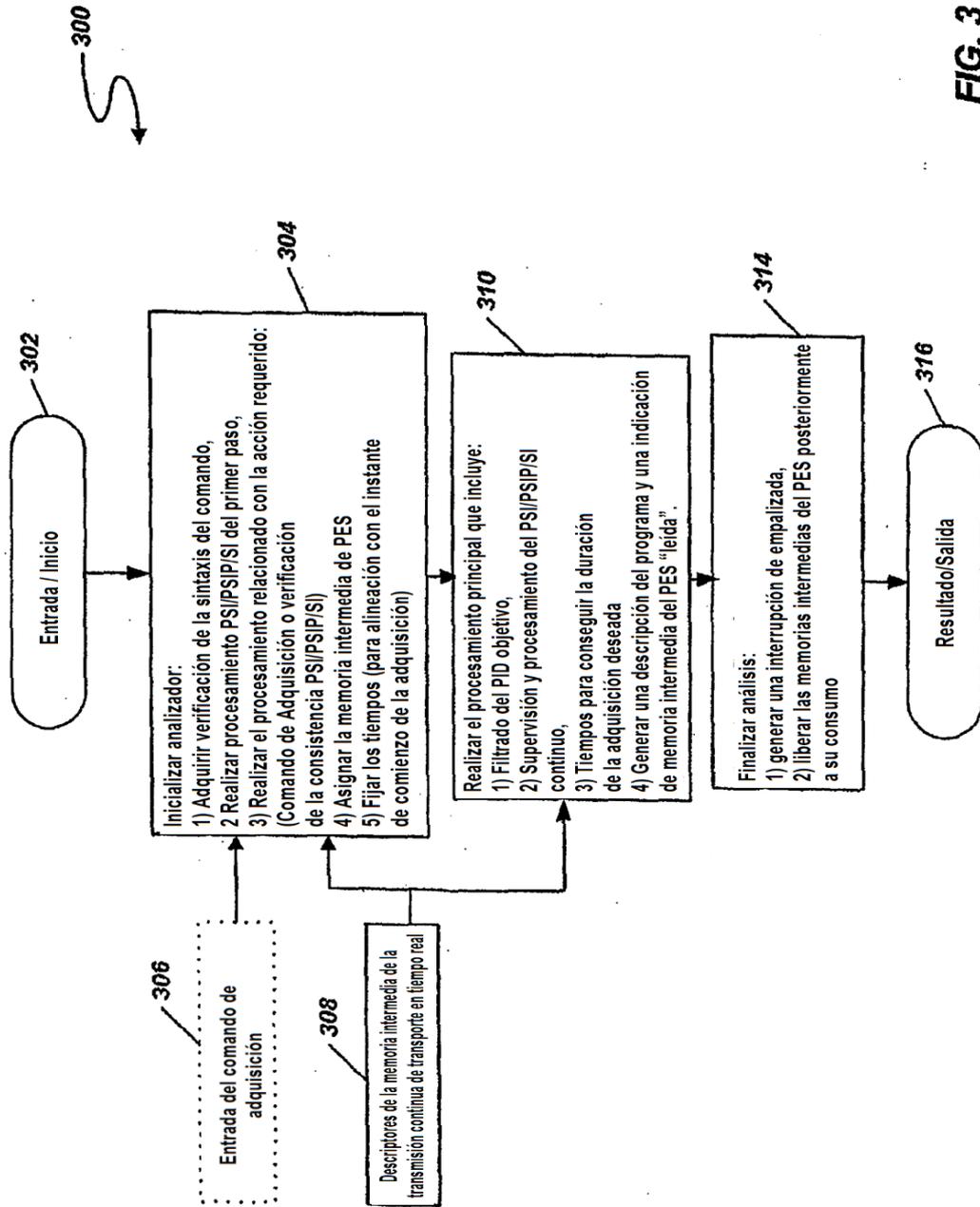


FIG. 3

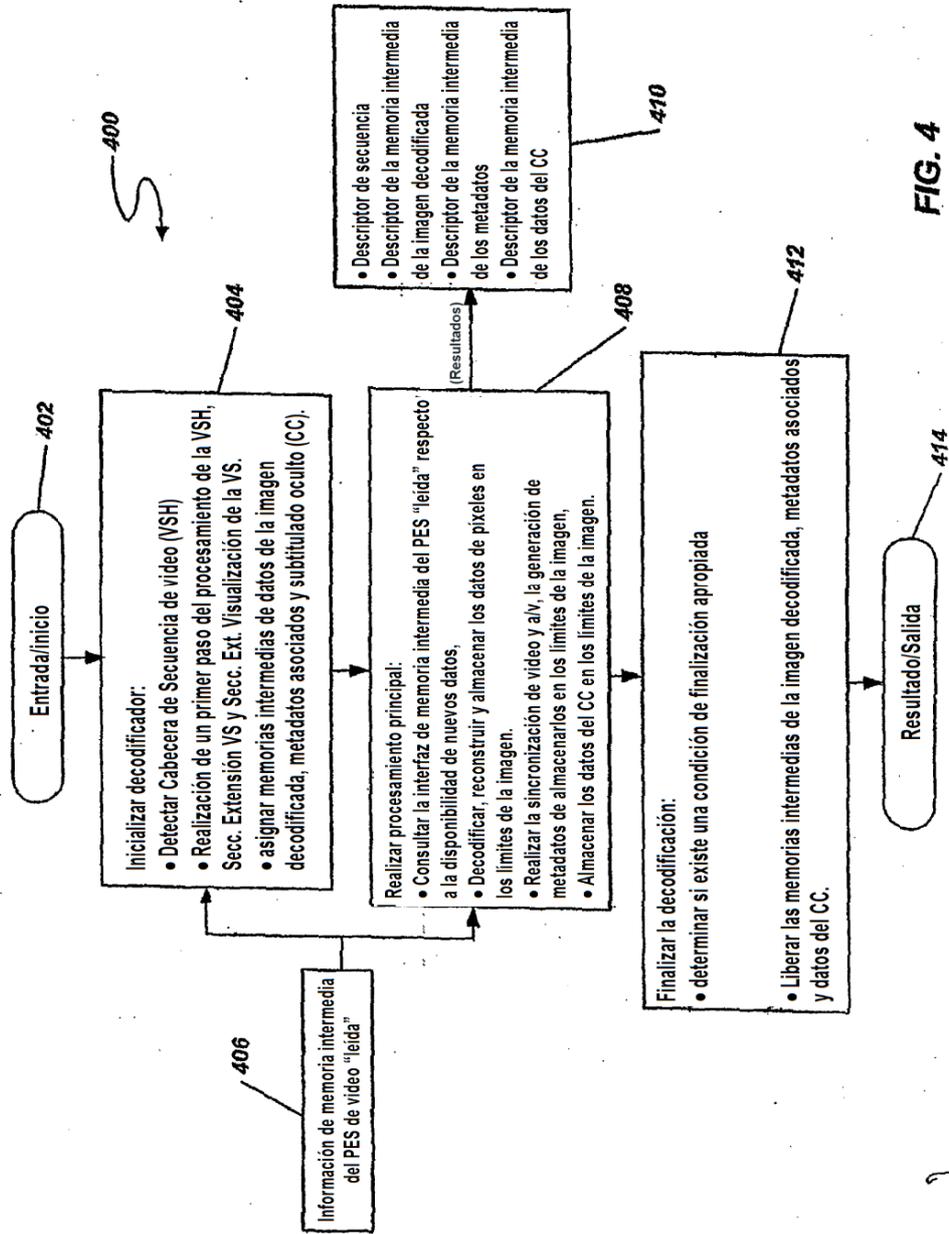


FIG. 4

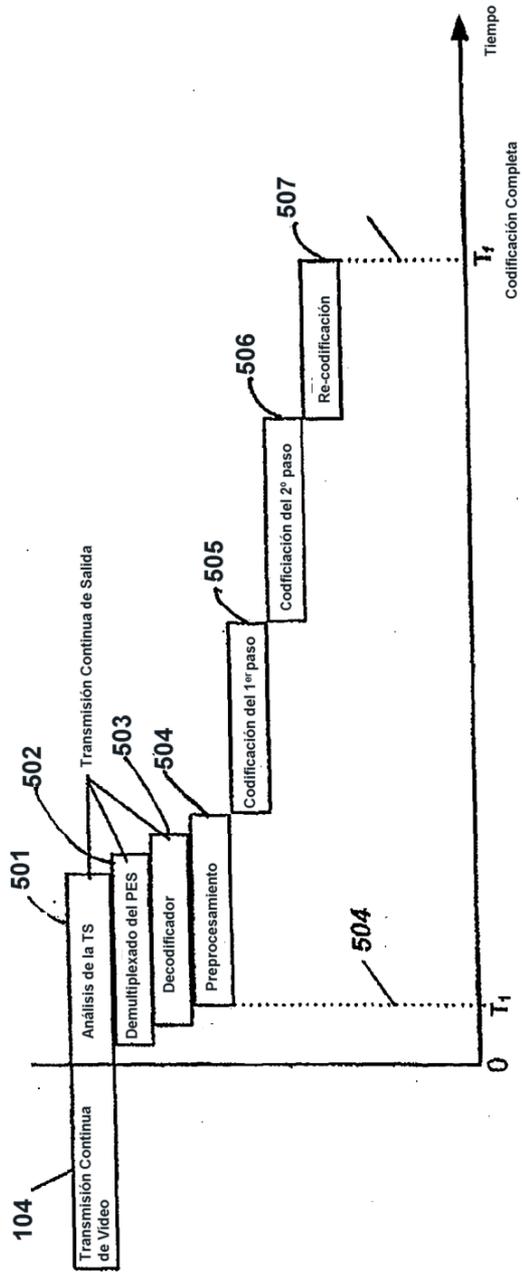


FIG. 5

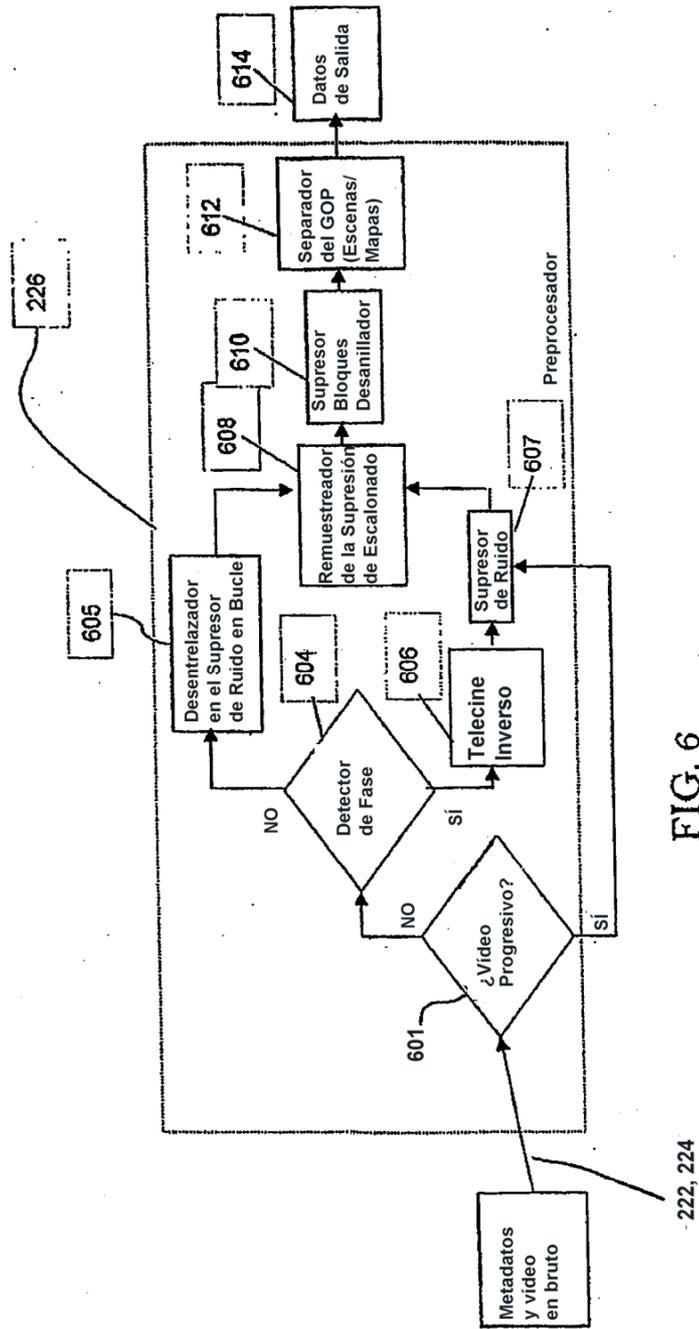


FIG. 6

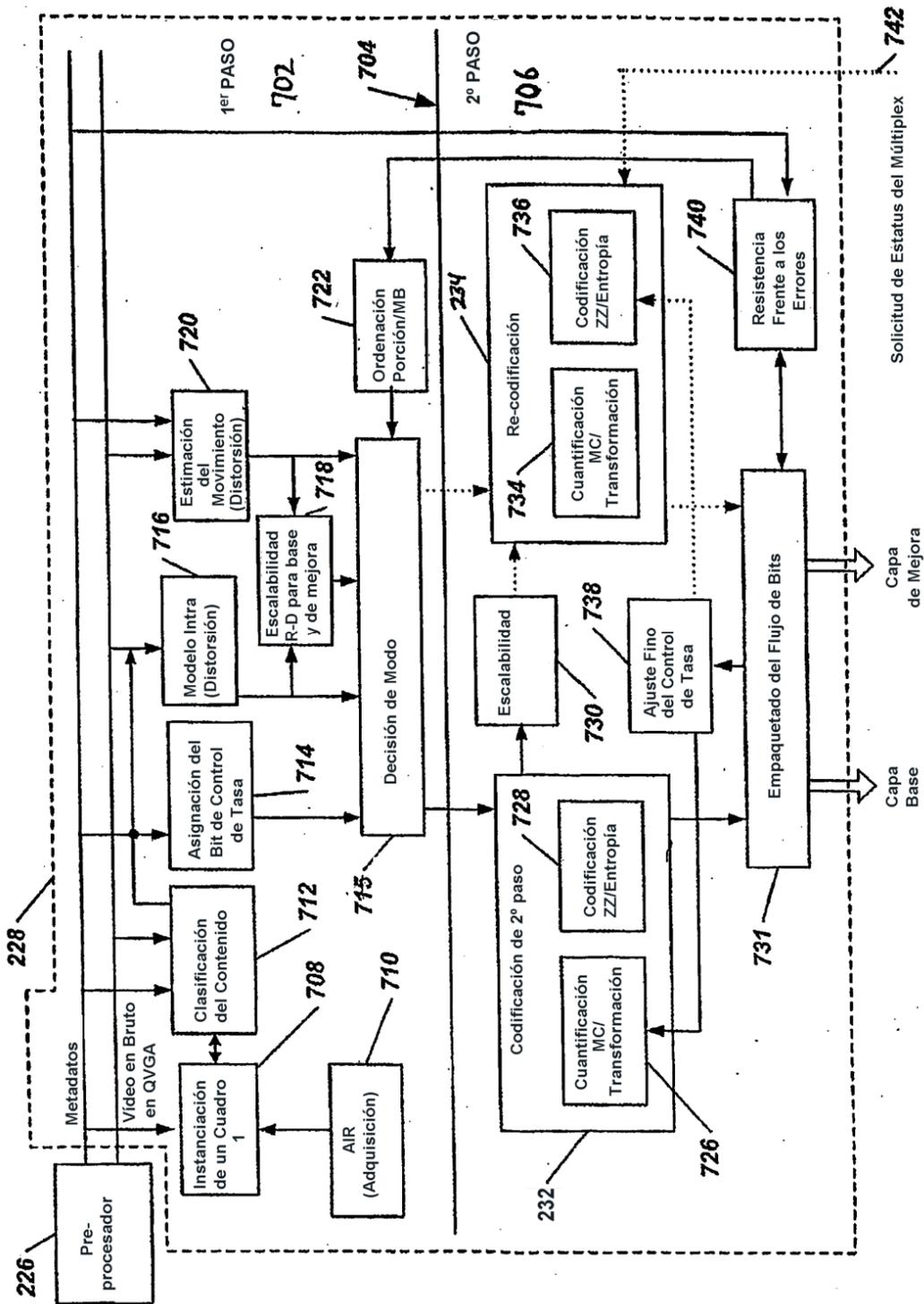


FIG. 7

		5	6	8	
