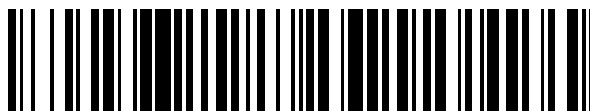


19



OFICINA ESPAÑOLA DE  
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 741 426**

51 Int. Cl.:

**H04N 19/147** (2014.01)  
**H04N 19/103** (2014.01)  
**H04N 19/154** (2014.01)  
**H04N 19/523** (2014.01)  
**H04N 19/895** (2014.01)  
**H04N 19/61** (2014.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

86 Fecha de presentación y número de la solicitud internacional: **19.10.2011 PCT/US2011/056835**

87 Fecha y número de publicación internacional: **26.04.2012 WO12054570**

96 Fecha de presentación y número de la solicitud europea: **19.10.2011 E 11774181 (9)**

97 Fecha y número de publicación de la concesión europea: **24.07.2019 EP 2630797**

54 Título: **Optimización de la tasa de distorsión resistente a errores para la codificación de imágenes y vídeos**

30 Prioridad:  
**20.10.2010 US 405025 P**

45 Fecha de publicación y mención en BOPI de la traducción de la patente:  
**11.02.2020**

73 Titular/es:  
**GUANGDONG OPPO MOBILE  
TELECOMMUNICATIONS CORP., LTD. (100.0%)  
No. 18 Haibin Road, Wusha, Chang'an  
Dongguan, Guangdong 523860, CN**

72 Inventor/es:  
**CHEN, ZHIFENG;  
PAHALAWATTA, PESHALA V. y  
TOURAPIS, ALEXANDROS**

74 Agente/Representante:  
**SÁEZ MAESO, Ana**

ES 2 741 426 T3

Aviso: En el plazo de nueve meses a contar desde la fecha de publicación en el Boletín Europeo de Patentes, de la mención de concesión de la patente europea, cualquier persona podrá oponerse ante la Oficina Europea de Patentes a la patente concedida. La oposición deberá formularse por escrito y estar motivada; sólo se considerará como formulada una vez que se haya realizado el pago de la tasa de oposición (art. 99.1 del Convenio sobre Concesión de Patentes Europeas).

**DESCRIPCIÓN**

Optimización de la tasa de distorsión resistente a errores para la codificación de imágenes y vídeos

5 Referencia cruzada con solicitudes relacionadas

Esta solicitud reivindica la prioridad de la Solicitud de Patente Provisional de Estados Unidos núm. 61/405,025 presentada el 20 de octubre de 2010.

10 Tecnología

Las modalidades de la presente invención se refieren generalmente a la codificación de datos de imágenes y vídeos. Más específicamente, las modalidades de la presente invención se refieren a la optimización de la tasa de distorsión resistente a errores para la codificación de imágenes y vídeos.

15 Antecedentes

Las aplicaciones de transmisión de imágenes y vídeos han aumentado en popularidad en los últimos años. Estas aplicaciones pueden abarcar desde software y dispositivos que permiten la transmisión de contenido bajo demanda a un usuario, para la transmisión de vídeos entre pares, para la emisión de vídeos en línea, así como también para compartir imágenes y vídeo dentro de una red doméstica del usuario. Estas aplicaciones pueden requerir la transmisión de imágenes y vídeos a través de redes cableadas así como también inalámbricas. Por ejemplo, la proliferación de dispositivos multimedia móviles, tales como teléfonos inteligentes que pueden grabar y reproducir datos de imagen y vídeo, ha generado un mayor interés en las tecnologías que permiten la transmisión inalámbrica de imágenes y vídeos de alta calidad, así como también en las tecnologías que permiten una conectividad perfecta entre tales dispositivos y otros dispositivos multimedia interconectados. Los nuevos desarrollos en vídeo, tal como la aparición de vídeo estereoscópico y multivista en 3D, así como también imágenes y vídeos de alto rango dinámico, también aumentarán la necesidad de una transmisión de imágenes y vídeos de alta calidad mientras que aumentarán los requisitos de ancho de banda. Por lo tanto, la capacidad de proporcionar imágenes y vídeos de alta calidad a través de redes de ancho de banda restringido es un factor importante para todas esas aplicaciones de transmisión.

Sin embargo, dos factores pueden afectar la capacidad de transmitir datos de imagen y vídeo de alta calidad. Son las limitaciones de ancho de banda y pérdidas aleatorias impredecibles en el canal.

35 Breve descripción de los dibujos

Los dibujos acompañantes, que se incorporan y constituyen una parte de esta especificación, ilustran una o más modalidades de la presente descripción y, junto con la descripción de las modalidades ilustrativas, sirven para explicar los principios y las implementaciones de la descripción.

40 La Figura 1 muestra un diagrama de bloques de un codificador de vídeo con un bloque de optimización de la tasa de distorsión.

La Figura 2 muestra un diagrama de bloques de un decodificador de vídeo.

45 La Figura 3 muestra un diagrama de bloques de un sistema de optimización de la tasa de distorsión que usa el cálculo analítico de los momentos de primer y segundo orden de distorsión de la transmisión.

Descripción de las modalidades ilustrativas

50 De acuerdo con un primer aspecto de la presente descripción, se proporciona un método para la optimización de distorsión de datos de imagen o de vídeo como en la reivindicación 1 independiente.

De acuerdo con un segundo aspecto de la presente descripción, se proporciona un codificador adaptado para recibir datos de imagen o de vídeo como en la reivindicación independiente 5.

55 La optimización de la tasa de distorsión resistente a errores (ERRDO) es un método que puede usarse para aliviar los efectos de las limitaciones de ancho de banda y las pérdidas aleatorias impredecibles en un canal. La optimización de la tasa de distorsión se usa en la compresión de vídeo para minimizar una estimación de una distorsión de vídeo sujeta a una restricción en la tasa de bits. Típicamente, esto se realiza a través de una selección cuidadosa de vectores de movimiento, modos de macrobloque, imágenes de referencia, etc., para reducir la redundancia que puede existir en un flujo de bits de manera que se ajustará a las restricciones prescritas. Sin embargo, una menor redundancia en el flujo de bits, aunque es relevante para la eficiencia de compresión, también reduce su resistencia a errores impredecibles.

