

UIT-R

Sector de Radiocomunicaciones de la UIT

Recomendación UIT-R BT.1885
(03/2011)

**Técnicas de medición objetiva de la calidad
de vídeo perceptual para la radiodifusión
de televisión digital de definición
convencional en presencia de
una referencia de anchura
de banda reducida**

Serie BT
Servicio de radiodifusión (televisión)



Prólogo

El Sector de Radiocomunicaciones tiene como cometido garantizar la utilización racional, equitativa, eficaz y económica del espectro de frecuencias radioeléctricas por todos los servicios de radiocomunicaciones, incluidos los servicios por satélite, y realizar, sin limitación de gamas de frecuencias, estudios que sirvan de base para la adopción de las Recomendaciones UIT-R.

Las Conferencias Mundiales y Regionales de Radiocomunicaciones y las Asambleas de Radiocomunicaciones, con la colaboración de las Comisiones de Estudio, cumplen las funciones reglamentarias y políticas del Sector de Radiocomunicaciones.

Política sobre Derechos de Propiedad Intelectual (IPR)

La política del UIT-R sobre Derechos de Propiedad Intelectual se describe en la Política Común de Patentes UIT-T/UIT-R/ISO/CEI a la que se hace referencia en el Anexo 1 a la Resolución UIT-R 1. Los formularios que deben utilizarse en la declaración sobre patentes y utilización de patentes por los titulares de las mismas figuran en la dirección web <http://www.itu.int/ITU-R/go/patents/es>, donde también aparecen las Directrices para la implementación de la Política Común de Patentes UIT-T/UIT-R/ISO/CEI y la base de datos sobre información de patentes del UIT-R sobre este asunto.

Series de las Recomendaciones UIT-R

(También disponible en línea en <http://www.itu.int/publ/R-REC/es>)

| Series | Título |
|------------|------------------------------------------------------------------------------------------------------------------|
| BO | Distribución por satélite |
| BR | Registro para producción, archivo y reproducción; películas en televisión |
| BS | Servicio de radiodifusión sonora |
| BT | Servicio de radiodifusión (televisión) |
| F | Servicio fijo |
| M | Servicios móviles, de radiodeterminación, de aficionados y otros servicios por satélite conexos |
| P | Propagación de las ondas radioeléctricas |
| RA | Radio astronomía |
| RS | Sistemas de detección a distancia |
| S | Servicio fijo por satélite |
| SA | Aplicaciones espaciales y meteorología |
| SF | Compartición de frecuencias y coordinación entre los sistemas del servicio fijo por satélite y del servicio fijo |
| SM | Gestión del espectro |
| SNG | Periodismo electrónico por satélite |
| TF | Emisiones de frecuencias patrón y señales horarias |
| V | Vocabulario y cuestiones afines |

Nota: Esta Recomendación UIT-R fue aprobada en inglés conforme al procedimiento detallado en la Resolución UIT-R 1.

Publicación electrónica
Ginebra, 2011

© UIT 2011

Reservados todos los derechos. Ninguna parte de esta publicación puede reproducirse por ningún procedimiento sin previa autorización escrita por parte de la UIT.

RECOMENDACIÓN UIT-R BT.1885

Técnicas de medición objetiva de la calidad de vídeo perceptual para la radiodifusión de televisión digital de definición convencional en presencia de una referencia de anchura de banda reducida

(2011)

Cometido

Esta Recomendación describe los métodos de evaluación objetiva de la calidad de vídeo para la radiodifusión de televisión digital de definición convencional que pueden medir la calidad de vídeo perceptual en condiciones de recepción móvil y fija cuando las características extraídas de la señal de vídeo de referencia pueden obtenerse fácilmente en el punto de medición.

La Asamblea de Radiocomunicaciones de la UIT,

considerando

- a) que es conveniente tener la capacidad de medir automáticamente las degradaciones de las señales de radiodifusión de vídeo;
- b) que la calidad de vídeo perceptual de la recepción móvil puede cambiar dinámicamente dependiendo de las propias condiciones de recepción;
- c) que la medición objetiva de la calidad de vídeo perceptual puede ser útil para complementar los métodos de evaluación subjetiva;
- d) que se han propuesto al UIT-R tres técnicas de medición de la calidad de vídeo para la radiodifusión de televisión de definición convencional en presencia de una referencia reducida y se ha demostrado que dichas técnicas proporcionan resultados equivalentes y coherentes;
- e) que las técnicas de medición objetiva de la calidad de vídeo perceptual para la radiodifusión de televisión digital de definición convencional en presencia de una referencia completa se han especificado en la Recomendación UIT-R BT.1683,

recomienda

1 que se utilice cualquiera de los modelos objetivos de calidad de vídeo indicados en el Anexo 1 para realizar mediciones objetivas de la calidad de vídeo perceptual de la radiodifusión de televisión digital de definición convencional en presencia de una referencia de anchura de banda reducida.

1 Introducción

Esta prueba de RRNR-TV aborda las imágenes definidas en la Recomendación UIT-R BT.601-6 y dos tipos de miedos: referencia reducida (RR) y sin referencia (NR). Los modelos RR tienen un acceso a anchura de banda limitado al vídeo fuente y los modelos NR no tienen acceso al vídeo fuente.

Los HCR en cada experimento engloban la codificación únicamente con efectos auxiliares y la codificación con errores de transmisión. Los esquemas de codificación examinados fueron el MPEG-2 y el H.264 (MPEG-4 Parte 10). Los codificadores MPEG-2 funcionaron con velocidades binarias comprendidas entre 1,0 y 5,5 Mbits/s y los codificadores H.264 con velocidades binarias

entre 1,0 y 3,98 Mbits/s. Cada experimento incluyó 12 secuencias de fuente de las cuales dos eran una fuente secreta. Cada experimento incluía 34 HCR y 156 secuencias de vídeo procesadas (PVS). De estas PVS, 40 contenían errores de transmisión y 116 contenían únicamente codificación.

1.1 Aplicación

La presente Recomendación proporciona unas estimaciones de la calidad de vídeo para las clases de vídeo TV3 a MM5B, como se definen en la Recomendación UIT-T P.911, Anexo B. Las aplicaciones para los modelos de estimación descritos en esta Recomendación incluyen, entre otras, las siguientes:

- 1) supervisión en tiempo real de la calidad potencial del servicio en la fuente;
- 2) supervisión a distancia de la calidad de destino cuando se dispone de canales laterales para las características extraídas de las secuencias del vídeo fuente;
- 3) medición de la calidad para supervisar un sistema de almacenamiento o de transmisión que utiliza técnicas de compresión y decompresión del vídeo, ya sea una sola vez o mediante concatenación de tales técnicas;
- 4) pruebas en laboratorio de los sistemas de vídeo.

1.2 Limitaciones

Los modelos de estimación descritos en esta Recomendación no pueden emplearse para sustituir completamente a las pruebas subjetivas. Los valores de correlación entre dos pruebas subjetivas cuidadosamente diseñadas y ejecutadas (es decir, en dos laboratorios distintos) normalmente caen en la gama de 0,95 a 0,98. Si se utiliza esta Recomendación para comparar los distintos códecs, conviene emplear un método cuantitativo (tal como el descrito en la Recomendación UIT-T J.149) para determinar la precisión del modelo en ese contexto particular.

Los modelos de esta Recomendación fueron validados midiendo la calidad de vídeo objetiva que presenta congelaciones de trama de hasta 2 s.

Los modelos de esta Recomendación no se validaron para mediciones objetivas de la calidad de vídeo con un retardo constantemente creciente (por ejemplo, vídeos que no descartan tramas perdidas tras una congelación de trama).

Cabe señalar que en el caso de nuevas tecnologías de codificación y transmisión que producen efectos auxiliares no incluidos en esta evaluación, los modelos de evaluación objetiva pueden dar lugar a resultados erróneos. En ese caso es necesario efectuar una evaluación subjetiva.

2 Referencias

Las siguientes Recomendaciones del UIT y otras referencias contienen disposiciones que, mediante su referencia en este texto, constituyen disposiciones de la presente Recomendación. Al efectuar esta publicación, estaban en vigor las ediciones indicadas. Todas las Recomendaciones y otras referencias son objeto de revisiones por lo que se preconiza que los usuarios de esta Recomendación investiguen la posibilidad de aplicar las ediciones más recientes de las Recomendaciones y otras referencias citadas a continuación. Se publica periódicamente una lista de las Recomendaciones UIT-T actualmente vigentes.

En esta Recomendación, la referencia a un documento, en tanto que autónomo, no le otorga el rango de una Recomendación.

2.1 Referencias normativas

- Recomendación UIT-R BT.500-12 – Metodología para la evaluación subjetiva de la calidad de las imágenes de televisión.
- Recomendación UIT-T P.910 (2008) – Métodos de evaluación subjetiva de la calidad de vídeo para aplicaciones multimedia.
- Recomendación UIT-T P.911 (1998) – Métodos de evaluación subjetiva de la calidad audiovisual para aplicaciones multimedia.
- Recomendación UIT-T J.143 (2000) – Requisitos de usuario para mediciones objetivas de la percepción de la calidad vídeo en televisión digital por cable.
- Recomendación UIT-T J.244 (2008) – Métodos para calibrar la desalineación constante de los dominios espacial y temporal con ganancia y desplazamientos constantes.

2.2 Referencias informativas

- Recomendación UIT-T J.149 (1998) – Método para especificar la precisión y la calibración cruzada de métricas de la calidad de vídeo.
- Recomendación UIT-T J.144 (2001) – Técnicas de medición objetiva de la percepción de la calidad vídeo en televisión por cable en presencia de una referencia completa.
- Recomendación UIT-T P.931 (1998) – Retardo de las comunicaciones multimedia, sincronización y medición de la velocidad de tramas.
- Recomendación UIT-T J.148 (2003) – Requisitos para un modelo objetivo de calidad perceptual de los multimedia.
- Recomendación UIT-T H.261 (1993) – Códec vídeo para servicios audiovisuales a p x 64 kbit/s.
- Recomendación UIT-T H.263 (1996) – Codificación de vídeo para comunicación a baja velocidad binaria.
- Recomendación UIT-T H.263 (1998) – Codificación de vídeo para comunicación a baja velocidad binaria (H.263+).
- Recomendación UIT-T H.264 (2003) – Codificación de vídeo avanzada para los servicios audiovisuales genéricos.
- VGEG – Validación de los modelos objetivos con referencia reducida y sin referencia para la televisión de definición convencional Fase I, 2009.

3 Definiciones

3.1 Términos definidos en otros textos:

Esta Recomendación utiliza los siguientes términos definidos en otros textos:

- 3.1.1 evaluación subjetiva (imagen)** (Recomendación UIT-T J.144): definición citada opcionalmente.
- 3.1.2 medición perceptual objetiva (imagen)** (Recomendación UIT-T J.144): definición citada opcionalmente.
- 3.1.3 proponente** (Recomendación UIT-T J.144): definición citada opcionalmente.

3.2 Términos definidos en esta Recomendación

Esta Recomendación define los siguientes términos:

3.2.1 Repetición de trama anómala: se define como un evento en el que el HCR produce una sola trama repetidamente en respuesta a un evento inusual o extraordinario. La repetición de trama anómala incluye, entre otros, los siguientes tipos de eventos: un error en el canal de transmisión, un cambio en el retardo a través del canal de transmisión, recursos informáticos limitados que afectan al comportamiento del decodificador y recursos informáticos limitados que repercuten en la presentación de la señal de vídeo.

3.2.2 Salto de trama constante: se define como un evento en el que el HCR produce tramas con contenido actualizado a una velocidad de trama efectiva fija e inferior a la velocidad de trama de la fuente.

3.2.3 Velocidad de trama efectiva: se define como el número de tramas únicas (es decir, total de tramas – tramas repetidas) por segundo.

3.2.4 Velocidad de trama: se define como el número de tramas únicas (es decir, total de tramas – tramas repetidas) por segundo.

3.2.5 Velocidad de trama prevista: se define como el número de tramas de vídeo por segundo físicamente almacenadas para alguna representación de una secuencia de vídeo. La velocidad de tramas deberá ser constante. Dos ejemplos de velocidades de trama previstas constantes son una cinta BetacamSP® que contiene 25 tramas/s y un fichero VQEG FR-TV Fase I de 625 líneas YUV que contiene 25 tramas/s; ambos tienen una velocidad de trama prevista de 25 tramas/s.

3.2.6 Condiciones de red en servicio: se definen como los errores impuestos en el tren de bits del vídeo digital como resultado de las condiciones de red en servicio.

3.2.7 Pausa con salto: se define como los eventos en los que el vídeo se detiene durante un cierto periodo de tiempo y a continuación arranca de nuevo produciéndose una cierta pérdida de información de vídeo. En la pausa con salto, el retardo temporal a través del sistema variará en torno a un retardo del sistema medio, a veces aumentando y a veces disminuyendo. Un ejemplo de pausa con salto es un par de videófonos IP, donde un intenso tráfico de red provoca que la pantalla del videófono IP se congele brevemente; cuando vuelve a funcionar dicha pantalla se ha perdido cierto contenido. El salto de trama constante y el salto de trama variable son subconjuntos de pausa con salto. Una secuencia de vídeo procesada que contenga pausa con salto tendrá aproximadamente la misma duración que la secuencia de vídeo original asociada.

3.2.8 Pausa sin salto: se define como cualquier evento en que el vídeo se detiene durante un cierto periodo de tiempo y a continuación arranca de nuevo sin que se haya perdido ninguna información de vídeo. Por consiguiente, el retardo temporal a través del sistema debe aumentar.

3.2.9 Velocidad de renovación: se define como la velocidad a la que se actualiza la pantalla.

3.2.10 Errores de transmisión simulados: se definen como los errores impuestos en el tren de bits del vídeo digital en un entorno altamente controlado. Entre los ejemplos pueden citarse las tasas de pérdidas de paquetes simuladas y los errores en los bits simulados.

3.2.11 Velocidad de trama de fuente (SFR): es la velocidad de trama prevista de las secuencias del vídeo de fuente original. La velocidad de trama de fuente es constante. Para la prueba VQEG RRNR-TV la SFR fue de 25 tramas/s o 30 tramas/s.

3.2.12 Errores de transmisión: se definen como los errores impuestos en la transmisión del vídeo. Como ejemplos de estos tipos de errores pueden citarse los errores de transmisión simulados y las condiciones de red en servicio.

3.2.13 Salto de trama variable: se define como un evento en el que el HCR produce tramas con contenido actualizado a una velocidad de trama efectiva que varía con el tiempo. El retardo temporal a través del sistema aumentará y disminuirá con el tiempo, variando en torno al retardo del sistema medio. Una secuencia de vídeo procesada que contiene un salto de trama variable tendrá aproximadamente la misma duración que la secuencia de vídeo original asociada.

4 Abreviaturas y acrónimos

Esta Recomendación utiliza las siguientes abreviaturas y acrónimos

| | |
|--------|-------------------------------------------------------------------------------------------------------------------------------------------------------|
| ACR | Índice por categoría absoluta (véase la Recomendación UIT-T P.910) (<i>absolute category rating</i>) |
| ACR-HR | Índice por categoría absoluta con referencia escondida (<i>absolute category rating with hidden reference</i>) (véase la Recomendación UIT-T P.910) |
| AVI | Entrelazado de audio vídeo (<i>audio video interleave</i>) |
| DMOS | Diferencia de la nota media de opinión de diferencias (<i>difference mean opinion score</i>) |
| FR | Referencia completa (<i>full reference</i>) |
| FRTV | Televisión de referencia completa (<i>full reference television</i>) |
| HRC | Circuito ficticio de referencia (<i>hypothetical reference circuit</i>) |
| NR | Sin referencia (o referencia cero) (<i>no (or zero) reference</i>) |
| PSNR | Relación señal de cresta/ruido (<i>peak signal-to-noise ratio</i>) |
| PVS | Secuencia de vídeo procesada (<i>processed video sequence</i>) |
| RMSE | Error cuadrático medio (<i>root mean square error</i>) |
| RR | Referencia reducida |
| SFR | Velocidad de trama de fuente (<i>source frame rate</i>) |
| SRC | Canal o circuito de referencia de fuente |
| VQEG | Grupo de Expertos en calidad del vídeo (<i>video quality Experts Group</i>) |
| YUV | Espacio cromático (<i>colour space</i>) |

5 Convenios

Ninguno.

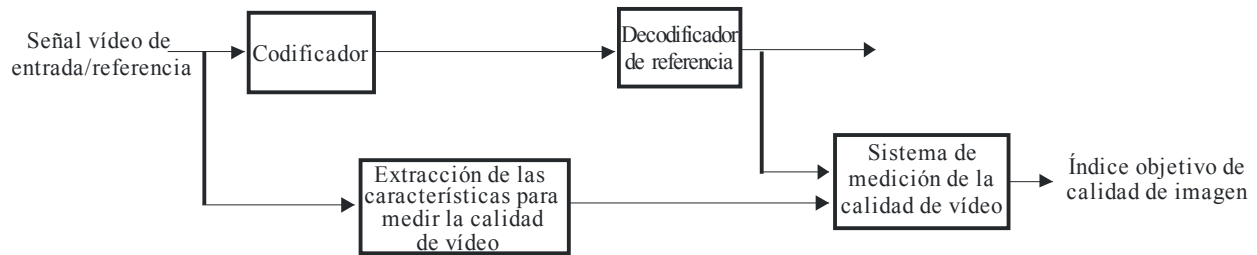
6 Descripción del método de medición con referencia reducida

El método de medición de doble extremo con referencia reducida para la medición objetiva de la calidad de vídeo perceptual, evalúa el comportamiento de los sistemas realizando una comparación entre las características extraídas a partir de la señal de vídeo de entrada, o referencia, sin distorsionar a la entrada del sistema y la señal degradada a la salida del sistema (véase la Fig. 1).

La Fig. 1 muestra un ejemplo de aplicación del método de referencia reducida para probar un códec en el laboratorio.

FIGURA 1

Aplicación del método de referencia reducida de medición de la calidad perceptual para probar un códec en el laboratorio



BT.1885-01

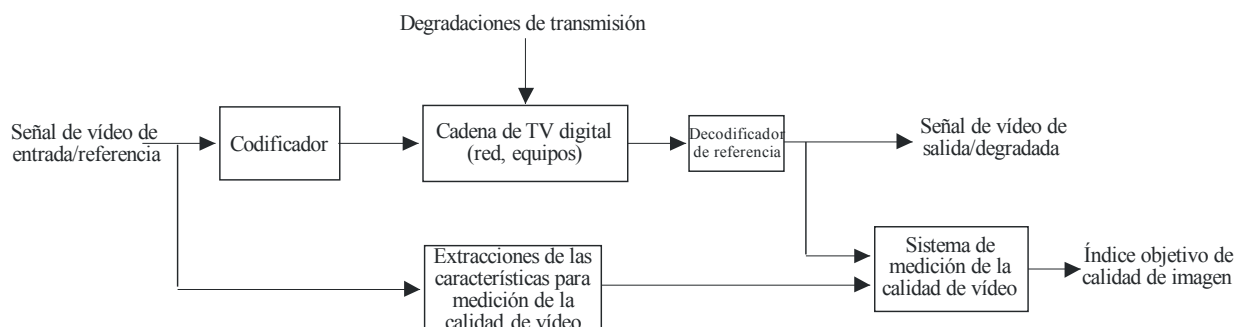
La comparación entre las señales de entrada y salida puede exigir un procedimiento de alineamiento temporal o de alineamiento espacial, este último tiene por objeto compensar cualquier desviación o recorte de la imagen horizontal o vertical. También puede que requiera la corrección de cualquier diferencia de desplazamiento o de ganancia, tanto en los canales de luminancia como en los de crominancia. A continuación se calcula el índice objetivo de calidad de imagen, normalmente aplicando un modelo perceptual de visión humana.

Los ajustes de alineamiento y ganancia se conocen con el nombre de normalización. Este proceso es necesario porque la mayoría de los métodos de referencia reducida comparan píxel a píxel las características extraídas de las imágenes de referencia y las imágenes procesadas. Un ejemplo sería el cálculo de la relación señal de cresta a ruido (PSNR). Únicamente se suprimen los cambios estáticos invariables en el tiempo que se producen en el vídeo. La Recomendaciones UIT-T J.244 y J.144 proporcionan métodos normalizados para informar sobre los valores necesarios a fin de normalizar el vídeo antes de evaluar la calidad objetiva. Las métricas de la calidad de vídeo descritas en el Anexo a la presente Recomendación incluyen métodos de normalización asociados. Los métodos de normalización alternativos pueden emplearse para los valores de la calidad de vídeo del Anexo siempre que ofrezcan la precisión en la normalización requerida.

