

19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 526 080**

51 Int. Cl.:

H04N 19/154 (2014.01)

H04N 19/177 (2014.01)

H04N 19/48 (2014.01)

H04N 7/64 (2006.01)

H04B 17/00 (2006.01)

H04N 17/00 (2006.01)

H04N 21/442 (2011.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

96 Fecha de presentación y número de la solicitud europea: **20.08.2012 E 12181015 (4)**

97 Fecha y número de publicación de la concesión europea: **08.10.2014 EP 2701397**

54 Título: **Modelo de calidad de vídeo dependiente del contenido para servicios de transmisión de vídeo**

45 Fecha de publicación y mención en BOPI de la traducción de la patente:
05.01.2015

73 Titular/es:

**DEUTSCHE TELEKOM AG (100.0%)
Friedrich-Ebert-Allee 140
53113 Bonn, DE**

72 Inventor/es:

**GARCIA, MARIE-NEIGE;
RAAKE, ALEXANDER;
ARGYROPOULOS, SAVVAS;
FEITEN, BERNHARD y
LIST, PETER**

74 Agente/Representante:

DE ELZABURU MÁRQUEZ, Alberto

ES 2 526 080 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

Modelo de calidad de vídeo dependiente del contenido para servicios de transmisión de vídeo

5 La invención se refiere a un procedimiento y un aparato para estimar la calidad percibida de una señal de vídeo digital, preferiblemente en el contexto de los servicios de transmisión de vídeo, tales como televisión bajo protocolo de Internet (Internet Protocol Television, IPTV) o vídeo bajo demanda (Vídeo on Demand, VoD) y, en particular, para
 10 estimaciones dependientes del contenido de la calidad percibida de una señal de vídeo digital proporcionando parámetros relacionados con la complejidad del contenido, y controlando los procedimientos de estimación de calidad de vídeo basados en parámetros, existentes o futuros, mediante los parámetros proporcionados relacionados con la complejidad del contenido. La invención está diseñada para secuencias de vídeo codificadas, pero funciona también sobre secuencias de vídeo no codificadas.

15 Con el propósito de garantizar un alto grado de satisfacción para el usuario de los servicios de vídeo, tal como transmisión de vídeo no interactiva (IPTV, VoD), la calidad de vídeo percibida de esos servicios debe ser estimada. El mantenimiento de la calidad de su servicio representa una importante responsabilidad del proveedor de difusión tanto con respecto al proveedor de contenidos como con respecto al cliente. En las grandes redes IPTV, solo las sondas de supervisión de la calidad, completamente automatizadas, pueden cumplir este requisito.

20 Con este propósito, se han desarrollado modelos de calidad de vídeo que proporcionan estimaciones de la calidad de vídeo tal como es percibida por el usuario. Esos modelos pueden proporcionar, por ejemplo, el grado de similitud entre el vídeo recibido en el lado del usuario y el vídeo no degradado original. Además, y de una manera más sofisticada, puede modelarse el sistema visual humano (Human Visual System, HVS). Por último, la salida del modelo puede ser asignada a los resultados de extensos ensayos subjetivos de la calidad, para proporcionar finalmente una estimación de la calidad percibida.

Los modelos de calidad de vídeo y, de esta manera, los sistemas de medición se clasifican generalmente de la manera siguiente:

Tipos de modelos de calidad

- 25
- Referencia completa (Full Reference, FR): se requiere una señal de referencia.
 - Referencia reducida (Reduced Reference, RR): se requiere información parcial extraída de la señal de origen.
 - Sin referencia (No Reference, NR): no se requiere señal de referencia.

Tipos de parámetros de entrada

- 30
- Basados en señal/medio: se requiere la imagen decodificada (información de píxel).
 - Basados en parámetros: se requiere información a nivel de secuencias de bits. La información puede ser información de cabecera de un paquete, que requiere un análisis sintáctico de las cabeceras de los paquetes, análisis de la secuencia de bits incluyendo la carga útil, es decir, la información de codificación, y la decodificación parcial o total de la secuencia de bits.

35 Tipo de aplicación

- Planificación de la red: el modelo o el sistema de medición se usa antes de la implementación de la red con el propósito de planificar la mejor implementación posible.
- Supervisión de servicio: el modelo se usa durante la operación del servicio.

La información relacionada de los tipos de modelos de calidad de vídeo puede encontrarse en la bibliografía [1-3].

40 Se han descrito varios modelos paramétricos de calidad de vídeo, basados en paquetes, en la bibliografía [4 - 6]. Sin embargo, una desventaja importante de estos modelos es que no tienen en cuenta el impacto sobre la calidad del contenido. En otros términos, y tal como se ha informado en estudios anteriores [7-12], la calidad de vídeo percibida depende de las características espacio-temporales del vídeo. Por ejemplo, es bien sabido que la pérdida de paquetes es ocultada generalmente mejor cuando no hay movimientos complejos en el vídeo, tal como en la
 45 transmisión de noticias. Cuando no hay pérdida de paquetes y para tasas de bit bajas y medias, el contenido con una baja complejidad espacio-temporal consigue una mejor calidad que el contenido espacio-temporalmente complejo.

Además, la técnica anterior tiene por objeto también incluir el impacto sobre la calidad del contenido en unos

modelos paramétricos de calidad de vídeo basados en parámetros, tanto para los casos con pérdida de paquetes como para los casos sin pérdida de paquetes, véanse las referencias [13a, 13b, 14, 15, 16].

Por ejemplo, en las referencias [13a, 13b, 14], la complejidad de los contenidos se determina para cada fotograma de vídeo comparando el tamaño del fotograma actual con un umbral adaptativo. Dependiendo de si el tamaño del fotograma actual es mayor, igual o menor que este umbral se producirá un aumento o una disminución de la calidad estimada asociada con el fotograma actual. Sin embargo, debido al uso de un valor de umbral y a si es mayor, igual o menor que este valor, el procedimiento descrito en estas referencias sólo proporciona una consideración relativamente aproximada del contenido de vídeo. En otras palabras, no hay una medida suave o continua de la complejidad de los fotogramas dentro de una ventana de medición determinada. Además, debido a que el umbral adaptativo se calcula sobre la totalidad o una parte de la ventana de medición, la complejidad de cada fotograma se determina en relación a la complejidad de otros fotogramas en la misma secuencia de vídeo, pero no en relación a la complejidad de otros contenidos.

En la referencia [15], se propone una solución para la inserción de parámetros relacionados con el contenido, es decir, parámetros que reflejan la complejidad espacio-temporal del contenido, tal como un parámetro de cuantificación y vectores de movimiento, en un modelo de calidad de vídeo basado en parámetros. Sin embargo, estos parámetros relacionados con el contenido no pueden ser extraídos de una secuencia de bits codificada, de manera que la referencia [15] no puede ser usada de la misma manera que la presente invención.

La referencia [16] presenta una solución para la estimación de la calidad de vídeo percibida en caso de pérdida de paquetes con un único parámetro, que representa la magnitud de la degradación de la señal debida a la pérdida de paquetes. Esta solución prevé la inclusión de un factor de corrección para ajustar la magnitud estimada de la degradación de la señal basada en la complejidad temporal o espacio-temporal del contenido. Sin embargo, no se propone ninguna solución para calcular este factor de corrección, por ejemplo en caso de vídeo codificado.

Por consiguiente, todavía hay una necesidad de un procedimiento para estimar la calidad percibida de una señal de vídeo digital. Por un lado, dicho procedimiento debería permitir una consideración bastante precisa del impacto sobre la calidad del contenido de la señal de vídeo y, por otro lado, debería ser aplicable también a un vídeo codificado, incluyendo tanto el caso de la degradación de codificación con y sin pérdida de paquetes. Asimismo, existe una necesidad de un aparato configurado para realizar un procedimiento con estas características.

Estos objetivos se consiguen mediante el procedimiento y el aparato con las características descritas en las reivindicaciones presentadas en la presente memoria.

La invención está dirigida a su uso con modelos de calidad de vídeo basados en parámetros en el caso de vídeo codificado, es decir, cuando sólo está disponible la información de cabecera de los paquetes. La invención funciona también en el caso de vídeo no codificado, pero puede ser, por diseño, menos precisa que un modelo de calidad de vídeo basado en una decodificación completa o una extracción de información más profunda a partir de la secuencia de bits no codificada. El uso solo de la información basada en paquetes ofrece la ventaja de mantener la complejidad computacional de la invención baja y, por supuesto, extiende el campo de aplicación a las secuencias no codificadas, así como a las secuencias codificadas.