60 Las técnicas de compresión de vídeo incluyen algunas técnicas para mejorar la resistencia a errores, tales como FMO (Ordenación flexible de los macrobloques), ASO (Ordenación arbitraria de slices) y slices redundantes. Sin embargo, estos han tenido un uso práctico limitado debido a la alta complejidad de la implementación, tanto en un codificador de vídeo como en un decodificador.

Un enfoque menos complejo para aumentar la resistencia a errores es el de la refresco intraaleatorio (por ejemplo, insertar aleatoriamente imágenes o macrobloques intracodificados en el flujo de bits) [ver referencia 1]. Esto reduce la dependencia de nuevas imágenes/macrobloques en imágenes/macrobloques previamente decodificados, y lo que reduce de esta manera la propagación de errores en un vídeo decodificado. Sin embargo, los bits adicionales usados para los macrobloques intracodificados aumentan el ancho de banda y, como resultado, también aumenta la susceptibilidad a los errores del flujo de bits.

Otros esquemas eligen regiones de una imagen o vídeo que tienen más probabilidades de verse afectados negativamente por errores, e insertan selectivamente bloques intracodificados solo en esas regiones. Además de la elección de bloques intracodificados, otras decisiones de modo, tal como una elección de una imagen de referencia (que incluye el uso de imágenes de referencia a largo plazo), que típicamente se realizan durante la optimización de la tasa de distorsión también pueden modificarse para considerar un efecto de las pérdidas de canal, lo que mejora de esta manera la resistencia a errores del flujo de bits mientras permanece dentro de un ancho de banda restringido. Además, también puede proporcionarse una protección contra errores desiguales a los paquetes de datos de origen que contienen información más importante, tal como en el caso de la codificación de vídeo escalable donde, para la decodificación, una capa puede depender de una capa previamente decodificada y, por lo tanto, más importante. Por lo tanto, una optimización de la tasa de distorsión resistente a errores puede formularse como:

$$\min_s J(s) = E[D(s,c)] + \lambda R(s), \quad (1)$$

donde  $s$  denota un modo seleccionado (en la presente descripción también denominado parámetros de origen) de acuerdo con el párrafo anterior,  $c$  denota las condiciones del canal,  $E[D(s,c)]$  denota la distorsión esperada dados los parámetros de origen seleccionados y las condiciones actuales del canal, y  $R(s)$  denota una tasa de bits dados los parámetros de origen seleccionados. El término "modo seleccionado" (parámetros de origen) se refiere a un método usado para la codificación (tal como la elección de bloques intracodificados, tamaño/forma de bloque, elección de imagen de referencia) y el término "condiciones del canal" se refiere a la calidad de un medio (tal como la tasa de error de paquete o bit) en la que el flujo de bits atraviesa desde un codificador a un decodificador. Las tasas de bits  $R$  pueden obtenerse como una tasa de bits verdadera de datos de origen de codificación, lo que puede aumentar la complejidad de un proceso de decisión de modo, o como una tasa de bits estimada.

Una diferencia entre la ecuación (1) y un esquema de optimización de la tasa de distorsión (RDO) convencional es un cálculo o estimación de una distorsión esperada dados los parámetros de origen y las condiciones del canal. Esta diferencia se debe a que la estimación de la distorsión ya no es determinista, sino que también depende de las pérdidas aleatorias durante la transmisión.

Los métodos para estimar la distorsión esperada en los procedimientos ERRDO se muestran en las [referencias 2-6]. El método propuesto en la [referencia 2] es un esquema basado en decodificadores hipotéticos múltiples que simula los errores de canal y emula un proceso de decodificación dentro de un codificador sin decodificar realmente un flujo de bits completo. Cada decodificador hipotético experimenta una realización de pérdida aleatoria diferente determinada por las condiciones del canal especificadas. Luego se usa una distorsión promedio en todos los decodificadores como una estimación de la distorsión esperada dados los parámetros de origen y del canal, y se usa en el marco de optimización de la tasa de distorsión.

El método propuesto en la [referencia 3] y mejorado aún más en la [referencia 4], que comúnmente se denomina como ROPE (Estimación de distorsión óptima recursiva por píxel), es un esquema recursivo menos intensivo en memoria que el esquema basado en múltiples decodificadores. Para calcular la distorsión esperada en un cuadro particular  $k$ , este método usa el primer y segundo momentos de los valores de píxeles reconstruidos en el decodificador para el cuadro  $k-1$  (asumiendo que el cuadro  $k$  depende del cuadro  $k-1$ ). El esquema puede extenderse a los casos en los que se usan múltiples cuadros de referencia [ver referencia 5].

El método en la [referencia 6] asume que la precisión de píxeles enteros se usa para la estimación de movimiento en el codificador, y que no se realiza ningún filtrado en el bucle. Los estándares de compresión de vídeo de última generación, tal como H.264/AVC [ver referencia 7] comúnmente usan interpolación de subpíxeles, filtrado en bucle, bipredicción, etc., donde no es óptimo asumir la precisión de píxeles enteros para la estimación de movimiento en el codificador.

Cuando se usan características con filtrado adicional sobre múltiples muestras de píxeles reconstruidos, los valores de muestras decodificadas después de la ocultación del error también se ven afectados por el filtrado, y esto puede tenerse en cuenta al estimar la distorsión.

La presente descripción propone una mejora a este esquema que extiende su aplicabilidad a los casos cuando el codificador usa la estimación de movimiento de subpíxeles, el filtrado en bucle, la bipredicción, etc., de manera que el método en la [referencia 6] también puede usarse en un marco de optimización de la tasa de distorsión resistente a errores.