Alto		HL	HM	HH	(CLASE)
Medio		2	3	7	
Bajo		1	2	4	
		Bajo	Medio	Alto	

MOVIMIENTO

TEXTURA

FIG. 8

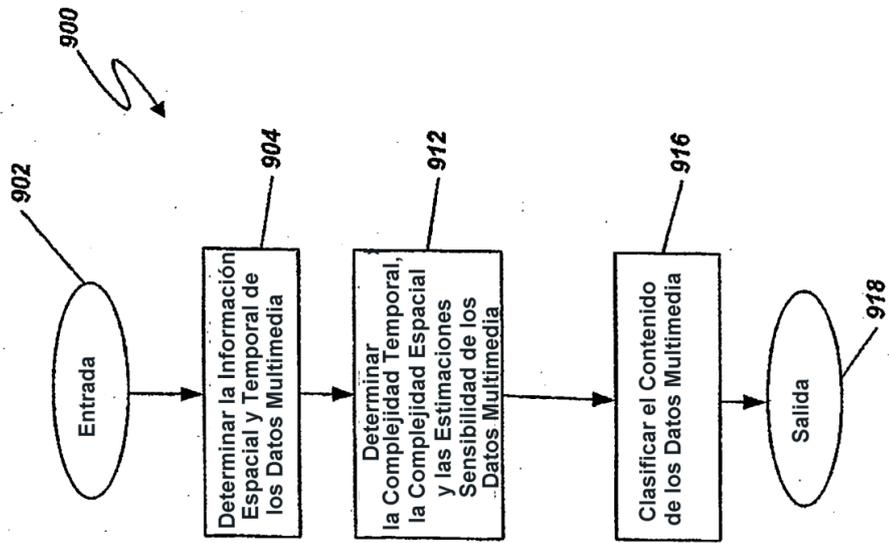


FIG. 9

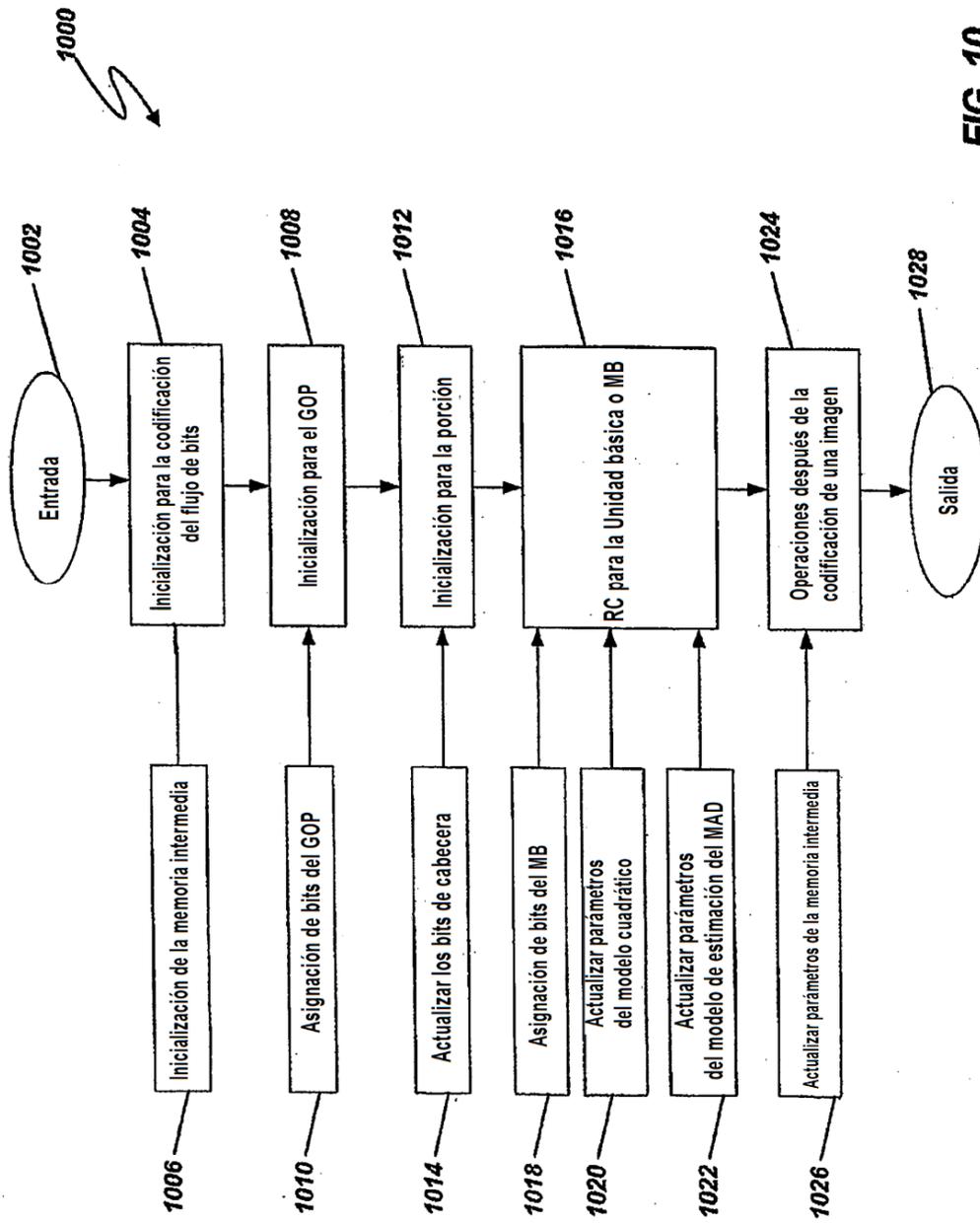


FIG. 10

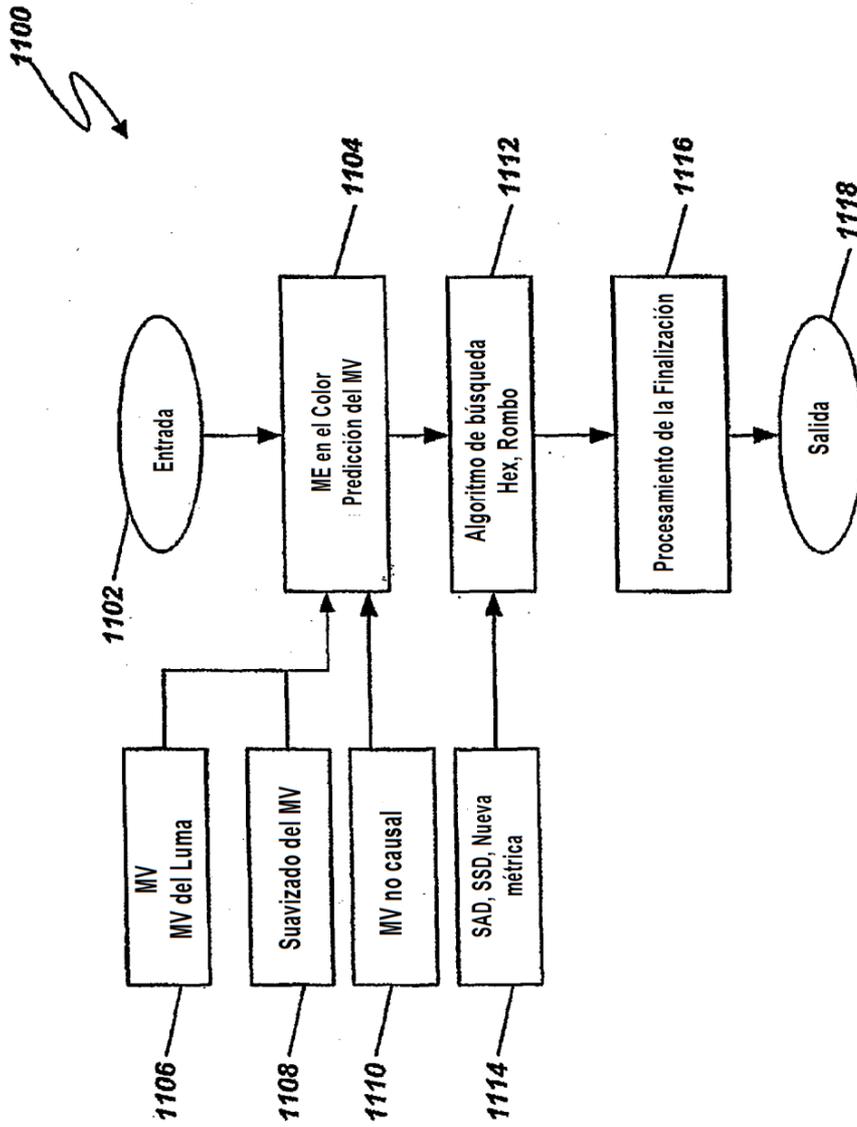


FIG. 11

