

19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 586 946**

51 Int. Cl.:

H04N 17/00 (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

96 Fecha de presentación y número de la solicitud europea: **30.07.2010 E 15151227 (4)**

97 Fecha y número de publicación de la concesión europea: **25.05.2016 EP 2866447**

54 Título: **Método y aparato para evaluar la calidad de una secuencia de vídeo mediante la sincronización temporal del flujo de bits encriptado de un codificador de vídeo con la secuencia de vídeo procesada obtenida por un decodificador de vídeo exterior**

45 Fecha de publicación y mención en BOPI de la traducción de la patente:
19.10.2016

73 Titular/es:

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

72 Inventor/es:

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

74 Agente/Representante:

DE ELZABURU MÁRQUEZ, Alberto

ES 2 586 946 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

Método y aparato para evaluar la calidad de una secuencia de vídeo mediante la sincronización temporal del flujo de bits encriptado de un codificador de vídeo con la secuencia de vídeo procesada obtenida por un descodificador de vídeo exterior

Campo de la Invención

La invención se refiere a métodos y aparatos para sincronizar temporalmente el flujo de bits de vídeo encriptado de un descodificador de vídeo exterior con la secuencia de vídeo procesada proporcionada por este descodificador de vídeo exterior.

Antecedentes de la Invención

La proliferación de aplicaciones de comunicación por vídeo a lo largo de los últimos años ha necesitado el desarrollo de fuertes medidas en la calidad del vídeo para evaluar la Calidad de la Experiencia (QoE), definida como la calidad del servicio percibida por el usuario. La evaluación de la calidad del vídeo es un aspecto crítico del diseño, la planificación y la monitorización eficientes de los servicios (por ejemplo, el Internet Protocol Television – IPTV) por parte de los proveedores del contenido.

Actualmente, los sistemas de evaluación de la calidad del vídeo híbridos emplean una combinación de información del paquete, información del flujo de bits y la secuencia de vídeo generada por un descodificador de vídeo exterior, por ejemplo, la secuencia de vídeo descodificada por una “set-top-box” (STB), o descodificador de televisión, y mostrada en el dispositivo de pantalla del usuario. En general, en un algoritmo de evaluación de la calidad del vídeo híbrido las características extraídas o calculadas del flujo de bits (por ejemplo, vectores de movimiento, tipos de macrobloques, coeficientes de transformación, parámetros de cuantificación, etc.) y la información extraída de las cabeceras de los paquetes (por ejemplo, velocidad de transferencia de los bits, pérdida de paquetes, retrasos, etc.) es combinada con las características extraídas del campo de los píxeles a partir de la secuencia de vídeo generada por un descodificador de vídeo exterior. Sin embargo, en muchas aplicaciones, el flujo de bits de vídeo transmitido está encriptado debido a cuestiones de derechos de autor y la clave de la descryptación sólo está incluida en un dispositivo de hardware (típicamente el STB del usuario final). Por tanto, los modelos de evaluación de la calidad no pueden tener acceso a la información útil del paquete y a las características relacionadas del flujo elemental (ES) que está encapsulado en los paquetes; en lugar de eso, usan información extraída de las cabeceras de los paquetes, esto es, de las cabeceras del flujo de transporte (MPEG-2 TS), del Real-Time Transport Protocol (RTP), o del flujo elemental empaquetado (PES), dependiendo del nivel de encriptación. Dichos modelos, llamados modelos paramétricos o basados en la cabecera del paquete, no están destinados explícitamente a aspectos tales como la calidad de la fuente o las implementaciones del codificador y del reproductor. El audio, vídeo y los modelos de la calidad audiovisual paramétricos están siendo desarrollados y estandarizados actualmente por el ITU-T Study Group 12 bajo el nombre provisional de P.NAMS.

Resultará evidente que las características extraídas de las cabeceras de los paquetes necesitan corresponderse temporalmente con las características extraídas de la secuencia de vídeo obtenida por el descodificador de vídeo exterior, de otra manera, debido a la pérdida de la sincronización temporal, la evaluación de la calidad del vídeo percibida no sería precisa. Así, el primer paso en cada algoritmo de evaluación de la calidad del vídeo híbrido es la sincronización del flujo de bits de vídeo (encriptado) con la secuencia de vídeo obtenida por el descodificador de vídeo exterior.

Se muestra un diagrama de bloques de un sistema de evaluación de la calidad del vídeo híbrido en la Figura 1. El dispositivo de descodificación exterior, por ejemplo, el STB situado en la parte del usuario final, descodifica el flujo de bits recibido y genera la secuencia de vídeo procesada (PVS) mostrada por el dispositivo de salida (típicamente un aparato de TV o un monitor). Es de destacar que el STB puede descodificar el flujo de bits encriptado de entrada ya que éste contiene la clave de descryptación requerida. En paralelo, un dispositivo de sondeo captura el flujo de bits encriptado de entrada, y a continuación lo examina y analiza para extraer y calcular características relacionadas con la calidad a partir del análisis de las cabeceras de los paquetes. El dispositivo de sondeo incorpora un módulo para la sincronización temporal entre el flujo de bits de entrada y la secuencia de vídeo obtenida del descodificador de vídeo exterior. Esta información de la sincronización es empleada basándose en el algoritmo de evaluación de la calidad del vídeo (junto con las características del flujo de bits de entrada y las características de la PVS) dentro del dispositivo de sondeo para evaluar la calidad de la secuencia de vídeo y calcular una estimación objetiva de la “Mean Opinion Score” (MOS) o Calidad Subjetiva Media.

A partir de ahora en adelante, la expresión “Processed Video Sequence” (PVS), o Secuencia de Vídeo Procesada, significa la secuencia de vídeo generada por el descodificador de vídeo exterior. En general, las razones principales para la asincronía temporal entre el flujo de bits y la PVS son el retraso y las pérdidas de paquetes que pueden ocurrir durante la transmisión del flujo de bits de vídeo por una red. Cuando el flujo de vídeo es transmitido por una red “best effort”, o “la mejor posible”, tal como Internet, el tiempo de llegada de cada paquete no es constante y puede variar significativamente. La variación a lo largo del tiempo de la latencia del paquete a través de la red se llama “jitter”, o fluctuación. Para asegurar una reproducción uniforme de la secuencia sin fluctuación, la mayoría de los sistemas de vídeo emplean un buffer o circuito intermedio de eliminación de fluctuaciones. El flujo de bits

recibidos es escrito en el buffer de entrada basándose en el tiempo de llegada de cada paquete, y los datos de la imagen correspondientes a un fotograma o cuadro son leídos en él y llevados al descodificador a intervalos de tiempo predeterminados correspondientes al período del fotograma. El tiempo en pantalla de cada imagen está determinado por el campo de la indicación de tiempo grabado en la cabecera del paquete. Esto es, el valor de la indicación de tiempo se corresponde con el período de tiempo que transcurre desde la detección del código de inicio de la imagen hasta el tiempo en pantalla de la imagen.

En el sistema de descodificación de vídeo descrito anteriormente, el tiempo en pantalla de cada imagen de vídeo se determina según los datos incluidos en el flujo de bits de vídeo para la determinación del tiempo en pantalla. Ya que el tiempo en pantalla de una imagen no es fijo, la PVS no siempre se puede hacer que se corresponda exactamente con el flujo de bits original. Además, la PVS es generada por un descodificador de vídeo exterior (que está, por ejemplo, integrado en el STB o en el dispositivo de pantalla del espectador) que no forma parte del sistema de evaluación de la calidad del vídeo y tiene propiedades y funcionalidad desconocidas ya que es típicamente proporcionado por un tercero.

