

# UIT-R

Sector de Radiocomunicaciones de la UIT

## **Recomendación UIT-R BT.1908-0** (01/2012)

# **Técnicas de de evaluación de la calidad de vídeo objetiva para las aplicaciones de radiodifusión que utilizan TVAD en presencia de una señal de referencia 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                                                                                                           |
|------------|------------------------------------------------------------------------------------------------------------------|
| <b>BO</b>  | Distribución por satélite                                                                                        |
| <b>BR</b>  | Registro para producción, archivo y reproducción; películas en televisión                                        |
| <b>BS</b>  | Servicio de radiodifusión sonora                                                                                 |
| <b>BT</b>  | <b>Servicio de radiodifusión (televisión)</b>                                                                    |
| <b>F</b>   | Servicio fijo                                                                                                    |
| <b>M</b>   | Servicios móviles, de radiodeterminación, de aficionados y otros servicios por satélite conexos                  |
| <b>P</b>   | Propagación de las ondas radioeléctricas                                                                         |
| <b>RA</b>  | Radio astronomía                                                                                                 |
| <b>RS</b>  | Sistemas de detección a distancia                                                                                |
| <b>S</b>   | Servicio fijo por satélite                                                                                       |
| <b>SA</b>  | Aplicaciones espaciales y meteorología                                                                           |
| <b>SF</b>  | Compartición de frecuencias y coordinación entre los sistemas del servicio fijo por satélite y del servicio fijo |
| <b>SM</b>  | Gestión del espectro                                                                                             |
| <b>SNG</b> | Periodismo electrónico por satélite                                                                              |
| <b>TF</b>  | Emisiones de frecuencias patrón y señales horarias                                                               |
| <b>V</b>   | 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, 2020

© UIT 2020

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.1908-0\*

**Técnicas de de evaluación de la calidad de vídeo objetiva  
para las aplicaciones de radiodifusión que utilizan  
TVAD en presencia de una señal  
de referencia reducida**

(2012)

**Cometido**

Esta Recomendación describe los métodos de evaluación de la calidad de vídeo percibida para las aplicaciones de radiodifusión que utilizan televisión digital de alta definición cuando se dispone de una señal de referencia reducida.

La Asamblea de Radiocomunicaciones de la UIT,

*considerando*

- a) que la capacidad de medir automáticamente la calidad del vídeo de radiodifusión se ha considerado desde hace tiempo como un activo de interés para la industria;
- b) que la Recomendación UIT-R BT.1683 describe métodos objetivos para medir la calidad de vídeo percibida para la radiodifusión de televisión digital de definición convencional cuando se dispone de una señal de referencia reducida;
- c) que la Recomendación UIT-R BT.709 describe valores de parámetro para las normas de TVAD para la producción y el intercambio internacional de programas y que la Recomendación UIT-R BT.500 describe métodos de evaluación subjetiva de la calidad de imagen incluida la televisión de alta definición;
- d) que la TVAD se utiliza cada vez más en la radiodifusión;
- e) que, basándose en los resultados del informe sobre TVAD enviado por el VQEG, la Comisión de Estudio 9 del UIT-T ha elaborado la Recomendación UIT-T J.342, que especifica la medición objetiva de la calidad de vídeo percibida de la TVAD en presencia de una señal de referencia reducida;
- f) que la medición objetiva de la calidad de vídeo percibida de la TVAD podrá complementar los métodos de evaluación subjetiva,

*recomienda*

- 1 que se utilice el modelo de medición objetiva de la calidad de vídeo que se presenta en el Anexo 1 para la medición objetiva de la calidad de vídeo percibida para las aplicaciones de radiodifusión que utilizan TVAD cuando se dispone de una señal de referencia reducida.

---

\* La Comisión de Estudio 6 de Radiocomunicaciones introdujo modificaciones redaccionales en esta Recomendación en febrero de 2020, de conformidad con la Resolución UIT-R 1.

## Anexo 1

### 1 Introducción

La presente Recomendación proporciona un método de medición de la calidad de vídeo para aplicaciones no interactivas de televisión de alta definición (TVAD) cuando se puede utilizar el método de medición de referencia reducida (RR). El modelo se comparó con los grados de calidad subjetiva obtenidos mediante la Recomendación UIT-R BT.500. Los análisis realizados han demostrado que la precisión de este modelo es equivalente a la de la PSNR.

Para que el modelo RR funcione correctamente, es preciso disponer del vídeo fuente intacto para que el modelo extraiga sus parámetros. Estos parámetros extraídos, junto con la secuencia de vídeo degradada, son los datos de partida para el modelo RR. El método de estimación realiza una evaluación tanto de la calibración (es decir, ganancia/desplazamiento y registro espacial/temporal) como de la calidad de vídeo objetiva.

Para la prueba de validación se utilizó tanto la Recomendación UIT-T H.264 como las degradaciones de codificación MPEG-2 y diversas condiciones de error de transmisión (por ejemplo, errores de bits, paquetes perdidos). El modelo de la presente Recomendación se puede utilizar para verificar la calidad de redes desplegadas con el fin de garantizar su disponibilidad operativa. Entre los efectos visuales de degradación se incluyen tanto degradaciones espaciales como temporales. Este modelo también se puede utilizar para pruebas de laboratorio de sistemas de vídeo. Cuando se use para comparar diferentes sistemas de vídeo, se aconseja emplear un método cuantitativo (como el de la Recomendación UIT-T J.149) para determinar la precisión del modelo en ese contexto concreto.

La presente Recomendación se considera adecuada para los servicios de radiodifusión entre 1 Mbit/s y 30 Mbit/s. En las pruebas de validación se consideraron las siguientes resoluciones y velocidades de trama:

- 1080/59,94/I
- 1080/25/P
- 1080/50/I
- 1080/29,97/P.

Durante las pruebas de validación se consideraron las siguientes condiciones para cada resolución:

| <b>Factores de prueba</b>                                                                                            |
|----------------------------------------------------------------------------------------------------------------------|
| Resolución de vídeo: 1 920 × 1 080 entrelazado y progresivo                                                          |
| Velocidades de trama de vídeo: 29,97 y 25 tramas por segundo                                                         |
| Velocidades binarias de vídeo: 1 a 30 Mbit/s                                                                         |
| Congelación temporal de trama (pausa con salto) de hasta dos segundos                                                |
| Errores de transmisión con pérdida de paquetes                                                                       |
| Conversión del SRC de 1 080 a 720/P, compresión, transmisión, descompresión y finalmente conversión de nuevo a 1 080 |
| <b>Tecnologías de codificación</b>                                                                                   |
| H.264/AVC (MPEG-4 Parte 10)                                                                                          |
| MPEG-2                                                                                                               |

Cabe señalar que, en el plan de pruebas de validación, se consideró 720/P como parte de las condiciones de prueba (HRC). Puesto que en la actualidad las pantallas suelen estar adaptadas a la resolución 720/P, se estimó que los circuitos HRC de 720/P serían más adecuados para este formato.

### 1.1 Aplicaciones

Las aplicaciones para los modelos de estimación descritos en esta Recomendación incluyen, entre otras:

- 1) la supervisión de la calidad de vídeo en el receptor cuando se dispone de canales laterales;
- 2) la supervisión de la calidad de vídeo en puntos de medición situados entre el punto de transmisión y el punto de recepción.

El modelo descrito en esta Recomendación proporciona unas prestaciones estadísticamente similares a las del modelo PSNR, por lo que se puede utilizar para la evaluación de la calidad de vídeo cuando se dispone de la señal de referencia reducida en el punto de medición.