La invención puede resumirse de la manera siguiente:

El objetivo de la presente invención es proporcionar un procedimiento para estimar la calidad percibida de una señal de vídeo digital proporcionando parámetros de complejidad de contenido y usando estos parámetros de complejidad de contenido para controlar los procedimientos arbitrarios de estimación de calidad de vídeo basados en parámetros (y, de esta manera, existentes o futuros). Por un lado, el procedimiento según la invención permite una consideración bastante precisa del impacto sobre la calidad del contenido de la señal de vídeo pero, por otro lado, es aplicable también al vídeo codificado y para ambos, en caso de pérdida de paquetes así como en caso sin pérdida de paquetes. Un objetivo adicional de la presente invención es proporcionar un aparato configurado para calcular los parámetros de complejidad de contenido e insertarlos en modelos arbitrarios de calidad de vídeo basados en parámetros, con todas las ventajas con las que están asociados dichos enfoques basados en el encabezado de los paquetes.

Cabe señalar también que la presente invención difiere sustancialmente de los enfoques de las referencias citadas anteriormente [13a, 13b, 14] tanto por los parámetros relacionados con el contenido que se calculan como por la manera en la que estos parámetros son incluidos en los modelos. En la presente invención, los parámetros relacionados con el contenido se proporcionan como valores absolutos, que no dependen del historial de fotogramas, tal como se describe en las referencias [13a, 13b, 14]. Por lo tanto, en principio, podrían ser usados para comparar la complejidad de dos contenidos diferentes, o de escenas o pasajes diferentes de un contenido. Además, los valores de los parámetros relacionados con el contenido, usados en la presente invención, son continuos (y no categóricos en términos de clases no indistintas como en [13a, 13b, 14]) y, de esta manera, permiten

una estimación muy precisa del impacto sobre la calidad del contenido. Además, en la presente invención, todos los parámetros se calculan a lo largo de toda la ventana de medición o por cada grupo de imagen (Group Of Picture, GOP) o por cada escena de vídeo, mientras que en la técnica anterior (véanse las referencias [13a, 13b, 14]), se calculan por cada fotograma.

5 Cabe señalar que en el caso de vídeo codificado, la estructura GOP puede ser estimada usando [20]. Además, cabe señalar que una escena (vídeo) comienza con un fotograma I y, generalmente, contiene varios GOPs. Los cortes de escena pueden ser detectados en caso de vídeo codificado usando la referencia [21] (no publicada con anterioridad a la fecha de presentación de la presente solicitud). Normalmente, dos escenas de vídeo difieren en sus contenidos semánticos. Además, la variación entre escenas de la complejidad espacio-temporal (ST) de la
10 señal de contenido es generalmente más baja que su variación entre escenas.

Dos formas comunes de expresar la calidad de vídeo Q_v estimada, en base a las contribuciones debidas de diferentes tipos de degradaciones, se muestran en las ecuaciones siguientes,

$$Q_v = Q_{vo} - I_{cod} - I_{tra}, \quad (1)$$

$$Q_v = Q_{vo} \times I_{cod} \times I_{tra}, \quad (2)$$

15

en las que I_{cod} e I_{tra} son ejemplos de "factores de deterioro" (Impairment Factors, IF). Un factor de deterioro cuantifica el impacto sobre la calidad de un tipo específico de degradación, y cada factor de deterioro puede ser calculado a partir de una descripción paramétrica de las señales y la ruta de transmisión. En las ecuaciones (1) y (2), I_{cod} representa el impacto sobre la calidad de los artefactos de compresión, e I_{tra} representa el impacto sobre
20 la calidad de los errores de transmisión (pérdida de paquetes). Cabe señalar que en la ecuación (2) ya lo largo de toda la solicitud, el símbolo "x" denotará la multiplicación normal entre dos números reales, que a veces se representa también con el símbolo ".".

Todos los términos en las ecuaciones (1) y (2) se expresan, por ejemplo, en una escala de 0 a 100, o de 1 a 5.

Q_{vo} es la calidad base y, típicamente, corresponde al valor más alto de la escala usada para expresar la calidad percibida, por ejemplo $Q_{vo} = 100$ o $Q_{vo} = 5$.

Según la invención, I_{cod} e I_{tra} y, de esta manera, Q_v pueden ser calculados para cada ventana de medición, en el que una ventana de medición tiene típicamente una duración de 10 a 20 segundos.

Otro enfoque, seguido por ejemplo por las referencias [13] y [14], consiste en calcular las contribuciones a la calidad relacionada con la imagen debidas a la codificación y la pérdida de paquetes por cada fotograma de vídeo. A
30 continuación, el conjunto obtenido de valores de calidad de fotogramas de vídeo se agrega a lo largo de la ventana de medición. Una manera sencilla de agregar los valores de calidad de vídeo por cada fotograma es calculando la media. En las referencias [17-19] se describen maneras más sofisticadas.

A continuación, se calculan I_{cod} , I_{tra} y Q_v por cada ventana de medición. Además, tanto I_{cod} como I_{tra} se calculan usando una función de la manera siguiente, a la cual se hará referencia en adelante como "función de impacto":

$$35 \quad f_{IF} : \mathbb{R}^m \times \mathbb{R}^n \times \mathbb{R}^u \rightarrow \mathbb{R}, \quad (\mathbf{p}^{IF}, \mathbf{q}^{IF}, \mathbf{a}^{IF}) \mapsto Imp := f_{IF}(\mathbf{p}^{IF}, \mathbf{q}^{IF}, \mathbf{a}^{IF}), \quad (3)$$

en la que $Imp \in \{I_{cod}, I_{tra}\}$, m , n y u son números enteros positivos, f_{IF} es una función de impacto que depende del índice (superior) IF que denota el factor de deterioro respectivo, y en el que

$$40 \quad \mathbf{p}^{IF} = (p_1^{IF}, \dots, p_m^{IF}) \in \mathbb{R}^m \quad (4)$$

denota un primer conjunto de parámetros que se refiere a las características técnicas de codificación o de la red, tales como la tasa de bits, el número de fotogramas por segundo o el porcentaje de pérdida de paquetes, y

$$45 \quad \mathbf{q}^{IF} = (q_1^{IF}, \dots, q_n^{IF}) \in \mathbb{R}^n \quad (5)$$

denota un segundo conjunto de parámetros, al cual sea referencia también en adelante como "parámetros dependientes del contenido", que se derivan de los parámetros GOP/complejidad de escena definidos más

adelante, y

$$\mathbf{a}^{IF} = (\alpha_1^{IF}, \dots, \alpha_u^{IF}) \in \mathbb{R}^u \quad (6)$$

denota un conjunto de coeficientes asociados a f_{IF} . En adelante, el superíndice IF se suprimirá a veces en la notación de las cantidades, proporcionadas por las ecuaciones (4) a (6) en aras de la simplicidad.

Aquí, p^{IF} y q^{IF} se calculan preferiblemente para cada ventana de medición, en el que una ventana de medición tiene típicamente una duración de 10 a 20 segundos. En adelante, el superíndice IF será denominado según el nombre respectivo de la variable usada para la cuantificación o medición de un factor de deterioro específico, es decir, por ejemplo, *lcod* o *ltra*. Además, la aplicación de la ecuación (3) no se limita a los casos de los factores de deterioro *lcod* e *ltra*; más bien, la ecuación (3) puede ser aplicada también a otros tipos de degradaciones de calidad, es decir, a otros factores de deterioro.

Cabe señalar que una función de impacto según la ecuación (3) constituye un concepto general para la estimación de la contribución relacionada con el contenido a los factores de deterioro. En otras palabras, la ecuación (3) no sólo es aplicable a diferentes factores de deterioro, tales como *lcod* o *ltra*, sino que también se aplica a diversos modelos (basados en parámetros) para estimar la degradación de la calidad debida a un factor de deterioro específico, por ejemplo, *lcod*. Usando los parámetros dependientes de contenido, descritos por un conjunto \mathbf{q}^{IF} en una realización específica de la ecuación (3) adaptada para un procedimiento de estimación elegido para un factor de deterioro, la estimación de este factor de deterioro es controlada mediante los parámetros dependientes del contenido. Cuando se realiza la última etapa de cálculo de una estimación de la calidad "global" Q_v percibida de la señal de vídeo, por ejemplo, empleando las ecuaciones (1) o (2) o cualquier otro procedimiento basado en estimaciones de uno o más factores de deterioro, la estimación de Q_v es controlada también por los parámetros dependientes del contenido. De esta manera, el procedimiento según la invención permite la consideración precisa, indicada anteriormente, del impacto sobre la calidad debido al contenido de la señal de vídeo.

Los parámetros GOP/complejidad de la escena usados para calcular los parámetros \mathbf{q}^{IF} dependientes de contenido son, todos ellos, parámetros que requieren el conocimiento del tipo y del tamaño (por ejemplo en bytes) de los fotogramas de vídeo. Normalmente, estos parámetros son calculados, aunque no necesariamente, por cada grupo de imagen (GOP) o escena de vídeo (SC), y los parámetros o la estimación de calidad resultante se agregan a continuación, sobre la ventana de medición.