En la bibliografía, el problema de la sincronización temporal entre una fuente y una secuencia de vídeo distorsionada ha sido estudiado previamente y se le denomina también registro de vídeo. En "Comparison of matching strategies for temporal frame registration in the perceptual evaluation of video quality", Proc. of the Second International Workshop on Video Processing and Quality Metrics for Consumer Electronics de enero de 2006, de M. Barkowsky, R. Bitto, J. Bialkowski y A. Kaup, se presenta y examina una comparación entre la correspondencia de bloques y la correlación de fase para el registro de vídeo en términos de resultados y de complejidad. También, se presentó un algoritmo de correspondencia de fotogramas para ocuparse de la retirada, inserción, mezcla y compresión de datos en "Video frame-matching algorithm using dynamic programming" del Journal of Electronic Imaging, SPIE, 2009, de Y. Y. Lee, C. S. Kim y S. U. Lee, basado en la minimización de una función de costo de la correspondencia usando una programación dinámica. En "Fast video temporal alignment estimation", (US-B1-6 751 360), de J. Lu, se presentó un método de estimación de alineación temporal rápida para alinear temporalmente un vídeo distorsionado con un vídeo original correspondiente para medidas de la calidad del vídeo. Cada secuencia de vídeo es transformada en una curva de firma calculando un punto de datos para cada fotograma como una relación cruzada entre dos fotogramas seguidos. La mala alineación temporal del vídeo distorsionado es determinada a continuación encontrando el valor máximo de la correlación cruzada normalizada entre las curvas de firma de las secuencias de vídeo examinadas. Otro método para identificar la correspondencia espacial, temporal y del histograma entre dos secuencias de vídeo ha sido descrito en "Video registration based on local prediction errors", (US-B2-7 366 361), de H. Cheng. La PVS es alineada con la secuencia de vídeo de referencia generando un mapeado a partir de un conjunto seleccionado de uno o más fotogramas originales con el conjunto procesado para que cada mapeado minimice un error de predicción local. En "Systems and methods for robust video temporal registration", (US-A-2008/0253689), de K. Ferguson se producen medidas de destilación de fotograma y subimagen a partir de secuencias de vídeo de referencia y prueba. A continuación, son alineadas linealmente usando el coeficiente de relación cruzada de Parson local entre fotogramas. Adicionalmente, en "Spatial and temporal alignment of video sequences", (US-A-2007/0097266), de C. Souchard, se define una función de movimiento para describir el movimiento de un conjunto de píxeles entre los fotogramas de la prueba y la secuencia de vídeo de referencia y se usa una transformación para alinear las dos imágenes.

En "Method for controlling digital television metrology equipment", US 6.618.077 B1, 2003, de J. Baina et al., se propone un método para la extracción de parámetros a partir de un MPEG-2 Transport Stream, o Flujo de Parámetros, para generar señales de sincronización. Sin embargo, este método es solamente aplicable cuando el flujo elemental de vídeo está paquetizado en un flujo de transporte MPEG-2 y no puede ser aplicado con un protocolo de transporte arbitrario. Por el contrario, el método propuesto puede ser aplicado a cualquier flujo de bits de vídeo sin necesidad de un protocolo específico de transporte o de aplicación. Además, el método de Baina et al. anteriormente mencionado proporciona señales de sincronización a un algoritmo de monitorización de la calidad del vídeo para indicar qué imágenes (fotogramas de vídeo) de la señal de vídeo deben ser usadas para la predicción de la calidad. En contraste con lo anterior, el método propuesto en esta memoria identifica la parte del flujo de bits (y de las características extraídas de él) que se corresponden con cada imagen que está siendo considerada de la secuencia de vídeo obtenida de un descodificador de vídeo exterior. Finalmente, el método de Baina et al. no explota la información del flujo de bits encriptada para sincronizar el flujo de bits de vídeo con la imagen procedente del descodificador de vídeo exterior, mientras que la invención propuesta explota el flujo de bits encriptado (y las características extraídas de él) para realizar la sincronización. La explotación del flujo de bits de vídeo permite la consideración de los efectos de las pérdidas de paquetes y puede ser aplicada en caso de errores de transmisión. Se muestra tecnología adicional relacionada en el documento US 2009/262198 A1.

Deberá tenerse en cuenta que a partir de ahora, las expresiones "fotograma" e "imagen" serán usadas de manera intercambiable y significarán una imagen (que puede ser un fotograma o un campo) de la PVS.

Compendio de la Invención

El objetivo de la invención es proporcionar un método y un aparato para sincronizar temporalmente el flujo de bits de vídeo encriptado de un descodificador de vídeo, por ejemplo, un STB, con la PVS de un descodificador de vídeo exterior y permitir su sincronización. Se consigue este objetivo a partir de las características de las reivindicaciones.

La información de la sincronización puede ser usada, por ejemplo, en un algoritmo de evaluación de la calidad del vídeo híbrido para asegurar que las características del flujo de bits extraídas temporalmente se corresponden con las imágenes en pantalla específicas durante la medida de la ventana de tiempo para la que se valora la calidad que percibe el espectador.

5 La invención proporciona un método para sincronizar temporalmente el flujo de bits de entrada encriptado de un descodificador de vídeo con la secuencia de vídeo procesada obtenida por un descodificador de vídeo exterior, el método comprende los pasos de:

- 10 a) capturar y suministrar el flujo de bits de entrada encriptado a un módulo de análisis;
 b) analizar en el módulo analizador del flujo de bits, la información de la cabecera del paquete del flujo de bits de entrada encriptado, y calcular el tamaño y el tipo de fotogramas contenidos en el flujo de bits;
 c) extraer, basándose en los tamaños y tipos de los fotogramas calculados, características que son representativas del contenido representado en los fotogramas de vídeo;
 15 d) extraer el tipo del contenido y las características respectivas representativas del contenido de las imágenes de la secuencia de vídeo procesada;
 e) comparar las características extraídas en el paso d) de la imagen en curso de la secuencia de vídeo procesada con las características extraídas en el paso c) de los fotogramas analizados para conseguir calcular la parte del flujo de bits que se corresponde con la PVS en curso.
 20 f) en el que la parte del flujo de bits que se corresponde con la PVS en curso es además introducida en un módulo de evaluación de la calidad del vídeo; y
 g) estimar, mediante el módulo de evaluación de la calidad del vídeo, una puntuación de la calidad subjetiva basada en la información obtenida del flujo de bits y de la secuencia de vídeo procesada.

25 Las características extraídas en los pasos c) y d) son preferiblemente características que varían con el tiempo, y esos pasos pueden realizarse en el campo de los píxeles.

El paso b) puede detectar además fotogramas que han sido afectados por la pérdida de paquetes, y puede comprender además los pasos de:

- 30 h) examinar la imagen en curso de la secuencia de vídeo procesada en busca de artefactos generados debido a pérdidas de paquetes; y
 i) hacer que la secuencia de vídeo procesada se corresponda con el fotograma correspondiente del flujo de bits, en donde, si se detectó que la imagen en curso de la secuencia de vídeo procesada en el paso g) contenía artefactos que fueron insertados como resultado de una pérdida de paquete, se hace corresponder esta imagen a continuación con la parte del flujo de bits correspondiente al fotograma en el que se encontró en el paso b) que contenía la pérdida del paquete.