### 1.2 Limitaciones

El modelo de estimación de la calidad de vídeo descrito en la presente Recomendación no puede sustituir a las pruebas subjetivas. Los valores de correlación entre dos pruebas subjetivas cuidadosamente diseñadas y realizadas (es decir, en dos laboratorios diferentes) normalmente se encuentran entre 0,95 y 0,98. La presente Recomendación no se puede utilizar para comparar sistemas de vídeo (por ejemplo, comparar dos códecs o dos implementaciones diferentes del mismo algoritmo de compresión). Las prestaciones del modelo de estimación de calidad de vídeo descrito en la presente Recomendación no son estadísticamente mejores que las de la PSNR.

Cuando existe congelación de trama, las condiciones de prueba normalmente tienen duraciones de congelación de trama inferiores a dos segundos. El modelo en esta Recomendación no se validó para medir calidades de vídeo en condiciones de almacenamiento temporal (es decir, vídeo con un retardo creciente constante o congelación sin salto). El modelo no se probó para velocidades de trama distintas de las utilizadas en los sistemas de televisión (es decir, 29,97 tramas por segundo y 25 tramas por segundo, en modo entrelazado o progresivo).

Cabe señalar que, en el caso de las 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.

Hay que destacar que el modelo de la presente Recomendación no se evaluó con contenido presencial típico de las videoconferencias.

## 2 Referencias

Las siguientes Recomendaciones del UIT-T 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.

Recomendación UIT-T J.244 (2008) – *Métodos de referencia completa y de referencia reducida para calibrar los sistemas de transmisión video con desalineación constante de los dominios espacial y temporal con ganancia y desplazamiento constantes.*

### 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):** Determinación de la calidad o del deterioro de imágenes tipo programa presentadas a un grupo de evaluadores humanos en sesiones de observación visual.

**3.1.2 Medición perceptual objetiva (imagen):** Medición de las características de una cadena de programa utilizando imágenes tipo programa y métodos de medición objetiva (con instrumentos) para obtener un indicio que se acerque a la nota que se obtendría con una prueba de evaluación subjetiva.

**3.1.3 Proponente:** Organización o empresa que propone un modelo de calidad de vídeo para la prueba de validación y su posible inclusión en una Recomendación de la UIT.

#### 3.2 Términos definidos en esta Recomendación

Esta Recomendación define los siguientes términos:

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

**3.2.2 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.

**3.2.3 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.

### 4 Abreviaturas y acrónimos

Esta Recomendación utiliza las siguientes abreviaturas y acrónimos

|        |                                                                                                                                                       |
|--------|-------------------------------------------------------------------------------------------------------------------------------------------------------|
| ACR    | Índice por categoría absoluta ( <i>absolute category rating</i> ) (véase la Recomendación UIT-R BT.500)                                               |
| 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> )                                                                             |
| ILG    | Grupo de laboratorio independiente del VQEG ( <i>VQEG's independent laboratory group</i> )                                                            |
| MOS    | Nota media de opinión ( <i>mean opinion score</i> )                                                                                                   |
| MOSp   | Nota media de opinión, predicción ( <i>mean opinion score, predicted</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 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.

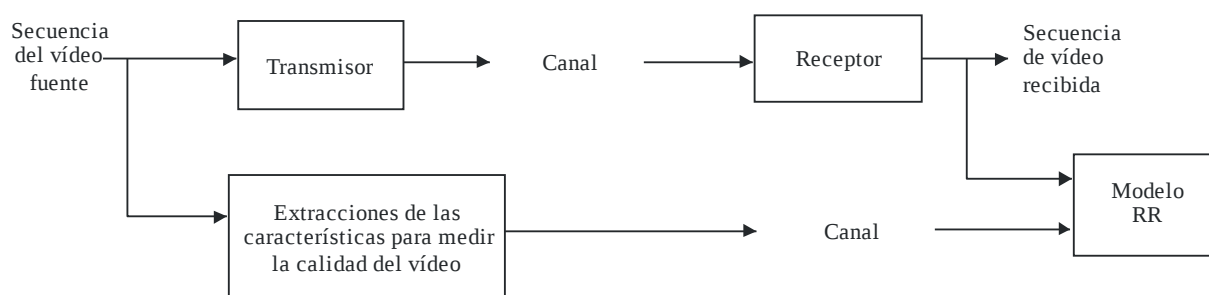
## 6 Descripción del método de medición con referencia reducida

### 6.1 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. 1 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 1 muestra las anchuras de banda del canal lateral para las características que se han comprobado en la prueba de TVAD del VQEG.

FIGURA 1  
Diagrama de bloques del modelo de referencia reducida



CUADRO 1

**Anchuras de banda del canal lateral**

| Formato del vídeo                               | Anchuras de banda probadas        |
|-------------------------------------------------|-----------------------------------|
| 1080/60 Hz (29,97 fps)<br>1080/30Pp (29,97 fps) | 56 kbit/s, 128 kbit/s, 256 kbit/s |
| 1080/25Pp (25 fps)<br>1080/50I Hz (25 fps)      | 56 kbit/s, 128 kbit/s, 256 kbit/s |

**6.2 Modelos de referencia reducida EPSNR****6.2.1 PSNR del borde (EPSNR)**

Los modelos RR miden principalmente las degradaciones en el borde. En los modelos, se aplica en primer lugar un algoritmo de detección del borde 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.

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 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. 2 a 6 ilustran este procedimiento. La Fig. 2 muestra una imagen fuente. La Fig. 3 representa una imagen de gradiente horizontal  $g_{horizontal}(m,n)$  obtenida aplicando un operador de gradiente horizontal a la imagen fuente de la Fig. 2. La Fig. 4 muestra una imagen del gradiente vertical  $g_{vertical}(m,n)$  que se obtiene aplicando un operador del gradiente vertical a la imagen fuente de la Fig. 2. La Fig. 5 representa la imagen de gradiente de magnitud (imagen de borde) y la Fig. 6 muestra la imagen de borde binaria (imagen de máscara) obtenida aplicando el umbral a la imagen del gradiente de magnitud de la Fig. 5.