Según la invención, pueden considerarse al menos los siguientes parámetros de GOP/complejidad de la escena:

- S_{sc}^I : tamaño medio de fotograma I para una escena *sc* determinada; en la realización preferida, preferiblemente, se ignora el primer fotograma I de la primera escena,
- S_{gop}^P : tamaño medio de fotograma P para un gop GOP determinado,
- S_{gop}^B : tamaño medio de referencia B (usado en caso de la codificación jerárquica) por cada GOP
- S_{gop}^b : Tamaño medio de los tamaños de fotograma b no de referencia por cada GOP,
- S_{gop}^{noI} : Tamaños medios de fotogramas P, B y b por cada GOP,
- B_{sc}^I : tasa de bits de los fotogramas I calculados por cada escena,
- B_{sc}^P : tasa de bits de los fotogramas P calculados por cada escena,
- B_{sc}^B : tasa de bits de los fotogramas B calculados por cada escena,
- B_{sc}^b : tasa de bits de los fotogramas b calculados por cada escena,
- B_{sc}^{noI} : tasa de bits conjunta de los fotogramas P, B y b calculados por cada escena.

En los símbolos anteriores, el tipo de secuencia de fotogramas, es decir, I, P, B, b o noI, se indicada mediante un superíndice, que no debe ser confundido con un exponente.

La tasa de bits por escena de los fotogramas con el tipo de fotograma T (B_{sc}^T , donde $T \in \{I, P, B, b, NoI\}$) se calcula como sigue:

45

$$B_{sc}^T = \frac{By_{sc}^T \times fr^T}{nfr^T \times br^T}, \quad (7)$$

donde

- 5
- By_{sc}^T : es la cantidad total de bytes para un fotograma T para cada escena,
 - fr^T es el número de fotogramas por segundo para los fotogramas T, es decir, el número de fotogramas T por segundo,
 - nfr^T es el número de fotogramas T por segundo en la escena,
 - br es la tasa de bits global, en Mbit/s.

10 Como una alternativa, fr^T podría ser sustituido por el número de fotogramas por segundo fr global y fr^T por el número nfr general de fotogramas en la escena.

Además, las siguientes relaciones pueden ser consideradas como parámetros GOP/complejidad de la escena. Cada relación se calcula por cada GOP a partir de los parámetros GOP/complejidad de la escena, tal como se ha definido anteriormente:

- 15
- $S^{P/I} = S_{gop}^P / S_{sc}^I$
 - $S^{b/I} = S_{gop}^b / S_{sc}^I$
 - $S^{b/P} = S_{gop}^b / S_{gop}^P$
 - $S^{noI/I} = S_{gop}^{noI} / S_{sc}^I$
 - $B^{P/I} = B_{sc}^P / B_{sc}^I$
- 20
- $B^{b/I} = B_{sc}^b / B_{sc}^I$
 - $B^{b/P} = B_{sc}^b / B_{sc}^P$
 - $B^{noI/I} = B_{sc}^{noI} / B_{sc}^I$

También aquí, el superíndice de los símbolos de la parte izquierda y derecha de las ecuaciones debe entenderse como un índice superior.

25 Un aspecto de la invención se refiere a un procedimiento para estimar la calidad de percepción de una señal de vídeo digital, en el que el procedimiento comprende las etapas de:

- (1a) extraer información a partir de la secuencia de bits de vídeo, que es capturada antes de su decodificación;
- (1b) obtener una estimación o estimaciones de uno o más factores de deterioro IF , usando, para cada una de las estimaciones, una función de impacto adaptada para el factor de deterioro respectivo;
- 30 (1c) estimar la calidad percibida de la señal de vídeo digital usando la estimación o las estimaciones obtenidas en la etapa (1b);

en el que el procedimiento está caracterizado por que cada una de las funciones de impacto usadas en la etapa (1b) acepta como entrada un conjunto de parámetros q dependientes del contenido calculados a partir de un conjunto de parámetros GOP/complejidad de la escena, en el que los parámetros GOP/complejidad de la escena pueden ser derivados a partir de la información de la cabecera del paquete y están disponibles en caso de secuencias de bits de vídeo codificadas.

35

Según el procedimiento de la invención, los parámetros GOP/complejidad de la escena pueden ser calculados por cada grupo de Imagen (GOP) o escena de vídeo.

40 Según una realización del procedimiento, cada una de las funciones de impacto usadas en la etapa (1b) depende además de:

las características técnicas de codificación o de red, por ejemplo la tasa de bits, el número de fotogramas por

segundo, el porcentaje de pérdida de paquetes o la proporción de pérdida en un GOP o escena; y /o

los coeficientes asociados con la función de impacto.

En una realización preferida de la invención, el conjunto de parámetros q dependientes del contenido se deriva a partir de al menos uno de los siguientes parámetros GOP/complejidad de la escena:

- 5 S_{sc}^I , que denota el tamaño medio de fotograma I por escena, en el que preferiblemente el primer fotograma I de la primera escena se ignora;
- S_{gop}^P , que denota el tamaño medio de fotograma P por cada GOP;
- S_{gop}^B , que denota los tamaños medios (de referencia) de fotograma B por cada GOP;
- S_{gop}^b , que denota los tamaños medios de fotograma b no de referencia por cada GOP;
- 10 S_{gop}^{noI} , que denota los tamaños medios conjuntos de los fotogramas P, B y b por cada GOP;
- B_{sc}^I , que denota la tasa de bits de los fotogramas I calculada por cada escena;
- B_{sc}^P , que denota la tasa de bits de los fotogramas P calculada por cada escena;
- B_{sc}^B , que denota la tasa de bits de los fotogramas B calculada por cada escena;
- B_{sc}^b , que denota la tasa de bits de los fotogramas b calculada por cada escena;
- 15 B_{sc}^{noI} , que denota la tasa de bits de los fotogramas P, B y b calculada por cada escena.

En una realización de la invención, el conjunto de parámetros q se deriva a partir de al menos uno de los parámetros GOP/complejidad de la escena

$$S^{P/I} = S_{gop}^P / S_{sc}^I$$

$$S^{b/I} = S_{gop}^b / S_{sc}^I$$

20 $S^{b/P} = S_{gop}^b / S_{gop}^P$

$$S^{noI/I} = S_{gop}^{noI} / S_{sc}^I$$

$$B^{P/I} = B_{sc}^P / B_{sc}^I$$

$$B^{b/I} = B_{sc}^b / B_{sc}^I$$

$$B^{b/P} = B_{sc}^b / B_{sc}^P$$

25 $B^{noI/I} = B_{sc}^{noI} / B_{sc}^I$

En una realización, se usa una función f_{IF} de impacto.

Preferiblemente, la función f_{IF} de impacto se usa para estimar el impacto sobre la calidad debido a los artefactos de compresión, que depende de un parámetro q_I dependiente del contenido que es calculado a partir de la inversa de la media ponderada del parámetro GOP/complejidad de la escena S_{sc}^I de las escenas sc multiplicado por un coeficiente. El coeficiente puede ser proporcional al número nx de píxeles por fotograma de vídeo y el número fr de fotogramas de vídeo por segundo.

30

En una realización preferida del procedimiento de la invención, cada escena sc tiene un peso de $w_{sc} \times N_{sc}$, donde N_{sc} es el número de GOPs por escena y w_{sc} es un factor de ponderación adicional, en el que para las escenas que tienen el valor S_{sc}^I más bajo: w_{sc} se establece a un valor mayor que 1, por ejemplo $w_{sc} = 16$, y para el resto de las escenas: w_{sc} se establece a un valor igual a 1.

35

En una realización, el parámetro q_I dependiente del contenido viene determinado por

$$q_I = \frac{\sum_{sc} w_{sc} \times N_{sc}}{\sum_{sc} S_{sc}^I \times w_{sc} \times N_{sc}} \times \frac{nx \times fr}{1000}$$

En el caso de un conjunto de parámetros unidimensional (vector de parámetros), en adelante, el símbolo del único elemento del conjunto será identificado con el símbolo del conjunto en aras de la simplicidad. Por ejemplo, si el conjunto de parámetros dependientes del contenido tiene sólo un parámetro, es decir, $\mathbf{q} = (q_i)$, se escribirá simplemente $\mathbf{q} = q_i$. De manera análoga, se establece $\mathbf{p} = (p_i) = p_i$ en el caso de un conjunto de parámetros unidimensional asociado con las características técnicas de codificación o de red.

En una realización del procedimiento de la invención, la función f_{IF} de impacto que depende del parámetro $\mathbf{q} = q_i$ dependiente del contenido viene determinada por

$$f_{IF}(\mathbf{p}, \mathbf{q}, \boldsymbol{\alpha}) = \alpha_1 \times \exp(\alpha_2 \times p_1) + \alpha_3 \times q_1 + \alpha_4 \quad , \quad |$$

en la que $\mathbf{p} = p_i$ es preferiblemente un parámetro que describe el número de bits por píxel y viene determinado, más preferiblemente, por

$$p_1 = \frac{\text{tasa de bits} \times 10^6}{n_x \times f_r} \quad , \quad y$$

donde $\boldsymbol{\alpha} = (\alpha_1, \alpha_2, \alpha_3, \alpha_4)$ es el conjunto de coeficientes asociados a la función de impacto.