Según un aspecto adicional, la invención proporciona un aparato para sincronizar temporalmente el flujo de bits de entrada encriptado de un descodificador de vídeo con la secuencia de vídeo procesada obtenida mediante un descodificador de vídeo exterior, el aparato comprende:

40 un analizador del flujo de bits que recibe el flujo de bits de entrada encriptado, el analizador de flujo de bits está configurado para examinar y analizar el flujo de bits de entrada, y
 45 un módulo de sincronización configurado para sincronizar el flujo de bits analizado con la secuencia de vídeo procesada,
 caracterizado por que
 el módulo de sincronización comprende:

50 un primer módulo de sincronización configurado para extraer, basándose en los tamaños y tipos de fotogramas calculados, características representativas del contenido mostrado en los fotogramas de vídeo,
 un segundo módulo de extracción configurado para extraer el tipo del contenido y las características respectivas representativas del contenido de las imágenes de la secuencia de vídeo procesada, y
 55 una unidad de comparación conectada a la primera unidad de extracción y a la segunda unidad de extracción, la unidad de comparación está configurada para

60 comparar las características extraídas de la imagen en curso de la secuencia de vídeo procesada con las características extraídas de los fotogramas analizados, y para determinar el fotograma correspondiente n*.

en donde la unidad de comparación está configurada para introducir la parte del flujo de bits que se corresponde con la PVS en curso en un módulo de evaluación de la calidad del vídeo; y
 en donde el módulo de evaluación de la calidad del vídeo está configurado para estimar una puntuación de la calidad subjetiva basada en la información obtenida del flujo de bits y de la secuencia de vídeo procesada.

Volviendo al objetivo de la invención, éste se puede conseguir en términos generales mediante la provisión de un dispositivo (sonda) que, según la invención presente, incluye un analizador del flujo de bits para examinar el flujo de bits de entrada, un módulo de sincronización responsable de sincronizar temporalmente el flujo de bits con la PVS.

5 El dispositivo responsable de la sincronización temporal del flujo de bits de vídeo y de la PVS comprende un buffer que almacena las características extraídas del flujo de bits analizado. Este buffer debe ser capaz de almacenar un número específico de dichas características (extraídas del flujo de bits de entrada encriptado) que se corresponden con número de imágenes N_{dec} para que las características extraídas de la imagen en curso de la PVS que está siendo examinada puedan ser comparadas con este número específico de características extraídas previamente.

10 El módulo responsable de la sincronización de la PVS y del flujo de bits de vídeo realiza las operaciones siguientes:

- a) compara las características extraídas de la PVS con las características extraídas de cada una de las imágenes contenidas en el buffer
- 15 b) calcula una métrica de distancia entre las características de la PVS y las características extraídas del buffer explotando la información extraída de las cabeceras de los paquetes
- c) determina la mejor imagen correspondiente del buffer de vídeo y la lleva (junto con sus características extraídas) al módulo de evaluación de la calidad del vídeo responsable de estimar la calidad del vídeo basándose en la información del flujo de bits y la PVS.

20 A continuación, se describe la invención con mayor detalle.

En primer lugar, se describe una realización de la sincronización temporal entre el flujo de bits y la PVS haciendo referencia a la Figura 2.

25 La sincronización temporal entre el flujo de bits y la secuencia de vídeo procesada consta de dos pasos. Inicialmente, en el primer paso, el aparato según la invención, esto es, un dispositivo de sondeo, captura el flujo de datos y lo alimenta al analizador del flujo de bits que extrae la siguiente información para cada imagen:

- 30 a. el tipo de fotograma y el tamaño de cada imagen de la secuencia de vídeo basados en la información de la cabecera del paquete
- b. los paquetes perdidos debido a errores de transmisión y las áreas correspondientes dentro de cada fotograma que han sido afectadas por la pérdida de los paquetes; y
- 35 c. características representativas del contenido de cada imagen

En el segundo paso se extraen características representativas del contenido representado en las imágenes de la secuencia de vídeo procesada, y a continuación son comparadas con las características extraídas del flujo de bits encriptado. El objeto de esta comparación es identificar los fotogramas correspondientes del flujo de bits encriptado (y los paquetes que transmiten la información de este fotograma) que se corresponden con las imágenes en curso de la PVS que están siendo examinadas para que las características del flujo de bits y la PVS puedan ser alineadas temporalmente.

Finalmente, se emplea un módulo de evaluación de la calidad del vídeo híbrido que combina información del flujo de bits encriptado y de la PVS para estimar la calidad de la secuencia de vídeo representada.

45 Sincronización basada en la correspondencia de características basadas en el fotograma
La primera realización del método para la sincronización entre el flujo de bits encriptado y la imagen descodificada (PVS) capturada de un descodificador de vídeo exterior está destinada al caso en el que no se pierden paquetes en el flujo de bits y que se muestra en la Figura 2.

50 En el primer paso, el flujo de bits de vídeo es alimentado a un módulo responsable del análisis de la información de la cabecera del paquete, y del cálculo del tamaño y del tipo de los fotogramas (intra, predictivo, o bipredictivo). El método para la extracción del tipo de fotogramas de vídeo encriptado basado en su tamaño no se reivindica como una invención de la patente presente. Por ejemplo, puede usarse el método propuesto en la patente europea EP 2 413 535 A1 del presente solicitante presentada el 30 de julio de 2010. Los detalles de la misma se encuentran en el Anexo 1 de esta descripción.

60 En el paso segundo, se usa la información sobre los tamaños y tipos de los fotogramas para extraer características (por ejemplo, que varían con el tiempo) que son representativas del contenido mostrado en los fotogramas de vídeo. Resultará evidente, que ya que la información útil no está disponible debido a la encriptación, solamente puede estimarse una aproximación de la clase del contenido. No se reivindica el método para estimar la clase del contenido basado en los tamaños de los fotogramas y en los tipos de los fotogramas como una invención de la patente presente. Con este objeto puede usarse cualquier método disponible, por ejemplo, el "Method and system for content estimation of packet video streams", WO 2009/012297 A1, de A. Clark. Los detalles de este método se encuentran en el Anexo 2 de esta descripción.

De manera similar, el tipo del contenido y las características respectivas (por ejemplo, que varían con el tiempo) representativas del contenido son extraídas de las imágenes de la secuencia de vídeo procesada. Ya que la invención presente está solamente destinada a los aspectos de la sincronización entre el flujo de bits y la secuencia de vídeo procesada, el método para la extracción de características de la imagen (en el campo de los píxeles) y la detección del tipo del contenido no se encuentran dentro de su alcance.

En el paso tercero, se hace que las características extraídas de la imagen en curso de la secuencia de vídeo procesada se correspondan con las características extraídas de los fotogramas que fueron analizados por el módulo de análisis de la cabecera del paquete. Ha de tenerse en cuenta que la característica correspondiente y la información de sincronización sobre la parte del flujo de bits que se corresponde con la PVS en curso que está siendo examinada pueden ser usadas como entrada al módulo responsable de la evaluación de la calidad del vídeo. Según se muestra en la Figura 2, un módulo de evaluación de la calidad del vídeo híbrido puede estar presente combinando información del flujo de bits y la secuencia de vídeo procesada para evaluar la calidad de la secuencia de vídeo.

Sincronización temporal basada en la activación del evento de pérdida del paquete

En esta realización de la invención presente, se supone que el flujo de bits y la secuencia de vídeo procesada están sincronizados cuando no ocurren errores (por ejemplo, debido a pérdidas de paquetes), basándose en el método descrito en la subsección previa y considera el caso en el que se requiere sincronización cuando se detecta una pérdida de paquete.

En el primer paso, el flujo de bits de vídeo es alimentado al módulo responsable del análisis de la información de la cabecera del paquete, del cálculo de los tipos de los fotogramas y de la detección de los fotogramas que han sido afectados por la pérdida de paquetes. El tipo de todos los fotogramas es detectado y de esta manera se determina si una pérdida de paquete se propaga o no a sus fotogramas siguientes debido al proceso de predicción.

En el paso siguiente, la imagen en curso de la secuencia de vídeo procesada es examinada por si existen distorsiones que pueden haber sido generadas debido a pérdidas de paquetes. Debe tenerse en cuenta que este algoritmo debe ser capaz de distinguir entre distorsiones causadas por pérdidas de paquetes y distorsiones que fueron producidas como resultado de una compresión "lossy", o sea, con pérdida de la secuencia del vídeo original. El algoritmo para la detección y puntuación de distorsiones está fuera del alcance de la invención presente. Por ejemplo, el método de G. Valensize et al., publicado en "Estimating channel-induced distortion in H.264/AVC video without bitstream information", QoMEX, 2010, podría ser usado con este objeto.