FIGURA 2

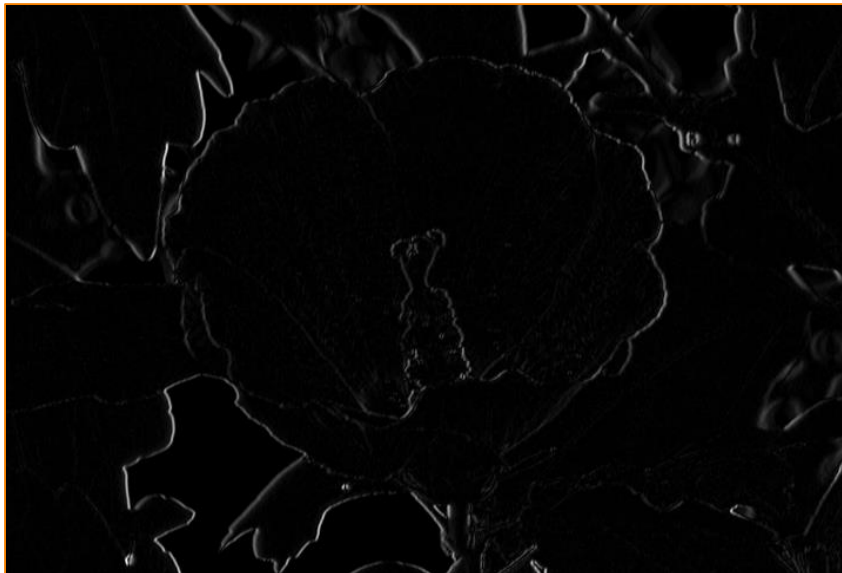
**Imagen fuente (imagen original)**



BT.1908-02

FIGURA 3

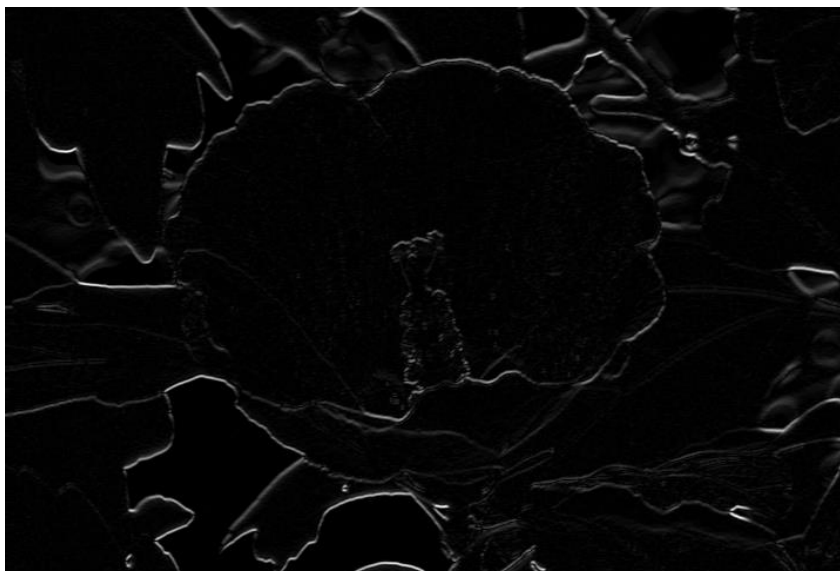
**Imagen del gradiente horizontal obtenida aplicando un operador de gradiente horizontal a la imagen fuente de la Fig. 2**



BT.1908-03

FIGURA 4

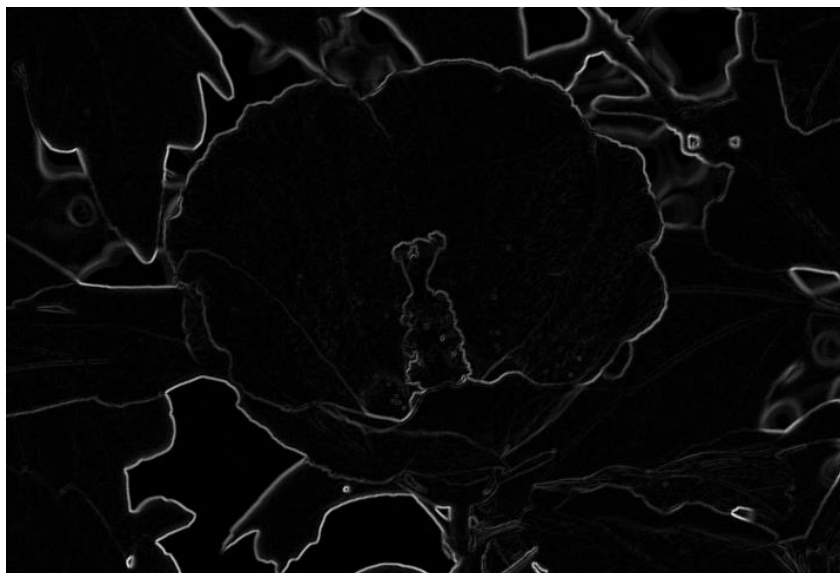
**Imagen del gradiente vertical obtenida aplicando un operador de gradiente vertical a la imagen fuente de la Fig. 2**



BT.1908-04

FIGURA 5

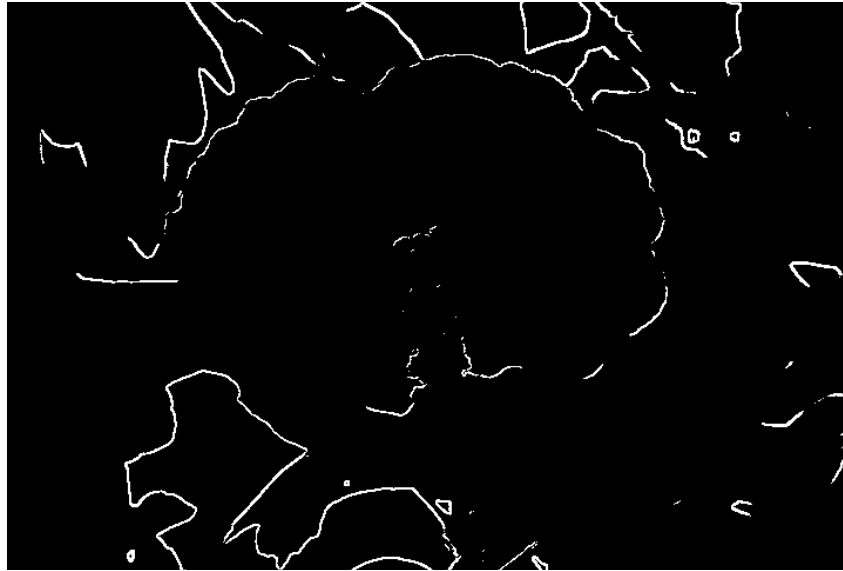
**Imagen del gradiente de magnitud**



BT.1908-05

FIGURA 6

Imagen de borde binaria (imagen de máscara) obtenida aplicando el umbral a la imagen del gradiente de magnitud de la Fig. 5

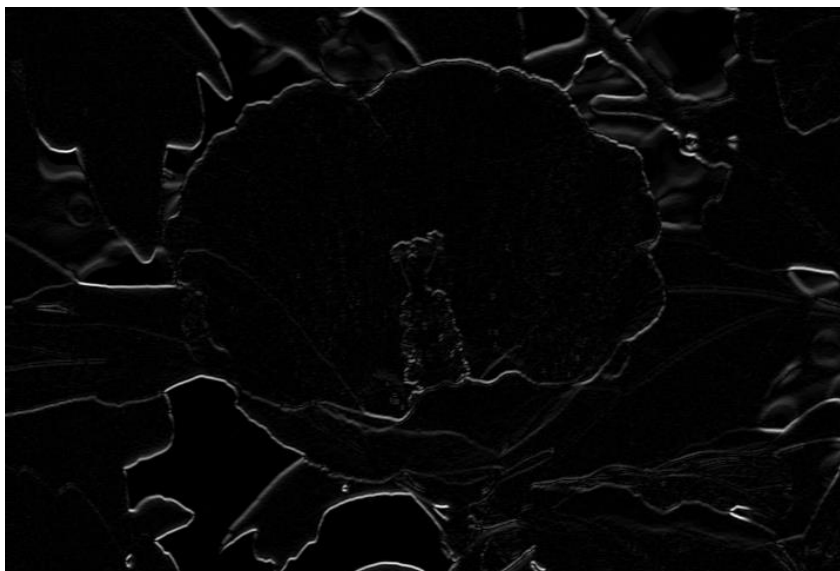


BT.1908-06

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 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 rebasen el valor umbral, se consideran como píxeles del borde. Las Figs. 7 a 9 ilustran el procedimiento modificado. La Fig. 7 muestra una imagen de gradiente vertical  $g_{vertical}(m,n)$  obtenida aplicando un operador de gradiente vertical a la imagen fuente de la Fig. 2. La Fig. 8 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. 7. La Fig. 9 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. 8.