En una realización del procedimiento de la invención, se usa una función f_{IF} de impacto, preferiblemente para estimar el impacto sobre la calidad debido a los artefactos de transmisión, que depende de un conjunto de parámetros $\mathbf{q} = (q_1, q_2)$ dependientes del contenido, en el que cada componente q_j con $j \in \{1, 2\}$ del conjunto se obtiene mediante una suma ponderada de los parámetros $\beta_{k,j}$ dependiente de los parámetros GOP/complejidad de la escena, en el que, preferiblemente, la suma ponderada para cada $j \in \{1, 2\}$ se calcula según

$$q_j = \sum_{k=1}^v \beta_{k,j} \times R_{k,j}$$

con pesos $R_{k,j}$.

Los pesos pueden venir determinados por

$$R_{k,j} = \sum_i r_i \times (T_k - t_i) \quad \text{para } j \in \{1, 2\}$$

en la que T_k es la duración de la pérdida de GOP k , t_i es la ubicación del GOP de un evento de pérdida y r_i denota la extensión espacial del evento de pérdida i .

Según una realización preferida, se usa:

en caso de un sector por fotograma,

$$r_i = \frac{nap}{np} \quad ;$$

y en el caso de más de un sector por fotograma,

$$r_i = \frac{nlp}{np} + nle \times \frac{1}{2 \times nsl} \quad ; \quad |$$

donde np es el número de paquetes en el fotograma, nap es el número de paquetes de secuencias de transporte

afectados (TS) en el fotograma afectado, nip es el número de paquetes perdidos en el fotograma, n/e es el número de eventos de pérdida en el fotograma y ns/l es el número de sectores en el fotograma.

El parámetro $\beta_{k,1}$ puede depender del parámetro GOP/complejidad de la escena $S^{nol/l}$.

El parámetro $\beta_{k,2}$ puede depender del parámetro GOP/complejidad de la escena $S^{b/P}$.

5 Según una realización del procedimiento, los parámetros $\beta_{k,1}$ cada $k \in \{1, \dots, v\}$ se obtienen mediante las etapas siguientes:

(12a) establecer $\beta_{k,1} = S^{nol/l}$;

(12b) en el caso $\beta_{k,1} \leq 0,5$, establecer $\beta_{k,1}$ a un valor $2x \beta_{k,1}$;

(12c) en el caso $\beta_{k,1} > 0,5$, establecer $\beta_{k,1}$ a un valor 1.

10 Preferiblemente, los parámetros $\beta_{k,2}$ para cada $k \in \{1, \dots, v\}$ se obtienen como $\beta_{k,2} = \max(0, -S^{b/P} + 1)$.

En una realización, la función f_{IF} de impacto que depende del conjunto de parámetros $\mathbf{q} = (q_1, q_2)$ dependientes del contenido es proporcionada por

$$f_{IF}(\mathbf{p}, \mathbf{q}, \boldsymbol{\alpha}) = \alpha_1 \times \log\left(1 + \frac{\alpha_2 \times q_1 + \alpha_3 \times q_2}{p_1 \times p_2}\right),$$

15

donde $\boldsymbol{\alpha} = (\alpha_1, \alpha_2, \alpha_3, \alpha_4)$ es el conjunto de coeficientes asociados a la función de impacto.

Preferiblemente, p_1 es un parámetro que describe el impacto sobre la calidad debido a los artefactos de compresión.

Preferiblemente, p_2 es el número de GOPs en la ventana de medición o la duración de la ventana de medición.

20 En una realización del procedimiento de la invención, la señal de vídeo es al menos parte de una secuencia de datos no interactiva, preferiblemente un vídeo no interactivo o secuencia audiovisual, o al menos parte de una secuencia de datos interactiva, preferiblemente un vídeo interactivo o secuencia audiovisual.

25 En una realización, el procedimiento se combina con uno o más procedimientos para estimar el impacto sobre la percepción de la calidad de una señal de vídeo digital por otros deterioros diferentes a la compresión y/o transmisión, en el que la combinación se realiza preferiblemente usando al menos una función lineal y/o al menos una función multiplicativa de los procedimientos a combinar.

30 En una realización, el procedimiento es combinado con uno o más de otros procedimientos para estimar la calidad de la percepción de un vídeo digital debido a la compresión y/o la transmisión, en el que la combinación es realizada preferiblemente usando al menos una función lineal y/o al menos una función multiplicativa de los procedimientos a combinar.

Un aspecto de la invención se refiere a un procedimiento para supervisar la calidad de una señal de vídeo digital transmitida con las etapas de:

(18a) transmitir la señal de vídeo desde un servidor al cliente;

35 (18b) en el lado del cliente, ejecutar el procedimiento para estimar la calidad de la percepción de una señal de vídeo digital según el procedimiento para estimar la calidad de la percepción de una señal de vídeo digital tal como se ha descrito anteriormente;

(18c) transferir el resultado de la estimación de la etapa (18b) al servidor;

(18d) en el lado del servidor, supervisar la estimación de la calidad de la señal de vídeo transmitida; y

en el que el procedimiento comprende preferiblemente las etapas adicionales de:

40 (18e) analizar la calidad supervisada de la señal de vídeo transmitida, preferiblemente en función de los parámetros de transmisión; y opcionalmente

(18f) cambiar los parámetros de transmisión en base a la etapa de análisis (18e) con el propósito de aumentar

la calidad de la señal de vídeo transmitida.

Un aspecto de la invención se refiere a un aparato para estimar la calidad de percepción de una señal de vídeo digital, en el que el aparato comprende:

5 unos medios configurados para extraer información a partir de una secuencia de bits de vídeo capturada antes de la decodificación;

al menos un estimador de impacto;

un estimador de calidad configurado para estimar la calidad de percepción Q_v de la señal de vídeo:

10 en el que el aparato está caracterizado por que cada uno de los estimadores de impacto está configurado para estimar el impacto sobre la calidad debido a un factor de deterioro por medio de una función de deterioro que acepta como entrada un conjunto de parámetros dependientes del contenido calculados a partir de un conjunto parámetros GOP/complejidad de la escena, en el que los parámetros de GOP/complejidad de la escena pueden derivarse a partir de la información de la cabecera del paquete y, de esta manera, están disponibles en el caso de secuencias de bits de vídeo codificado.

15 Preferiblemente, el aparato está configurado además para estimar la calidad de percepción de una señal de vídeo digital usando un procedimiento según una cualquiera de las realizaciones del procedimiento para estimar la calidad de percepción de una señal de vídeo digital, tal como se ha descrito anteriormente.

Un aspecto de la invención se refiere a un decodificador que se puede conectar a un receptor para recibir una señal de vídeo digital, en el que el decodificador comprende el aparato según la invención.

20 Un aspecto de la invención se refiere a un sistema para supervisar la calidad de una señal de vídeo digital transmitida, en el que el sistema comprende un servidor y un cliente, y el sistema está configurado para ejecutar el procedimiento para supervisar la calidad de una señal de vídeo digital transmitida según la invención, tal como se ha descrito anteriormente.

En una realización del sistema, el cliente está configurado como el aparato según la invención.

En una realización del sistema, el cliente comprende un aparato según la invención.

25 En una realización alternativa del sistema de la invención, el sistema comprende además el decodificador según la invención, en el que el decodificador está conectado al cliente.

Lista de figuras:

Fig. 1: Ilustración del cálculo de la ecuación (10) usada como un ejemplo para explicar el impacto sobre la calidad del contenido en el caso sin pérdidas. Véase el texto anterior para más detalles.

30 Fig. 2: Ilustración de las ecuaciones (17a) a (17c) usadas como un ejemplo para explicar el impacto sobre la calidad del contenido en caso de pérdida de paquetes. Véase el texto anterior para más detalles.

Fig. 3: Ilustración de la ecuación (18) usada como un ejemplo para explicar el impacto sobre la calidad del contenido en caso de pérdida de paquetes. Véase la descripción para más detalles.

35 Otros aspectos, características y ventajas serán evidentes a partir del resumen anterior, así como a partir de la descripción siguiente, incluyendo las figuras y las reivindicaciones.