Las Figuras 1-2 son diagramas de alto nivel de una arquitectura de codificador y decodificador de vídeo, respectivamente. En el diagrama del codificador mostrado en la Figura 1, los valores de muestra de entrada en la localización espacial  $u$ , y tiempo  $k$  se denotan  $f(u,k)$  (100). Una predicción residual en el codificador después de la transformación/cuantización y la cuantización/transformación inversa se denota  $e_c(u,k)$  (110). Una predicción residual es un residuo que se espera en el decodificador, como se predice durante una etapa de codificación. Este residuo se añade a una predicción temporal,  $f_c(u+mv(u,k), k-1)$  (120), donde  $mv(u,k)$  denota un vector de movimiento estimado, y luego se recorta y se filtra para obtener muestras de imágenes reconstruidas finales en el codificador, denotado  $f_c(u,k)$  (140). En un caso más general, un vecino temporal que no es necesariamente un cuadro anterior  $k-1$ , representado como  $j$  puede usarse como referencia para generar la predicción temporal. También se muestra en la Figura 1 un bloque de optimización de la tasa de distorsión (160), que luego se describirá con referencia a la Figura 3.

En el diagrama del decodificador mostrado en la Figura 2, se recibe un flujo de bits de entrada (230), que puede contener errores aleatorios o pérdidas de paquetes, y se realiza la decodificación de entropía en el flujo de bits. También puede realizarse una etapa de corrección de errores antes de la decodificación de entropía para recuperarse de algunas pérdidas en el canal. Cuando tales intentos de recuperación de pérdidas en el canal no son suficientes con la corrección de errores, una etapa de ocultación de errores intenta recuperar información perdida (por ejemplo, modos de macrobloque, vectores de movimiento, información residual, etc.) del flujo de bits. La ocultación de errores corrige los píxeles dañados en un cuadro determinado que usa la información obtenida de otros píxeles del mismo cuadro (ocultación de errores intracuadros) o un cuadro anterior (ocultación de errores intracuadros). La etapa de ocultación de errores puede ocurrir como un proceso posterior después de que se decodifiquen todas las partes de una imagen o región recibidas correctamente.

Un residuo oculto en el decodificador se denota  $e_d(u,k)$  (200), y los vectores de movimiento ocultos se denotan  $mv_d(u,k)$  (210). Debido a los errores en el flujo de bits, ambos valores pueden diferir de los valores correspondientes calculados en el codificador. Los vectores de movimiento ocultos se combinan con el cuadro  $k$ , desde un memoria intermedia de imagen de referencia (240) para crear una estimación temporal, denotada como  $f_d(u+mv_d(u,k), k-1)$  (250). La estimación temporal  $f_d(u+mv_d(u,k), k-1)$  (250) se añade al residuo oculto  $e_d(u,k)$  (200), luego se recorta y se filtra para obtener las muestras finales reconstruidas en el decodificador, denotada  $f_d(u,k)$  (220).

Dado que la distorsión es una medida de la calidad de una muestra en una salida, dada una entrada particular, la distorsión esperada de un píxel en la posición  $u$  en la imagen  $k$ , denotada por  $E[D(u,k)]$  puede calcularse como:

$$E[D(u,k)] = E\left[(f(u,k) - f_d(u,k))^2\right], \quad (2)$$

donde la distorsión se representa como una diferencia al cuadrado de la muestra  $f(u,k)$  en una entrada del codificador y la muestra  $f_d(u,k)$  en una salida del decodificador. La ecuación anterior se puede reescribir como:

$$E[D(u,k)] = E\left[\left\{(f(u,k) - f_c(u,k)) + (f_c(u,k) - f_d(u,k))\right\}^2\right], \quad (3)$$

al sumar y restar el término  $f_c(u,k)$  para obtener una diferencia entre la entrada y la salida del codificador en el lado izquierdo de la ecuación (3), y una diferencia entre la salida del codificador y la salida del decodificador en el lado derecho de la ecuación. La diferencia en el codificador puede reescribirse como una distorsión del codificador  $D_c(u,k)$  y la diferencia en el decodificador puede reescribirse como una distorsión del decodificador  $D_d(u,k)$ , que se muestra como:

$$E[D(u,k)] = E\left[\left(D_c(u,k) + D_d(u,k)\right)^2\right], \quad (4)$$

Al aplicar álgebra simple a la ecuación (4), obtenemos:

$$E[D(u,k)] = \left[D_c(u,k)\right]^2 + D_c(u,k)E\left[D_d(u,k)\right] + E\left[\left(D_d(u,k)\right)^2\right], \quad (5)$$

donde  $D_c(u,k)$  representa la distorsión en el codificador (causada por la cuantización, el recorte, etc.) y  $D_d(u,k)$  representa la distorsión de la transmisión en el decodificador (causada por errores de canal, recorte, etc.).  $D_c(u,k)$  es determinista y se conoce en el codificador, por lo tanto se elimina la "expectativa".

Basado en la ecuación (5), para calcular la distorsión esperada, se calculan primero  $E[D_d(u,k)]$  (primer momento) y  $E\left[\left(D_d(u,k)\right)^2\right]$  (segundo momento). Suponiendo que el píxel en la localización  $u$  de la imagen  $k$  se basa en la imagen  $j$  como

una referencia,  $E[D_d(u,k)]$  y  $E[(D_d(u,k))^2]$  puede calcularse como funciones de la probabilidad de pérdida de paquetes que contienen la información del vector de movimiento y residual relacionada con el píxel en la posición  $u$  y  $E[D_d(u+mv_d(u,k),j)]$ ,  $E[D_d(u+mv(u,k),j)]$ ,  $E[(D_d(u+mv_d(u,k),j))^2]$ , y  $E[(D_d(u+mv(u,k),j))^2]$ .