FIGURA 7

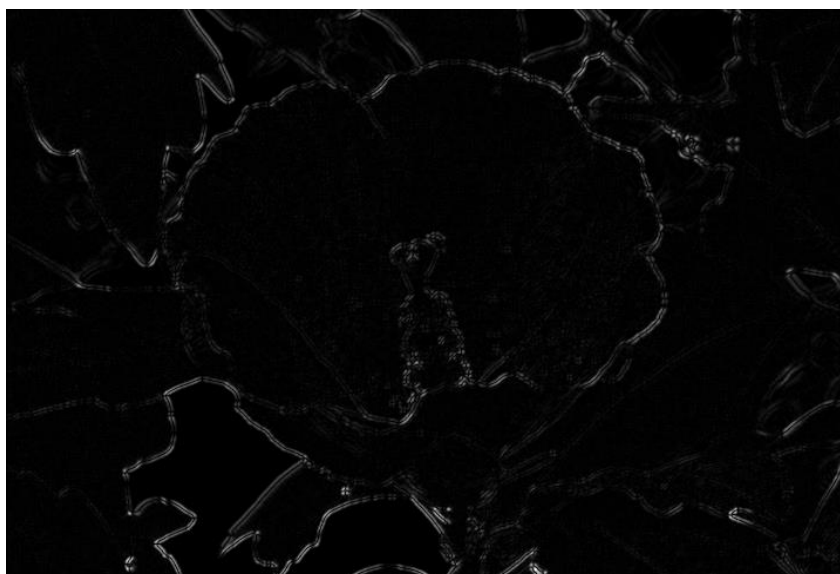
**Imagen de gradiente vertical obtenida aplicando un operador de gradiente vertical a la imagen fuente de la Fig. 2**



BT.1908-07

FIGURA 8

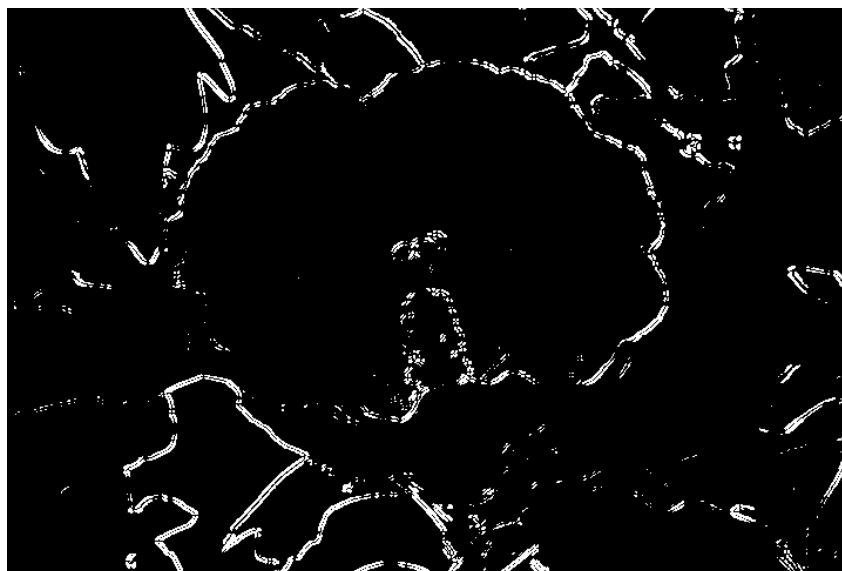
**Imagen de gradiente sucesivo modificada (imagen de gradiente horizontal y vertical) obtenida aplicando un operador de gradiente horizontal a la imagen de gradiente vertical de la Fig. 7**



BT.1908-08

FIGURA 9

Imagen de borde binario (imagen de máscara) obtenida aplicando el umbral a la imagen de gradiente sucesivo modificada de la Fig. 8



BT.1908-09

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. 5 y 8). A continuación, se produce una imagen de máscara (imagen de borde binario) aplicando el umbral a la imagen de borde (véanse las Figs. 6 y 9). 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. 6 y 9 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.

### 6.2.2 Selección de las características a partir de las secuencias del vídeo fuente

Como se trata de un modelo de referencia reducida (RR), debe extraerse un conjunto de características de cada imagen de una secuencia de vídeo fuente. En el modelo EPSBR RR, se selecciona un cierto número de píxeles de borde de cada imagen. A continuación, se codifican y transmiten los datos de los emplazamientos y los valores de los píxeles. Sin embargo, para algunas secuencias de vídeo, el número de píxeles 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 píxeles de borde de una imagen es más pequeño que un valor determinado, el usuario puede reducir el valor umbral hasta que el número de píxeles de borde sea superior a un valor concreto. Alternativamente, pueden seleccionarse píxeles de borde que correspondan a los mayores valores de la imagen de gradiente horizontal y vertical. Cuando no hay píxeles de borde (por ejemplo, imágenes en blanco) en una trama, puede seleccionarse aleatoriamente el número requerido de píxeles o saltar la trama. Por ejemplo, si se seleccionan 10 píxeles de borde de cada trama, pueden clasificarse los píxeles 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 píxeles de borde en emplazamientos idénticos. A fin de superar este problema, puede seleccionarse en primer lugar varias veces el número deseado de píxeles de la imagen de gradiente horizontal y vertical y a continuación elegir aleatoriamente el número deseado de píxeles de borde entre los píxeles seleccionados de la imagen de gradiente horizontal y vertical. En los modelos probados en la prueba de TVAD del VQEG, el número deseado de píxeles de borde se elige aleatoriamente entre un gran conjunto de píxeles de borde. Dicho conjunto se obtiene aplicando una operación umbral a la imagen de gradiente.

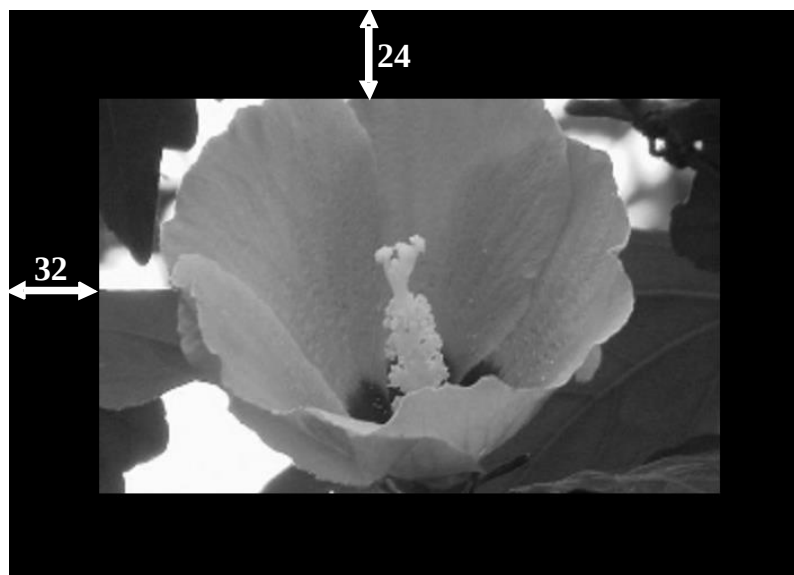
En los modelos EPSNR RR, los emplazamientos y los valores de los píxeles de borde se codifican tras aplicar un filtro gaussiano paso bajo a los emplazamientos de los píxeles seleccionados. Aunque en la prueba de TVAD del VQEG se utilizó el filtro paso bajo gaussiano ( $7 \times 3$ ), pueden emplearse diferentes filtros paso bajo dependiendo de los formatos de vídeo. Cabe señalar que durante el proceso de codificación, puede aplicarse un recorte. Para evitar la selección de píxeles de borde en las zonas recortadas, el modelo selecciona píxeles de borde en la zona media (véase la Fig. 10). El Cuadro 2 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 2

**Requisitos de bits por cada píxel de borde**

| <b>Formato del vídeo</b> | <b>Tamaño</b> | <b>Tamaño tras el recorte</b> | <b>Bits por emplazamiento</b> | <b>Bits por valor de píxel</b> | <b>Número total de bits por píxel</b> |
|--------------------------|---------------|-------------------------------|-------------------------------|--------------------------------|---------------------------------------|
| AD progresivo            | 1 920 × 1080  | 1 856 × 1 032                 | 21                            | 8                              | 29                                    |
| AD entrelazado           | 1 920 × 540   | 1 856 × 516                   | 20                            | 8                              | 28                                    |