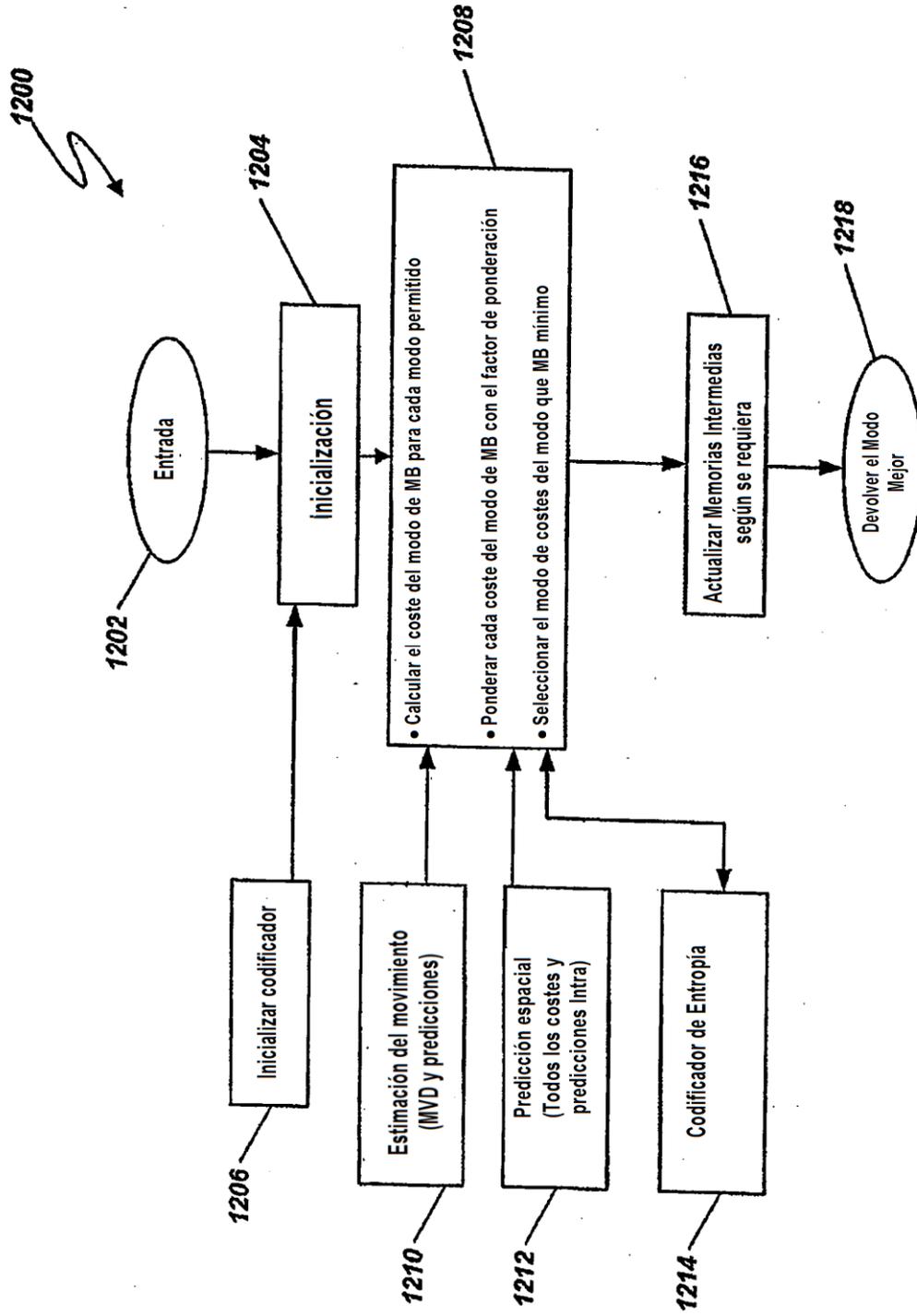


FIG. 12

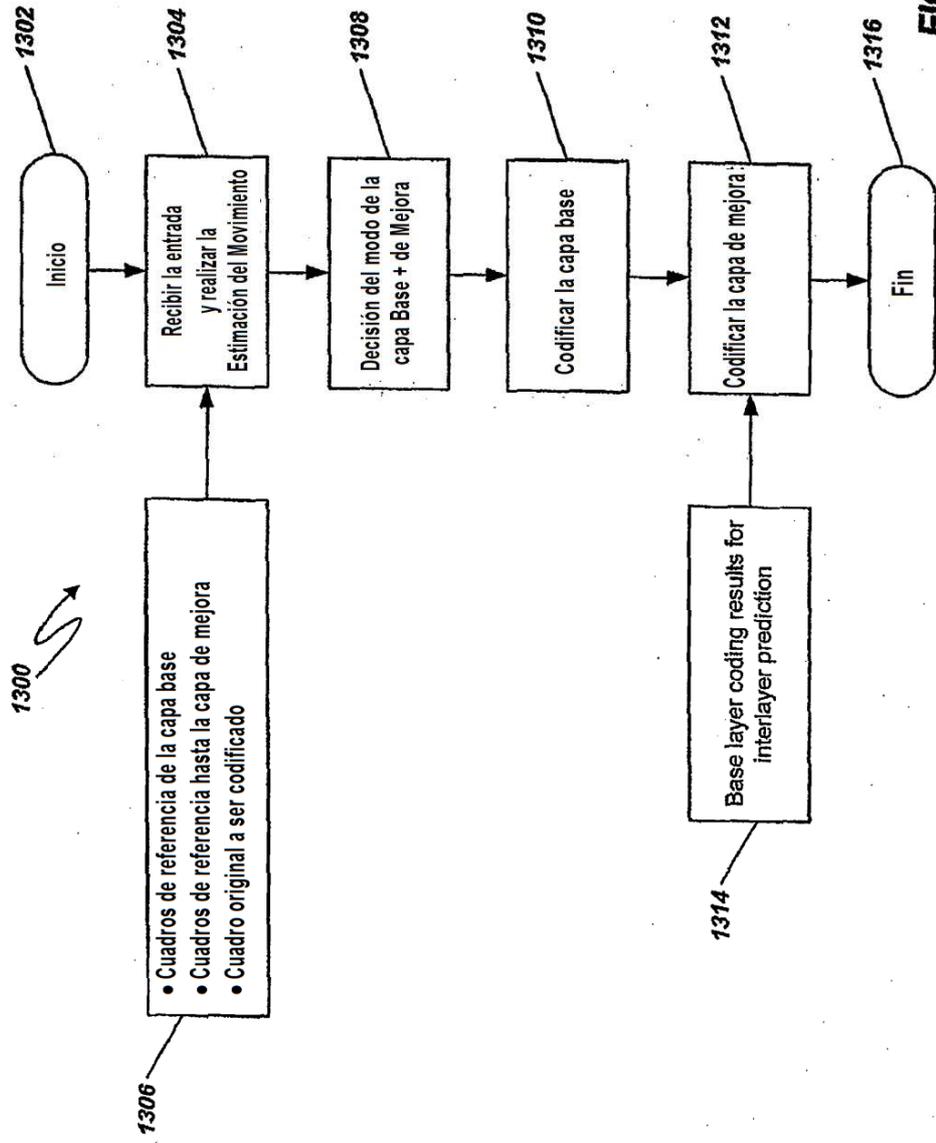


FIG. 13

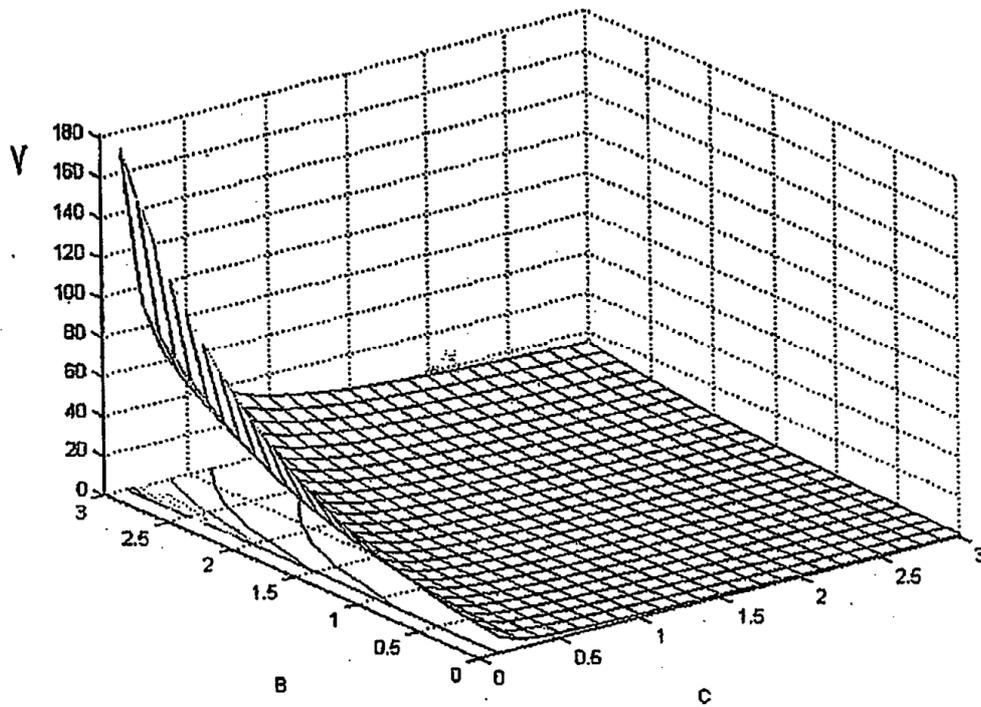


FIG. 15

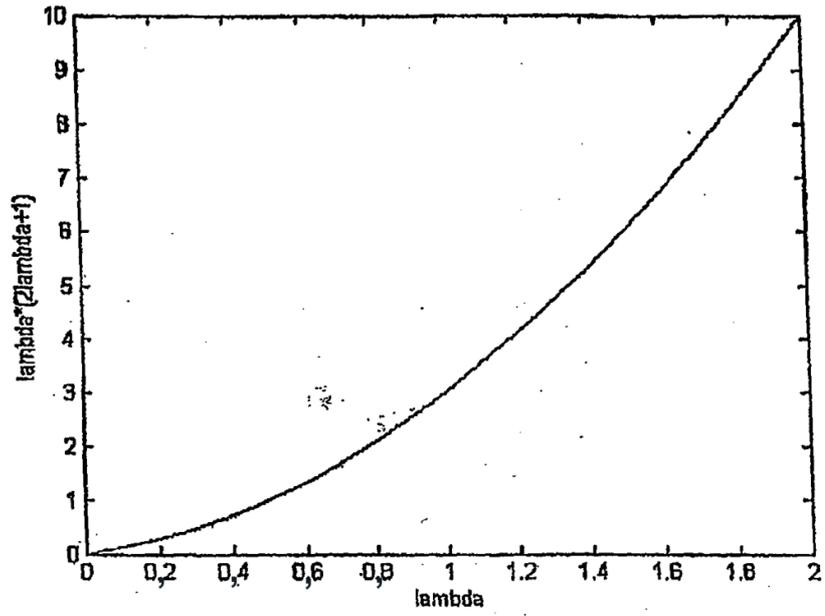


FIG. 16

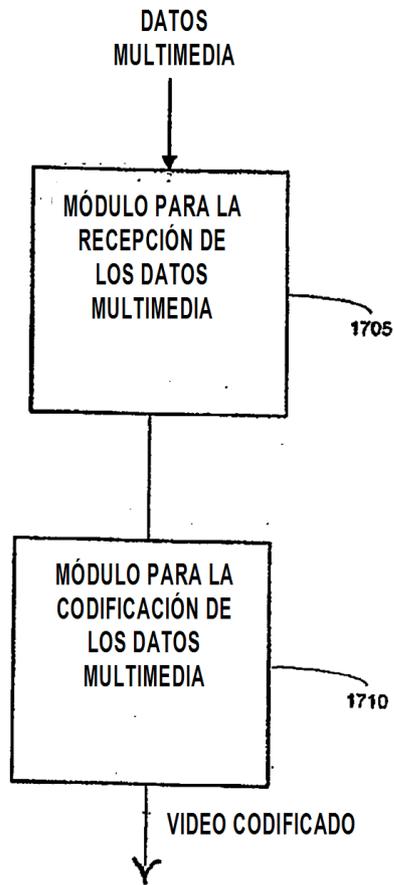


FIG. 17