5 La base para tal suposición es proporcionada por la Figura 2, desde donde puede verse que la cantidad de estimación temporal proporcionada durante la ocultación del error es una medida de distorsión en el decodificador. En particular, una localización  $(u+mv_d(u,k), j)$  representa a las localizaciones de un píxel en la imagen reconstruida del decodificador  $j$  que se usaría para ocultar el píxel en la localización  $u$  en la imagen  $k$  con ocultación de errores si se pierde el paquete correspondiente. De manera similar,  $(u+mv(u,k), j)$  representa una localización del píxel en la imagen reconstruida del decodificador  $j$  que se usaría para predecir el píxel en la localización  $u$  en la imagen  $k$  si se recibe el paquete correspondiente.

Dado que la imagen  $j$  se usa como una referencia para la imagen  $k$ , la optimización de la tasa de distorsión resistente a errores ya se ha realizado en la imagen  $j$ , y luego codificado. Por lo tanto, en el caso de que los vectores de movimiento  $mv(u,k)$  y  $mv_d(u,k)$  se refieren a las localizaciones de píxeles enteros, los valores para  $E[D_d(u+mv_d(u,k),j)]$ ,  $E[D_d(u+mv(u,k),j)]$ ,  $E[(D_d(u+mv_d(u,k),j))^2]$ , y  $E[(D_d(u+mv(u,k),j))^2]$  ya se habrían calculado durante la optimización de la tasa de distorsión resistente a errores de la imagen  $j$ . Por lo tanto, los valores resultantes de  $E[D_d(u+mv_d(u,k),j)]$ ,  $E[D_d(u+mv(u,k),j)]$ ,  $E[(D_d(u+mv_d(u,k),j))^2]$ , y  $E[(D_d(u+mv(u,k),j))^2]$  pueden almacenarse y leerse de la memoria.

20 En códecs de vídeo avanzados, sin embargo,  $mv(u,k)$  y  $mv_d(u,k)$  pueden referirse a localizaciones de subpíxeles, y luego, los valores anteriores generalmente se vuelven a calcular durante la optimización de la tasa de distorsión de la imagen  $k$  en base a la información disponible de las localizaciones de píxeles enteros de la imagen  $j$ . Los cálculos de los primeros momentos,  $E[D_d(u+mv(u,k),j)]$  y  $E[D_d(u+mv_d(u,k),j)]$ , son relativamente sencillos incluso para los vectores de movimiento de subpíxeles. El cálculo de los segundos momentos (los valores al cuadrado), sin embargo, puede ser más complejo. La presente descripción describe un método para usar un enfoque de baja complejidad para el cálculo de los segundos momentos en el caso de subpíxeles.

Puede suponerse que una interpolación usada para generar un píxel en la posición de subpíxel  $u$  comprende un filtro de respuesta de impulso finito (FIR) de  $N$  taps aplicado en píxeles en localizaciones de píxeles enteros  $u_i$  ( $i=0, \dots, N-1$ ) en la imagen  $j$ . Para cada localización  $u_i$ , se aplica un peso  $w_i$ . Por lo tanto, el segundo momento de la distorsión esperada después de la compensación de movimiento puede escribirse como:

$$E[(D_d(u, j))^2] = E\left[\left(\sum_{i=0}^{N-1} w_i D_d(u_i, j)\right)^2\right], \quad (6a)$$

donde,  $w_i$  representa el tap del filtro en la localización de píxeles  $u_i$ . Escribiendo la ecuación (6) como:

$$E[X^2] = E\left[\left(\sum_{i=0}^{N-1} w_i X_i\right)^2\right], \quad (6b)$$

45 y suponiendo que  $\sum_{i=0}^{N-1} w_i = 1$ , se obtiene la siguiente expresión:

$$E[X^2] = \sum_{i=0}^{N-1} w_i E[(X_i)^2] - \sum_{i=0}^{N-2} \sum_{j=i+1}^{N-1} w_i w_j E[(X_i - X_j)^2], \quad (7).$$

50 De acuerdo con una modalidad de la presente descripción, puede suponerse que dado que  $i$  y  $j$  representan localizaciones de píxeles que están extremadamente cerca, las distorsiones representadas por  $X_i$  y  $X_j$  se correlacionan muy bien. Por lo tanto, el segundo término en el lado derecho de la ecuación (7) es insignificante. Esto permite que el cálculo del segundo momento se realice con la misma complejidad que el del primer momento, de manera que:

$$E[X^2] \approx \sum_{i=0}^{N-1} w_i E[(X_i)^2], \quad (8).$$

De acuerdo con otra modalidad, el término  $E[(X_i - X_j)^2]$  puede modelarse como una función (por ejemplo, una función lineal o una función exponencial) de la distancia entre las localizaciones de píxeles  $i$  y  $j$ , permitiendo de esta manera una estimación de baja complejidad de tal término.

5 En otra modalidad, el término puede calcularse con precisión, una vez para cada subpíxel, y luego almacenarse en la memoria. Cada vez que se usa un macrobloque del mismo cuadro como referencia en un cuadro futuro, el valor calculado almacenado en la memoria puede usarse para el cálculo de distorsión esperada, lo que reduce de esta manera la complejidad computacional de los cuadros futuros. Por lo tanto, aunque se realiza una vez un cálculo complejo, los cálculos futuros se simplifican.

10 Un diagrama de bloques como se muestra en la Figura 3 ilustra una manera en la que el método descrito anteriormente puede usarse dentro de un marco ERRDO, donde todo el proceso ilustrado puede producirse, por ejemplo, dentro del bloque de optimización de la tasa de distorsión (160) del codificador en la Figura 1. En la Figura 3, el bloque selector de modo (300) prueba todos los modos de codificación posibles (tal como intercodificación, intracodificación, varios tamaños/formas de bloque para la estimación de movimiento, etc.). Para cada modo, los bloques de cálculo (340, 350, 360, 370) calculan el primer y segundo momentos de distorsión de la transmisión para la imagen  $j$  (es decir, la imagen de referencia correspondiente para el modo seleccionado) y el bloque selector de modo (300) usa esos valores calculados del primer y segundo momentos de distorsión de la transmisión para obtener la distorsión estimada para la imagen actual. Para cada modo, también se calcula una tasa de bits estimada. El modo deseado se selecciona luego mediante el bloque selector de modo (300), por ejemplo, mediante el uso de un método tal como el que se muestra en la ecuación (1), y luego se usa para la codificación enviando parámetros de movimiento estimados del modo seleccionado desde el bloque selector de modo (300) a la unidad de estimación de movimiento (170) y la unidad de compensación de movimiento (180) de la Figura 1. El primer y segundo momentos calculados de la distorsión para el modo seleccionado también se almacenan en la memoria intermedia del primer momento (320) y las memorias intermedias del segundo momento (330) para su uso en la codificación de imágenes posteriores. Los modos que no se seleccionan se descartan.