Como las métricas de la calidad de vídeo se basan normalmente en aproximaciones a la respuesta visual humana, en vez de en mediciones de fenómenos de codificación específicos, en principio son igualmente válidas para los sistemas analógicos y para los sistemas digitales. También son en principio válidas para las cadenas donde se combinan los sistemas analógicos y digitales, o donde se concatenan los sistemas de compresión digital.

La Fig. 2 muestra un ejemplo de aplicación del método de referencia reducida para probar una cadena de transmisión.

FIGURA 2
Aplicación del método de referencia reducida de medición de la calidad perceptual para probar una cadena de transmisión



BT.1885-02

En este caso, se introducen en un decodificador de referencia señales procedentes de varios puntos de la cadena de transmisión; por ejemplo, el decodificador puede estar situado en un punto de la red, como en la Fig. 2, o directamente a la salida del decodificador como en la Fig. 1. Si la cadena de transmisión digital es transparente, la medición del índice objetivo de calidad de imagen en la fuente es igual a la medición en cualquier otro punto de la cadena.

Se acepta generalmente que el método de referencia completa proporciona la mejor precisión en las mediciones de la calidad de imagen perceptual. Se ha demostrado que el método tiene el potencial de una alta correlación con las evaluaciones subjetivas realizadas de conformidad con los métodos ACR-HR especificados en la Recomendación UIT-T P.910.

7 Conclusiones del Grupo de Expertos en calidad del vídeo

Los estudios de las mediciones de la calidad de vídeo perceptual se han llevado a cabo en un grupo informal, denominado Grupo de Expertos en calidad del vídeo (VQGE), que informa de sus resultados a las Comisiones de Estudio 9 y 12 del UIT-T y a la Comisión de Estudio 6 de Radiocomunicaciones. La prueba RRNR-TV recientemente completada por el VQEG evaluó el comportamiento de los algoritmos de medición de la calidad de vídeo perceptual de referencia para los formatos de imagen de la Recomendación UIT-R 601-6.

Basándose en la evidencia actual, el UIT-T puede recomendar por ahora seis métodos RR (Modelo_A 15k, Modelo_A 80k, Modelo_A 256k, Modelo_C 80k, Modelo_B 80k (525 líneas únicamente) y Modelo_B 256k (525 líneas únicamente)).

Las descripciones técnicas de estos modelos figuran en los Anexos A a C, respectivamente. Obsérvese que el orden de aparición de los Anexos es puramente arbitrario y no ofrece indicación alguna sobre el comportamiento de la predicción de la calidad.

Los Cuadros 1 y 2 muestran resultados significativos de la prueba VQEG RRNR-TV. Para el formato de 525 líneas, cuatro modelos (Modelo_A 15k, Modelo_A 80k, Modelo_A 256k y Modelo_C 80k) son estadísticamente mejores que la PSNR y dos modelos (Modelo_B 80k y Modelo_B 256k) son estadísticamente equivalentes a la PSNR. Cabe señalar que la PSNR fue calculada por la NTIA utilizando una búsqueda exhaustiva de los límites de calibración. Para el formato de 625 líneas, cuatro modelos (Modelo_A 15k, Modelo_A 80k, Modelo_A 256k y Modelo_C 80k) son estadísticamente equivalentes y estadísticamente mejores que la PSNR.

CUADRO 1

Prueba significativa para el formato de 525 líneas

| Formato de 525 líneas | Mejor comparación | Comparación con la PSNR | Correlación |
|------------------------------|--------------------------|--------------------------------|--------------------|
| Modelo_A 15k | 1 | 1 | 0,906 |
| Modelo_A 80k | 1 | 1 | 0,903 |
| Modelo_A 256k | 1 | 1 | 0,903 |
| Modelo_C 80k | 1 | 1 | 0,882 |
| Modelo_B 80k | 0 | 1 | 0,795 |
| Modelo_B 256k | 0 | 1 | 0,803 |
| PSNR_NTIA | 0 | 1 | 0,826 |

NOTA 1 – «1» en «Mejor comparación» indica que este modelo es estadísticamente equivalente al modelo de mejor comportamiento. «0» indica que este modelo no es estadísticamente equivalente al modelo de mejor comportamiento. «1» en «Comparación con la PSNR» indica que este modelo es estadísticamente equivalente al modelo de mejor comportamiento. «0» indica que este modelo no es estadísticamente equivalente al modelo de mejor comportamiento.

CUADRO 2

Prueba significativa para el formato de 625 líneas

| Formato de 625 líneas | Mejor comparación | Comparación con la PSNR | Correlación |
|------------------------------|--------------------------|--------------------------------|--------------------|
| Modelo_A 15k | 1 | 1 | 0,894 |
| Modelo_A 80k | 1 | 1 | 0,899 |
| Modelo_A 256k | 1 | 1 | 0,898 |
| Modelo_C 80k | 1 | 1 | 0,866 |
| PSNR_NTIA | 0 | 1 | 0,857 |

NOTA 1 – «1» en «Mejor comparación» indica que este modelo es estadísticamente equivalente al modelo de mejor comportamiento. «0» indica que este modelo no es estadísticamente equivalente al modelo de mejor comportamiento. «1» en «Comparación con la PSNR» indica que este modelo es estadísticamente equivalente al modelo de mejor comportamiento. «0» indica que este modelo no es estadísticamente equivalente al modelo de mejor comportamiento.

Los Cuadros 3 y 4 ofrecen información sobre los comportamientos del modelo en la prueba VQEG RRNR-TV.

CUADRO 3

**Descripción informativa sobre los comportamientos
del modelo en la prueba VQEG RRNR-TV
(formato de 525 líneas)**

| Formato de 525 líneas | Correlación | RMSE | OR |
|-----------------------|-------------|-------|-------|
| Modelo_A 15k | 0,906 | 0,418 | 0,385 |
| Modelo_A 80k | 0,903 | 0,423 | 0,378 |
| Modelo_A 256k | 0,903 | 0,424 | 0,378 |
| Modelo_B 80k | 0,795 | 0,598 | 0,667 |
| Modelo_B 256k | 0,803 | 0,587 | 0,647 |
| Modelo_C 80k | 0,882 | 0,465 | 0,513 |
| PSNR_NTIA | 0,826 | 0,556 | 0,571 |

CUADRO 4

**Descripción informativa sobre los comportamientos
del modelo en la prueba VQEG RRNR-TV
(formato de 625 líneas)**

| Formato de 625 líneas | Correlación | RMSE | OR |
|-----------------------|-------------|-------|-------|
| Modelo_A 15k | 0,894 | 0,524 | 0,468 |
| Modelo_A 80k | 0,899 | 0,513 | 0,462 |
| Modelo_A 256k | 0,898 | 0,516 | 0,468 |
| Modelo_C_80k | 0,866 | 0,585 | 0,583 |
| PSNR_NTIA | 0,857 | 0,605 | 0,564 |

Anexo A

Modelo A: Método de referencia reducida de la Universidad de Yonsei

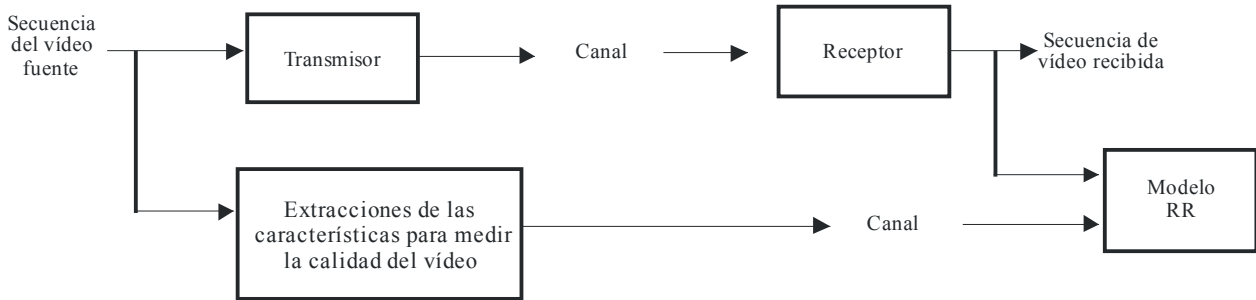
1 Introducción

Aunque la PSNR ha sido ampliamente utilizada como medida objetiva de la calidad del vídeo, también se ha informado de que no representa adecuadamente la calidad de vídeo perceptual. Analizando la forma en que los seres humanos perciben la calidad de vídeo se observa que el sistema visual humano es sensible a la degradación en torno a los bordes. En otras palabras, cuando los píxeles de los bordes de un vídeo aparecen borrosos, los evaluadores tienden a dar notas bajas a la calidad del vídeo aunque la PSNR sea elevada. Basándose en esta observación, se han desarrollado modelos de referencia reducida que miden fundamentalmente las degradaciones en los bordes.

La Fig. 3 ilustra la manera en que funciona un modelo de referencia reducida. Las características que se utilizarán para medir la calidad de vídeo en un punto de comprobación se extraen de la secuencia del vídeo fuente y se transmiten. El Cuadro 5 muestra las anchuras de banda del canal lateral para las características que se han probado en la prueba VQEG RRNR-TV.

FIGURA 3

Diagrama de bloques del modelo de referencia reducida



BT.1885-03

CUADRO 5

Anchuras de banda del canal lateral

| Formato del vídeo | Anchuras de banda probadas |
|-------------------|----------------------------------|
| Formato 525 | 15 Kbit/s, 80 Kbit/s, 256 Kbit/s |
| Formato 625 | 15 Kbit/s, 80 Kbit/s, 256 Kbit/s |

2 Modelos de referencia reducida EPSNR

2.1 PSNR del borde (EPSNR)

Los modelos RR miden principalmente las degradaciones en el borde. En los modelos, un algoritmo de detección del borde se aplica en primer lugar a la secuencia del vídeo fuente para localizar los píxeles del borde. A continuación, se mide la degradación de los píxeles calculando el error cuadrático medio. A partir del valor de este error se determina la PSNR del borde (EPSNR).

Se puede utilizar cualquier algoritmo de detección del borde aunque puede haber alguna pequeña diferencia en los resultados. Por ejemplo, puede emplearse cualquier operador de gradiente para localizar los píxeles del borde. Se ha propuesto un cierto número de operadores de gradiente. En muchos algoritmos de detección del borde la imagen del gradiente horizontal $g_{horizontal}(m,n)$ y la imagen del gradiente vertical $g_{vertical}(m,n)$ se calculan en primer lugar utilizando los operadores de gradiente. A continuación, puede calcularse la imagen del gradiente de magnitud $g(m,n)$ como sigue:

$$g(m,n) = |g_{horizontal}(m,n)| + |g_{vertical}(m,n)|$$

Por último, se aplica una operación de umbral a la imagen del gradiente de magnitud $g(m,n)$ para determinar los píxeles del borde. En otras palabras, los píxeles cuyos gradientes de magnitud rebasan un valor umbral se consideran píxeles del borde.

Las Figs. 4 a 8 ilustran este procedimiento. La Fig. 4 muestra una imagen de fuente. La Fig. 5 representa una imagen de gradiente horizontal $g_{horizontal}(m,n)$ obtenida aplicando un operador de gradiente horizontal a la imagen de fuente de la Fig. 4. La Fig. 6 muestra una imagen del gradiente vertical $g_{vertical}(m,n)$ que se obtiene aplicando un operador del gradiente vertical a la imagen de fuente de la Fig. 4. La Fig. 7 representa la imagen de gradiente de magnitud (imagen de borde) y la Fig. 8 muestra la imagen de borde binaria (imagen de máscara) obtenida aplicando el umbral a la imagen del gradiente de magnitud de la Fig. 7.

FIGURA 4

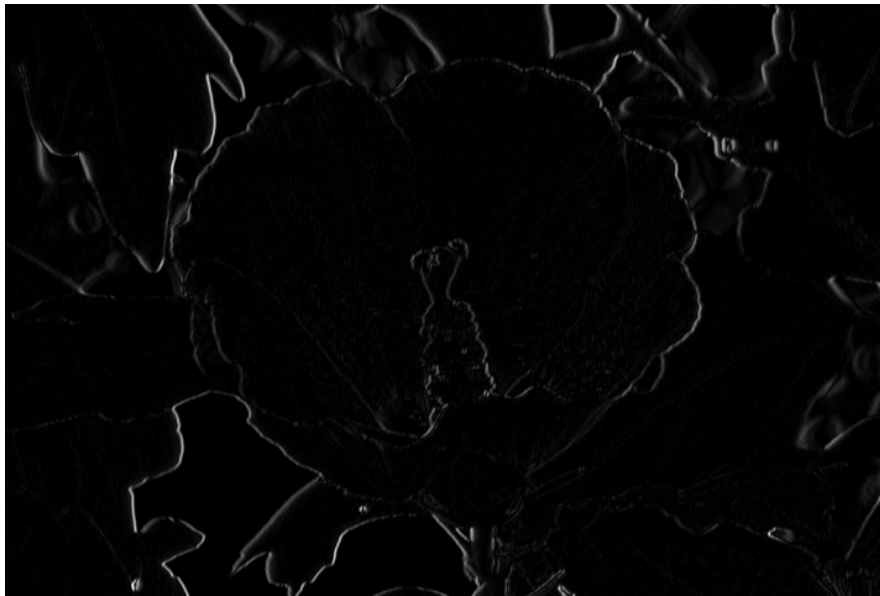
Imagen de fuente (imagen original)



BT.1885-04

FIGURA 5

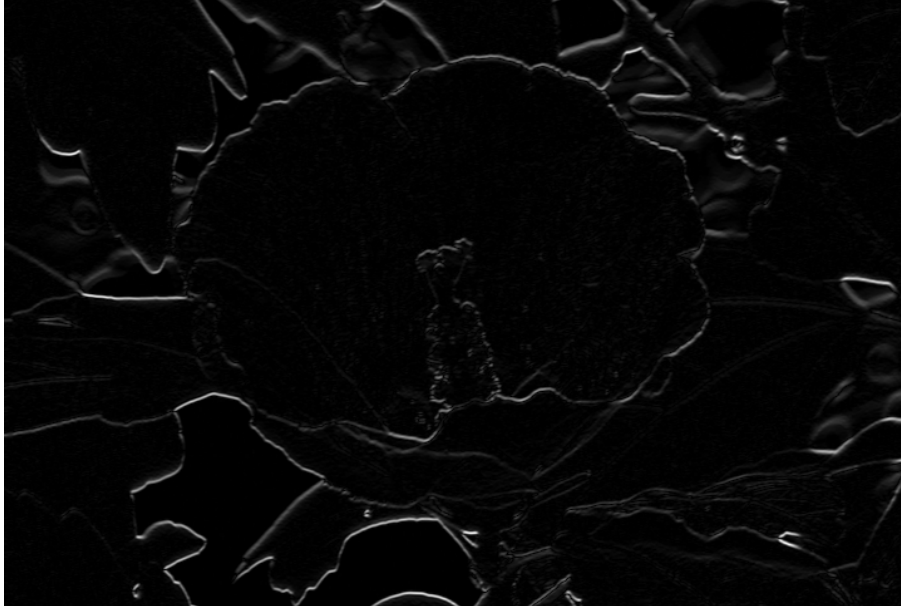
Imagen del gradiente horizontal obtenida aplicando un operador de gradiente horizontal a la imagen de fuente de la Fig. 4



BT.1885-05

FIGURA 6

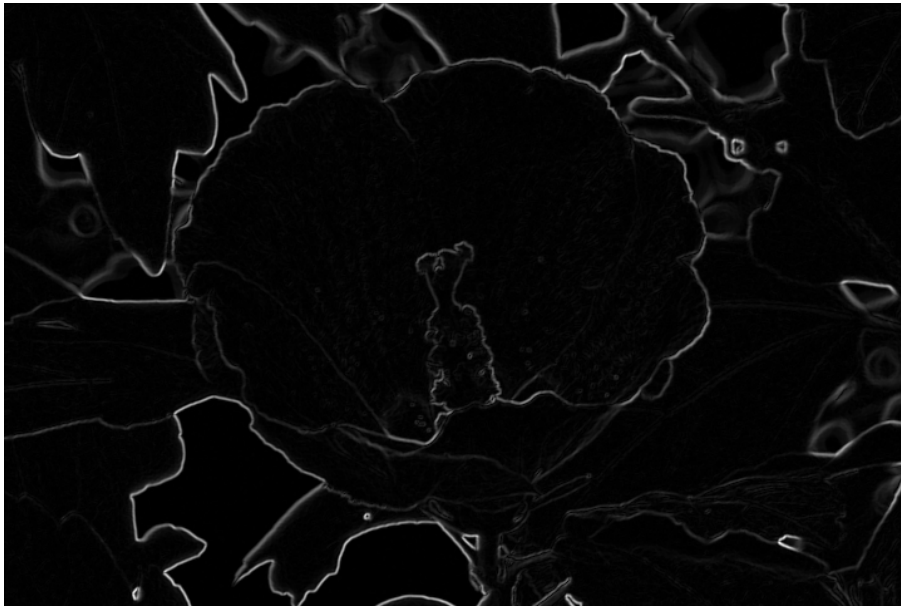
Imagen del gradiente vertical obtenida aplicando un operador de gradiente vertical a la imagen de fuente de la Fig. 4



BT.1885-06

FIGURA 7

Imagen del gradiente de magnitud



BT.1885-07

FIGURA 8

Imagen de borde binaria (imagen de máscara) obtenida aplicando el umbral a la imagen del gradiente de magnitud de la Fig. 7



BT.1885-08

Alternativamente, puede emplearse un procedimiento modificado para encontrar los píxeles del borde. Por ejemplo, puede aplicarse en primer lugar un operador de gradiente vertical a la imagen de fuente produciendo una imagen de gradiente vertical. A continuación, se aplica un operador de gradiente horizontal a la imagen de gradiente vertical produciendo una imagen de gradiente sucesivo modificada (imagen de gradiente horizontal y vertical). Por último, puede aplicarse una operación de umbral a la imagen de gradiente sucesivo modificada para encontrar los píxeles del borde. En otras palabras, los píxeles de la imagen del gradiente sucesivo modificada que rebasan el valor umbral, se consideran como píxeles del borde. Las Figs. 9 a 12 ilustran el procedimiento modificado. La Fig. 9 muestra una imagen de gradiente vertical $g_{vertical}(m,n)$ obtenida aplicando un operador de gradiente vertical a la imagen de fuente de la Fig. 4. La Fig. 10 muestra una imagen de gradiente sucesivo modificada (imagen de gradiente horizontal y vertical) que se obtiene aplicando un operador de gradiente horizontal a la imagen de gradiente vertical de la Fig. 9. La Fig. 11 representa la imagen de borde binario (imagen de máscara) obtenida mediante la aplicación del umbral a la imagen de gradiente sucesivo modificada de la Fig. 10.

Cabe señalar que ambos métodos pueden considerarse como un algoritmo de detección del borde. Puede seleccionarse cualquier algoritmo de detección del borde dependiendo de la naturaleza de los vídeos y de los algoritmos de compresión. Sin embargo, algunos métodos pueden resultar más eficaces que otros.

Por consiguiente, en el modelo, se aplica en primer lugar un operador de detección de borde produciendo imágenes de borde (véanse las Figs. 7 y 10). A continuación, se produce una imagen de máscara (imagen de borde binario) aplicando el umbral a la imagen de borde (véanse las Figs. 8 y 11). En otras palabras, los píxeles de la imagen de borde cuyo valor es más pequeño que el umbral t_e se fijan a cero y los píxeles cuyo valor es igual o superior al umbral se fijan a un valor distinto de cero. Las Figs. 8 y 11 muestran algunas imágenes de máscara. Dado que un vídeo puede considerarse como una secuencia de tramas o campos, el anterior procedimiento puede aplicarse a cada trama o campo de vídeo. Como el modelo puede utilizarse para vídeos basados en campo o en trama, el término «imagen» se empleará para indicar un campo o una trama.