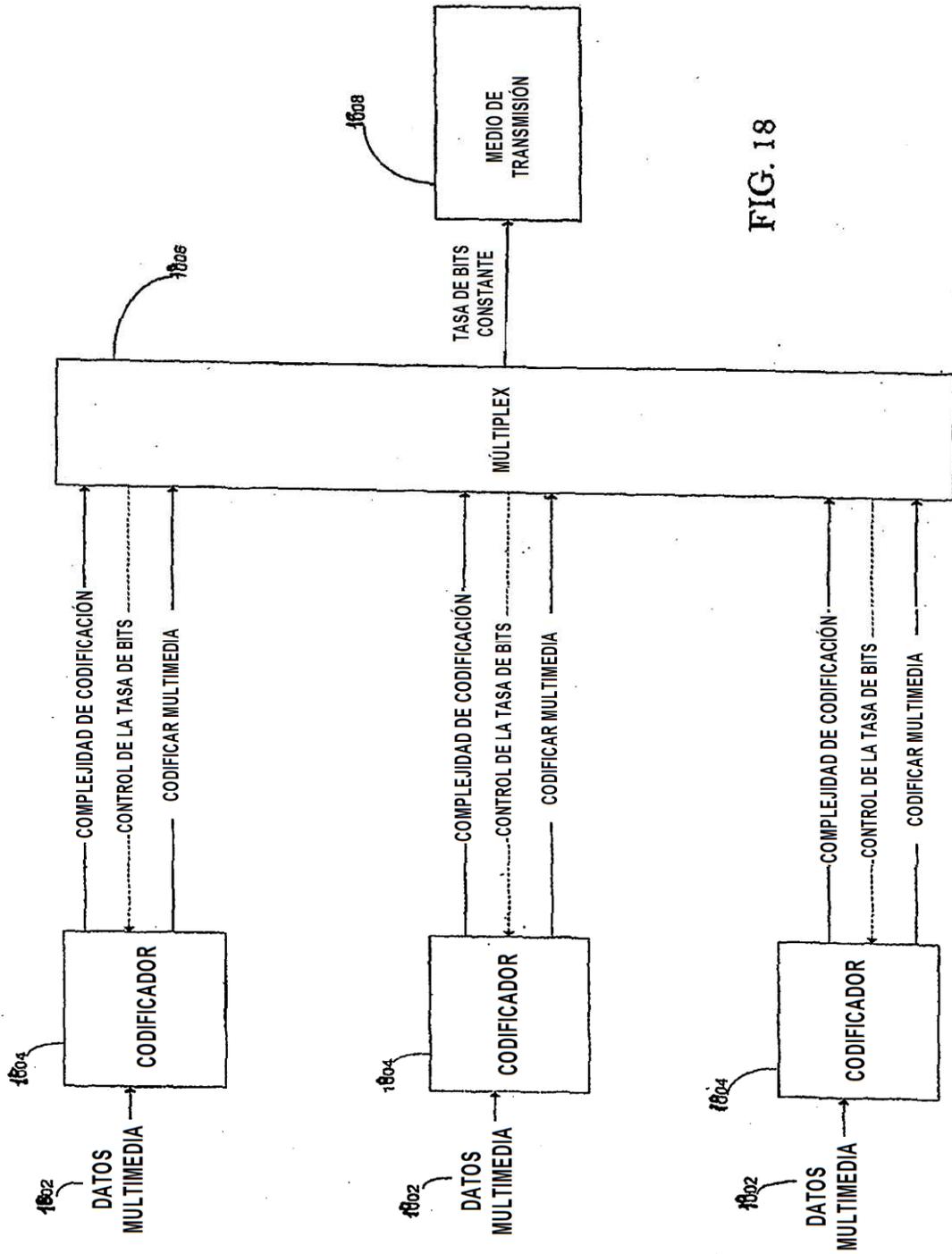


FIG. 18

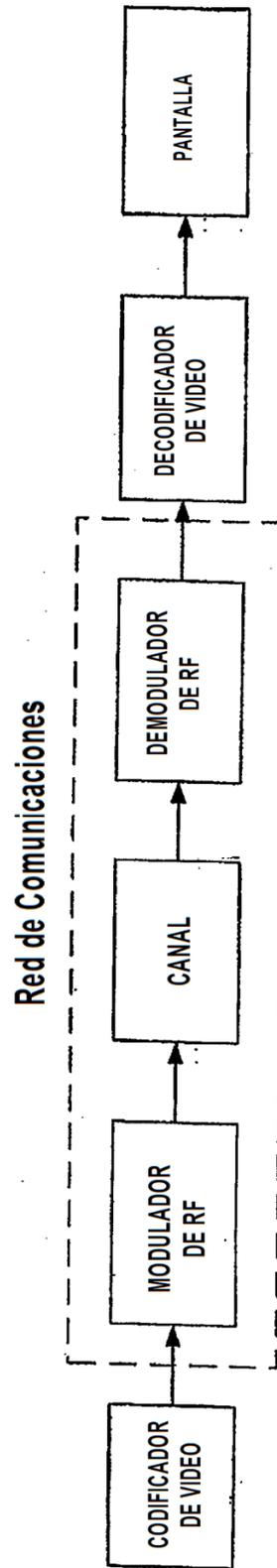


FIG. 19

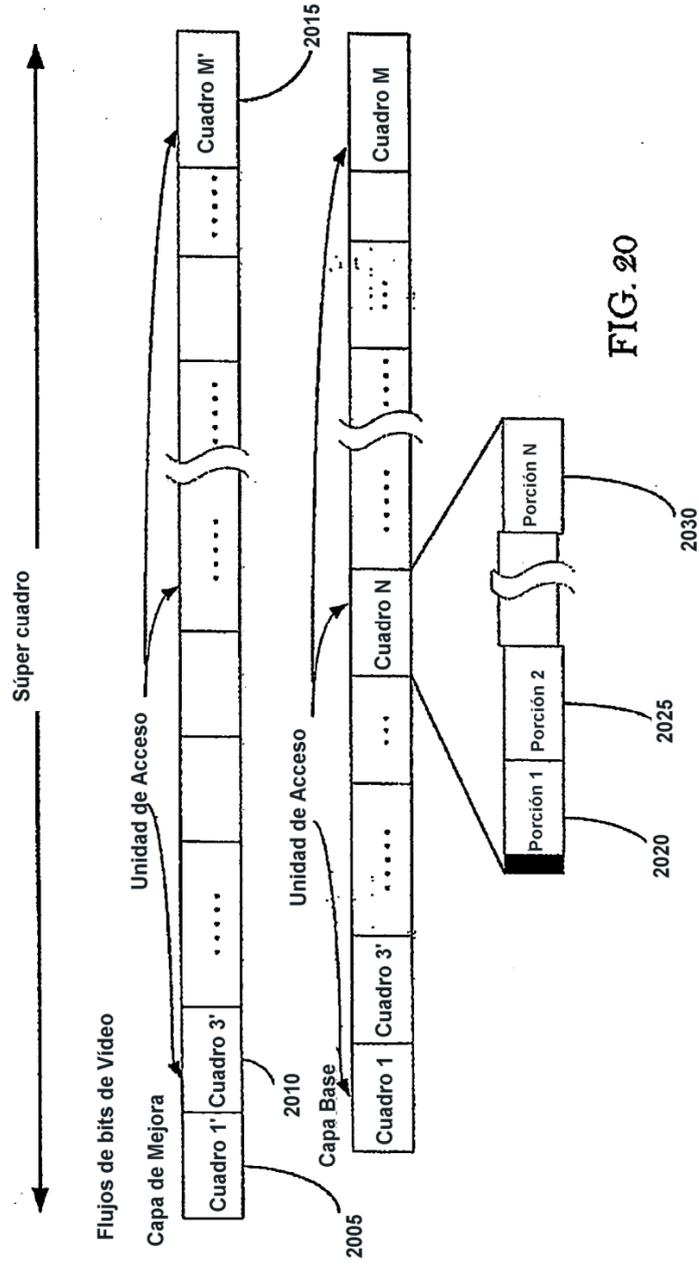


FIG. 20

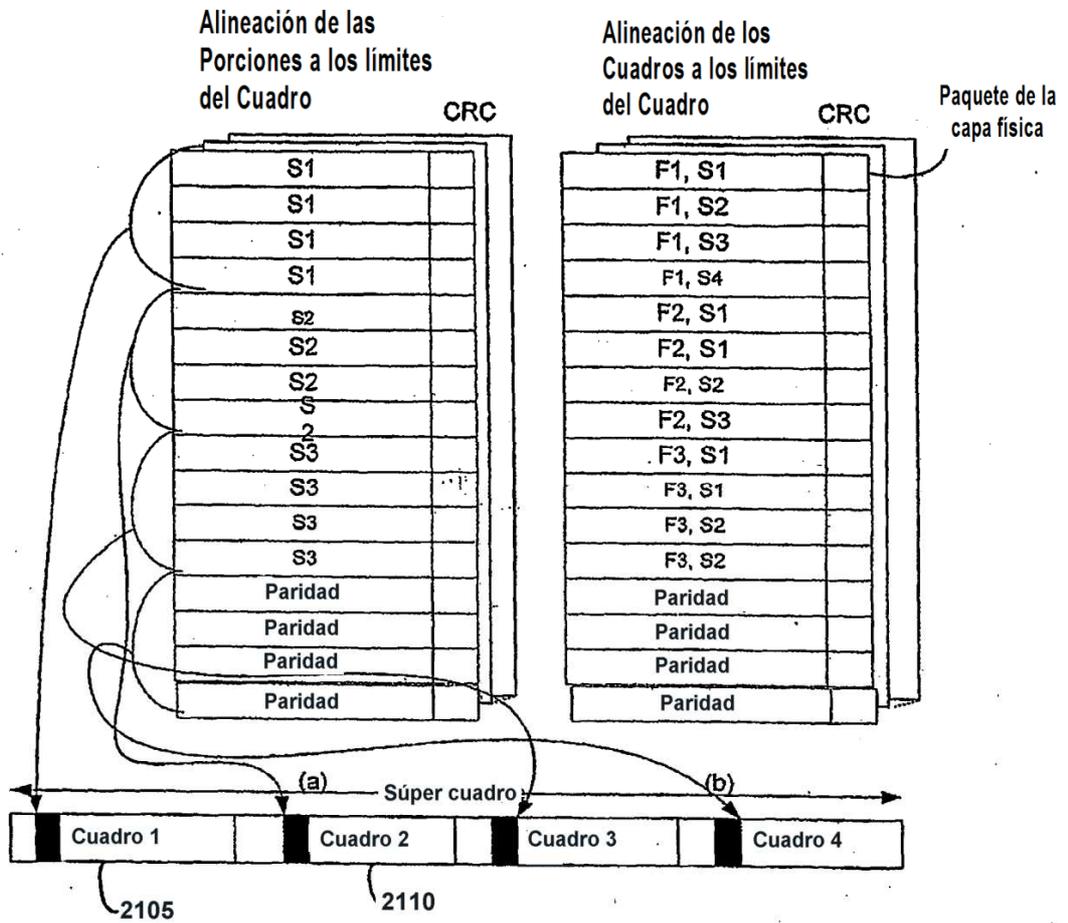


FIG. 21

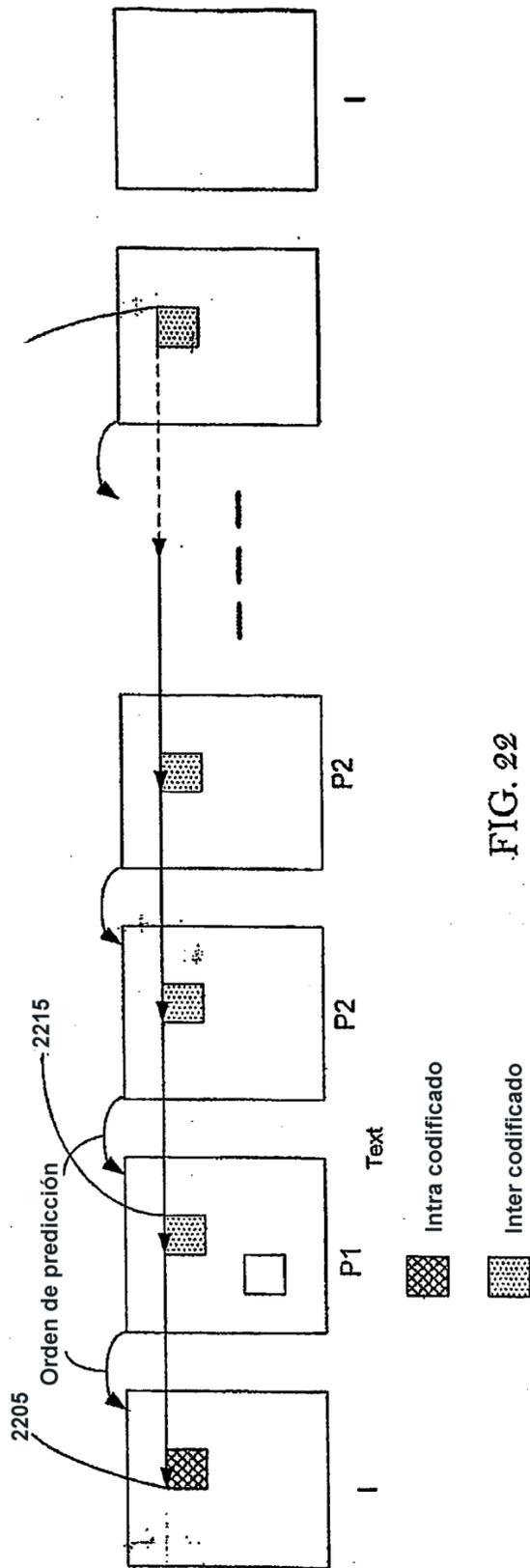


FIG. 22