FIGURA 10

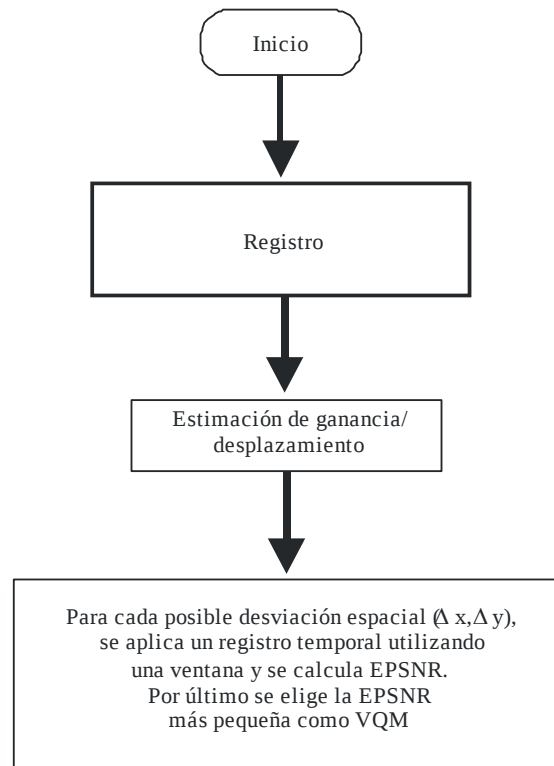
**Ejemplo de recorte y la zona media**

El modelo selecciona los píxeles de borde de cada trama de conformidad con la anchura de banda permitida (véase el Cuadro 1). El Cuadro 3 muestra el número de píxeles de borde por trama que puede transmitirse en las anchuras de banda probadas.

**CUADRO 3**  
**Número de píxeles de banda por trama**

| Formato del vídeo | 56 kbit/s | 128 kbit/s | 256 kbit/s |
|-------------------|-----------|------------|------------|
| AD progresivo     | 46        | 105        | 211        |
| AD entrelazado    | 24        | 54         | 109        |

**FIGURA 11**  
**Organigrama del modelo**



### 6.2.3 Registro espacial/temporal y ajuste ganancia/desplazamiento

Antes de calcular la diferencia entre los píxeles de borde de la secuencia del vídeo 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 de TVAD del VQEG. 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, y se aplicó el método de calibración a las tramas en secuencias de vídeo progresivo. 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 de TVAD del VQEG el valor umbral se fijó a 2 dB.

En el punto de comprobación, la secuencia de vídeo procesada debe alinearse con los píxeles de borde extraídos de la secuencia del vídeo fuente. Sin embargo, si la anchura de banda del canal lateral es pequeña, solo están disponibles unos pocos píxeles de borde de la secuencia de vídeo fuente (véase la Fig. 12). Por tanto, el registro temporal puede ser inexacto si se lleva a cabo utilizando una sola trama (véase la Fig. 13). 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. 14 ilustra el procedimiento. El error cuadrático medio en la ventana se determina como sigue:

$$MSE_{window} = \frac{1}{N_{win}} \sum (E_{SRC}(i) - E_{PVS}(i))^2$$

donde  $MSE_{ventana}$  es el error cuadrático medio de la ventana,  $E_{SRC}(i)$  es un píxel de borde dentro de la ventana que tiene su correspondiente píxel en la secuencia de vídeo procesada,  $E_{PVS}(i)$  es un píxel de la secuencia de vídeo procesada correspondiente al píxel de borde y  $N_{ventana}$  es el número total de píxeles de borde utilizados para calcular el  $MSE_{ventana}$ . 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 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. 18).

FIGURA 12

Selección de píxel de borde de la secuencia de vídeo fuente

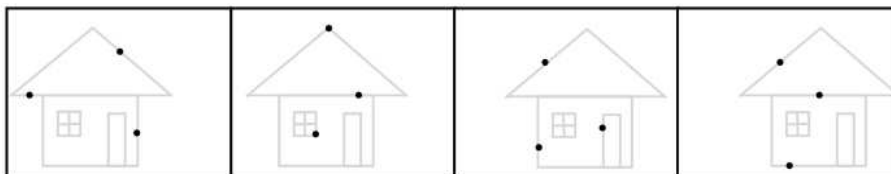
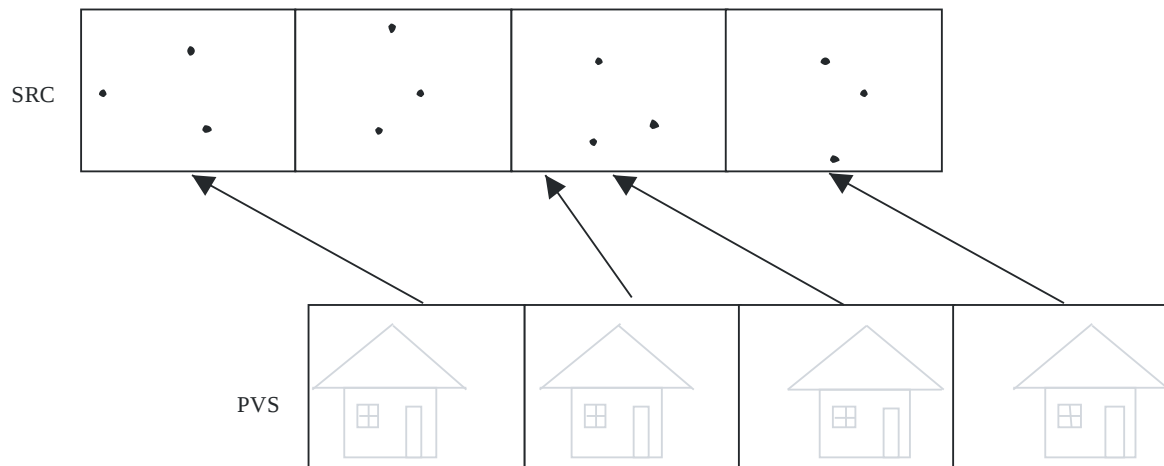


FIGURA 13

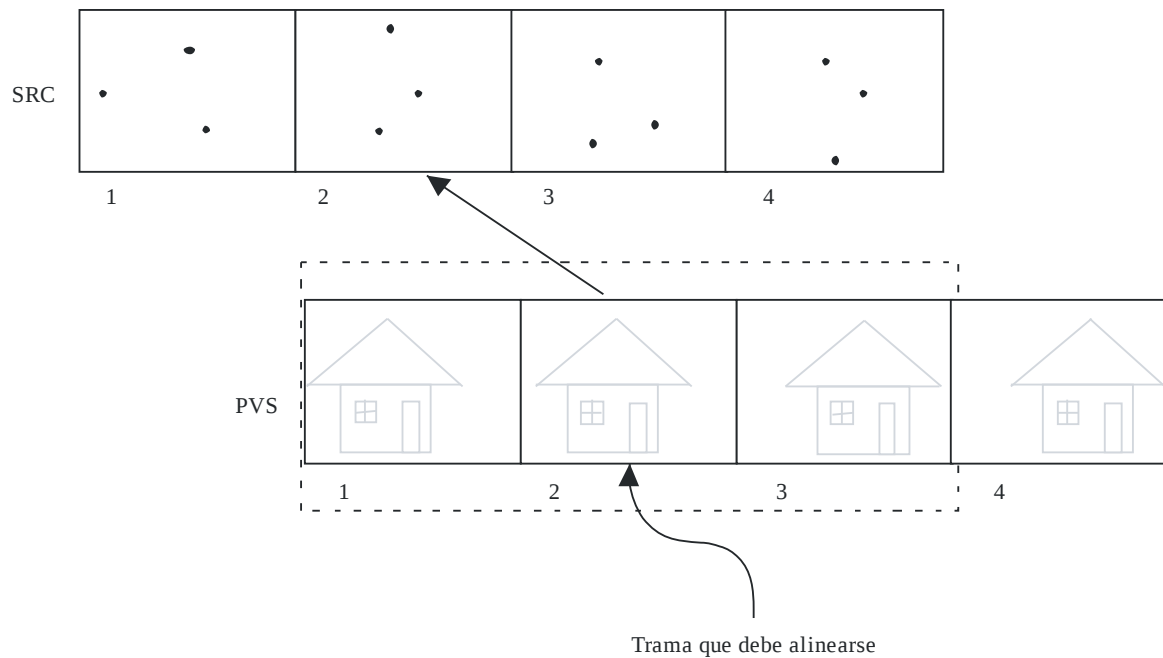
**Alineamiento de la secuencia de vídeo procesada con los píxeles de borde de la secuencia de vídeo fuente**