FIGURA 9

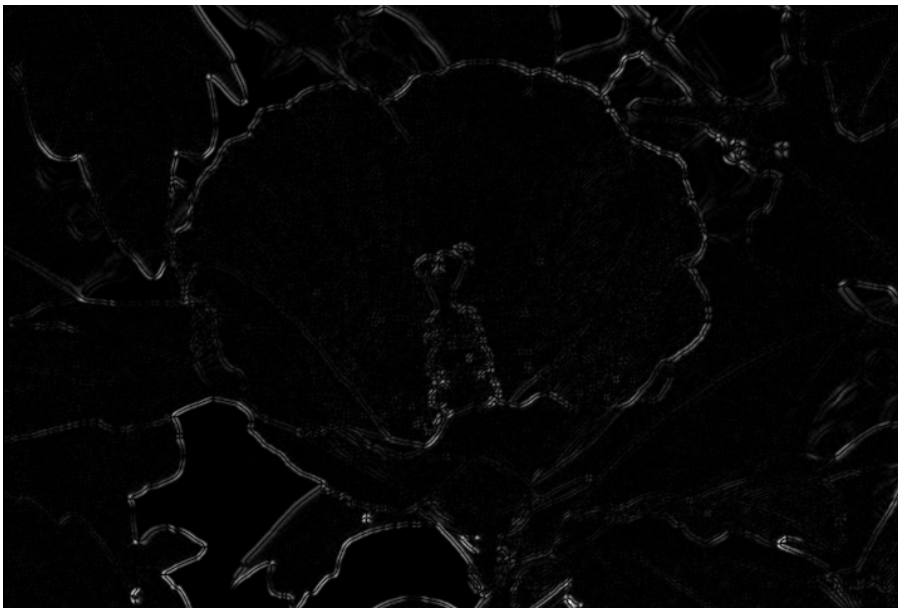
Imagen de gradiente vertical obtenida aplicando un operador de gradiente vertical a la imagen de fuente de la Fig. 4



BT.1885-09

FIGURA 10

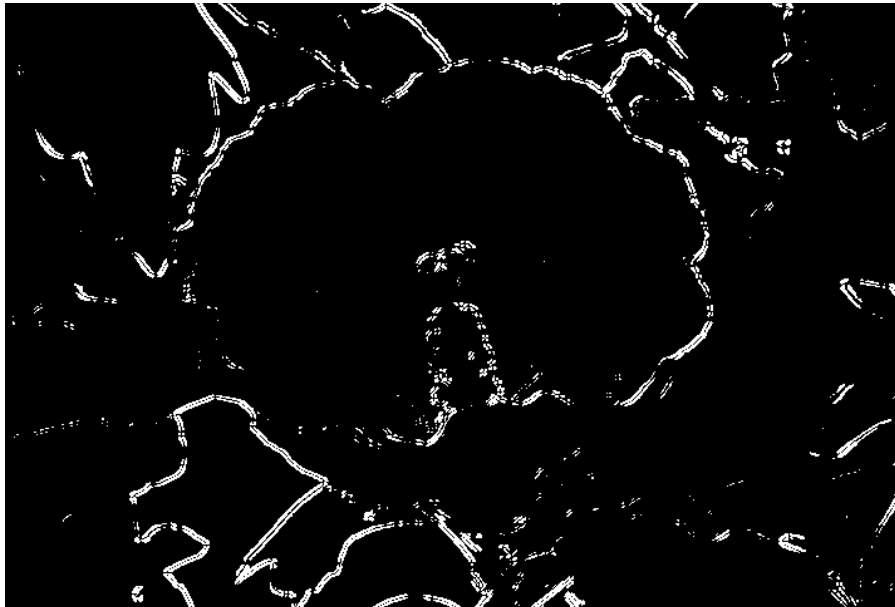
Imagen de gradiente sucesivo modificada (imagen de gradiente horizontal y vertical) obtenida aplicando un operador de gradiente horizontal a la imagen de gradiente vertical de la Fig. 9



BT.1885-10

FIGURA 11

Imagen de borde binario (imagen de máscara) obtenida aplicando el umbral a la imagen de gradiente sucesivo modificada de la Fig. 10



BT.1885-11

2.2 Selección de las características a partir de las secuencias del vídeo de fuente

Como se trata de un modelo RR, debe extraerse un conjunto de características de cada imagen de una secuencia de vídeo de fuente. En el modelo EPSBR RR, se selecciona un cierto número de píxels de borde de cada imagen. A continuación, se codifican y transmiten los datos de los emplazamientos y los valores de los píxels. Sin embargo, para algunas secuencias de vídeo, el número de píxels de borde puede ser muy pequeño cuando se utiliza un valor umbral fijo. En el caso más desfavorable, puede ser cero (imágenes en blanco o imágenes de frecuencia muy baja). Para abordar este problema, si el número de píxels de borde de una imagen es más pequeño que un valor determinado, el usuario puede reducir el valor umbral hasta que el número de píxels de borde sea superior a un valor concreto. Alternativamente, pueden seleccionarse píxels de borde que correspondan a los mayores valores de la imagen de gradiente horizontal y vertical. Cuando no hay píxels de borde (por ejemplo, imágenes en blanco) en una trama, puede seleccionarse aleatoriamente el número requerido de píxels o saltar la trama. Por ejemplo, si se seleccionan 10 píxels de borde de cada trama, pueden clasificarse los píxels de la imagen de gradiente horizontal o vertical de acuerdo con sus valores y seleccionar los 10 valores más elevados. No obstante, este procedimiento puede producir múltiples píxels de borde en emplazamientos idénticos. A fin de superar este problema, puede seleccionarse en primer lugar varias veces el número deseado de píxels de la imagen de gradiente horizontal y vertical y a continuación elegir aleatoriamente el número deseado de píxels de borde entre los píxels seleccionados de la imagen de gradiente horizontal y vertical. En los modelos probados en la prueba VQEG RRNR-TV, el número deseado de píxels de borde se elige aleatoriamente entre un gran conjunto de píxels de borde. Dicho conjunto se obtiene aplicando una operación umbral a la imagen de gradiente.

En los modelo EPSNR RR, los emplazamientos y los valores de los píxels de borde se codifican tras aplicar un filtro gaussiano paso bajo a los emplazamientos de los píxels seleccionados. Aunque en la prueba el VQEG RRNR-TV se utilizó el filtro paso bajo gaussiano (5×3), pueden emplearse diferentes filtros paso bajo dependiendo de los formatos de vídeo. Cabe señalar que durante el proceso de decodificación, puede aplicarse un recorte. Para evitar la selección de píxels de borde en

las zonas recortadas, el modelo selecciona píxeles de borde en la zona media (véase la Fig. 12). El Cuadro 6 muestra los tamaños tras el recorte. Dicho Cuadro también presenta el número de bits necesarios para codificar el emplazamiento y el valor de píxel de un píxel de borde.

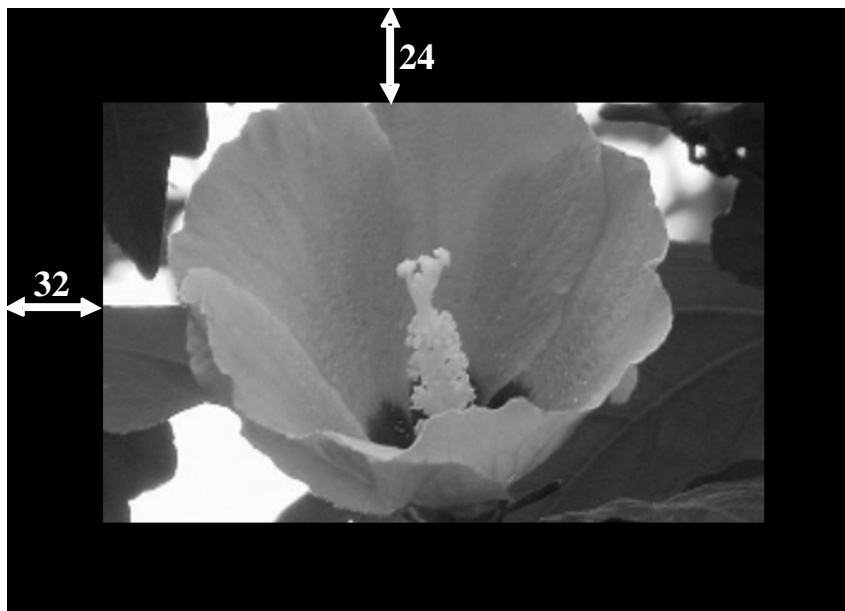
CUADRO 6

Requisitos de bits por cada píxel de borde

| Formato del vídeo | Tamaño | Tamaño tras el recorte | Bits por emplazamiento | Bits por valor de píxel | Número total de bits por píxel |
|-------------------|-----------|------------------------|------------------------|-------------------------|--------------------------------|
| 525 | 720 × 486 | 656 × 438 | 19 | 8 | 27 |
| 625 | 720 × 576 | 656 × 528 | 19 | 8 | 27 |

FIGURA 12

Ejemplo de recorte y la zona media



BT.1885-12

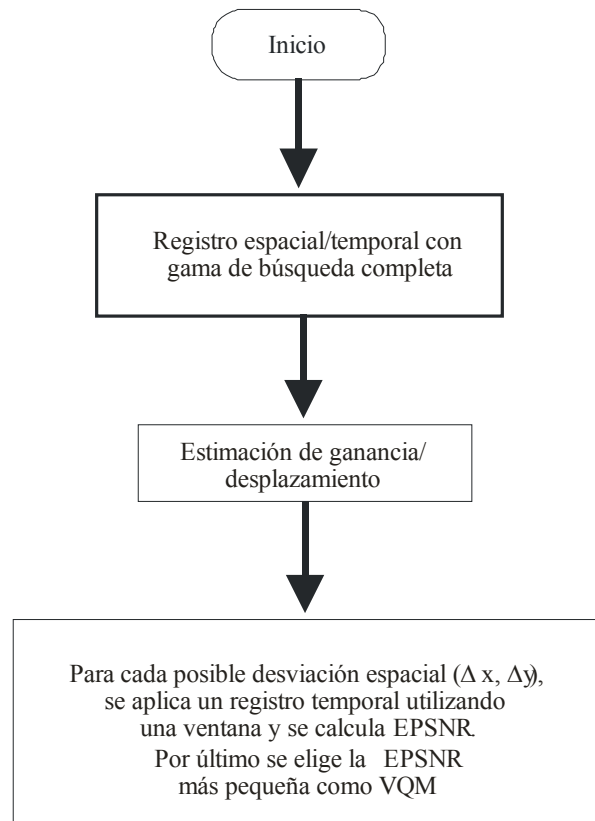
El modelo selecciona los píxeles de borde de cada trama de conformidad con la anchura de banda permitida (véase el Cuadro 5). El Cuadro 7 muestra el número de píxeles de borde por trama que puede transmitirse en las anchuras de banda probadas.

CUADRO 7

Número de píxeles de banda por trama

| Formato del vídeo | 15 Kbit/s | 80 Kbit/s | 256 Kbit/s |
|-------------------|-----------|-----------|------------|
| 525 | 16 | 74 | 238 |
| 625 | 20 | 92 | 286 |

FIGURA 13

Organigrama del modelo

BT.1885-13

2.3 Registro espacial/temporal y ajuste ganancia/desplazamiento

Antes de calcular la diferencia entre los píxeles de borde de la secuencia del vídeo de fuente y los de la secuencia del vídeo procesado que es la secuencia de vídeo recibida en el receptor, el modelo aplica en primer lugar un registro espacial/temporal y un ajuste de ganancia/desplazamiento. Se empleó el método de calibración (Anexo B) de la Recomendación UIT-T J.244. Para transmitir las características de ganancia y desplazamiento de la citada Recomendación (Anexo B) se utilizó el 30% de las anchuras de banda disponibles en la prueba VQEG RRNR-TV. Como la secuencia de vídeo está entrelazada, el método de calibración se aplicó tres veces: los campos pares, los campos impares y las tramas combinadas. Si la diferencia entre el error en el campo par (PSNR) y el error en el campo impar era mayor que un valor umbral, se utilizaron los resultados del registro (desviación x y desviación y) con la PSNR más baja. En los otros casos, se emplearon los resultados del registro con las tramas combinadas. En la prueba VQEG RRNR-TV el valor umbral se fijó a 2 dB.

En el punto de comprobación, la secuencia de vídeo procesada debe alinearse con los píxels de borde extraídos de la secuencia del vídeo de fuente. Sin embargo, si la anchura de banda del canal lateral es pequeña, solo están disponibles unos pocos píxels de borde de la secuencia de vídeo fuente (véase la Fig. 14). Por tanto, el registro temporal puede ser inexacto si se lleva a cabo utilizando una sola trama (véase la Fig. 15). A fin de solucionar este problema, el modelo utiliza una ventana para el registro temporal. En vez de emplear una sola trama de la secuencia de vídeo procesada, el modelo construye una ventana que consiste en un cierto número de tramas adyacentes para encontrar la desviación temporal óptima. La Fig. 16 ilustra el procedimiento. El error cuadrático medio en la ventana se determina como sigue:

$$\text{MSE}_{\text{ventana}} = \frac{1}{N_{\text{viento}}} \sum (E_{\text{SRC}}(i) - E_{\text{PVS}}(i))^2$$

donde:

- $\text{MSE}_{\text{ventana}}$: es el error cuadrático medio de la ventana
- $E_{\text{SRC}}(i)$: es un píxel de borde dentro de la ventana que tiene su correspondiente píxel en la secuencia de vídeo procesada
- $E_{\text{PVS}}(i)$: es un píxel de la secuencia de vídeo procesada correspondiente al píxel de borde
- N_{viento} : es el número total de píxels de borde utilizados para calcular el $\text{MSE}_{\text{ventana}}$.

Este error cuadrático medio de la ventana se emplea como la diferencia entre una trama de la secuencia de vídeo procesada y la correspondiente trama de la secuencia de vídeo de fuente.

El tamaño de la ventana puede determinarse considerando la naturaleza de la secuencia de vídeo procesada. Para una aplicación típica, se recomienda una ventana correspondiente a dos segundos. Alternativamente, pueden aplicarse varios tamaños de ventana y utilizarse el que proporciona el error cuadrático medio más pequeño. Además, pueden usarse diferentes centros de ventana para considerar los saltos de trama debidos a los errores de transmisión (véase la Fig. 20).

FIGURA 14

Selección de píxel de borde de la secuencia de vídeo de fuente

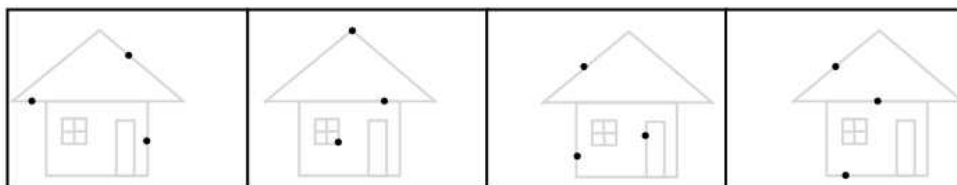
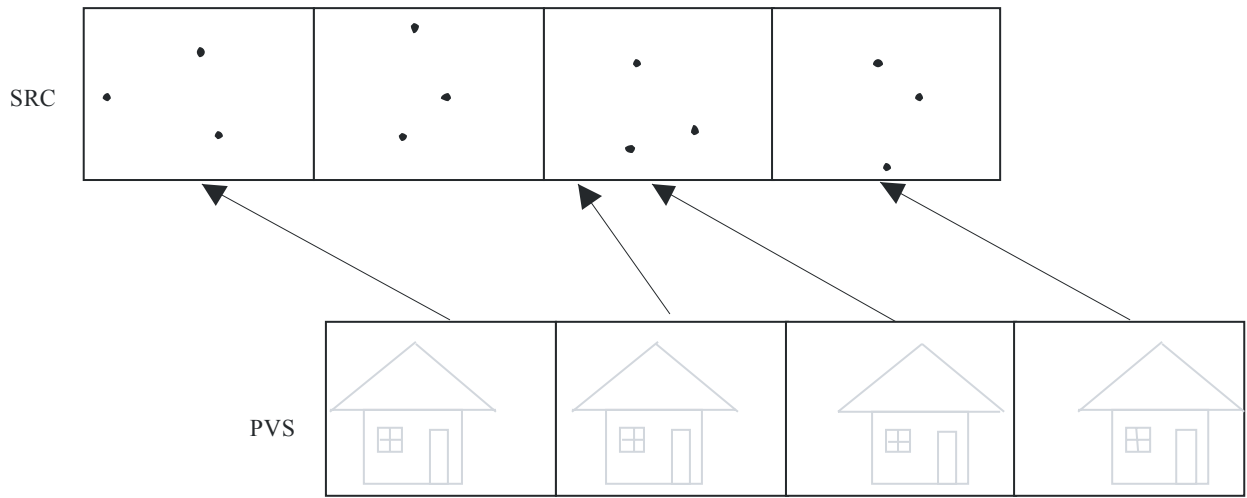


FIGURA 15

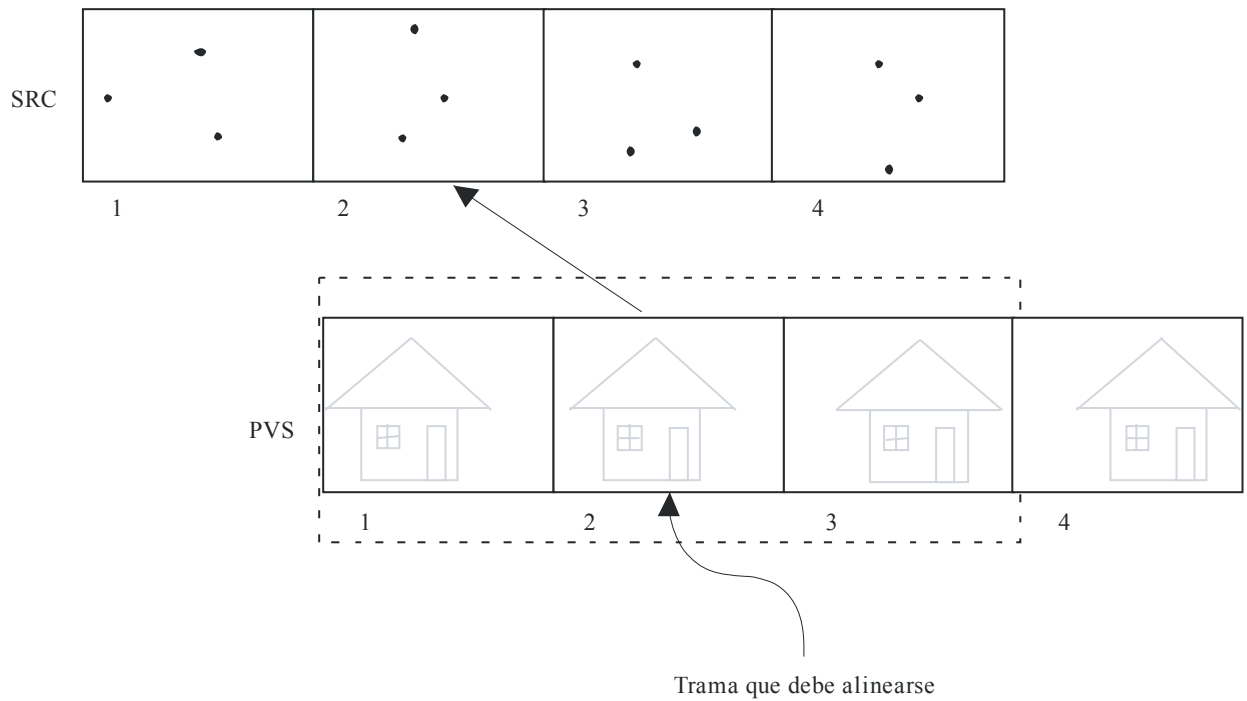
Alineamiento de la secuencia de vídeo procesada con los píxels de borde de la secuencia de vídeo de fuente



BT.1885-15

FIGURA 16

Alineamiento de la secuencia de vídeo procesada con los píxels de borde utilizando una ventana

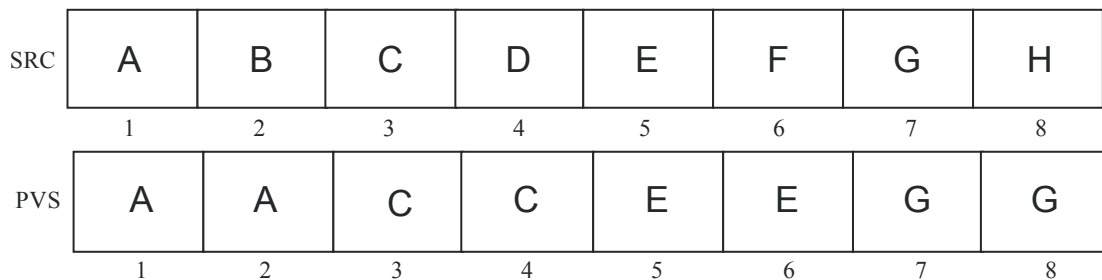


BT.1885-16

Cuando se codifica la secuencia de vídeo de fuente con altas relaciones de compresión, el codificador puede reducir el número de tramas por segundo y la secuencia de vídeo procesada tiene tramas repetidas (véase la Fig. 17). En la Fig. 17 la secuencia de vídeo procesada no tiene tramas correspondientes a algunas tramas de la secuencia de vídeo de fuente (tramas 2^a, 4^a, 6^a y 8^a). En este caso, el modelo no utiliza tramas repetidas para calcular el error cuadrático medio. En otras palabras, el modelo lleva a cabo un registro temporal utilizando la primera trama (trama válida) de cada bloque repetido. Por consiguiente, en la Fig. 18 sólo tres tramas (3^a, 5^a y 7^a) en la ventana se utilizan para el registro temporal.