En el tercer paso se realiza la correspondencia de la secuencia de vídeo procesada con los fotogramas correspondientes del flujo de bits. Si se detectaron que las imágenes en curso de la secuencia de vídeo procesada contenían distorsiones que fueron causadas como resultado de una pérdida de paquete, entonces se hace corresponder esa secuencia de imágenes con la parte del flujo de bits correspondiente a los fotogramas en los que se encontró (en el paso inicial) que contenían las degradaciones por pérdida de paquetes. En este paso, la correspondencia puede no ser realizada necesariamente basándose en una imagen, sino más bien basándose en hacer que se corresponda la secuencia de fotogramas erróneos con su parte correspondiente del flujo de bits. Esto permite una estimación más precisa ya que considera el efecto de la propagación temporal de errores basándose en los tipos de fotogramas.

Finalmente, se calcula la información de la sincronización (esto es, el fotograma en curso de la PVS junto con su parte correspondiente del flujo de bits de vídeo). Esta información puede ser dirigida como entrada al módulo responsable de la evaluación de la calidad del vídeo.

La técnica anterior mencionada en la parte introductoria precedente considera que el problema de la sincronización temporal entre dos secuencias de vídeo debido al salto, repetición, supresión, mezcla y compresión de datos de los fotogramas que puede resultar debido a la transmisión de la secuencia de vídeo por una red errónea o por el procesamiento de la secuencia. Por el contrario, la invención presente considera la sincronización de la secuencia de vídeo procesada con el flujo de bits encriptado, por ejemplo, en la parte del usuario final, y, de esta manera, las únicas fuentes de asincronía son el retraso introducido por el dispositivo de descodificación y el analizador del flujo de bits y las diferentes técnicas de enmascaramiento del error aplicadas por el descodificador exterior (STB) y el descodificador dentro del dispositivo de sondeo cuando se detectan las pérdidas de paquetes en el flujo de bits de entrada. Además, la explotación del flujo de bits y la extracción de la característica respectiva permiten considerar los efectos de las pérdidas de paquetes en la sincronización.

La Figura 1 muestra un dispositivo de sondeo adicional para la sincronización temporal del flujo de bits de entrada encriptado y la secuencia de vídeo obtenida por un descodificador de vídeo exterior; La Figura 2 muestra la sincronización entre el flujo de bits encriptado y la imagen descodificada obtenida por un descodificador de vídeo exterior, basada en la extracción de características del flujo de bits y de la PVS según una realización preferida de la invención.

Aunque la invención ha sido ilustrada y descrita en detalle en los dibujos y en la descripción precedente, dicha

ilustración y descripción deben ser consideradas como ilustrativas o ejemplares y no como restrictivas. Deberá entenderse que personas con una experiencia ordinaria en la materia pueden hacer cambios y modificaciones dentro del alcance de las reivindicaciones siguientes. En particular, la invención presente cubre realizaciones adicionales con cualquier combinación de características de diferentes realizaciones descritas anterior y posteriormente.

Además, en las reivindicaciones la expresión “comprende” no excluye otros elementos o pasos, y el artículo indefinido “un o una” no excluye una pluralidad. Una sola unidad puede cumplir las funciones de varias características expuestas en las reivindicaciones. Las expresiones “esencialmente”, “sobre”, “aproximadamente” y otras similares en conexión con un atributo o un valor en particular definen también exactamente el atributo o exactamente el valor, respectivamente. Cualquier signo de referencia de las reivindicaciones no debe ser tomado como limitador del alcance de la invención.

Anexo 1

Método para la extracción del tipo de fotogramas de vídeo encriptado basado en sus tamaños (Citación de los párrafos [0013] a [0021] y [0026] a [0061] de la patente europea EP 2 413 535 A1)

Un método para estimar el tipo de la estructura del Grupo de Imagen, GoP, de una pluralidad de fotogramas de vídeo de un flujo de vídeo, puede comprender los pasos de:

- a. capturar los tamaños de los fotogramas en bytes de cada fotograma de vídeo siguiente a un Intra-fotograma inicial, fotograma I, para obtener un grupo de tamaños de fotogramas;
- b. convertir, después de un número de fotogramas, el grupo de tamaños de fotogramas obtenido en el paso a) en un grupo de ceros y unos, en donde el cero significa tamaños pequeños de fotogramas, según se asume para fotogramas bidireccionales, fotogramas B, y uno significa tamaños grandes de fotogramas, según se asume para fotogramas predichos, fotogramas P;
- c. hacer que se correspondan el grupo de ceros y de unos obtenido en el paso b) con un número de pautas predefinidas de números binarios, dichas pautas predefinidas de números binarios presentan estructuras GoP posibles;
- d. convertir el resultado de dichas correspondencias del paso c) para formar un valor de puntuación único; y
- e. determinar la pauta particular del número de pautas predefinidas de números binarios que tienen el mejor valor de puntuación, según una métrica predefinida.

En el paso a), los límites de los fotogramas de vídeo del flujo de bits pueden ser detectados o bien observando los cambios de ciertos indicadores de tiempo de la capa de transporte, o por ciertos bits de las estructuras de las cabeceras que son usados para señalar nuevos fotogramas y decidir Intra-fotogramas a partir de otros fotogramas.

Se prefiere que en el paso b) el número de fotogramas comprenda todos los fotogramas hasta el siguiente fotograma I.

En el paso b) la conversión de los tamaños de fotogramas continuos a números binarios comprende de preferencia el uso de un valor de referencia que depende de características constantes o de fluctuación lenta, o de un valor de referencia adaptativo dinámicamente, que son de preferencia dependientes de los valores de los tamaños de los fotogramas capturados previamente y/u otras características dinámicas.

Las pautas básicas predefinidas de los números binarios muestran estructuras GoP posibles para números diferentes de fotogramas B consecutivos de codificación basada en el fotograma, codificación basada en el campo, codificación GoP abierta, o codificación GoP cerrada.

En el paso c) se prefiere que las pautas básicas predefinidas cortas estén repetidas con un posible truncamiento de la última pauta básica hasta que la serie de pautas básicas repetida tenga la misma longitud que la serie de tamaños de fotogramas expresada en binario del grupo de los tamaños de fotogramas.

Además, en el paso c) la correspondencia puede ser realizada por una operación “nor exclusivo” destinada al elemento, en donde en el paso d) el resultado de cada “nor exclusivo” es añadido para formar el valor de la puntuación, y en donde en el paso e) la métrica del valor de la puntuación se corresponde con el valor máximo.

Según una realización, en el paso e) para la pauta que tiene el mejor valor de la puntuación se selecciona la estructura GoP, el estatus del campo/fotograma y el estatus del GoP abierto/GoP cerrado asociados como resultado de la estimación de la pauta del GoP del GoP en curso.

De preferencia, para los siguientes GoPs o bien se capturan nuevas series de tamaños de fotogramas, o se usa cualquier tipo de medias del GoP haciendo la media de los tamaños de los fotogramas de los grupos previos y en curso de los tamaños de fotogramas.

Después de la estimación de la pauta del GoP se lleva a cabo una puntuación adicional del tipo del fotograma, en donde cada conjunto de fotogramas B consecutivos entre dos fotogramas P que consiste en al menos 3 fotogramas B es examinado adicionalmente por si los tamaños de los fotogramas según el resultado del paso a) en estas posiciones de dichos conjuntos que pueden contener fotogramas B de referencia son significativamente mayores que en todas las otras posiciones de los fotogramas B de dichos conjuntos, en donde, si éste es el caso de un número adecuado de dichos conjuntos de fotogramas B del grupo de tamaños de fotogramas, se asume que dichos fotogramas son verdaderamente fotogramas B de referencia y se asume que la estructura GoP estimada es una estructura GoP de una así llamada secuencia de codificación jerárquica.