BT.1908-13

FIGURA 14

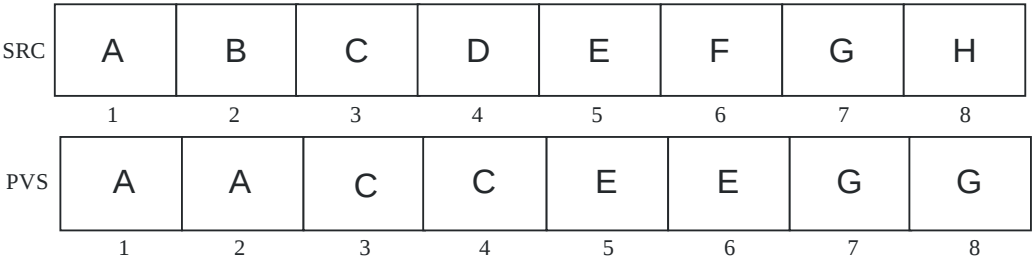
**Alineamiento de la secuencia de vídeo procesada con los píxeles de borde utilizando una ventana**



BT.1908-14

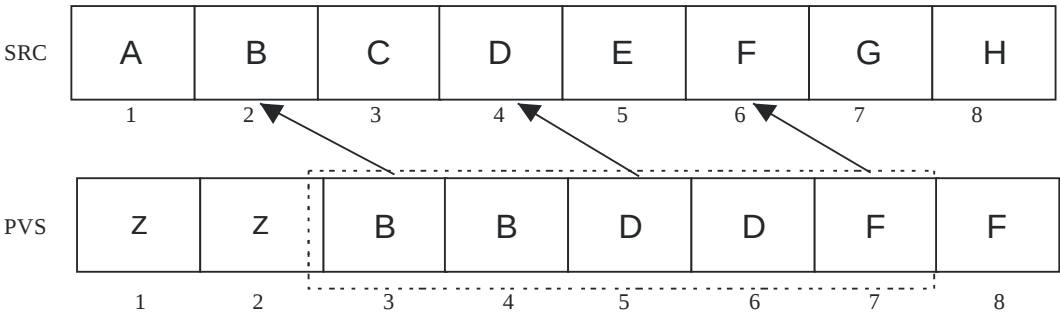
Cuando se codifica la secuencia de vídeo fuente con altas relaciones de compresión, el codificador puede reducir el número de tramas por segundo y la secuencia de vídeo procesada tiene tramas repetidas (véase la Fig. 15). En la Fig. 15 la secuencia de vídeo procesada no tiene tramas correspondientes a algunas tramas de la secuencia de vídeo fuente (tramas 2ª, 4ª, 6ª y 8ª). 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. 16 sólo tres tramas (3ª, 5ª y 7ª) en la ventana se utilizan para el registro temporal.

FIGURA 15  
Ejemplo de tramas repetidas



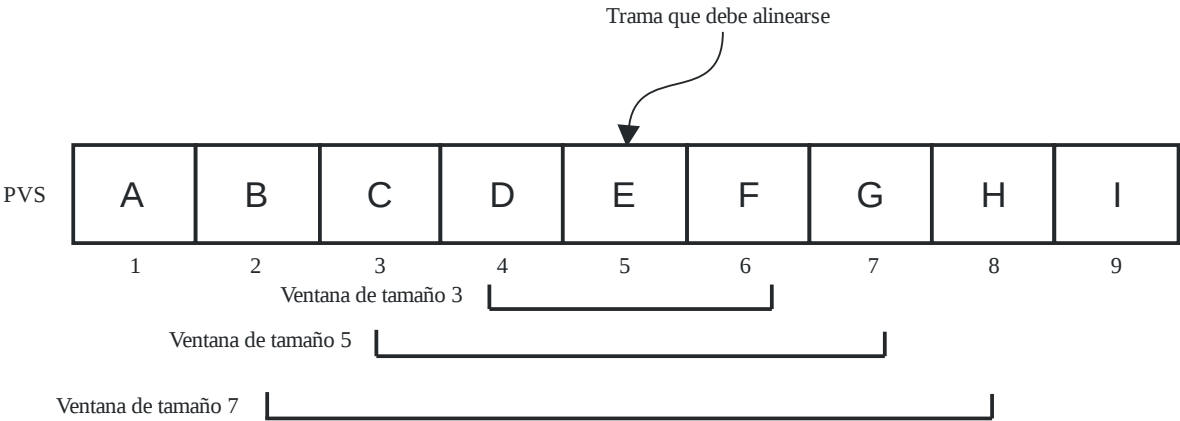
BT.1908-15

FIGURA 16  
Manejo de tramas repetidas



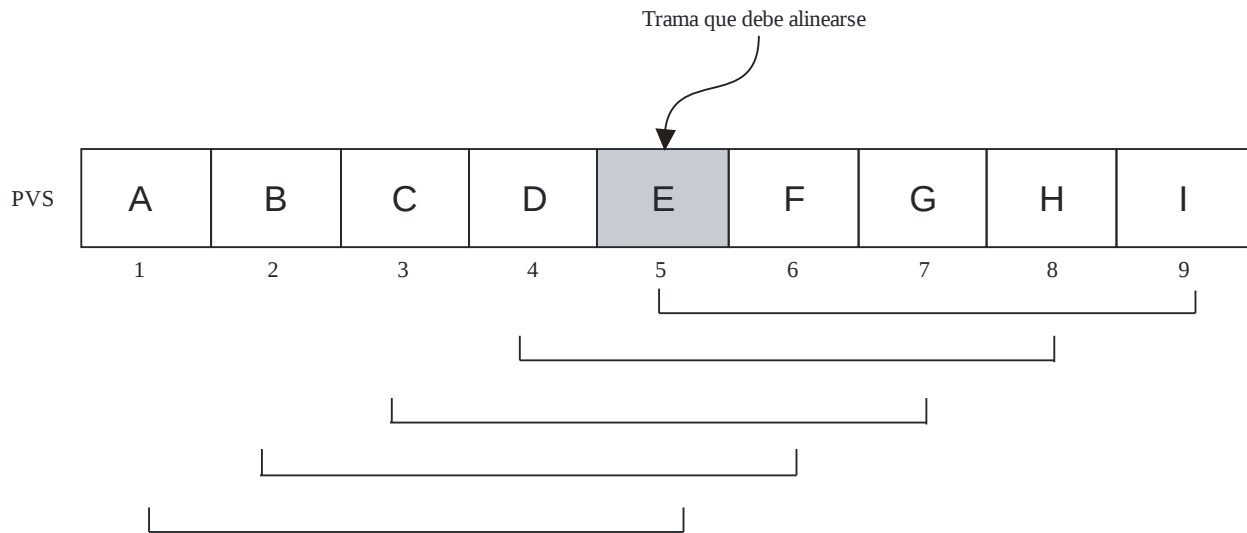
BT.1908-16

FIGURA 17  
Ventanas con varios tamaños



BT.1908-17

FIGURA 18  
Centros de ventana



BT.1908-18

#### 6.2.4 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.

Puesto existen factores que pueden reducir la calidad del vídeo, se ajusta el valor de la EPSNR para tener en cuenta esos efectos que se explican a continuación.