Según la invención, el impacto sobre la complejidad del contenido tanto en el deterioro I_{cod} de la calidad relacionado con la compresión como en el deterioro I_{tra} de la calidad relacionado con la transmisión pueden estimarse usando el esquema descrito a continuación:

Caso sin pérdidas - I_{cod}

40 Una realización de la invención se refiere a la inclusión de los parámetros de GOP /complejidad de la escena en la ecuación (3), en la que $I_{mp} = I_{cod}$, $m = 1$, $n = 1$, $u = 4$, y en el que I_{mp} se obtiene haciendo que $f_{I_{cod}}$ sea una función exponencial:

$$f_{I_{cod}}(\mathbf{p}^{I_{cod}}, \mathbf{q}^{I_{cod}}, \boldsymbol{\alpha}^{I_{cod}}) = \alpha_1^{I_{cod}} \times \exp(\alpha_2^{I_{cod}} \times p_1^{I_{cod}}) + \alpha_3^{I_{cod}} \times q_1^{I_{cod}} + \alpha_4^{I_{cod}}. \quad (8)$$

45 Como un ejemplo del conjunto de coeficientes $\boldsymbol{\alpha}^{I_{cod}}$ en la ecuación (8), se tiene:

$$\alpha_1^{Icod} = 47,78 ,$$

$$\alpha_2^{Icod} = -21,46 ,$$

$$\alpha_3^{Icod} = 7,61 ,$$

$$\alpha_4^{Icod} = 7,71 ,$$

y, preferiblemente, p_1^{Icod} es el número medio de bits por píxel que viene determinado más preferiblemente por

$$p_1^{Icod} = \frac{br \times 10^6}{nx \times fr} , \quad (9)$$

en la que nx y fr son el número de píxeles por fotograma de vídeo y el número de fotogramas de vídeo por segundo, respectivamente, Además, br es la tasa de bits de vídeo en Mbit/s.

En una realización preferida, q_1^{Icod} es una función del parámetro GOP/complejidad de la escena S_{sc}^I y se expresa como sigue:

$$q_1^{Icod} = \frac{\sum_{sc} w_{sc} \times N_{sc}}{\sum_{sc} S_{sc}^I \times w_{sc} \times N_{sc}} \times \frac{nx \times fr}{1000} , \quad (10)$$

en la que nx y fr son el número de píxeles por fotograma de vídeo y el número de fotogramas de vídeo por segundo, respectivamente, y N_{sc} el número de GOPs por escena. Para la escena que tiene el valor S_{sc}^I más bajo, $w_{sc} > 1$, en el que preferiblemente $w_{sc} = 16$, si no $w_{sc} = 1$.

La Fig. 1 ilustra, como un ejemplo, el cálculo de la ecuación (10) con una secuencia de vídeo compuesta de dos escenas (se supone que la ventana de medición corresponde a la duración de esta secuencia de vídeo). El formato de la secuencia de vídeo es 1080p25. Como consecuencia, $nx = 1920 \times 1080 = 2.073.600$ y $fr = 25$.

La primera escena ($sc = 1$) contiene dos GOPs (gop1 y gop2), es decir, $N_1 = 2$, y su tamaño medio de fotograma I es $S_1^I = 0,1$ (por ejemplo, en Megabytes).

La segunda escena ($sc = 2$) contiene tres GOPs (gop3 a gop5), es decir, $N_2 = 3$, y su tamaño medio de fotograma I es $S_2^I = 0,3$ (por ejemplo, en Megabytes).

El S_{sc}^I mínimo en la secuencia de vídeo es S_1^I . Como consecuencia,

$$w_1 = 16 ,$$

$$w_2 = 1 ,$$

y

$$q_1^{Icod} = \frac{16 \times 2 + 3 \times 1}{0,1 \times 10^6 \cdot 16 \times 2 + 0,3 \times 10^6 \times 3 \times 1} \times \frac{2073600 \times 25}{1000} = 0,4425 .$$

Caso con pérdidas – Itra

Una realización de la invención se refiere a la inclusión de parámetros GOP/complejidad de la escena en la ecuación (3), en la que $Imp = Itra$, $m = 2$, $n = 2$, $u = 3$, y en la que Imp se obtiene haciendo que f_{Itra} sea una función logarítmica:

$$f_{lra}(\mathbf{p}^{lra}, \mathbf{q}^{lra}, \boldsymbol{\alpha}^{lra}) = \alpha_1^{lra} \times \log\left(1 + \frac{\alpha_2^{lra} \times q_1^{lra} + \alpha_3^{lra} \times q_2^{lra}}{p_1^{lra} \times p_2^{lra}}\right). \quad (11)$$

Como un ejemplo del conjunto de coeficientes α^{lra} en la ecuación (11), se tiene:

$$\begin{aligned} \alpha_1^{lra} &= 17,95, \\ \alpha_2^{lra} &= \alpha_3^{lra} = 59,02. \end{aligned}$$

Preferiblemente,

$$\begin{aligned} p_1^{lra} &= Icod, \\ p_2^{lra} &= v, \end{aligned}$$

en la que v es el número de GOPs en la ventana de medición. De manera alternativa, v es la duración de la ventana de medición.

En la realización preferida, q_1^{lra} y q_2^{lra} se derivan a partir de los parámetros GOP/complejidad de la escena y se obtienen por cada ventana de medición usando las relaciones siguientes:

$$q_1^{lra} = \sum_{k=1}^v \beta_{k,1} \times R_{k,1}, \quad (12)$$

$$q_2^{lra} = \sum_{k=1}^v \beta_{k,2} \times R_{k,2}, \quad (13)$$

en la que v es el número de GOPs en la ventana de medición, y $R_{k,1}$ y $R_{k,2}$ son descriptores espacio-temporales de la pérdida calculada para cada GOP k , que se calculan como sigue:

$$R_{k,1} = R_{k,2} = R_k = \sum_i r_i \times (T_k - t_i), \quad (14)$$

en la que T_k es la duración de la pérdida de GOP k , t_i es la ubicación en el GOP de un evento de pérdida i y r_i denota la extensión espacial del evento de pérdida i , y en la que preferiblemente:

en caso de un sector por fotograma,

$$r_i = \frac{nap}{np}; \quad (15)$$

y, en caso de más de un sector por fotograma,

$$r_i = \frac{nip}{np} + nle \times \frac{1}{2 \times nsl}; \quad (16)$$

en la que np es el número de paquetes en el fotograma, nap es el número de paquetes en la secuencia de transporte (TS) afectada en el fotograma afectado (derivado usando cualquier procedimiento que implica la información del encabezado del paquete, tal como números de secuencia, marcas de tiempo, etc.), nlp es el número de paquetes perdidos en el fotograma, nle es el número de eventos de pérdida en el fotograma, y nsl es el número de sectores en el fotograma.

Cabe señalar que r_k es xl_k/T_k de la ecuación (5) en la referencia [16]. De manera similar, r_i de la ecuación (15) corresponde a xl_i de la ecuación (7c) en la referencia [16], y r_i en la ecuación (16) corresponde a xl_i en la

ecuación (7) de la referencia [16]. Finalmente, la suma de $\beta_{k,1}$ y $\beta_{k,2}$ de las ecuaciones (12) y (13) corresponde al factor de corrección $\alpha_{1,k}$ en la ecuación (9a) de la referencia [16]. Sin embargo, tal como se ha indicado anteriormente, no se propone ninguna solución para calcular este factor de corrección en caso de un vídeo codificado.

- 5 Además, los parámetros $\beta_{k,1}$ y $\beta_{k,2}$ se derivan a partir de los parámetros GOP/complejidad de la escena y se calculan para cada GOP k .

En una realización preferida, $\beta_{k,1}$ se obtiene usando las etapas siguientes (véase la Fig. 2):

- (a) establecer $\beta_{k,1} = S^{nol/}$;
 (b) en el caso de $\beta_{k,1} \leq 0,5$, establecer $\beta_{k,1}$ a $2 \times \beta_{k,1}$;
 10 (c) en caso de $\beta_{k,1} > 0,5$, establecer $\beta_{k,1}$ a 1.

En una realización preferida, $\beta_{k,2}$ se obtiene usando (véase la Fig. 3):

$$\beta_{k,2} = \max(0, -S^{h/P} + 1). \quad (18)$$

15 Aunque la invención ha sido ilustrada y descrita en detalle en los dibujos y la descripción precedentes, dicha ilustración y dicha descripción deben ser consideradas como ilustrativas o ejemplares y no restrictivas. Se entenderá que las personas con conocimientos en la materia pueden realizar cambios y modificaciones dentro del alcance de las reivindicaciones siguientes. En particular, la presente invención abarca realizaciones adicionales con cualquier combinación de características procedentes de diferentes realizaciones descritas anteriormente y a continuación.

20 Además, en las reivindicaciones la palabra "comprende" no excluye otros elementos o etapas, y los artículos indefinidos "un" o "una" no excluyen una pluralidad. Una única unidad puede cumplir las funciones de varias características expuestas en las reivindicaciones. De manera particular, los términos "esencialmente", "sobre", "aproximadamente" y similares, en conexión con un atributo o un valor definen también exactamente el atributo o exactamente el valor, respectivamente. Cualquier signo de referencia en las reivindicaciones no debería interpretarse como limitativo del alcance.