FIGURA 17

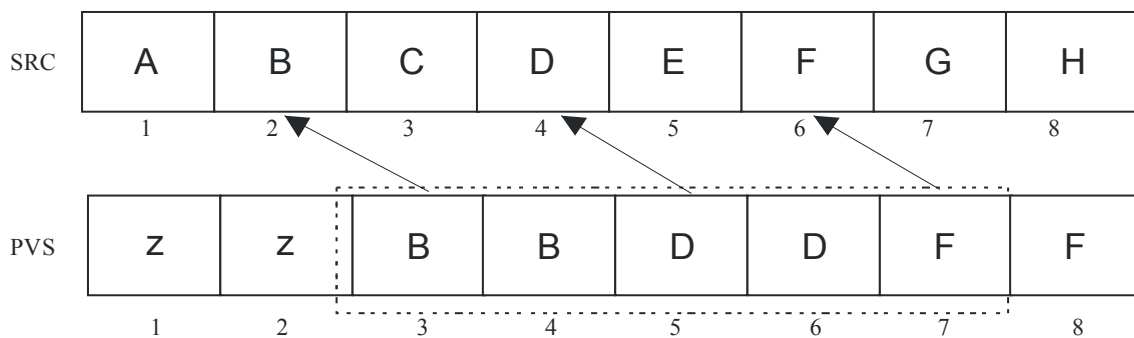
Ejemplo de tramas repetidas



BT.1885-17

FIGURA 18

Manejo de tramas repetidas

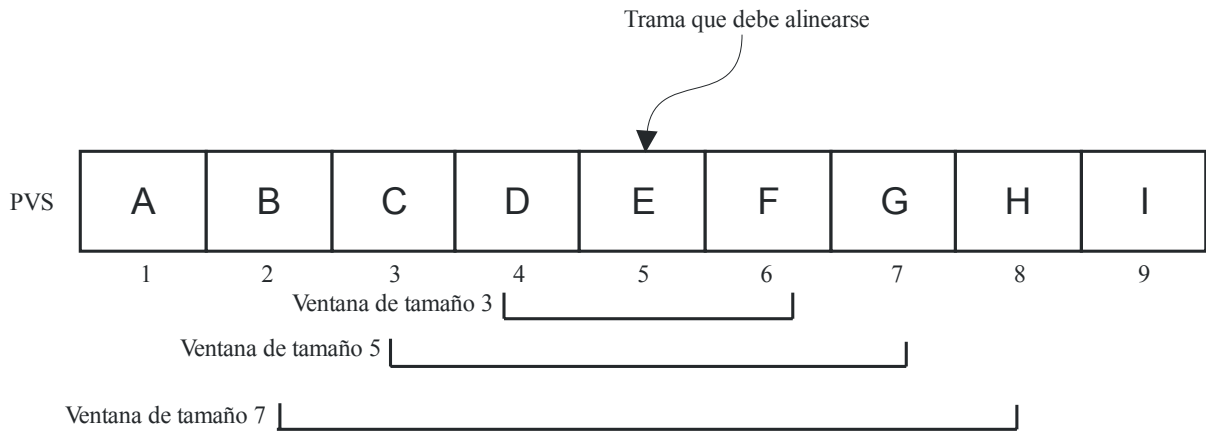


BT.1885-18

Es posible tener una secuencia de vídeo procesada con repetición de trama irregular que puede dar a que el método de registro temporal utilice una ventana produciendo resultados imprecisos. Para abordar este problema, es posible ajustar localmente cada trama de la ventana a un valor determinado (por ejemplo ± 1) como muestra la Fig. 21 tras el registro temporal que utiliza una ventana. A continuación, se utiliza el ajuste local que proporciona el mínimo MSE para calcular la EPSNR.

FIGURA 19

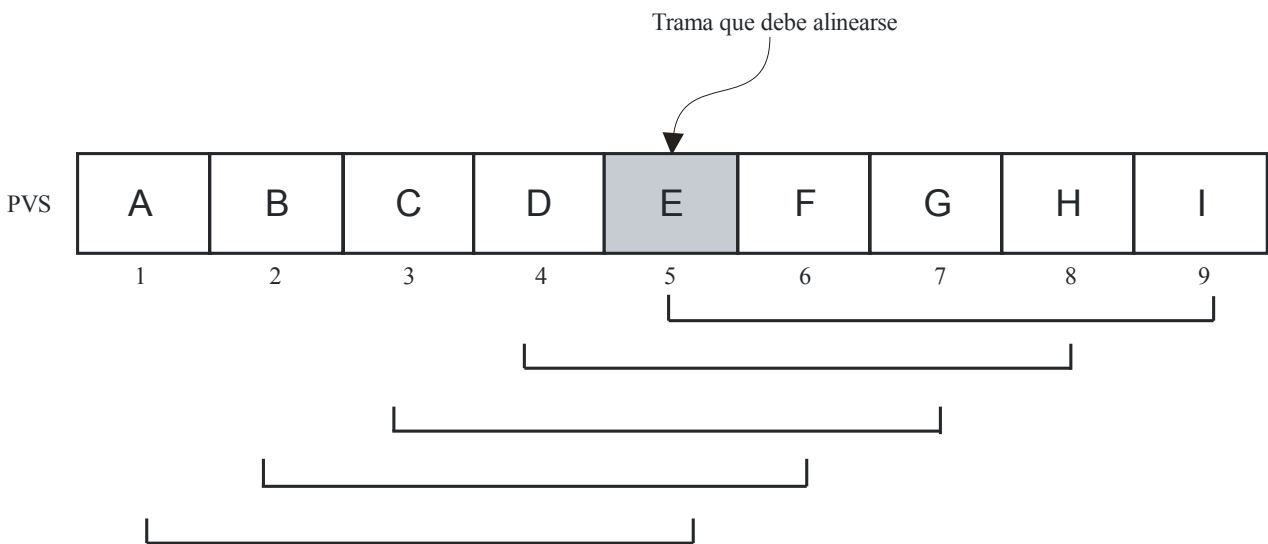
Ventanas con varios tamaños



BT.1885-19

FIGURA 20

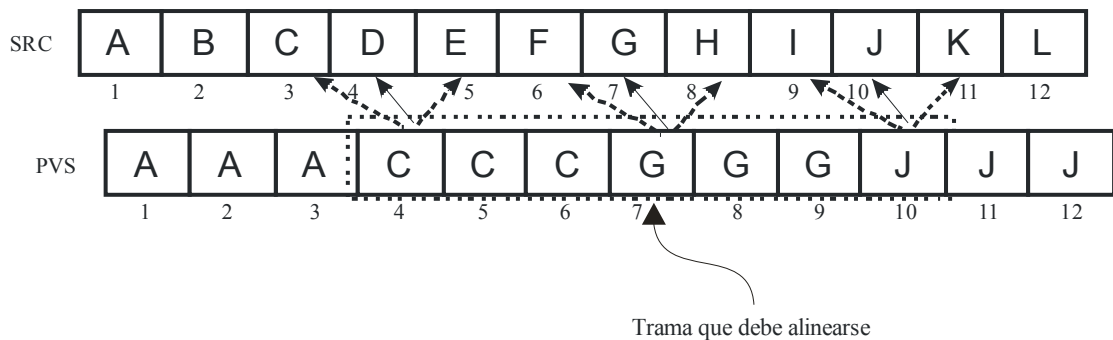
Centros de ventana



BT.1885-20

FIGURA 21

Ajuste local del registro temporal utilizando una ventana



BT.1885-21

2.4 Cálculo de la EPSNR y postprocesamiento

Una vez realizado el registro temporal, se calcula la media de las diferencias entre los píxeles de borde de la secuencia de vídeo de fuente y los correspondientes píxeles de la secuencia de vídeo procesada, esta media puede considerarse el error cuadrático medio de borde de la secuencia de vídeo procesada (MSE_{borde}). Por último, se calcula la EPSNR (PSNR de borde) como sigue:

$$EPSNR = 10 \log_{10} \left(\frac{P^2}{MSE_{borde}} \right)$$

siendo p el valor de cresta de la imagen.

1) Tramas congeladas

Puede haber repetición de tramas debido a las velocidades de trama reducidas y congelación de trama debido al error de transmisión que degrada la calidad de vídeo perceptual. Para tratar este efecto, el modelo aplica el siguiente ajuste antes de calcular la EPSNR.

$$MSE_{trama_congelada_considerada} = MSE_{borde} \times \frac{K \times N_{total_tramas}}{N_{total_tramas} - N_{total_tramas_congeladas}}$$

donde:

$MSE_{trama_congelada_considerada}$: es el error cuadrático medio que tiene en cuenta las tramas repetidas y congeladas

N_{total_tramas} : es el número total de tramas, $N_{total_tramas_congeladas}$

K : es una constante.

En el modelo probado en la prueba VQEG RRNR-TV, se fijó K al valor 1.

2) Alta frecuencia y movimiento rápido

Si la secuencia de vídeo contiene una cantidad elevada de altas frecuencia y movimientos rápidos, la calidad perceptual tiende a aumentar para el mismo MSE. A fin de considerar este efecto, se define como sigue la diferencia de trama normalizada (NFD) y la energía de alta frecuencia normalizada (NHFE):

$$NFD = \frac{FD}{energía\ media\ por\ píxel}$$

donde $FD = \frac{1}{N_F} \sum_i \sum_{k=1}^{altura\ anchura} \sum_{j=1} (Trama_i[j,k] - Trama_{i-1}[j,k])^2$ y N_F es el número de tramas utilizado

en el sumatorio. Cabe señalar que se excluyen las tres diferencias de trama más altas a la hora de calcular FD con objeto de eliminar los cambios de escena en el cálculo de la diferencia de trama media, suponiendo secuencias de vídeo de 8 s. La energía de alta frecuencia normalizada (NHFE) se determina calculando el valor medio de las energías de alta frecuencia (véase la Fig. 22) tras aplicar la transformada de Fourier 2D:

$$NHFD = \frac{media\ de\ las\ energías\ de\ alta\ frecuencia}{energía\ media\ por\ píxel}$$

Por último, se utilizan las siguientes ecuaciones:

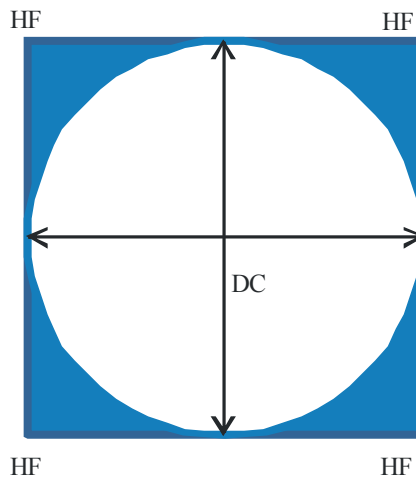
```

IF(SNFD > 0.35 && SNHFE > 2.5) {
    IF(EPSNR < 20) EPSNR = EPSNR+3
    ELSE IF(EPSNR < 35) EPSNR = EPSNR+5
}
ELSE IF((SNFD > 0.2 && SNHFE > 1.5) || (SNFD>0.27) && SNHFE > 1.3)) {
    IF(28 < EPSNR < 40) EPSNR = EPSNR + 3
    IF(EPSNR > 40) EPSNR = 40
}
    
```

donde SNDF es la NFD de fuente y SNHFE es la NHFE de fuente. Hay que tener en cuenta que SND y SNHFE se calculan a partir de SRC y se transmiten como parte de los datos sobre las características (1 byte por cada uno).

FIGURA 22

Cálculo de la energía de alta frecuencia normalizada (NHFE). Las energías de alta frecuencia se determinan a partir de la zona sombreada



BT.1885-22

3) Borrosidad

Para considerar los efectos de borrosidad se emplean las siguientes ecuaciones:

```

IF (NHFE/SNHFE < 0.5)
    IF(EPSNR>26)     EPSNR = 26
ELSE IF (NHFE/SNHFE < 0.6)
    IF(EPSNR>32)     EPSNR = 32
ELSE IF (NHFE/SNHFE < 0.7)
    IF(EPSNR>36)     EPSNR = 36
ELSE IF (NHFE/SNHFE > 1.2)
    IF(EPSNR>23)     EPSNR = 23
ELSE IF (NHFE/SNHFE > 1.1)
    IF (EPSNR>25)     EPSNR = 25
    
```

donde NHFE es PVS NHFE

4) Bloqueo

Para considerar los efectos de bloqueo se determinan los valores medios de las diferencias de las columnas. Suponiendo módulo 8, el grado de bloqueo para la i-ésima trama se calcula como sigue:

$$Blk[i] = \frac{\text{mayor diferencia en las columnas}}{\text{segunda mayor diferencia en las columnas}}$$

El grado de bloqueo final (*Bloqueo*) se determina promediando los grados de bloqueo de trama:

$$\text{Bloqueo} = \frac{1}{\text{número de tramas}} \sum_i \text{Blk}[i]$$

Por último, se utilizan las siguientes ecuaciones:

```
IF(BLOCKING > 1.4) {
  IF (20≤EPSNR<25) EPSNR = EPSNR-1.086094*BLOCKING-0.601316
  ELSE IF (EPSNR<30) EPSNR = EPSNR-0.577891*BLOCKING-3.158586
  ELSE IF (EPSNR<35) EPSNR = EPSNR-0.223573*BLOCKING-3.125441
}
```

5) Máximas tramas congeladas

Los errores de transmisión pueden causar largas tramas congeladas. Para considerar esta tramas se utilizan las siguientes ecuaciones:

```
IF(MAX_FREEZE > 22 AND EPSNR>28) EPSNR = 28
ELSE IF(MAX_FREEZE > 10 AND EPSNR>34) EPSNR = 34
```

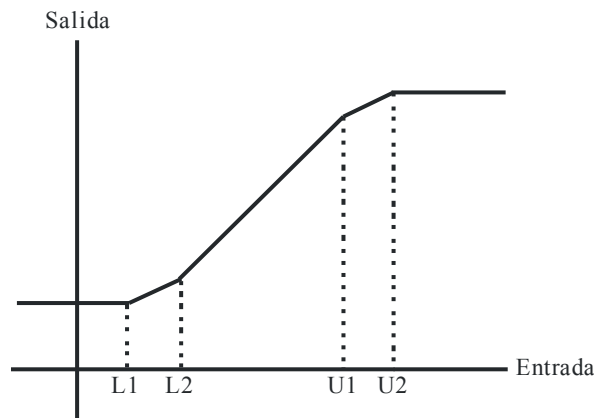
donde MAX_FREEZE es la mayor duración de las tramas congeladas. Cabe señalar que si la secuencia de vídeo no es de 8 s, deben utilizarse umbrales distintos.

6) Ajuste lineal por tramos

Cuando la EPSNR rebasa un cierto valor, la calidad perceptual se satura. En este caso, es posible fijar el límite superior de la EPSNR. Además, cuando se desea una relación lineal entre la EPSNR y la DMOS (diferencia de la nota media de opinión) puede aplicarse una función lineal por tramos como la que se representa en la Fig. 23. En el modelo probado en la prueba VQEG RRNR-TV, el límite superior se fija a 48 y el límite inferior a 15.

FIGURA 23

Función lineal por tramos para la relación lineal entre la EPSNR y la DMOS



BT.1885-23

Los modelos de referencia reducida de EPSNR para la medición objetiva de la calidad de vídeo se basan en la degradación del borde. Los modelos pueden implementarse en tiempo real mediante utilización moderada de la potencia de cálculo. Los modelos son adecuados para aplicaciones que requieren comprobación de la calidad de vídeo en tiempo real cuando se dispone de canales laterales.

Anexo B

Modelo B: Método de referencia reducida NEC

Este Anexo proporciona una descripción funcional completa del modelo RR en el que son los valores de actividad en vez de los valores de píxel para bloque de píxels de un tamaño determinado los que se transmiten al extremo del cliente. La calidad de vídeo se estima basándose en la diferencia de actividad entre el canal de referencia de la fuente (SRC) y la secuencia de vídeo procesada (PVS). Las ponderaciones psicovisuales con respecto a la diferencia de actividad se realizan para mejorar la precisión de la estimación.

Este modelo no necesita registros espaciales y de ganancia y desplazamiento que exijan una gran potencia de cálculo. Además, puede implementarse mediante un programa de 30 líneas y un programa de 250 líneas en los extremos del servidor y del cliente, respectivamente. Por tanto, es adecuado para supervisar la calidad de vídeo en tiempo real de los servicios de radiodifusión que son los máximos beneficiados de las implementaciones fáciles y de baja complejidad.

1 Resumen

En el modelo RR, se transmiten al extremo de cliente los valores de actividad en vez de los valores de píxel para los bloques de píxels de un tamaño determinado. La calidad de vídeo se estima basándose en la diferencia de actividad entre el SRC y la PVS. Las ponderaciones psicovisuales con respecto a la diferencia de actividad se realizan para mejorar la precisión de la estimación.

Este modelo no necesita registros espaciales y de ganancia y desplazamiento que exijan una gran potencia de cálculo. Además, puede implementarse mediante un programa de 30 líneas y un programa de 250 líneas en los extremos del servidor y del cliente, respectivamente. Por tanto, es adecuado para supervisar la calidad de vídeo en tiempo real de los servicios de radiodifusión que son los máximos beneficiados de las implementaciones fáciles y de baja complejidad.

2 Definiciones

Actividad – Valor medio de la diferencia absoluta entre cada valor de luminancia y la media de los valores de luminancia para un bloque de tamaño determinado.

Bloque – Conjunto de $M \times N$ píxels (M columnas por N filas).

Trama – Una imagen de televisión completa.

Ganancia – Factor de escala multiplicativo aplicado por el circuito ficticio de referencia (HRC) a todos los píxels de un plano de imagen individual (por ejemplo, crominancia, luminancia). La ganancia de la señal de luminancia se conoce normalmente como contraste.

Circuito ficticio de referencia (HRC, *hypothetical reference circuit*) – Sistema de vídeo sometido a prueba tal como un códec o un sistema de transmisión de vídeo digital.

Luminancia (Y) – Porción de la señal de vídeo que transporta predominantemente la información de luminancia (es decir, la parte en blanco y negro de la imagen).

National Television Systems Committee (NTSC) – Sistema compuesto de vídeo de color analógico de 525 líneas [1].

Desplazamiento o desplazamiento de nivel – Factor aditivo aplicado al circuito ficticio de referencia (HRC) a todos los píxels de un plano de imagen individual (por ejemplo, luminancia, crominancia). El desplazamiento de la señal de luminancia se conoce normalmente como brillo.

Relación señal de cresta/ruido (PSNR, *peak signal-to-noise ratio*) – Relación entre la máxima potencia posible de la señal y la potencia de ruido perturbador.

Líneas con alternancia de fase (PAL, *phase-altering line*) – Sistema compuesto de vídeo de color analógico de 625 líneas.

Exploración de barrido por cuadrículas («*raster scan*») – Conversión de un modelo bidimensional rectangular en un modelo unidimensional de forma que el primer punto del modelo unidimensional procede de la primera fila superior del modelo bidimensional explorada de izquierda a derecha, seguida de manera similar por las filas segunda, tercera, etc., del modelo (desplazamiento hacia abajo) explorada cada una de ellas de izquierda a derecha.

Referencia reducida (RR) – Método de medición de la calidad de vídeo que utiliza características de anchura de banda baja, extraídas de los trenes de vídeo original o procesados, en contraposición al vídeo de referencia completa que requiere un conocimiento completo de los trenes de vídeo original y procesado [2]. Los métodos de referencia reducida presentan ventajas a la hora de supervisar la calidad del servicio de extremo a extremo ya que la información de referencia reducida se transmite fácilmente a cualquier punto de la red de telecomunicaciones.

Región de interés (ROI, *region of interest*) – Retícula de imagen (especificada en coordenadas rectangulares) utilizada para denominar una subregión particular de una campo o una trama de vídeo.

Escena – Secuencia de tramas de vídeo.

Registro espacial – Proceso utilizado para estimar y corregir las derivas espaciales de la secuencia de vídeo procesada con respecto a la secuencia de vídeo original.

Registro temporal – Proceso empleado para estimar y corregir una deriva temporal (es decir, un retardo del vídeo) de la secuencia de vídeo procesada con respecto a la secuencia de vídeo original.

Métrica de la calidad de vídeo (VQM, *video quality metric*) – Medición global de la degradación del vídeo. La VQM se expresa como un solo número y tiene una gama de salida nominal que va de cero a uno, correspondiendo el cero a la ausencia de percepción de degradación y el uno a la máxima degradación percibida.