##### 1) Medición de bloqueo I

Para determinar los efectos de bloqueo se calculan 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:

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

Por último, se utilizan las siguientes ecuaciones:

IF(BLOCKING > 12 and  $25 \leq EPSNR < 30$ )    adjust\_EPSNR\_blk1=3

IF(BLOCKING > 5 and  $30 \leq EPSNR < 35$ )    adjust\_EPSNR\_blk1=5

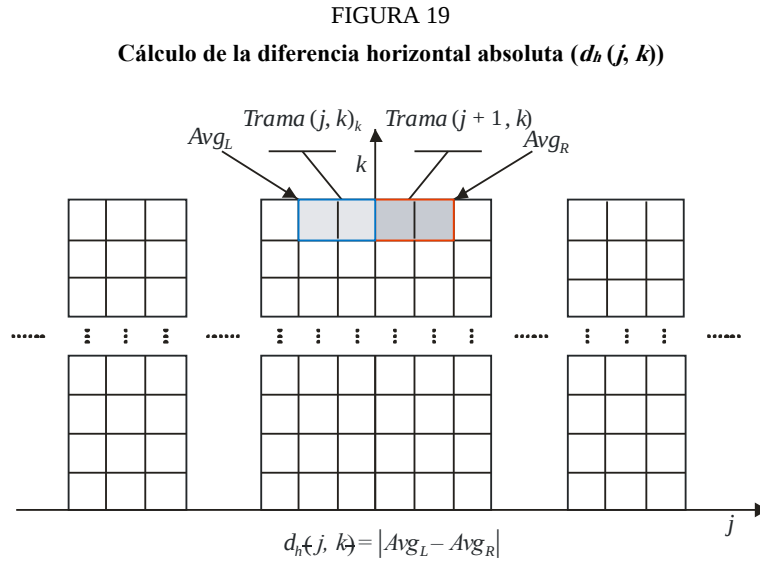
## 2) Medición de bloqueo II

En el caso de que se produzca un deterioro por bloqueo en cada columna 8-ésima (por ejemplo en MPEG2), también se utiliza otra medición de bloqueo. Para esta medición de bloqueo, se calcula en primer lugar la diferencia horizontal absoluta de la forma siguiente (Véase la Fig. 19):

$$d_h(j, k) = |Avg_L - Avg_R|$$

donde:

$$Avg_L = \frac{1}{2} \sum_{p=-1}^0 Trama(j+p, k), \quad Avg_R = \frac{1}{2} \sum_{p=1}^2 Trama(j+p, k)$$



BT.1908-19

A continuación se define la suma de los bloqueos horizontales ( $SB_h$ ) en la posición  $j$  de la forma siguiente:

$$SB_h[j] = \left( \sum_{1 \leq k \leq altura} (|Trama(j, k) - Trama(j+1, k)| \times u(d_h(j, k) - \Phi(Avg_L))) \right)^2$$

donde  $u(\cdot)$  representa la función de paso unitario y:

$$\Phi(s) = \begin{cases} 17(1 - \sqrt{s/127}) + 3 & \text{si } s \leq 127 \\ 3(s - 127)/128 + 3 & \text{en otro caso} \end{cases}$$

Tras repetir el procedimiento para todas las tramas, se calcula el bloqueo horizontal de trama ( $FB_h$ ) de la forma siguiente:

$$FB_h = \left( \sum_{\substack{1 \leq j \leq anchura \\ j \equiv 0 \pmod{8}}} SB_h(j) \right)^{1/2}$$

Para cada trama, se calcula la diferencia de columnas ( $NFB_h$ ) excluyendo cada columna 8-ésima de la forma siguiente:

$$NFB_h = \frac{1}{7} \sum_{l=1}^7 \left( \sum_{\substack{1 \leq j \leq anchura \\ j \equiv l \pmod{8}}} \left( \sum_{1 \leq k \leq altura} (Trama(j, k) - Trama(j+1, k)) \times u(d_h(j, k) - \Phi(Avg_L)) \right)^2 \right)^{1/2}$$

A continuación, se calcula la característica de bloqueo horizontal final,  $BLK_H$ , mediante la ecuación:

$$BLK_H = \ln(FB_h / NFB_h)$$

La característica de bloqueo vertical  $BLK_V$  se calcula de forma similar. Para secuencias de vídeo entrelazadas, la característica de bloqueo vertical se calcula en la secuencia de campo. El grado de bloqueo para la trama  $i$ -ésima se calcula como sigue:

$$TramaBLK(i) = 0,5 \times BLK_H + 0,5 \times BLK_V$$

El grado de bloqueo final ( $BLOCKING2$ ) se obtiene promediando el 10% superior de los grados de bloqueo de trama.

Finalmente, se utilizan las ecuaciones siguientes:

|                                       |                     |
|---------------------------------------|---------------------|
| IF(BLOCKING2 > 1.5 and 25 ≤ EPSNR<30) | adjust_EPSNR_blk2=2 |
| IF(BLOCKING2 > 1.3 and 30 ≤ EPSNR<35) | adjust_EPSNR_blk2=2 |
| IF(BLOCKING2 > 1.5 and 35 ≤ EPSNR<40) | adjust_EPSNR_blk2=2 |
| IF(BLOCKING2 > 1 and 40 ≤ EPSNR<45)   | adjust_EPSNR_blk2=2 |
| IF(BLOCKING2 > 0.5 and 45 ≤ EPSNR<55) | adjust_EPSNR_blk2=2 |

Como se puede observar en las ecuaciones anteriores, este ajuste tiene poco efecto en el valor final de EPSNR. Si no se producen los acontecimientos de bloqueo en cada columna 8-ésima, se puede renunciar a este ajuste o buscar en primer lugar las ubicaciones del bloqueo. También es posible utilizar una función diferente para  $\Phi(s)$ .

### 3) Tramas congeladas máximas y tramas congeladas totales

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

|                                        |                           |
|----------------------------------------|---------------------------|
| IF(MAX_FREEZE ≥ 8 and 25 ≤ EPSNR<30)   | adjust_EPSNR_max_freeze=3 |
| IF(MAX_FREEZE ≥ 6 and 30 ≤ EPSNR<35)   | adjust_EPSNR_max_freeze=3 |
| IF(MAX_FREEZE ≥ 3 and 35 ≤ EPSNR<40)   | adjust_EPSNR_max_freeze=3 |
| IF(MAX_FREEZE ≥ 1.5 and 40 ≤ EPSNR<45) | adjust_EPSNR_max_freeze=2 |
| IF(MAX_FREEZE ≥ 1 and 45 ≤ EPSNR<95)   | adjust_EPSNR_max_freeze=2 |

donde MAX\_FREEZE es la mayor duración de las tramas congeladas. Cabe señalar que si la secuencia de vídeo no es de 10 s, deben utilizarse umbrales distintos.

Así mismo, las tramas congeladas totales se consideran de la forma siguiente:

|                                         |                               |
|-----------------------------------------|-------------------------------|
| IF(TOTAL_FREEZE ≥ 80 and 25 ≤ EPSNR<30) | adjust_EPSNR_total_freeze=3   |
| IF(TOTAL_FREEZE ≥ 40 and 30 ≤ EPSNR<35) | adjust_EPSNR_total_freeze=4   |
| IF(TOTAL_FREEZE ≥ 10 and 35 ≤ EPSNR<40) | adjust_EPSNR_total_freeze=3.5 |
| IF(TOTAL_FREEZE ≥ 2 and EPSNR ≥ 40)     | adjust_EPSNR_total_freeze=1.5 |

donde TOTAL\_FREEZE es la duración total de las tramas congeladas. Cabe señalar que si la secuencia de vídeo no es de 10s, deben utilizarse umbrales distintos.