De acuerdo con las modalidades adicionales de la presente descripción, la ecuación (7) también puede aplicarse en casos distintos a la interpolación de subpíxeles, tal como, por ejemplo, la predicción ponderada o el desbloqueo ponderado, para determinar la distorsión esperada de una muestra filtrada linealmente en función de la distorsión esperada de las muestras usadas para el filtrado. En general, la ecuación (7) y sus simplificaciones descritas anteriormente también pueden usarse en otras aplicaciones que podrían beneficiarse del cálculo simplificado de un segundo momento de una variable aleatoria que es una suma ponderada de un conjunto de variables aleatorias.

35 En conclusión, las modalidades de la presente descripción extienden un método existente (ecuación 6) para estimar la distorsión de la transmisión de vídeo en una forma (ecuación 7) que puede usarse con un proceso de decisión de modo resistente a errores. Estas modalidades también mejoran la aplicabilidad del algoritmo de estimación de distorsión a los estándares de codificación de vídeo de última generación.

Los métodos y sistemas descritos en la presente descripción pueden implementarse en hardware, software, microprogramas o sus combinaciones. Las características descritas como bloques, módulos o componentes pueden implementarse juntas (por ejemplo, en un dispositivo lógico tal como un dispositivo lógico integrado) o por separado (por ejemplo, como dispositivos lógicos conectados por separado). La parte de software de los métodos de la presente descripción puede comprender un medio legible por ordenador que comprende instrucciones que, cuando se ejecutan, realizan, al menos en parte, los métodos descritos. El medio legible por ordenador puede comprender, por ejemplo, una memoria de acceso aleatorio (RAM) y/o una memoria de solo lectura (ROM). Las instrucciones pueden ejecutarse por un procesador (por ejemplo, un procesador de señal digital (DSP), un circuito integrado específico de la aplicación (ASIC) o una matriz lógica de campo programable (FPGA)).

Los ejemplos expuestos anteriormente se proporcionan para dar a los expertos en la técnica una descripción completa y una descripción de cómo hacer y usar las modalidades de los métodos y sistemas para ERRDO para la codificación de imágenes y vídeos de la descripción, y no pretenden limitar el alcance de lo que los inventores consideran como su descripción. Los expertos en la técnica pueden usar las modificaciones de los modos descritos anteriormente para llevar a cabo la descripción.

55 Se entenderá que esta descripción no se limita a métodos o sistemas particulares, que pueden, por supuesto, variar. Se debe también entender que la terminología usada en la presente descripción es para propósitos de describir las modalidades particulares solamente, y no pretende ser limitativa. Como se usa en esta descripción y en las reivindicaciones adjuntas, las formas singulares "un", "una", "el/la" incluyen referentes plurales a menos que el contexto lo indique claramente de cualquier otra manera. El término "pluralidad" incluye dos o más referentes a menos que el contenido lo indique claramente de cualquier otra manera. A menos que se defina de cualquier otra manera, todos los términos técnicos y científicos que se usan en la presente descripción tienen el mismo significado que el que se conoce comúnmente por los expertos en la técnica a la que pertenece esta descripción.

Se han descrito varias modalidades de la descripción. Sin embargo, se debe entender que pueden hacerse varias modificaciones. El alcance de la invención se define por las reivindicaciones adjuntas.

65

Lista de referencias

- [1] T. Stockhammer, MM Hannuksela y T. Wiegand, "H.264/AVC in Wireless Environments," IEEE Trans. on Circuits and Systems for Video Technology, vol. 13, núm. 7, pág. 657-673, julio de 2003.
- 5 [2] T. Stockhammer, T. Wiegand y S. Wenger, "Optimized Transmission of H.26L/JVT Coded Video Over Packet-Lossy Networks," Proc. of the International Conference on Image Processing (ICIP), 2002.
- [3] R. Zhang, SL Regunathan y K. Rose, "Video Coding with Optimal Intra/Inter-Mode Switching for Packet Loss Resilience," IEEE Journal on Selected Areas in Communications, vol. 18, núm 6, pág. 966-976, junio de 2000.
- 10 [4] H. Yang y K. Rose, "Advances in Recursive Per-Pixel End-to-End Distortion Estimation for Robust Video Coding in H.264/AVC," IEEE Trans. on Circuits and Systems for Video Technology, vol. 17, núm. 7, pág. 845-856, julio de 2007.
- [5] A. Leontaris y PC Cosman, "Video Compression for Lossy Packet Networks with Mode Switching and a Dual-Frame Buffer," IEEE Trans. on Image Processing, vol. 13, núm. 7, pág. 885-897, julio de 2004.
- [6] Z. Chen, D. Wu, "Study of video distortion for real-time video transmission over wireless channels," Proceedings of SPIE Defense & Security Symposium 2009, Orlando, FL, USA, 13 al 17 de abril de 2009.
- 15 [7] Advanced video coding for generic audiovisual services,  
<http://www.itu.int/rec/recommendation.asp?type=folders&lang=e&parent=T-REC-H.264>, marzo de 2009.