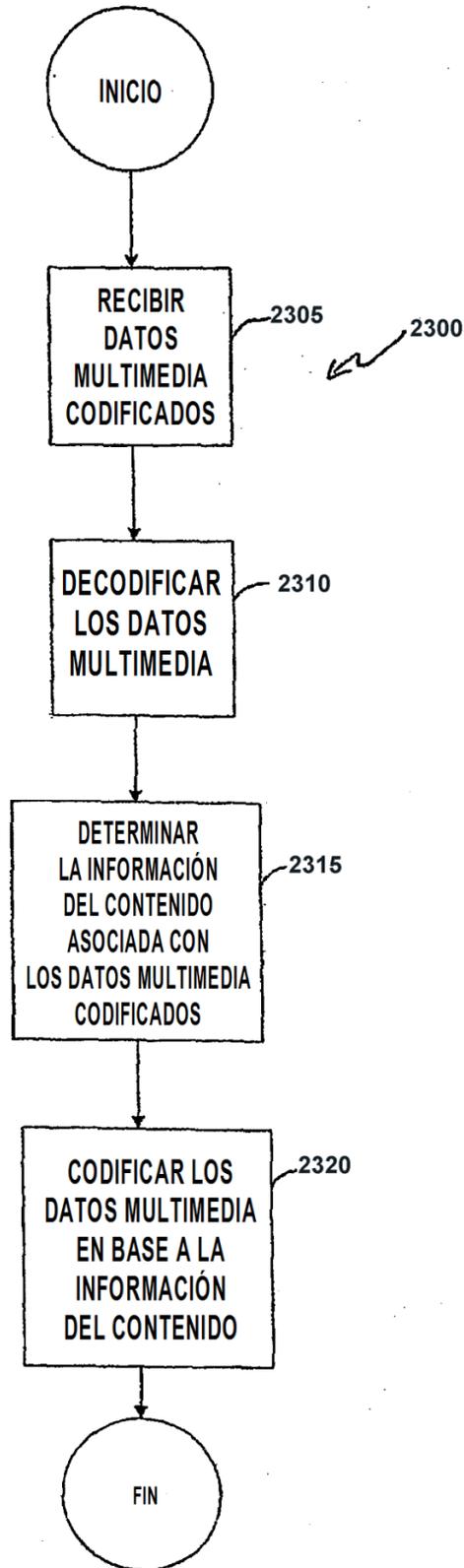


FIG. 23

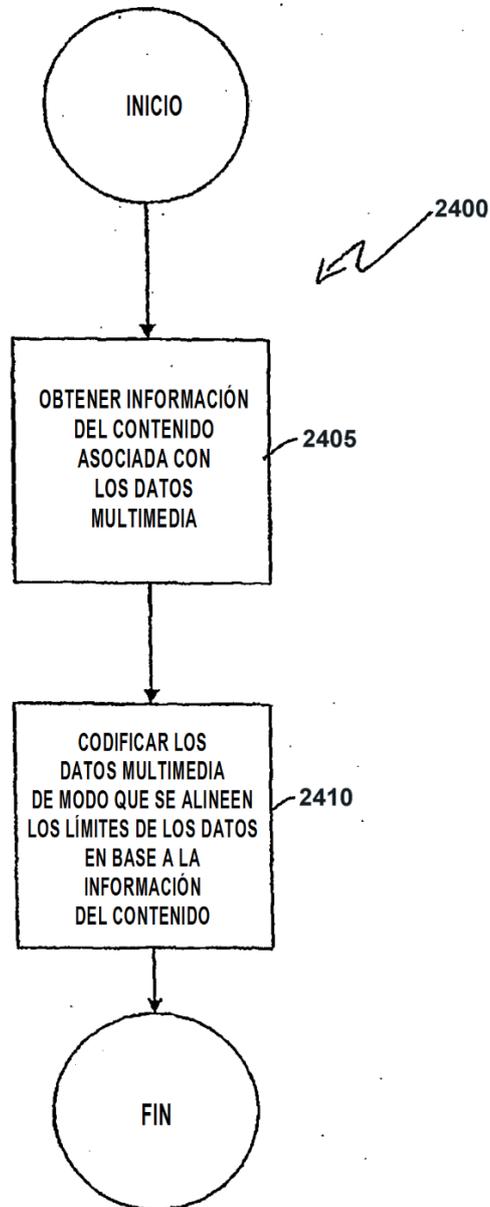


FIG. 24



2510

2505

FIG. 25



FIG. 26

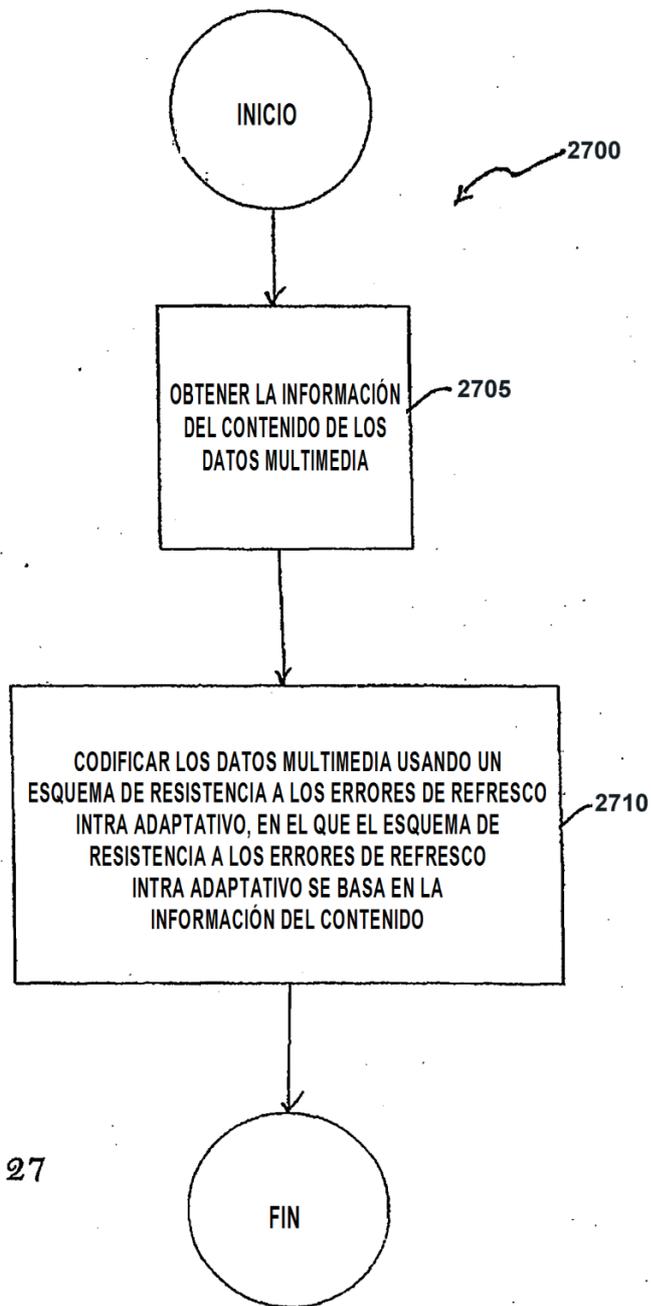


FIG. 27

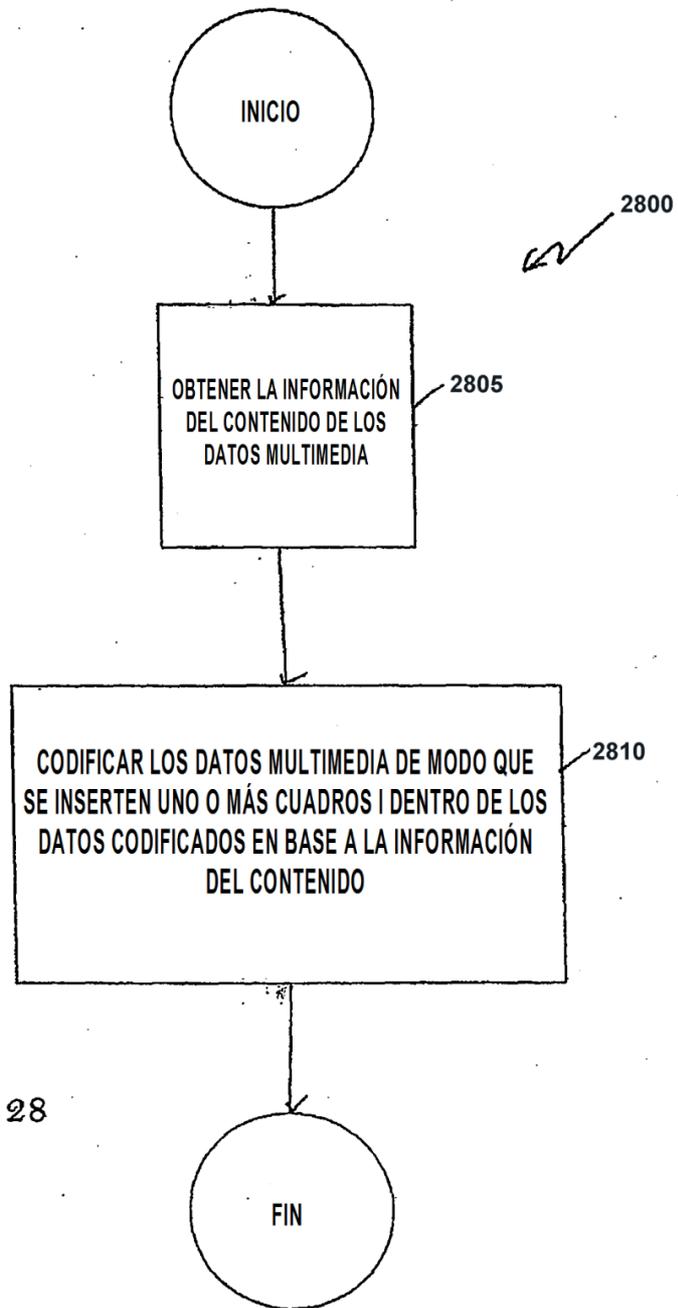


FIG. 28

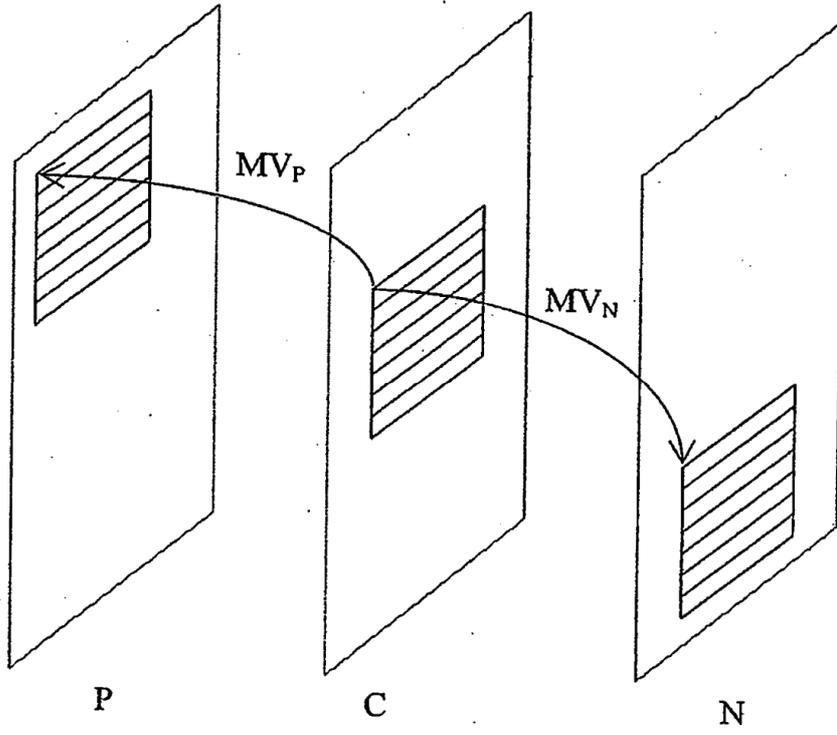