#### 4) Bloque de errores de transmisión

Se pueden producir bloques congelados locales debido a errores de transmisión. En escenas estáticas, algunos bloques también son idénticos a los bloques de tramas anteriores en la misma posición. Para considerar los bloques congelados locales debidos a errores de transmisión, se clasifican los bloques que contienen los píxeles de borde transmitidos ya sea como bloques idénticos (es decir, los bloques son idénticos a los bloques de las tramas anteriores) o en bloques diferentes. A continuación, se calculan dos EPSNR para los bloques idénticos y los bloques diferentes. Si la diferencia de los dos EPSNR (EPSNR\_diff) es grande, significa que se pueden producir errores de transmisión. Por ello, la EPSNR se ajusta de la forma siguiente:

|                                                                            |                      |
|----------------------------------------------------------------------------|----------------------|
| IF( $8 \leq \text{EPSNR\_diff} \leq 30$ and $25 \leq \text{EPSNR} < 30$ )  | adjust_EPSNR_diff= 3 |
| IF( $9 \leq \text{EPSNR\_diff} \leq 30$ and $30 \leq \text{EPSNR} < 35$ )  | adjust_EPSNR_diff= 4 |
| IF( $10 \leq \text{EPSNR\_diff} \leq 30$ and $35 \leq \text{EPSNR} < 40$ ) | adjust_EPSNR_diff= 6 |
| IF( $9 \leq \text{EPSNR\_diff} < 10$ and $35 \leq \text{EPSNR} < 40$ )     | adjust_EPSNR_diff= 2 |
| IF( $9 \leq \text{EPSNR\_diff} \leq 30$ and $40 \leq \text{EPSNR} < 45$ )  | adjust_EPSNR_diff= 4 |

Sin embargo, si el número total de bloques idénticos es menor que 100, no se realiza ningún ajuste.

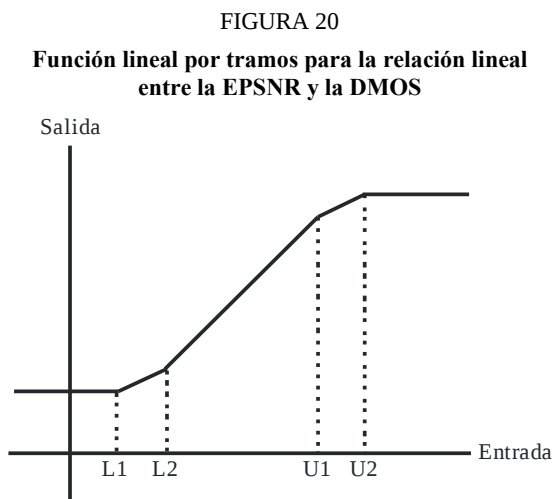
#### 5) Ajuste final del EPSNR

Finalmente, el valor de la EPSNR se ajusta de la forma siguiente:

$$\text{EPSNR} \leq \text{EPSNR} - \text{MAX}(\text{adjust\_EPSNR\_blk1}, \text{adjust\_EPSNR\_blk2}, \text{adjust\_EPSNR\_max\_freeze}, \text{adjust\_EPSNR\_total\_freeze}, \text{adjust\_EPSNR\_diff})$$

#### 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. 20. En el modelo probado en la prueba VQEG TVAD, el límite superior se fija a 50 y el límite inferior a 19.



## Apéndice 1 (Informativo)

### Conclusiones del Grupo de Expertos en calidad del vídeo

Los estudios de las mediciones de la calidad de vídeo percibida 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. La prueba de fase I de televisión de alta definición recientemente completada por el VQGE evaluó el comportamiento de los algoritmos de medición de la calidad de vídeo percibida de referencia propuestos.

Las estadísticas siguientes proceden del informe final del VQGE sobre TVAD (Informe del VQGE). Cabe señalar que el texto del informe del VQGE incluye otros métodos de medición entre ellos la correlación Pearson y el RMSE calculados en experimentos individuales, intervalos de confianza, pruebas de importancia estadística sobre experimentos individuales, análisis de los subconjuntos de datos que incluyen degradaciones concretas (por ejemplo, solamente para la codificación de la Recomendación UIT-T H.264), diagramas de dispersión y coeficientes para los que los valores son válidos.

#### Análisis primario

Los resultados del modelo RR se resumen en el Cuadro 4. La PNSR se calcula según lo establecido en la Recomendación UIT-T J.340 y se incluye en este análisis con fines comparativos. «RMSE del superconjunto» identifica la medición primaria (RMSE) realizada en el superconjunto combinado (es decir, se incluyen los seis experimentos en una única escala). «Total de grupos de mejor comportamiento» identifica el número de experimentos (de 0 a 6) en los que este modelo fue el modelo de mejor comportamiento o fue equivalente estadísticamente al modelo con mejores prestaciones. «Mejor que la PNSR total» identifica el número de experimentos (de 0 a 6) para los que este modelo fue estadísticamente mejor que el de la PNSR. «Mejor que la PNSR del superconjunto» indica si cada modelo es estadísticamente mejor que la PNSR en el superconjunto combinado. «Correlación del superconjunto» identifica la correlación de Pearson calculada en el superconjunto combinado.

CUADRO 4

| Métrica                                 | PSNR | Yonsei56k | Yonsei128k | Yonsei256k |
|-----------------------------------------|------|-----------|------------|------------|
| RMSE del superconjunto                  | 0,71 | 0,73      | 0,73       | 0,73       |
| Total de grupos de mejor comportamiento | 6    | 4         | 4          | 4          |
| Equivalente a o mejor que el PSNR total | 6    | 4         | 4          | 4          |
| Equivalente a la PSNR del superconjunto | Sí   | Sí        | Sí         | Sí         |
| Correlación del superconjunto           | 0,78 | 0,77      | 0,77       | 0,77       |

Puesto que las características de este modelo son estadísticamente idénticas para las tres anchuras de banda, se recomienda utilizarlo con por lo menos una anchura de banda de canal lateral de 56 kbit/s.

### Análisis secundario

El Cuadro 5 indica el RMSE del modelo RR, para subdivisiones del superconjunto. Estas subdivisiones fraccionan los datos por tipos de codificación (UIT-T H.264 o MPEG-2) y por la presencia de errores de transmisión (Errors) o para indicar cuándo incluye el HRC sólo objetos de codificación (Coding). Puesto que los experimentos no estaban diseñados para que estas variables se distribuyeran en la gama total de calidad, sólo se han presentado el RMSE para esas subdivisiones.

CUADRO 5

| Tipo de HRC         | PSNR | Yonsei56k | Yonsei128k | Yonsei256k |
|---------------------|------|-----------|------------|------------|
| Codificación H.264  | 0,75 | 0,65      | 0,65       | 0,65       |
| Error H.264         | 0,67 | 0,86      | 0,85       | 0,86       |
| Codificación MPEG-2 | 0,78 | 0,81      | 0,81       | 0,80       |
| Error MPEG-2        | 0,66 | 0,68      | 0,68       | 0,68       |
| Codificación        | 0,75 | 0,69      | 0,69       | 0,69       |
| Error               | 0,67 | 0,79      | 0,78       | 0,79       |

### Bibliografía

- Recomendación UIT-R BT.500, *Metodología para la evaluación subjetiva de la calidad de imágenes de televisión*.
- Recomendación UIT-T P.910 (2008), *Métodos de evaluación subjetiva de la calidad de vídeo para aplicaciones multimedios*.
- Recomendación UIT-T P.911 (1998), *Métodos de evaluación subjetiva de la calidad audiovisual para aplicaciones multimedios*.
- 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.340 (2010), *Algoritmo de referencia para calcular la relación señal/ruido de cresta (PNSR) de una secuencia vídeo procesada con desplazamientos espaciales constantes y un retardo constante*.