5
10 A continuación se describe el método con más detalle.

15 En un primer paso, se estima o calcula el número de bytes de cada fotograma de vídeo. Existen varios métodos diferentes apropiados para realizar esta tarea. Por ejemplo, pueden explotarse los cambios de ciertos indicadores de tiempo de la capa de transporte para detectar el comienzo de fotogramas nuevos, o se utilizan ciertos bits de las estructuras de las cabeceras que se usan para señalar nuevos fotogramas. En general, estas propiedades de los flujos de bits están disponibles incluso en los flujos de bits encriptados. Por tanto, o bien se pueden calcular los tamaños de los fotogramas con el número exacto de bytes, o se pueden estimar lo suficientemente bien para realizar los pasos siguientes.

20 En un segundo paso, el flujo de bits de vídeo es escaneado hasta que se encuentra el siguiente fotograma I, lo que representa el inicio del siguiente grupo GoP.

25 En flujos no encriptados, el siguiente fotograma I (como con todos los tipos de fotogramas) puede ser encontrado mirando en la cabecera del fotograma de vídeo. En flujos encriptados, o si los límites del fotograma aparecen en algún lugar en puntos desconocidos dentro de la información útil, en vez de hacer lo anteriormente indicado, se puede acceder a ciertas banderas que marcan los fotogramas I como "puntos de acceso aleatorio". Si por alguna razón esas banderas no aparecen, todavía se puede usar un análisis de los tamaños de los fotogramas encontrados para detectar fotogramas I, ya que su tamaño es en general mucho mayor que el de los fotogramas P y B.

30 En un tercer paso, los tamaños en bytes de todos los fotogramas de vídeo siguientes hasta el siguiente fotograma I son capturados en ciertos grupos hasta que es encontrado el siguiente fotograma I y por tanto se completa el GoP.

35 En un cuarto paso estos grupos son hechos corresponder con un número de pautas predefinidas que representan estructuras GoP típicas. La mejor correspondencia es asociada a una estructura GoP específica que en pasos posteriores sirve como prototipo para las siguientes GoPs y puede por tanto habilitar una fácil estimación a priori de los tipos de fotogramas de los fotogramas de entrada.

40 Para los siguientes GoPs, el grupo de tamaños de fotogramas puede ser reajustado, para que solamente se usen datos del GoP en curso, o pueden añadirse los tamaños de fotogramas de un cierto número de GoPs previos para obtener valores medios. En el último caso puede necesitarse un grupo adicional para que se haga notar el número de valores de tamaños de fotogramas añadidos a los elementos del grupo. De preferencia, las medias móviles pueden ser calculadas aplicando:

45 para (todos los fotogramas n dentro del GoP en curso)

$$\text{FrameSizeArray}[n] = (M - 1)/M * \text{FrameSizeArray}[n] + 1/M * \text{FrameSize}[n]$$

en donde M es el número de GoPs al que se le va a hacer la media.

50 Se pueden usar diferentes esquemas de ponderación

Posibles pautas de GoPs

55 Según se ha dicho anteriormente, un GoP tiene usualmente una estructura específica que se compone de una cierta secuencia de fotogramas P y B entre dos fotograma I que marcan los límites del GoP. Ejemplos de estructuras GoP típicas son:

- I, b, b, P, b, b, P, b, b, P...
 - I, b, B, b, P, b, B, b, P, b, B, b, P
- (la estructura GoP "clásica" conocida de la MPEG2)
(posible estructura GoP para codificación jerárquica, siendo B = referencia al fotograma B b = no hace referencia al fotograma B)

60 Debe tenerse en cuenta que con frecuencia se codifica el vídeo entrelazado como "secuencia de campo" y que comprende dos campos independientes por fotograma de vídeo. En este caso, las estructuras anteriores tendrían la forma siguiente:

- I/P, b/b, b/b, P/P, b/b, b/b, P/P...
- I/P, b/b, B/B, b/b P/P, b/b, B/B, b/b, P/P...

formando diferentes pautas que, si son correctamente detectadas, permiten tomar la decisión entre la codificación de fotograma y la de campo incluso en flujos encriptados. Se ha de tener en cuenta aquí que con frecuencia en las secuencias de campo sólo el primer campo de un fotograma clave es realmente codificado como campo intra. El segundo campo del fotograma clave puede ser un campo predicho. En la notación anterior esto daría como resultado un fotograma I/P. Incluso son posibles fotogramas P/I.

Las pautas anteriores son visualizadas en el "orden de presentación". De hecho, los fotogramas son transmitidos en un orden diferente, el llamado "orden del flujo de bits" u "orden de descodificación". Ya que los fotogramas B tienen referencias en el pasado y en el futuro, estas referencias futuras tienen que ser transmitidas por adelantado dando como resultado un orden del flujo de bits como se muestra a continuación:

	$I_0, P_1, b_2, b_3, P_4, b_5, b_6, P_7, b_8, b_9 \dots$	orden del flujo de bits
	$I_0, b_2, b_3, P_1, b_5, b_6, P_4, b_8, b_9, P_7 \dots$	orden de presentación (con índice de orden del flujo de bits)
	$I_0, P_1, B_2, b_3, b_4, P_5, B_6, b_7, b_8, \dots$	orden del flujo de bits
	$I_0, b_3, B_2, b_4, P_1, b_7, B_6, b_8, P_5, \dots$	orden de presentación (con índice de orden del flujo de bits)

Incluso son posibles diferentes pautas para la misma estructura GoP. Esto se debe a que existen técnicas de "transmisión del GoP cerrado" y de "transmisión del GoP abierto".

En GoPs cerrados, la transmisión del contenido de los diferentes GoPs no se superpone nunca. En estructuras GoP abiertas el primer fotograma I es transmitido antes de los fotogramas B precediéndolos en el orden de presentación (los que cierran el GoP precedente). Los ejemplos anteriores se han mostrado con el tipo de GoP cerrado. El mismo ejemplo con el tipo de GoP abierto tiene esta forma:

	$b_{-2}, b_{-1}, I_0, b_2, b_3, P_1, b_5, b_6, P_4, \dots$	orden de presentación (GoP abierto)
	$I_0, b_{-2}, b_{-1}, P_1, b_2, b_3, P_4, b_5, b_6, \dots$	orden del flujo de bits
	$b_{-3}, B_{-2}, b_{-1}, I_0, b_3, B_2, b_4, P_1, b_7, B_6, b_8, P_5 \dots$	orden de presentación (GoP abierto)
	$I_0, B_{-2}, b_{-3}, b_{-1}, P_1, B_2, b_3, b_4, P_5, B_6, b_7, b_8 \dots$	orden del flujo de bits

con los primeros dos (tres) fotogramas B (marcados aquí con letras en negrita) que pertenecen al GoP precedente, que son transmitidos después del fotograma I del GoP en curso.

Resulta fácil de entender que el orden del flujo de bits y por tanto la siguiente pauta GoP resultante difieren de una manera clara que, si es correctamente detectada, permite tomar la decisión entre las secuencias del GoP cerrado y del GoP abierto.

Otras pautas posibles incluyen:

	$I, P, P, P, P \dots$	(estructura GoP sin fotogramas B como las usadas en conferencias de vídeo y otras aplicaciones de bajo retraso)
	I, b, P, b, P, \dots	(pauta del GoP sólo con un fotograma B)

Está dentro del alcance de este método estimar la estructura GoP de cualquier flujo de vídeo analizado, ya sea totalmente legible, encriptado o corrompido por la pérdida de paquetes. La única información necesaria para conseguir este objetivo es los tamaños en bytes de cada fotograma de vídeo encontrado y el conocimiento de qué fotogramas son del tipo del fotograma I.