REIVINDICACIONES

1. Un método para la optimización de la distorsión de datos de imagen o de vídeo mientras se codifican los datos de imagen o de vídeo, que comprende:

proporcionar datos de imagen o de vídeo;  
 evaluar una distorsión esperada de los datos de imagen o de vídeo codificados en función de distorsión de la transmisión  $D_d(u, j)$  de los datos de imagen o de vídeo, en donde dicha distorsión de la transmisión representa la diferencia entre una muestra de imagen o de vídeo reconstruida a la salida del codificador y una muestra de imagen o de vídeo reconstruida a la salida del decodificador para una posición de subpíxel  $u$  y un índice de imágenes  $j$ ; y  
 realizar la optimización de la tasa de distorsión resistente a errores de los datos de imagen o de vídeo para la selección del modo de codificación mientras se codifican los datos de imagen o de vídeo, caracterizado porque:

evaluar la distorsión esperada incluye calcular un segundo momento aproximado de  $D_d(u, j)$  basado en la siguiente ecuación:

$$E[X^2] \approx \sum_{i=0}^{N-1} w_i E[(X_i)^2]$$

en donde:

$X = D_d(u, j)$ ,  
 $X_i = D_d(u_i, j)$ , y

$w_i$  denota un parámetro de peso para un filtro de respuesta de impulso finito de  $N$  taps usado para la interpolación de valores correspondientes a las posiciones de subpíxeles  $u$  de los valores correspondientes a las posiciones de píxeles enteros  $u_i$  en donde

$$\sum_{i=0}^{N-1} w_i = 1$$

y

$N$  denota un número entero de las localizaciones de píxeles enteros  $u_i$ .

2. El método como se mencionó en la reivindicación 1, en donde los datos de imagen o de vídeo de salida después de la decodificación comprenden una variable aleatoria desconocida.
3. El método como se mencionó en la reivindicación 1, en donde distorsión de la transmisión comprende una función de pérdidas de canal.
4. El método como se mencionó en cualquiera de las reivindicaciones 1 a 3, en donde la evaluación de distorsión esperada comprende evaluar la distorsión esperada en términos de un término lineal y un término cuadrático, cada uno del término lineal y del término cuadrático que expresa la distorsión esperada de los datos de imagen o de vídeo.
5. Un codificador adaptado para recibir datos de imagen o de vídeo, el codificador que comprende:

una entrada para recibir los datos de imagen o de vídeo;  
 una unidad de memoria intermedia de imagen de referencia (150), la unidad de memoria intermedia de imagen de referencia que contiene los datos de imagen o de vídeo almacenados;  
 una unidad de optimización de la tasa de distorsión (160), la unidad de optimización de la tasa de distorsión adaptada para recibir los datos de imagen o de vídeo desde la entrada, para realizar la optimización de distorsión de datos de imagen o de vídeo mientras codifica los datos de imagen o de vídeo que comprende:

evaluar una distorsión esperada de los datos de imagen o de vídeo codificados como una función de distorsión de la transmisión  $D_d(u, j)$  de los datos de imagen o de vídeo, en donde dicha distorsión de la transmisión representa la diferencia entre una muestra de imagen o de vídeo reconstruida a la salida del codificador y una muestra de imagen o de vídeo reconstruida a la salida del decodificador para una posición de subpíxel  $u$  y un índice de imágenes  $j$   
 realizar la optimización de la tasa de distorsión resistente a errores de los datos de imagen o de vídeo para la selección del modo de codificación mientras se codifican los datos de imagen o de vídeo, en donde evaluar la distorsión esperada incluye calcular un segundo momento aproximado de  $D_d(u, j)$  basado en la siguiente ecuación:



$$E[X^2] \approx \sum_{i=0}^{N-1} w_i E[(X_i)^2]$$

en donde:

$$X = D_d(u, j),$$

$$X_i = D_d(u_i, j) \text{ y}$$

$w_i$  denota un parámetro de peso para un filtro de respuesta de impulso finito de  $N$  taps usado para la interpolación de valores correspondientes a las posiciones de subpíxeles  $u$  o de valores correspondientes a las posiciones de píxeles enteros  $u_i$  en donde