3 Consideraciones generales sobre el cálculo de la métrica de calidad del vídeo

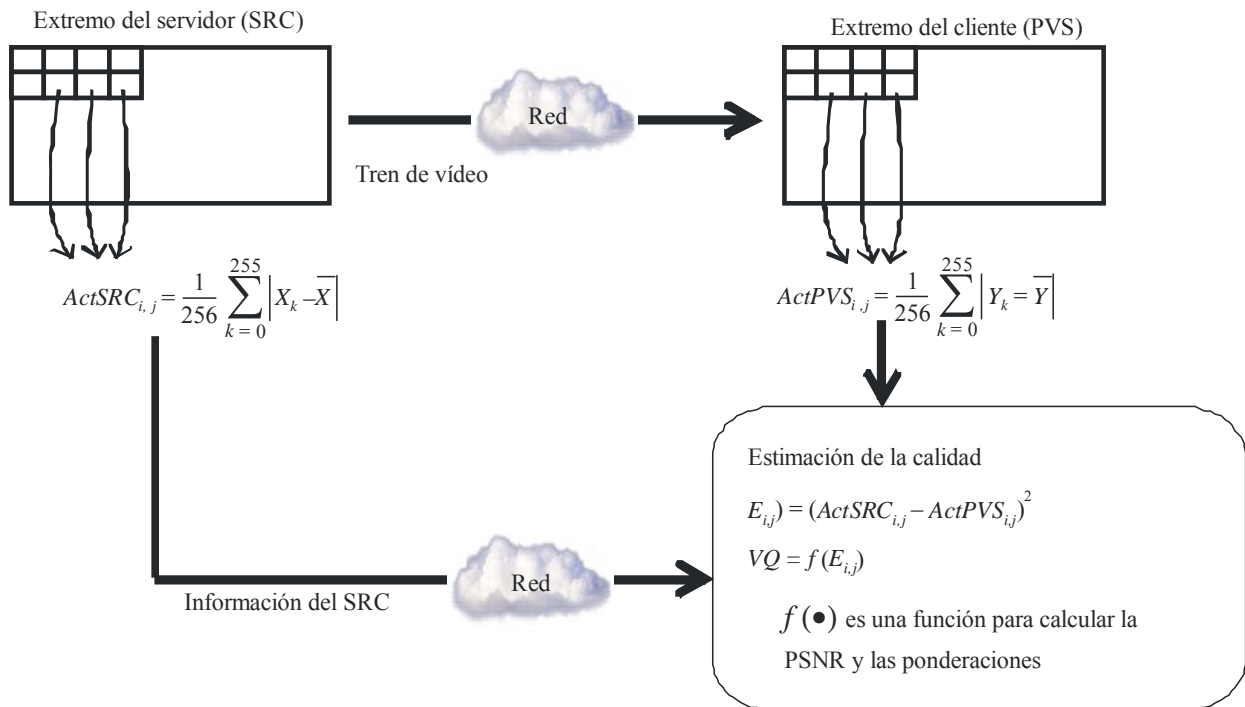
El modelo RR transmite los valores de actividad para bloques de píxels individuales de un tamaño determinado al extremo del cliente. Este valor indica la varianza de los valores de luminancia en el bloque. La Fig. 1 resume el modelo RR. Como muestra la Fig. 24, la calidad de vídeo se determina basándose en la diferencia de actividad entre el SRC y la PVS. Además, se aplican ponderaciones psicovisuales con respecto a la diferencia de actividad para lograr una mayor precisión en la estimación. La estimación de la calidad de vídeo se lleva a cabo mediante los siguientes pasos:

- 1) El valor de actividad para cada bloque de píxels de luminancia de 16×16 del SRC se calcula en el extremo del servidor. Todos los valores de actividad se transmiten a continuación al extremo del cliente. El valor de actividad de un bloque se define como la diferencia absoluta media de los valores individuales y su valor medio.
- 2) Los correspondientes valores de actividad se calculan en el extremo del cliente con respecto a la PVS.
- 3) En el extremo del cliente, cada bloque se evalúa inicialmente con su error cuadrático, que es la diferencia cuadrática entre los valores de actividad del SRC y la PVS.

- 4) Se aplican ponderaciones psicovisuales a los errores cuadráticos en los bloques con componente de frecuencia espacial elevada, un color específico, una gran diferencia entre tramas y un cambio de escena.
- 5) Se obtiene una nota provisional de la calidad de vídeo a partir de la suma de los errores cuadráticos ponderados de la misma forma que en el cálculo de la PSNR.
- 6) La nota se modifica para reflejar las degradaciones perceptualmente muy dañinas debidas al bloqueo y a las degradaciones locales. Por último, la nota modificada representa la calidad de vídeo medida de la PVS en el modelo RR.

FIGURA 24

Estimación de la calidad de vídeo basada en la diferencia de actividad



BT.1885-24

4 Algoritmo detallado

4.1 Extremo del servidor

- 1) Los píxeles de luminancia de un SRC se dividen en bloques de 16×16 píxeles en cada trama trascurrido un segundo desde la parte superior de la secuencia de vídeo. En el primer segundo, la información del SRC no se transmite ya que es difícil para el sistema visual humano detectar la degradación de la calidad de vídeo en las escenas que aparecen inmediatamente después de la primera trama.
- 2) En cada bloque, salvo los del borde de la trama, se calculan los valores de actividad (actividad SRC: $ActSRC_{i,j}$). La Fig. 25 describe los bloques cuyos valores de actividad se calculan y transmiten. La actividad SRC se calcula como sigue:

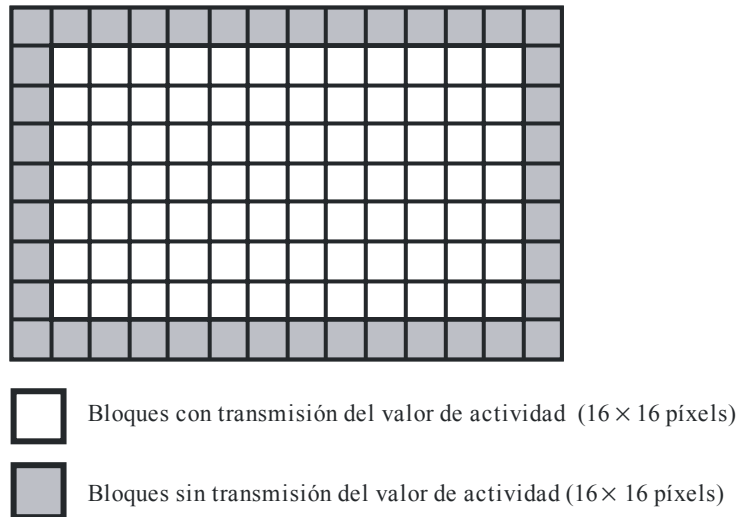
$$ActSRC_{i,j} = \frac{1}{256} \sum_{k=0}^{255} |X_k - \bar{X}|$$

donde X_k es un valor de luminancia en un bloque de tamaño determinado del SRC, \bar{X} es su valor medio, i es un número de trama y j es un número de bloque en la trama.

- 3) Los valores de actividad, que se expresan con ocho bits por bloque, se transmiten al extremo del cliente en el orden de exploración de barrido de cuadrículas («raster scan») transcurrido un segundo desde la parte superior de la secuencia de vídeo. Para transmitir la información del SRC a una velocidad de 256 Kbit/s, los valores de actividad se transmiten en todas las tramas. Cuando la velocidad se reduce a 80 Kbits/s, los valores de actividad se transmiten cada cuatro tramas.

FIGURA 25

Bloques con y sin transmisión del valor de actividad



BT.1885-25

4.2 Extremo de cliente

4.2.1 Cálculo del error cuadrático de los valores de actividad

- 1) Los píxeles de luminancia de una PVS se dividen en bloques de 16 x 16 píxeles en cada trama transcurrido un segundo desde la parte superior de la secuencia de vídeo.
- 2) En cada bloque, salvo en los del borde de trama, se calculan los valores de actividad (actividad PVS: $ActPVS_{i,j}$). A una velocidad de 256 Kbits/s de la información del SRC, los valores de actividad se calculan en todas las tramas. Cuando la velocidad de la información del SRC se reduce a 80 Kbits/s, los valores de actividad se calculan cada cuatro tramas.

$$ActPVS_{i,j} = \frac{1}{256} \sum_{k=0}^{255} |Y_k - \bar{Y}|$$

donde:

Y_k : es el valor de luminancia en un bloque de tamaño determinado de la PVS

\bar{Y} : es su valor medio

i : es un número de trama

j : es un número de bloque en la trama.

- 3) Los errores cuadráticos entre la actividad SRC y la actividad PVS se calculan como sigue:

$$E_{i,j} = (ActSRC_{i,j} - ActPVS_{i,j})^2$$

4.2.2 Ponderación psicovisual para el error cuadrático

Para tener en cuenta las características del sistema visual humano, se aplican al $E_{i,j}$ tres tipos de ponderación, a saber, ponderación para la diferencia en la frecuencia espacial, ponderación para la diferencia en la región de color específica y ponderación para la diferencia entre tramas de la luminancia.

1) Ponderación para la diferencia en la frecuencia espacial

Para esta ponderación se utiliza un factor de ponderación W_{SF} y un umbral Th_{SF} (véanse en el Cuadro 8 los valores de W_{SF} y Th_{SF}).

$$E_{i,j} \Leftarrow \begin{cases} E_{i,j} \times W_{SF}, & ActPVS_{i,j} > Th_{SF} \\ E_{i,j}, & \text{en cualquier otro caso} \end{cases}$$

2) Ponderación para la diferencia en la región de color específica

Para un bloque determinado y sus ocho bloques circundantes, si el número de píxels ($NumROIPixels$) en $48 \leq Y \leq 224$, $104 \leq Cb \leq 125$, y $135 \leq Cr \leq 171$ es mayor que un umbral, se lleva a cabo la siguiente ponderación utilizando un factor de ponderación W_{CR} y un umbral Th_{CR} .

$$E_{i,j} \Leftarrow \begin{cases} E_{i,j} \times W_{CR}, & NumROIPixels > Th_{CR} \\ E_{i,j}, & \text{en cualquier otro caso} \end{cases}$$

Véanse en el Cuadro 8 los valores de W_{CR} y Th_{CR} .

3) Ponderación para la diferencia entre tramas de la luminancia

Se calcula la diferencia absoluta media ($MAD_{i,j}$) de la luminancia entre un bloque determinado y el bloque de la trama precedente. $MAD_{i,j}$ se define como:

$$MAD_{i,j} = \frac{1}{256} \sum_{k=0}^{255} |Y_k - Y'_k|$$

donde Y_k es un valor de luminancia en un bloque de 16×16 píxels de la PVS e Y'_k es un valor de luminancia en la misma posición de la trama precedente.

La siguiente ponderación se realiza utilizando los factores de ponderación W_{MAD1} , W_{MAD2} y los umbrales Th_{MAD1} , Th_{MAD2} .

$$E_{i,j} \Leftarrow \begin{cases} E_{i,j} \times W_{MAD1}, & MAD_{i,j} > Th_{MAD1} \\ E_{i,j} \times W_{MAD2}, & MAD_{i,j} \leq Th_{MAD2} \\ E_{i,j}, & \text{en cualquier otro caso} \end{cases}$$

Véanse en el Cuadro 8 los valores de W_{MAD1} , W_{MAD2} , Th_{MAD1} y Th_{MAD2} .

4.2.3 Ponderación en el caso de detección de cambio de escena

Para cada trama se calcula el valor medio de $MAD_{i,j}$ (MAD_{Ave_i}) como sigue:

$$MAD_{Ave_i} = \frac{1}{M} \sum_{j=0}^{M-1} MAD_{i,j}$$

siendo M el número de bloques en una trama.

Si MAD_{Ave_i} es mayor que un umbral Th_{SC} , se considera como un cambio de escena. Cuando se detecta un cambio de escena, $E_{i,j}$ se ajusta a 0 durante 15 tramas y tras el cambio de escena.

$$SceneChange = \begin{cases} TRUE, & MAD_{Ave} > Th_{SC} \\ FALSE & \text{en cualquier otro caso} \end{cases}$$

$$E_{i,j} \Leftarrow \begin{cases} E_{i,j} \times W_{SC} & 15 \text{ tramas tras } SceneChange = TRUE \\ E_{i,j}, & \text{en cualquier otro caso} \end{cases}$$

Véanse en el Cuadro 8 los valores de W_{SC} y Th_{SC} .

4.2.4 PSNR basada en el error cuadrático de la actividad

Se calcula una PSNR basándose en la diferencia de actividad como:

$$VQ = 10 \times \log_{10} \frac{255 \times 255}{E_{Ave}}$$

$$E_{Ave} = \frac{1}{N \times M} \sum_{i=0}^{N-1} \sum_{j=0}^{M-1} E_{i,j}$$

donde N y M son el número de tramas y de bloques utilizados para el cálculo de la PSNR.

4.2.5 Ponderación para los efectos auxiliares del bloqueo

Para esta ponderación se utiliza un factor de ponderación W_{BL} , un umbral Th_{BL} , y la información de nivel de bloqueo BL_{Ave} . (Véanse en el Cuadro 8 los valores de W_{BL} Y Th_{BL} .)

$$VQ \Leftarrow \begin{cases} VQ \times W_{BL}, & BL_{Ave} > Th_{BL} \\ VQ, & \text{en cualquier otro caso} \end{cases}$$

BL_{Ave} se calcula mediante los siguientes pasos:

Paso 1: Se determinan los valores de actividad para bloques de 8×8 píxels. Como muestra la Fig. 26, el valor medio (Act_{Ave}) de los dos valores de actividad en bloques adyacentes ($ActBlock_1$, $ActBlock_2$) se calcula como sigue:

$$Act_{Ave} = \frac{1}{2} (ActBlock_1 + ActBlock_2)$$

Paso 2: Se calcula la diferencia absoluta de los valores de luminancia a lo largo del límite entre los dos bloques. Como ilustra la Fig. 26, $Y_{1,0}$ y $Y_{2,0}$ representan los valores de luminancia en los bloques derecho e izquierdo a lo largo del límite. Un valor medio de la diferencia de luminancia absoluta, $DiffBound$, se expresa como sigue:

$$DiffBound = \frac{1}{8} \sum_{i=0}^7 |Y_{1,i} - Y_{2,i}|$$

Paso 3: El nivel de bloqueo ($BL_{i,j}$) se define mediante la relación entre $DiffBound$ y Act_{Ave} , es decir:

$$BL_{i,j} = \frac{DiffBound}{Act_{Ave} + 1}$$

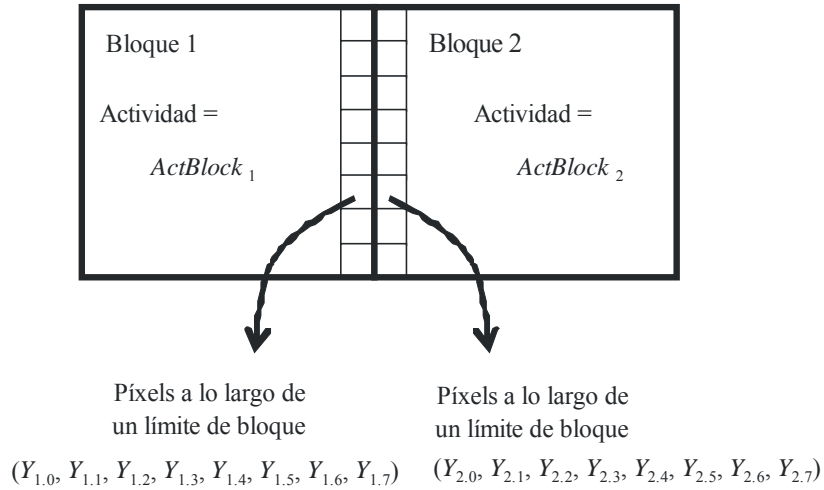
Paso 4: El valor medio de BL se calcula así:

$$BL_{Ave} = \frac{1}{N \times M} \sum_{i=0}^{N-1} \sum_{j=0}^{M-1} BL_{i,j}$$

Para los bloques situados en el extremo derecho de las tramas, el valor de $BL_{i,j}$ se fija a cero. Si BL_{Ave} es mayor que un umbral predeterminado, se considera que la secuencia de vídeo incluye un elevado nivel de bloqueo y se aplica ponderación al valor de la calidad de vídeo calculada.

FIGURA 26

Píxel y valores de actividad utilizados para calcular el nivel de bloqueo



BT.1885-26

4.2.6 Ponderación para los efectos auxiliares de las degradaciones locales

Para esta ponderación se utiliza un factor de ponderación W_{LI} , un umbral Th_{LI} , y la degradación local LI . (Véanse los valores de W_{LI} y Th_{LI} en el Cuadro 8.)

$$VQ \leftarrow \begin{cases} VQ \times W_{LI}, & LI < Th_{LI} \\ VQ & \text{en cualquier otro caso} \end{cases}$$

LI se calcula mediante los siguientes pasos. La diferencia de la varianza de la actividad se utiliza para detectar la degradación local debida a los errores de transmisión.

- 1) Para un bloque determinado y sus ocho bloques circundantes, la varianza de la actividad se calcula para el SRC ($ActVar_{SRC}$) y la PVS ($ActVar_{PVS}$) y la diferencia absoluta de estos valores de la varianza se determina como sigue:

$$\Delta ActVar = |ActVar_{SRC} - ActVar_{PVS}|$$

- 2) Se calcula para cada trama el valor medio de estos valores de diferencia absoluta.
- 3) LI se calcula como la relación entre los valores máximo ($\Delta ActVar_{Máx}$) y mínimo ($\Delta ActVar_{Mín}$) de la media.

$$LI = \begin{cases} \Delta ActVar_{Mín} / \Delta ActVar_{Máx} & \Delta ActVar_{Máx} \neq 0 \\ 1 & \Delta ActVar_{Máx} = 0 \end{cases}$$

VQ representa la nota de calidad de vídeo.

4.2.7 Parámetros para las ponderaciones

El Cuadro 8 muestra los valores de los parámetros para las ponderaciones. Estos valores se determinan mediante un experimento preliminar con el conjunto de datos de prueba.

CUADRO 8

Parámetros para las ponderaciones

| Tipo de operación de ponderación | Valores de los parámetros | |
|-------------------------------------------|---------------------------|-------|
| Ponderación de la frecuencia espacial | W_{SF} | 0,36 |
| | Th_{SF} | 25 |
| Ponderación de un color específico | W_{CR} | 4,0 |
| | Th_{CR} | 175 |
| Ponderación de la diferencia entre tramas | W_{MAD1} | 0,06 |
| | Th_{MAD1} | 17 |
| | W_{MAD2} | 25 |
| | Th_{MAD2} | 13 |
| Detección de cambio de escena | W_{SC} | 0,0 |
| | Th_{SC} | 35 |
| Ponderación del bloqueo | W_{BL} | 0,870 |
| | Th_{BL} | 1,0 |
| Ponderación de la degradación local | W_{LI} | 0,870 |
| | Th_{LI} | 1,67 |

4.2.8 Registro

1) Registro espacial

El modelo RR no necesita ningún registro espacial. Ello se debe a que el error cuadrático se calcula basándose en los valores de actividad que son más robustos frente a las derivas espaciales que los basados en los valores de los píxels.

2) Registro de ganancia y desplazamiento

El modelo RR no necesita registro de ganancia y desplazamiento. Los valores de actividad son inherentemente libres de desplazamiento (es decir, componentes de cc) e insensibles a la ganancia.

3) Registro temporal

La secuencia PVS se divide en subsecuencias de 1 segundo. Para cada subsecuencia, se calculan los errores cuadráticos medios de la actividad con cinco variaciones de retardo de hasta ± 2 SRC tramas. El mínimo valor de estos errores cuadráticos medios se utiliza finalmente como el error cuadrático medio en esta subsecuencia. El retardo resultante de este mínimo error cuadrático medio se ajusta como registro temporal.

5 Códigos de muestra

A continuación aparecen los códigos de muestra en lenguaje C para el modelo RR.