Cálculo/estimación de los tamaños de los fotogramas

Como este método se basa en el análisis de los tamaños de los bytes de todos los fotogramas de una secuencia de vídeo, estos tamaños tienen que ser extraídos del flujo de bits para cada fotograma. Existen muchos esquemas de transporte diferentes, que pueden requerir acciones diferentes para realizar esta tarea. Con mucho, el esquema de transporte más importante por redes basadas en el IP es sin embargo el "protocolo de tiempo real" (RTP). Por tanto, la extracción del tamaño del fotograma es descrita aquí solamente para el RTP, pero funciona de manera similar en otros esquemas de transporte. Las personas expertas en la materia pueden adoptar fácilmente el concepto para cualquier esquema de transporte diferente que tenga este uso.

Existen dos grandes enfoques para transmitir el contenido de vídeo con el RTP:

- El vídeo es llevado directamente como una información útil de los paquetes de RTP. En este caso el audio y otra información posible (información del "sistema") es llevado en flujos de RTP diferentes que tienen diferentes números de puerto. Por tanto, la porción de vídeo del flujo de bits puede fácilmente ser separada del resto de los paquetes.

Un fotograma de vídeo consiste en general en la información útil de un número de paquetes RTP, cada uno de ellos lleva una indicación de tiempo. En el caso del vídeo, estas indicaciones de tiempo de RTP son hechas corresponder con el tiempo de descodificación relativo del fotograma de vídeo que llevan consigo. Ya que cada uno de los paquetes RTP puede tener un tamaño diferente, un fotograma de vídeo está en general en un número de paquetes RTP completos. Para calcular en este caso el tamaño de un fotograma de vídeo, sólo tienen que ser añadidos los tamaños de la información útil de todos los paquetes RTP con la misma indicación de tiempo.

Incluso si aparecen límites de fotogramas dentro de una única información útil RTP, se pueden estimar bastante bien los tamaños de los fotogramas por el valor antedicho. No es necesario el acceso a la información útil en sí misma, que podría estar encriptada.

Se puede acceder al número de secuencia RTP, que forma parte de la cabecera RTP, para detectar pérdidas de paquetes RTP. Ya que no se puede conocer el tamaño de un paquete perdido, es necesario que sea estimado. Esto se puede conseguir haciendo algún tipo de media del tamaño de los paquetes anteriores.

- El vídeo es transportado conforme al RTP, pero la información útil es un “flujo de transporte” (TS). En un TS, el vídeo, audio y otras informaciones se multiplexan en un flujo único. Cada uno de los paquetes multiplexados tiene un tamaño de 188 bytes incluyendo la pequeña cabecera del TS. Un paquete RTP transporta 7 de estos paquetes TS, que están marcados por el así llamado “Programa ID” (PID) como pertenecientes a un subflujo específico. Uno de estos subflujos es el vídeo a ser analizado. Incluso en los flujos encriptados, las cabeceras TS no están en general encriptadas, y se puede por tanto tener acceso a ellas. Estas cabeceras permiten realizar una medida exacta de los tamaños de los fotogramas. En el caso de pérdidas de paquetes RTP, se puede utilizar un llamado “contador de continuidad” dentro de la cabecera TS para calcular el número de paquetes del subflujo perdidos separadamente para cada subflujo con pérdida de RTP particular. Ya que los paquetes TS son mucho más pequeños que los paquetes RTP, y (en ausencia del raro caso del “stuffing” o relleno) la información útil tiene el mismo tamaño, el número de bytes perdidos puede ser estimado con mayor precisión. El mencionado contador de continuidad de la cabecera TS tiene solamente 4 bits. Por tanto, puede detectarse la falta de hasta 16 paquetes TS. Junto con el número conocido de los paquetes RTP que faltan y un valor para el número medio de paquetes del subflujo TS por paquete RTP que pueden ser fácilmente calculados, se puede hacer una estimación precisa de los paquetes TS perdidos cuando sean más de 16.

Ya que la indicación de tiempo RTP está sincronizada con el tiempo de descodificación relativo del fotograma de vídeo que éstos llevan consigo, se pueden detectar incluso fotogramas perdidos completos, porque en tal caso la indicación de tiempo RTP aumenta en más de uno la diferencia de tiempos entre fotogramas sucesivos.

Detección de los fotogramas I

Para iniciar el análisis, se escanea el flujo de bits para encontrar el comienzo del siguiente fotograma I, que marca el comienzo de un nuevo GoP. Los fotogramas I pueden ser detectados mediante diferentes medios. En el caso de un flujo no encriptado, se puede tener un fácil acceso a las cabeceras de los fotogramas del flujo elemental para todos los fotogramas de vídeo. Las cabeceras de los fotogramas de vídeo contienen explícitamente el tipo del fotograma.

Si el vídeo es transportado directamente como información útil RTP, siempre deben aparecer cabeceras de fotogramas en el primer byte de la información útil de cada paquete RTP que sigue después de un paquete RTP con el bit M de la cabecera RTP puesto a 1. Si las cabeceras de los indicadores aparecen también dentro de la información útil de un paquete RTP, éstas se pueden encontrar fácilmente, porque contienen una secuencia de bytes única.

En el caso de que el TS se use siguiendo el RTP, se puede explotar la llamada “bandera de acceso aleatorio” dentro del campo de adaptación de la cabecera TS. Esta bandera señala que el fotograma que sigue fue codificado como fotograma I. Incluso en flujos encriptados el campo de adaptación como parte de la cabecera TS no está usualmente encriptado.

En cuanto se conocen los límites del fotograma, los fotogramas I pueden también ser detectados empíricamente. Su tamaño medio es usualmente mucho mayor que el de los fotogramas P y B.

En cuanto se encuentra el comienzo de un fotograma I, se encuentra también el comienzo del siguiente GoP. A partir de ese punto, los tamaños de todos los fotogramas siguientes son recogidos de varias maneras posibles. En la descripción siguiente este proceso de recogida es continuado durante un GoP completo hasta que se inicia el análisis de los datos reunidos. En otras realizaciones, el período de recogida puede tener una longitud diferente.

Realización de la recogida de datos y de la correspondencia de pautas

Un grupo largo de tamaños de fotogramas

Después de haber sido encontrado el fotograma I inicial del GoP en curso, se guardan los tamaños de los fotogramas de todos los fotogramas siguientes en los elementos de un grupo, siendo el índice del grupo igual al

número de fotogramas que han pasado tras dicho fotograma I. De preferencia, se hace esto hasta que se detecta el siguiente fotograma I, que señala el final del primer GoP a ser analizado. El grupo de tamaños de fotogramas así construido es la entrada a un proceso de correspondencia de pautas que compara este grupo con un conjunto de pautas típicas, y produce una pauta estimada del GoP y otros datos cuando sea aplicable.

Si el flujo de bits analizado no está encriptado, y por tanto los tipos de fotogramas son conocidos por el algoritmo, los tamaños reales de los fotogramas pueden ser reemplazados por valores estandarizados como '1' para fotogramas B no referenciados, '2' para fotogramas B referenciados y '4' para fotogramas P, que tienen el mismo cometido que los tamaños reales de los fotogramas, pero que no introducen ninguna incertidumbre estadística.