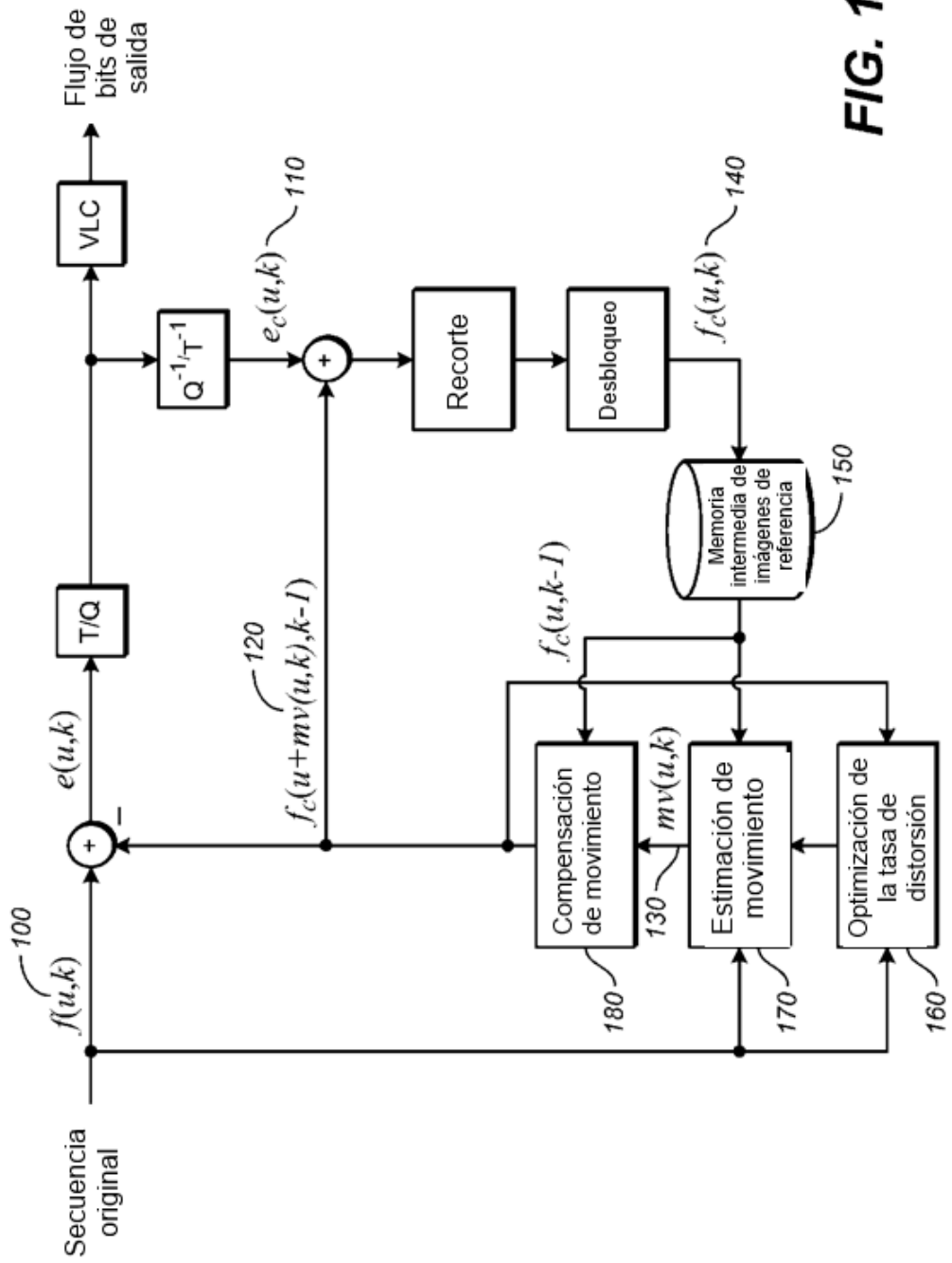
$$\sum_{i=0}^{N-1} w_i = 1$$

y

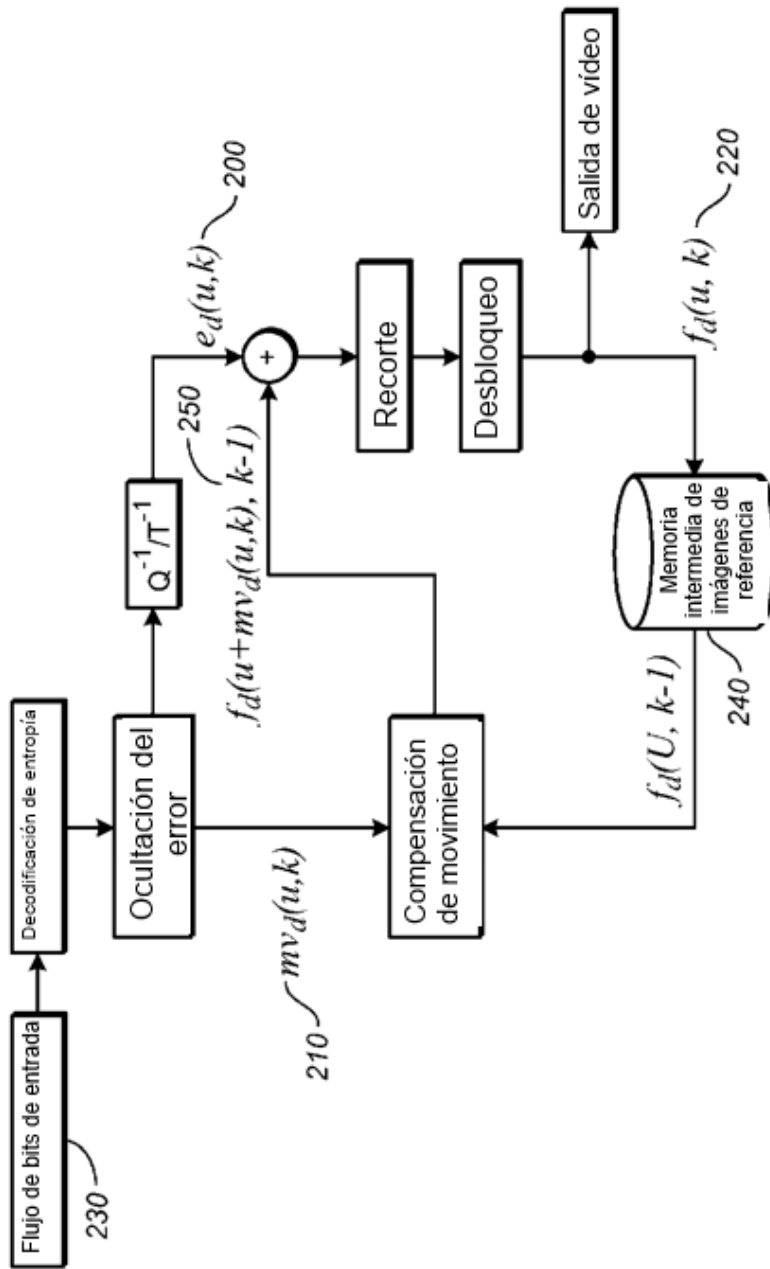
$N$  denota un número entero de las localizaciones de píxeles enteros  $u_i$ ;

una unidad de estimación de movimiento (170), la unidad de estimación de movimiento adaptada para recibir los datos de imagen o de vídeo desde la entrada, el modo de codificación seleccionado con sus parámetros de movimiento estimados de la unidad de optimización de la tasa de distorsión (160) y adaptada para recibir los datos de imagen o de vídeo almacenados desde la unidad de memoria intermedia de imágenes de referencia (150); una unidad de compensación de movimiento (170), la unidad de compensación de movimiento adaptada para recibir los datos de imagen o de vídeo desde la unidad de estimación de movimiento (170) y los datos de imagen o de vídeo almacenados desde la unidad de almacenamiento de imágenes de referencia (150); una unidad de transformación; y una unidad de cuantización, la unidad de cuantización y la unidad de transformación adaptadas para recibir los datos de imagen o de vídeo combinados desde la entrada y la unidad de compensación de movimiento (180).