FIG. 29

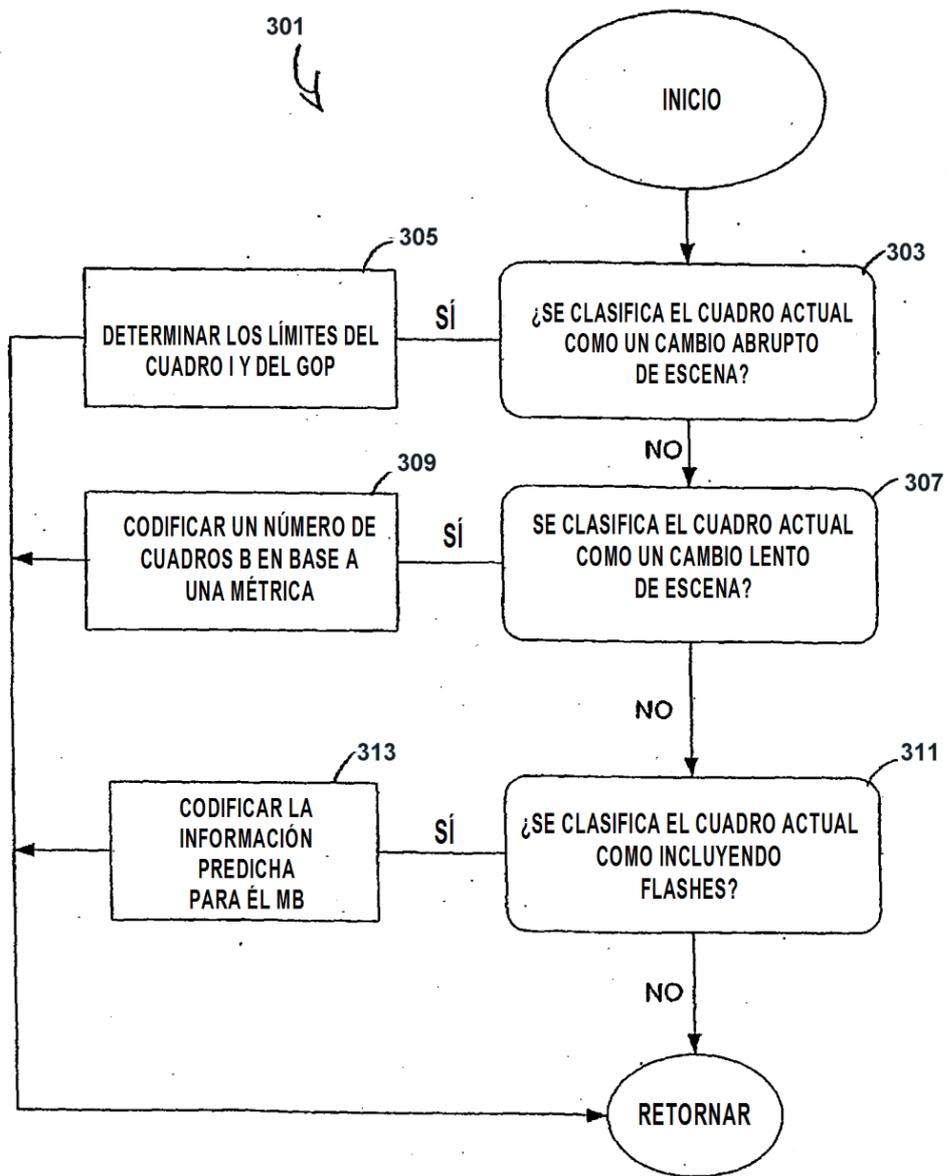


FIG. 30

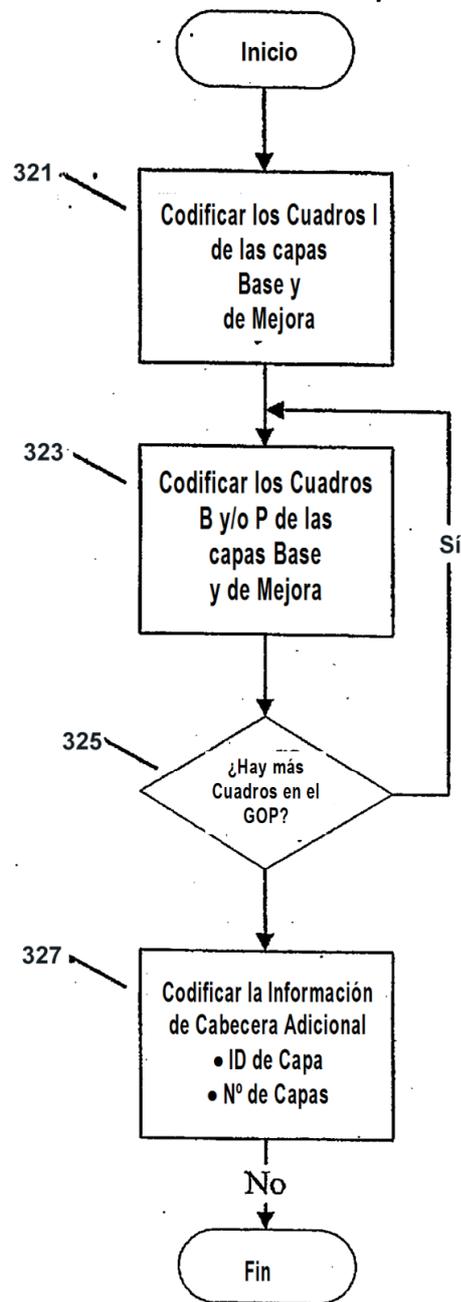


FIGURA 31

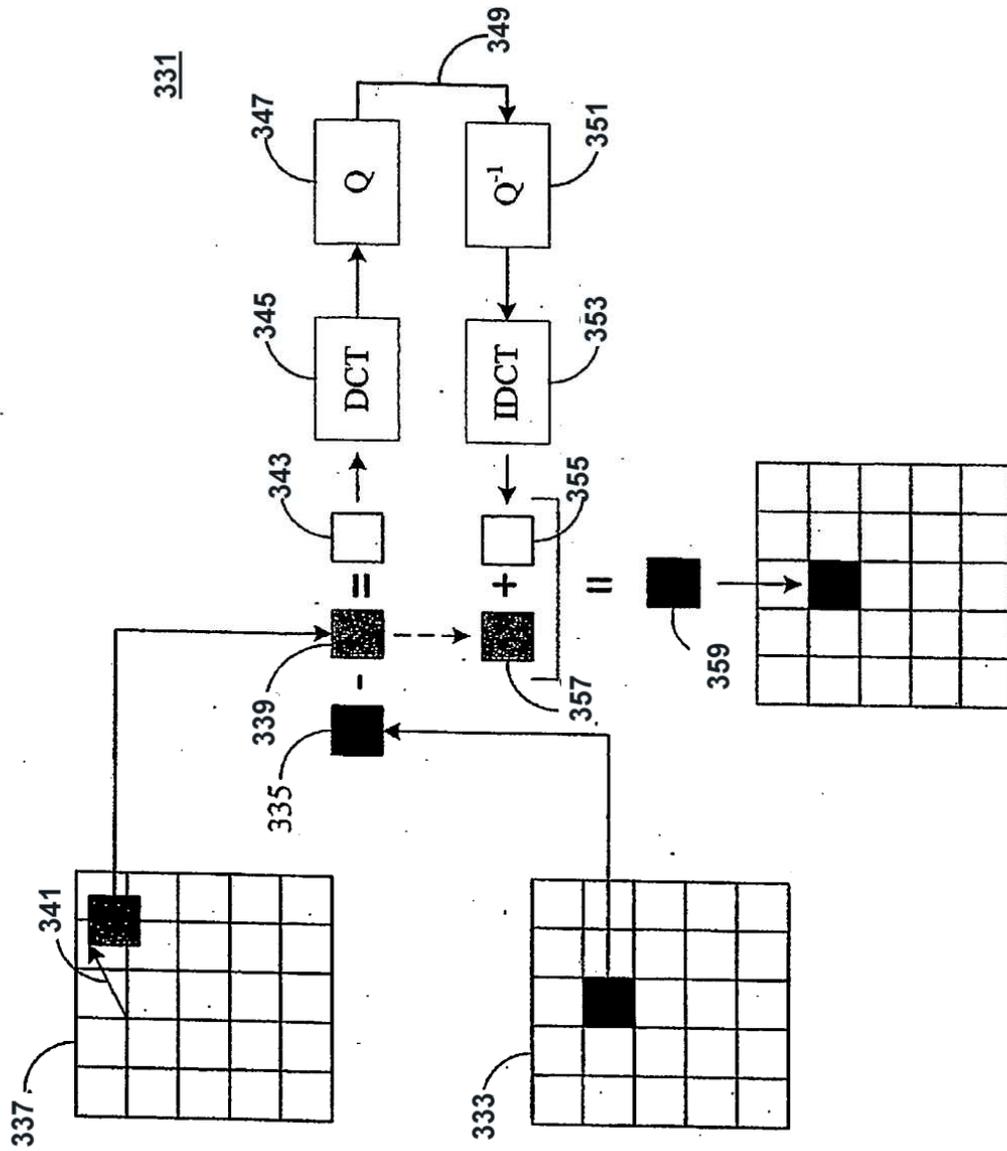


Figura 32

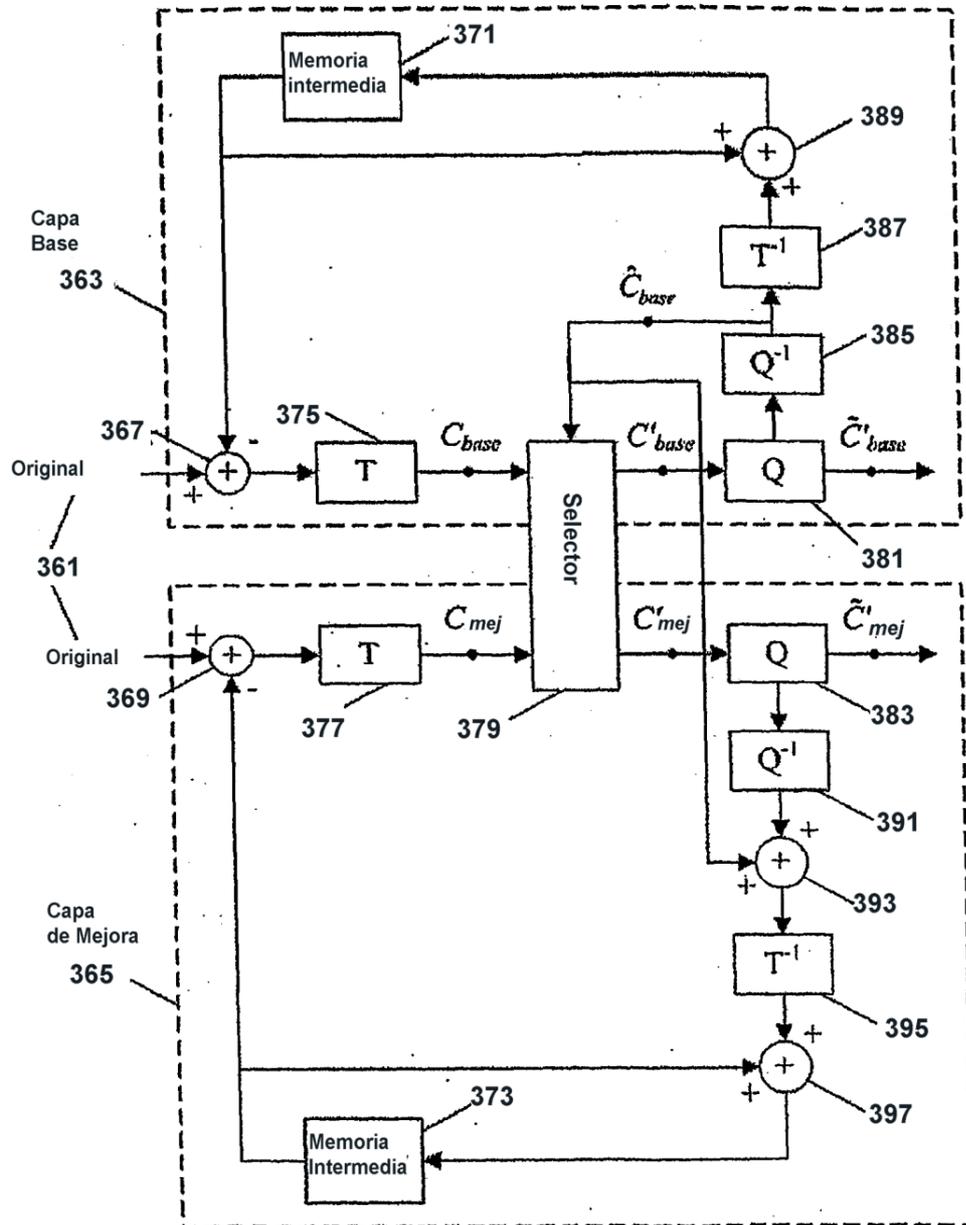


Figura 33

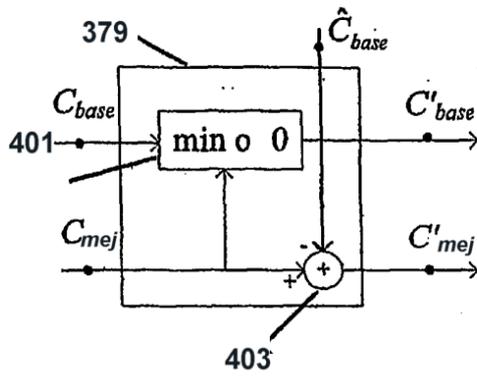