Correspondencia de pautas

En una realización, el grupo de tamaños de fotogramas es hecho corresponder con pautas binarias mediante un '0' para los fotogramas B y un '1' para los fotogramas P. El ejemplo anterior de las estructuras GoP es hecho corresponder por tanto con pautas como:

0, 0, 1, 0, 0, 1, ...	(l, b, b, P, b, b, P estructura con el GoP abierto)
1, 0, 0, 1, 0, 0, ...	(l, b, b, P, b, b, P estructura con el GoP cerrado)
0, 0, 0, 1, 0, 0, 0, 1, ...	(l, b, B, b, P, b, B, b, P estructura con el GoP abierto)
1, 0, 0, 0, 1, 0, 0, 0, ...	(l, b, B, b, P, b, B, b, P estructura con el GoP cerrado)

En esta realización, el "grupo de tamaños de fotogramas" es por consiguiente convertido en un grupo binario también. En el caso más sencillo, se usa un valor de referencia único para reemplazar los valores de cada elemento de dicho grupo por un '0' para "fotograma(s) de tamaño(s) pequeño(s)" y por un '1' para "grandes fotograma(s)". El valor de referencia para decidir entre "grande" y "pequeño" puede ser simplemente la media de todos los elementos multiplicados por un factor > 1. En realizaciones alternativas el umbral puede ser el valor medio de todos los elementos del grupo o una combinación ponderada de los elementos más pequeños y de los más grandes u otras combinaciones de los elementos del grupo. El umbral puede ser adoptado también para un número pequeño de valores de los tamaños de los fotogramas. Esto puede ser realizado variando uno de los valores de umbral descritos por un factor o un término aditivo o calculando el umbral solamente a partir de valores del tamaño del fotograma en una cierta proximidad al valor en curso.

En general, puede ser razonable no usar unos pocos de los primeros valores del grupo porque inmediatamente después de un fotograma I los fotogramas P y B tienen con frecuencia tamaños pequeños irregulares.

Proceso de correspondencia

En general, no se conoce anticipadamente la longitud de un GoP. Valores típicos son un fotograma I por cada 1 o 2 segundos, dando lugar a una longitud del GoP de, por ejemplo, 25 a 50 fotogramas (o 50 hasta 100 campos en los modos de "codificación de campo") para la velocidad típica de los fotogramas de 25 fotogramas/s. Usualmente, el codificador tiene una detección de escena de corte, y coloca un fotograma I en cada posición de las escenas de corte. Por tanto, si se encuentran escenas de corte, los GoPs pueden tener cualquier tamaño hasta 2 e incluso 1 fotogramas.

En contraste con la longitud global, las pautas básicas de cada una de las estructuras GoP posibles son constantes y cortas (por ejemplo, 0, 0, 1 para el ejemplo l, b, b, P, ...). Con el objeto de hacer que se correspondan con los elementos de un GoP completo, estas pautas básicas simples tienen que ser repetidas hasta que se alcance el tamaño real del GoP.

En las realizaciones aquí explicadas sólo se aplica la correspondencia binaria, el proceso de correspondencia en sí mismo puede ser una simple operación de "nor exclusivo" sobre cada elemento del grupo convertido de tamaños de fotogramas con los elementos de pauta básica concatenados. Los resultados de estas operaciones de "nor exclusivo" son añadidos a continuación a una puntuación correspondiente de la pauta usada. Esto se hace con todas las pautas predefinidas disponibles.

La pauta específica, que consigue el número más alto de puntuación, es considerada la mejor correspondencia y su estructura GoP, estado de fotograma/campo, estado abierto/cerrado asociados son adoptados entonces para realizar análisis posteriores.

Anexo 2

Método para la estimación de la clase del contenido basado en los tamaños de los fotogramas y en los tipos de los fotogramas.

(Tomado de los párrafos [0012] a [0022] del documento WO 2009/012297 A1)

A partir de ahora se describe un método para la estimación del contenido de fotogramas de un flujo de vídeo de paquetes encriptado sin que los paquetes sean desencriptados. El concepto puede ser usado también como un

método alternativo para estimar el contenido de flujos de vídeo de paquetes sin encriptar.

5 El método opera examinando en primer lugar las cabeceras de los paquetes para determinar dónde empieza y acaba un fotograma dado. (Debido a que los grandes fotogramas están divididos en múltiples paquetes, no se puede suponer que cada fotograma está contenido en un paquete único). Un analizador de red puede realizar esta tarea, por ejemplo, examinando los campos de la indicación de tiempo de las cabeceras RTP (Protocolo de Transporte en Tiempo Real) de los paquetes del flujo de vídeo. Los paquetes con indicaciones de tiempo idénticas comprenden un fotograma único.

10 A continuación, el analizador clasifica el fotograma observado dentro del intervalo como un fotograma I, P o B. En flujos de vídeo no encriptados, el analizador puede leer simplemente el tipo de fotograma de la cabecera de la imagen. En los flujos de vídeo encriptados, el analizador puede leer el tipo de fotograma directamente si éste tiene la clave de desencriptación apropiada. En flujos de vídeo encriptado, el analizador puede leer directamente el tipo del fotograma. Alternativamente, el analizador puede estimar el tipo de fotograma basándose en el tamaño (en bytes) del fotograma. Según se ha descrito anteriormente, los fotogramas I, P, y B son respectivamente grandes, medios y pequeños uno con relación a otro.

20 En realizaciones en donde el analizador estima el tipo de fotograma basándose en el tamaño de los fotogramas, el analizador de la red empieza por contar el número de bytes del fotograma. Esto se puede hacer determinando el tamaño de la información útil de los datos de cada paquete y sumando este valor a todos los paquetes que comprenden el fotograma. El analizador estima también el tamaño de cualquier paquete que fuera descartado e incluye esta estimación en su cuenta de bytes global del fotograma.

25 El analizador puede detectar un paquete descartado examinando los números de secuencia del paquete, por ejemplo, el número de secuencia del RTP, para encontrar cualquier lapso en la secuencia. (Como los paquetes pueden viajar siguiendo un orden no secuencial, el analizador tiene que mantener una lista de números de secuencia recientemente observados. El analizador puede a continuación clasificar un número de secuencia que falta como un paquete descartado basándose en que no ha sido visto después de un tiempo suficientemente amplio). El analizador puede estimar el tamaño de un paquete descartado basándose en el tamaño medio de los paquetes. Dicha media puede ser calculada, por ejemplo, como una media ponderada de los últimos n paquetes, como una media de todos los paquetes recibidos o como una media de los paquetes dentro del GoP dado. Además, la media puede ser limitada a paquetes que contienen datos del mismo tipo de fotograma (I, P o B). El analizador de la red puede calcular además la velocidad de pérdida de paquetes media determinando la proporción de paquetes del flujo de vídeo que han sido descartados.

35 El analizador de la red mantiene uno o dos valores “umbrales del fotograma” para que éste pueda estimar la variación de tamaños de los fotogramas I, P y B. Debido a que los fotogramas I son mucho más grandes que los fotogramas P y los B, estos fotogramas con una cuenta de bytes por encima de un cierto umbral (“umbral del fotograma I”) son clasificados como fotogramas I. De manera similar, algunas realizaciones utilizan un umbral (más bajo) separado (“umbral del fotograma P”) para discriminar entre los fotogramas P y B porque los primeros son usual y significativamente más grandes que los últimos. Los umbrales de los fotogramas I y P pueden ser actualizados continuamente basándose en la variación de los tamaños de los fotogramas del flujo de vídeo.

45 El analizador de la red puede mantener también valores para los tamaños máximo, mínimo y medio de los fotogramas I, P, y B respectivamente. Estos valores pueden ser actualizados continuamente. Para cada fotograma encontrado, el analizador de la red puede comparar su tamaño con el tamaño del fotograma medio de su tipo y calcular la varianza del fotograma dado. El analizador de la red puede mantener además valores medios de las varianzas de los fotogramas.

50 Debido a que cada fotograma I inicia un nuevo GoP, el analizador de la red puede calcular además la longitud de cada GoP que observa. El analizador puede mantener también los valores máximo, mínimo, medio y de la varianza de la longitud del GoP.