5.1 Código común para los extremos de servidor y cliente

```
// Calculate the activity value
unsigned int CalcActivitybyRect(unsigned char * lpBuffer, int nWidth, int iRectWidth, int iRectHeight)
{
    // lpBuffer: Luminance Frame Buffer
    // nWidth: Frame Buffer Width
    // iRectWidth: Width of the rectangle to calculate an activity value.
    // iHeightWidth: Height of the rectangle to calculate an activity value.
    unsigned int i, j, nTmp, nSum;
    unsigned char *pSrc;
    pSrc = lpBuffer; nSum = 0;
    for (j = 0; j < iRectHeight; j++){
        for (i = 0; i < iRectWidth; i++){
            nSum += pSrc[i];
        }
        pSrc += nWidth;
    }
    nSum /= (iRectWidth*iRectHeight);

    pSrc = lpBuffer; nTmp = 0;
    for (j = 0; j < iRectHeight; j++){
        for (i = 0; i < iRectWidth; i++){
            nTmp += abs(pSrc[i] - nSum);
        }
        pSrc += nWidth;
    }
    return nTmp/iRectWidth/iRectHeight;
}
```

5.2 Extremo servidor

```
// Server side
int nStart = 30; // the frame number to start transmission (30 or 25)
int nMaxFrame = 240; // the number of total video frames (240 or 200)
int nFrameIncrement = 1; // 1 for 256kbps, 4 for 80kbps
void ReadOneFrame(unsigned char, int, unsigned char *, int, int); // function to read one frame data
int nRim = 16 // 16 or 32 (use 32 to avoid the problem in HRC9)
// nWidth: Frame Buffer Width
// nHeight: Frame Buffer Height
// lpSrc: Frame Buffer
for(int nFrame = nStart; nFrame < nMaxFrame; nFrame+=nFrameIncrement){
    ReadOneFrame(SRC_file_name, nFrame, lpSrc, nWidth, nHeight);
    for (j= 16; j<nHeight-32; j+=16) {
        for (i= nRim; i<nWidth- nRim; i+=16) {
            lpOrg = lpSrc + i + j * nWidth;
            nActSrc = CalcActivitybyRect(lpOrg, nWidth, 16, 16);
            // OutputSRCInfo(nActSrc); // Output or transmission the SRC information
        }
    }
}
```

5.3 Extremo cliente

```
// Client Side
int nStart = 30; // the frame number to start transmission (30 or 25)
int nMaxFrame = 240; // the number of total video frames (240 or 200)
int nFrameIncrement = 1; // 1 for 256kbps, 4 for 80kbps
int nFrameRate = 30; //30 or 25
void ReadOneFrame(unsigned char, int, unsigned char **, int, int); // function to read one frame data
void ReadRRData(unsigned char, int, unsigned char *); // function to read RR-data
// nWidth: Frame Buffer Width
// nHeight: Frame Buffer Height
// lpPvsByte[3]: Frame Buffer (0:Y, 1:Cb, 2:Cr)
// lpRRData: RR-data Buffer
// double ddActivityDifference[][]: Store the activity-difference
// double ddActivityVariance[][]: Store the activity-variance
// double ddBlock[][]: Store the blockiness level
// int nSceneChange: Scene change detection
```

```

for(int nTemporalAlign = -2; nTemporalAlign <=2; nTemporalAlign++){ // Changing temporal alignment
    for(int nFrame = 0; nFrame < nMaxFrames; nFrame++){
        if(nFrame+nTemporalAlign >= nMaxFrames || nFrame+nTemporalAlign < 0){
            continue;
        }
        ReadOneFrame(PVS_file_name, nFrame+nTemporalAlign, lpPvsByte, nWidth, nHeight);
        if(((nFrame-(nFrameRate+nStart)) % nFrameIncrement) == 0
            && nFrame >= nStart ){
            ReadRRData(RR_file_name, nFrame, lpRRData);
            ddActivityDifference[nTemporalAlign+2][nFrame]
                = RRCalcObjectiveScore(lpPvsByte, lpRRData, nWidth, nHeight);
            ddActivityVariance[nTemporalAlign+2][nFrame] = gnActVar;
        }else{
            ddActivityDifference[nTemporalAlign+2][nFrame] = 0.0;
            ddActivityVariance[nTemporalAlign+2][nFrame] = 0.0;
        }
        // Blockiness Level
        if(nTemporalAlign ==0){
            ddBlock[nFrame] = BlockinessLevelEstimation(lpPvsByte[0], nWidth, nHeight);
        }
        // Pixel copy for inter-frame difference calculation
        memcpy(lpPrev, lpPvsByte[0], sizeof(char)*nWidth*nHeight);
        if(nSceneChange){
            nSceneChange--;
        }
    }
}

double ddSum[8][5]; // Sum of the Activity-difference for each second
double ddActVarSum[8][5]; // Sum of the Activity-variance for each second
double ddActVarMax[8][5]; // Maximum of the Sum of the Activity-variance
double ddActVarMin[8][5]; // Minimum of the Sum of the Activity-variance
int nnMin[8];
int nnNumFrames[8][5];
#define LARGENUMBER 100000
for(int nTemporalAlign = -2; nTemporalAlign <=2; nTemporalAlign++){
    for(int j=0;j<8;j++){ // for each one second
        nnNumFrames[j][nTemporalAlign+2] = 0;
        ddActVarMax[j][nTemporalAlign+2] = 0.0;
        ddActVarMin[j][nTemporalAlign+2] = LARGENUMBER;
        ddActVarSum[j][nTemporalAlign+2] = 0.0;
        ddSum[j][nTemporalAlign+2] = 0.0;
        for(int i=nFrameRate*;i< (j+1)*nFrameRate; i++){
            if(ddActivityDifference[nTemporalAlign+2][i]){
                ddSum[j][nTemporalAlign+2] += ddActivityDifference[nTemporalAlign+2][i];
                nnNumFrames[j][nTemporalAlign+2]++;
            }
            ddActVarSum[j][nTemporalAlign+2] += ddActivityVariance[nTemporalAlign+2][i];
            if(ddActivityVariance[nTemporalAlign+2][i]){
                if(ddActivityVariance[nTemporalAlign+2][i] >
                    ddActVarMax[j][nTemporalAlign+2]){
                    ddActVarMax[j][nTemporalAlign+2] =
                        ddActivityVariance[nTemporalAlign+2][i];
                }

                if(ddActivityVariance[nTemporalAlign+2][i] <
                    ddActVarMin[j][nTemporalAlign+2]){
                    ddActVarMin[j][nTemporalAlign+2] =
                        ddActivityVariance[nTemporalAlign+2][i];
                }
            }
        }
    }
}

// Local Impariment Level Calculation
double dSum = 0.0;
double dActMax = 0.0;
double dActMin = LARGENUMBER;
int nNumFrames = 0;

```

```

for(int j=1; j<8; j++){
    double dMin = LARGENUMBER;
    double dMinSum = LARGENUMBER;
    for(int nTemporalAlign = -2; nTemporalAlign <=2; nTemporalAlign++){
        if(ddActVarSum[j][nTemporalAlign+2] < dMin){
            dMin = ddActVarSum[j][nTemporalAlign+2];
            dMinSum = ddSum[j][nTemporalAlign+2];
            nnMin[j] = nTemporalAlign+2;
        }
    }
    dSum += dMinSum;
    nNumFrames += nnNumFrames[j][nnMin[j]];
    if(ddActVarMax[j][nnMin[j]] > dActMax){
        dActMax = ddActVarMax[j][nnMin[j]];
    }
    if(ddActVarMin[j][nnMin[j]] < dActMin){
        dActMin = ddActVarMin[j][nnMin[j]];
    }
}
double dTransError = dActMax/dActMin;
// Blockiness Level Calculation
double dBlockinessLevel = 0.0;
for(int i=0; i<nMaxFrames; i++){
    dBlockinessLevel += ddBlock[i];
}
dBlockinessLevel = dSumBlock / (double)(nMaxFrames-nFrameRate);
// Calculating the Video Quality Score
if(nNumFrames && nNumberOfBlocks && dSum){
    dSum = dSum / (double)(nNumFrames)/(double)nNumberOfBlocks;
    dSum = 10*log10(255.0*255.0/dSum); //PSNR based on the activity difference
    if(dBlockinessLevel > dBlokinessTh){
        dSum /= nBlockinessWeighting; // Weighting for blockiness level
    }
    if(dTransError > nErrorTh){
        dSum /= nErrorWeightin; // Weighting for transmission errors
    }
}
}
return dSum;
}
-----
// Calculating the MAD value
unsigned int CalcMAD(unsigned char *lpSrc, unsigned char *lpSrc2, int nWidth, int nHeight)
{
    // lpSrc: Frame Buffer of the current frame
    // lpSrc2: Frame Buffer of the previous frame
    unsigned int nSum = 0;
    for (y = 0; y < nHeight; y++) {
        for (x = 0; x < nWidth; x++) {
            nSrc = lpSrc[x + y*nWidth];
            nSrc2 = lpSrc2[x + y*nWidth];
            nSum += abs(nSrc - nSrc2);
        }
    }
    return nSum/nWidth/nHeight;
}
// Calculating a mean squared error with weightings
double RRCalcObjectiveScore(unsigned char *lpBuffer[], unsigned char *lpRRData, int nWidth, int nHeight)
{
    int i, j, nActSrc, nActSrc2, nY, nCb, nCr, nYMin, nYMax, nCbMin, nCbMax, nCrMin, nCrMax;
    int nMBX, nMBY, nMB, nStart;
    unsigned int nMAD;
    double e2, dMADFrame;
    unsigned char *lpRec, *lpRecCb, *lpRecCr, *lpPrev1;
    int nRim = 16 // 16 or 32 (use 32 to avoid the problem in HRC9)
    nYMin = 48; nYMax = 224; nCbMin = 104; nCbMax = 125; nCrMin = 135; nCrMax = 171;
    nMB = nMBY = nStart = 0;
    e2 = dMADFrame = 0.0;
    for (j=16+nStart; j<ilimageHeight-32; j+=16) {
        nMBX = 0;

```

```

for (i= nRim; i<nWidth- nRim; i+=16) {
    lpRec = lpBuffer[0] + i + j * nWidth;
    lpRecCb = lpBuffer[1] + i/2 + (j/2) * nWidth/2;
    lpRecCr = lpBuffer[2] + i/2 + (j/2) * nWidth/2;
    lpPrev1 = lpPrev + i + j * nWidth;
    nActSrc = lpRRData[nMB]; // SRC activity
    nActSrc2 = CalcActivitybyRect(lpRec, nWidth, 0, 16, 16); // PVS activity
    nActArray[nMB] = (double)nActSrc;
    nActArray2[nMB] = (double)nActSrc2;
    e2 += (double)(nActSrc - nActSrc2)*(nActSrc - nActSrc2); // Mean squared error
    nMAD = CalcMAD(lpRec, lpPrev1, 16, 16, nWidth); // Inter-frame differenece
    dMADFrame += (double)nMAD;
    int nNumROIPixels=0;
    for(int jj=-16;jj<32; jj++){
        for(int ii=-16;ii<32; ii++){
            nY = lpRec[ii];
            nCb = lpRecCb[ii/2];
            nCr = lpRecCr[ii/2];
            if(nY >= nYMin && nY <= nYMax
            && nCb >= nCbMin && nCb <= nCbMax
            && nCr >= nCrMin && nCr <= nCrMax){
                nNumROIPixels++;
            }
        }
        lpRec += nWidth;
        if((jj & 1) == 1){
            lpRecCb += nWidth/2;
            lpRecCr += nWidth/2;
        }
    }

    // Weighting for spatial frequency
    if(nActSrc2 > gdwActThHigh){
        e2 *= dActWeightingHigh;
    }else if(nActSrc2 > gdwActThLow){
        e2 *= dActWeightingMiddle;
    }else {
        e2 *= dActWeightingLow;
    }
    // Weighting for specific color region
    if(nNumROIPixels > dwROITh){
        e2 *= dROIWeighting;
    }

    // Weighting for inter-frame difference
    if(nMAD > dwMADThHigh){
        e2 *= dMADWeightingHigh;
    }else if(nMAD > dwMADThLow){
        e2 *= dMADWeightingMiddle;
    }else {
        e2 *= dMADWeightingLow;
    }
    nMB++;      nMBX++;
}
nMBY++;
}

// Calculating Activity-Variance for Surrounding Nine Blocks.
double nSumActSrc, nSumActPvs, nActVar, nActVar2;
nSumActSrc = nSumActPvs = nActVar = nActVar2 = 0.0;
gnActVar = 0.0;
for (j=1; j<nMBY-1; j++) {
    for (i=1; i<nMBX-1; i++) {
        nSumActSrc = 0.0;
        nSumActPvs = 0.0;
        for(int jj=-1; jj<2; jj++){
            for(int ii=-1; ii<2; ii++){
                nSumActSrc += nActArray[i+ii+nMBX*(j+jj)];
                nSumActPvs += nActArray2[i+ii+nMBX*(j+jj)];
            }
        }
    }
}

```

```

        nSumActSrc /= 9.0;
        nSumActPvs /= 9.0;
        nActVar = 0.0;
        nActVar2 = 0.0;
        for(int jj=-1; jj<2; jj++){
            for(int ii=-1; ii<2; ii++){
                nActVar += (nActArray[i+ii+nMBX*(j+jj)]-nSumActSrc)*(nActArray[i+ii+nMBX*(j+jj)]-nSumActSrc);
                nActVar2 += (nActArray2[i+ii+nMBX*(j+jj)]-
                    nSumActPvs)*(nActArray2[i+ii+nMBX*(j+jj)]-nSumActPvs);
            }
        }
        nActVar /= 9.0;
        nActVar2 /= 9.0;
        gnActVar += abs(nActVar- nActVar2);
    }
}
// Average of the Activity-Variance for the Frame
gnActVar = gnActVar/(double)(nMBY-2)/(double)(nMBY-2);
// Scene change detection
if(dMADFrame/ nMB > 35){
    nSceneChange = 15;
}
if(nSceneChange){
    e2 = 0.0;
}
gnFrame++;
return e2;
}
// Calculate Blockiness Level
double BlockinessLevelEstimation(unsigned char *lpBuffer, int nWidth, int nHeight)
{
    // lpBuffer: Frame Buffer
    int i, j, nActSrc, nActSrc2, nDiff, nMB;
    unsigned char *lpRec = lpBuffer;
    double dBlock=0.0;
    nMB = 0;
    for (j=0; j<nHeight-16; j+=8) {
        for (i=0; i<nWidth-16; i+=8) {
            lpRec = lpBuffer + i + j * nWidth;
            nActSrc = CalcActivitybyRect(lpRec, nWidth, 0, 8, 8); // Activity of the left block
            nActSrc2 = CalcActivitybyRect(lpRec+8, nWidth, 0, 8, 8); // Activity of the right block
            nActSrc = (nActSrc + nActSrc2)/2; // Average of the activity values
            nDiff = 0;
            for(int jj=0;jj<8; jj++){
                nDiff += abs(lpRec[7+jj*nWidth] - lpRec[8+jj*nWidth]);
                // Difference of the luminance values at the boundary
            }
            nDiff/= 8;
            dBlock += (double)nDiff/(double)(nActSrc+1);
            nMB++;
        }
    }
    return (double)dBlock/(double)nMB;
}
}

```

6 Referencias informativas

- [1] SMPTE 170M, «SMPTE Standard for Television – Composite Analog Video Signal – NTSC for Studio Applications», Society of Motion Picture and Television Engineers.
- [2] Recomendación UIT-T J.143 – Requisitos de usuario para mediciones objetivas de la percepción de la calidad vídeo en televisión digital por cable.

Anexo C

Modelo C: Método de referencia reducida NTIA

1 Antecedentes

En el periodo de tiempo 2003-2004 la National Telecommunications and Information Administration (NTIA) de EE.UU. desarrolló dos modelos de calidad de vídeo (VQM) con una anchura de banda de referencia reducida (RR) de aproximadamente 12 a 14 Kbits/s para el vídeo muestreado de la Recomendación UIT-R BT.601 [1]. Estos modelos se denominaron «VQM de anchura de banda baja» y «VQM de anchura de banda baja rápido». Este último era una versión informativamente eficaz del VQM de anchura de banda baja. El VQM de anchura de banda baja rápido es aproximadamente cuatro veces más rápido ya que extrae las características espaciales de las tramas de vídeo que se han promediado previamente en vez de extraer las características espaciales directamente de las tramas de vídeo UIT-R BT. 601. Se logran ahorros de cálculo adicionales para el VQM de anchura de banda baja rápido determinando las características de información temporal (es decir, el movimiento) mediante un submuestreo aleatorio de los píxels en el canal de luminancia Y en vez de utilizar todos los píxels en los tres canales de vídeo (Y, Cb, y Cr). Ambos VQM han estado disponibles en nuestras herramientas informáticas VQM durante varios años y pueden utilizarse gratuitamente para aplicaciones comerciales y no comerciales. Pueden descargarse las versiones ejecutables binarias de estas herramientas VQM y sus códigos de fuente asociados [2].

Como la NTIA deseó presentar ambos VQM a las pruebas de TV de referencia reducida (RRTV) para una evaluación independiente por el Grupo de Expertos en calidad del vídeo (VQEG), escogió presentarlos con diferentes categorías de velocidades binarias aun cuando tengan idénticos requisitos de velocidad binaria RR. La NTIA decidió presentar el VQM de anchura de banda baja a la categoría de 256 k y el VQM de anchura de banda baja rápido a la categoría de 80 k ya que cabía esperar que el comportamiento del primero de ellos fuese superior al del segundo. Ambos VQM utilizaron el algoritmo de calibración RR de la NTIA que aparece en la Recomendación UIT-T J.244[3]. Este algoritmo de calibración exige aproximadamente de 22 a 24 Kbits/s de anchura de banda RR a fin de producir estimaciones del retardo temporal, la deriva espacial, la escala espacial, la ganancia y la desviación de nivel.

Un interesante resultado de las pruebas de evaluación VQEG RRTV [4] fue que el VQM de anchura de banda baja rápido se comportó mejor que el VQM de anchura de banda baja en las pruebas de 525 líneas y 625 líneas. Es un resultado interesante puesto que demuestra que es mejor extraer las características espaciales de las tramas promediadas que extraerlas de las tramas no promediadas. Debe investigarse si este resultado sigue siendo cierto o no para otros conjuntos de datos. Actualmente, la NTIA no ve razón alguna para normalizar ambos modelos por lo que este Anexo solo describe el VQM de anchura de banda baja rápido.

2 Introducción

Este Anexo presenta una descripción y un código de referencia para el VQM de anchura de banda baja rápido NTIA. Este modelo utiliza técnicas similares a los del VQM general NTIA, cuya descripción figura en las Recomendaciones UIT-T J.44 [5] y UIT BT.1683 [6]. El VQM de anchura de banda baja rápido utiliza las características RR con una anchura de banda muy inferior que el VQM general NITA, pero la extracción de las características y el proceso de comparación es similar en ambos. Para el vídeo muestreado de la Recomendación UIT-R BT.601 [1] el VQM de anchura de banda baja rápido emplea unas características RR que requieren aproximadamente de 12 a 14 Kbits/s de anchura de banda de transmisión. Este Anexo describe sólo el VQM de anchura de

banda baja rápido ya que sus algoritmos de calibración de anchura de banda baja complementarios están plenamente documentados en la Recomendación UIT-T J.244 [3]. No obstante, para completar la información, el código de referencia en este Anexo incluye el algoritmo tanto del VQM de anchura de banda baja rápido como el de calibración de anchura de banda baja asociado. El código de referencia también contiene ejemplos de funciones de cuantificación para las características utilizadas por la calibración de anchura de banda baja (estas funciones de cuantificación no forman parte de la Recomendación UIT-T J.244).

3 Descripción del VQM de anchura de banda baja rápido

3.1 Consideraciones generales sobre el VQM

La descripción del VQM engloba tres áreas primarias:

- 1) las características de anchura de banda baja extraídas de los trenes de vídeo original y procesado;
- 2) los parámetros resultantes de la comparación de las características de los trenes original y procesado;
- 3) el cálculo del VQM que combina varios parámetros cada uno de los cuales mide un aspecto distinto de la calidad de vídeo.

Esta descripción utiliza referencias fácilmente disponibles relativas a los detalles técnicos.

3.2 Descripción de las características

3.2.1 Consideraciones generales de las características

El VQM de anchura de banda baja utiliza tres tipos de características: color, espacial y temporal. Cada uno de estos tipos de características cuantifica las distorsiones perceptuales en sus respectivos dominios. La subrutina del código de referencia «`model_fastlowbw_features`» proporciona una descripción matemática completa de las características empleadas por el VQM de anchura de banda baja rápido.

3.2.2 Características de color

Las características de color son las mismas características f_{COHER_COLOR} utilizadas por el VQM General de la NTIA. Estas características se describen detalladamente en el Anexo D.7.3 a la Recomendación UIT-T J.144. Las características f_{COHER_COLOR} proporcionan una medición de vector bidimensional de la cantidad de información de crominancia azul y roja (C_B , C_R) en cada región S-T. Por consiguiente, las características f_{COHER_COLOR} son sensibles a las distorsiones de color. Las características f_{COHER_COLOR} del VQM de anchura de banda baja rápido de la NTIA se extraen de la región espacial-temporal (S-T) de un tamaño de 30 líneas verticales por 30 píxels horizontales en 1 segundo de tiempo (es decir, $30 \times 30 \times 1s$) mientras que el VQM general de la NTIA utiliza la región S-T de tamaño $8 \times 8 \times 1$.