Figura 34

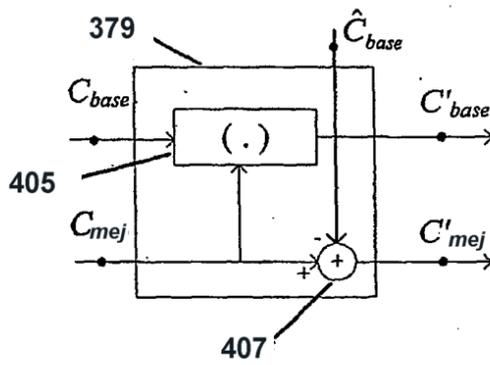


Figura 35

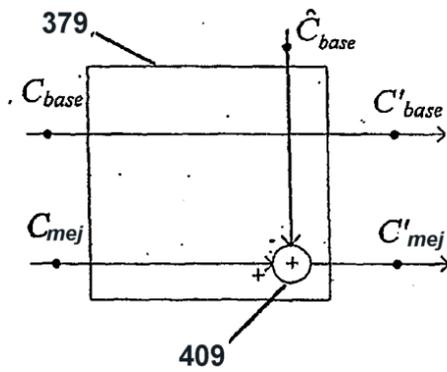


Figura 36

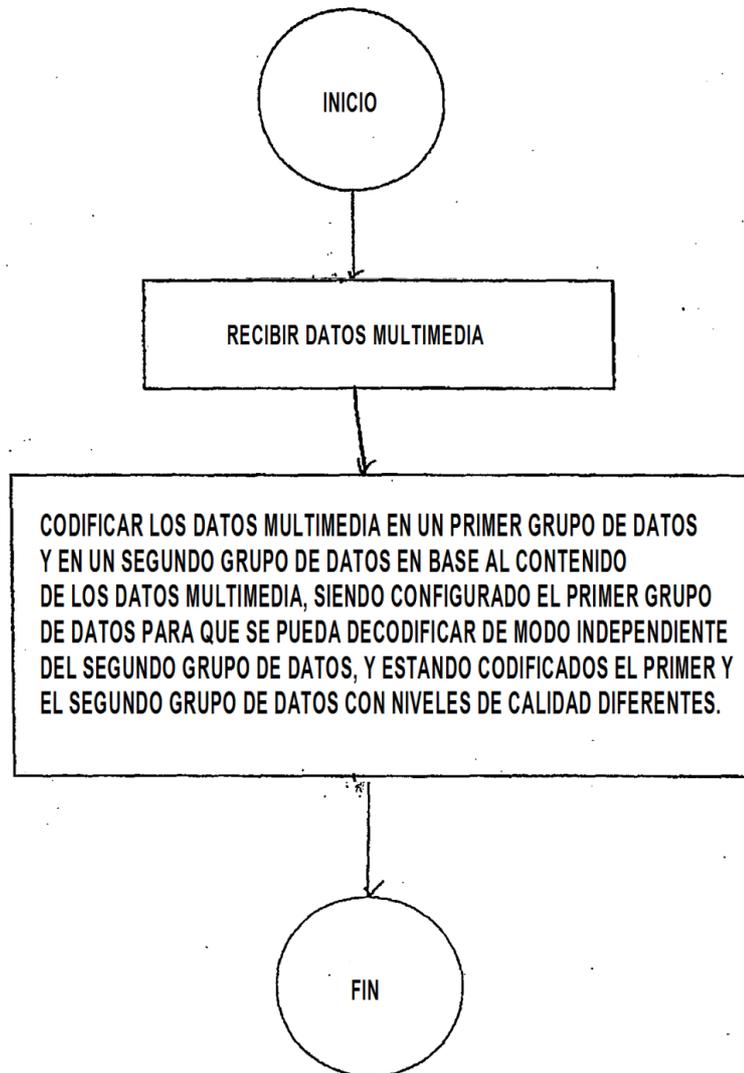
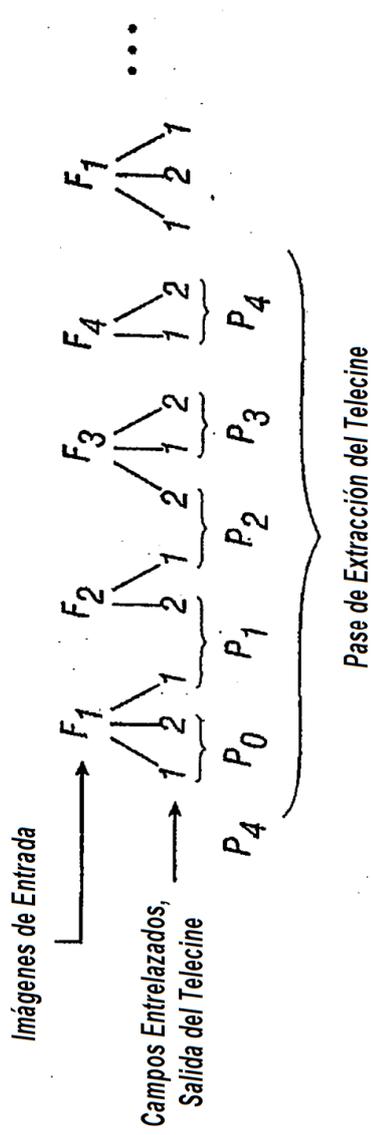
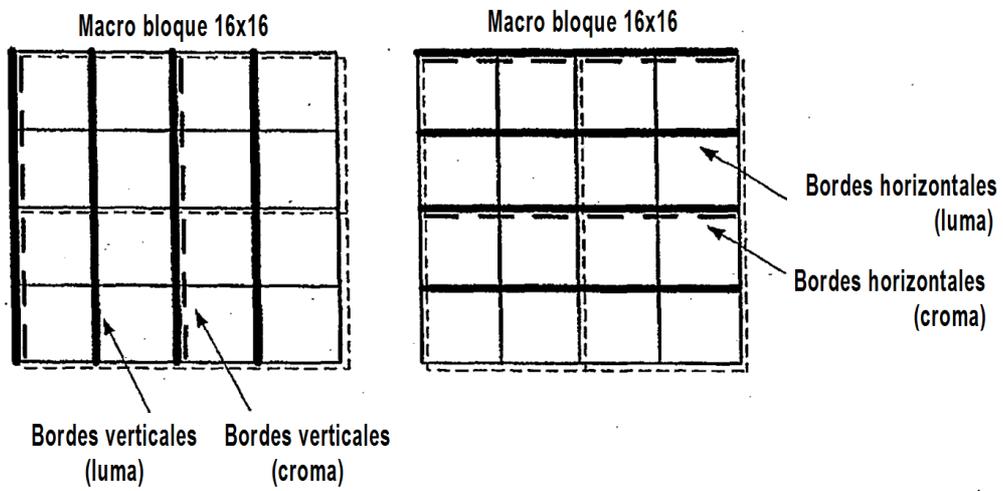