Referencias:

- [1] A. Takahashi, D. Hands y V. Barriac, "Standardization Activities in the ITU for a QoE Assessment of IPTV," en IEEE Communication Magazine, 2008.
- [2] S. Winkler y P. Mohandas, "The Evolution of Vídeo Quality Measurement: From PSNR to Hybrid Metrics," in IEEE Trans. Broadcasting, 2008.
- 30 [3] A. Raake, M.N. Garcia, S. Moeller, J. Berger, F. Kling, P. List, J. Johann, y C. Heidemann, "T-V-MODEL: Parameter-based prediction of IPTV quality," en Proc. of ICASSP, 2008.
- [4] O. Verscheure, P. Frossard y M. Hamdi, "User-oriented QoS analysis in MPEG-2 vídeo delivery," en Real-Time Imaging, 1999.
- 35 [5] K. Yamagishi y T. Hayashi, "Parametric Packet-Layer Model for Monitoring Vídeo Quality of IPTV Services," en Proc. of ICC, 2008.
- [6] M-N. Garcia y A. Raake, "Parametric Packet-Layer Vídeo Quality Model for IPTV," en Proc. of ISSPA, 2010.
- [7] S. Pechard, D. Barba y P. Le Callet, "Vídeo quality model based on a spatio-temporal features extraction for H.264-coded HDTV sequences," en Proc. of PCS, 2007.
- 40 [8] Y. Liu, R. Kurceren y U. Budhia, "Vídeo classification for vídeo quality prediction," en Journal of Zhejiang University Science A, 2006.
- [9] M. Ries, C. Crespi, O. Nemethova y M. Rupp, "Content-based Vídeo Quality Estimation for H.264/AVC Vídeo Streaming," en Proc. of Wireless Communications and Networking Conference, 2007.
- 45 [10] A. Khan, L. Sun. y E. Ifeachor, "Content clustering based vídeo quality prediction model for MPEG4 vídeo streaming over wireless networks," en Proc. of ICC, 2009.

- [11] Garcia, M.-N., Schleicher, R. y Raake, A. "Towards A Content-Based Parametric Video Quality Model For IPTV", in Proc. Of VPQM, 2010.
- [12] Guangtao Zhai et al, Cross-dimensional Quality Assessment for Low Bitrate Video, en IEEE Transactions on Multimedia, 2008.
- 5 [13a] Clark, A. (Telchemy), WO 2009012297 (AI), Method and system for content estimation of packet video streams.
- [13b] Clark, A. (Telchemy), US 2009/004114 (AI), Method and system for viewer quality estimation of packet video streams.
- 10 [14] Liao, Ning et al, "A packet-layer video quality assessment model with spatiotemporal complexity estimation", EURASIP Journal on Image and Video Processing 2011, 2011:5 (22 de Agosto 2011)
- [15] Garcia, M.-N., Schleicher, R. y Raake, A. (2010). Towards A Content-Based Parametric Video Quality Model For IPTV. Fifth International Workshop on Video Processing and Quality Metrics for Consumer Electronics (VPQM 2010). Intel, 20-25.
- 15 [16] WO 2012/076202 ("Method and apparatus for assessing the quality of a video signal during encoding and transmission of the video signal")
- [17] Rosenbluth, J. H. (AT&T) "ITU-T Delayed Contribution D.064: Testing the quality of connections having time varying impairments", 1998
- [18] Gros, L., Chateau, N. "Instantaneous and Overall Judgements for Time- Varying Speech Quality: Assessments and Relationships, Acta Acustica, Volumen 87, Número 3, Mayo/Junio 2001, pp. 367-377(11)
- 20 [19] Weiss, B., Moller, S., Raake, A., Berger, J., Ullmann, R. (2009). Modeling Conversational Quality for Time-varying Transmission Characteristics, Acta Acustica united with Acustica 95, 1140-1151.
- [20] WO/2012/013655 ("Method for estimation of the type of the group of picture structure of a plurality of video frames in a video stream")
- 25 [21] PCT/EP2011/067741 (Argyropoulos, S. et al, "Scene change detection for perceptual quality evaluation in video sequences"). PCT/EP201 1/067741 es un documento cuya fecha de presentación es anterior a la fecha de presentación de la presente solicitud, pero que no se publicó antes de la fecha de presentación de la presente solicitud.

REIVINDICACIONES

1. Procedimiento para estimar la calidad de percepción de una señal de vídeo digital, en el que el procedimiento comprende las etapas de:

(1a) extraer información de la secuencia de bits de vídeo, que es capturada antes de su decodificación;

5 (1b) obtener una estimación o estimaciones de uno o más factores de deterioro IF , para cada una de las estimaciones, una función de impacto adaptada para el factor de deterioro respectivo;

(1c) estimar la calidad percibida de la señal de vídeo digital usando la estimación o las estimaciones obtenidas en la etapa (1b);

10 en el que el procedimiento está caracterizado por que cada una de las funciones de impacto usadas en la etapa (1b) acepta como entrada un conjunto de parámetros \mathbf{q} dependientes del contenido calculados a partir de un conjunto de parámetros GOP/complejidad de la escena, en el que los parámetros GOP/complejidad de la escena pueden ser derivados a partir de la información de la cabecera del paquete y están disponibles en caso de secuencias de bits de vídeo codificadas.

15 en el que el conjunto de parámetros \mathbf{q} dependientes del contenido se deriva al menos a partir de un parámetro GOP/complejidad de la escena S'_{sc} , que denota el tamaño medio de fotograma I por cada escena, en el que, preferiblemente, el primer fotograma I de la primera escena se ignora;

20 en el que para estimar al menos uno de los factores de deterioro, preferiblemente, el impacto sobre la calidad debido a los artefactos de la compresión, se usa una función f_{IF} de impacto que depende de un parámetro q_1 dependiente del contenido calculado a partir de la inversa de la media ponderada del parámetro GOP/complejidad de la escena S'_{sc} sobre las escenas sc multiplicado por un coeficiente, y

en el que cada escena sc tiene un peso de $w_{sc} \times N_{sc}$, donde N_{sc} es el número de GOPs por cada escena y w_{sc} es un factor de ponderación adicional, en el que para las escenas que tienen el valor S'_{sc} más bajo: w_{sc} se establece a un valor mayor que 1, por ejemplo $w_{sc} = 16$, y para el resto de las escenas: w_{sc} se establece a un valor igual a 1

25 2. Procedimiento según la reivindicación 1, en el que:

el coeficiente es proporcional al número de píxeles por fotograma de vídeo nx y el número fr de fotogramas de vídeo por segundo.

3. Procedimiento según la reivindicación 2, en el que el parámetro q_1 dependiente del contenido viene determinado por

30

$$q_1 = \frac{\sum_{sc} w_{sc} \times N_{sc}}{\sum_{sc} S'_{sc} \times w_{sc} \times N_{sc}} \times \frac{nx \times fr}{1000} \cdot \left| \right.$$

35 4. Procedimiento según una cualquiera de las reivindicaciones 1 a 3, en el que los parámetros GOP/complejidad de la escena se calculan por cada grupo de imagen (GOP) o por cada escena de vídeo.

5. Procedimiento según una cualquiera de las reivindicaciones 1 a 4, en el que cada una de las funciones de impacto usadas en la etapa (1b) depende además de:

características técnicas de codificación o de red, por ejemplo, la tasa de bits, el número de fotogramas por segundo, el porcentaje de pérdida de paquetes o la proporción de pérdida en un GOP o una escena, y/o

40 coeficientes asociados con la función de impacto.

6. Procedimiento según una cualquiera de las reivindicaciones 1 a 5, en el que el conjunto de parámetros \mathbf{q} dependientes del contenido es derivado además desde al menos uno de los parámetros GOP/complejidad de la escena siguientes:

S^P_{gop} , que denota el tamaño medio de fotograma P por cada GOP;

S_{gop}^B , que denota los tamaños medios (de referencia) de fotograma B por cada GOP;

S_{gop}^b , que denota los tamaños medios de fotograma b no de referencia por cada GOP;

S_{gop}^{nol} , que denota los tamaños medios conjuntos de los fotogramas P, B y b por cada GOP;

B_{sc}^I , que denota la tasa de bits de los fotogramas I calculada por cada escena;

5 B_{sc}^P , que denota la tasa de bits de los fotogramas P calculada por cada escena;

B_{sc}^B , que denota la tasa de bits de los fotogramas B calculada por cada escena;

B_{sc}^b , que denota la tasa de bits de los fotogramas b calculada por cada escena;

B_{sc}^{nol} , que denota la tasa de bits de los fotogramas P, B y b calculada por cada escena.