3.2.3 Características espaciales

Las características espaciales son las mismas características f_{SI13} y f_{HV13} que se utilizan en el VQM General de la NTIA. Estas características se describen detalladamente en el Anexo D.7.2.2 a la Recomendación UIT-T J.144. Las características f_{SI13} y f_{HV13} miden el volumen y la distribución angular de los gradientes espaciales en las subregiones espacial-temporal (S-T) de las imágenes de luminancia (Y). Por consiguiente, las características f_{SI13} y f_{HV13} son sensibles a las distorsiones espaciales tales como la borrosidad y el bloqueo.

Las características f_{SI13} y f_{HV13} del VQM de anchura de banda baja rápido de la NTIA se extraen de la región temporal-espacial (S-T) de un tamaño de 30 líneas verticales por 30 píxels horizontales en 1 segundo de tiempo (es decir, $30 \times 30 \times 1s$) mientras que el VQM General de la NTIA utiliza la región S-T de tamaño $8 \times 8 \times 0,2s$. Además, para ahorrarse algunos cálculos, se obtiene en primer lugar previamente la media de los periodos de 1 segundo de las imágenes de luminancia Y a lo largo del tiempo antes de aplicar los filtros de realce del borde de 13×13 bidimensionales indicados en el Anexo D.7.2.1 a la Recomendación UIT-T J.144.

Una característica espacial adicional se extrae de los periodos de 1 s de las imágenes de luminancia (Y) previamente promediadas. Esta característica es el nivel de luminancia *media* (Y) de cada región S-T de $30 \times 30 \times 1s$ (aquí denominada f_{MEAN}). El objeto de la característica f_{MEAN} es proporcionar una función de ponderación perceptual del nivel de luminancia para las pérdidas de información espacial (SI) medidas por las características f_{SI13} y f_{HV13} . Esto se descubrirá en el punto sobre descripción de los parámetros.

3.2.4 Características temporales

Pueden obtenerse unas estimaciones bastante precisas de la calidad de vídeo percibida a partir del conjunto de características de color y espaciales descrito anteriormente. Sin embargo, como las regiones S-T de las que se extraen estas características abarcan muchas tramas de vídeo (es decir, periodos de 1s de tramas de vídeo), tienden a ser insensibles a las breves perturbaciones temporales que aparecen en la imagen. Tales perturbaciones pueden proceder de ruido o de errores de transmisión digital; y, si bien son breves por naturaleza, pueden tener repercusiones significativas sobre la calidad de imagen percibida. Por tanto, se utiliza una característica RR temporal para cuantificar los efectos perceptuales de las perturbaciones temporales. Esta característica mide la información temporal absoluta (ATI), o movimiento, en el plano de imagen de luminancia Y y se calcula como sigue:

$$f_{ATI} = rms\{rand5\%|Y(t) - Y(t - 0.2s)|\}$$

Para lograr mayor eficacia en los cálculos, Y se submuestra aleatoriamente para que contenga sólo el 5% de píxels de imagen (representado aquí por la función $rand5\%$). La imagen Y submuestreada en el instante $t-0,2s$ se sustrae de la imagen Y idénticamente submuestreada en el instante t y el error cuadrático medio (rms) del resultado se utiliza como medida de la ATI. Empleando el convenio indicado en el Anexo D.8 a la Recomendación UIT-T J.144 esto también puede representarse como sigue:

$$f_{ATI} \cong Y_rand5\%_ati0.2s_rms$$

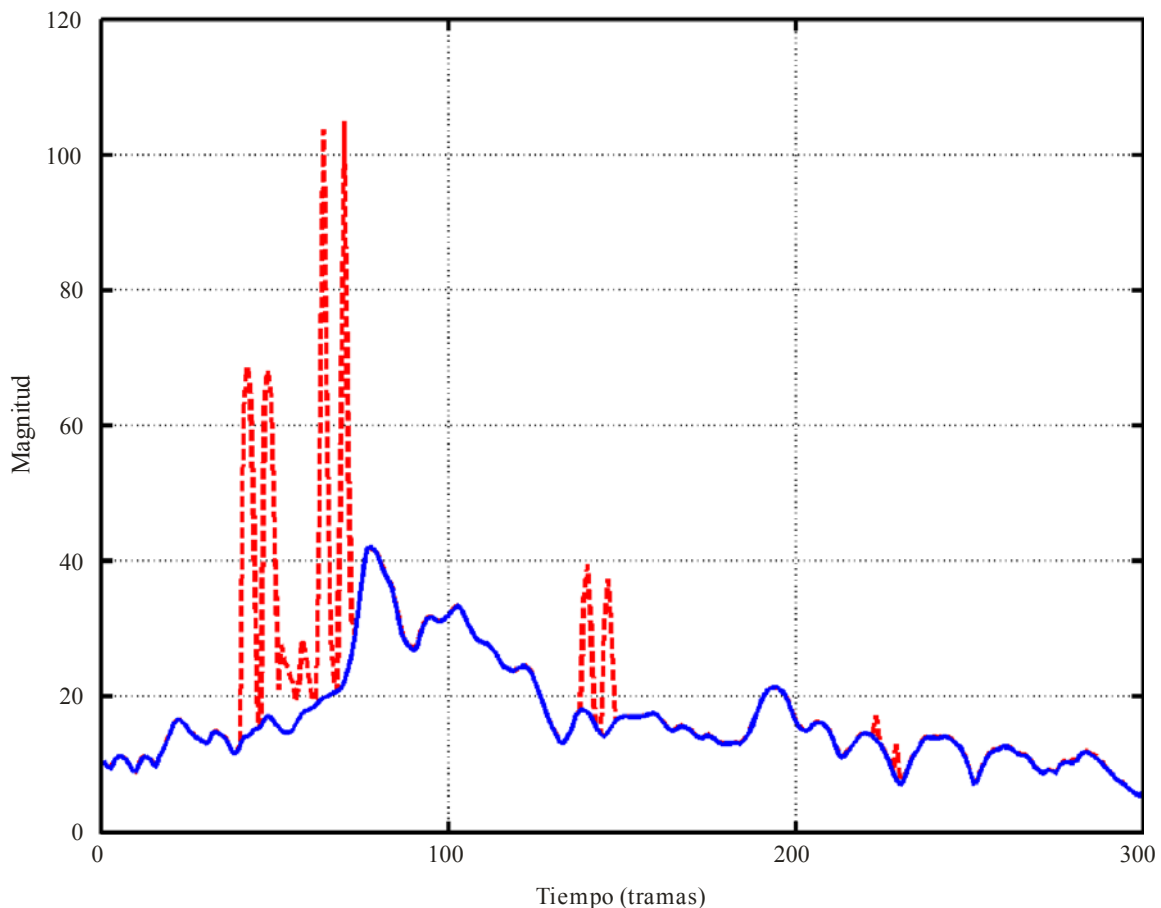
La característica f_{ATI} es sensible a las perturbaciones temporales. Para un vídeo de 30 tramas/s, 0,2s representan 6 tramas de vídeo mientras que para un vídeo de 25 tramas/s, 0,2s representan 5 tramas de vídeo. La sustracción de las imágenes que se encuentran separadas 0,2s hace que la característica sea insensible a los sistemas de vídeo en tiempo real de 30 tramas/s y 25 tramas/s que tienen velocidades de actualización de trama de al menos 5 tramas/s. Los aspectos de calidad de estos sistemas de vídeo de velocidad de trama baja, muy comunes en aplicaciones multimedia, son suficientemente detectados por las características f_{SI13} , f_{HV13} , y f_{COHER_COLOR} . La separación de 0,2s también se adapta mejor a la respuesta de cresta temporal del sistema visual humano que la diferenciación de dos imágenes que están separadas una trama en el tiempo.

La Fig. 27 proporciona un ejemplo gráfico de la característica f_{ATI} para una escena de vídeo original (línea azul continua) y procesado (línea roja de puntos) de un sistema de vídeo digital con errores de ráfaga transitorios en el canal de transmisión digital. Los errores transitorios en la imagen procesada crean crestas en la característica f_{ATI} . La anchura de banda requerida para transmitir la característica f_{ATI} es extremadamente baja puesto que requiere sólo 30 muestras por segundo para

vídeos de 30 tramas/s. Otros tipos de ruido aditivo en el vídeo procesado, como pueda ser el generado por un sistema de vídeo analógico, aparecerán como una deriva de cc positiva en la historia temporal del tren de características procesado con respecto al tren de características original. Los sistemas de codificación de vídeo que eliminan el ruido provocarán una deriva de cc negativa.

Antes de extraer un parámetro de error transitorio del tren de característica f_{ATI} mostrado en la Fig. 27, es conveniente aumentar la anchura de las crestas de movimiento (crestas rojas en la Fig. 27). El motivo es que estas breves crestas de movimiento causadas por los errores transitorios no representan adecuadamente la repercusión perceptual de estos tipos de errores. Un método para incrementar la anchura de las crestas de movimiento consiste en aplicar un máximo filtrado a los trenes de características original y procesado antes de calcular la función del parámetro de error entre las dos formas de onda. Para el parámetro de error basado en f_{ATI} , se utilizó un filtro máximo de 7 puntos de anchura (aquí denominado función *max7pt* que produjo una muestra de salida en cada trama que es el máximo entre el propio valor y los 3 vecinos más próximos a cada lado (es decir, las muestras en los instantes anterior y posterior).

FIGURA 27

Ejemplo de historia temporal de la característica f_{ATI} 

BT.1885-27

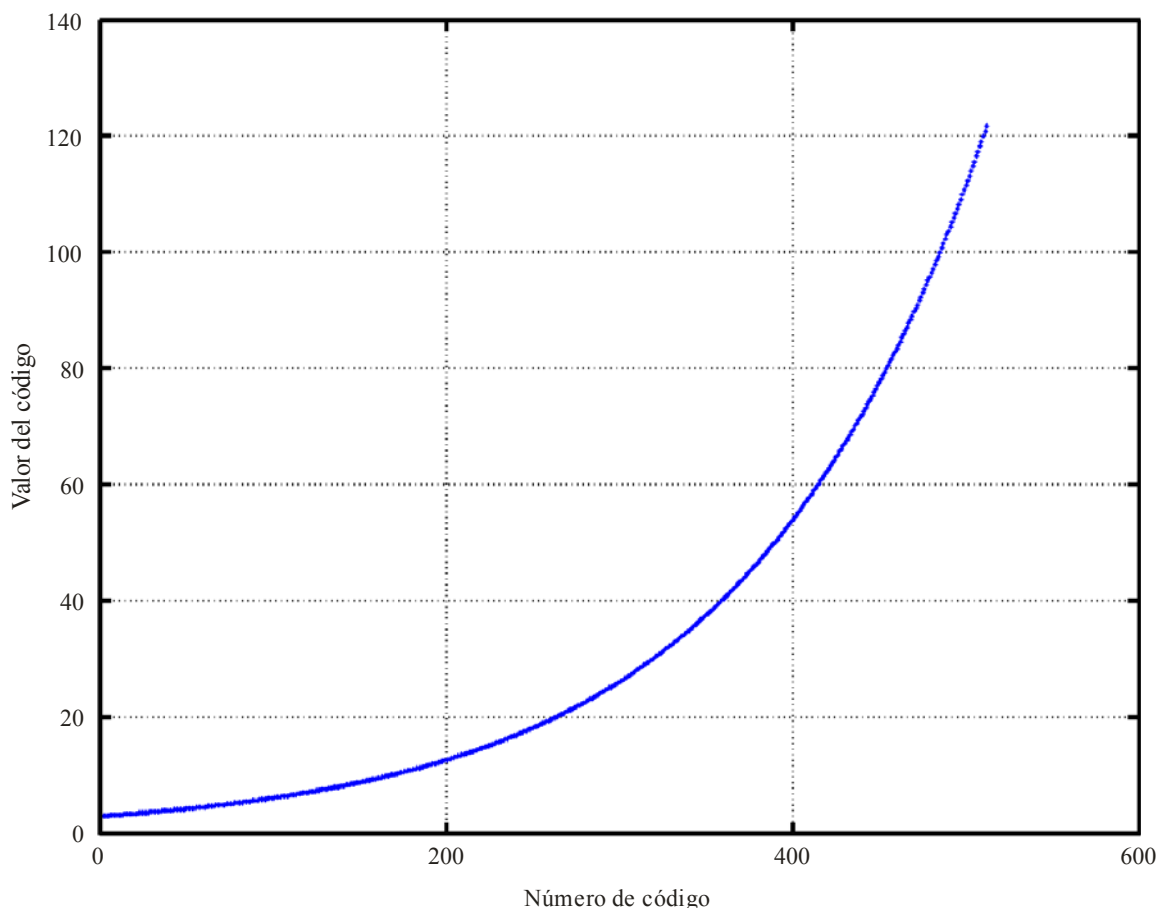
3.2.5 Cuantificación de las características

La cuantificación con 9 bits de precisión es suficiente para las características Y_{MEAN} , f_{SI13} , f_{HV13} , y f_{COHER_COLOR} mientras que la característica f_{ATI} debe cuantificarse con 10 bits. Para que el efecto sea mínimo en el cálculo de los parámetros de la calidad de vídeo, debe utilizarse un diseño de

cuantificado no lineal en el que el error de cuantificación es proporcional a la magnitud de la señal que va a cuantificarse. Los valores muy bajos se cuantifican de manera uniforme a algún valor de corte, por debajo del cual no hay información útil para evaluar la calidad. Este diseño del cuantificador minimiza el error en los cálculos de los parámetros correspondientes porque estos cálculos se basan normalmente en una tasa de errores o tasa logarítmica de los trenes de características procesado y original (véase el punto siguiente sobre descripción de los parámetros).

La Fig. 28 representa la gráfica del cuantificador no lineal de 9 bits empleado para la característica f_{S113} original. La subruina del código de referencia «model_lowbw_compression» proporciona una descripción matemática completa de los cuantificadores recomendados utilizados por el VQM de anchura de banda baja rápido. Si las características caen fuera de la gama de cuantificadores recomendados en el extremo bajo o alto (muy improbable), los parámetros S-T obtenidos de estas características se ponen a cero de manera que no tengan influencia sobre el VQM global.

FIGURA 28

Cuantificador de 9-bits no lineal para la característica f_{S113} 

BT.1885-28

3.3 Descripción de los parámetros

3.3.1 Consideraciones generales sobre los parámetros

El cálculo de los parámetros supone varios pasos que siguen los diversos aspectos perceptuales de la calidad de vídeo. Estos pasos pueden implicar:

- la aplicación de un umbral perceptual a las características extraídas desde cada subregión S-T;

- el cálculo de una función de error entre las características procesadas y las correspondientes características originales;
- la agrupación del error resultante en el espacio y en el tiempo.

En el Anexo D.8 a la Recomendación UIT-T J.144 figura una descripción detallada de estas técnicas y su notación matemática correspondiente para los nombres de los parámetros, que también se utilizará aquí. La subrutina del código de referencia «model_fastlowbw_parameters» proporciona una descripción matemática completa de los parámetros utilizados por el VQM de anchura de banda baja rápido. Para mayor sencillez, la descripción de los parámetros que se hace en este punto no considera los efectos de la cuantificación de las características (por ejemplo, el tratamiento de los valores de las características que pueden caer fuera de las gamas de cuantificación recomendadas).

3.3.2 Nuevos métodos

Este punto resume nuevos métodos establecidos para mejorar la correlación objetiva a subjetiva de los parámetros basados en las características RR con anchuras de banda de transmisión muy bajas tales como las utilizadas por el VQM de anchura de banda baja rápido de la NTIA (es decir, nuevos métodos que no figuran en la Recomendación UIT-T J.144). Conviene señalar que no se han encontrado mejoras para la forma básica de las funciones de error en los parámetros indicados en el Anexo D.8.2.1 a la Recomendación UIT-T J.144. Las dos funciones de error que dan lugar de forma consistente a los mejores resultados de los parámetros (para parámetros espaciales y temporales) son una tasa logarítmica $\{\log_{10} [f_p(s,t) / f_o(s,t)]\}$ y una tasa de errores $\{[f_p(s,t) - f_o(s,t)] / f_o(s,t)\}$, donde $f_p(s,t)$ y $f_o(s,t)$ son la característica procesada y la correspondiente característica original extraídas de la región S-T con las coordenadas espaciales s y las coordenadas temporales t , respectivamente. Los errores deben separarse en ganancias y pérdidas, ya que los seres humanos responden de forma distinta a las degradaciones aditivas (por ejemplo, bloqueo) y sustractivas (por ejemplo, borrosidad). Aplicando un umbral perceptual más bajo a las características antes de la aplicación de estas dos funciones de error se evita la división por cero.

Tras calcular los parámetros S-T utilizando una de las funciones de error, dichos parámetros pueden agruparse en el espacio y en el tiempo para producir un valor de parámetro para el videoclip. Esta agrupación de errores puede producirse en múltiples etapas (por ejemplo, en el espacio y en el tiempo). Un nuevo método de agrupación de errores utilizado por el VQM de anchura de banda baja rápido se denomina agrupación de errores en macrobloque (MB). Esta agrupación de errores MB reúne un número contiguo de subregiones S-T y aplica una función de agrupación de errores a este conjunto. Por ejemplo, la función denominada «MB(3,3,2)max» llevara a cabo una función max sobre los valores de los parámetros para cada grupo de 18 subregiones S-T apiladas de la forma 3 verticales por 3 horizontales por 2 temporales. Para las subregiones S-T de $32 \times 32 \times 1s$ de las características f_{SI13} , f_{HV13} , y f_{COHER_COLOR} descritas antes, cada región MB(3,3,2) engloba una porción del tren de vídeo que abarca 96 líneas verticales por 96 píxels horizontales por 2 segundos de tiempo. Se ha observado que la agrupación de errores MB es útil para realizar un seguimiento de la repercusión perceptual de las degradaciones localizadas en el espacio y en el tiempo. Tales degradaciones localizadas dominan a menudo el proceso de decisión de calidad. La agrupación de errores MB también puede implementarse como un proceso de filtrado de manera que en vez de producir un solo valor de salida para cada MB, cada una de las muestras S-T se sustituye por su valor filtrado MB, estando el MB centrado en la muestra S-T. Esto se denomina agrupación de errores MB solapados (OMB).

Un segundo método de agrupación de errores es el sumatorio de Minkowski generalizado (P, R), definido como sigue:

$$Minkowski(P, R) = \sqrt[R]{\frac{1}{N} \sum_{i=1}^N |v_i|^P}$$

En esta expresión v_i representa los valores de los parámetros incluidos en el sumatorio. Este sumatorio puede, por ejemplo, incluir todos los valores de los parámetros en un instante determinado en el tiempo (agrupación espacial) o puede aplicarse a los macrobloques antes descritos. El sumatorio de Minkowski en el que la potencia P es igual a la raíz R ha sido utilizado por muchos diseñadores de métricas de calidad de vídeo para la agrupación de errores. El sumatorio de Minkowski generalizado, donde $P \neq R$, proporciona flexibilidad adicional para linealizar la respuesta de los parámetros individuales a los cambios en la calidad percibida. Se trata de un paso necesario antes de combinar múltiples parámetros en una sola estimación de la calidad de vídeo percibida, que se lleva a cabo con un ajuste lineal.

3.3.3 Parámetros de color

Se extraen dos parámetros de las características f_{COHER_COLOR} . Uno de estos parámetros, denominado *color_extreme*, mide las distorsiones de color extremas que pueden causar los bloques coloreados de los errores de transmisión. El otro parámetro, denominado *color_spread*, proporciona una indicación de la varianza o dispersión de los errores de color. En vez de usar la medición de la distancia euclidiana para cuantificar las distorsiones (como se hace en el Anexo D.8.2.2 a la Recomendación UIT-T J.144) estos dos parámetros utilizan la raíz cuadrada de la distancia Manhattan. Siguiendo la notación matemática del citado Anexo D.8.2.2 a la Recomendación UIT-T J.144, donde $f_p(s,t)$ y $f_o(s,t)$ representan la característica f_{COHER_COLOR} bidimensional extraída de la región S-T de los trenes de vídeo procesado y original, esta función de comparación de características viene dada por la expresión:

$$sqrmanhat(s,t) = \sqrt{\sum_{C_B, C_R} \left| \frac{f_p(s,t)}{-P} - \frac{f_o(s,t)}{-P} \right|}$$

La medición de la distancia Manhattan parece ser más adecuada que la de la distancia euclidiana y la función de raíz cuadrada es necesaria para linealizar la respuesta de los parámetros a los cambios de calidad. Siguiendo las notaciones matemáticas que aparecen en el Anexo D.8 de la Recomendación UIT-T J.144 los parámetros de color vienen dados por :

$$color_extreme = color_coher_color_30x30_1s_mean_sqrmanhat_OMB(3,3,2)above99\%_Minkoski(0.5,1)$$

$$color_spread = color_coher_color_30x30_1s_mean_sqrmanhat_OMB(3,3,2)Minkoski(2,4)_90\%$$

Un parámetro de color combinado (*color_comb*) que contiene la combinación óptima de los dos parámetros *color_extreme* y *color_spread* se calcula como sigue:

$$color_comb = 0.691686 * color_extreme - 0.617958 * color_spread$$

Este valor positivo del parámetro *color_comb* se recorta a continuación en el extremo inferior, lo que se representa matemáticamente (siguiendo la notación del Anexo D.8.5 de la Recomendación UIT-T J.144) por:

$$color_comb = color_comb_clip_0.114$$

Este parámetro *color_comb* se incluye en la combinación lineal para el cálculo de VQM.