FIG. 37



Possible Decisions of the System

FIG. 38



Límites en un macro bloque a ser filtrados (los límites de luma se muestran con línea continua y los límites del croma se muestran con líneas discontinuas)

FIG. 39

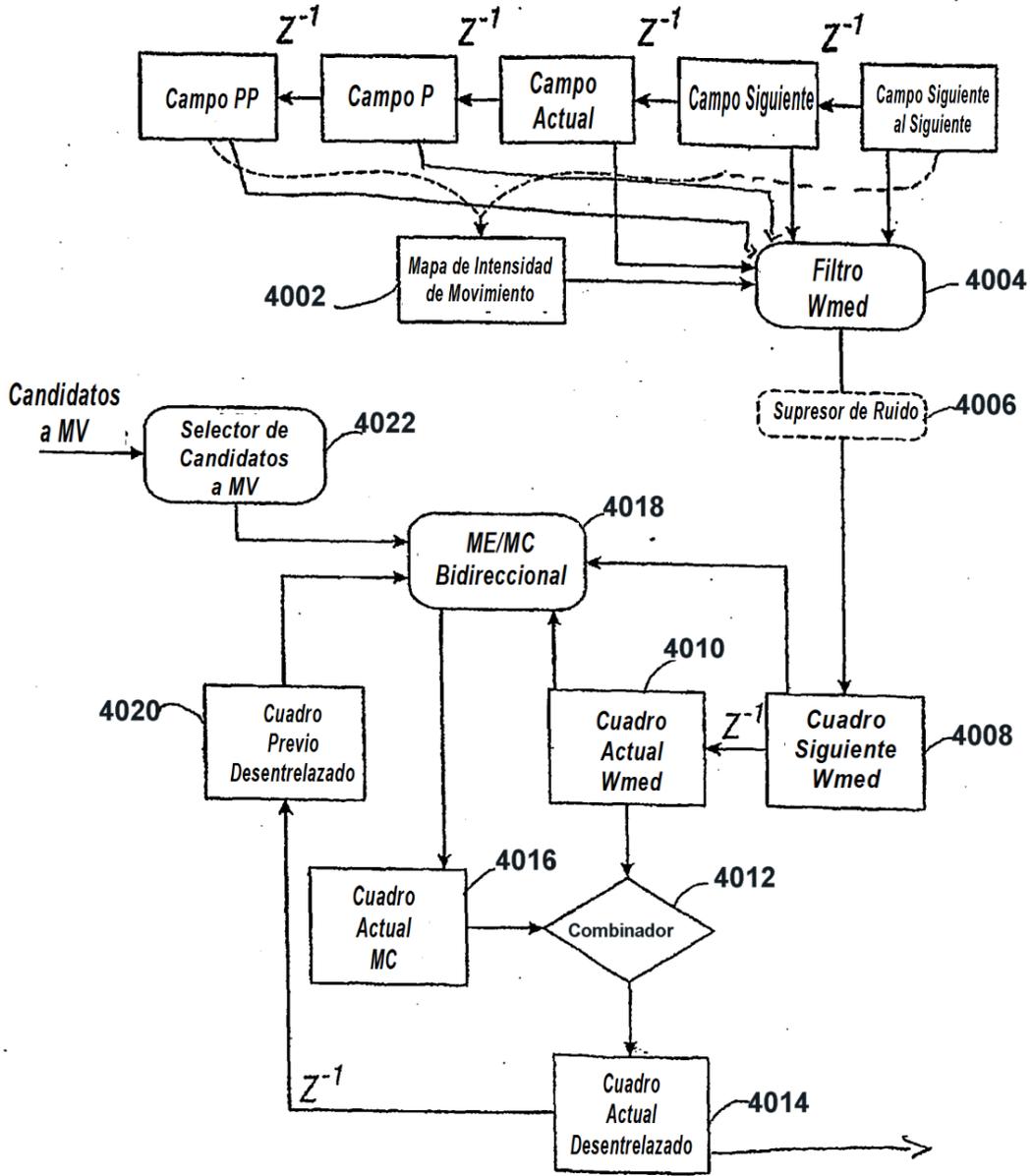
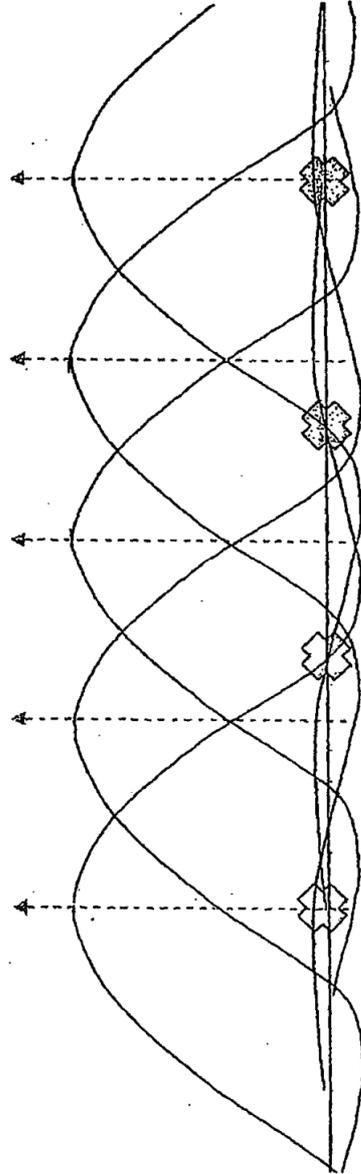


FIG. 40



Remuestreador polifásico 1-D

FIG. 41

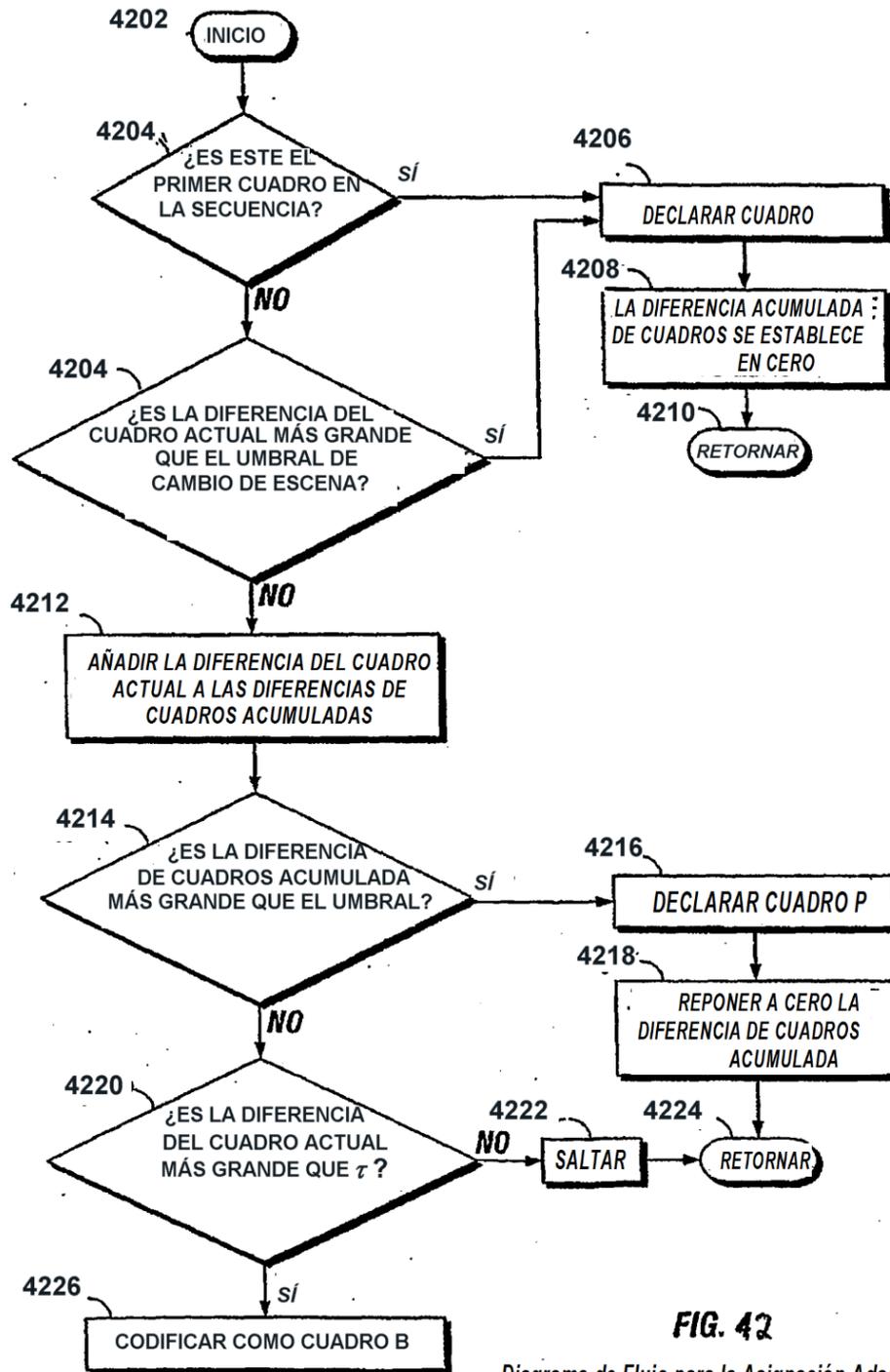


FIG. 42

Diagrama de Flujo para la Asignación Adaptativa de los Tipos de Cuadro a los Cuadros