10 7. Procedimiento según la reivindicación 6, en el que el conjunto \mathbf{q} de parámetros es derivado a partir de al menos uno de los parámetros GOP/complejidad de la escena siguientes:

$$S^{P/I} = S_{gop}^P / S_{sc}^I$$

$$S^{b/I} = S_{gop}^b / S_{sc}^I$$

$$S^{b/P} = S_{gop}^b / S_{gop}^P$$

$$S^{nol/I} = S_{gop}^{nol} / S_{sc}^I$$

15 $B^{P/I} = B_{sc}^P / B_{sc}^I$

$$B^{b/I} = B_{sc}^b / B_{sc}^I$$

$$B^{b/P} = B_{sc}^b / B_{sc}^P$$

$$B^{nol/I} = B_{sc}^{nol} / B_{sc}^I$$

20 8. Procedimiento según una cualquiera de las reivindicaciones 1 a 7, en el que la función f_{IF} de impacto que depende del parámetro $\mathbf{q} = q_1$ dependiente del contenido viene determinada por

$$f_{IF}(\mathbf{p}, \mathbf{q}, \boldsymbol{\alpha}) = \alpha_1 \times \exp(\alpha_2 \times p_1) + \alpha_3 \times q_1 + \alpha_4 \quad ,$$

en la que, preferiblemente, $\mathbf{p} = p_1$ es un parámetro que describe el número de bits por píxel y viene determinado, más preferiblemente, por

25
$$p_1 = \frac{\text{tasa de bits} \times 10^6}{nx \times fr} \quad , y$$

donde $\boldsymbol{\alpha} = (\alpha_1, \alpha_2, \alpha_3, \alpha_4)$ es el conjunto de coeficientes asociados a la función de impacto.

30 9. Procedimiento según una cualquiera de las reivindicaciones 1 a 8, en el que se usa una función f_{IF} de impacto, preferiblemente para estimar el impacto sobre la calidad debido a los artefactos de transmisión, que depende de un conjunto de parámetros dependientes del contenido $\mathbf{q} = (q_1, q_2)$, en el que cada componente q_j con $j \in \{1, 2\}$ del conjunto que se obtiene mediante una suma ponderada de los parámetros $\beta_{k,i}$ dependientes de los parámetros GOP/complejidad de la escena, en el que la suma ponderada para cada $j \in \{1, 2\}$ es calcula preferiblemente según

35
$$q_j = \sum_{k=1}^v \beta_{k,j} \times R_{k,j}$$

con pesos $R_{k,j}$.

10. Procedimiento según la reivindicación 9, en el que las ponderaciones vienen determinadas por

$$R_{k,j} = \sum_i r_i \times (T_k - t_i) \text{ para } j \in \{1, 2\}$$

5 en la que T_k es la duración de la pérdida de GOP k , t_i es la ubicación en el GOP de un evento de pérdida i y r_i denota la extensión espacial del evento de pérdida i , y en el que preferiblemente:

en caso de un sector por fotograma,

$$r_i = \frac{nap}{np} ; \text{ y}$$

en el caso de más de un sector por fotograma,

10

$$r_i = \frac{nlp}{np} + nle \times \frac{1}{2 \times nsl} ; \left| \right.$$

15

donde np es el número de paquetes en el fotograma, nap es el número de paquetes de secuencias de transporte afectados (TS) en el fotograma afectado, nlp es el número de paquetes perdidos en el fotograma, nle es el número de eventos de pérdida en el fotograma y nsl es el número de sectores en el fotograma

11. Procedimiento según las reivindicaciones 9 o 10, en el que:

el parámetro $\beta_{k,1}$ depende del parámetro GOP/complejidad de la escena $S^{nol/l}$; y/o

el parámetro $\beta_{k,2}$ depende del parámetro GOP/complejidad de la escena $S^{b/P}$.

20

12. Procedimiento según una cualquiera de las reivindicaciones 9 a 11, en el que los parámetros $\beta_{k,1}$ para cada $k \in \{1, \dots, v\}$ se obtienen mediante las etapas siguientes:

(12a) establecer $\beta_{k,1} = S^{nol/l}$;

(12b) en el caso $\beta_{k,1} \leq 0,5$, establecer $\beta_{k,1}$ a un valor $2x \beta_{k,1}$;

(12c) en el caso $\beta_{k,1} > 0,5$, establecer $\beta_{k,1}$ a un valor 1.

25

13. Procedimiento según una cualquiera de las reivindicaciones 9 a 12, en el que los parámetros $\beta_{k,2}$ para cada $k \in \{1, \dots, v\}$ se obtienen como $\beta_{k,2} = \max(0, -S^{b/P} + 1)$.

14. Procedimiento según una cualquiera de las reivindicaciones 9 a 13, en el que la función f_{IF} de impacto dependiente del conjunto de parámetros $\mathbf{q} = (q_1, q_2)$ dependientes del contenido viene determinada por

30

$$f_{IF}(\mathbf{p}, \mathbf{q}, \boldsymbol{\alpha}) = \alpha_1 \times \log\left(1 + \frac{\alpha_2 \times q_1 + \alpha_3 \times q_2}{p_1 \times p_2}\right) ,$$

en el que, preferiblemente, p_1 es un parámetro que describe el impacto sobre la calidad debido a los artefactos de compresión, preferiblemente, p_2 es el número de GOPs en la ventana de medición o la duración de la ventana de medición, y $\boldsymbol{\alpha} = (\alpha_1, \alpha_2, \alpha_3)$ es el conjunto de coeficientes asociados con la función de impacto.

35

15. Procedimiento según una cualquiera de las reivindicaciones anteriores, en el que la señal de vídeo es al menos parte de una secuencia de datos no interactiva, preferiblemente una secuencia de vídeo o audiovisual no interactiva, o al menos parte de una secuencia de datos interactiva, preferiblemente una secuencia de vídeo o audiovisual interactiva.

40

16. Procedimiento según una cualquiera de las reivindicaciones anteriores, en el que el procedimiento se combina con uno o más procedimientos para estimar el impacto sobre la calidad de percepción de una señal de vídeo digital por otros deterioros diferentes a la compresión y/o la transmisión, en el que la combinación es realizada preferiblemente usando al menos una función lineal y/o al menos una función multiplicativa de los procedimientos a

combinar.

17. Procedimiento según una cualquiera de las reivindicaciones anteriores, en el que el procedimiento se combina con uno o más procedimientos diferentes para estimar la calidad de la percepción de un vídeo digital debida a la compresión y/o la transmisión, en el que la combinación se realiza, preferiblemente, usando al menos una función lineal y/o al menos una función multiplicativa de los procedimientos a combinar.

18. Procedimiento para supervisar la calidad de una señal de vídeo digital transmitida con las etapas de:

(18a) transmitir la señal de vídeo desde un servidor al cliente;

(18b) en el lado del cliente, ejecutar el procedimiento para estimar la calidad de la percepción de una señal de vídeo digital según una cualquiera de las reivindicaciones 1 a 17;

(18c) transferir el resultado de la estimación de la etapa (18b) al servidor;

(18d) en el lado del servidor, supervisar la estimación de la calidad de la señal de vídeo transmitida; y

en el que el procedimiento comprende preferiblemente las etapas adicionales de:

(18e) analizar la calidad supervisada de la señal de vídeo transmitida, preferiblemente dependiendo de los parámetros de transmisión; y opcionalmente

(18f) cambiar los parámetros de transmisión en base al análisis de la etapa (18e) con el propósito de aumentar la calidad de la señal de vídeo transmitida.

19. Aparato para estimar la calidad de la percepción de una señal de vídeo digital, en el que el aparato comprende:

unos medios configurados para extraer información desde una secuencia de bits de vídeo capturada antes de la decodificación;

al menos un estimador de impacto;

un estimador de calidad configurado para estimar la calidad de percepción Q_v de la señal de vídeo:

en el que el aparato está caracterizado por que cada uno de los estimadores de impacto está configurado para estimar el impacto sobre la calidad debido a un factor de deterioro por medio de una función de deterioro que acepta como entrada un conjunto de parámetros \mathbf{q} dependientes del contenido calculados a partir de un conjunto parámetros GOP/complejidad de la escena, en el que los parámetros GOP/complejidad de la escena pueden derivarse a partir de la información de la cabecera del paquete y, de esta manera, están disponibles en el caso de secuencias de bits de vídeo codificado;

en el que el conjunto de parámetros \mathbf{q} dependientes del contenido es derivado al menos a partir de un parámetro GOP/ complejidad de la escena S'_{sc} , que denota el tamaño medio de fotograma l por cada escena, en el que, preferiblemente, el primer fotograma l de la primera escena se ignora;

en el que para estimar al menos uno de los factores de deterioro, preferiblemente el impacto sobre la calidad debido a los artefactos de compresión, se usa una función f_{IF} de impacto que depende de un parámetro q_1 dependiente del contenido que se calcula a partir de la inversa de la media ponderada del parámetro GOP/complejidad de la escena S'_{sc} a través de las escenas sc multiplicado por un coeficiente; y

en el que cada escena sc tiene un peso de $w_{sc} \times N_{sc}$, donde N_{sc} es el número de GOPs por escena y w_{sc} es un factor de ponderación, en el que para las escenas que tienen el valor S'_{sc} más bajo: w_{sc} se establece a un valor mayor que 1, por ejemplo $w_{sc} = 16$, y para el resto de las escenas: w_{sc} se establece a un valor igual a 1.