3.3.4 Parámetros espaciales

A partir de la característica f_{SI13} se calculan dos parámetros espaciales, uno que mide una pérdida de información espacial (*si_loss*) y otro que mide una ganancia de información espacial (*si_gain*). Siguiendo la notación matemática del Anexo D.8 a la Recomendación UIT-T J.144 estos parámetros vienen dados por :

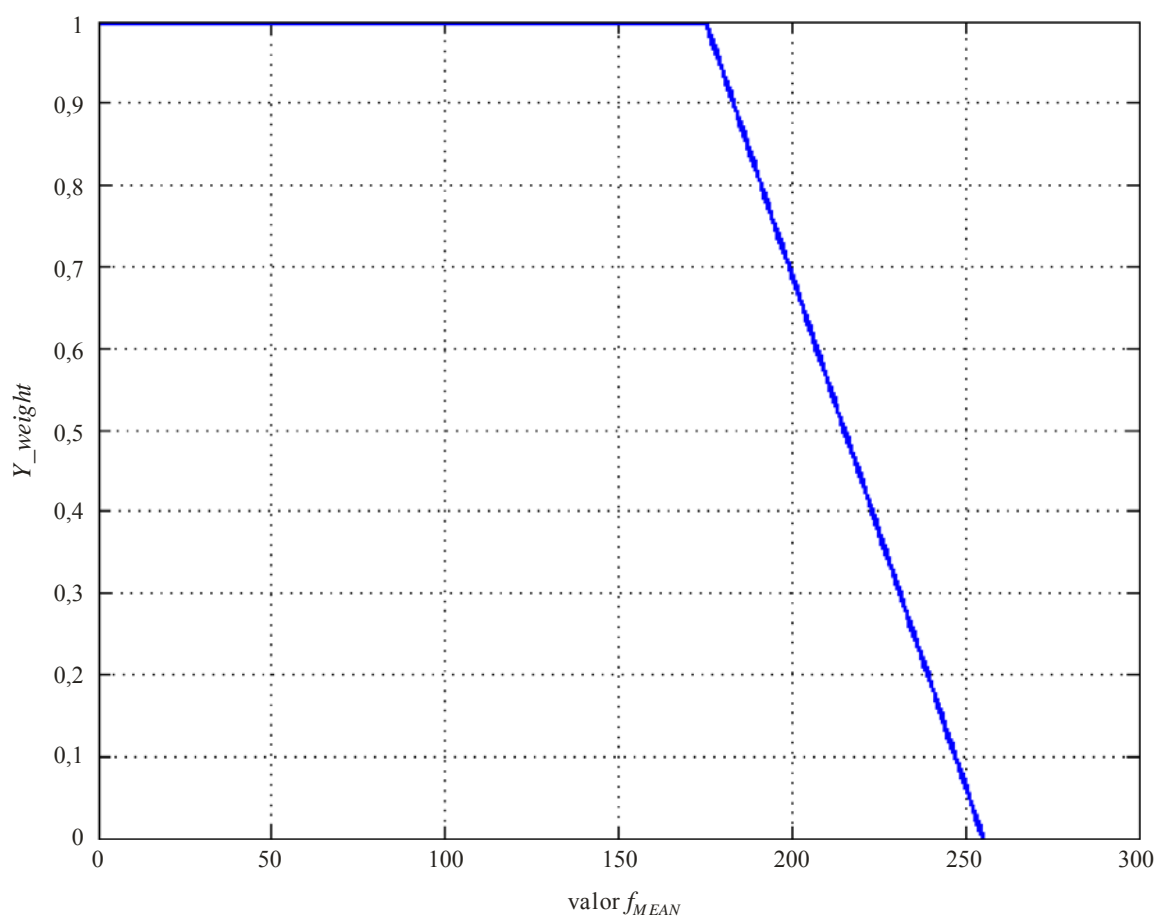
$$si_loss = avg1s_Y_sil3_30x30_std_3_ratio_loss_OMB(3,3,2)Minkoski(1,2)_Minkoski(1.5,2.5)_clip_0.12$$

$$si_gain = avg1s_Y_sil3_30x30_std_3_log_gain_clip_0.1_above99\%tail_Minkoski(1.5,2)$$

A medida que aumenta el nivel medio de luminancia (Y) de la subregión S-T (medido mediante la característica f_{MEAN}), disminuye la capacidad de percibir cambios en el detalle espacial (por ejemplo, tal como la borrosidad medida por si_loss). Esto puede representarse introduciendo una función de ponderación (Y_weight) como muestra la Fig. 29 a los valores si_loss de cada subregión S-T (es decir, los valores si_loss tras llevar a cabo la función de comparación de tasa de pérdidas en cada subregión S-T pero antes de las funciones de colapso espacial y temporal). La función de ponderación Y_weight es igual a uno (es decir, ponderación total) hasta que se alcanza un nivel medio de luminancia de 175 y a continuación disminuye linealmente hasta cero a medida que los valores de luminancia aumentan de 175 a 255. Esta corrección intermedia se aplica únicamente a los valores si_loss y no a los valores si_gain .

FIGURA 29

Función de ponderación Y_weight para modificar los parámetros si_loss S-T



BT.1885-29

A partir de la característica f_{HV13} se calculan dos parámetros espaciales, uno que mide la pérdida de información espacial Horizontal y Vertical (HV) relativa (hv_loss) y otro que mide la ganancia (hv_gain). Siguiendo la notación matemática del Anexo D.8 a la Recomendación UIT-T J.144 estos parámetros vienen dados por las expresiones:

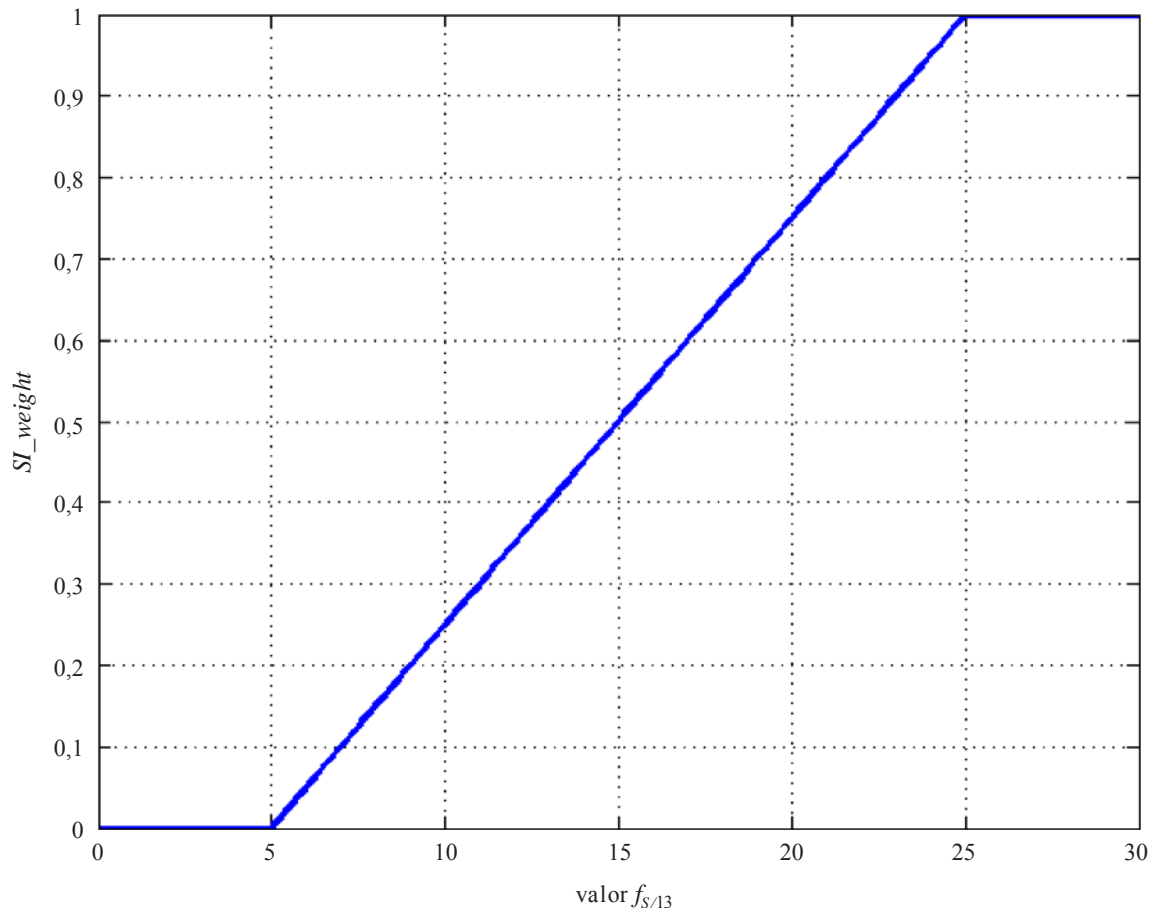
$$hv_loss = \text{avg1s_Y_hv13_angle0.225_rmin20_30x30_mean_4_ratio_loss_...} \\ \text{OMB}(3,3,2)\text{below1\%_Minkoski}(1,1.5)_clip_0.08$$

$$hv_gain = \text{avg1s_Y_hv13_angle0.225_rmin20_30x30_mean_4_log_gain_...} \\ \text{clip_0.06_OMB}(3,3,2)\text{above99\%tail_Minkoski}(1.5,3)$$

En las ecuaciones anteriores no se muestra que la función Y_weight representada en la Fig. 29 también se aplica a los dos valores hv_loss y hv_gain de cada subregión S-T antes de las funciones de colapso espacial y temporal (tras calcular, respectivamente, los valores de $ratio_loss$ y log_gain). Se aplica una función de ponderación adicional (SI_weight mostrada en la Fig. 30) a los valores de hv_loss de cada subregión S-T. Ello fue necesario para reducir la sensibilidad de hv_loss en las regiones S-T que tienen muy poca información espacial (es decir, valores bajos de la característica original f_{S13}).

FIGURA 30

Función de ponderación SI_weight para modificar los parámetros hv_loss S-T



BT.1885-30

Los parámetros de distorsión espacial pueden aplastarse (es decir, las excursiones que sobrepasan claramente las que aparecen en los datos de prueba se limitan o comprimen) utilizando funciones tales como la función de aplastamiento VQM que figura en el punto relativo al cálculo de VQM.

3.3.5 Parámetros temporales

A partir de la característica f_{ATI} se calculan dos parámetros temporales, uno que mide el ruido aleatorio añadido (ati_noise) y otro que mide las perturbaciones del movimiento causadas por los errores de transmisión (ati_error). Siguiendo la notación matemática del Anexo D.8 a la Recomendación UIT-T J.144 estos parámetros vienen dados por:

$$ati_noise = Y_rand5\%_ati0.2s_rms_5_ratio_gain_between25\%50\%$$

$$ati_error = Y_rand5\%_ati0.2s_rms_max7pt_12_ratio_gain_above90\%$$

Para que los parámetros *ati_noise* y *ati_error* sean más robustos contra los desalineamientos temporales, los parámetros se calculan para todos los alineamientos temporales del vídeo procesado que se encuentren en el margen de $\pm 0,4$ s con respecto al mejor alineamiento temporal estimado del vídeo original, y a continuación se selecciona el mínimo valor del parámetro.

3.4 Cálculo del VQM

De forma similar al VQM General de la NTIA que aparece en el Anexo D a la Recomendación UIT-T J.144, el cálculo del VQM de anchura de banda baja rápida combina linealmente dos parámetros de la característica f_{HV13} (*hv_loss* y *hv_gain*), dos parámetros de la característica f_{SI13} (*si_loss* y *si_gain*) y dos parámetros de la característica f_{COHER_COLOR} (excepto que los dos parámetros han sido combinados en un solo parámetro de distorsión de color denominado *color_comb*). El parámetro de ruido en el VQM General de la NTIA ha sido sustituido por 2 parámetros basados en la característica f_{ATI} de anchura de banda baja aquí descritos (*ati_noise* y *ati_error*).

Por consiguiente, VQM_{FLB} (abreviatura de VQM de anchura de banda baja rápida) consiste en una combinación lineal de ocho parámetros. VQM_{FLB} viene dado por:

$$VQM_{FLB} = \{ 0,38317338378290 * hv_loss + 0,37313218013131 * hv_gain + \\ 0,58033514546526 * si_loss + 0,95845512360511 * si_gain + \\ 1.07581708014998 * color_comb + \\ 0,17693274495002 * ati_noise + 0,02535903906351 * ati_error \}$$

El VQM total (tras la adición de las contribuciones de todos los parámetros) se recorta al umbral más bajo de 0,0 a fin de evitar los números VQM negativos. Por último, se aplica a los valores de VQM por encima de 1,0 una función de aplastamiento que permite una máxima sobreoscilación del 50% para limitar los valores de VQM de vídeo excesivamente distorsionado que caen fuera de la gama de los datos de prueba.

Si $VQM_{FLB} > 1,0$, entonces $VQM_{FLB} = (1 + c) * VQM_{FLB} / (c + VQM_{FLB})$, donde $c = 0,5$.

VQM_{FLB} calculado de la forma anterior tendrá valores superiores o iguales a cero y un máximo valor nominal de uno. El VQM_{FLB} puede ocasionalmente rebasar el valor uno en el caso de escenas de vídeo extremadamente distorsionadas.

Para lograr que el VQM_{FLB} sea más robusto frente a los desalineamientos espaciales, el VQM_{FLB} se calcula para todos los alineamientos espaciales del vídeo procesado que se encuentran a más o menos 1 píxel del mejor alineamiento espacial estimado del vídeo original, y a continuación se selecciona el mínimo VQM_{FLB} .

4 Referencias

- [1] Recomendación UIT-R BT.601-6 (01/07) – Parámetros de codificación de televisión digital para estudios con formatos de imagen normal 4:3 y de pantalla ancha 16:9.
- [2] Video Quality Model (VQM) Software Tools – Binary executables and source code, available from the National Telecommunications and Information Administration (NTIA) at http://www.its.bldrdoc.gov/n3/video/VQM_software.php.
- [3] Recomendación UIT-T J.244 (04/08) – Métodos de calibración de referencia completa y referencia reducida para los sistemas de transmisión de vídeo con desalineamiento constante de los dominios espacial y temporal con ganancia y desplazamientos constantes.

- [4] VQEG Final Report of MM Phase I Validation Test (2008), «Final report from the Video Quality Experts Group on the validation of objective models of multimedia quality assessment, phase I», Video Quality Experts Group (VQEG), <http://www.its.bldrdoc.gov/vqeg/projects/multimedia>, ITU-T Study Group 9 TD923, Study Period 2005-2008.
- [5] Recomendación UIT-T J.144 (03/04) – Técnicas de medición objetiva de la percepción de la calidad de vídeo en televisión digital por cable en presencia de una referencia completa.
- [6] Recomendación UIT-R BT.1683 (06/04) – Técnicas de medición objetiva de la calidad de vídeo perceptual para la radiodifusión de televisión digital de definición convencional en presencia de una referencia completa.

5 Código de referencia para implementar el VQM de anchura de banda baja rápido

El objeto de este código de referencia es asistir al usuario a realizar la adecuada implementación del VQM de anchura de banda baja rápido. Aunque se utiliza el código MATLAB® como código de referencia, puede emplearse cualquier código de software que reproduzca los resultados indicados aquí. Cada subsección del § 5 contiene código MATLAB para la función indicada en el encabezamiento de la sección (por ejemplo, salvar el contenido del § 5.1 en un fichero denominado «fastlowbw_ref.m»). Ejecutar fastlowbw_ref sin argumentos para recibir información de ayuda sobre la forma de llamar a una rutina. Este código contiene la flexibilidad necesaria para ejecutar el modelo en un videoclip corto (es decir, de 5 a 15 s de duración) dentro de una secuencia de vídeo más larga (por ejemplo, una secuencia de 1 minuto). Ello se hace desplazando el videoclip corto un segundo y volviendo a calcular el modelo para cada desplazamiento temporal. Aunque no se demuestra esta funcionalidad en lo que sigue, los comentarios en el código y los argumentos devueltos de «model_fastlowbw_parameters.m» se referirán a esta capacidad. Esta capacidad escondida puede ser útil para implementar un sistema de comprobación de la calidad de vídeo en servicio.

Cuando los vectores de prueba de muestra (es decir, los videoclips) se procesan con el código de referencia VQM de anchura de banda baja rápido (función «fastlowbw_ref.m»), se crean ficheros de texto que contienen la calibración y los resultados del modelo. En el siguiente ejemplo de llamadas de función MATLAB, deben obtenerse ficheros de salida similares a los indicados a continuación (debido al proceso aleatorio utilizado por el VQM de anchura de banda baja rápido, los resultados pueden variar ligeramente de los aquí presentados):



5.Reference
Code.doc

Apéndice

Análisis de los errores de transmisión

Las pruebas de validación de VQEG para el proyecto RRNR-TV incluían los formatos de 525 líneas (NTSC) y 625 líneas (PAL). Cada experimento incluyó 12 secuencias de fuente y 156 secuencias de vídeo procesadas (PVS). De estas 156 PVS, 40 contenían errores de transmisión y 116 únicamente errores de codificación. Los Cuadros 9 y 10 muestran el RMSE y el OR para las PVS con errores de transmisión. Cabe señalar que los RMSE y los OR se calcularon utilizando las rectas de regresión obtenidas de todos los datos. En otras palabras, las rectas de regresión se determinaron utilizando todos los datos. A continuación, el RMSE y el OR para los errores de transmisión se calcularon mediante las PVS con errores de transmisión.

CUADRO 9

**RMSE y OR para la prueba de validación RRNR-TV (formato 525).
ET: errores de transmisión**

| Formato 525 | TODOS | | Con ET | | Sin ET | |
|--------------|-------|-------|--------|-------|--------|-------|
| | RMSE | OR | RMSE | OR | RMSE | OR |
| Model_A_15k | 0,418 | 0,385 | 0,574 | 0,500 | 0,362 | 0,293 |
| Model_A_80k | 0,423 | 0,378 | 0,582 | 0,475 | 0,366 | 0,293 |
| Model_A_256k | 0,424 | 0,378 | 0,584 | 0,475 | 0,367 | 0,293 |
| Model_B_80k | 0,598 | 0,667 | 0,768 | 0,650 | 0,544 | 0,586 |
| Model_B_256k | 0,587 | 0,647 | 0,763 | 0,600 | 0,530 | 0,578 |
| Model_C_80k | 0,465 | 0,513 | 0,557 | 0,550 | 0,440 | 0,405 |
| Model_C_256k | 0,511 | 0,609 | 0,584 | 0,450 | 0,495 | 0,578 |
| PSNR_NTIA | 0,556 | 0,571 | 0,549 | 0,500 | 0,568 | 0,491 |

CUADRO 10

**RMSE y OR para la prueba de validación RRNR-TV (formato 625).
ET: errores de transmisión**

| Formato 625 | TODOS | | Con ET | | Sin ET | |
|--------------|-------|-------|--------|-------|--------|-------|
| | RMSE | OR | RMSE | OR | RMSE | OR |
| Model_A_15k | 0,524 | 0,468 | 0,597 | 0,450 | 0,508 | 0,414 |
| Model_A_80k | 0,513 | 0,462 | 0,594 | 0,500 | 0,494 | 0,379 |
| Model_A_256k | 0,516 | 0,468 | 0,593 | 0,500 | 0,499 | 0,379 |
| Model_B_80k | 0,887 | 0,724 | 0,545 | 0,500 | 0,986 | 0,716 |
| Model_B_256k | 0,864 | 0,744 | 0,523 | 0,600 | 0,962 | 0,716 |
| Model_C_80k | 0,585 | 0,583 | 0,282 | 0,200 | 0,663 | 0,647 |
| Model_C_256k | 0,657 | 0,590 | 0,292 | 0,175 | 0,747 | 0,638 |
| PSNR_NTIA | 0,605 | 0,564 | 0,338 | 0,250 | 0,678 | 0,517 |