55 Para cada GoP, el analizador puede estimar la cantidad de detalle, movimiento o “panning”, (o paneo: seguimiento de la imagen con la cámara) presentes en el GoP. De manera similar, el analizador puede estimar si se ha producido un cambio de escena desde el GoP anterior hasta el GoP en curso. El analizador puede conseguir esto comparando los tamaños de los fotogramas y la longitud del GoP con los diversos máximos, mínimos y medias descritos anteriormente.

60 Por ejemplo, un fotograma I que sea relativamente grande en comparación con otros fotogramas I de la secuencia de vídeo indica un alto nivel de detalle del GoP del que forma parte. Si un GoP contiene fotogramas P o B que son relativamente grandes, entonces el GoP está mostrando una gran cantidad de movimiento. Se puede detectar el paneo en una secuencia de vídeo por una transmisión relativamente más frecuente de fotogramas I. Se indica un cambio de escena cuando un fotograma I es enviado en medio de un GoP normal y reinicia efectivamente el GoP.

65 Después de calcular estas diversas métricas, el analizador de la red puede usarlas como entradas para un algoritmo

de estimación de la calidad del vídeo. Dicho algoritmo puede calcular la relación de la señal de cresta estimada al ruido (PSNR) y dar una "mean opinion score" (MOS), o puntuación subjetiva media.

REIVINDICACIONES

- 5 1. Método para evaluar la calidad de una secuencia de vídeo sincronizando temporalmente el flujo de bits de entrada encriptado de un descodificador de vídeo con la secuencia de vídeo procesada (PVS) obtenida por un descodificador de vídeo exterior, comprendiendo el método los pasos de:
- 10 a) suministrar el flujo de bits de entrada encriptado a un módulo de análisis;
 b) analizar, en el módulo de análisis, la información de la cabecera del paquete del flujo de bits de entrada encriptado, y calcular el tamaño y el tipo de las imágenes contenidas en el flujo de bits;
 15 c) extraer, basándose en los tamaños y tipos de la imagen calculados, características que son representativas del contenido presentado en las imágenes de vídeo;
 d) extraer el tipo del contenido y las características respectivas representativas del contenido de las imágenes de la secuencia de vídeo procesada;
 e) comparar las características extraídas en el paso d) de la imagen en curso de la secuencia de vídeo procesada con las características extraídas en el paso c) de las imágenes analizadas para calcular la parte del flujo de bits que se corresponde con la PVS en curso;
 20 f) en donde la parte del flujo de bits que se corresponde con la PVS en curso es introducida adicionalmente en un módulo de evaluación de la calidad del vídeo; y
 g) estimar, mediante el módulo de evaluación de la calidad del vídeo, una puntuación de la calidad subjetiva basada en la información obtenida del flujo de bits y de la secuencia de vídeo procesada.
- 25 2. El método de la reivindicación 1, en el que las características extraídas en los pasos c) y d) son características que varían con el tiempo.
- 30 3. El método de la reivindicación 1, en donde el paso b) detecta además fotogramas que han sido afectados por la pérdida de paquetes, y comprende además los pasos de:
- 35 h) examinar la imagen en curso de la secuencia de vídeo generada por el descodificador de vídeo exterior para artefactos generados debido a las pérdidas de paquetes; y
 i) hacer que la secuencia de vídeo procesada se corresponda con el fotograma correspondiente del flujo de bits, en donde, si se detectó que la imagen en curso de la secuencia de vídeo procesada en el paso f) contenía artefactos que fueron insertados a consecuencia de una pérdida de paquete, entonces esa imagen es hecha corresponder con la parte del flujo de bits que se corresponde con el fotograma que se encontró en el paso b) que contenía la pérdida de paquete.
- 40 4. Aparato para evaluar la calidad de una secuencia de vídeo sincronizando temporalmente el flujo de bits de entrada encriptado de un descodificador de vídeo con la secuencia de vídeo procesada (PVS) descodificada por un descodificador de vídeo exterior, comprendiendo el aparato:
- 45 un analizador de flujo de bits que recibe el flujo de bits de entrada encriptado, estando configurado el analizador del flujo de bits para examinar y analizar el flujo de bits de entrada, y
 un módulo de sincronización que está configurado para sincronizar el flujo de bits analizado con la secuencia de vídeo procesada, y
 un módulo de evaluación de la calidad del vídeo,
caracterizado por que
 el módulo de sincronización comprende:
- 50 un primer módulo de extracción que está configurado para extraer, basándose en los tamaños y tipos de fotogramas calculados, características que son representativas del contenido presentado en los fotogramas de vídeo,
 un segundo módulo de extracción que está configurado para extraer el tipo del contenido y características respectivas representativas del contenido de las imágenes de la secuencia de vídeo procesada, y
 55 una unidad de comparación conectada a la primera unidad de extracción y a la segunda unidad de extracción, estando configurada la unidad de comparación para
- 60 comparar las características extraídas de la imagen en curso de la secuencia de vídeo procesada con las características extraídas de los fotogramas analizados, y para determinar el fotograma correspondiente n*, en donde la unidad de comparación está configurada para introducir la parte del flujo de bits que se corresponde con la PVS en curso en un módulo de evaluación de la calidad del vídeo; y
 en donde el módulo de evaluación de la calidad del vídeo está configurado para estimar una puntuación de la calidad subjetiva basada en la información obtenida del flujo de bits y de la secuencia de vídeo procesada.

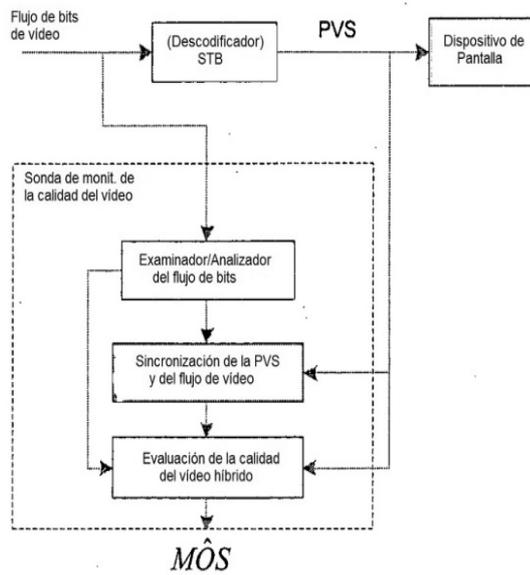


Fig. 1

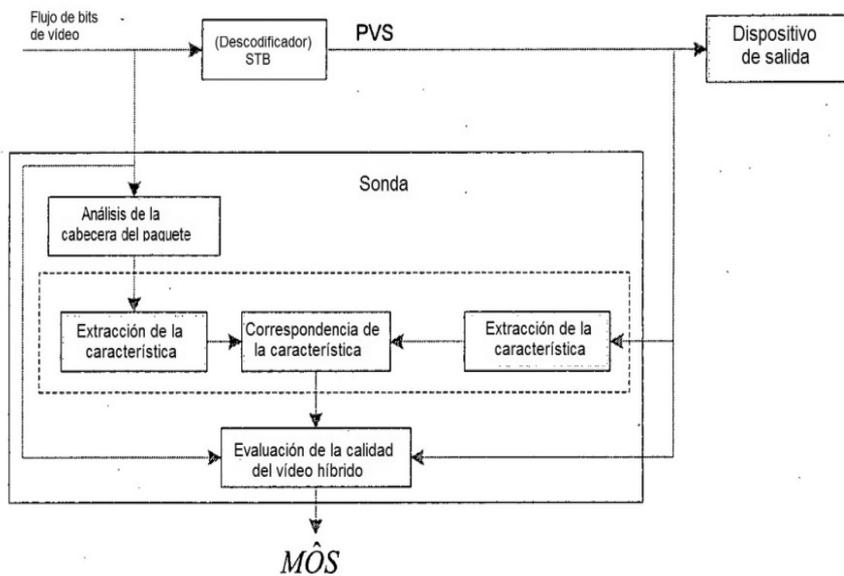


Fig. 2