6. El codificador como se mencionó en la reivindicación 5, en donde la unidad de optimización de la tasa de distorsión (160) se adapta para recibir retroalimentación desde la unidad de compensación de movimiento (180).
7. El codificador como se mencionó en la reivindicación 5 o la reivindicación 6, en donde la unidad de optimización de la tasa de distorsión (160) comprende además:
  - un primer momento de la unidad de cálculo de distorsión de la transmisión;
  - un primer momento de la unidad de memoria intermedia de distorsión de la transmisión (310), el primer momento de la unidad de memoria intermedia de transmisión adaptado para recibir un primer momento calculado del valor de distorsión de la transmisión desde el primer momento de la unidad de cálculo de distorsión de la transmisión;
  - un segundo momento de la unidad de cálculo de distorsión de la transmisión;
  - un segundo momento de la unidad de memoria intermedia de distorsión de la transmisión (330), el segundo momento de la unidad de memoria intermedia de distorsión de la transmisión adaptado para recibir un segundo momento calculado del valor de distorsión de la transmisión desde el segundo momento de la unidad de distorsión de la transmisión; y
  - una unidad selectora de modo (300), la unidad selectora de modo adaptada para recibir el primer y segundo momentos calculados del valor de distorsión de la transmisión, la entrada de datos de imagen o de vídeo y la retroalimentación de la unidad de compensación de movimiento (180).
8. El codificador como se mencionó en cualquiera de las reivindicaciones 5 a 7, en donde uno o más codificadores o la unidad de optimización de la tasa de distorsión se adaptan para reducir la distorsión de la transmisión en un decodificador.
9. Un medio de almacenamiento legible por ordenador que contiene un conjunto de instrucciones que hace que un ordenador realice el método como se mencionó en cualquiera de las reivindicaciones 1-4, cuando las instrucciones se ejecutan en el ordenador.



**FIG. 1**



**FIG. 2**

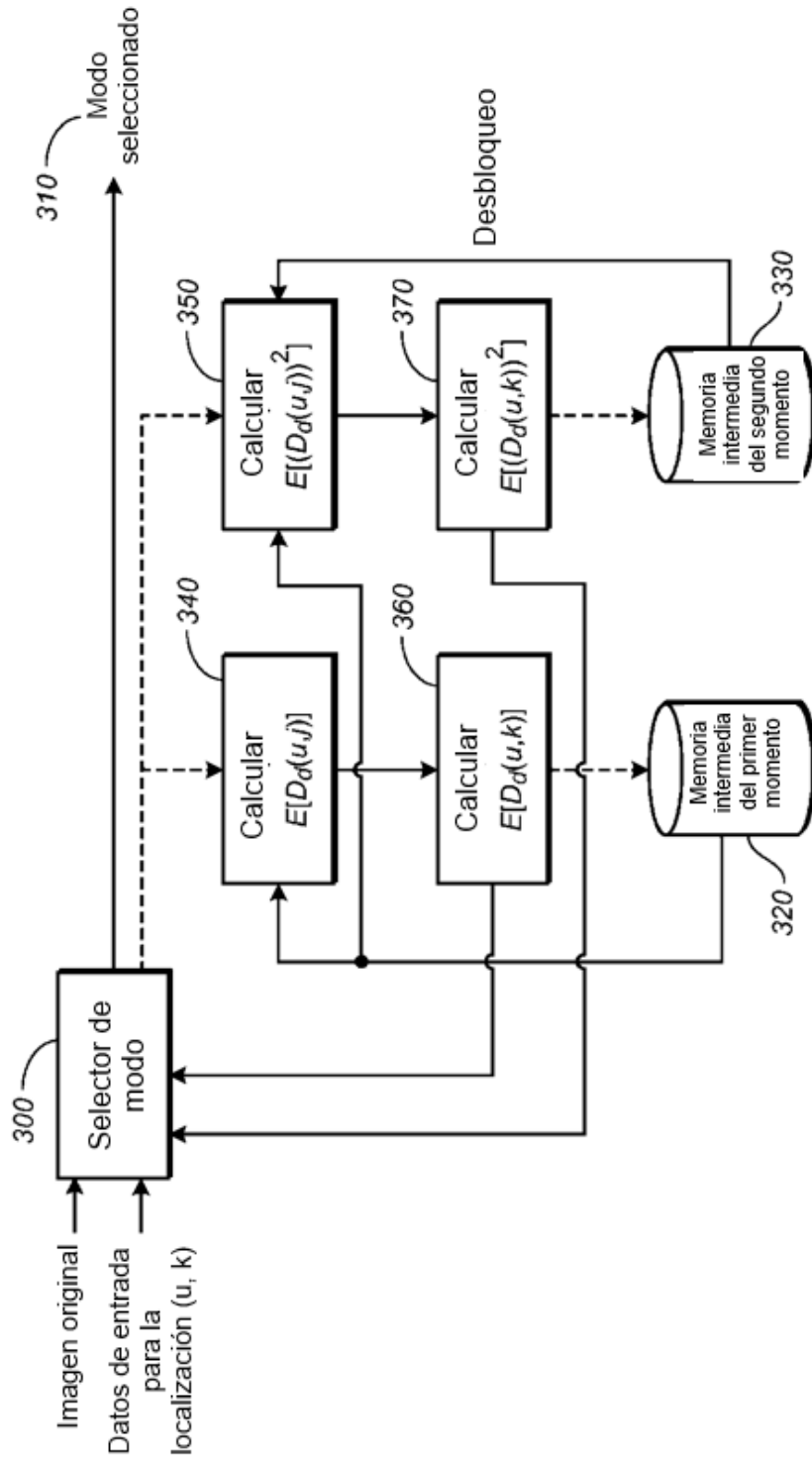


FIG. 3