20. Aparato según la reivindicación 19, que además está configurado para estimar la calidad de percepción de una señal de vídeo digital usando un procedimiento según una cualquiera de las reivindicaciones 2 a 17.

21. Decodificador conectable a un receptor para recibir una señal de vídeo digital, en el que el decodificador comprende el aparato según las reivindicaciones 19 o 20.

22. Sistema para supervisar la calidad de una señal de vídeo digital transmitida, en el que el sistema comprende un servidor y un cliente, y el sistema está configurado para ejecutar el procedimiento según la reivindicación 18.

23. Sistema según la reivindicación 22, en el que:

el cliente está configurado como el aparato según la reivindicación 19 o 20; y /o

el cliente comprende un aparato según la reivindicación 19 o 20.

24. Sistema según la reivindicación 22, que comprende además el decodificador según la reivindicación 21, en el que el decodificador está conectado al cliente.

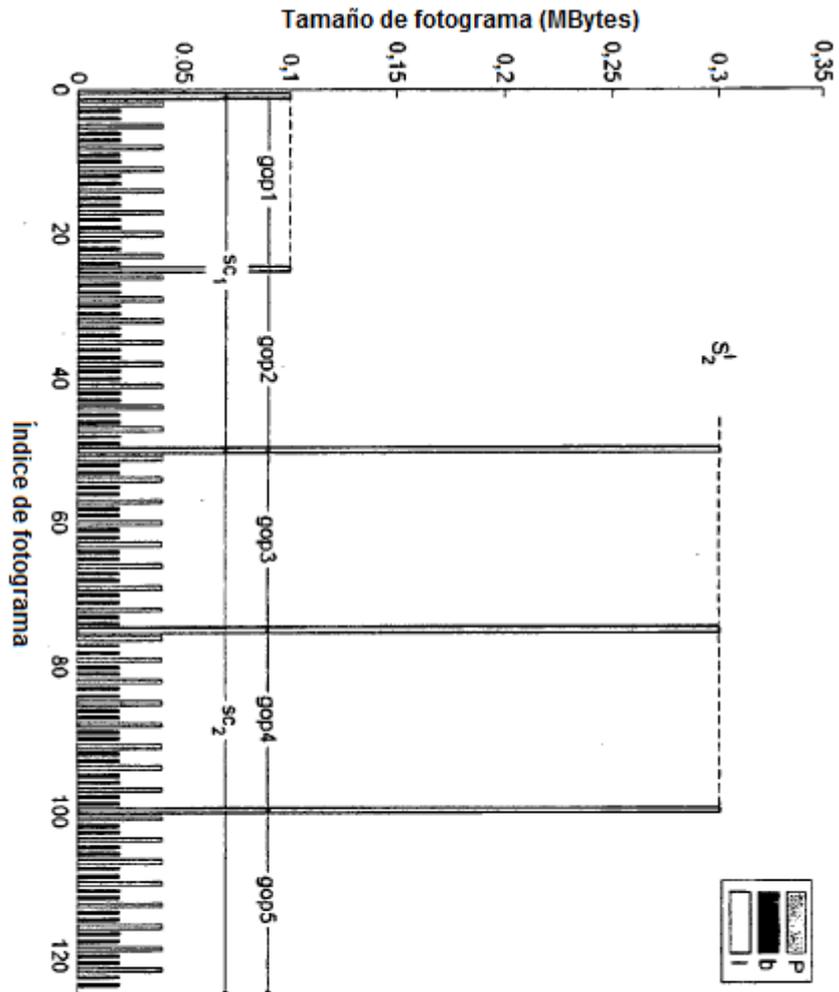


Fig. 1

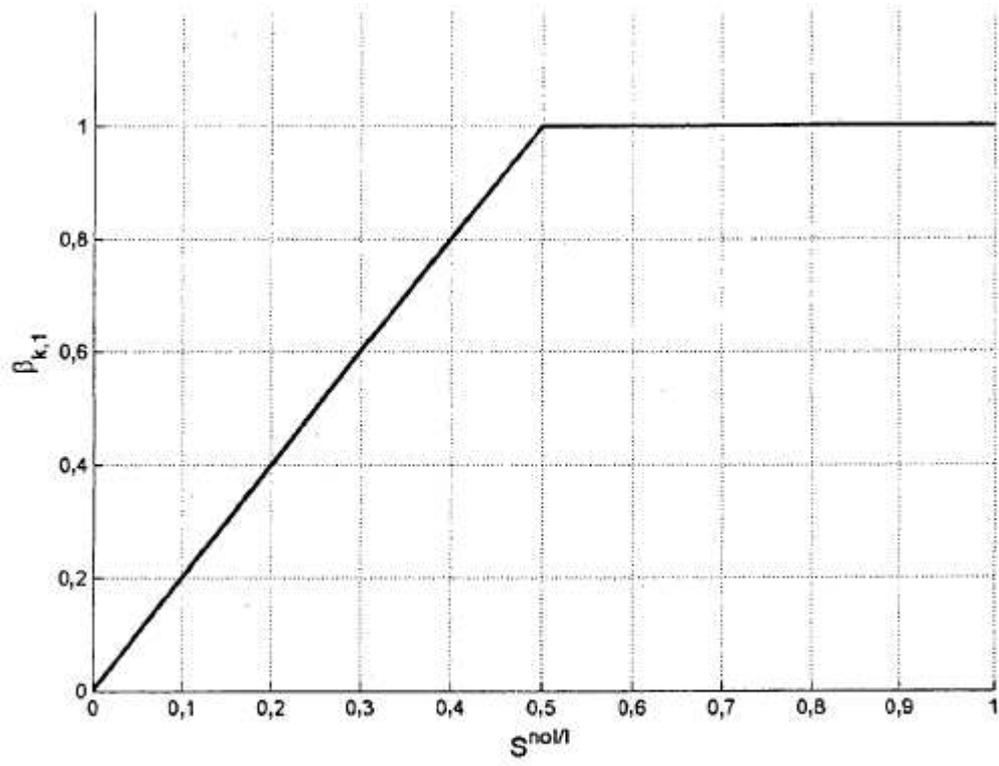


Fig. 2

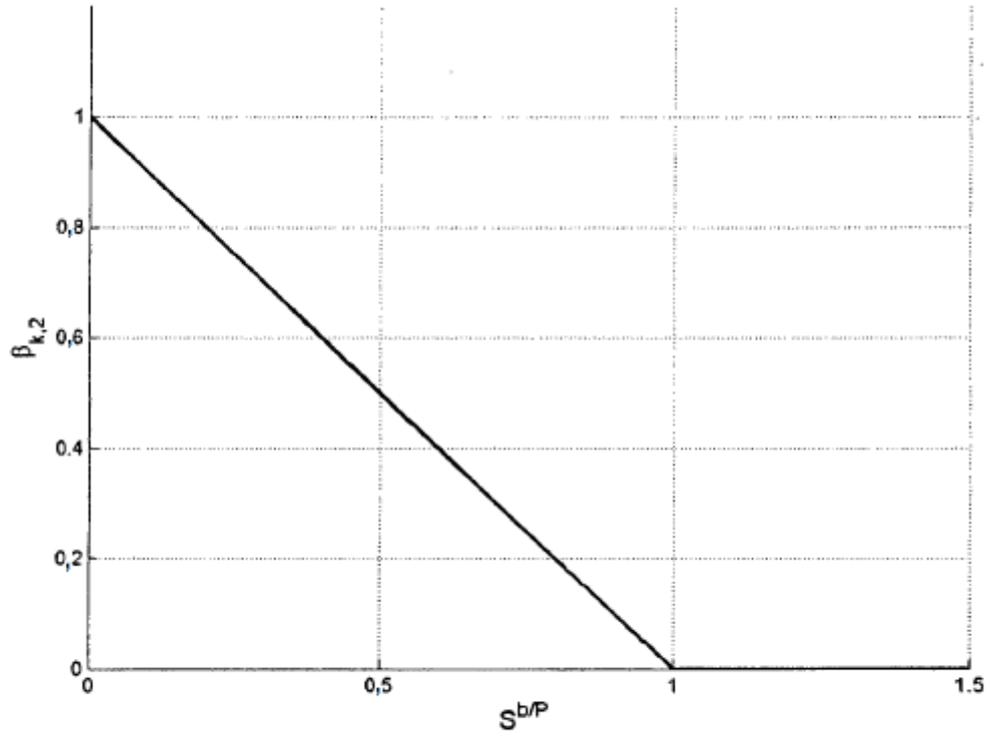


Fig. 3